

Юрий Иванович Рыжиков

Теория очередей и управление запасами

Серия «Учебники для вузов»

Главный редактор
Заведующий редакцией
Выпускающий редактор
Верстка
Корректор

*В. Усманов
Л. Волкова
В. Земских
Ю. Рыжиков
Е. Пряникова*

ББК 22.183.5я7 УДК 519.8(075)

Рыжиков Ю. И.

Р 93 Теория очередей и управление запасами. — СПб: Питер, 2001. — 384 с.: ил. — (Серия «Учебники для вузов»).

ISBN 5-318-00073-8

В книге известного ученого, одного из тех, кто стоял у истоков отечественной школы теории запасов, Ю. И. Рыжикова, представлены методы массового обслуживания и управления запасами. В новой работе автору удалось не только системно и последовательно изложить базовые теоретические положения теории очередей, но и дать ценные практические рекомендации по оптимизации сложных систем снабжения, снижению времени обслуживания, ликвидации избыточных запасов.

Мы надеемся, что книга будет полезна не только преподавателям, аспирантам, студентам экономических вузов, изучающих логистику, маркетинг, менеджмент, но и практикам-менеджерам, организаторам производства, специалистам по логистике.

© Рыжиков Ю. И., 2001

© Серия, оформление, Издательский дом «Питер», 2001

Все права защищены. Никакая часть данной книги не может быть воспроизведена в какой бы то ни было форме без письменного разрешения владельцев авторских прав.

ISBN 5-318-00073-8

ЗАО «Питер Бук», 196105, Санкт-Петербург, Благодатная ул., д. 67.

Лицензия ИД № 01940 от 05.06.00.

Налоговая льгота — общероссийский классификатор продукции ОК 005-93,
том 2; 95 3000 — книги и брошюры.

Подписано в печать 07.12.2000. Формат 60 × 90/16. Усл. п. л. 24. Тираж 5000.
Заказ № 1334.

Отпечатано с готовых диапозитивов в АОТ «Типография „Правда“».
191119, С.-Петербург. Социалистическая ул., 14

Оглавление

Предисловие	9
I Элементы модели управления запасами	13
1 Постановка задачи	15
1.1. Проблемы с запасами	15
1.2. Модель управления запасами и ее элементы	18
1.3. Из истории теории запасов	22
1.4. Классификация моделей УЗ	30
1.5. Выбор варианта модели	39
2 Экономические параметры модели	49
2.1. Экономика снабжения	49
2.2. Транспортные расходы	51
2.3. Расчет стоимости хранения	52
2.4. Стоимость поставки	58
2.5. Определение величины штрафа	60
3 Элементы теории вероятностей и очередей	64
3.1. Основные понятия теории распределений	64
3.1.1. Определения	64
3.1.2. Числовые характеристики	65
3.1.3. Показательное распределение	66
3.2. Аппроксимация распределений	67
3.3. Типовые распределения	69
3.3.1. Распределения фазового типа	69
3.3.2. Гамма-распределение	71

3.3.3. Нормальное распределение	73
3.3.4. Распределение Вейбулла	74
3.3.5. Дискретные распределения	75
3.4. Преобразования распределений	76
3.5. Задачи теории очередей	78
3.6. Поток заявок	79
3.6.1. Основные определения	79
3.6.2. Число событий на фиксированном интервале	80
3.6.3. Число событий на случайном интервале	81
3.6.4. Случайное прореживание потоков	83
3.6.5. Регулярное прореживание потоков	84
3.6.6. Суммирование потоков	84
3.7. Процесс обслуживания	85
3.8. Организация и продвижение очереди	86
3.9. Классификация и обозначение систем обслуживания	87
3.10. Показатели эффективности	87
3.11. Законы сохранения в теории массового обслуживания	89
3.12. Расчет марковских систем по законам сохранения	93
3.13. Система $M/G/1$	94
3.14. Многоканальные системы с фазовыми распределениями	95
3.14.1. Многофазное представление сложных СМО	95
3.14.2. Уравнения глобального баланса	98
3.14.3. Итерационный метод	99
3.14.4. Метод матрично-геометрической прогрессии	102
3.15. Сети обслуживания	104
3.15.1. Определения и допущения	105
3.15.2. Разомкнутая сеть с однородным потоком	108
3.15.3. Расчет сети с неоднородными потоками	109
3.15.4. Замкнутая сеть	109
3.16. Распределение времени пребывания в сети	111
3.16.1. Решение в средних	111
3.16.2. Преобразование Лапласа и высшие моменты	111
3.17. Пакет МОСТ по расчету систем с очередями	113
3.17.1. История ППП для расчета систем обслуживания	115
3.17.2. «Персональный» МОСТ	115
3.17.3. Профессиональная версия МОСТа	115
3.17.4. Общая характеристика МОСТа-F1	117

4	Прогнозирование данных	120
4.1.	Проблемы прогнозирования	120
4.2.	Характеристики временных рядов	121
4.3.	Метод скользящей средней	123
4.4.	Экспоненциальное сглаживание	123
4.5.	Аналитическое выравнивание	124
4.6.	Определение качества прогноза	125
4.7.	Корреляция	126

II Базовые модели 127

5	Однопродуктовые модели	129
5.1.	Детерминированный спрос	129
5.1.1.	Общий случай	129
5.1.2.	Предельные варианты	132
5.1.3.	Ошибки в параметрах и функция затрат	133
5.1.4.	Зависимые параметры	135
5.2.	Хранение продукта при естественной убыли	139
5.3.	Стареющий продукт	143
5.4.	Объем заказа с учетом инфляции	143
5.5.	Детерминированный нестационарный спрос	146
5.5.1.	Постановка задачи	146
5.5.2.	Динамическое программирование	147
5.6.	Случайный спрос, периодическая стратегия	150
5.6.1.	Непрерывный спрос	150
5.6.2.	Учет погрешностей в параметрах	155
5.6.3.	Дискретный спрос	157
5.6.4.	«Импульсные» расходы	158
5.6.5.	Штраф по вероятности дефицита	160
5.6.6.	Штраф по времени дефицита	160
5.6.7.	Многопериодные обобщения	163
5.7.	Неизвестная интенсивность спроса	163
5.8.	Расчет пороговых стратегий	166
5.8.1.	Модель без переноса дефицита	167
5.8.2.	Модель с переносом дефицита	168
5.8.3.	Показательно распределенный спрос	169
5.8.4.	Гамма-распределенная задержка	169
5.9.	Комбинированные стратегии	170

6	Многономенклатурные задачи	171
6.1.	Классификация номенклатур	171
6.2.	Детерминированный спрос	174
6.3.	Взаимозаменяемые продукты	185
6.4.	Планирование запасов при ограничениях	187
6.4.1.	Метод множителей Лагранжа	187
6.4.2.	Многономенклатурные закупки	188
6.5.	Периодические поставки, вероятностный спрос	189
6.5.1.	Заданная периодичность	189
6.5.2.	Выбираемая периодичность	192
6.6.	Зонная стратегия	193
6.6.1.	Показательно распределенный спрос	195
6.6.2.	Дискретно распределенный спрос	195

7	Системы снабжения	198
7.1.	Общие положения	198
7.2.	Факты и тенденции	200
7.3.	Детерминированный спрос	204
7.4.	Синтез системы при детерминированном спросе	210
7.5.	Децентрализованная система при вероятностном спросе	215
7.5.1.	Однородный запас	218
7.5.2.	Многономенклатурный запас	223
7.6.	Линейные многокаскадные системы снабжения	227
7.7.	АСУ снабжением. опыт и проблемы	231

III Проблема запасных частей 235

8	Снабжение запасными частями	237
8.1.	Специфика проблемы	237
8.2.	Системы обеспечения запчастями	241
8.2.1.	Структура	241
8.2.2.	Организация поставок	243
8.3.	Военные системы снабжения	246
8.4.	Экономические параметры	249
8.5.	Спрос на запчасти	253
8.5.1.	Общие соображения	253
8.5.2.	Распределение спроса для резервированных систем	256

8.6.	Расчет распределения времени ремонта	260	10	«Метрический» анализ систем	306
8.6.1.	О распределении длительности диагностики	260	10.1.	Специфика спроса в иерархии	307
8.6.2.	Учет организации рабочего цикла	262	10.1.1.	Особенности генерации заявок	307
8.7.	Политика заказов и процедуры управления	266	10.1.2.	Составной пуассоновский поток	308
8.8.	Параметры моделей	267	10.2.	Показатели эффективности	311
8.9.	Проблема штрафов для военных систем	268	10.3.	Задержки	313
8.9.1.	Авиация	269	10.4.	Основы METRIC	314
8.9.2.	Ракетные комплексы	270	10.5.	Базовая METRIC	315
9	Локальная задача о восстанавливаемом ЗИПе	275	10.6.	MOD-METRIC	315
9.1.	Предварительные соображения	275	10.7.	METRIC реального времени	318
9.2.	Основная модель	278	10.8.	Двухэшелонная система с экстренными поставками	320
9.3.	Многолинейная марковская система	282	10.9.	Трехуровневая METRIC	321
9.4.	Учет дополнительных задержек	283	10.10.	Децентрализованная система	323
9.5.	Восстановление в сети	283	10.11.	Стратегии, отличные от $(S - 1, S)$	325
9.6.	Минимизация затрат за время эксплуатации	284	10.11.1.	Спрос в верхнем эшелоне	325
9.7.	Восстановление с порогом включения и разогревом	284	10.11.2.	Нормальная аппроксимация	326
9.7.1.	Вложенная цепь Маркова	285	10.12.	Стратегия с непрерывным просмотром	327
9.7.2.	Переход к стационарным вероятностям	286	11	Оптимизация систем обеспечения запчастями	330
9.7.3.	Распределение времени пребывания	288	11.1.	Задача о рюкзаке	330
9.8.	Замкнутая система восстановления	289	11.1.1.	Постановка задачи	330
9.8.1.	Расчет вложенной цепи Маркова	290	11.1.2.	Основная расчетная схема	335
9.8.2.	Расчет стационарных вероятностей	290	11.1.3.	Ограничения и обобщения метода	336
9.8.3.	Вероятности перехода	291	11.2.	«Рюкзак» и двухуровневая система	338
9.9.	Учет погрешности исходных данных	294	11.3.	Лагранжевы методы для систем снабжения	339
9.9.1.	Однолинейная система со штрафом по вероятности дефицита	297	11.3.1.	Постановка задачи для двухуровневой системы	339
9.9.2.	Однолинейная система со штрафом по ожидаемому дефициту	297	11.3.2.	Процедура Фокса и Лэнди	340
9.9.3.	Многоканальные марковские системы	298	11.3.3.	Половинное деление	342
9.9.4.	Пример	298	11.3.4.	Метод Мукштадта	342
9.9.5.	Интервальная оценка частных параметров	299	11.4.	Сатисфакционный подход	345
9.10.	Многономенклатурная задача	301	Закключение		348
9.10.1.	Общая цель	301	Литература		350
9.10.2.	Общее восстановление	303	Приложения		369
9.11.	Частично восстанавливаемый ЗИП	304	А Расчет системы кратных периодов		370
			Б Расчет зонной стратегии		375

Предисловие

Необходимость ускорения темпов прироста совокупного общественного продукта требует преодоления ряда объективных трудностей, отражающих особенности современного уровня производства. Постепенно возрождаются прежние масштабы экономической системы, продолжающаяся специализация усложняет внутренние связи. Вместе с ростом масштаба производства и повышением производительной силы труда посредством кооперации, разделения труда, применения машин и т.д. увеличивается и масса сырья и вспомогательных материалов, входящих в ежедневный процесс воспроизводства.

В современных машиностроительных и приборостроительных предприятиях, выпускающих массовую или крупносерийную продукцию, синхронизация требуется не только при сборочно-транспортных, но и при всех остальных операциях, включая работу заготовительных цехов. Общеизвестно, что календарные планы производства и закупки начинают корректироваться, как только они составлены. Задержка с одной деталью ломает весь график сборки. Оснастка не оправдывает ожиданий. Изменения технологии, трудность освоения новой продукции, брак, переделки, прогулы, забастовки — все это будни промышленности. Равномерное обеспечение потребителей дает как прямой экономический эффект (снижение затрат на дополнительно привлекаемую в пиках нагрузки рабочую силу, в частности на оплату сверхурочных, и возможность избавиться от лишних фондов оборудования, простаивающего вне пиков), так и косвенный — через улучшение условий труда и качества продукции.

Неизбежная рассогласованность ритма производства поставщиков и потребителей, дискретность процесса поставок, возможность случайных колебаний в интенсивности потребления или длительности интервалов между поставками относительно среднего (расчетного)

уровня вынуждают создавать в системах снабжения *запасы*. Хранение запасов во многих случаях¹ обходится дешевле, чем любой другой способ обеспечения ритмичного производства: резервные производственные мощности (их лучше использовать для освоения новой продукции), трудовые ресурсы, экстренные поставки и т.п. В состав запасов могут входить сырье, материалы, полуфабрикаты, комплектующие изделия, готовая продукция, отходы производства. Отметим, что капитал тоже может рассматриваться как запас, причем цена его хранения определяется темпом инфляции.

Нельзя не согласиться с Ю.А. Беляевым [4] в том, что «без запасов никто и ничто существовать не может: ни машина, ни человек, ни государство, ни вселенная. Без запаса прочности мост разрушится, а парашют порвется при малейшей перегрузке. Надежность технической системы создается запасом прочности конструкций, дублированием элементов, приданием запчастей. Портфель заказов редакции обеспечивает равномерную загрузку сотрудников и работу типографии. Запас мудрости руководителей позволяет им предвидеть будущие заботы и разглядеть пророков в своем отечестве; оценить неординарные идеи и справедливую критику; понять пользу плюрализма мнений и гражданских свобод. Запасы скудоумия, инерции, профессиональной неграмотности могут привести к не менее поразительным результатам».

Колоссальный объем средств, вложенных в запасы, придает проблеме научного управления ими первостепенную важность. На рубеже 1990-х гг. суммарная стоимость запасов в СССР превышала 450 млрд. рублей в тогдашнем масштабе цен. В США 90% капитала фирм вложены в запасы. В 1976 г. товарные и производственные запасы США составляли 276 млрд. долл. (17% валового национального дохода) [200]. В эту статистику не включены огромные запасы органов управления — местных, штатов и правительства, в том числе военные, которые в 1969 г. оценивались в 50 млрд. долл. [89].

Избыточные запасы были причиной многих неудач в бизнесе, оказывали дестабилизирующее влияние при кризисах. Излишние запасы являются тормозом на пути научно-технического прогресса. Переход на новый вид продукции обычно сдерживается требованием выработки всех имеемых запасов или дорогостоящим списанием их. Все эти проблемы обострились в связи с ускорением научно-технического прогресса, ди-

¹Существуют задачи (в электроэнергетике), где создание буферных запасов невозможно, а в газовой промышленности и гидроэнергетике — сложно. Здесь для покрытия меняющихся потребностей используется резерв мощностей.

версификацией спроса и индивидуализацией производства, сокращением сроков морального старения техники. Возросли сложность решаемых задач и цена ошибочных решений.

В таких случаях полезно вспомнить, что «нет ничего лучше хорошей теории» (Л. Больцман). Принятие решений должно быть основано на *модели* реальной ситуации. Совокупность правил построения моделей и работы с ними образует соответствующую прикладную теорию. Общеизвестно (см. [124]), что из всех специализированных ветвей исследования операций *управление запасами* (УЗ), с одной стороны, имеет наибольшие возможности для практического применения и с другой — наиболее развитую теорию. Сегодня по УЗ ежегодно только на английском языке публикуется свыше 100 журнальных статей, монографии, учебники, труды конференций. Обсуждение проблем часто идет в более широком контексте (производство, логистика, транспортные аспекты). Разумеется, существует отставание практики от теории (и наоборот), но применение данной теории уже в течение 80 лет возможно и оправдано. Согласно [190], типичная компания может удвоить объем продаж при сохранении объема запаса или сохранить этот объем, уменьшив запасы вдвое. В главе 1 будут приведены многие другие примеры эффективного использования рекомендаций теории управления запасами (ТУЗ) в отечественной и зарубежной практике, а также обзор ее развития.

Автор стоял у истоков отечественной школы теории запасов [56] и надеется, что содержимое предлагаемой книги также окажется полезным широкому кругу преподавателей, студентов и организаторов производства. Ее отличием от [56] является более детальная проработка экономических аспектов и специфики систем снабжения военного назначения, а также обстоятельное рассмотрение проблемы восстанавливаемого ЗИПа с применением современных методов теории массового обслуживания (она же — теория очередей). Как и авторы [190],

«мы полагаем, что достигли разумного баланса между обзором теоретических проблем и методов и практическими подходами, имеющими теоретическое звучание. В книге сочетаются теория, практические советы (когда она не работает) и философия — для перспективы. Именно:

- важность запасов для принятия решений на высших уровнях;
- взаимосвязь между агрегированными и индивидуальными решениями;

- недостоверность спроса на отдельные изделия и учет ожидаемого риска для страховых запасов;
- соображения реализации, включая компьютерные программы».

Добавим к этому, что использование сложных алгоритмов может оказаться не оправданным из-за погрешностей построения модели, погрешностей исходных данных к ней и малого относительного выигрыша при переходе к усложненной модели.

Автор приносит извинения за некоторую устарелость отечественных экономических данных — их не публикуют уже много лет, а журнал «Материально-техническое снабжение» «почил в бозе».

Разумеется, существующая теория не решает всех проблем, в особенности применительно к оптимизации сложных систем снабжения. Здесь всегда остается поле для индивидуального подхода к специфике задачи. Представляется, однако, что набор *частных* моделей является достаточно полным. Возникающие серьезные проблемы следует рассматривать как вызов и приглашение к самостоятельному развитию теории. Современному менеджеру, часто сбежавшему в бизнес из точных наук (пример — член-корреспондент РАН Б.А. Березовский), это вполне по силам.

Рассмотренные в книге математические модели зачастую заимствовались из первоначально разобщенных областей исследования со сложившейся в каждой из них системой обозначений. Чтобы облегчить работу с первоисточниками, пришлось отказаться от полной унификации обозначений и, соответственно, от словаря таковых. Сами обозначения были по возможности упрощены, в частности, убраны индексы и значки, поясняющие размерность экономических параметров (цена хранения, «штраф», суммарные затраты). Эта размерность либо поясняется при построении очередной модели, либо легко устанавливается по контексту или анализом структуры целевой функции.

Компьютерный набор и верстка книги выполнены автором в издательской системе \LaTeX с применением кириллических шрифтов высокого разрешения семейства LH.

Часть I

Элементы модели управления запасами

Глава 1

Постановка задачи

1.1. Проблемы с запасами

Материальные запасы подразделяются на товарные и производственные. Товарные запасы — это готовая продукция у поставщиков (сбытовая), на складах и базах (складская). Производственными считаются запасы, уже находящиеся у потребителей, но еще не вступившие в процесс переработки.

С другой точки зрения запасы можно классифицировать на расходные и резервные. Расходные предназначены для обеспечения потребностей производства при несовпадении сроков и размеров поступления и потребления ресурсов. Резервные должны удовлетворять потребности к пику продаж, страховать от забастовок, задержек и срыва поставок, учитывать возможные чрезвычайные обстоятельства (мобилизационный запас). Иногда дополнительно выделяют запас, находящийся в пути между звеньями системы или фазами производства.

Основными факторами образования запасов готовой продукции являются задержки, необходимые для:

- накопления готовой продукции до размеров отгружаемых партий;
- укомплектования партий поставки, если потребителю одновременно отгружается несколько номенклатур;
- упаковки и затаривания продукции;
- оформления отгрузочной документации;

- погрузки в транспортные средства

Необходимость надежного обеспечения спроса подкрепляется серьезными экономическими последствиями недопоставок, вынуждает увеличивать запасы, что приводит к временному исключению из оборота значительных материальных ценностей. При необоснованном сокращении запасов (например вследствие нарушения поставщиком сроков или объема поставок) возможна остановка производства. Применяемые к неаккуратному поставщику санкции лишь отчасти компенсируют убытки предприятия. Недостаток нужных населению товаров в розничной торговой сети возрождает забытые очереди, ставит на грань выживания обнищавшие группы населения и может привести к тяжелейшим социально-экономическим последствиям.

Всякого рода перестраховки тяжелым грузом ложатся на бюджет системы снабжения и включающих ее организаций. Многочисленные примеры такого рода по данным на середину 1960-х гг. приводились в монографии автора [56]. Тогда уровень запасов в торговле по многим позициям приближался к годовому объему продаж. В последующие годы ситуация менялась мало [10] (см. табл. 1.1 и 1.2).

Т а б л и ц а 1.1

Динамика запасов в торговле, млрд руб.

Сфера	1966	1969	1970	1971	1972
Оптовая торговля и промышленность	34	32	29	26	23
Розничная торговля	96	84	85	88	91
Всего	130	116	114	114	114

Авторы [10] объясняют рост запасов несоответствием спроса предложению и ростом доходов населения, становящегося более разборчивым. Снижение цен уменьшало запасы. Заметим, что норматив товарных запасов в 1966–1971 гг. сокращался трижды (с 84 до 76 дней). Хронический характер проблемы раздувания запасов иллюстрируется данными [4]: в конце 1980-х гг. в магазинах потребительской кооперации имелось лопат и киноаппаратов на 430 дней торговли, кос — на 353, а фотоаппаратов — на полгода. Неудивительно, что в 1977 г. размер оборотных средств в запасах товарно-материальных ценностей превышал 230 млрд руб., а к 1991 г. достиг почти 500 млрд руб. [30].

Т а б л и ц а 12

Торговые запасы, в днях оборота

Вид товара	1966	1970	1972
Хлопчатобумажные ткани	201	211	157
Шерстяные ткани	227	102	103
Шелковые ткани	298	175	142
Льняные ткани	216	131	108
Одежда и белье	130	114	139
Меховые изделия	73	46	52
Трикотаж верхний	109	67	91
Трикотаж нижний	-	171	176
Кожаная обувь	118	107	138
Резиновая обувь	147	113	97
Часы	154	90	77
Холодильники	11	17	21
Пылесосы	63	37	41
Стиральные машины	16	68	207
Электроутюги	138	23	33
Металлическая посуда	205	150	133
Стекло, фарфор, фаянс	209	98	66
Телевизоры	21	108	90
Фотоаппараты	268	142	135
Итого	145	128	132

Одной из основных причин этой ситуации было то, что учет и планирование велись в агрегированных показателях и оперативное управление осуществлялось на основе усредненных нормативов т е практически без обратной связи Совершенствовать *нормативы* бесполезно Комплексное объяснение торговых безобразий Ю А Беляев [4] находит в рукописном наследии Ф Э Дзержинского

- 1) «товар проходит слишком большое число звеньев между производством и потреблением,
- 2) слишком долго идет,
- 3) слишком много лиц питается из торговли, слишком дорого все это обходится»

Не намного лучше обстояло дело и в промышленности. Относительно свежий пример [20]: на зеркально-фурнитурном комбинате объединения «Севзапмебель» на 1 июня 1980 г. образовалось сверхнормативных остатков сырья на 750 тыс. руб. Из-за недостаточности складских площадей, нарушений условий хранения было перерасходовано оконного и полированного стекла на 537 тыс. руб.

По проблемам материально-технического снабжения в промышленности и торговле принимались многочисленные постановления. В частности, на XXIV и XXV съездах КПСС указывалось на необходимость совершенствовать механизм планирования, образования и использования запасов материальных ресурсов, на недопустимость образования сверхнормативных и излишних запасов и материалов и оборудования Позже, как отмечается в [7],

«в ходе перестройки возникла насущная необходимость в осуществлении такой концепции управления народным хозяйством, которая объективно обеспечила бы сочетание интересов всех субъектов производства на разных уровнях управления».

В частности, возникли проблемы, связанные с противоречивостью действия на практике ряда положений Закона о государственном предприятии и постановлений о переходе на хозяйственный расчет, оптовую торговлю средствами производства, о внешнеэкономических связях, приватизации экономических объектов. Децентрализация и демократизация экономики, курс к капитализму существенно повысили роль экономических факторов в организации снабжения и способствовали переносу центра тяжести на локальные экономические структуры и решения.

Разумеется, решение возрастающих по своей актуальности проблем управления запасами дает не минимизация, а *оптимизация* запасов.

1.2. Модель управления запасами и ее элементы

Обеспечение потребностей хозяйственных, социальных и военных объектов в различных материальных средствах (топливе, продоволь-

ствии, сырье, полуфабрикатах, комплектующих деталях, расходных материалах и т.п.) включает в себя три фазы: планирование, производство и распределение. Как правило, к моменту реализации поставки данные, положенные в основу заявки, оказываются устаревшими и объем поставки уже не соответствует фактической потребности. Для предотвращения остановки производства в случае недостаточности поставки у потребителей и в системе снабжения создаются *запасы*. К необходимости создания запасов приводят следующие факторы:

- 1) Дискретность поставок.
- 2) Случайные колебания
 - в спросе за интервал между поставками,
 - в объеме поставок (например, при планировании поставок сельскохозяйственной продукции, зависящих от урожая и организации его сбора),
 - в длительности интервалов между поставками.
- 3) Предполагаемые изменения конъюнктуры:
 - сезонность спроса,
 - сезонность производства,
 - инфляционные ожидания,
 - ожидаемое повышение цен.

Перечисленные факторы, действуя порознь или группируясь в различных сочетаниях, создают тенденцию к увеличению запасов.

Имеется, однако, ряд соображений в пользу *минимизации* запасов. К их числу относятся.

- 1) плата за физическое хранение запаса;
- 2) упущенный доход, который мог бы быть получен при вложении омертвленных в запасе средств в предприятия с твердым доходом (в простейшем случае — учетный процент);
- 3) потери в количестве запаса (испарение, усушка, утруска, радиоактивный распад, хищения);
- 4) качественные изменения (ухудшение потребительских свойств вследствие необратимых процессов в хранимом продукте — разложение, гниение, ухудшение внешнего вида, старение изоляции);

- 5) устаревание (моральный износ), особенно характерный для модных товаров, бытовой электроники, персональных компьютеров, литературы по ним и их программному обеспечению.

Управление запасами заключается в установлении моментов и объемов заказа на восполнение их и распределении вновь прибывшей партии по нижестоящим звеньям системы снабжения. Совокупность правил, по которым принимаются эти решения, назовем *стратегией* управления запасами. Каждая такая стратегия связана с определенными (чаще всего в вероятностном смысле) затратами по доведению материальных средств до потребителей. Будем считать оптимальной ту стратегию, которая минимизирует эти затраты. Отыскание оптимальных стратегий и является предметом *теории оптимального управления запасами*.

При сравнении стратегий учитываются только переменные составляющие функции затрат, зависящие от выбора стратегии. Таким образом, во многих моделях управления запасами удается игнорировать большую часть затрат на содержание управленческого аппарата (кроме расходов по оформлению поставок), а также пропорциональную объему партии стоимость производства материальных средств, которая на достаточно длительном отрезке времени определяется суммарным спросом и не зависит от организации снабжения.

Математическая формулировка задачи о нахождении оптимальной стратегии существенно зависит от исследуемой ситуации. Однако общность учитываемых факторов позволяет говорить о единой модели управления запасами. Приведем ее качественное описание, ограничившись для простоты одним складом, на который поступает случайный поток качественно однородных требований — заявок от потребителей.

Заявки немедленно удовлетворяются до тех пор, пока их суммарный объем (с начала планируемого периода) не превысит начального запаса. Все последующие требования не могут быть обслужены тотчас же, вследствие чего потребитель простаивает и несет некоторый убыток. Этот убыток по справедливости относится на счет системы снабжения — она выплачивает *штраф*. Время от времени запас хранимого имущества *пополняется* со склада вышестоящего объединения, центральной базы или из промышленности, причем с каждым таким пополнением связаны определенные дополнительные затраты. Наконец, склад несет расходы по *хранению* находящегося в нем имущества. Требуется так выбрать момент и объем заказа на восполнение, чтобы суммарные затраты на

хранение, штраф и поставки были минимальны. На работу склада могут быть наложены некоторые *ограничения* (например, максимальный запас не должен превышать вместимость склада, а его стоимость — заданной суммы). В этих случаях разыскивается условный минимум затрат.

Элементами задачи управления запасами, таким образом, являются:

- 1) система снабжения;
- 2) спрос на предметы снабжения;
- 3) возможность восполнения запасов;
- 4) функция затрат,
- 5) ограничения;
- 6) стратегия управления запасами.

Напомним, что здесь и далее «стратегия» понимается в смысле терминологии принятия решений, т.е. как выбранная менеджером линия поведения, полностью определяющая его действия в рамках рассматриваемой модели.

Системы УЗ можно классифицировать по многим признакам:

- вид запасов (сырье, полуфабрикаты, готовая продукция, инструменты, запчасти);
- место хранения (производитель, потребитель, снабженческая база или другие элементы товаропроводящей сети);
- структура системы (изолированный склад, последовательная система складов, иерархическая система, с ремонтными возможностями или без них);
- свойства запасов (одно- или многономенклатурные запасы, их взаимозаменяемость, ограниченность срока годности, порча при хранении);
- статистические характеристики процессов спроса и поставок (стационарность, коррелированность спроса, управляемость, случайность поставок);
- цели системы (стоимостные и вероятностные критерии, многокритериальность);

- ограничения (на объем и номенклатуру запасов, размеры партий, надежность и экономические характеристики процесса снабжения);
- информационные характеристики (периодичность сбора данных, наблюдаемость спроса, полнота знаний о коэффициентах потерь).

1.3. Из истории теории запасов

Управление экономикой на всех уровнях должно стать точной наукой, исключающей всякую произвольность и волюнтаризм. Приведенные выше соображения и цифры позволяют утверждать, что достижение наивыгоднейшего компромисса между противоречивыми требованиями сокращения расходов на хранение и надежного обеспечения платежеспособного спроса является делом достаточно сложным и требует тщательного количественного анализа. Методы такого анализа и составляют предмет *теории управления запасами* (ТУЗ).

В течение первого полувека советского периода отсутствие достаточно разработанной теории и технических средств управления экономическими системами не позволяло полностью реализовать преимущества социалистического способа ведения хозяйства. Известную роль сыграло и нигилистическое отношение к применению математических методов в экономике, под которое даже подводили «философскую» базу, искусственно противопоставляя качественные и количественные связи между явлениями, математический и диалектический подход, причинные и функциональные зависимости.

В начале шестидесятых годов ставшая, наконец, очевидной необходимость конкретного руководства, основанного на количественном анализе, привела к перелому в отношении к математико-экономическим исследованиям. И математики, и экономисты вспомнили замечательные слова П.Л. Чебышева о том, что «математика создалась и развивалась под влиянием общей задачи всей человеческой деятельности: распорядиться имеющимися под руками средствами для достижения наибольшей выгоды». Начали интенсивно развиваться как макроэкономические теории (различные матричные модели и балансовые схемы), так и микроэкономические (локальные), охватывающие менее масштабные организации и/или отдельные стороны их деятельности. Особый размах получили исследования по линейному программированию (в частности,

транспортным задачам), нелинейному и динамическому программированию, теории игр.

Новая волна повышенного интереса к экономико-математическим методам была связана с интенсивным развитием *теории больших систем*. Системный анализ — это не что иное, как просвещенный здравый смысл, на службу которого поставлены современные аналитические методы. Хотя системный анализ — прежде всего искусство, это искусство требует и специальной математической подготовки. Системный анализ рассматривается как методология упорядочения и уяснения проблемы, которая затем уже может решаться с применением или без применения математики и ЭВМ. Системный анализ ограничивает задачу без существенных отклонений от строгой ее формулировки с помощью таких средств, как

- математические методы, позволяющие сократить последовательный анализ всех возможных комбинаций;
- оценка чувствительности задачи к различным факторам и отбрасывание несущественных переменных;
- агрегирование переменных, решений и действий;
- рациональная формулировка критериев оптимизации.

Преимуществом математических моделей является прежде всего то, что они позволяют проверять эффективность принимаемых решений, не прибегая к практической реализации. Для воплощения рекомендуются только решения, соответствующие цели, чем исключаются потери на опробование (и от опробования) неудачных вариантов. Разумеется, приходится считаться с неполнотой соответствия между моделью и реальностью.

Характерной чертой большой системы является то, что работа ее звеньев и системы в целом оценивается с точки зрения включающей их системы более высокого ранга. В связи с этим теоретически ни для одной конкретной системы нельзя точно и строго определить критерий эффективности, не поднимаясь при решении этой задачи до самого высокого уровня, при котором в рассмотрение вводятся социальные факторы глобального масштаба. Практически оказывается необходимым оценивать *экономический* выигрыш, приносимый каждой такой системой, сопоставляя его с затратами на функционирование данного комплекса. Таким образом, в теорию управления, исследование операций, теорию надеж-

ности и другие дисциплины, издавна использовавшие математические методы, начинает входить экономика.

Однако, как писал академик В.С. Немчинов, «примитивное понимание взаимоотношений между большими и малыми экономическими системами может создать лишь такую окостенелую систему, в которой все параметры управления заданы заранее. Такая система будет тормозить социальный и технический прогресс и под напором реального процесса хозяйственной жизни рано или поздно будет сломана». Академик «как в воду глядел» — прогноз исполнился с превышением ожиданий.

Выход из прежнего и нового тупиков нужно искать в предоставлении объединениям предприятий, отдельным предприятиям и организациям широкой экономической самостоятельности и задании им в директивном порядке лишь минимального числа показателей. Совпадение же интересов предприятия с государственными достигается «дополнением плана товарно-денежным механизмом с приведением в действие широкой системы экономических рычагов и материального стимулирования... Двойной механизм планирования и управления позволяет соединить свойственную централизованной системе научную обоснованность программы развития, большую избирательность решений с высокой гибкостью, приспособляемостью и инициативностью».

Очевидно, что и разработка математических моделей управления должна идти по двум направлениям:

- 1) централизованного управления ключевыми экономическими позициями,
- 2) локальных систем с управлением через товарно-денежные отношения.

В моделях централизованного управления, кроме основных натуральных показателей, должны вырабатываться и параметры, косвенно регулирующие деятельность локальных систем: правила ценообразования, нормы прибыли, принципы налогообложения и ставки налогов, уровень отчислений в бюджет и т.п. Локальные системы, считая упомянутые параметры заданными, могут планировать свои действия по критерию максимальной экономичности при соблюдении наложенных планом немногочисленных (в сравнении с недавним прошлым) ограничений. Разумеется, «правила игры» предполагаются достаточно стабильными, чтобы можно было планировать на перспективу и делать долгосрочные капиталовложения.

Модели управления запасами заложили основы синтеза математической экономики и теории управления, внося в первую динамические и вероятностные представления, а во вторую — экономические [40, с. 12].

Попытки наладить если не научное, то хотя бы разумное управление запасами предпринимались давно. В частности, представляет интерес опыт 1920-х гг. по хозрасчетным снабженческо-сбытовым государственным синдикатам. Они связывали промышленность с рынком, не вмешиваясь непосредственно в производственную деятельность. Взаимодействуя с потребителями, они ослабляли монопольный диктат производителей. Устойчивую работу обслуживаемых трестов и предприятий они обеспечивали за счет создаваемых запасов, отпуская в случае необходимости дефицитные предметы сверх плана — на коммерческой основе. Сбытовые синдикаты получали с потребителей комиссионную долю в зависимости от надежности поставок [28].

Предпринимались и попытки автоматизации систем управления снабжением. Еще в 1930-е гг. в Москве в ЦУМе вводились товарные ярлыки для машинной обработки на перфорационной технике. Эти попытки уже на базе ЭВМ пытались возобновить в 1960-х гг., но также неудачно. Ни одна АСУ запасами (и не только ими) не была доведена до конца. Не окупались даже эксплуатационные расходы. Основными причинами провала были следующие:

- Ставилась неверная цель — «эффективность АСУ» (загрузка ЭВМ, число решаемых задач) вместо «эффективности управляемого процесса».
- Проекты ориентировались не на перспективную, а на существующую технику (отсталая и дорогостоящая техническая база) и устаревали «на корню».
- Колоссальные затраты на АСУ сковывали все дальнейшие новации.
- На ЭВМ переводилась домашняя организация работ (в частности, перфоленты возили в ВЦ на автотранспорте, фиксировали и тиражировали промежуточные результаты счета на бумаге).
- Операторские пульта и технология разрабатывались в расчете на дефицитных в ту пору программистов.

- Процесс создания АСУ жестко регламентировался, принудительно насаждались « типовые решения »¹.
- Игнорировались оптимизационные задачи, решались в основном учетные.
- Отсутствовала должная информационная база (прежде всего по спросу).
- Не контролировалась достоверность вводимой информации.

Теория управления запасами считается относительно молодой отраслью исследования операций. В первой трети XX века появился ряд статей по определению оптимального объема заказа — Ф. Харриса (1915 г.), К. Стефаник-Алмейера (1927), К. Андлера (1929) и Р. Уилсона (1934), с именем которого обычно и связывается соответствующая группа формул. Эти зависимости исключительно робастны и являются компонентами алгоритмов решения многих более сложных задач, в том числе со стохастическим спросом. Шведская компания «Вольво» даже в наши дни триумфа электронной вычислительной техники снабжает своих агентов и дилеров специальной счетной линейкой, разработанной на основе формулы Уилсона.

Формирование теории управления запасами как научной дисциплины началось в середине 1950-х гг. обстоятельный обзор этого этапа ее развития был сделан в [56]. В книге [161] исторический обзор захватывает еще два десятилетия и дается по следующим разделам:

- экономичный объем заказа;
- однопериодные стохастические задачи;
- стохастические задачи с пороговым уровнем;
- непрерывный контроль уровней;
- задачи обеспечения надежности;
- многономенклатурные модели;
- иерархические системы снабжения;
- имитационное моделирование;
- новые результаты.

¹Этот процесс достоверно и художественно описан в романе А. Морозова «Программист» (М.: Молодая гвардия, 1979).

Здесь (и в списке литературы) мы ограничимся лишь доступными для российского читателя переводами иностранных монографий [5, 14, 85, 89].

Внимание отечественных исследователей — математиков, экономистов, военных, практиков — теория запасов привлекла в начале 1960-х гг. Среди наиболее активных авторов назовем Е В Булинскую В 1969 г. вышла написанная автором данной книги первая отечественная монография [56], позже переведенная в ГДР. В ней приводится обзор ранних этапов развития этой теории за рубежом и в СССР, а также обширная аннотированная библиография, к которой (а также к [161]) мы и отсылаем читателя, интересующегося историей вопроса. К этому следует добавить и отечественную предысторию за 1920–1950-е гг. [77, с. 28–32]

Позже появился целый ряд учебных пособий [4, 7, 8, 9, 13, 18, 19, 20, 21, 24, 27, 29, 33, 34, 35, 36, 37, 41, 42, 43, 47, 77, 79, 80, 87, 88, 96, 97], в связи с которыми необходимо отметить:

- широкую географию издательств (Москва, Ленинград, Минск, Рига, Одесса, Караганда),
- общность основного содержания,
- сделанный в ряде случаев заметный акцент на прикладные задачи по профилю вуза (торговля, организация производства широкого профиля, металлургия, сельское хозяйство).

Наличие в некоторых из них обширных подборок задач говорит о том, что управление запасами является объектом серьезного изучения в лучших экономических вузах страны. Сошлемся также на то, что задачи управления запасами рассматриваются и во всех современных руководствах по исследованию операций, например в [1, 6, 94].

Под руководством Е А. Хруцкого в ЦЭМИ АН СССР в 1969 г. был проведен первый, а в 1972 г. — второй симпозиум по управлению запасами [44]. В 1982, 1984 и 1989 гг. Венгерской академией наук были проведены Международные симпозиумы по этой проблематике [193] с участием советских ученых [46].

Из недавних отечественных монографий отдельного упоминания заслуживает [30]. Здесь в качестве главного показателя использовано условие обеспечения заданной надежности снабжения. Предлагаемые алгоритмы разработаны с помощью аппарата классической теории управления современных методов теории адаптации, математического

программирования, стохастической оптимизации, принципа максимума, динамического программирования, марковских процессов с доходами. Отмечена также существенная специфика задач УЗ, для которых не типичен широко используемый в общей теории управления квадратичный критерий.

В печати приводятся данные об экономической эффективности рекомендаций теории запасов. В монографии Ф. Хэнссменна [93] сообщается, что некоторым компаниям удавалось снизить уровень запасов наполовину при улучшении обслуживания потребителей. На крупной железной дороге снижение запасов составило 20%. По Джонсону [162], ежегодный выигрыш от рационального планирования запасов в ВВС Канады составил 300 тыс. долларов при одновременном увеличении боеготовности самолетов на 24%. Серия работ, проведенных лабораторией систем снабжения RAND Corporation по заказу американских ВВС, позволила вдвое сократить запасы по дорогостоящим запчастям при сохранении числа недостатков по ним практически на прежнем уровне.

Выше уже отмечалось активное участие венгерских ученых в работах по теории запасов. Отчасти по этой причине уровень запасов в Венгрии за 1970–1982 гг. снизился. Там получила значительное развитие система планового распределения средств производства путем оптовой торговли, за счет которой покрывалось 60% потребности. Свидетельством серьезного отношения к теории запасов в Венгрии является вышедший в 1990 г. справочник [161], содержащий 336 моделей теории запасов².

К сожалению, в отечественной практике, как отмечалось в [30, с. 5], зарубежные разработки в области управления запасами имели крайне ограниченную применимость. Это было связано с сильной зависимостью моделей управления запасами от основных хозяйственных механизмов, принципов ценообразования и налогообложения, организации рынка и т.д. Было практически невозможно определить издержки из-за потери предпочтения, волевого назначения норм потерь при хранении, существования целого ряда ограниченно конвертируемых денег — наличных и безналичных, валютных нескольких категорий. К этому перечню другой автор [77] добавляет

- отсутствие комплексного подхода к составным частям совокупных

²Из них 36 предложены автором данной книги.

запасов на всем пути их движения и их распределению между звеньями системы;

- представление источников снабжения в виде бункера с неограниченным наличием;
- неучет динамики производства и потребления;
- недостаточно гибкий и в общем примитивный учет последствий дефицита;
- неприспособленность существующих форм статистической и бухгалтерской отчетности к выделению параметров, необходимых для решения задач управления запасами.

Даже сейчас, в начавшийся и для России век автоматизации, состояние информационного обеспечения торговли согласно [98] характеризуется

- низким уровнем автоматизации и информатизации информационных потоков;
- большой избыточностью информации в бухгалтерских, статистических и оперативных торговых документах;
- практической невозможностью получения оперативной информации о текущем состоянии товарооборота;
- несовершенством информационных связей между отделами и службами торговых предприятий, самими предприятиями и различными уровнями управления торговлей;
- многократностью обработки и размножения одних и тех же документов;
- большой трудоемкостью ведения документооборота.

В общем же причины разрыва между теорией и практикой управления одинаковы на всех широтах и меридианах. Необходимо:

- 1) прежде всего решать задачи, значимые для практики, и исходить из практических потребностей, а не из математических возможностей;
- 2) добиваться «прозрачности» аргументации и решения;
- 3) пользоваться реальными и *первичными* данными;
- 4) учитывать затраты на управление;

- 5) убедить персонал в пользе (для организации и лично для него) от внедрения рекомендаций ТУЗ.

Основными объектами дальнейших исследований авторы [30] считают:

- интегрированную логистику (производство, хранение и транспорт совместно);
- то же — в эшелонированных системах;
- исследование игровых постановок и подходов, известных из теории принятия решений, в частности, адаптивного подхода и методов управления по неполным данным;
- частично наблюдаемый спрос;
- многономенклатурные задачи с коррелированным спросом;
- замкнутые по спросу системы;
- переменный объем требований;
- имеющую существенную специфику задачу снабжения запчастями.

Некоторые из этих проблем будут обсуждаться в дальнейшем.

1.4. Классификация моделей УЗ

Многообразие реальных ситуаций вызвало необходимость в рассмотрении огромного числа вариантов задачи управления запасами, которые систематизированы лишь частично. Использование богатейшего материала, накопленного теорией УЗ (Inventory Control), немыслимо без его упорядочения в рамках единой классификации. Попытки такой классификации и введения унифицированных обозначений (по типу известной нотации Кендалла в теории массового обслуживания — разд. 3.9) предпринимались неоднократно, но оказались малопродуктивными.

Г.И. Феклисов [87] предлагал характеризовать модель системы управления запасами 17-компонентным вектором, задающим стратегию формирования заказа, объемы пачек требований и поставок, статистические распределения интервалов между пачками каждого рода, наличие складской и тразитной форм поставок, количество уровней системы, характер связей между уровнями, формирование расходов на хранение и штрафы, наличие ограничений (по типичным их классам) и т.д.

В справочнике [161] классификация моделей проводится уже по 45 элементам. Укрупненно она различает модели по

- числу номенклатур;
- числу складов,
- характеру восполнения;
- характеру спроса;
- способу рассмотрения динамики;
- целевой функции;
- стратегии восполнения;
- способу контроля уровня запаса;
- учету недостатков;
- задержке поставок.

На стр. 56 справочника [161] приведена типовая карточка модели. Часть элементов не определяется с достаточной точностью (например, спрос), так что требуются дополнительные пояснения. Соответственно кодирование системы связано с некоторой потерей существенной информации. Дополнительно модель характеризуется ее математической проработкой:

- оптимальное решение не приводится;
- дается в замкнутом виде (формула);
- дается соотношение, определяющее итерации;
- описан имитационный алгоритм;
- получено приближенное решение;
- рекомендован другой вычислительный процесс

Сразу же отметим, что между этими пунктами не всегда удастся провести строгое различие, а последний неинформативен.

В главе 5 [161] проводится статистический анализ 336 собранных в книге моделей в разрезе предложенной классификации. В частности, обнаружилось, что

- 82% моделей однородные, 18% — многономенклатурные; только в 1% работ рассматриваются отношения замены/дополнения; лишь в 7% работ учитываются общие ограничения;

- только 10% были связаны с конкретным видом продукта;
- 79% работ рассматривают детерминированный спрос;
- 77% работ посвящены статическим моделям (с неизменными параметрами);
- в 60% работ используются непрерывные управляющие переменные;
- стратегии управления вида (s, q) рассмотрены в 25% работ, а пороговые вида (T, S) — в 20%;
- в 24% случаев рассматривается бездефицитное снабжение, в 55% — отложенный спрос, в 14% — потерянный;
- вероятностные модели рассмотрены в 51% работ, методы математического анализа использованы в 26% случаев, динамическое программирование в 11%;
- оптимальное решение получено в 30% случаев; в 8% работ предложены итерационные схемы, в 4% — аппроксимации; менее чем в 3% работ для сложных задач приведены детальная схема алгоритма или программа.

Эта статистика чрезвычайно поучительна. Она дает ясную и довольно грустную картину распределения усилий исследователей. Преимущественно что полегче, в абстрактной постановке, без серьезной проработки вычислительных аспектов.

Детальная классификация моделей УЗ имеет реальную полезность лишь при создании для последних компьютерной базы знаний. Тогда можно построить диалоговую систему, опрашивающую пользователя и последовательно формирующую код нужной модели — или ближайшей к ней (аналогично описанному в разд. 3.17 пакету программ МОСТ). По коду модели (конъюнкции признаков) можно войти в базу знаний, найти библиографический источник, посмотреть подробности, лучше уяснить допущения и метод получения результата, при необходимости и способности — модифицировать модель. Затем можно ввести релевантную задаче количественную информацию.

Для описания теоретических основ управления запасами классификацию достаточно построить на перечисленных выше (с. 21) основных элементах модели

Под **системой снабжения** понимается совокупность источников заявок и складов, между которыми в ходе операций снабжения осуществляются перевозки хранимого имущества. В состав системы могут входить звенья, обеспечивающие восстановление отказавших у пользователей устройств. Тогда в системе должны храниться как сборки (оборотный запас), так и используемые при их восстановлении элементы. В простейшем (и наиболее изученном) случае система сводится к единственному складу. Функция затрат составляется и минимизируется для системы в целом.

Возможны три варианта построения систем снабжения: децентрализованная, линейная и эшелонированная. В первом случае все склады непосредственно обслуживают потребителей, и недостача на одном или нескольких складах по решению органа управления снабжением может быть покрыта за счет избытка запасов на других складах. Источник восполнения запасов для всех складов принимается неисчерпаемым. Во втором рассматривается производственная цепочка (часто — конвейер) и рассчитывается распределение буферных запасов по степеням готовности продукта. В третьем случае каждая недостача покрывается за счет конечных запасов склада высшей ступени.

Системы снабжения классифицируются также по числу хранимых номенклатур (однородные и многономенклатурные) и по стабильности свойств хранимого имущества. Многономенклатурность является основным фактором, усложняющим модель, и разнообразно проецируется на все остальные элементы модели. В частности, спрос на разные номенклатуры может быть независимым, комплектным и коррелированным; поставки — частично либо полностью совмещаемыми и отдельными (аналогично — восстановление ремонтпригодных элементов ЗИПа); вычисление штрафов может быть независимым или производиться по ожидаемому максимуму дефицита либо вероятности дефицита; ограничения задаются независимо или совместно по группе номенклатур.

Чаще всего предполагается, что ни количество, ни свойства хранимого продукта естественным изменениям не подвержены. Однако могут быть случаи его убыли (испарение, утечка, усушка, утриска), естественной порчи (продукты питания) или, наоборот, возрастания ценности предметов хранения со временем (вина, сыры, антиквариат). При изменении свойств предметов хранения со временем и при наличии нескольких партий с различными датами выпуска задача приобретает дополнительный аспект — необходимо решить, за счет какой партии удовлетворить очередное требование.

Наконец, все системы снабжения в зависимости от постоянства их параметров и значений управляющих переменных можно разделить на статические и динамические. В первом случае рассматривается минимизация затрат за единственный период или в единицу времени, во втором — за указанное (возможно, бесконечное) число периодов, причем сумма затрат приводится к начальному периоду.

Спрос на предметы снабжения может быть

- стационарным и нестационарным;
- детерминированным или стохастическим;
- непрерывно распределенным или дискретным;
- зависящим от спроса на другие номенклатуры или независимым.

Типичными примерами нестационарных ситуаций являются торговля сезонными и модными товарами, а также периоды пикового (предпраздничного) спроса.

В случае дискретного спроса каждое отдельное требование дополнительно характеризуется своим объемом (числом заказанных единиц). Объем требования может быть постоянной или переменной (в частности, случайной с известным распределением) величиной. Требования постоянного объема без потери общности сводятся к единичным, требования переменного объема задаются распределением объема пачки и особенно характерны для пирамидальных систем со спросом, накапливаемым в нижних звеньях. Нестационарный спрос в очередной период может быть зависимым или независимым в смысле связи с предысторией процесса. Практически исследованы случаи стационарного и независимого в обоих смыслах спроса.

Детали с зависимым спросом (комплекты) нельзя планировать независимо, поскольку спрос на них будет иметь сильную положительную корреляцию. В случае взаимозаменяемых изделий (характерный пример — литература по компьютерным технологиям) имеется *отрицательная* корреляция.

Пополнение запасов всегда происходит с некоторой случайной задержкой относительно момента выдачи требования. Однако роль и длина этой задержки сильно зависят от конкретных условий, что позволяет в ряде случаев упростить задачу. Степень возможного упрощения определяет один из следующих вариантов:

- мгновенная поставка;
- задержка поставок на фиксированный срок (в частности, кратный длине периода);
- случайная задержка с известным распределением длительности.

Задержка поставок может увеличиваться в период низкого спроса, когда поставщик накапливает заказы перед запуском производства. Эту модель хорошо описывает система массового обслуживания с порогом включения и разогревом (см. разд. 9.7). Тот же эффект может наблюдаться и при очень высоком спросе, создающем очередь заявок.

В некоторых моделях с задержкой, кроме обычной, вводится *экстренная* поставка, которая, как правило, принимается мгновенной. Возможность такой поставки исключает отрицательные уровни запаса (задолженность). Ситуации с отложенным спросом типичны для военных и государственных организаций, а с потерянным — для торговли. Наконец, может существовать различие в объеме поставок:

- поставка равна требуемому количеству;
- поставка равна случайной величине с характеристиками закона распределения, в общем случае зависими от величины заказа.

С принципиально случайными поставками связаны процессы, включающие взаимодействие с природой: заполнение водохранилищ (осадки, испарение, фильтрация), сбор сельскохозяйственной продукции (здесь, однако, имеется определенная зависимость от «заказа», т.е. подготовленных и засеянных площадей). Случайный объем поставок может иметь место и в эшелонированных системах снабжения, в особенности с элементами производства. Здесь ненадежность взаимных поставок является следствием принятой структуры управления с ее характерными чертами: дискретностью контроля, укрупненной отчетностью, доминантой внеэкономических стимулов. Существенную роль играют и внутренние флюктуации из-за аварий, брака, внеплановых ремонтов и т.п.

Если случайность является следствием плохой организации снабжения, необходимо организационными мерами добиваться своевременного и полного выполнения заказов.

Функция затрат образует показатель эффективности принятой стратегии и учитывает следующие издержки:

- расходы на хранение;

- транспортные расходы и затраты, связанные с заказом каждой новой партии;
- затраты на штрафы.

Иногда в минимизируемую функцию включаются (с отрицательным знаком) доходы, полученные от продажи остатков запаса в конце каждого периода. В некоторых случаях ставится задача *максимизации доходов*.

В зависимости от особенностей исследуемой ситуации рассматриваются следующие варианты выбора отдельных составляющих функции затрат:

Издержки хранения:

- пропорциональные среднему уровню положительного запаса за период и времени существования положительного запаса;
- пропорциональные положительному остатку к концу периода;
- пропорциональные максимальному запасу;
- нелинейные функции одного из вышеуказанных количеств.

Стоимость поставки (допускаются любые комбинации перечисленных ниже вариантов):

- пропорциональная объему поставки;
- постоянная независимо от объема и числа номенклатур;
- сумма фиксированных составляющих — по числу номенклатур в заявке;
- пропорциональная необходимому приросту интенсивности производства.

Суммарный штраф:

- пропорциональный среднему уровню положительной недостачи за период и времени существования недостачи;
- пропорциональные недостаче к концу периода;
- нелинейные функции одного из вышеуказанных количеств;
- постоянный (выплачивается при ненулевой недостаче).

В многономенклатурных задачах штрафы могут суммироваться или назначаться по максимальному дефициту (требование комплектного обеспечения спроса).

Ограничения в задачах УЗ могут быть различного характера. Укажем следующие виды ограничений:

- по максимальному объему (весу, стоимости) запасов;
- по средней стоимости;
- по числу поставок в заданном интервале времени;
- по максимальному объему (весу, стоимости) поставки или кратности этого объема некоторой минимальной величине (целое число стандартных «упаковок» — вагонов, цистерн, бочек, коробок);
- по доле требований, удовлетворяемых из наличного запаса (без дополнительных задержек).

Введение ограничений может существенно изменить формулировку задачи УЗ. В частности, в стохастической модели без ограничений оптимальный запас, обращая в минимум сумму затрат на поставки, хранение и штрафы, автоматически дает наиболее выгодную вероятность недостачи. Ограничение же последнего типа полностью определяет сумму штрафа, что позволяет исключить ее из функции затрат и минимизировать только расходы на поставки и хранение. Если расходы на хранение и поставки заданы, то отыскивается стратегия, максимизирующая вероятность обеспечения спроса. Такой вариант особенно часто встречается в многономенклатурных задачах

Иногда в задаче управления запасами минимизируются не денежные расходы, а какой-либо другой дефицитный ресурс (вес, объем) — обычно при заданной вероятности обеспечения многономенклатурного спроса. На математической стороне вопроса такая замена по существу не отражается.

Опыт решения многих задач исследования операций и управления запасами в частности свидетельствует о том, что целевая функция в окрестности оптимума меняется медленно. В сочетании с неизбежной погрешностью исходных данных это оправдывает приближенный расчет оптимальных параметров системы и различные допущения, которые приходится делать для получения решения.

В задаче с учетом случайных факторов (вариабельность спроса, задержки поставок, случайное восполнение) ожидаемые затраты всегда больше, чем в строго детерминированном случае при сопоставимых исходных данных. Это превышение следует рассматривать как вынужденную плату за работу в условиях неопределенности.

Стратегия управления запасами, т.е. структура правил определения момента и объема заказа, в приложениях обычно считается известной, и задача сводится к определению нескольких констант (параметров стратегии). В *периодических* стратегиях заказ производится в каждом периоде T , в стратегиях с критическими уровнями («оперативных», «с непрерывным контролем») — при снижении запаса до порога s или ниже. Второе различие между простейшими стратегиями определяется правилом определения *объема* заказа. постоянная партия объема q или партия, восполняющая наличный запас (в сумме с ранее сделанными заказами) до верхнего критического уровня S . Перечисленные правила относятся к стационарным ситуациям и могут временно корректироваться для нестационарных (например, накануне очередного предпраздничного пика в торговле)

Необходимо отметить, что область применения теории управления запасами отнюдь не ограничивается складскими операциями. В частности, под запасом можно подразумевать.

- наличие товара;
- рабочую силу, планируемую для выполнения конкретного задания;
- размер капитала страховой компании;
- емкость складских помещений;
- объем информации в базе данных;
- грузоподъемность транспортных средств;
- производственную мощность предприятия;
- напор воды в водохранилище ГЭС;
- численность персонала данной квалификации (при планировании подготовки кадров) и т.д.

Ясно, что в соответствии с этим меняется содержание слова «заказ», которое теперь подразумевает более общий акт управления: закупка, договор о найме рабочей силы, начало строительства новых производственных цехов, под «поставкой» же понимается реализация соответствующего акта. «Штраф» за недостачу определенного ресурса, естественно, рассчитывается специфическим образом. Таким образом, при переосмысливании элементов модели методами теории УЗ может

быть решен очень широкий круг задач оптимального планирования. Однако для удобства и компактности изложения мы сохраним чисто снабженческую терминологию

1.5. Выбор варианта модели

Выше были перечислены элементы, сочетание вариантов которых образует конкретную модель теории запасов. Выясним, как классифицировать варианты этих элементов в практических ситуациях.

Система снабжения. Соответствующий вариант определяется спецификой и размещением потребителей и складов. *Изолированным* можно считать любой склад с единственным источником восполнения по каждой номенклатуре при условии, что вероятностью отсутствия запасов у поставщика можно пренебречь.

Если между несколькими такими складами в критических ситуациях возможен обмен запасами, имеет место децентрализованная система снабжения. При источнике снабжения ограниченной мощности по необходимости приходится требовать запас из более высокого звена. В результате модель приобретает эшелонированную структуру — как правило, пирамидального типа (каждый склад высшего ранга обслуживает несколько складов низшего ранга). При рассмотрении запасов сырья и полуфабрикатов для многоступенчатого производственного процесса пригоден линейный (цепной) вариант эшелонированной системы. Если некоторые из промежуточных продуктов могут использоваться в нескольких процессах, проходящих одновременно, система вновь приобретает пирамидальное строение.

В подавляющем большинстве случаев на складах систем снабжения хранится несколько или много (вплоть до сотен тысяч) номенклатур. Тем не менее, часто удается свести задачу к однономенклатурной, решение которой существенно проще. Таких ситуаций три

- 1) Каждая номенклатура поставляется независимо (объединение поставок невозможно), а штрафы по отдельным номенклатурам суммируются. В этом случае задача распадается на N (по числу номенклатур) однокомпонентных.
- 2) Поставки осуществляются комплектами, каждый элемент может быть использован только в данном комплекте (детали и узлы

сложных технических устройств). Такой комплект можно считать одной обобщенной номенклатурой.

- 3) Спрос на группу номенклатур, приходящих от одного поставщика, имеет сильную положительную корреляцию. Эти номенклатуры вновь объединяются в обобщенный предмет снабжения.

Классификация по *стабильности* свойств предметов хранения проводится в зависимости от быстроты изменения этих свойств (скорости естественной убыли или порчи) в сопоставлении с периодичностью поставок. Если эта убыль за период (а ухудшение свойств хранимого обычно приводится к эквивалентному уменьшению количества при условном сохранении качества) значительно меньше ожидаемого спроса, хранимые материалы и продукты можно считать стабильными. Исключение представляет случай *очень дорогостоящих продуктов* (например, короткоживущих радиоактивных изотопов). Даже малая естественная убыль их за период между поставками вносит существенный вклад в затраты на хранение, и их нестабильность должна обязательно учитываться в модели. Тем более это относится к ситуации, когда продукты распада затрудняют использование хранимого вещества, «отравляя» его.

Все модели делятся на статические и динамические. К статическим задачам можно отнести ситуации, где все параметры не зависят от времени, а *момент* затраты определенной суммы с точки зрения итогового показателя безразличен. В частности, это модели с минимизацией затрат в единицу времени. В динамических задачах приходится считаться с экономической выгодой от рассрочки платежей, что проявляется в стремлении оттягивать затраты на конец планируемого периода. Так, если L_k — затраты в k -й интервал этого периода, то общие затраты будут подсчитываться согласно $L = \sum_{k=1}^n \alpha^{k-1} L_k$, где $\alpha \leq 1$ — так называемый дисконт-фактор, а n — общее число интервалов. Этим приемом все затраты приводятся к настоящему моменту времени (на первый период). Второй особенностью динамических задач является зависимость начального запаса в каждом периоде от решений, принятых в предшествующих периодах. Решение динамических задач значительно сложнее, чем статических.

Спрос на предметы снабжения. В этом пункте мы рассмотрим только выбор варианта задания спроса. Способы исследования количественных характеристик спроса разбираются в главе 4.

Необходимо различать две ситуации. Первая характеризуется тем, что потребительский спрос, не удовлетворенный на данном этапе, можно удовлетворить на одном из последующих (предъявленные потребителями заказы ставятся на учет и выполняются в дальнейшем). Она соответствует закреплению потребителей за поставщиком. Доказано (H. Scarf), что в этом случае оптимальное управление определяется только одной кумулятивной переменной состояния, равной сумме текущего запаса и всех ранее сделанных, но еще не реализованных заявок на восполнение. Для систем с потерями оптимальный заказ должен зависеть от каждой из упомянутых переменных *porozнь*.

Вторая ситуация порождает отказ в обслуживании, и потребитель обращается к другому поставщику. Типичным примером является торговля. Следствие отказа для торгового предприятия — упущенный розничный доход, возможная полная потеря данного клиента и его знакомых и потенциального «настрига» с них.

Стационарность спроса определяется прежде всего условиями работы потребителя. Строго говоря, трудно ожидать действительно стационарного спроса в течение длительного промежутка времени, но в целях удобства анализа малыми изменениями параметров можно пренебречь и, периодически пересматривая последние, считать спрос кусочно-стационарным. Характеристиками спроса, близкими к стационарным, обладают промышленные объекты, введенные в нормальный режим эксплуатации, в особенности с многономенклатурным производством, когда изменения производственной программы по отдельным позициям, требующим общих исходных материалов, частично нивелируют колебания в спросе на последние.

Детерминированность спроса (иначе говоря, его предсказуемость) определяется ролью случайных факторов (и действием закона больших чисел) в процессе потребления материальных средств. Для завода с жесткой производственной программой или магазина продовольственных товаров эта роль относительно невелика, и спрос может быть спрогнозирован с достаточной для практики точностью. Наоборот, планирование поставок запасных частей, как правило, приводит к управлению запасами с вероятностным спросом. Критерием перехода к таким моделям можно считать коэффициент вариации (отношение среднеквадратического отклонения к среднему спросу), превышающий значение 0.2.

Особым классом задач управления запасами является исследование ситуаций, где в связи с новизной техники и отсутствием достаточного

опыта эксплуатации конкретных образцов известен только тип распределения отказов, но не его параметры. Здесь приходится применять *адаптивные*, т.е. совершенствуемые в процессе их реализации, стратегии управления запасами.

Дискретность спроса характерна для дорогостоящего оборудования и предметов потребления длительного пользования. Дискретным представлением спроса пользуются преимущественно при малой интенсивности спроса (малом спросе за планируемый период). При достаточно большом спросе за период (не менее четырех единиц) применяют более удобное с точки зрения обычных вычислительных методов *непрерывное* представление спроса, а дробные результаты округляют до ближайшего целого числа. Единицей в зависимости от задачи может быть ящик молотков, дюжина молотков или один молоток: важно лишь, чтобы выбор единицы не менялся на всех этапах решения задачи. Возможность округления базируется на малости производной целевой функции по ее аргументу в окрестности оптимума. При определении страховых запасов, уменьшение которых приводит к быстрому росту суммы штрафов, округление целесообразно проводить в большую сторону.

При решении многопериодных или многономенклатурных задач методами динамического программирования для сокращения объема вычислений, наоборот, проводят дискретизацию спроса.

Пополнение запасов. Критерием выбора варианта здесь является ожидаемое значение спроса за время t задержки между подачей заявки и выполнением заказа. Если эта величина пренебрежимо мала, допустимо использование модели с мгновенной поставкой. В противном случае при малом коэффициенте вариации задержки выбирается модель с фиксированной задержкой, а при большом — со случайной. Распределение случайной задержки устанавливается статистически теми же методами, что и распределение спроса.

Объем восполнения *случаен*, если возможности поставщика в момент заказа недостаточно предсказуемы.

Функция затрат (она же — целевая функция) для статической модели представляет собой затраты за период. При осуществлении динамической программы, рассчитанной на небольшое число n сравнительно продолжительных периодов, чаще всего применяется взвешенное (дисконтированное) суммирование затрат в отдельные периоды. Приведение обычно делается к нулевому периоду, на который приходится

капиталовложения. При большой продолжительности операций по снабжению в качестве целевой функции выбирают предельное (при $n \rightarrow \infty$) значение упомянутой взвешенной суммы или минимум затрат в единицу времени.

В зависимости от этого выбора устанавливается способ расчета затрат на хранение и штрафы. При минимизации расходов за период эти затраты чаще всего исчисляются по ожидаемому остатку (дефициту) к концу периода. В противном случае предпочтителен расчет по «импульсу» этих величин — произведению средней недостачи (запаса) на время ее (его) существования. Этот же подход применяется в моделях с повышенной ролью фактора времени, где необходимо учитывать *время простоя* потребителей. Иногда величина штрафа не зависит не только от момента появления недостачи, но и от ее величины, так что некоторая постоянная сумма «выплачивается» при первом же необслуженном требовании. В этом случае штраф определяется *вероятностью* недостачи. Как правило, затраты на хранение или штрафы являются линейной функцией упомянутых величин. Применительно к особо ответственным системам (объекты энергетики, жизнеобеспечения, системы военного назначения) строятся специфические (обычно нелинейные) штрафные функции.

Ограничения обычно задаются руководителем системы при постановке задачи об оптимальном управлении запасами. В особенности это относится к ограничениям, исключающим функцию штрафов. Они классифицируются на ограничения

- поставщика (минимальный и максимальный размеры заказа, кратность его стандартной партии);
- рынка (уровень обслуживания, стоимостный спрос);
- внутренние (вместимость склада, бюджет)

Наиболее типичны ограничения по вместимости складов, суммарной стоимости запасов и моментам выдачи заказов. Приходится считаться и с понятием «транзитная норма отгрузки» — объем партии, поставка которой считается экономически целесообразной. Этот объем обычно определяется емкостью типового транспортного средства (танкер, вагон, цистерна, платформа, контейнер, грузовик и т.п.) и меняется в зависимости от характера перевозимого груза.

Выбор конкретного ограничения определяется условиями работы системы. Возможен и одновременный учет нескольких ограничений.

Трудности вычисления оптимальных параметров значительно возрастают с введением каждого нового ограничения. Поэтому имеет смысл сначала решить задачу без ограничений и, получив решение, проверить существенность каждого из них. Далее нужно последовательно вводить эти ограничения, начиная с имеющих наибольшую относительную жесткость (степень их относительного нарушения «свободным» решением).

Стратегии управления запасами обычно выбираются из перечисленных выше простейших, оптимальность которых доказана теорией для весьма широкого круга ситуаций. На военных и промышленных объектах в недавнем прошлом гораздо чаще использовались системы с периодической проверкой, но с появлением АСУ снабжением усиливается тенденция к переходу на системы с оперативной информацией.

Периодические стратегии используются при фиксированных сроках подачи заявок (например, при снабжении центральных и региональных складов из промышленности или от экспортера). Периодичность проверок может колебаться от одного дня до двух недель для группы А, от недели до месяца — для В и от двух недель до 6 месяцев для группы С³.

Важными аспектами стратегий управления запасами являются учет задержки поставки и возможных вариаций ее объема (урожаем сельскохозяйственной продукции, улов рыбы, поступление воды в водохранилище, ограниченность возможностей вышележащих уровней системы). Реальная задержка лежит в пределах от двух-трех часов до нескольких месяцев и определяется географическими факторами и организацией снабжения.

Иногда заведомо ограничиваются каким-либо естественным классом политик управления, зависящих от небольшого числа параметров. Тогда задача сводится к определению оптимальных значений этих параметров. Обычно она может быть разделена на две: задача программного управления на основе детерминированной компоненты спроса и задача управления с обратной связью, призванная обеспечить компенсацию его флуктуаций. Первая, детерминированная, определяется ценой заказа и средними затратами на хранение. Вторая служит для поддержания страхового уровня, обеспечивающего определенную гарантию удовлетворения спроса. Сама гарантия зависит от цены штрафа и с ростом ее стремится к абсолютной. Однако при таком подходе остаются сомнения, не

³Номенклатуры группируются в зависимости от их доли стоимостного спроса в суммарном по убыванию этих долей (см. главу 6).

существует ли более сложное правило подачи заказов, выходящего за рамки рассматриваемого класса и обеспечивающего меньшие издержки. В некоторых случаях удается выделить условия, при которых простые политики будут оптимальными в глобальном смысле. Несколько таких ситуаций со ссылкой на первоисточники рассмотрено в [56].

Качественные особенности работы систем управления запасами, реализующих изложенные выше принципы, очевидны. На рис. 1.1 приведены графики изменения запаса для трех основных стратегий при следующих допущениях:

- восполнение мгновенно;
- уровни запаса контролируются периодически через интервалы τ ;
- допускается отрицательный запас (backorders — задолженность, которая гасится при очередной поставке);
- спрос непрерывен по времени и объему;
- объем поставки детерминирован.

Не показанная на рисунке стратегия (T, q) с периодическим восполнением и постоянным объемом заказа q фактически не содержит элементов обратной связи, и процесс оказывается неуправляемым. Она соответствует чисто нормативному снабжению и может быть применена лишь в условиях стабильного спроса.

Периодическая стратегия с предельным верхним уровнем (T, S) легко встраивается в обычные схемы календарного планирования, гибко реагирует на изменения спроса. Обычная область ее применения — при высокой цене штрафа. Модели с периодическим пополнением имеют нерегулируемую частоту заказов. Это вызывает излишние транспортные и административные расходы после периодов с низким спросом и увеличивает средний дефицит при высоком спросе. С другой стороны, их преимуществом является совпадение моментов просмотра и заказа, что сводит к минимуму расходы на просмотр и связанную с ним канцелярскую работу. Объем гарантийного запаса здесь должен обеспечивать функционирование на интервале длительностью $T + \tau$, тогда как при пороговой (с непрерывным контролем) — только на время задержки поставок τ . Поэтому страховые запасы при периодических стратегиях больше.

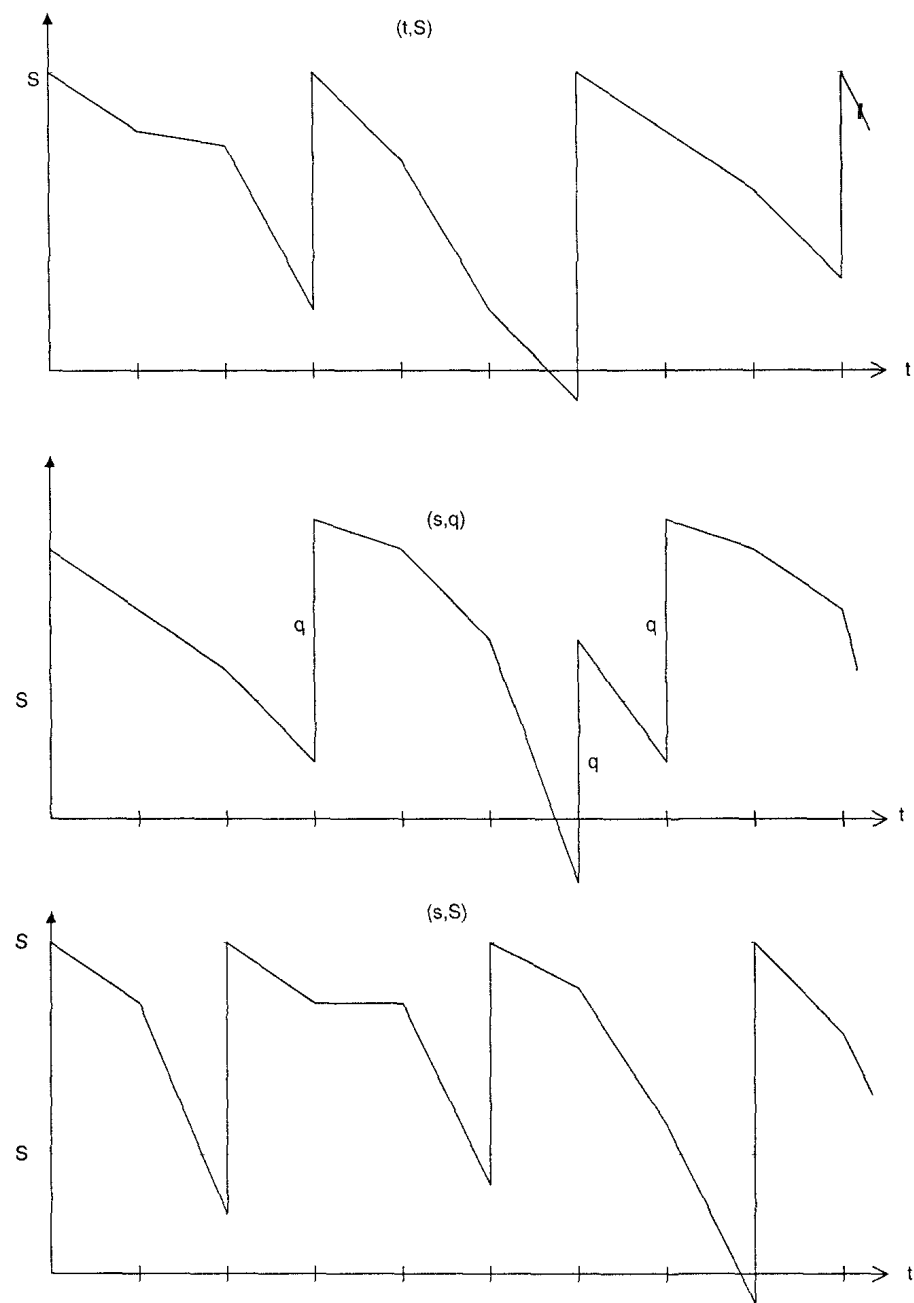


Рис. 1.1. Типовые стратегии управления запасами

В многономенклатурных задачах для достижения экономии на поставках необходимо совмещать моменты заказов. Для них наиболее подходящей будет стратегия (T, S) в ее векторной форме, т.е. с различными, вообще говоря, значениями $\{T_i, S_i\}$ для каждой номенклатуры, причем периоды $\{T_i\}$ для обеспечения совмещения заказов должны быть кратны некоторому базисному периоду T .

Стратегии с пороговым уровнем заказа реализуют принцип обратной связи по состоянию процесса. Модель с критическим уровнем (s, q) реагирует на спрос более медленно, чем (T, S) , так как спрос с момента последней поставки и до перехода критического уровня накапливается, не вызывая реакции системы. Сама эта реакция не является достаточно тонкой и при временной стабилизации спроса на аномально высоком уровне приводит к частым недостаткам.

Система «двух уровней» (или двухбункерная) (s, S) является наиболее гибкой по отношению к спросу и позволяет поддерживать относительно постоянное постоянство запаса вблизи критического уровня при достаточно редких поставках. Здесь при снижении запаса до уровня $y \leq s$ выдается заказ на партию объема $S - y$. В практическом использовании она сложнее, чем (s, q) . Употребительным частным случаем является ее вариант с $s = S - 1$ (при дискретном спросе). Здесь заказ производится при поступлении каждого очередного требования. Еще более частный случай — отсутствие запаса на складе данного уровня ($S = 0$).

Необходимо отметить, что при непрерывном контроле уровней и постоянном объеме требований стратегии (s, S) и (s, q) при $S = s + q$ тождественны. Различие между этими стратегиями по определению объема заказа возникает только при дискретизации моментов контроля текущего уровня (в частности, именно так обстояло дело в неавтоматизированных системах снабжения) или при переменной величине требований.

При поступлении требований в дискретные моменты времени нет смысла контролировать уровни непрерывно — достаточно сравнивать с порогом s остаток после удовлетворения каждого очередного требования. Учет этого обстоятельства позволяет считать, что для однопродуктовой (или приводящейся к таковой) задачи стратегия (s, S) всегда является наилучшей.

Из более сложных стратегий имеет смысл рассматривать трехпараметрические стратегии вида (T, s, S) , комбинирующие особенности периодических и пороговых стратегий. В неавтоматизированных системах снабжения стоимость проверки может быть соизмерима со стоимостью

подачи заказа. Нежелательно проводить проверки, не заканчивающиеся выдачей заказа — это было бы слишком расточительно. Это означает, что в системе (T, s, S) с периодическим контролем и уровнем заказа после проверки должен всегда следовать заказ, а потому она будет практически эквивалентной (T, S) .

Другой возможный вариант стратегии для многономенклатурных задач — заказ партий объема $S_i - y_i$ по всем номенклатурам i , для которых в момент обнаружения уровня $y_j \leq s'_j$ оказалось, что наличный запас $y_i \leq s''_i$, причем $s''_i > s'_i$.

В разных звеньях складской системы возможно использование разных стратегий.

При сравнении стратегий должна учитываться и цена управления (обработки данных). Современные АСУ снабжением, исключая самые большие, могут быть реализованы на базе персональных ЭВМ и широко распространенных типовых систем управления базами данных для них. Это определяет умеренные начальные капиталовложения и малые операционные расходы на обработку текущей информации. При управлении запасами по современной информационной технологии (с помощью электронных баз данных) реализация пороговых стратегий не сложнее, чем периодических.

В заключение отметим стратегию управления Just in Time (точно в срок), реализованную в Японии на производстве автомашин «Тойота». Японцы обходятся двухчасовым запасом деталей; американцы в аналогичных условиях содержат месячный, что сказывается на цене и конкурентоспособности продукции. Поставки «с колес» требуют идеальной организации, полностью определяются производственным графиком и с точки зрения теории запасов являются вырожденной ситуацией — здесь запасы отсутствуют.

Глава 2

Экономические параметры модели

2.1. Экономика снабжения

Когда проблема (управления запасами) сформулирована, центр тяжести переносится на поиск, сбор, анализ и отсеив исходной информации, которая может быть неполной, недостоверной и даже искаженной сознательно. Известно, что постановка каждой новой задачи требует для своего решения и такой информации, которая раньше отсутствовала или не выделялась (пример — фиксированная составляющая транспортных затрат и затрат на оформление заказа).

Для хранения, перемещения, обеспечения сохранности и учета продуктов труда требуются складские емкости, подвижной состав, подъемно-транспортное и электронно-вычислительное оборудование, соответствующие сооружения и передаточные устройства (продуктопроводы, зарядные и заправочные станции, эстакады, дороги, мосты и т.п.). Основные производственные фонды требуют значительных единовременных капитальных вложений для первоначального ввода их в действие, периодического капитального, среднего и текущего ремонта, а также организации системы амортизационных отчислений для накопления средств, необходимых для ремонта и полного обновления физически или морально износившихся основных фондов. *Методики* их расчета практически не зависят от предметной области.

Текущие затраты специфичны — по крайней мере, на разных этапах кругооборота продуктов труда их классифицируют по-разному (с учетом специфики этапа). На транспорте, например, рассматривают начально-конечные операции (эта составляющая не зависит от расстояния) и собственно передвижение. На стадии производства затраты, связанные с производственным запасом и с запасом готовой продукции, интерпретируются соответственно как затраты на снабженческую и сбытовую деятельность.

Для решения задач управления запасами появляется необходимость изменить методический подход к подсчету затрат в соответствии со следующими принципами:

- 1) Источником возникновения затрат являются материальные ресурсы — хранимые, перемещаемые и сохраняемые.
- 2) Системный подход требует единого состава затрат для любой стадии существования запасов. Состав затрат определяется элементами затрат живого и овеществленного труда, используемого на различного рода работах с материальными ресурсами.
- 3) Классификация должна разделять работу с материальными ресурсами по непересекающимся видам деятельности в соответствии с технологией формирования, хранения, перемещения и сохранения материальных ресурсов.
- 4) Из затрат должны исключаться элементы, вызванные низким уровнем планирования и управления (потери от простоя транспортных средств по вине склада, сверхнормативной естественной убыли и хищений, выплаты процентов по ссудам, неустойки за нарушение договорных обязательств по срокам, ассортименту и объему поставок).

В текущем учете и периодической отчетности конкретные данные о расходах на управление снабжением не выделяются, а бухгалтерская документация содержит их в «котловых» статьях: «Сырье и материалы», «Топливо», «Общезаводские расходы». Для получения нужных устойчивых данных необходим большой объем выборки по различным министерствам, ведомствам, фирмам, предприятиям разных типов. Поэтому специалисты по теории запасов предлагают из существующей отчетности определять лишь порядок величин затрат, дополнив их выборочными обследованиями и экспертными оценками.

Затраты необходимо учитывать по месту их возникновения в кругообороте, исключая повторный счет. При наличии реальных тарифов снабженческо-сбытовой организации нет необходимости контролировать и оплачивать чужую деятельность. В частности, в составе транспортных расходов должны остаться только те, которые возникают непосредственно внутри нее в результате перемещений грузов ее технологическим транспортом. Обсуждаемые ниже данные в современных экономических условиях интересны и полезны главным образом перечнем и группировкой статей расхода и предполагают проведение экономического анализа конкретной ситуации.

2.2. Транспортные расходы

Во многих задачах управления запасами (прежде всего при синтезе систем снабжения) необходимо рассматривать альтернативные способы транспортировки продукции и связанные с ними затраты. К концу существования СССР во внутрисоюзном грузообороте удельный вес *железнодорожного* транспорта составлял около 80%. Грузопоток *трубопроводного* транспорта за 1965–1975 гг. увеличился более чем в два раза. Это связано с его эффективностью — при его использовании почти отсутствуют затраты на начально-конечные операции. Доля *автомобильного* транспорта несколько повышается — в частности, вследствие роста парка большегрузных автомобилей. Данный вид транспорта имеет бесспорное преимущество при перевозках на короткие расстояния, при подвозе грузов к станциям, причалам и от них, а также при перевозках мелких партий грузов, так как не требует длительного накопления грузов и не несет больших убытков от недоиспользования грузоподъемности подвижного состава.

При расчете стоимости доставки товара учитывается только пропорциональная расстоянию составляющая, при определении фиксированной компоненты стоимости поставок — начально-конечная. Приведем доли (в процентах) расходов по начально-конечным операциям с погрузкой и выгрузкой [27] для перевозок на расстояние 500 км (в скобках — на 100 км):

- железнодорожный транспорт — 45 (80);
- речной — 55 (86);
- морской каботаж — 91 (94);

- автомобильный — 3 (12);
- воздушный — 25 (63).

Все транспортные затраты, разумеется, позже относятся на счет получателя.

2.3. Расчет стоимости хранения

Стоимость хранения материальных средств — единственный фактор в модели управления запасами, определяющий целесообразность сокращения последних. Как правило, она является линейной функцией характеристик движения запаса. Коэффициентом пропорциональности служит цена хранения единицы данного продукта в единицу времени или за период между поставками. По Постановлению ЦК КПСС и Совмина от 12 июля 1979 г плата за сверхнормативные и непрокредитованные запасы материальных средств и не установленного оборудования вносилась из прибыли предприятия. Независимо от этого постановления так же обстоит дело с запасами предприятий негосударственных форм собственности.

Потери от хранения (собственно хранение, страховка, налоги, убыль, устаревание) исчисляются от среднего уровня запаса, омертвление — от максимального (краткосрочно освобождаемые средства выгодно вложить не удастся). Стоимость хранения обычно считается пропорциональной стоимости имущества («как в ломбарде» [80]). Этот последний подход является сугубо приближенным. Он основан на большом удельном весе в цене хранения платы за омертвление средств и введении к последней усредненных по множеству номенклатур процентных надбавок. В хозяйственной практике запасы расцениваются в зависимости от места их нахождения: для производителя — по себестоимости, в оптовой торговле — по оптовым ценам покупателя (без налога с оборота), у отдельных потребителей — по оптовым ценам промышленности (включая налог с оборота), в розничной торговле — по розничным ценам.

Классификация затрат в системах снабжения определяется целями анализа. Ниже приводятся данные [27, гл. 3] о распределении относительных издержек содержания производственных запасов по набору статей, «классическому» для экономики советского периода.

Расходы по управлению снабжением (13%), в том числе на оплату собственно управленческого персонала — 5.6; на работников по приему материалов — 4.2; на израсходованные при приемке материалы — 2.3.

Транспортные расходы (34%): оплата железнодорожных тарифов транзитной поставки — 1.7; складская наценка при поставках с базы — 2.9, расходы на подачу вагонов и погрузо-разгрузочные работы, выполняемые транспортным и крановым цехами — 11.4; стоимость доставки автотранспортом предприятия — 4.5; автотранспортом сторонних организаций — 6.7; прочие транспортные — 6.8.

Издержки содержания запасов (52.6%). основная и дополнительная зарплата работников складов и сектора складского хозяйства отдела снабжения — 7.0; услуги вспомогательных цехов — 20.5; амортизационные отчисления от стоимости основных фондов и инвентаря складов — 1.5; плата за основные фонды складов — 3.9; текущие расходы по содержанию складов (отопление, освещение, водоснабжение) — 2.3.

Затраты на иммобилизацию запасов (убытки от омертвления денежных средств) оценивались в 17.3%.

В цитированной работе далее сформулированы предложения по изменению перечня статей. Транспортные тарифы переложены на потребителя. В новой группировке (60.3% от первоначальных затрат) расходы на управление составляют 21.5%; на перемещение — 26.3%; на содержание — 52.2%. Приведем соответственно исправленные данные из того же источника.

Расходы по управлению снабжением составляют 58.6%, в том числе на оплату собственно управленческого персонала — 28.5; служебные командировки — 2.3; канцелярские расходы и услуги связи — 1.5; содержание персонала складов — 13.7; упаковку — 11.6.

Транспортные расходы внутри предприятия — всего 33%.

Издержки содержания запасов (8.6%): содержание основных фондов — 5.1; текущий ремонт и амортизация — 3.5.

С позиций теории управления запасами, в образовании цены хранения принимают участие следующие факторы:

- стоимость складского помещения и техники;
- эксплуатационные расходы и затраты на управление;
- транспортные операции;

- затраты на регламентные работы, проводимые с хранимым имуществом;
- потери от естественной убыли имущества;
- убытки от снижения его потребительских качеств;
- потери от омертвления денежных средств, вложенных в запас.

Рассмотрим способ учета каждого из них.

Стоимость помещения. Стоимость складского помещения включает в себя плату за основные фонды (земля, собственно здание, ограждение и оборудование), амортизационные отчисления (стоимость склада, деленная на ориентировочный срок его службы), охрану, расходы на отопление или охлаждение (в зависимости от климатических условий и времени года), уборку, освещение, вентиляцию, подачу воды и т.п. Строго говоря, все эти расходы не зависят от величины запаса (при условии, что весь запас полностью размещается на данном складе), так что при хранении у себя однородного продукта на складе достаточной вместимости с ними иногда можно не считаться («оценка» по Л.В. Канторовичу емкости склада равна нулю). В случае аренды оплата складских помещений должна начисляться пропорционально максимальному уровню запасов.

В многономенклатурных задачах, где возникает вопрос о распределении ограниченных складских площадей между отдельными продуктами, эти затраты должны обязательно учитываться, что, в частности, может привести к нецелесообразности создания запаса по отдельным номенклатурам. Для расчета цены аренды все связанные с ней расходы делят на полезную площадь (емкость) склада и полученную цифру умножают на площадь (объем), занимаемый единицей хранимого продукта. Плата за аренду здесь должна учитываться независимо от того, имеет ли место формальная аренда или предприятие является владельцем склада.

Эксплуатационные расходы. Здесь должны учитываться только те расходы, которые непосредственно связаны с обслуживанием потребителей: подготовка нарядов, комплектация груза, его упаковка, погрузка и отправка потребителю, а также расходы на машинно-счетные работы, тару (за вычетом доходов по операциям с тарой), вспомогательные материалы, спецодежду, инвентаризацию, профилактические осмотры хранимого имущества. В зависимости от статей расхода они либо определяются статистически по каждой номенклатуре и затем относят-

ся к годовому спросу на нее, либо раскладываются пропорционально площади, занимаемой каждой единицей хранения. Оплата персонала рассчитывается отдельно по каждой группе операций с учетом дополнительной оплаты, расходов на социальное страхование и накладных расходов (управляющий персонал). Учитываются также канцелярские и почтово-телеграфные расходы (включая e-mail), оплата междугородных переговоров по телефону.

Транспортные расходы. В эту статью ранее включались все затраты базы, связанные с перемещением хранимого имущества: оплата железнодорожных тарифов транзитной поставки; складская наценка при поставках с базы; расходы на подачу вагонов и погрузо-разгрузочные работы, выполняемые транспортным и крановым цехами; стоимость доставки автотранспортом предприятия; автотранспортом сторонних организаций; прочие транспортные. Их общая доля в стоимости хранения достигала 34% [26]. По современным воззрениям, подавляющая их часть относится на счет потребителя в составе постоянной или пропорциональной стоимости заказа.

Все затраты, связанные с организацией восполнения и отгрузки запасов на складе, учитываются при расчете стоимости поставок и в данную статью расходов не включаются.

К регламентным работам можно отнести любые дополнительные мероприятия, проводимые на складах с целью обеспечения сохранности имущества, материалов и товаров. Затраты на них играют весьма значительную роль на складах технического имущества и в расчете на единицу хранения определяются как произведение стоимости одного регламента (с учетом квалификации проверяющего, амортизации используемой аппаратуры и затраченного времени) на число проверок в год. Принципиально методика расчета затрат на регламентные работы будет общей для склада радиоэлектронного оборудования и для овощехранилища.

Потери от естественной убыли хранимого. В некоторых случаях хранение продукта связано с его естественной убылью (летучие жидкости, сжиженные газы, радиоактивные изотопы). Убыль происходит по экспоненциальному закону: $y(t) = y_0 e^{-\gamma t}$. В типичных условиях малости произведения γt имеет место $y(t) \approx y_0(1 - \gamma t)$, так что относительные потери в единицу времени (по количеству и по стоимости) равны γ . Для летучих жидкостей убыль пропорциональна поверхности испарения. Поэтому в случае полусферического котла (в отличие от поставленной вертикально цилиндрической бочки) скорость испарения

будет убывать по мере уменьшения содержимого.

Убытки от снижения потребительских качеств. Цены на сырье, используемые изготовителем, могут подвергаться значительным сезонным колебаниям. Когда цена низка, выгодно создавать запасы сырья, которых хватило бы на весь сезон высокого спроса и которые можно было бы по мере надобности использовать в производстве. Потери при хранении могут возникнуть по физико-химическим причинам (например, снижение сахаристости сахарной свеклы) и из-за морального устаревания продукта. Эти расходы принимаются пропорциональными стоимости единицы продукта и снижению потребительских качеств. Последнее определяется по уменьшению ресурса (надежности, концентрации полезного компонента), приросту процента отбраковки в единицу времени (за год) или вынужденной уценке. Потери от морального устаревания определяются по относительному уменьшению спроса в единицу времени и компенсирующей его уценке.

Потери от омертвления средств. Эти потери фактически означают упущенную выгоду и должны вычисляться в соответствии с нормой эффективности E денежных вложений в оборотные средства. К определению E существуют различные подходы — см., например, [77, с. 223–226]. Во времена «развитого социализма» она составляла 10–15%. По-видимому, в нынешней нестабильной экономической ситуации нижней гранью оценки следует считать процент Сбербанка по вкладам населения. Составляющая стоимости годового хранения, связанная с омертвлением денежных средств, определяется как произведение стоимости единицы хранимого на принятую норму эффективности оборотных средств.

С нормой эффективности связано значение дисконт-фактора $\alpha = 1/(1 + E) \approx 1 - E$. С его помощью затраты L_i , сделанные в i -й год работы системы, пересчитываются к начальному моменту капитальных вложений согласно $\tilde{L}_i = \alpha^{i-1} L_i$.

В сборнике «Организация производства на промышленных предприятиях США» [38, гл. 18] приводятся сводные результаты анализа стоимости хранения на 50 предприятиях США — табл. 2.1. По многочисленным данным зарубежной печати, на практике допустимы значения от 0.15 до 0.35 от цены изделия. Согласно [190], цена хранения в стабильных условиях назначается как общий по многим продуктам процент от цены и в среднем составляет 25 центов на доллар. Для мелких компаний, ведущих рискованные операции, она повышается вдвое. Военные организации долгое время использовали норматив 0.03 и изменили его лишь

в конце 60-х гг [89, с. 478]. Эту безусловно устаревшую информацию мы приводим, чтобы показать хроническую недооценку экономических факторов военными всего мира.

Расчеты отечественных экономистов показывали, что уровень затрат на складах, связанных с запасами, составляет около 25 коп. на рубль стоимости запаса, из них 60% — затраты на иммобилизацию, 16% — на основные производственные фонды, 24% — текущие расходы.

Т а б л и ц а 2.1

Издержки хранения на предприятиях США

Статьи издержек	Издержки в % от стоимости		
	минимальные	максимальные	средние
Процент на капитал	3.5	6.0	3.5
Налоги и страховка	0.5	3.0	1.0
Аренда помещения	0.7	2.5	2.5
Подъемно-транспортн. операции и учет	2.5	6.7	4.5
Порча и хищения	0.5	3.0	1.5
Устаревание	3.5	20.0	4.0
Итого	11.2	41.2	17.0

Оборачиваемость *производственных* запасов меньше и уровень механизации на них ниже. Для них называлась другая цифра — 40 коп. на рубль. На ГПЗ-1 по [81] издержки содержания производственных запасов составляли 18.25 копейки на рубль стоимости (без омертвления). Учитывая приведенное выше соотношение затрат от омертвления средств и прочих и уменьшение оборачиваемости, из этой цифры выводим как нижнюю оценку стоимости хранения 45 коп. на рубль.

При складской форме снабжения когда поставка организуется не непосредственно от производителя, а через промежуточный склад (оптовую базу), цена хранения входит в складскую наценку, добавляемую к оптовой цене поставляемой продукции.

2.4. Стоимость поставки

В функцию расходов на снабжение включается только та доля затрат на восполнение запасов, величина которой зависит от принятой организации снабжения. В частности, при минимизации затрат в единицу времени нет смысла учитывать составляющую расходов, пропорциональную объему перевозок. Ведь за достаточно длительное время суммарный объем перевозок окажется очень близким к суммарному спросу, т.е. к величине, не зависящей от организации поставок, и соответствующее слагаемое функции затрат не повлияет на выбор оптимальной стратегии.

В общем случае в стоимость поставки входят:

- постоянная составляющая;
- слагаемые, пропорциональные суммарному объему (весу) партии и/или количеству затребованных номенклатур либо (в неоднородном случае) сумма соответствующих фиксированных составляющих.

К постоянной составляющей относят расходы, связанные с самим актом оформления поставки (просмотр базы данных, выписка и оформление наряда, почтовые расходы). Размещение заказа на восполнение запаса является довольно дорогостоящим делом. Этот процесс может включать следующие операции:

- изучение картотеки движения запасов в целях определения объема заказа;
- диктовка, написание или печатание заказа на бланке;
- отправка заказа;
- принятие заказа по его прибытии;
- проверка соответствия товара сопроводительному документу;
- при необходимости — оформление претензии по несоответствию состава, количеству, повреждениям;
- распаковка и размещение на складе;
- сверка сопроводительного документа со счетом поставщика;
- проверка цен и срока платежа, указанных в счете;
- регистрация полученного заказа в картотеке контроля запасов;

- учет счетов поставщика;
- проверка и оплата счета;
- перенос соответствующих проводок из ведомости ежедневного учета закупок в общую ведомость;
- проверка согласованности данных о запасах и счетов, и т.д.

В ценах 1970-х гг. подобного вида «канцелярские расходы» составляли в розничной торговле до 5 руб., в оптовой 3–4 руб.

Если поставка связана с организацией специального рейса транспортного средства (самолета, автомобиля, железнодорожного вагона, цистерны, контейнера), то стоимость этого рейса включается в постоянную составляющую, а из транспортных затрат исключается слагаемое, пропорциональное объему перевозки. При полной загрузке, наоборот, из транспортных затрат оставляют только пропорциональное слагаемое. Коэффициенты пропорциональности определяются действующими тарифами в расчете на весовую единицу.

Составляющая функции затрат, пропорциональная числу номенклатур в поставке, появляется в процессе подготовки наряда. Для ее вычисления необходимо просуммировать стоимостные выражения трудозатрат персонала, связанных с поиском информации о наличном запасе по данной номенклатуре, определением объема заказа и расчетом стоимости партии, приемом пополнения и его размещением в хранилище. При резком различии в затратах, связанных с дополнительным заказом каждой новой номенклатуры, целесообразно усреднение не проводить и рассматривать эти затраты как добавочные фиксированные составляющие, выплачиваемые при заказе соответствующих номенклатур. Заработную плату работников, занятых содержанием запасов и организацией заказов, делят пропорционально их занятости каждым видом работ.

Необходимо иметь в виду, что организация поставок стоит денег как получателю, так и поставщику, причем последний отнесет связанные с этим собственные расходы на счет получателя. В дополнение к статьям затрат, аналогичных возникающим у получателя, к ним в производственных условиях добавятся усредненные затраты на запуск или на переналадку станков. Они вызваны необходимостью остановки производства для переналадки технологических линий (оплата наладчикам, подготовка новой документации, убытки от простоя оборудования за время переналадки и первоначального снижения его производительности).

Обычно переналадки связываются с работой металлообрабатывающих агрегатов и ассоциируются с дополнительными затратами на установку (смену) приспособлений. В поточных системах (например, мельницах и трубопроводах) при переходе на новый продукт в процессе переналадки выпускается смесь старого и нового продукта. Эта смесь либо идет в брак, либо реализуется как продукция низшего качества — следовательно, с потерей дохода.

Разумеется, получатель не в состоянии выполнить строгий расчет этих затрат — для этого у него нет исходных данных, но он должен запросить итоговые сведения у поставщика.

Можно учесть и непостоянство стоимости запуска (в зависимости от числа параллельных технологических линий). При хранении жидких и газообразных продуктов могут возникнуть ограничения, связанные с объемом буферных емкостей. Иногда «постоянная» составляющая рассматривается как ступенчатая функция целочисленного аргумента (например, количества вагонов, необходимого для реализации поставки). При этом суммарные затраты становятся разрывной функцией объема поставки, что существенно усложняет нахождение оптимального объема последней.

Одновременный заказ различной продукции может потребовать значительных денежных средств на заказ, а также слишком емких перевозочных средств.

2.5. Определение величины штрафа

Важнейшим элементом задачи управления запасами является регулирование дефицита. Времена тотального дефицита хорошо помнит старшее поколение граждан России. Некоторые деятели (А.И. Левин, [10, с. 9]) теоретически обосновывали закономерность превышения роста спроса над ростом предложения при социализме. Гарантированный сбыт продукции не способствовал ориентации на обеспечение реального спроса. Имеет основания мнение [4], что «в СССР дефицит создавался искусственно и был инструментом управления массами и кадрами. Недопроизводство позволяло всучить потребителю что угодно... Спекуляция и коррупция, бюрократизм и снижение качества, потеря профессионального мастерства — все это последствия дефицита сырья и готовой продукции, возможностей и гражданских прав, здравого смысла и профессиональной компетентности, свободы принятия реше-

ний и материальной заинтересованности в труде, конкуренции и свободы предпринимательства».

Надо признать, однако, что дефицит всего вышеперечисленного сохранился (более того — усугубился) в постперестроечное время. Кроме того, как будет показано ниже, полностью бездефицитное снабжение целесообразно лишь в предельном случае (при бесконечной цене штрафа). Такая ситуация является теоретической абстракцией, и определению реального значения упомянутой цены (хотя бы и приближенного) должно уделяться первостепенное внимание. На наш взгляд, одной из причин существования у потребителей политики излишнего запаса является отсутствие возможности и экономической необходимости сопоставления потерь от сверхнормативных запасов с ущербом из-за дефицита продукции — прежде всего производственно-технического назначения. Погрешность в цене штрафа сравнительно слабо сказывается на параметрах стратегии, но существенно влияет на величину операционных затрат.

Для установления цены штрафа используются два основных метода:

- определение прямых убытков и/или упущенной выгоды;
- «через обратную задачу».

Сущность первого метода состоит в детальном экономическом анализе последствий недостачи. В руководящих документах недавнего времени приводились стандартные штрафные санкции (8–12% от стоимости недопоставленной продукции [28]). НИИ ЦСУ рекомендовал 30 коп. на рубль дефицита, ЦЭМИ и НИИ МТС увеличивали их до 60 коп. Был указан предельный процент недопоставок, при превышении которого включалась шкала прогрессивного снижения премий руководителям предприятий-поставщиков — вплоть до полного лишения премий при любой недопоставке по позициям, утверждаемым Госпланом. Однако, как отмечали многие исследователи и практики, эти санкции отнюдь не компенсировали ущерба, причиненного потребителям. Если график поставок и фиксировался в договоре, то запаздывания в поставках сроком до 10 суток вообще не наказывались, а штрафы за изменение сроков поставок были слишком незначительны, чтобы существенно влиять на политику поставщика. К тому же поставщик отвечал перед вышестоящими органами за выполнение *суммарного* плана реализации, план же поставок конкретному потребителю (в особенности мелкому) часто оказывался невыполненным.

По указанным причинам для оценки последствий дефицита обычно используется подсчет убытков в конкретной ситуации. Если угроза дефицита ликвидируется экстренными мерами (например, поставки продукции самолетом вместо обычных железнодорожных, сверхурочные работы, раскомплектование изделий, изготовление недостающих предметов на неспециализированном оборудовании, использование заготовок из более дорогого материала или с избыточными припусками на обработку), то ценой штрафа является превышение стоимости экстренной поставки над обычной (включая стоимость оформления экстренной поставки, которая также значительно возрастает). Ущерб в результате реального простоя для цеха (предприятия) складывается из прямых убытков (оплата рабочим вынужденного простоя, сверхурочных работ на других производствах, отклонения от оптимальных размеров партии) и упущенной выгоды (процент на замороженные в незавершенном производстве вследствие дефицита оборотные средства). Если из-за недопоставок останавливается устройство приносящее известный доход или могущее его приносить (например, главный конвейер автозавода или ЭВМ, рабочее время которой сдается в аренду внешним заказчикам), то к сумме штрафа добавляется потерянный доход. Для большинства перечисленных ситуаций в ЦЭМИ АН СССР были разработаны соответствующие методики — см [77, с 233–234].

Второй способ оценки штрафа в значительной степени базируется на субъективном подходе, но представляется менее трудоемким. Его сущность заключается в том, что для исследуемой ситуации подбирается наиболее близкая модель управления запасами с известным решением. Далее считаем, что управление запасами происходило *оптимальным* образом, и, решая обратную задачу, находим, какой цене штрафа соответствуют принятые параметры стратегии. Так, если для «задачи газетчика» оптимальный запас S^* определяется по формуле (5.6.2), то цена штрафа может быть найдена по формуле

$$d = \frac{c + hF(S^*)}{1 - F(S^*)}$$

Для модели с постоянным спросом интенсивности λ оптимальный период между поставками вычисляется согласно (5.1.6), так что

$$d = \left(\frac{\lambda T^2}{2g} - \frac{1}{h} \right)^{-1}$$

Найденные величины могут использоваться при управлении запасами в

условиях, близких к условиям исходной задачи. Если же они окажутся в явном противоречии с разумным представлением о цене штрафа, необходимо пересмотреть параметры исходной стратегии, что позволит приблизиться к более реалистической оценке искомого параметра.

Некоторые задачи расчета штрафов при необеспеченности спроса на запасные части будут рассмотрены в разд. 8.9.

В условиях рынка важной составляющей издержек из-за потери требования является утрата предпочтения потребителей. Она может вызвать снижение доходов от будущих продаж, так как потребитель будет вести дело с другим поставщиком и может сообщить другим возможным клиентам о неудовлетворительном обслуживании.

Глава 3 .

Элементы теории вероятностей и очередей

В данной главе приводится (с минимумом доказательств) необходимый для дальнейшего набор сведений из указанных теорий, «спроецированных» на проблематику управления запасами.

3.1. Основные понятия теории распределений

3.1.1. Определения

Под *дискретным* событием понимается реализация значения неотрицательной случайной величины из множества $k = 0, 1, \dots$ — например, числа заявок на некоторое изделие в планируемый период. *Вероятность* его p_k определяется как отношение кратности n_k появления события к полному числу n произведенных независимых опытов при устремлении последнего к бесконечности (статистическое определение) или как степень уверенности в появлении этого события. *Непрерывная* случайная величина задается *плотностью распределения* $f(x)$. Вероятность того, что такая величина примет значения из

интервала $[x, x + \Delta x]$, приближенно равна $f(x)\Delta x$.

Функция распределения (ФР)

$$F(x) \stackrel{\text{def}}{=} \Pr\{k < x\}$$

для дискретных распределений определяется суммированием соответствующих вероятностей, а для непрерывных — интегрированием плотности. Соответственно вероятность конкретного значения в дискретном случае вычисляется как разность смежных значений ФР, а в непрерывном — плотность распределения — дифференцированием ФР.

Почти все используемые в данной книге распределения определены на полуоси $[0, \infty)$ или множестве натуральных чисел. Это обстоятельство в дальнейшем дополнительно не оговаривается.

Практические проблемы теории очередей и управления запасами обычно связываются с достижением высоких вероятностей своевременного решения задачи (обеспечения спроса), мало чувствительных к изменению параметров. В таких случаях удобнее работать с *дополнительными* функциями распределения (ДФР)

$$\bar{F}_T(t) \stackrel{\text{def}}{=} \Pr\{T > t\}.$$

Индекс имени случайной величины там, где это не ведет к неоднозначности, мы будем опускать.

3.1.2. Числовые характеристики

Для сравнения распределений удобно пользоваться их числовыми характеристиками. *Квантилью* q случайной величины по вероятности p называется решение уравнения $F(q) = p$. Практически используются квантили по вероятности 0.5 (медиана), 0.25 и 0.75 (квартили), реже — децили.

Начальный момент k -го порядка непрерывной случайной величины

$$\bar{x}_k = \int_0^{\infty} x^k f(x) dx, \quad k = 1, 2, \dots$$

Это же понятие для дискретной случайной величины определяется через взвешенное суммирование вероятностей. Для экономии места мы будем

записывать аналогичные выражения с помощью интеграла Стильтьеса, объединяющего дискретный и непрерывный случаи

$$\bar{x}_k = \int_0^{\infty} x^k dF(x), \quad k = 1, 2, \dots$$

Особую роль играет первый момент \bar{x}_1 , называемый математическим ожиданием. Его статистическим аналогом является среднее значение случайной величины.

В процессах обслуживания важную роль играет распределение *остатка* обслуживания при равномерном выборе контрольной точки на интервале его случайной длительности. Моменты остаточного распределения $\{\tilde{b}_k\}$ выражаются через моменты исходного $\{b_k\}$ согласно

$$\tilde{b}_k = \frac{b_{k+1}}{(k+1)b_1}, \quad k = 1, 2, \dots \quad (3.1.1)$$

В частности,

$$\tilde{b}_1 = \frac{b_2}{2b_1}. \quad (3.1.2)$$

Центральные моменты вычисляются относительно математического ожидания:

$$\bar{x}_k^c = \int_0^{\infty} (x - \bar{x}_1)^k dF(x), \quad k = 2, 3, \dots$$

Второй центральный момент $\bar{x}_2^c = D$ называется *дисперсией* распределения и служит мерой разброса случайной величины относительно ее среднего значения. В тех же целях используются *среднеквадратическое отклонение* $\sigma = \sqrt{D}$, размерность которого совпадает с размерностью случайной величины, а также безразмерный относительный показатель $v = \sigma/\bar{x}_1$ — *коэффициент вариации*. Обозначения $M[\cdot]$ и $D[\cdot]$ ниже будут использоваться как операторы вычисления математического ожидания и дисперсии соответственно.

3.1.3. Показательное распределение

Показательным (экспоненциальным) называется распределение с плотностью вида

$$f(t) = \mu e^{-\mu t}, \quad (3.1.3)$$

дополнительной функцией распределения

$$\bar{F}(t) = e^{-\mu t} \quad (3.1.4)$$

и начальными моментами

$$f_i = i!/\mu^i, \quad i = 1, 2, \dots \quad (3.1.5)$$

Для определенности обсудим процесс обслуживания. Условная плотность распределения длительности остатка обслуживания, которое продолжается уже в течение τ , описывается формулой

$$\tilde{f}(t|\tau) = \frac{f(t+\tau)}{\bar{F}(\tau)} = \frac{\mu e^{-\mu(t+\tau)}}{e^{-\mu\tau}} = \mu e^{-\mu t}$$

независимо от уже истекшей длительности обслуживания (!!!). Говорят, что показательное распределение обладает *марковским* свойством — отсутствием последствия (памяти). Параметр потока обслуживаний с функцией распределения $B(t)$

$$\mu(\tau) \stackrel{\text{def}}{=} \lim_{\Delta t \rightarrow 0} \frac{B(\tau + \Delta t) - B(\tau)}{\bar{B}(\tau)\Delta t} = -\frac{\bar{B}'(\tau)}{\bar{B}(\tau)} \quad (3.1.6)$$

в случае показательного распределения длительности равен μ и не зависит от уже истекшего времени τ . Соответственно вероятность завершения обслуживания на малом интервале длины Δt

$$\Pr_e\{t, t + \Delta t\} = \mu\Delta t + o(\Delta t)$$

не зависит от положения этого интервала на оси времени.

Отмеченные уникальные свойства показательного распределения делают его исключительно удобным в аналитических выкладках, связанных с описанием процессов обслуживания

3.2. Аппроксимация распределений

Выравнивание статистических распределений, т.е. подбор теоретических зависимостей, описывающих фактически наблюдавшиеся данные, обычно проводится при условии сохранения значений интегральных числовых характеристик распределений, аккумулирующих его основные свойства. В качестве таких характеристик чаще всего выступают начальные моменты. Как отмечается в [23, с 127],

«можно ожидать, что если два распределения имеют некоторое число одинаковых моментов, то они в какой-то степени схожи и аппроксимируют друг друга. Обоснованием этого суждения является совпадение их аппроксимаций по методу наименьших квадратов при конечном размахе и разложений по ортогональным многочленам — при бесконечном ... Практически аппроксимация такого рода оказывается очень хорошей, даже если совпадают только первые три или четыре момента».

В той же книге приводятся необходимые и достаточные условия возможности построения по моментам *характеристической функции*, однозначно определяющей распределение во всех точках его непрерывности. Эти условия, как правило, выполняются, хотя известны и отрицательные примеры (логарифмически нормальное распределение).

При всех своих недостатках (чаще потенциальных, чем реальных) метод моментов остается наиболее удобным в работе и обычно применяемым методом выравнивания распределений. Количество сохраняемых моментов рассматривается как порядок аппроксимации. Оно равно числу свободных параметров теоретической кривой. Отметим также, что

- наращивание числа учитываемых моментов не всегда сопровождается монотонным улучшением качества аппроксимации;
- любое число учтенных моментов не может обеспечить хорошего согласия на «хвосте» распределения — в области малых значений ДФР (по опыту расчетов — менее 0.01);
- относительная погрешность высших моментов, найденных статистическими методами, быстро возрастает по мере увеличения порядка моментов.

Поэтому практически ограничиваются выравниванием не менее двух (см. разд. 9 2), но не более четырех моментов.

Существуют методы, позволяющие автоматически выбрать тип выравнивающего распределения в заданном семействе (например, одну из кривых К. Пирсона выбирают по четырем моментам — см. [56]). Критерии согласия (χ^2 Р Фишера, $n\omega^2$ Н.В. Смирнова, А.Н. Колмогорова) статистических и теоретических распределений с несколькими общими моментами мало чувствительны к виду распределений. Это позволяет в классе функций с заданными несколькими моментами выбирать дающие те или иные вычислительные преимущества — например, допускающие

эффективное вычисление ДФР или облегчающие расчет функционалов специального вида. Именно такие распределения (в первую очередь показательное и родственные ему) и будут использоваться в дальнейшем.

При выборе типа аппроксимации важно знать степень отличия исходного распределения от показательного при равенстве средних значений. В первом приближении мерой отличия может служить коэффициент вариации (для показательного закона $v = 1$). Однако с целью единообразного представления отличий в высших моментах удобно ввести набор *коэффициентов немарковости*

$$\xi_i = f_i/f_1^i - i!, \quad i = 2, 3, \dots \quad (3.2.1)$$

С помощью (3.1.5) легко проверить, что для показательного распределения все они равны нулю.

3.3. Типовые распределения

3.3.1. Распределения фазового типа

Пусть процесс проходит систему фаз с показательно распределенной задержкой в каждой из них.

Распределение Эрланга r -го порядка имеет плотность

$$f(t) = \frac{\mu(\mu t)^{r-1}}{(r-1)!} e^{-\mu t}, \quad (3.3.1)$$

дополнительную функцию распределения

$$\bar{F}(t) = \sum_{i=0}^{r-1} \frac{(\mu t)^i}{i!} e^{-\mu t} \quad (3.3.2)$$

и моменты

$$f_k = r(r+1) \dots (r+k-1)/\mu^k, \quad k = 1, 2, \dots \quad (3.3.3)$$

Оно предполагает последовательное прохождение r фаз (см. рис. 3.1).

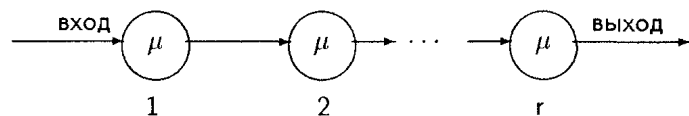


Рис. 3.1. Эрланговское распределение

Дисперсия эрланговского распределения $D = r/\mu^2$.

Распределение Эрланга — двухпараметрическое, причем r должно быть целым. Это обстоятельство определяет следующую схему подбора параметров распределения Эрланга:

- 1) найти $r = [f_1^2/D + 0.5]$ (ближайшее целое к указанному отношению);
- 2) вычислить вещественный параметр $\mu = r/f_1$.

При этом строго выравнивается первый момент и приближенно — второй.

Гиперэкспоненциальное распределение. Представление ДФР в виде

$$\bar{F}(t) = \sum_{i=1}^k y_i e^{-\mu_i t} \quad (3.3.4)$$

позволяет считать исходный процесс проходящим одну из n альтернативных фаз. Здесь $\{y_i\}$, $0 \leq y_i \leq 1$, $\sum_{i=1}^n y_i = 1$ интерпретируются как вероятности выбора i -й фазы, а $\{\mu_i\}$ — как параметры показательного распределения времени пребывания в ней. Диаграмма гиперэкспоненциального распределения приведена на рис. 3.2.

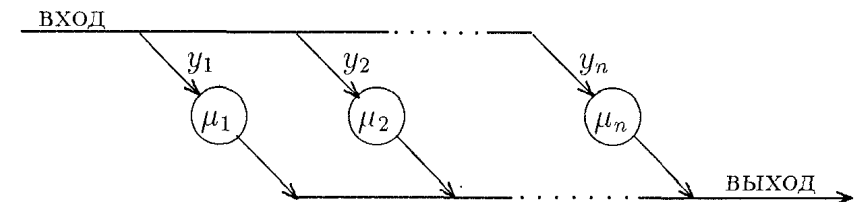


Рис. 3.2. Гиперэкспоненциальное распределение

Среди названных величин могут быть комплексные (попарно сопряженные), что подчеркивает фиктивный характер расщепления процесса на фазы. Допустимость таких параметров была впервые отмечена Д. Коксом в 1955 г.

Использование формулы для моментов показательного распределения приводит к системе уравнений относительно параметров аппрок-

симации

$$\begin{aligned} y_1 + y_2 + \dots + y_n &= 1, \\ y_1/\mu_1 + y_2/\mu_2 + \dots + y_n/\mu_n &= f_1, \\ y_1/\mu_1^2 + y_2/\mu_2^2 + \dots + y_n/\mu_n^2 &= f_2/2!, \\ \dots &\dots \\ y_1/\mu_1^i + y_2/\mu_2^i + \dots + y_n/\mu_n^i &= f_i/i!, \quad i = \overline{0, 2n-1}. \end{aligned} \quad (3.3.5)$$

Здесь f_i — i -й момент исходного распределения. Введем обозначения $x_i = \mu_i^{-1}$; $\tilde{f}_i = f_i/i!$. Тогда система (3.3.5) переходит в

$$\sum_{j=1}^n y_j x_j^i = \tilde{f}_i, \quad i = \overline{0, 2n-1}. \quad (3.3.6)$$

Эта система решается методом, описанным в книге Р.В. Хемминга [90, с. 139–141].

Численные эксперименты показали, что особыми случаями для H_n -аппроксимации являются любые распределения Эрланга порядка $k \leq n$. Это вполне естественно, поскольку последние предполагают последовательную систему фаз, тогда как гиперэкспонента — параллельную.

При H_2 -аппроксимации задача подбора ее параметров может быть сведена к решению квадратного уравнения. Детальный анализ показывает, что при H_2 -аппроксимации обсуждаемых ниже гамма-распределений с параметром $0 < \alpha < 1$ все параметры вещественны и положительны. В случае $1 < \alpha < 2$ одна из «вероятностей» $\{y_j\}$ будет отрицательной, а другая превысит единицу. Как показали вычислительные эксперименты, столь парадоксальные промежуточные результаты не мешают успешному расчету систем обслуживания. Наконец, при $\alpha > 2$ параметры H_2 -аппроксимации будут комплексными сопряженными. Эта возможность порождает комплексные вероятности состояний рассчитываемых систем и соответственно удваивает расход памяти при машинном счете.

3.3.2. Гамма-распределение

Гамма-распределение имеет плотность

$$f(t) = \frac{\mu(\mu t)^{\alpha-1}}{\Gamma(\alpha)} e^{-\mu t}, \quad (3.3.7)$$

где $\Gamma(\alpha)$ — гамма-функция со свойствами

$$\begin{aligned} \Gamma(\alpha) &\stackrel{\text{def}}{=} \int_0^{\infty} x^{\alpha-1} e^{-x} dx, \\ \Gamma(1) &= 1, \\ \Gamma(\alpha+1) &= \alpha\Gamma(\alpha). \end{aligned} \quad (3.3.8)$$

Оно является обобщением распределения Эрланга на случай нецелых $\alpha > 0$. Поэтому его нельзя считать фазовым. На рис. 3.3 показано несколько типичных графиков гамма-плотности с одинаковым средним $\alpha/\mu = 1$ и различными α . Случай $\alpha = 1$ соответствует показательному закону.

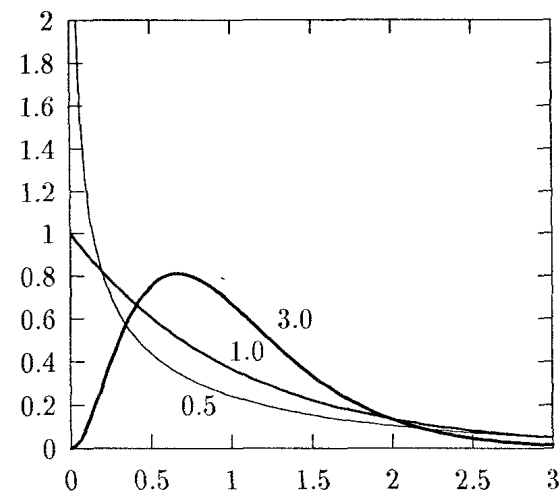


Рис. 3.3. Гамма-плотности

Моменты Γ -распределения вычисляются по формулам типа (3.3.3) с заменой r на α . Его дисперсия $D = f_2 - f_1^2 = \alpha/\mu^2$, а коэффициент вариации $v = 1/\sqrt{\alpha}$.

Снятие требования целочисленности α позволяет обеспечить точное выравнивание двух моментов, если положить

$$\mu = f_1/D, \quad \alpha = \mu f_1. \quad (3.3.9)$$

Для выравнивания большего числа моментов приходится применять гамма-плотность с поправочным многочленом. Удобно подобрать α и μ

согласно (3.3.9), после чего нетрудно сформировать систему линейных уравнений относительно коэффициентов поправочного многочлена¹.

3.3.3. Нормальное распределение

На рис. 3.4 приведены графики играющего важнейшую роль в теории вероятностей и математической статистике *нормального* распределения с плотностью $f(x) = \exp[-(x-a)^2/2\sigma^2]/\sigma\sqrt{2\pi}$ при общем среднем $a = 3$ и различных σ . Параметр σ — это среднеквадратическое отклонение (квадратный корень из дисперсии). ФР нормального распределения выражается через функцию Лапласа

$$\Phi(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \quad (3.3.10)$$

согласно

$$F(x) = \frac{1}{2} \left[\Phi \left(\frac{x-a}{\sigma\sqrt{2}} \right) + 1 \right]. \quad (3.3.11)$$

Рисунок наглядно иллюстрирует концентрацию значений случайной величины в окрестности среднего значения (точнее, математического ожидания) при уменьшении среднеквадратического отклонения. Площади под каждой кривой одинаковы, поскольку для любой плотности $\int_{-\infty}^{\infty} f(x) dx = 1$. Отличия нормального распределения от других, имеющих те же среднее и дисперсию, можно охарактеризовать асимметрией $s = \mu_3/\sigma^3$ и эксцессом $\varepsilon = \mu_4/\sigma^4 - 3$. В этих формулах $\{\mu_i\}$ суть *центральные* моменты, вычисленные относительно среднего.

¹Эта технология ниже рассматривается для распределения Вейбулла.

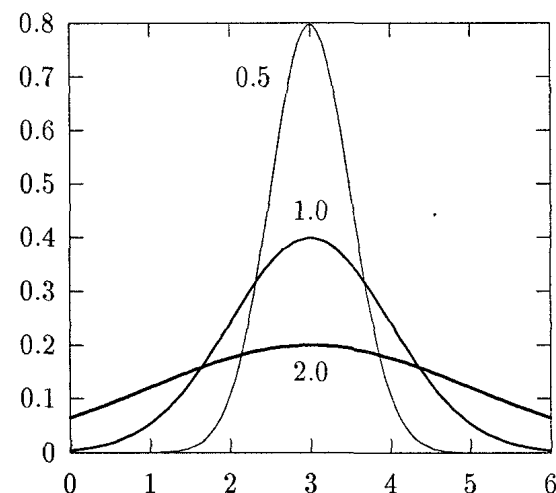


Рис. 3.4. Нормальные плотности

3.3.4. Распределение Вейбулла

Основным показателем качества функционирования многих СМО является ДФР времени пребывания заявки в системе, получаемая в виде таблицы для заданных значений аргумента. Неинтегрируемость гамма-плотности вынуждает искать для вычисления ДФР по заданным моментам другую аппроксимацию. Удобно применение распределения Вейбулла

$$\bar{F}(t) = e^{-t^k/W}. \quad (3.3.12)$$

Его моменты

$$f_j = \left(\frac{j}{k}\right) \Gamma\left(\frac{j}{k}\right) W^{j/k} = W^{j/k} \Gamma(1 + j/k), \quad j = 1, 2, \dots, \quad (3.3.13)$$

а отношение

$$\alpha = f_2/f_1^2 = 2k\Gamma(2/k)/\Gamma^2(1/k) = 2\Gamma(2u)/(u\Gamma^2(u)), \quad (3.3.14)$$

где $u = 1/k$. Воспользовавшись формулой удвоения аргумента гамма-функции, можно переписать (3.3.14) в виде

$$\alpha = \frac{2^{2u}\Gamma(u)\Gamma(u+1/2)}{u\sqrt{\pi}\Gamma^2(u)} = \frac{2^{2u}\Gamma(u+1/2)}{\sqrt{\pi}\Gamma(u+1)},$$

откуда следует обеспечивающая быстро сходящийся итерационный процесс уточнения u формула

$$u_i = \frac{1}{2 \ln 2} \ln \frac{\alpha \sqrt{\pi} \Gamma(u_{i-1} + 1)}{\Gamma(u_{i-1} + 1/2)}, \quad i = 1, 2, \dots \quad (3.3.15)$$

Начальное приближение

$$u_0 = \ln 2\alpha / (2 \ln 2). \quad (3.3.16)$$

При необходимости учета более двух моментов можно применить ту же базовую аппроксимацию с поправочным многочленом:

$$\bar{F}(t) = e^{-t^k/W} \sum_{i=0}^N g_i t^i. \quad (3.3.17)$$

Моменты этого распределения

$$f_j = \frac{j}{k} \sum_{i=0}^N g_i \Gamma\left(\frac{i+j}{k}\right) W^{(i+j)/k}, \quad j = \overline{0, N}. \quad (3.3.18)$$

Теперь ясен алгоритм подбора параметров аппроксимации (3.3.17) по моментам $\{f_j\}$, $j = \overline{1, N}$:

- 1) Вычислить $\alpha = f_2/f_1^2$.
- 2) Определить u_0 согласно (3.3.16).
- 3) Решить уравнение (3.3.15) методом итераций.
- 4) Вычислить $k = 1/u$, $W = (f_1/\Gamma(u+1))^k$.
- 5) Сформировать систему (3.3.18) линейных алгебраических уравнений относительно $\{g_i\}$.
- 6) Решить эту систему любым стандартным методом.

3.3.5. Дискретные распределения

Дискретные распределения обычно задаются набором вероятностей появления допустимых значений.

Распределение Пуассона описывает вероятности редких событий:

$$p_x = \frac{(\lambda t)^x}{x!} e^{-\lambda t}, \quad x = 0, 1, \dots \quad (3.3.19)$$

Здесь λ — интенсивность потока редких событий, а t — время наблюдения. Для этого распределения

$$M[X] = D[X] = \lambda t. \quad (3.3.20)$$

Биномиальное распределение получается при повторении независимых испытаний, каждое из которых имеет только два возможных исхода. Без потери общности эти исходы можно классифицировать как «успех» и «неудачу», наблюдаемые с вероятностями ρ и $1 - \rho$ соответственно. Вероятность появления x успехов в n испытаниях равна

$$p_x = \binom{n}{x} \rho^x (1 - \rho)^{n-x}, \quad x = \overline{0, n}. \quad (3.3.21)$$

Для биномиального распределения

$$M[X] = n\rho, \quad D[X] = n\rho(1 - \rho). \quad (3.3.22)$$

Отрицательное биномиальное распределение

$$p_x = \binom{x+n-1}{n-1} \rho^n (1 - \rho)^x, \quad x = 0, 1, \dots \quad (3.3.23)$$

дает вероятности того, что потребуется ровно $x + n$ испытаний для получения n успехов. Здесь

$$M[X] = n(1 - \rho)/\rho, \quad D[X] = n(1 - \rho)/\rho^2. \quad (3.3.24)$$

Геометрическое распределение является его частным случаем ($n = 1$). Здесь

$$p_x = \rho(1 - \rho)^x, \quad x = 0, 1, \dots, \quad (3.3.25)$$

а числовые характеристики

$$M[X] = (1 - \rho)/\rho, \quad D[X] = (1 - \rho)/\rho^2. \quad (3.3.26)$$

3.4. Преобразования распределений

Для некоторых вероятностных расчетов удобно пользоваться преобразованиями распределений. В случае непрерывных распределений

это преобразование Лапласа — Стильбеса (ПЛС)

$$\varphi(s) = \int_0^{\infty} e^{-sx} dF(x). \quad (3.4.1)$$

Моменты распределения выражаются через ПЛС согласно

$$f_k = (-1)^k \varphi(s)^{(k)}|_{s=0}, \quad k = 1, 2, \dots \quad (3.4.2)$$

При известном правиле вычисления ПЛС дифференцирование может быть выполнено аналитически либо с помощью многократного численного дифференцирования в нуле на основе полученной алгоритмически таблицы. Соответствующая процедура входит в состав пакета МОСТ — см. разд. 3.17.

Для дискретных распределений вводится производящая функция

$$P(z) = \sum_{i=0}^{\infty} z^i p_i, \quad 0 \leq z \leq 1. \quad (3.4.3)$$

Производящая функция пуассоновских вероятностей

$$P(z) = e^{-\lambda t(1-z)}. \quad (3.4.4)$$

Для отрицательного биномиального распределения производящая функция

$$P(z) = \left[\frac{\rho}{1 - (1 - \rho)z} \right]^n. \quad (3.4.5)$$

В задачах теории очередей и управления запасами часто приходится выполнять свертку распределений, т.е. находить распределение суммы независимых случайных величин. Примером такой задачи служит построение распределения времени пребывания заявки в системе $V(t)$ по распределениям $W(t)$ времени ожидания и $B(t)$ чистой длительности обслуживания. В терминах ПЛС

$$\nu(s) = \omega(s)\beta(s) \quad (3.4.6)$$

(ПЛС свертки равно произведению ПЛС составляющих). Свертка может быть выполнена непосредственно в моментах на основе символического разложения

$$v^k = (w + b)^k, \quad (3.4.7)$$

в котором после развертывания бинома показатели степени переводятся в индексы соответствующих моментов. Простота этих соотношений — серьезный аргумент в пользу применения метода моментов.

Производящая функция суммы дискретных случайных величин (т.е. их свертки) равна произведению производящих функций слагаемых. Если производящая функция распределения слагаемого есть $P_1(z)$, а производящая функция числа слагаемых — $P_2(z)$, то производящая функция суммы

$$P(z) = P_2(P_1(z)). \quad (3.4.8)$$

Свертка дискретных распределений в моментах выполняется так же, как для непрерывных распределений. Для получения отдельных вероятностей свертки применяется формула

$$p_k = \sum_{i=0}^k p'_i p''_{k-i}. \quad (3.4.9)$$

3.5. Задачи теории очередей

Теория очередей в русскоязычной литературе чаще именуется теорией массового обслуживания (ТМО) и здесь рассматривается потому, что является основой математического аппарата для расчета восстанавливаемого ЗИПа. Термин «массовое» предполагает многократную повторяемость ситуаций в том или ином смысле (много прибывших в систему и обслуженных заявок, большое число находящихся в эксплуатации аналогичных систем) и статистическую устойчивость картины. Выводы и рекомендации, получаемые методами ТМО, применимы лишь при наличии одного или обоих названных факторов.

Модель задачи массового обслуживания включает в себя:

- поток заявок;
- каналы обслуживания;
- организацию очереди и дисциплину обслуживания;
- показатели эффективности.

Дадим содержательное описание и перечень возможных вариантов задания этих элементов и установим математическое содержание соответствующих понятий.

3.6. Поток заявок

3.6.1. Основные определения

Входящий поток задан, если для каждого номера n задано совместное распределение интервалов $\{z_1, z_2, \dots, z_n\}$ между смежными заявками: $z_i = t_i - t_{i-1}$, $t_0 = 0$. Если длины упомянутых интервалов независимы в совокупности, то зависимость момента прибытия очередной заявки t_i от предыстории процесса сводится к фиксации t_{i-1} (поток с *ограниченным последствием*). Такой поток может быть задан функциями распределения $A_i(t) \stackrel{\text{def}}{=} \text{Pr}\{z_i < t\}$. Если $A_i(t) = A(t)$ при всех $i \geq 1$, поток называется *рекуррентным*. Это наиболее общий случай из реально используемых в расчетах.

Обозначим $Z_k(t, \tau)$ событие, состоящее в появлении ровно k заявок на полуинтервале $[t, t + \tau)$. Свойства потока заявок могут быть охарактеризованы через вероятности $\{p_k(t, \tau)\}$ таких событий. Поток называется *стационарным*, если эти вероятности определяются только длиной интервала τ и не зависят от его положения на оси времени (переменная t). Поток называется потоком *без последствия*, если события $Z_{k_1}(t_1, \tau_1)$ и $Z_{k_2}(t_2, \tau_2)$ для неперекрывающихся интервалов времени независимы. Поток считается *ординарным*, если вероятность появления на элементарном участке $[t, t + \Delta t)$ более чем одного события имеет порядок малости $o(\Delta t)$, т.е. выше Δt . Поток, одновременно удовлетворяющий всем перечисленным требованиям, именуется *простейшим*.

Указанные свойства наблюдаются часто, но не всегда. Например, интенсивность потока заявок может зависеть от времени суток или года, заявки могут поступать группами постоянного или случайного объема. Сразу же заметим, что в случае неординарного потока требований в виде «пачек» постоянного объема удобнее переходить к ординарному потоку групповых заявок.

Для определения количественных характеристик потоков введем вероятность $\Pi_1(t, \tau)$ появления хотя бы одного требования за интервал длины τ , прилегающий справа к t . Параметр потока вычисляется по формуле

$$\lambda(t) \stackrel{\text{def}}{=} \lim_{\tau \rightarrow 0} \frac{\Pi_1(t, \tau)}{\tau}. \quad (3.6.1)$$

3.6.2. Число событий на фиксированном интервале

Известно, что интервалы между требованиями стационарного ординарного потока без последствия (простейшего потока) подчиняются показательному распределению. Его параметр $\lambda = 1/a_1$, где $a_1 = a$ — средний интервал между требованиями.

Для подсчета распределения числа требований простейшего потока за время t выполним свертку показательных распределений. Их свертка k -го порядка есть распределение Эрланга того же порядка (см. разд. 3.3.1). Вероятность появления на интервале длины t ровно k заявок равна $\bar{F}_{k+1}(t) - \bar{F}_k(t)$. Подставляя в это выражение формулу (3.3.2), убеждаемся, что вероятность прихода за $[0, t)$ ровно k требований

$$p_k(t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \quad k = 0, 1, \dots \quad (3.6.2)$$

Эта формула задает распределение Пуассона с параметром λt , отчего простейший поток называют также *пуассоновским*.

В теории массового обслуживания простейший поток занимает особое место по следующим причинам:

- 1) Сумма конечного числа независимых простейших потоков образует простейший поток с интенсивностью, равной сумме интенсивностей составляющих.
- 2) Сумма n независимых стационарных потоков с ограниченным последствием при условии малой интенсивности составляющих в сравнении с суммарной интенсивностью при $n \rightarrow \infty$ сходится к простейшему потоку.
- 3) Случайное прореживание произвольного стационарного ординарного потока с ограниченным последствием, т.е. выбрасывание каждого очередного требования независимо с некоторой вероятностью, при увеличении вероятности выбрасывания приближает поток к простейшему.
- 4) Вероятность наступления события простейшего (и только простейшего) потока на малом интервале длины Δt пропорциональна длине этого интервала и не зависит от его положения на оси времени, что дает колоссальные расчетные преимущества.

Наряду с простейшим потоком часто рассматривают так называемый *примитивный* поток, связанный с понятием о замкнутых системах

массового обслуживания. В таких системах имеется конечное число R источников заявок, причем суммарное количество действующих источников и необслуженных заявок постоянно. Если в системе находится k заявок, то входящий поток считается простейшим с мгновенной интенсивностью $\lambda(R - k)$, где λ — интенсивность простейшего потока в расчете на один источник. При $k = R$ поток заявок прерывается.

3.6.3. Число событий на случайном интервале

Пусть λ — параметр входящего потока. Тогда вероятность появления ровно j событий потока за случайное время, подчиненное распределению $B(t)$, равна

$$q_j = \int_0^{\infty} \frac{(\lambda t)^j}{j!} e^{-\lambda t} dB(t), \quad j = 0, 1, \dots \quad (3.6.3)$$

Указанные вероятности играют важную роль в расчете сложных СМО, и необходимо уметь эффективно их вычислять для возможно более широкого круга распределений. Специальный интерес вызывает вероятность q_0 , которая может рассматриваться как ПЛС от распределения $B(t)$ с параметром λ .

При равномерном распределении случайного времени на отрезке $[a - l, a + l]$ искомые вероятности

$$q_j = \frac{1}{2\lambda l} \left[\sum_{i=0}^j \frac{[\lambda(a-l)]^i}{i!} e^{-\lambda(a-l)} - \sum_{i=0}^j \frac{[\lambda(a+l)]^i}{i!} e^{-\lambda(a+l)} \right], \quad j = 0, 1, \dots$$

В частности,

$$q_0 = (e^{-\lambda(a-l)} - e^{-\lambda(a+l)}) / (2\lambda l).$$

Если t распределено на отрезке $[a - l, a + l]$ по треугольному закону,

то

$$\begin{aligned} q_j &= \left\{ \left[\frac{j+1}{\lambda} - (a-l) \right] \sum_{i=0}^j \frac{[\lambda(a-l)]^i}{i!} e^{-\lambda(a-l)} \right. \\ &+ \left[\frac{j+1}{\lambda} - (a+l) \right] \sum_{i=0}^j \frac{[\lambda(a+l)]^i}{i!} e^{-\lambda(a+l)} \\ &+ \frac{1}{\lambda} \frac{[\lambda(a-l)]^{j+1}}{j!} e^{-\lambda(a-l)} + \frac{1}{\lambda} \frac{[\lambda(a+l)]^{j+1}}{j!} e^{-\lambda(a+l)} \\ &\left. + 2 \left(a - \frac{j+1}{\lambda} \right) \sum_{i=0}^j \frac{(\lambda a)^i}{i!} e^{-\lambda a} - \frac{2}{\lambda} \frac{(\lambda a)^{j+1}}{j!} e^{-\lambda a} \right\} / (\lambda l^2), \\ & \quad j = 0, 1, \dots \end{aligned}$$

Начальный коэффициент

$$q_0 = \frac{1}{(\lambda l)^2} \left[e^{-\lambda(a-l)} + e^{-\lambda(a+l)} - 2e^{-\lambda a} \right].$$

Для гамма-распределения с параметром формы r

$$\begin{aligned} q_j &= \int_0^{\infty} \frac{(\lambda t)^j}{j!} e^{-\lambda t} \frac{\mu(\mu t)^{r-1}}{\Gamma(r)} e^{-\mu t} dt = \frac{\lambda^j \mu^r}{j! \Gamma(r)} \int_0^{\infty} t^{r+j-1} e^{-(\lambda+\mu)t} dt \\ &= \frac{\lambda^j \mu^r}{j! (\lambda + \mu)^{r+j} \Gamma(r)} \int_0^{\infty} u^{r+j-1} e^{-u} du \\ &= \left(\frac{\mu}{\lambda + \mu} \right)^r \left(\frac{\lambda}{\lambda + \mu} \right)^j \frac{\Gamma(r+j)}{j! \Gamma(r)}, \quad j = 0, 1, \dots \end{aligned} \quad (3.6.4)$$

Получим рекуррентные формулы вычисления $\{q_j\}$ при времени обслуживания, подчиненном гамма-распределению. Прежде всего, из (3.6.4) следует $q_0 = (\mu/(\lambda + \mu))^r$. Далее,

$$\frac{q_j}{q_{j-1}} = \frac{\lambda}{\lambda + \mu} \frac{\Gamma(r+j) \cdot (j-1)!}{j! \Gamma(r+j-1)} = \frac{\lambda}{\lambda + \mu} \frac{r+j-1}{j}.$$

Итак, при гамма-распределении

$$\begin{aligned} q_0 &= \left(\frac{\mu}{\lambda + \mu} \right)^r, \\ q_j &= q_{j-1} \frac{\lambda}{\lambda + \mu} \frac{r+j-1}{j}, \quad j = 1, 2, \dots \end{aligned} \quad (3.6.5)$$

Частным случаем гамма-распределения при $r = 1$ является показательное распределение. При этом

$$q_j = \frac{\mu}{\lambda + \mu} \left(\frac{\lambda}{\lambda + \mu} \right)^j, \quad j = 0, 1, \dots \quad (3.6.6)$$

Наконец, для гамма-плотности с поправочным многочленом, записываемой аналогично формуле (3.3.17) для ДФР Вейбулла, искомые вероятности

$$q_j = \left(\frac{\mu}{\lambda + \mu} \right)^\alpha \left(\frac{\lambda}{\lambda + \mu} \right)^j \frac{1}{j!} \sum_{i=0}^N \frac{g_i}{(\lambda + \mu)^i} \frac{\Gamma(\alpha + j + i)}{\Gamma(\alpha)}, \quad j = 0, 1, \dots \quad (3.6.7)$$

Расчет $\{q_j\}$ и здесь может быть организован рекуррентно, но по более сложной схеме.

3.6.4. Случайное прореживание потоков

Пусть в рекуррентном входящем потоке с ПЛС интервалов между заявками $\alpha(s)$ каждая заявка сохраняется в потоке с вероятностью z независимо от остальных заявок. Тогда для просеянного потока ПЛС распределения интервалов между заявками

$$\varphi(s) = \sum_{n=1}^{\infty} z(1-z)^{n-1} \alpha^n(s) = \frac{z\alpha(s)}{1 - (1-z)\alpha(s)}. \quad (3.6.8)$$

Моменты результирующего распределения

$$f_1 = a_1/z, \quad f_k = [a_k + (1-z) \sum_{i=1}^{k-1} \binom{k}{i} f_i a_{k-i}] / z, \quad k = 2, 3, \dots \quad (3.6.9)$$

В частности,

$$f_2 = [a_2 + 2(1-z)f_1 a_1] / z.$$

Второй коэффициент немарковости ξ_2 для просеянного потока

$$\xi_2^{(f)} = z\xi_2^{(a)}. \quad (3.6.10)$$

При $z \rightarrow 0$ коэффициент немарковости стремится к нулю, т.е. результирующий поток сходится к простейшему. Простейший поток при случайном прореживании остается простейшим — меняется только его интенсивность.

3.6.5. Регулярное прореживание потоков

«Справедливое» (циклическое) распределение поступающих заявок между n обслуживающими устройствами порождает для каждого из них поток с регулярным прореживанием (остается n -я заявка исходного потока). Соответственно распределение интервалов между заявками оказывается n -кратной сверткой исходного распределения, а его ПЛС — n -й степенью исходного.

Получить моменты просеянного таким образом потока можно последовательной сверткой в моментах или численным дифференцированием упомянутой ПЛС в нуле. Однако зависимость $\xi_2(n)$ можно вывести из элементарных соображений. Поскольку средние и дисперсии интервалов между оставшимися заявками суммируются, а второй момент равен квадрату первого плюс дисперсия, имеем

$$\begin{aligned} \xi_2(n) &= [f_1^2(n) + D(n)] / f_1^2(n) - 2! \\ &= nd / (na_1)^2 + 1 - 2 = (d/a_1^2) / n - 1 = v^2 / n - 1. \end{aligned}$$

Здесь v — коэффициент вариации исходного распределения. Очевидно, предел $\lim_{n \rightarrow \infty} \xi_2(n) = -1$ и совпадает со значением ξ_2 для вырожденного распределения, так что регулярно просеиваемый поток в указанных условиях приближается к детерминированному.

3.6.6. Суммирование потоков

Рассмотрим предложенную в [70] методику суммирования двух рекуррентных потоков. Смысл операции суммирования иллюстрирует рис. 3.5.

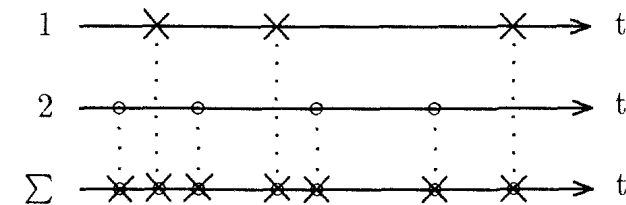


Рис. 3.5. Схема суммирования потоков

Момент появления очередной заявки суммарного потока — это минимум из моментов появления ближайших заявок составляющих. Если предыдущей была заявка первого потока, то распределение времени ожидания

заявки второго заменяется на соответствующее остаточное распределение, и наоборот. Частота выбора вариантов определяется удельным весом заявок каждого типа в суммарном потоке, т.е. отношениями $\{\lambda_i/\Lambda\}$. Итак, ДФР интервалов между заявками суммарного потока

$$\bar{A}_\Sigma(t) = \frac{\lambda_1}{\Lambda} \bar{A}_1(t) \bar{A}_2^*(t) + \frac{\lambda_2}{\Lambda} \bar{A}_2(t) \bar{A}_1^*(t). \quad (3.6.11)$$

Эта формула особенно легко реализуется при гиперэкспоненциальной аппроксимации составляющих распределений, для которых случайная модификация также дает гиперэкспоненту. Перепишем выражение для моментов распределения в виде

$$f_k = \int_0^\infty t^k f(t) dt = k \int_0^\infty t^{k-1} \bar{F}(t) dt.$$

Теперь составляющие моментов распределения с ДФР типа (3.3.4) могут быть вычислены согласно

$$\begin{aligned} f_k &= k \int_0^\infty t^{k-1} \left(\sum_{i=1}^2 y_i e^{-\mu_i t} \right) \left(\sum_{j=1}^2 u_j e^{-\lambda_j t} \right) dt \\ &= k \sum_{i=1}^2 y_i \sum_{j=1}^2 u_j \int_0^\infty t^{k-1} e^{-(\mu_i + \lambda_j)t} dt = \sum_{i=1}^2 \sum_{j=1}^2 y_i u_j k! / (\mu_i + \lambda_j)^k. \end{aligned} \quad (3.6.12)$$

Последовательное суммирование любого числа потоков можно выполнить, аппроксимируя H_2 -законами распределения интервалов суммарного потока и очередного слагаемого. При большом числе слагаемых имеет смысл организовать суммирование по схеме двоичного дерева.

3.7. Процесс обслуживания

Системы обслуживания по числу установленных устройств делятся на одно- и многоканальные. Количество требований, одновременно могущих находиться на обслуживании, не превышает числа каналов n . При очень большом числе каналов можно считать $n = \infty$. Каналы могут быть однородными, специализированными по типам заявок, различающимися интенсивностью обслуживания и т.п.

В случае показательного распределения нетрудно найти вероятность окончания хотя бы одного обслуживания для многоканальной системы. Пусть обслуживанием заняты k каналов. Тогда каждый из них за малый интервал длины Δt закончит обслуживание с вероятностью $\mu \Delta t + o(\Delta t)$. По определению, параметр потока обслуживаний

$$\mu_k = \lim_{\Delta t \rightarrow 0} \frac{1 - (1 - \mu \Delta t)^k + o(\Delta t)}{\Delta t} = \lim_{\Delta t \rightarrow 0} \frac{k \mu \Delta t + o(\Delta t)}{\Delta t} = k \mu,$$

а вероятность обслуживания за Δt двух и более заявок имеет порядок $o(\Delta t)$. Таким образом, в многоканальной системе при экспоненциально распределенной длительности обслуживания поток окончаний обслуживания при нахождении в каналах k требований является *ординарным*, а его параметр

$$\mu_k = \begin{cases} k \mu, & k = \overline{0, n-1}, \\ n \mu, & k = n, n+1, \dots \end{cases} \quad (3.7.1)$$

(интенсивность обслуживания ограничивается числом каналов).

3.8. Организация и продвижение очереди

Заявки, пришедшие в занятую систему, не могут быть обслужены немедленно и образуют очередь. Очередь может быть ограничена максимальной длиной R или максимальным временем W пребывания в ней. Примером задачи с временным ограничением является прибытие на стройку самосвала с бетонной смесью. При нарушении ограничения заявка получает отказ. Введение ограничения автоматически исключает очень большие задержки, но связано с дополнительными «штрафами» за отказ в обслуживании.

Новь прибывшая заявка в зависимости от организации и назначения системы становится либо в конец очереди (дисциплина FCFS: First Come — First Served), либо в ее начало (LCFS: Last Come — First Served). Последний вариант иначе называется стековым («магазинным») принципом.

При неоднородных заявках может вводиться приоритетное обслуживание. В этом случае заявки выстраиваются в несколько очередей, и в освободившийся канал поступает заявка из непустой очереди с наивысшим приоритетом. В некоторых ситуациях (абсолютный приоритет)

срочная заявка может прервать уже начатое обслуживание. Снятая заявка поступает в одну из очередей или теряется. Примером заявки с абсолютным приоритетом является поломка обслуживаемого устройства. Обслуживание подобной заявки состоит в отыскании и устранении неисправности. Поток поломок каналов обслуживания формируется замкнутым источником заявок, что не позволяет рассчитывать подобные СМО простой сменой интерпретации заявок с наивысшим абсолютным приоритетом.

3.9. Классификация и обозначение систем массового обслуживания

Для сокращенного наименования абстрактных СМО Д. Кендалл предложил использовать обозначения вида $A/B/n/R$, где A указывает распределение интервалов между требованиями, B — распределение времени обслуживания, n — число каналов, R — предельное число заявок в очереди или в системе (последний вариант удобнее). Типы распределений обозначаются следующим образом:

M — показательное (с марковским свойством);

E_r — эрланговское порядка r ;

D — детерминированное (постоянное время обслуживания или регулярный поток);

H_k — гиперэкспоненциальное с k составляющими;

G — произвольное распределение.

При $R \rightarrow \infty$ четвертую позицию принято опускать. Таким образом, запись вида $M/G/1$ означает одноканальную СМО с простейшим входящим потоком, произвольным (точнее, не конкретизируемым) распределением времени обслуживания и неограниченной очередью.

3.10. Показатели эффективности

Перечень показателей. Показатели эффективности процессов обслуживания обычно устанавливаются для стационарного (предельного при $t \rightarrow \infty$) режима. Их можно разделить на две группы — счетные

и временные. Счетные показатели связаны со стационарным распределением $\{p_j\}$ числа заявок в системе или финальным распределением $\{\pi_j\}$ числа заявок перед прибытием очередной заявки (эти распределения, вообще говоря, не совпадают). К *счетным* показателям относятся:

1) сами упомянутые распределения (в частности, они нужны для расчета емкости буферных накопителей из условия размещения очереди с вероятностью не менее заданной);

2) вероятность π_R отказа в приеме заявки на обслуживание;

3) вероятность нулевого ожидания $P_0 = \sum_{j=0}^{n-1} \pi_j$;

4) среднее число заявок в системе $L = \sum_{j=1}^R j p_j$;

5) среднее число заявок в очереди $q = \sum_{j=n+1}^R (j-n) p_j$;

6) среднее число свободных каналов $f = \sum_{j=0}^{n-1} (n-j) p_j$;

7) среднее число занятых каналов $z = n - f$.

Из *временных* показателей наиболее существенны:

1) ДФР $\bar{V}(t)$ времени пребывания заявки в системе;

2) ДФР времени виртуального (ненулевого) ожидания;

3) моменты названных распределений;

4) распределение и моменты длительности непрерывной занятости системы.

Оперативность системы оценивают по характеристикам распределения времени пребывания. Характеристики ожидания и, в частности, его средняя длительность отражают цену, которую клиент должен заплатить за совместное с другими клиентами (заявками) использование обслуживаемой системы.

Имея распределения числа заявок в системе, можно получить большинство интересующих исследователя временных характеристик с помощью принципа сохранения стационарной очереди (см. с. 90).

Показатель *пропускной способности* системы — интенсивность потока обслуженных заявок — имеет смысл определять для замкнутых

систем и СМО с ограниченной очередью. В остальных случаях она равна интенсивности входящего потока.

Выбор показателей. Часть названных показателей характеризует СМО с точки зрения потребителя, другие — с позиций эксплуатационного персонала. Улучшение оперативности обслуживания, в котором заинтересованы клиенты системы, достигается путем увеличения мощности системы и ухудшает показатели загрузки. Поэтому говорить об оптимизации системы можно только при комплексном подходе к ней. В качестве обобщенного показателя эффективности обычно берется взвешенная сумма показателей разных групп — по одному от каждой. Примером комплексного подхода и источником полезных аналогий может служить задача о выборе оптимального оборотного запаса (см. главу 9).

Часто системы проектируются из условия обеспечения заданных вышестоящим органом показателей обслуживания при минимальных затратах. Однако обосновать требуемые показатели очень трудно.

Выбранный показатель эффективности должен быть достаточно чувствителен к варьируемым параметрам системы. Это требование, в частности, делает работу с ДФР предпочтительнее, чем с обычной функцией распределения, так как последняя в практически интересной области высоких вероятностей успешного решения задачи меняется чрезвычайно медленно.

3.11. Законы сохранения в теории массового обслуживания

Существование стационарных режимов в системах массового обслуживания при стационарном входящем потоке возможно лишь при выполнении фундаментальных соотношений типа законов сохранения между некоторыми количествами, характеризующими состояние системы. Эти законы имеют отчетливое физическое истолкование, а их применение упрощает анализ СМО. При кажущейся очевидности вербальных формулировок упомянутых законов из них удается извлечь далеко не тривиальные и весьма конструктивные следствия. Рассмотрим наиболее важные из законов сохранения ТМО.

Сохранение заявок. Закон сохранения заявок формулируется в следующем виде:

Частота поступления заявок в канал обслуживания в среднем равна частоте выходов из этого канала.

Покажем, как использовать этот принцип для расчета вероятности свободного состояния однолинейной системы $GI/G/1$. Пусть a и b — средний интервал между прибывающими заявками и средняя длительность обслуживания соответственно, средняя частота прибытия заявок $\lambda = 1/a$. Частота обслуживаний равна вероятности занятости $1 - p_0$, деленной на b . Значит, в стационарном режиме должно быть $1/a = (1 - p_0)/b$, откуда следует

$$p_0 = 1 - b/a. \quad (3.11.1)$$

Для n -канальной системы условие баланса заявок сводится к

$$\sum_{j=0}^{n-1} (n-j)p_j = n - \lambda b = n(1 - \lambda b/n). \quad (3.11.2)$$

Отношение

$$\rho = \lambda b/n \quad (3.11.3)$$

называется *коэффициентом загрузки* системы и для однолинейной СМО совпадает с вероятностью ее занятости

Сохранение очереди. Зафиксируем число заявок в очереди перед прибытием очередной заявки и в момент приема ее на обслуживание. Очевидно, что при дисциплине очереди FCFS

распределение числа заявок, прибывших за время ожидания начала обслуживания, совпадает с распределением длины очереди перед прибытием заявки.

Сформулированный принцип справедлив для системы $G/G/n$ и всех частных вариантов ее. Этот же принцип в некоторых случаях может быть усилен и применен к системе в целом:

распределение числа заявок, прибывших за время ожидания окончания обслуживания, совпадает с распределением числа заявок в системе перед прибытием заявки.

Поскольку условие «первый пришел — первый обслужен» должно быть сохранено, усиленная формулировка работает для более узкого класса систем $G/G/1$ и $G/D/n$, в которых гарантируется совпадение порядка завершения обслуживания с порядком выборки из очереди

Законы сохранения очереди позволяют установить связь между распределением времени пребывания заявки в очереди и производящей функцией распределения числа заявок в ней перед прибытием очередной заявки

Производящей функции можно придать прямой вероятностный смысл если считать, что каждая заявка с вероятностью z независимо от остальных обладает некоторым свойством (например, является «красной»), то $\Pi(z)$ есть вероятность иметь все заявки «красными»

Пусть

$w(t)$ — плотность распределения времени ожидания начала обслуживания,

$\Pi(z)$ — производящая функция распределения числа заявок в очереди в момент прибытия новой заявки

Поскольку $\Pi(z)$ можно истолковать как вероятность отсутствия в системе синих заявок в момент прибытия новой, из сформулированного выше закона сохранения очереди следует равенство

$$\Pi(z) = \int_0^{\infty} e^{-\lambda(1-z)t} w(t) dt,$$

которое после подстановки $z = 1 - s/\lambda$ можно переписать как уравнение для ПЛС распределения времени ожидания начала обслуживания

$$\omega(s) = \Pi(1 - s/\lambda). \quad (3.11.4)$$

Оно устанавливает связь между «штучными» и «временными» показателями работы СМО. Приравняв множители при одинаковых степенях s в степенных разложениях обеих частей равенства (3.11.4), убеждаемся, что

$$w_k = q_{[k]}/\lambda^k, \quad k = 1, 2, \dots, \quad (3.11.5)$$

где

$$q_{[k]} \stackrel{\text{def}}{=} \sum_{n=k}^{\infty} n(n-1) \dots (n-k+1) \pi_n$$

есть k -й факториальный момент распределения числа заявок в очереди. Этот результат получен в статье [115]. В частности, среднее время ожидания в очереди

$$w_1 = q_{[1]}/\lambda = q_1/\lambda \quad (3.11.6)$$

(известная формула Литтла).

Аналогичная связь существует между факториальными моментами распределения числа заявок в системе и моментами распределения времени пребывания заявки в ней для систем $M/G/1$ и $M/D/n$. Усиленная формула Литтла

$$v_1 = L/\lambda \quad (3.11.7)$$

верна для всех систем и сетей обслуживания независимо от выполнения принципа FCFS.

Сохранение вероятностей состояний. Рассмотрим процесс переходов через разрез АВ в марковской системе (диаграмма переходов рис. 3.6).

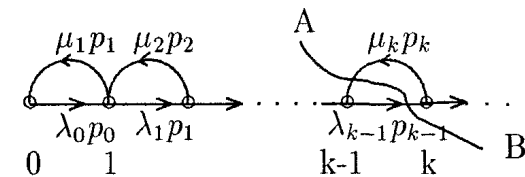


Рис. 3.6. К закону сохранения вероятностей для марковских систем

Состояния системы характеризуются числом k находящихся в ней заявок. Очевидно,

в стационарном режиме средние частоты переходов через разрез в противоположных направлениях равны.

Для диаграммы рис. 3.6 с переходами только между соседними состояниями применение этого закона непосредственно приводит к равенству

$$p_{k-1} \lambda_{k-1} = p_k \mu_k, \quad (3.11.8)$$

так что необходимые вероятности можно определить рекуррентно:

$$p_k = \frac{\lambda_{k-1}}{\mu_k} p_{k-1}, \quad k = 1, 2, \dots, \quad (3.11.9)$$

отправляясь от начальной вероятности p_0

Частным случаем разреза является кольцевой разрез, выделяющий одно из состояний системы. При этом противоположным направлениям соответствуют входящие и выходящие стрелки

Сохранение объема работы. Выделим класс так называемых «консервативных» дисциплин обслуживания, в котором возобновление прерванного обслуживания производится без потери ранее затраченных ресурсов. В указанном классе дисциплин

распределение объема невыполненной работы, находящегося в СМО, постоянно и не зависит от выбора конкретной дисциплины.

Покажем применение этого принципа к расчету среднего времени ожидания заявки в однолинейной системе. Прежде всего, в данном случае средний объем работы совпадает со средним значением w времени ожидания начала обслуживания вновь прибывшей заявкой. Это время будет складываться из времени завершения \bar{f}_1 начатого обслуживания и времени обслуживания находящихся в очереди ранее пришедших заявок. Средний остаток начатого обслуживания определяется формулой (3.1.2) и должен учитываться с вероятностью занятости системы, равной $\rho = \lambda b_1$.

$$\bar{f}_1 = \rho b_2 / (2b_1) = \lambda b_2 / 2.$$

Среднее число заявок в очереди на основании формулы Литтла составит λw , причем каждая из них в среднем обслуживается b_1 единиц времени. Итак, закон сохранения объема работы приводит к равенству

$$w = \lambda b_2 / 2 + \lambda w b_1,$$

откуда следует формула (Полячека-Хинчина)

$$w = \lambda b_2 / [2(1 - \rho)] \quad (3.11.10)$$

для среднего времени ожидания в $M/G/1$.

3.12. Расчет марковских систем по законам сохранения

Применим законы сохранения к расчету основных показателей системы $M/M/1$. Прежде всего отметим, что для этой СМО $\lambda(k) = \lambda$ и

$\mu(k) = \mu$ независимо от k . Следовательно, для всех k имеет место $\mu p_{k+1} = \lambda p_k$, или $p_{k+1} = (\lambda/\mu)p_k = \rho p_k$, и стационарные вероятности

$$p_k = (1 - \rho)\rho^k, \quad k = 0, 1, \dots \quad (3.12.1)$$

образуют геометрическую прогрессию со знаменателем $\rho = \lambda/\mu$. Условием существования стационарного режима является неравенство $\rho < 1$. Вероятность $p_0 = 1 - \rho$, найденная здесь из условия нормировки, совпадает с выражением (3.11.1) при соответствующей замене обозначений, что подтверждает закон сохранения требований.

Для n -канальной системы распределение числа требований в системе задается вероятностями

$$\begin{aligned} p_0 &= \left[\sum_{i=0}^{n-1} \frac{(\lambda/\mu)^i}{i!} + \frac{(\lambda/\mu)^n}{(n-1)!} \frac{1}{1-\rho} \right]^{-1}, \\ p_i &= p_0 \frac{(\lambda/\mu)^i}{i!}, \quad i = \overline{1, n}, \\ p_i &= p_n \rho^{i-n}, \quad i = n+1, n+2, \dots \end{aligned} \quad (3.12.2)$$

3.13. Система $M/G/1$

Выделим смежные моменты $\{\eta_n\}$, $n = 1, 2, \dots$, окончания обслуживания очередного требования в системе $M/G/1$ и рассмотрим состояния системы в моменты регенерации $\{\eta_n + 0\}$. В эти моменты система либо свободна, либо только начинает обслуживание, так что фиксировать значение истекшего времени обслуживания не нужно. Такой процесс является вложенным в исходный и, поскольку полностью (в вероятностном смысле) определяется своим текущим состоянием, называется *вложенной цепью Маркова*.

Обозначим q_j вероятность прибытия j новых требований за время обслуживания. Нетрудно видеть, что $\{q_j\}$ определяются согласно (3.6.3) и при аппроксимации распределения времени обслуживания гамма-плотностью проблем с их вычислением не будет. В установленном режиме застать в системе k требований в очередной момент регенерации можно, если в ней:

- а) в предыдущий момент регенерации находилась $j + 1$ заявка ($j \leq k$), одна обслужилась и пришло еще $k - j \geq 0$ заявок;

б) за время обслуживания первой заявки, прибывшей в свободную систему, пришло еще k требований.

Следовательно, *финальные* (предельные при устремлении номера момента регенерации к бесконечности) вероятности состояний вложенной цепи связаны системой уравнений

$$\pi_k = \sum_{j=0}^k \pi_{j+1} q_{k-j} + \pi_0 q_k = \sum_{j=1}^{k+1} \pi_j q_{k+1-j} + \pi_0 q_k, \quad k = 0, 1, \dots \quad (3.13.1)$$

Из этих формул легко получить рекуррентные выражения для расчета вероятностей:

$$\pi_k = (\pi_{k-1} - \pi_0 q_{k-1} - \sum_{j=1}^{k-1} \pi_j q_{k-j}) / q_0, \quad k = 1, 2, \dots \quad (3.13.2)$$

Можно показать, что для рассматриваемой системы вероятности состояний вложенной цепи совпадают со стационарными.

3.14. Многоканальные системы с фазовыми распределениями

3.14.1. Многофазное представление сложных СМО

Реализация любого метода расчета процессов в системе с очередями требует ее предварительной марковизации, т.е. представления текущего состояния системы параметрами, полностью определяющими (в вероятностном смысле) будущее поведение процесса. Эффективным методом марковизации сложных СМО является фазовая аппроксимация составляющих распределений. Каждая комбинация распределений (и их порядков) порождает специфическую диаграмму, которая может быть представлена различными способами. Разумеется, при изображении диаграмм порядки распределений приходится конкретизировать.

На диаграмме переходов для $M/E_3/2$ (рис. 3.7) состояние системы идентифицируется полным количеством заявок в ней (номер яруса) и распределением проходящих обслуживание заявок по его фазам («ключ» соответствующего микросостояния). Здесь и на последующих диаграммах сумма значений позиций ключа равна числу задействованных каналов обслуживания.

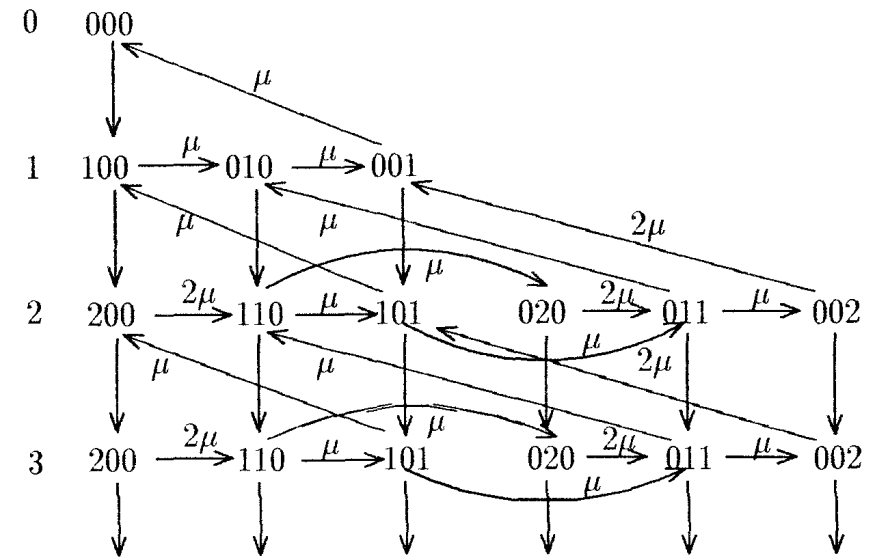


Рис. 3.7. Переходы в системе $M/E_3/2$

Порядок следования ключей обеспечивает «вертикальность» переходов по прибытию заявки, что упрощает построение матриц интенсивностей переходов между микросостояниями ярусов. Переходы по завершению фазы обслуживания переводят заявку либо в микросостояние того же яруса, либо (если фаза последняя) — на вышележащий ярус, так как заявка считается обслуженной полностью и покидает систему.

Работа системы $M/H_k/n$ может быть интерпретирована как процесс обслуживания неоднородного потока заявок [68], причем тип заявки определяет параметр показательного распределенного обслуживания. Теперь ключ микросостояния указывает количество находящихся в каналах обслуживания заявок каждого типа (рис. 3.8, 3.9). Суммарный входящий поток имеет интенсивность λ ; прибывающая (или выбираемая из очереди) заявка с вероятностью y_i относится к i -му типу, $i = 1, 2$. На последнем рисунке при $j > n$ параметр потока обслуживаний заявок i -го типа равен $m_i \mu_i$, где m_i — содержимое i -й позиции ключа. Завершение обслуживания с вероятностями $\{y_i\}$ в зависимости от типа выбранной из очереди заявки приводит в одно из микросостояний вышележащего яруса.

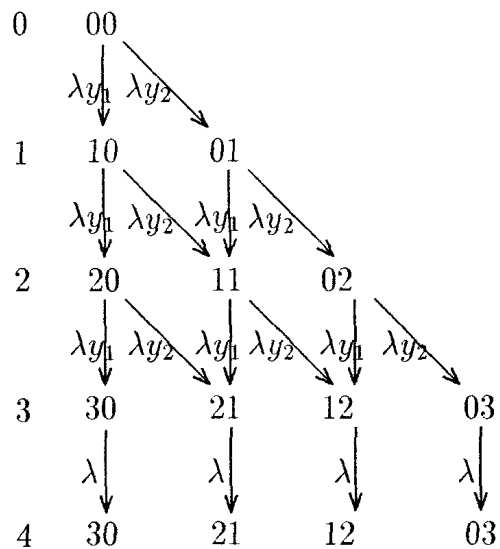


Рис. 3.8 Прибытие заявки в систему $M/H_2/3$

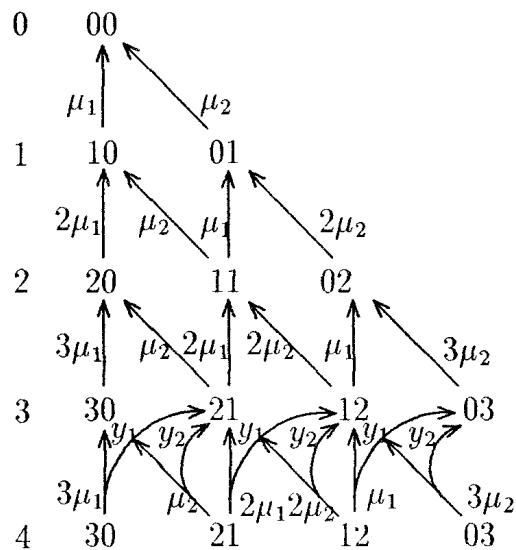


Рис. 3.9. Уход заявки из системы $M/H_2/3$

3.14.2. Уравнения глобального баланса

Обозначим через S_j множество всех возможных микросостояний системы, при которых на обслуживании находится ровно j заявок, а через σ_j — количество элементов в S_j . Далее в соответствии с диаграммой переходов для выбранной модели построим матрицы интенсивностей переходов из микросостояний j -го яруса:

$A_j[\sigma_j \times \sigma_{j+1}]$ — в S_{j+1} (прибытие заявки);

$C_j[\sigma_j \times \sigma_j]$ — в S_j (конец промежуточной фазы обслуживания);

$B_j[\sigma_j \times \sigma_{j-1}]$ — в S_{j-1} (полное завершение обслуживания);

$D_j[\sigma_j \times \sigma_j]$ — ухода из состояний яруса j

(в квадратных скобках здесь и далее указывается размер матриц).

Введем векторы-строки $\gamma_j = \{\gamma_{j,1}, \gamma_{j,2}, \dots, \gamma_{j,\sigma_j}\}$ нахождения СМО в состоянии (j, i) , $j = 0, 1, \dots$. Теперь можно записать векторно-матричные уравнения баланса переходов между состояниями

$$\begin{aligned} \gamma_0 D_0 &= \gamma_0 C_0 + \gamma_1 B_1, \\ \gamma_j D_j &= \gamma_{j-1} A_{j-1} + \gamma_j C_j + \gamma_{j+1} B_{j+1}, \quad j = 1, 2, \dots \end{aligned} \quad (3.14.1)$$

В связи с зависимостью структуры матриц переходов от типа и порядка аппроксимирующих фазовых распределений ценность программной реализации расчетной схемы определяется возможностью автоматического построения матриц переходов. Эту проблему можно решить следующим образом:

- для каждого j -го яруса системы, $j = \overline{0, n}$, автоматически сгенерировать последовательность ключей микросостояний;
- сформировать матрицы переходов, сопоставляя ключи-«источники» j -го яруса и совместимые с ними по выбранному типу переходов ключи-«преемники»: для матрицы B на $(j-1)$ -м ярусе, для матрицы C — на j -м, для A — на $(j+1)$ -м ярусе.

Именно такой подход применен в описанном ниже (разд. 3.17) пакете подпрограмм МОСТ.

Система (3.14.1), дополненная условием нормировки, даже для моделей с ограниченной очередью характеризуется чрезвычайно высокой размерностью, и стандартные методы решения систем линейных алгебраических уравнений применительно к ней оказываются малоэффективными. Мы рассмотрим два метода ее решения: итерационный и матрично-геометрической прогрессии.

3.14.3. Итерационный метод

Идея этой схемы была впервые предложена Такахаси и Таками [209]. Мы опишем ее в более общей форме и с детальной проработкой расчетных зависимостей [66, 67, 68].

Положим $t_j = \gamma_j/p_j$, где p_j — суммарная вероятность наличия в системе ровно j заявок, и обозначим

$$x_j = p_{j+1}/p_j, \quad z_j = p_{j-1}/p_j. \quad (3.14.2)$$

Тогда систему (3.14.1) можно переписать относительно векторов условных вероятностей $\{t_j\}$, нормированных к единице в пределах яруса:

$$\begin{aligned} t_0 D_0 &= t_0 C_0 + x_0 t_1 B_1, \\ t_j D_j &= z_j t_{j-1} A_{j-1} + t_j C_j + x_j t_{j+1} B_{j+1}, \quad j = 1, 2, \dots \end{aligned} \quad (3.14.3)$$

С помощью векторов-столбцов $\mathbf{1}_j = \{1, 1, \dots, 1\}^T$ размера σ_j для всех j могут быть записаны дополняющие систему (3.14.3) условия нормировки

$$t_j \mathbf{1}_j = 1 \quad (3.14.4)$$

и баланса суммарных интенсивностей переходов между смежными ярусами

$$t_j A_j \mathbf{1}_{j+1} = x_j t_{j+1} B_{j+1} \mathbf{1}_j. \quad (3.14.5)$$

Алгоритм расчета набора векторов $\{t_j\}$ и чисел $\{x_j\}$ и $\{z_j\}$, удовлетворяющих соотношениям (3.14.3)–(3.14.5), в случае разомкнутой системы с неограниченной очередью опирается на существование предельного вектора условных вероятностей $t_\infty = \lim_{j \rightarrow \infty} t_j$, которое является следствием стабилизации переходных матриц при $j > n$. Алгоритм основан на последовательном приближении к искомому характеристикам для ограниченного множества индексов $j = \overline{0, N}$ и по существу является блочным вариантом известного итерационного метода Гаусса-Зейделя.

Перепишем уравнения системы (3.14.3) для $j \geq 1$ в виде

$$t_j^{(m)} (D_j - C_j) = z_j^{(m)} t_{j-1}^{(m-1)} A_{j-1} + x_j^{(m)} t_{j+1}^{(m)} B_{j+1}, \quad j = 1, 2, \dots,$$

где верхний индекс указывает номер итерации (по причинам, которые поясняются ниже, итерации предпочтительно вести от больших индексов к меньшим). Теперь ясно, что

$$t_j^{(m)} = z_j^{(m)} \beta_j' + x_j^{(m)} \beta_j'', \quad (3.14.6)$$

где

$$\begin{aligned} \beta_j' &= t_{j-1}^{(m-1)} A_{j-1} (D_j - C_j)^{-1}, \\ \beta_j'' &= t_{j+1}^{(m)} B_{j+1} (D_j - C_j)^{-1}. \end{aligned} \quad (3.14.7)$$

При $j = N$ считается, что

$$\beta_N'' = t_{N-1}^{(m-1)} B_N (D_N - C_N)^{-1}. \quad (3.14.8)$$

Осталось указать способ расчета $\{z_j^{(m)}\}$ и $\{x_j^{(m)}\}$. Перепишем (3.14.5) с учетом (3.14.6):

$$(z_j^{(m)} \beta_j' + x_j^{(m)} \beta_j'') A_j \mathbf{1}_{j+1} = x_j^{(m)} t_{j+1}^{(m)} B_{j+1} \mathbf{1}_j.$$

Отсюда следует пропорциональность

$$z_j^{(m)} = c x_j^{(m)} \quad (3.14.9)$$

с коэффициентом

$$c = \frac{t_{j+1}^{(m)} B_{j+1} \mathbf{1}_j - \beta_j'' A_j \mathbf{1}_{j+1}}{\beta_j' A_j \mathbf{1}_{j+1}}. \quad (3.14.10)$$

В этой и последующих формулах произведения матриц перехода на вектор $\mathbf{1}_j$ равны суммам строк соответствующих матриц и могут быть вычислены до начала итераций.

Подстановка (3.14.8) в (3.14.6) и умножение обеих частей результата на $\mathbf{1}_j$ дают

$$1 = t_j^{(m)} \mathbf{1}_j = x_j^{(m)} (c \beta_j' + \beta_j'') \mathbf{1}_j.$$

Итак,

$$x_j^{(m)} = 1 / [(c \beta_j' + \beta_j'') \mathbf{1}_j]. \quad (3.14.11)$$

Удобным критерием прекращения итераций является условие

$$\max_j |x_j^{(m)} - x_j^{(m-1)}| \leq \varepsilon_1.$$

Описанный алгоритм в связи с усечением числа ярусов — см. формулу (3.14.8) — будет давать надежные результаты лишь при условии

$$\|t_N - t_\infty\| \leq \varepsilon_2. \quad (3.14.12)$$

Целесообразно выполнить итерации для умеренного значения N , после чего проверить выполнение условия (3.14.12).

Упомянутое условие указывает естественный путь задания начального приближения: $t_N^{(0)} = t_\infty$. Рассмотрим способ нахождения такого приближения. Будем обозначать предельные при $j \rightarrow \infty$ матрицы, векторы и отношения смежных вероятностей прежними символами, но без индексов. Если существует $t_\infty = t$, то существуют и предельные значения отношений смежных вероятностей x и $z = 1/x$, причем из (3.14.5) следует

$$x = tA1_n/tB1_n \quad (3.14.13)$$

и стационарные вероятности состояний с большими индексами образуют геометрическую прогрессию.

Расчеты свидетельствуют о хорошей аппроксимации x формулой

$$x = \rho^{2/(v_A^2 + v_B^2)}. \quad (3.14.14)$$

Полагая x известным, как следствие (3.14.6) имеем предельное равенство

$$t = (x^{-1}tA + xtB)(D - C)^{-1} = t(x^{-1}A + xB)(D - C)^{-1} = tQ,$$

где

$$Q = (x^{-1}A + xB)(D - C)^{-1}.$$

Обозначим $(Q - I)_1$ матрицу, полученную из $Q - I$ заменой ее первой строки на единичную, и положим $\delta_1 = \{1, 0, 0, \dots, 0\}$. Тогда искомым вектор получается как решение системы линейных алгебраических уравнений

$$t(Q - I)_1 = \delta_1. \quad (3.14.15)$$

В качестве начальных приближений к векторам $\{t_j\}$ в ходе описанных выше итераций, охватывающих все ярусы диаграммы, для $j \geq n$ следует воспользоваться решением (3.14.15) при значении x , найденном по формуле (3.14.14). На вышележащих ярусах можно принять все состояния равновероятными. Опыт расчетов свидетельствует о слабом влиянии выбора начальных приближений (в рамках рассмотренных подходов) на требуемое число итераций.

После прекращения итераций можно переходить к нахождению абсолютных значений вероятностей. Прежде всего отметим, что из определения чисел $\{x_j\}$ следуют равенства

$$p_{j+1} = p_j x_j, \quad j = \overline{0, N-1}. \quad (3.14.16)$$

Подставив их в условие (3.11.2) баланса прибытия и ухода заявок

$$\sum_{j=0}^{n-1} (n-j)p_j = n - b/a,$$

получаем формулу для вероятности свободного состояния системы

$$p_0 = \frac{n - b/a}{n + \sum_{j=1}^{n-1} (n-j) \prod_{i=0}^{j-1} x_i}. \quad (3.14.17)$$

Последующие вероятности для $j = \overline{1, N}$ определяются рекуррентно с помощью (3.14.16). При необходимости та же формула может быть применена для больших значений j с использованием $x_j = x_\infty$.

Итерационный метод применим и для расчета замкнутых систем.

3.14.4. Метод матрично-геометрической прогрессии

Для расчета «фазовых» систем обслуживания весьма эффективен предложенный Ивэнсом [137] метод матрично-геометрической прогрессии. Идея его заключается в представлении векторов вероятностей микросостояний полностью занятой системы соотношением типа

$$\gamma_j = \gamma_n R^{j-n}, \quad j = n, n+1, \dots, \quad (3.14.18)$$

где R — матричный знаменатель прогрессии. Выпишем одно из уравнений системы (3.14.1) для $j > n$, опуская индексы у стабилизировавшихся к этому ярусу матриц переходов:

$$\gamma_j D = \gamma_{j-1} A + \gamma_{j+1} B + \gamma_j C. \quad (3.14.19)$$

Подставив выражения векторов вероятностей состояний из (3.14.18), можно переписать (3.14.19) в виде

$$\gamma_{j-1} R(D - C) = \gamma_{j-1} A + \gamma_{j-1} R^2 B,$$

откуда следует, что искомым знаменатель прогрессии должен удовлетворять матричному квадратному уравнению

$$R^2 B - R(D - C) + A = 0. \quad (3.14.20)$$

При известном R векторы $\{\gamma_j\}$, $j = \overline{0, n}$, находим решением системы уравнений глобального баланса для микросостояний соответствующих ярусов, дополненной условием баланса заявок:

$$\begin{aligned} \sum_{j=0}^{n-1} (n-j)\gamma_j \mathbf{1}_j &= n - b/a, \\ \gamma_0(D_0 - C_0) &= \gamma_1 B_1, \\ \gamma_j(D_j - C_j) &= \gamma_{j-1} A_{j-1} + \gamma_{j+1} B_{j+1}, \quad j = \overline{1, n-1}, \\ \gamma_n(D_n - C_n) &= \gamma_{n-1} A_{n-1} + \gamma_n R B_{n+1}. \end{aligned} \quad (3.14.21)$$

Практически эту систему перед решением необходимо расписать по компонентам упомянутых в ней векторов. Решение может быть проведено как прямым методом типа гауссова исключения, так и методом итераций. Последующие векторы вероятностей определяются на основе (3.14.18) рекуррентно — домножением предыдущего вектора на R .

Основная проблема реализации матричного метода заключается в определении знаменателя прогрессии R . Уравнение вида (3.14.20) может быть решено лишь численными методами. Наилучшим решением вопроса оказалось вычисление поправок к знаменателю прогрессии в линейном приближении. Для упрощения обозначений перепишем (3.14.20) с учетом равенства $C = 0$ (при гиперэкспоненциальном обслуживании переходы между микросостояниями в пределах яруса отсутствуют — см. рис. 3.9) и определим поправку Δ из условия

$$(R + \Delta)(R + \Delta)B - (R + \Delta)D + A = 0.$$

Пренебрегая членом, содержащим квадрат поправки, имеем

$$R^2 B + \Delta R B + R \Delta B - R D - \Delta D + A \approx 0,$$

или

$$\Delta(RB - D) + R \Delta B \approx RD - R^2 B - A.$$

Последнее уравнение приводится к виду

$$\Delta F + R \Delta B = G \quad (3.14.22)$$

и является *линейным* относительно поправочной матрицы Δ . Коэффициенты развернутой формы системы уравнений (3.14.22) могут быть выражены через элементы известных матриц A, B и D , после чего она решается стандартным методом. Размер системы может оказаться весьма внушительным (для модели $H_2/H_2/5$ — 144 неизвестных), что предъявляет значительные требования к оперативной памяти и делает трудоемким каждый шаг итерации. Однако число итераций резко сокращается.

Дальнейшее уменьшение числа итераций достигается получением лучшего начального приближения. В этих целях перепишем уравнение (3.14.20) в форме $xRB - RD + A = 0$. Теперь ясно, что можно принять

$$R_0 = A(D - xB)^{-1}.$$

Совместное применение этих приемов оказалось исключительно эффективным. Так, для системы $H_2/H_2/3$ с коэффициентом загрузки $\rho = 0.9$ и коэффициентами вариации исходных распределений $v_A = 0.4$ и $v_B = 3$ (данные, близкие к наихудшим из проверенных в смысле сходимости) при начальной норме невязки 19.280 в трех последовательных итерациях были получены невязки 3.350, $2.30 \cdot 10^{-3}$ и $7.23 \cdot 10^{-10}$. Заметим, что исходный вариант метода для снижения нормы невязки до 10^{-6} потребовал 266 шагов, из которых 104 пришлось на последний порядок. Ни в одном из полусотни обчисленных случаев не потребовалось более четырех шагов итераций.

3.15. Сети обслуживания

Анализ реальных процессов обслуживания (в том числе ремонта) часто показывает, что заявка получает «полное удовлетворение» лишь после прохождения нескольких фаз обслуживания в специализированных узлах. В таких случаях принято говорить о *сетях* обслуживания. Создание теории сетей обслуживания стимулировалось как возвратом к старым задачам на новом (более детальном) уровне, так и принципиально новыми приложениями, среди которых прежде всего нужно отметить вычислительные сети и сети передачи данных с использованием вычислительных устройств.

3.15.1. Определения и допущения

Сеть обслуживания состоит из рабочих узлов, занумерованных от 1 до M , источника (узел «0») и стока (узел « $M+1$ »). Новая заявка рождается в источнике; с попаданием заявки в сток фиксируется окончание ее пребывания в сети. Если суммарная интенсивность входящего потока не зависит от количества находящихся в сети заявок, сеть считается *открытой*, в противном случае — *замкнутой*. Практически исследован единственный частный случай замкнутых сетей — сеть с постоянной популяцией (числом заявок) K . Здесь вместо каждой попавшей в сток заявки в источнике мгновенно генерируется (и передается в один из рабочих узлов) новая. Замкнутые сети этого вида хорошо моделируют локальные вычислительные системы коллективного пользования. При наличии в сети неоднородных заявок она может быть замкнутой по одним типам и разомкнутой по другим. Такая сеть считается *смешанной*.

Маршрут заявки в сети, вообще говоря, случаен и определяется неразложимой матрицей передач $R = \{r_{i,j}\}$, $i, j = \overline{0, M+1}$, образованной вероятностями перехода из i -го в j -й узел. Эти вероятности не зависят от маршрута, уже пройденного заявкой. Неразложимость (неприводимость) сети означает невозможность ее разделения на не связанные допустимыми переходами компоненты.

Для каждого j -го узла задаются моменты распределения чистой длительности обслуживания $\{b_{j,l}\}$, $l = \overline{1, L}$, число каналов n_j и дисциплина обслуживания.

При наличии заявок нескольких типов $q = \overline{1, Q}$ они разбиваются на классы «замкнутых» Q_c и «открытых» Q_o , а всем перечисленным выше величинам приписывается дополнительный (верхний) индекс типа заявки q .

Как было показано в предыдущих разделах, основные показатели работы СМО можно рассчитать, получив распределение вероятностей ее состояний. Разумеется, предварительно должна быть проведена маркировка процесса.

Состояние *сети* обслуживания в общем случае приходится характеризовать *совместным* распределением вероятностей состояний, учитывающим взаимную зависимость происходящих в узлах процессов. Совместные распределения крайне неудобны в работе, в связи с чем при анализе сетей делаются допущения, сводящие упомянутую зависимость к минимуму:

- вероятности $\{r_{i,j}\}$ перехода заявки из узла i в узел j не зависят от предыстории заявки (в частности, от кратности прохождения узла) и от состояния узла-преемника j ;
- распределения длительности обслуживания в узлах определяются только типом заявки и номером узла и не зависят от обслуживания других заявок.

Режим работы сети мы будем предполагать стационарным. Очевидный критерий возможности такого режима — это докритическая загрузка всех его узлов, что обеспечивает конечные средние длины очередей в узлах.

В качестве *показателей работы* сетей обслуживания используют:

- «маргинальные» распределения числа заявок в узлах, т.е. усредненные по распределениям заявок в дополняющих подсетях (указанные распределения нужны для обоснования требований к емкости накопителей заявок);
- распределение времени пребывания заявки в сети и его моменты (дифференцированно по видам заявок);
- производительность сети (интенсивность потока обслуженных заявок) — только по замкнутым типам.

Строго говоря, независимый расчет узлов возможен только для мультипликативных сетей (П-сетей), для которых совместное распределение вероятностей состояний узлов представимо произведением частных распределений на нормирующий множитель. Соответствующие условия сформулированы в известной теореме ВСМР [108].

Тридцатилетнее развитие методов строгой декомпозиции сетей обслуживания практически исчерпало их возможности. Ряд особенностей реальных задач (узлы типа FCFS с немарковским обслуживанием, с различными распределениями обслуживания для заявок разных видов, с приоритетным обслуживанием) исключает возможность П-решения. Как будет показано в главе 9, характеристики работы СМО сильно зависят от вариации распределений, неучет которой дает намного большие ошибки, чем «силовая» декомпозиция сети. Кроме того, «классические» методы расчета сетей имеют слишком быстрый рост трудоемкости при увеличении размерности задачи (в особенности по объему популяции и числу типов заявок) и во многих случаях вынуждают ограничиться средними характеристиками, недостаточными для большинства приложений.

Все приближенные методы расчета СеМО опираются на ту или иную форму *декомпозиции* сети. Декомпозиция предоставляет возможность:

- сведения большой задачи к серий задач меньшей размерности;
- однократного анализа типовой подструктуры общей модели,
- аналитического исследования простых подмоделей;
- упрощения параметрических исследований (разделяются фиксированные и изменяемые подмодели);
- применения к подмоделям различных временных масштабов;
- применения к разным подмоделям различных методов исследования, наиболее подходящих в конкретных условиях.

Расчет сети выполняется в следующей очередности:

- 1) Решением уравнений баланса или Q таких систем в неоднородном случае определяются средние потоки через узлы с точностью до постоянного множителя
- 2) Из сети последовательно выделяются по одному узлы $j = \overline{1, M}$ с присущими им числовыми характеристиками обслуживания, причем дополняющая подсеть заменяется пуассоновским источником заявок постоянной интенсивности. Выделенный узел рассчитывается как изолированная система.
- 3) В случае ограниченной популяции по результатам второго этапа проводится коррекция потоков, после чего этот этап повторяется (до выполнения условия сходимости итераций).
- 4) После стабилизации потоков каждый узел рассчитывается как изолированная система со своей дисциплиной обслуживания
- 5) Определяются агрегированные характеристики распределения времени пребывания заявок в сети в целом.

Наличие этапов агрегации 1,3 и 5, обеспечение баланса потоков (этап 1) и постоянства популяций (этап 3), независимость этапа балансировки от дисциплины обслуживания, а также применение итераций, исключающих длительное накопление погрешностей, обеспечивают высокую устойчивость (робастность) методов указанного типа.

3.15.2. Разомкнутая сеть с однородным потоком

Для разомкнутой сети все потоки между узлами считаются простейшими. Расчет начинается с определения из уравнений баланса заявок

$$\lambda_i = \Lambda r_{0,i} + \sum_{j=1}^M \lambda_j r_{j,i}, \quad i = \overline{1, M}, \quad (3.15.1)$$

интенсивностей $\{\lambda_i\}$ входящих в узлы потоков. Здесь Λ — суммарная интенсивность потока от внешних источников. Далее для всех узлов должно быть проверено условие отсутствия перегрузки

$$\lambda_i b_{i,1} / n_i < 1, \quad (3.15.2)$$

обеспечивающее существование в сети стационарного режима. Положительный результат проверки позволяет продолжить вычисления в зависимости от алгоритмических и вычислительных возможностей, которыми располагает исследователь. В частности, предположение о том, что все потоки в сети — простейшие, позволяет непосредственно приступить к расчету маргинальных распределений числа заявок в узлах с заданными для них распределениями длительности обслуживания, числом каналов и полученными из системы (3.15.1) интенсивностями $\{\lambda_i\}$ входящих потоков.

Отказ от упомянутой гипотезы предполагает детальный учет:

- свойств внешних потоков;
- расщепления выходящих из узлов потоков вследствие случайного (с вероятностями $\{r_{i,j}\}$) либо регулярного распределения обслуженных заявок по узлам-преемникам;
- суммирования на входе преемника потоков от нескольких источников;
- преобразования входящих потоков в узлах обслуживания.

Этот учет при наличии циклических маршрутов требует итерационного уточнения распределений интервалов между входящими в узлы заявками, причем внутри итераций приходится использовать более трудоемкие алгоритмы расчета узлов — с немарковскими входящими потоками.

3.15.3. Расчет сети с неоднородными потоками

Алгоритм разд. 3.15.2 естественно обобщается на случай Q типов неоднородных заявок с дифференцированными характеристиками беспriorитетного обслуживания каждого типа. Обобщенный алгоритм состоит из следующих шагов:

- 1) Решение системы уравнений баланса (3.15.1) по каждому q -му типу заявок относительно средних интенсивностей потоков $\{\lambda_j^{(q)}\}$ в узлах $j = \overline{1, M}$; заметим, что матрица передач R , вообще говоря, зависит от q .
- 2) Проверка отсутствия перегрузки: $\max_j \left\{ \sum_q \lambda_j^{(q)} b_{j,1}^{(q)} / n_j \right\} < 1$
(при его нарушении необходима коррекция исходных данных).
- 3) Расчет удельного веса заявок каждого типа во всех узлах и средневзвешенных моментов распределений длительности обслуживания по каждому узлу.
- 4) Расчет моментов распределения времени ожидания, общих для всех типов заявок, по каждому из узлов.
- 5) Для всех типов заявок:
 - получение для всех узлов моментов распределения времени пребывания при одном посещении посредством свертки моментов распределений времени ожидания и чистой длительности обслуживания заявок этого типа;
 - расчет моментов распределения времени пребывания заявки этого типа в сети;
 - построение ДФР времени пребывания в сети.
- 6) Конец алгоритма.

3.15.4. Замкнутая сеть

Для интересующей нас замкнутой однородной сети, моделирующей процессы ремонта, сохраняется постоянное значение K суммарного числа заявок в источнике и в рабочих узлах сети. Обозначим через $\{\bar{k}_i\}$ среднее число заявок в рабочих узлах, $\sigma = \sum_{i=1}^M \bar{k}_i$ — их сумму,

$\bar{k}_0 = K - \sigma$ — среднее число исправных агрегатов, l_0 — удельную интенсивность их отказов. Тогда условия баланса заявок можно записать в виде

$$\begin{aligned} \sum_{i=1}^M \lambda_i r_{i,j} + l_0 \bar{k}_0 r_{0,j} - \lambda_j &= 0, & j = \overline{1, M-1}, \\ \sum_{i=1}^M \lambda_i r_{i,M+1} &= l_0 \bar{k}_0. \end{aligned} \quad (3.15.3)$$

Последнее уравнение играет роль нормировки — оно приравнивает средние интенсивности новых отказов и завершений ремонта.

Расчет системы производится следующим образом:

- 1) Выбрать начальное приближение $\bar{k}_0 = K/2$.
- 2) Решением системы (3.15.3) найти средние интенсивности $\{\lambda_j\}$ входящих в узлы потоков. Положить среднее число заявок в рабочих узлах сети $\sigma = 0$.
- 3) Для всех узлов $j = \overline{1, M}$:
 - рассчитать распределение числа заявок в узле как в разомкнутой СМО с интенсивностью потока заявок λ_j ;
 - вычислить среднее число \bar{k}_j заявок в ней;
 - пересчитать сумму: $\sigma = \sigma + \bar{k}_j$.
- 4) Вычислить новое приближение $\bar{k}'_0 = K - \sigma$.
- 5) Если $|\bar{k}_0 - \bar{k}'_0| > \varepsilon K$, заменить \bar{k}_0 на \bar{k}'_0 и вернуться к этапу 2.
- 6) При необходимости в расчете распределения времени пребывания заявок в сети повторить этапы 2 и 3 с дополнительным вычислением моментов распределения времени пребывания заявки в каждом из узлов сети.
- 7) Конец алгоритма.

3.16. Распределение времени пребывания в сети

3.16.1. Решение в средних

На основании формулы Литтла для сети в целом среднее время пребывания заявки в разомкнутой сети

$$v = \sum_{i=1}^M \bar{k}_i / \Lambda. \quad (3.16.1)$$

Здесь среднее число заявок в сети делится на суммарную интенсивность входящего потока. Для замкнутой сети применяется та же формула с заменой $\sum_{i=1}^M \bar{k}_i$ на K и

$$\Lambda = \sum_{i=1}^M \lambda_i r_{i,M+1} \quad (3.16.2)$$

(это суммарная интенсивность потока обслуженных заявок, по закону сохранения заявок равная интенсивности входящего потока).

3.16.2. Преобразование Лапласа и высшие моменты

Для наиболее ответственных применений знать среднее время пребывания заявки в сети недостаточно: в подобных случаях обычно встает вопрос и о высших моментах и/или построении функции распределения.

Выделим из матрицы передач R :

- вектор-строку $P = \{r_{0,1}, r_{0,2}, \dots, r_{0,M}\}$ вероятностей перехода из источника в конкретные рабочие узлы;
- вектор-столбец $T = \{r_{1,M+1}, r_{2,M+1}, \dots, r_{M,M+1}\}^T$ вероятностей перехода из рабочих узлов в сток;
- матрицу $Q = \{r_{i,j}\}$, $i, j = \overline{1, M}$, вероятностей переходов между рабочими узлами.

Кроме того, определим диагональную матрицу $N(s)$ преобразований Лапласа $\{\nu_i(s)\}$ распределений времени пребывания в рабочих узлах сети и сформируем матрицу

$$\Gamma(s) = N(s)Q. \quad (3.16.3)$$

ПЛС распределения длительности одношаговых переходов в сток

$$\gamma_1(s) = PN(s)T.$$

Для двухшаговых переходов

$$\gamma_2(s) = P\Gamma(s)N(s)T$$

и вообще для k -шаговых

$$\gamma_k(s) = P\Gamma^{k-1}(s)N(s)T.$$

Полное ПЛС распределения времени пребывания заявки в сети

$$\gamma(s) = \sum_{k=1}^{\infty} \gamma_k(s) = P \left[\sum_{k=0}^{\infty} \Gamma^k(s) \right] N(s)T = P[I - \Gamma(s)]^{-1} N(s)T.$$

Существенным элементом этой технологии является вычисление преобразований Лапласа распределений времени пребывания заявок в узлах по моментам этих распределений — см. расчет q_0 для различных аппроксимаций, подобранных из условия выравнивания моментов. Моменты $\{g_i\}$ распределения времени пребывания в сети можно получить численным дифференцированием $\gamma(s)$ в нуле.

Пользователей сети обычно интересует вопрос о вероятности пребывания заявки в сети *дольше* заданного времени, причем само это время указывается со значительным элементом произвола. Практически в таких случаях нужно строить по найденным моментам ДФР времени пребывания заявки в сети для значений аргумента, перекрывающих диапазон возможных изменений директивного срока. Таблица ДФР может быть построена на основе аппроксимации ее ДФР Вейбулла. Напомним, что не следует ожидать хорошей относительной точности для значений ДФР, меньших 0.01.

3.17. Пакет МОСТ по расчету систем с очередями

Методы теории очередей применимы к любым случайным процессам скопления и рассасывания очередей на транспорте, в промышленности, технике (прежде всего вычислительной), здравоохранении, бытовом обслуживании, военном деле, образовании, организации научных экспериментов на дефицитном оборудовании и т.п. Как будет показано в части III этой книги, важнейшая проблема расчета восстанавливаемого ЗИПа решается методами названной теории. Предварительная оценка достижимых при данных условиях и конфигурации системы показателей качества обслуживания (средней длины очереди, вероятности превышения допустимого времени ожидания или потери заявки и т.п.) позволяет избежать как дорогостоящей перестраховки из-за создания систем избыточной производительности по числу или мощности каналов обслуживания, так и неоправданного снижения качества обслуживания. Последнее особенно опасно в ситуациях с высоким риском (системы управления посадкой самолетов, аварийные службы, группы быстрого реагирования, экстренная медицинская помощь), а в более обычных случаях гарантирует как минимум потерю клиентуры.

Заметим, что подобный расчет требуется не только при проектировании системы обслуживания: он необходим при каждом серьезном изменении интенсивностей потоков заявок, их маршрутизации, трудоемкости обработки, требований к качеству обслуживания. Таким образом, необходимыми расчетными средствами должны быть оснащены не только проектировщики, но и управляющий персонал реально эксплуатируемых систем обслуживания.

Общеизвестна возможность расчета произвольных систем обслуживания с помощью имитационного моделирования, реализуемого средствами типа GPSS. Однако эти средства негибки, требуют большого расхода машинного времени и, что самое неприятное, дают низкую точность определения вероятностей редких событий. Поэтому предпочтительно использование аналитических методов — при необходимости в сочетании с численными. Простейшие предположения об экспоненциальных распределениях интервалов между смежными заявками и длительности обслуживания приводят к известным формулам Эрланга (около 1910 г.), погрешность которых для распределений, заметно отличающихся от показательного, может быть сколь угодно велика.

Теория массового обслуживания имеет дело со сложными случайными процессами. Реалистические расчетные методики для решения достаточно широкого круга задач, часть которых описана в этой главе, требуют системной разработки численных методов ТМО и их реализации в пакете прикладных программ (ППП). Общая организация ППП и подходы к их оценке обсуждаются, например, в [76]. Организация, приобретающая готовый пакет, получает следующие преимущества [190]:

- 1) собственная разработка может быть неподъемной (по дефициту времени и квалифицированных кадров) для малой фирмы;
- 2) до внедрения пройдет меньше времени, поскольку проектирование, программирование, отладка, документирование уже выполнены;
- 3) готовый пакет обойдется дешевле собственной разработки;
- 4) его качество наверняка будет выше, а число ошибок и потребность в ресурсах компьютера — меньше;
- 5) пакет будет хорошо документирован и удобен в работе;
- 6) будет возможен обмен информацией и опытом с другими пользователями;
- 7) обновление версий и развитие пакета обойдутся сравнительно дешево.

Разумеется, имеются аргументы и в пользу самостоятельной разработки:

- 1) Трудна сравнительная оценка пакетов (нужно учитывать очень много факторов).
- 2) Практически неизбежны различия в терминологии, подходах, обозначениях, форматах.
- 3) Всегда в какой-то степени (часто — весьма существенной) необходима подстройка покупного пакета, которая может оказаться очень сложной.
- 4) Разработчиков трудно уговорить что-то изменить.
- 5) Сопровождение чужого продукта собственными силами очень трудоемко.

3.17.1. История ППП для расчета систем обслуживания

Работы по созданию численных методов расчета систем обслуживания публикуются в мировой печати с конца 1970-х гг. (RESQ — США, QNAP — Франция, PERFORMS — Япония и др.). Тогда же появились и первые отечественные разработки аналогичного назначения (коллектив В.Ф. Матвеева, Ю.И. Митрофанов, В.М. Порховник, Ю.И. Рыжиков, Т.Н. Яблокова).

Ниже дано описание состава, возможностей и технологии применения пакета МОСТ² (Массовое Обслуживание — СТАционарные задачи), разработанного автором и более 20 лет применяемого при чтении дисциплины «Моделирование вычислительных систем» в Военном инженерно-космическом университете им. А.Ф. Можайского.

Первая версия пакета была реализована в 1977 г. на языке Алгол-60 для машин типа М-220 [66]. Дальнейшие разработки велись на языке ПЛ/1 для машин типа ЕС, программно совместимых с IBM/370. Версия из 83 функциональных процедур в 1987 г. была передана для тиражирования и распространения в Государственный фонд алгоритмов и программ (Эстонское НПО ВТИ) и эксплуатируется по неполным данным приблизительно в 30 организациях. Версия МОСТ-2 (1993 г.) в составе 135 процедур и программной оболочки реализована на ЕС-1066 и работает под управлением VM/SP. Работа с ней описана в руководстве [74].

3.17.2. «Персональный» МОСТ

В 1995 г. выполнен перенос пакета со стационарных машин типа ЕС на IBM-совместимые ПЭВМ. В качестве языка программирования был выбран Фортран-77 (Microsoft, версия 5.0) как единственная на тот момент из распространенных систем программирования, имеющая встроенные средства работы с комплексными переменными и динамического выделения памяти под объекты заданной структуры (ALLOCATABLE-массивы) — см. [75].

3.17.3. Профессиональная версия МОСТа

Профессиональный пакет (МОСТ-FP) позволяет решать не только практические, но и исследовательские задачи теории очередей с наи-

²К одноименной финансовой группе никакого отношения не имеет.

большей гибкостью, точностью и эффективностью. Он открывает пользователю доступ на свой внутренний уровень (137 процедур) Режим работы пользователя с пакетом — сборочное программирование Пакет состоит из следующих групп функциональных процедур (укрупненно):

- основные (анализ конкретных типов СМО);
- приоритетные;
- временные (переход к временным показателям);
- сетевые (расчет сетей обслуживания);
- аппроксимационные (подбор параметров фазовых и иных удобных для последующих вычислений распределений по моментам исходных распределений);
- «фазовые» (построение матриц интенсивностей переходов между микросостояниями фазовых моделей);
- математические разной степени общности (решение систем линейных алгебраических уравнений, обращение матриц специальной структуры, вычисление производящих функций дискретных распределений вероятностей, многократного численного дифференцирования в нуле и т.д.), избавляющие от необходимости привлекать другие библиотеки.

Благодаря такой структуре пакет обеспечивает максимальную гибкость применения при умеренном объеме. Функциональная избыточность его позволяет в каждом конкретном случае использовать процедуру минимальной общности, т.е. решить задачу с минимумом затрат машинных ресурсов, а также организовать взаимное тестирование процедур на областях пересечения их функциональных возможностей.

Пакет соответствует мировому уровню теории очередей и в значительной степени построен на личных научных результатах автора.

Процедуры пакета оформлены в виде библиотеки объектных модулей и вызываются из составленной пользователем главной процедуры, записанной на Фортране-77. В общем случае расчет СМО проходит следующие этапы:

- аппроксимацию исходных распределений по методу моментов (в ряде случаев распределения могут подставляться непосредственно своими моментами или параметрами);

- расчет распределения числа заявок в СМО конкретного типа (в том числе многоканальных немарковских);
- расчет моментов распределения времени ожидания по нему;
- расчет моментов распределения времени пребывания заявки в системе;
- построение по моментам таблицы значений дополнительной функции распределения.

Часть этапов в отдельных случаях может быть опущена. Для приоритетных одноканальных систем моменты распределения времени пребывания в системе заявок каждого типа вычисляются непосредственно.

3.17.4. Общая характеристика МОСТа-F1

Полное использование возможностей пакета (работа с продвинутой его версией) требует углубленного изучения теории вопроса по подготовленной автором монографии «Алгоритмический подход к задачам массового обслуживания». Кроме того, необходимо знать программирование на Фортране-77 для ПЭВМ (справочные сведения приводятся в Руководстве к МОСТу-FP). Однако весьма широкий круг практически важных задач можно решать и без перечисленных познаний. Наряду с полным МОСТом для квалифицированных пользователей, реализована его усеченная версия МОСТ-F1 для непрофессионалов. Она позволяет рассчитывать замкнутые и разомкнутые системы и сети обслуживания однородных заявок с учетом трех моментов исходных распределений (с согласия пользователя МОСТ-F1 по двум моментам может самостоятельно подобрать третий). Допускается также указание конкретных типов распределений (показательное, эрланговское, детерминированное).

По данным, введенным в режиме диалога, и требуемым показателям МОСТ-F1 автоматически формирует Фортран-программу, вызывающую нужные модули пакета, и запускает ее на компиляцию и счет. Выходные показатели (включая ДФР времени пребывания) заносятся в ответный файл в форме, удобной для использования современными графическими средствами.

Пакет использует фазовую аппроксимацию распределений и непосредственно вызывает 21 процедуру МОСТа. Всего в него входят 50 процедур (их список был построен автоматически как транзитивное замыкание первичного).

Документация к МОСТу-F1 состоит из:

- сокращенной версии упомянутой монографии «Задачи и методы расчета систем с очередями», вводящей пользователя в круг основных идей и базовых понятий;
- Руководства, включающего, в частности, описание цикла из трех лабораторных работ для освоения пакета, уяснения его возможностей и использования в учебном процессе вузов и курсов переподготовки специалистов (в профессиональной версии предлагаются 11 таких работ).

Пакет реализован на ПЭВМ, программно совместимых с IBM PC, и работает под управлением MS DOS. Для его применения необходима система программирования Фортран-77 фирмы Microsoft версии 5.0 и старше, в которых реализованы элементы Фортрана-90. Пакет поставляется на одной дискете, упакованной архиватором ARJ. Распакованный МОСТ-F1 содержит шесть подкаталогов.

В каталог BASE включены .exe-модули четырех генераторов Фортран-программ (по основным классам решаемых задач) — всего около 150 Кб. Здесь же находятся библиотека mf1 объектных модулей МОСТа (303 Кб) и ее стандартная опись

Каталог FRG содержит четыре файла заготовок — фрагментов будущих Фортран-программ (опять же по классам задач)

Каталог GO включает два пакетных файла: mogen1 для запуска системы и используемый им fc1g, взаимодействующий с Фортраном.

В каталоге INS находятся 50 файлов с блоками INTERFACE TO, обеспечивающими сопряжение вызывающих программ с вызываемыми

В каталог RES помещены эталонные результаты прогона МОСТ-F1 по четырем основным ветвям, а в TST — протоколы диалога, эти результаты порождающего.

Для запуска пакета следует войти в директорию GO и вызвать файл mogen1.bat. МОСТ-F1 поздоровается с вами и предложит выбрать класс решаемой задачи (расчет изолированной системы или сети — замкнутых либо разомкнутых). Сделанный выбор определит генератор, который будет вести дальнейший диалог. В процессе диалога вводимая с клавиатуры информация

- контролируется на корректность,
- используется для планирования подлежащей вызову цепочки процедур,

- накапливается для непосредственного использования при расчете (записывается в файл `most.dat`, откуда будет считана построенной генератором Фортран-программой).

Контролю подвергаются:

- дисперсии временных распределений (на неотрицательность);
- вычисляемый коэффициент загрузки системы (должен быть меньше единицы);
- элементы строк матрицы вероятностей перехода между узлами сети обслуживания (сумма элементов строки должна быть равна единице).

При нарушении указанных требований система просит повторить ввод или принудительно завершает диалог.

По завершении ввода генератор формирует Фортран-программу, вызывающую нужную цепочку процедур и содержащую требуемые описания и операторы ввода-вывода (в последних используются стандартные имена файлов `most.dat` и `most.res`). Затем запускается пакетный файл `fc1g.bat`, под управлением которого

- программа компилируется,
- из библиотеки к ней присоединяются вызываемые (в том числе транзитивно) модули МОСТА и формируется `exe`-модуль,
- последний запускается на счет,
- после завершения его работы удаляются все рабочие файлы, кроме `most.res`.

Результаты выдаются с необходимыми пояснениями, причем массивы вероятностей выводятся в один столбец (нумерация состояний подразумевается с нуля), а таблица ДФР в два (аргумент и значение функции). Такая редакция (после копирования нужных частей файла стандартными средствами) облегчает взаимодействие с автоматическими построителями графиков типа GRAPHER или GNUPLOT.

Глава 4

Прогнозирование данных

4.1. Проблемы прогнозирования

Исходные данные для задач управления запасами (спрос, цены, временные задержки, ограничения) подвержены заметному влиянию случайных факторов и, кроме того, могут меняться во времени на интервале планирования. Для решения реальных задач управления запасами необходимо *прогнозирование* перечисленных данных. Стратегия предсказаний зависит от компетентности прогнозиста, возможностей компьютера, его математического обеспечения, наличия и представительности данных, срока прогноза. Быстрая реакция на ошибки прогнозирования предпочтительнее сверхточного прогноза.

Методика прогнозирования характеристик временных последовательностей слабо связана с предметной областью и освещена в любом руководстве по математической статистике. Мы опишем лишь основные идеи прогнозирования применительно к спросу, а за математическими подробностями отошлем читателя к книге [84], содержащей описание не только методов, но и реализующих их пакетов статистических подпрограмм.

В условиях товарного производства потребности трансформируются в *спрос*. Спрос на средства производства формируется под воздействием множества факторов: научно-технический прогресс, темпы роста

объемов производства, нормы расхода, взаимозаменяемость и т.д. При решении задач управления запасами спрос надо *прогнозировать*. Эту работу следует проводить в двух направлениях:

- 1) уточнение объемов производства и строительства и разработка прогрессивных норм расхода продукции на производственных предприятиях;
- 2) статистическое прогнозирование спроса у потребителей, на складах и базах территориальных управлений и управлений технологической комплектации — в системах оптово-розничной торговли.

Расчетно-аналитический способ должен сочетаться со статистическими методами.

Трудность применения расчетных методов заключается в том, что к моменту составления прогноза предприятия еще не имеют планов-заказов (между прочим и потому, что спрос на их продукцию является основой для составления планов). Дополнительные сложности возникают в моделях с «замкнутым» спросом, зависящим от ценовой политики (а спрос всегда зависит от ценовой политики) и от степени удовлетворения потребителей, т.е. от стратегии управления запасами.

Свою сложность — неполную наблюдаемость спроса — имеют и статистические методы. Причинами этого являются:

- укрупненная фиксация спроса в стоимостном или натуральном исчислении (штуки, тонны, погонные или квадратные метры) без детальной разбивки по номенклатурам;
- неучет неудовлетворенных заявок (в особенности в торговле и для дефицитных товаров).

В настоящее время в связи с внедрением электронных баз данных создались объективные условия для улучшения наблюдаемости статистических объектов.

4.2. Характеристики временных рядов

Временные ряды описываются двумя группами показателей: характеристики сечений и характеристики динамики. К первым можно отнести средний уровень, дисперсию, среднее квадратическое отклонение, коэффициент вариации для выбранного момента времени; ко вторым

— абсолютный прирост (конечную разность), коэффициент (темп) роста, автокорреляцию и др. Обычно прогнозирование ведется на уровне средних и дисперсий. Если таковые оценены, то далее принимают дополнительную гипотезу о виде функции распределения.

При недостатке статистической или аналитической информации используются *экспертные* методы, которые опираются на опыт и интуицию специалистов.

Оценки могут быть точечными, интервальными и квантильными. Статистический анализ экспертных оценок «по полной программе» предполагает сложную обработку мнений экспертов. В качестве результирующей точечной оценки, например, рекомендуется применять медиану.

В случае двухточечных оценок эксперт определяет x_{\min} и x_{\max} . Функция плотности бета-распределения имеет вид

$$f(x) = C(x - x_{\min})^{\alpha}(x_{\max} - x)^{\gamma}.$$

В практике сетевого планирования (при оценке длительности работ) рекомендуется принять $\alpha = 1$, $\gamma = 2$. Тогда $C = 12/(x_{\max} - x_{\min})$, а матожидание и среднее квадратическое отклонение подсчитываются согласно

$$\bar{x} = (2x_{\max} + 3x_{\min})/5, \quad \sigma = (x_{\max} - x_{\min})/5.$$

Имеются варианты получения тех же характеристик с учетом наиболее вероятного значения случайной величины x_{prob} (трехточечные оценки [40]):

$$\bar{x} = \frac{1}{6}(x_{\text{opt}} + 4x_{\text{prob}} + x_{\text{pes}}), \quad \sigma = \frac{1}{6}(x_{\text{opt}} - x_{\text{pes}}).$$

Здесь x_{opt} — оптимистическая оценка, x_{pes} — пессимистическая, x_{prob} — наиболее вероятная.

Наиболее эффективным методом краткосрочного прогнозирования спроса является анализ временных рядов. Изучение изменения потребностей во времени является основой информационного обеспечения задач управления запасами.

Большинство временных рядов характеризуется колебаниями, которые можно разделить на «вековые» (тренд), сезонные и случайные. Тренд описывает усредненную *тенденцию* изучаемого процесса (его детерминированную компоненту). Линию тренда можно отождествить с линией регрессии в корреляционном анализе. Тренд учитывается лучше, если информация берется по нескольким периодам:

$$\bar{x}_k = \alpha x^{k-1} + 2(1 - \alpha)\bar{x}^{k-1} - (1 - \alpha)\bar{x}^{k-2}.$$

Здесь через x_k обозначен спрос в k -й период, а \bar{x}_k — его прогноз.

Другие динамические показатели устанавливаются сравнением данных о спросе в последовательные периоды, длительность которых определяется спецификой задачи. Коэффициент роста показывает, во сколько раз уровень (среднее значение) заданного периода больше уровня базисного периода. Средний темп роста определяется как геометрическое среднее коэффициентов роста $\{k_i\}$: $\bar{k} = \sqrt[k_1 k_2 \dots k_n]$. Прогнозируемый уровень $x_{M+r} = x_M \cdot \bar{k}^r$.

4.3. Метод скользящей средней

Наиболее простым методом выявления тенденции временного ряда является сглаживание его уровней. В методе скользящей средней фактические уровни заменяются рядом средних, которые рассчитываются для подвижных интервалов фиксированной длины и относятся к середине каждого из них. Чем продолжительнее интервал сглаживания, тем сильнее усреднение и больше поглощаются колебания. В практике вычисления чаще всего используются трех-, пяти- и семичленные формулы. В общем случае

$$\bar{x}_t = \frac{1}{2m+1} \sum_{i=t-2m-1}^{t-1} x_i.$$

При большом числе наблюдений удобнее пользоваться рекуррентной формулой

$$\bar{x}_t = \bar{x}_{t-1} + (x_{t-1} - x_{t-2m-2}) / (2m+1).$$

Метод скользящей средней целесообразно применять лишь для процессов с незначительным изменением средних во времени, а также для краткосрочного прогнозирования. Здесь все данные имеют одинаковый вес $1/(2m+1)$. Он требует хранения данных за $2m+1$ периодов, что при работе с многими тысячами номенклатур становится весьма обременительным. Известны варианты этого метода с отнесением результата усреднения к концу периода, с многофазным сглаживанием и др [77].

4.4. Экспоненциальное сглаживание

Этот метод состоит в сглаживании временного ряда с помощью взвешенной скользящей средней, в которой веса подчиняются экспонен-

циальному закону. Значение прогнозируемого параметра определяется по формуле

$$\bar{x}_t = \bar{x}_{t-1} + \alpha(x_{t-1} - \bar{x}_{t-1}), \quad (4.4.1)$$

где коэффициент α , $0 < \alpha < 1$, отражает степень учета рассогласования между прогнозом \bar{x}_{t-1} и фактической потребностью x_{t-1} в предыдущий период. Эту формулу можно переписать в рекуррентном виде:

$$\bar{x}_t = \alpha x_{t-1} + (1 - \alpha)\bar{x}_{t-1}. \quad (4.4.2)$$

Выполнив последовательные подстановки, приходим к формуле

$$\bar{x}_t = \alpha \sum_{i=1}^t (1 - \alpha)^{i-1} x_{t-i} + (1 - \alpha)^t \bar{x}_0, \quad (4.4.3)$$

которая объясняет название и смысл метода — показательными убывающими весами «старых» наблюдений. Метод обеспечивает устойчивую реакцию системы на изменение потребностей, скорость реакции регулируется коэффициентом α . Для медленно меняющейся потребности можно рекомендовать $\alpha = 0.1$, для более динамичной — $\alpha = 0.3-0.5$. Очевидно, α должна убывать с увеличением длины периодов. Возможным подходом к определению оптимального α является минимизация суммы квадратов отклонений прогноза от фактического спроса по имеющимся реальным данным.

4.5. Аналитическое выравнивание

Аналитическое выравнивание временных рядов аналогично определению теоретической линии регрессии в корреляционном анализе. Первая задача состоит в выборе типа кривой: многочлены, дробно-рациональные функции, экспоненты, логистические кривые и др. Вид кривых предпочтительно определить из теоретических соображений: внутренней логики процесса и его связей с окружающим миром. Помогает также анализ конечных разностей и их относительных значений [77, с. 266]. При выравнивании многочленами полезно предварительное вычисление конечных разностей: порядок многочлена равен наивысшему порядку ненулевых разностей. На приемлемость экспоненциальной аппроксимации указывает близкая к линейной зависимость от времени логарифмов исходных данных.

Циклические (точнее, гармонические) кривые роста призваны отразить поведение спроса, имеющего сезонные или иные периодические колебания. Любые циклические колебания можно представить в виде суммы простых гармоник, т.е. разложить в ряд Фурье:

$$x = a_0 + \sum_{i=1}^m (a_i \cos i\omega t + b_i \sin i\omega t).$$

Сезонные временные ряды разлагаются на следующие компоненты:

$$x = u + w + \varepsilon,$$

где u — тенденция (тренд); w — сезонная волна; ε — случайные колебания с нулевым средним. Начало сезона может определяться особыми условиями (погодными, датой Пасхи в текущем году и т.п.).

Затем решается вторая задача — определяются параметры регрессии (методом наименьших квадратов). Выбранная кривая проверяется на значимость по критерию Фишера. После окончательного выбора кривой прогнозирование осуществляется на основе экстраполяции тенденции.

Среди экономико-математических методов моделирования потребностей следует выделить многофакторный анализ. Согласно [89], процедура предсказания для военных самолетов базируется на определении по прошлому опыту среднего числа запасных частей каждого вида, необходимых на час полета. Общее число часов полета для всех самолетов оценивается на основе оперативных планов и умножается на удельный расход деталей. Последний очень трудно надежно оценить при низком спросе на запасные части — тип самолетов может быть снят с эксплуатации до того, как будет собрана достаточная информация.

4.6. Определение качества прогноза

Для оценки надежности предсказаний делается проверка их на контрольной последовательности, *не вошедшей в использованное при расчете параметров множество*. Целесообразно планировать расстановку экспериментальных точек так, чтобы при ограниченном числе экспериментов получить оценки для искомых зависимостей с наименьшей погрешностью. Расположение точек тем лучше, чем больше отличается от нуля определитель системы нормальных уравнений в методе наименьших квадратов.

4.7. Корреляция

Коррелированность спроса в системах снабжения проявляется в трех основных формах:

- автокорреляция по времени;
- корреляция спроса на одном складе между разными номенклатурами;
- коррелированность спроса на разных складах

Зависимость спроса при управлении запасами внутри промышленного предприятия практически неизбежна. Она определяется технологией производства, необходимостью комплектных поставок деталей, узлов, компонент, реагентов, зависимостью спроса на запасные части от спроса на основной продукт и т.д.

Спрос может быть коррелированным и в торговле. Положительная корреляция реализуется при подборе комплектов, ансамблей и аксессуаров (готовая одежда, автомобили, компьютеры), отрицательная — для взаимозаменяемых предметов. Многономенклатурные модели с коррелированным спросом сложны и редко обсуждаются в литературе

Часть II

Базовые модели

Глава 5

Однопродуктовые модели

5.1. Детерминированный спрос

5.1.1. Общий случай

Рассмотрим случай постоянной интенсивности спроса λ и поставок μ . Эта модель отличается растянутостью процесса поставок во времени и хорошо описывает работу трубопроводов. Попутно отметим, что их доля в общем объеме транспортных услуг за 1965–1985 гг. возросла в 6 раз, а средняя дальность перекачки нефти и газа — в 3.2 раза. Технически решены проблемы передвижения по продуктопроводам и сыпучих грузов: угля, руды в виде пульпы, цемента, песка, гравия, щебня, зерна. График изменения уровня запаса показан на рис. 5.1.

Полный цикл работы системы имеет продолжительность T . Обозначим через S предельный запас на складе. Считая расходы на хранение и на штрафы пропорциональными среднему запасу (дефициту) и времени их существования с коэффициентами h и d соответственно,

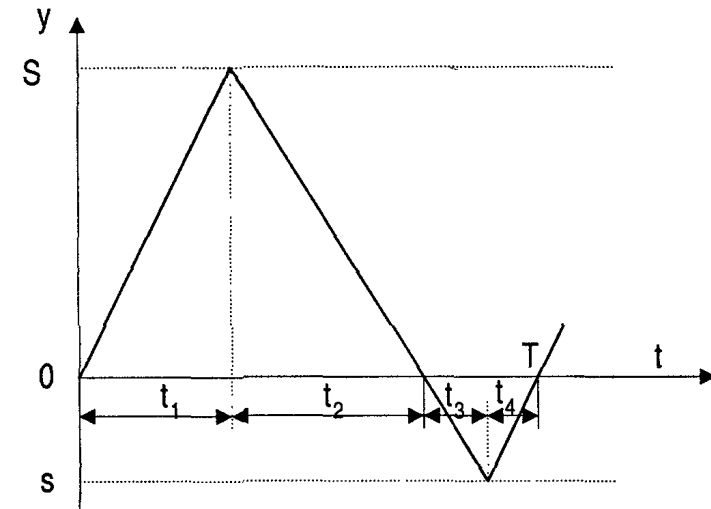


Рис. 5.1. Динамика запаса при детерминированном спросе

получаем для затрат за цикл выражение

$$L_T = g + h \int_0^{t_1+t_2} y(t) dt - d \int_{t_1+t_2}^T y(t) dt,$$

где g — фиксированные расходы, связанные с запуском производства (организацией поставки). Очевидно, что текущий запас

$$y(t) = \begin{cases} (\mu - \lambda)t & \text{при } 0 \leq t \leq t_1, \\ S - \lambda(t - t_1) & \text{при } t_1 < t \leq t_1 + t_2 + t_3, \\ s + (\mu - \lambda)(t - t_1 - t_2 - t_3) & \text{при } t_1 + t_2 + t_3 < t \leq T. \end{cases}$$

Основания треугольников на рис. 5.1 равны высоте, умноженной на $[1/(\mu - \lambda) + 1/\lambda] = \mu/(\lambda(\mu - \lambda))$. Соответственно

$$T = (S - s) \frac{\mu}{\lambda(\mu - \lambda)},$$

откуда

$$s = S - \frac{\lambda(\mu - \lambda)}{\mu} T. \quad (5.1.1)$$

«Импульс» положительного запаса (площадь треугольника) равен

$$\frac{S^2}{2} \left(\frac{1}{\mu - \lambda} + \frac{1}{\lambda} \right).$$

Аналогично выражается импульс отрицательного запаса. С учетом (5.1.1) затраты за цикл составят

$$L_T = g + \frac{h\mu S^2}{2\lambda(\mu - \lambda)} + \frac{d\mu}{2\lambda(\mu - \lambda)} \left[\frac{\lambda}{\mu}(\mu - \lambda)T - S \right]^2,$$

а в единицу времени

$$\begin{aligned} L &= \frac{g}{T} + \frac{h\mu S^2}{2T\lambda(\mu - \lambda)} + \frac{d\mu}{2T\lambda(\mu - \lambda)} \left[\frac{\lambda}{\mu}(\mu - \lambda)T - S \right]^2 \\ &= \frac{1}{T} \left[g + \frac{(d+h)\mu S^2}{2\lambda(\mu - \lambda)} \right] + \frac{d\lambda}{2\mu}(\mu - \lambda)T - dS. \end{aligned}$$

Найдем частные производные от L по S и T и приравняем их нулю:

$$\begin{aligned} \frac{\partial L}{\partial S} &= \frac{(d+h)\mu S}{\lambda(\mu - \lambda)T} - d = 0, \\ \frac{\partial L}{\partial T} &= \frac{d\lambda}{2\mu}(\mu - \lambda) - \frac{1}{T^2} \left[g + \frac{(d+h)\mu S^2}{2\lambda(\mu - \lambda)} \right] = 0. \end{aligned}$$

Совместное решение этих уравнений дает для оптимальных S и T условия

$$\begin{aligned} S^* &= \sqrt{\frac{2\lambda g(1 - \lambda/\mu)}{h(1 + h/d)}}, \\ T^* &= \sqrt{\frac{2g(1 + h/d)}{\lambda h(1 - \lambda/\mu)}}. \end{aligned}$$

При этом достигается минимум затрат в единицу времени

$$L^* = \sqrt{\frac{2\lambda g h(1 - \lambda/\mu)}{1 + h/d}}.$$

Момент запуска производства определяется достижением дефицита

$$s^* = -\frac{1}{d} \sqrt{\frac{2\lambda g}{s} \frac{1 - \lambda/\mu}{1 + h/d}}.$$

5.1.2. Предельные варианты

Из полученных соотношений как частные случаи легко выводятся более известные формулы теории запасов. При высоком штрафе можно принять $h/d \approx 0$. Тогда

$$S^* = \sqrt{\frac{2\lambda g}{h}(1 - \lambda/\mu)}, \quad (5.1.2)$$

$$T^* = \sqrt{\frac{2g}{\lambda h(1 - \lambda/\mu)}}, \quad (5.1.3)$$

$$L^* = \sqrt{2\lambda g h(1 - \lambda/\mu)}, \quad (5.1.4)$$

а недостачи полностью исключаются. Другой частный случай соответствует очень высокой интенсивности восполнения запаса (например, при поставке всей партии с вышестоящего склада). В этой модели $\lambda/\mu \rightarrow 0$ и

$$S^* = \sqrt{\frac{2\lambda g}{h(1 + h/d)}}, \quad (5.1.5)$$

$$T^* = \sqrt{\frac{2g(1 + h/d)}{\lambda h}}, \quad (5.1.6)$$

$$L^* = \sqrt{\frac{2\lambda g h}{1 + h/d}}. \quad (5.1.7)$$

Наибольшее применение получили формулы Уилсона, выведенные при обоих рассмотренных допущениях

$$S^* = \sqrt{2\lambda g/h}, \quad (5.1.8)$$

$$T^* = \sqrt{2g/(\lambda h)}, \quad (5.1.9)$$

$$L^* = \sqrt{2\lambda g h}. \quad (5.1.10)$$

Помимо рассмотренных выше показателей, представляют интерес еще два — объем заказываемой партии q и точка заказа s при задержке τ

между заказом и началом поставки. Первый из них равен спросу λT за период, так что для общего случая

$$q^* = \sqrt{\frac{2\lambda g(1+h/d)}{h(1-\lambda/\mu)}},$$

а при $\lambda/\mu \rightarrow 0$

$$q^* = \sqrt{\frac{2\lambda g}{h}(1+h/d)}. \quad (5.1.11)$$

График функции $L(q)$ в окрестности минимума затрат является весьма пологим. Это позволяет подстраивать q с учетом дополнительных соображений (кратность стандартным упаковкам, удобная периодичность $T = q/\lambda$) практически без увеличения расходов. В моделях с высоким штрафом $q \approx S$.

Точка заказа при задержке поставок определяется как $s + \lambda\tau$, где τ — средняя задержка.

5.1.3. Ошибки в параметрах и функция затрат

Расчет параметров стратегий управления запасами по формулам данной главы обеспечивает минимум затрат при интенсивностях спроса и восстановления и стоимостных параметрах, известных с достаточной точностью. В противном случае погрешности в их определении приводят к выбору параметров стратегии, отличных от оптимальных, и как следствие — к некоторому увеличению затрат. С другой стороны, иногда приходится идти на заведомое отклонение параметров стратегии от теоретически оптимальных, например, при ограниченной вместимости складов, заданной вышестоящими органами периодичности; необходимости обеспечить полную загрузку транспортных средств (вагонов, контейнеров) большой емкости, не совпадающей с оптимальным объемом партии.

В модели с детерминированным и полностью удовлетворяемым спросом постоянной интенсивности затраты в единицу времени подсчитываются согласно

$$L = \frac{\lambda g(1-\lambda/\mu)}{S} + \frac{hS}{2} \quad (5.1.12)$$

(эта формула легко получается как частный случай модели, рассмотренной в начале данной главы). Соответственно оптимальное решение S^* дается формулой (5.1.2), а минимум затрат — (5.1.4)

Пусть максимальный запас выбран с относительной погрешностью δ_S , так что затраты

$$L = \frac{\lambda g(1-\lambda/\mu)}{S(1+\delta_S)} + \frac{hS}{2}(1+\delta_S)$$

Абсолютное приращение затрат составит

$$\begin{aligned} \Delta L &= L - L^* = \frac{\lambda g(1-\lambda/\mu)}{S^*} [(1+\delta_S)^{-1} - 1] + hS^*\delta_S/2 \\ &= S^*\delta_S \left[\frac{h}{2} - \frac{1}{1+\delta_S} \frac{\lambda g(1-\lambda/\mu)}{(S^*)^2} \right]. \end{aligned}$$

Подставив сюда значение S^* из (5.1.2), убеждаемся, что

$$\begin{aligned} \Delta L &= \delta_S \sqrt{\frac{2\lambda g}{h}(1-\lambda/\mu)} \left[\frac{h}{2} - \frac{1}{1+\delta_S} \frac{\lambda g(1-\lambda/\mu)}{2\lambda g(1-\lambda/\mu)/h} \right] \\ &= \delta_S \sqrt{\frac{2\lambda g}{h}(1-\lambda/\mu)} \left(\frac{h}{2} - \frac{1}{1+\delta_S} \frac{h}{2} \right) \end{aligned}$$

Для малых δ_S можно принять $1/(1+\delta_S) \approx 1 - \delta_S$. В этом случае

$$\Delta L = \frac{1}{2}\delta_S^2 \sqrt{2\lambda g h(1-\lambda/\mu)}.$$

Обратившись к (5.1.4), замечаем, что множитель при $\delta_S^2/2$ равен L^* . Следовательно, относительное увеличение затрат при неточном выборе L в окрестности оптимума

$$\delta L = \frac{\Delta L}{L^*} = \frac{1}{2}\delta_S^2. \quad (5.1.13)$$

Найдем зависимость δL от ошибок в исходных данных для расчета S^* . Прологарифмировав (5.1.4), имеем

$$\ln S^* = \frac{1}{2} \{ \ln 2 + \ln g - \ln h + \ln[\lambda(1-\lambda/\mu)] \}.$$

Дифференцирование этого выражения и замена дифференциалов конечными приращениями дают

$$\begin{aligned} \frac{\Delta S}{S^*} &= \frac{1}{2} \left[\frac{\Delta g}{g} - \frac{\Delta h}{h} + \frac{(1 - 2\lambda/\mu)\Delta\lambda}{\lambda(1 - \lambda/\mu)} + \frac{(\lambda/\mu)^2 \cdot \Delta\lambda}{\lambda(1 - \lambda/\mu)} \right] \\ &= \frac{1}{2} \left(\frac{\Delta g}{g} - \frac{\Delta h}{h} + \frac{\mu - 2\lambda}{\mu - \lambda} \frac{\Delta\lambda}{\lambda} + \frac{\lambda}{\mu - \lambda} \frac{\Delta\mu}{\mu} \right), \end{aligned}$$

откуда

$$\delta_S = \frac{1}{2} \left[\delta_g - \delta_h + \frac{(\mu - 2\lambda)\delta_\lambda + \lambda\delta_\mu}{\mu - \lambda} \right].$$

Если знак допущенной погрешности неизвестен, то при δ_h следует поставить плюс.

5.1.4. Зависимые параметры

Входящие в формулы данного раздела экономические параметры можно считать постоянными лишь в первом приближении — в некотором диапазоне объемов партий q . Так, цена заказа g и цена хранения h могут быть ступенчатыми возрастающими функциями q (при увеличении q могут потребоваться дополнительные затраты на организацию производства, новые складские емкости). В подобных случаях необходимо задаться начальным значением q_0 (например, середина разумного диапазона), рассчитать $h(q_0)$ и $g(q_0)$ и по приведенным выше формулам найти q_1 . Если $h(q_0) = h(q_1)$ и $g(q_0) = g(q_1)$, полученное значение q является окончательным. В противном случае процесс повторяется при $h(q_1)$ и $h(q_1)$, и т.д. Расчет, как правило, сходится быстро.

Поставщик в отличие от потребителей заинтересован в отгрузке максимальной партии, так как это освобождает его от значительной части затрат на содержание запаса. Поэтому он делает ступенчатые скидки, которые отражаются в модели через цену h хранения запаса. Вопрос о скидках обычно рассматривается в двух вариантах [89, с. 82–87]. В первом скидка назначается на каждую единицу закупаемого товара в зависимости от общего объема партии q . Для $q \in [y_i, y_{i+1})$ цена штуки равна b_i , последовательность $\{b_i\}$ — убывающая. Без потери общности $y_1 = 0$, $y_{n+1} = \infty$. Целевая функция определяется по интервалам:

$$L_i(q) = \frac{f \cdot b_i q}{2} + \frac{\lambda g}{q} + \lambda b_i, \quad y_i \leq q < y_{i+1}.$$

Здесь f — цена хранения (в долях цены) и g — стоимость заказа, не зависящая от объема партии. Интенсивность спроса λ считается постоянной, заказ выдается при снижении запаса до нуля. Функция затрат получается разрывной комбинацией кривых $L(q)$, и оптимальное решение достигается в минимуме одной из них (q^* , рис. 5.2а) или в точке разрыва (q' , рис. 5.2б).

Алгоритм оптимизации:

- 1) Вычислить $q_n = \sqrt{2\lambda g / (f \cdot b_n)}$. Если $q_n \geq y_{n-1}$, q_n оптимально.
- 2) Если $q_n < y_{n-1}$, рассчитать q_{n-1} .
- 3) Если $q_{n-1} \geq y_{n-2}$, сравнить $L(q_{n-1})$ с $L(y_{n-1})$, и минимум из них даст оптимальное q .
- 4) Если $q_{n-1} < y_{n-2}$, вычислить q_{n-2} .
- 5) Если $q_{n-2} \geq y_{n-3}$, минимум из $L(q_{n-2})$, $L(y_{n-1})$ и $L(y_{n-2})$ даст оптимальное q , и т.д.

Процедура требует не более чем n шагов.

Во втором случае дается «дифференциальная скидка» на каждую дополнительную единицу, приобретаемую сверх очередного порога. Здесь оплата хранения «пороговой» партии может быть отнесена к фиксированной составляющей стоимости заказа, и формула Уилсона дает оптимальное приращение относительно этого порога. Итоговая функция затрат имеет вид нижней огибающей семейства кривых, а глобальный минимум совпадает с наименьшим из минимумов образующих кривых. Примеры соответствующих графиков показаны на рис. 5.3.

В общем случае все параметры модели могут зависеть от объема партии. Пусть цена закупки есть $b(q)$, цена хранения $h = f \cdot b(q)$. Стоимость выполнения $\hat{g} = g + qb(q)$. Целевая функция

$$L(q) = fb(q)\frac{q}{2} + \frac{g + qb(q)}{q/\lambda} = fb(q)\frac{q}{2} + \frac{g\lambda}{q} + \lambda b(q).$$

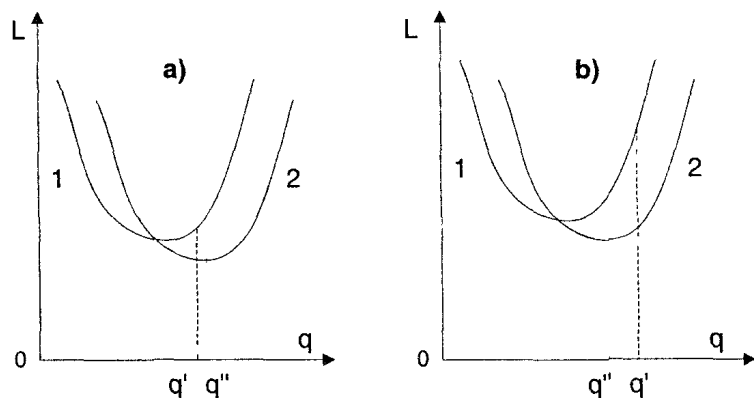


Рис. 5.2. Минимальные затраты при интегральной скидке

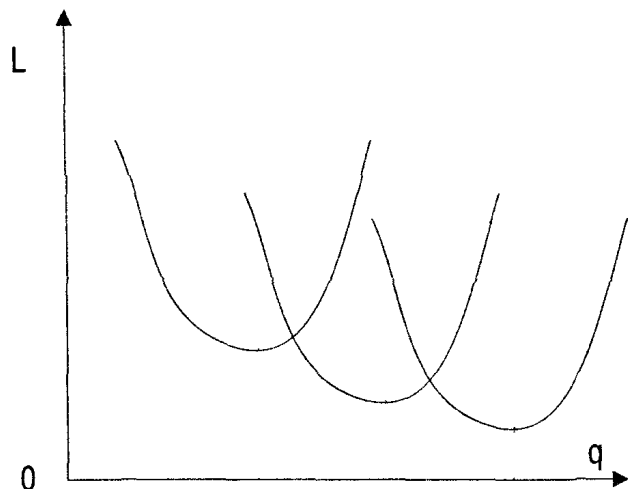


Рис. 5.3. Минимальные затраты при дифференциальной скидке

Решение зависит от вида $b(q)$. В случае $b(q) = b_0 - b_1q$, $b_0 \gg b_1$,

$$q = \sqrt{\frac{2\lambda g + 2fb_1q^3}{fb_0 - 2\lambda b_1}}$$

и находится методом итераций.

Практический интерес вызывает задача определения продажной цены изделия v с учетом зависимости от нее интенсивности спроса λ .

Будем считать, что спрос обеспечивается полностью, а себестоимость единицы продукции составляет u . Используя для затрат на снабжение формулу Уилсона (5.1.10), можно для дохода в единицу времени записать выражение

$$R = (v - u)\lambda(v) - \sqrt{2gh\lambda(v)}.$$

Максимальный доход достигается при $\partial R/\partial v = 0$, или при

$$v = u - \lambda(v)/\lambda'(v) + \sqrt{gh/2\lambda(v)}.$$

Это уравнение решается методом итераций

Рассмотрим конкретизацию данного подхода. В модели с дефицитом и мгновенными поставками $\partial \text{доход} / \partial \text{цикл}$

$$R_T = c\lambda T - g - \frac{hS^2}{2\lambda},$$

длина цикла $T = (S - s)/\lambda$. Соответственно доход в единицу времени

$$R = c\lambda - \frac{\lambda}{S - s} \left[g + \frac{hS^2}{2\lambda} \right].$$

Будем считать, что интенсивность спроса определяется относительной удовлетворенностью потребителей $-s/(S - s)$ согласно

$$\lambda = \frac{\lambda_0}{1 - \frac{\alpha s}{S - s}} = \frac{\lambda_0(S - s)}{S - (\alpha + 1)s},$$

где α — постоянный множитель. С учетом этой зависимости

$$R = \frac{c\lambda_0(S - s) - \lambda_0 g}{S - (\alpha + 1)s} - \frac{hS^2}{2(S - s)}.$$

Дифференцирование $R(s, S)$ дает условия оптимальности

$$\begin{aligned} \frac{c\lambda_0}{S - (\alpha + 1)s} - \frac{c\lambda_0(S - s) - \lambda_0 g}{(S - (\alpha + 1)s)^2} - \frac{S}{S - s} + \frac{h}{2} \left(\frac{S}{S - s} \right)^2 &= 0, \\ -\frac{c\lambda_0}{S - (\alpha + 1)s} + \frac{(\alpha + 1)(c\lambda_0(S - s) - \lambda_0 g)}{(S - (\alpha + 1)s)^2} - \frac{h}{2} \left(\frac{S}{S - s} \right)^2 &= 0. \end{aligned}$$

С помощью математического пакета Maple V удалось установить, что оптимальный верхний порог S^* является корнем кубического уравнения

$$cahz^3 - (\alpha + 1)ghz^2 - 2\lambda g^2 = 0.$$

Начальное приближение для s^* выберем из условия компенсации потерь cs от неудовлетворенного спроса за цикл экономией на организации поставок. Максимизируя сумму $gS/(S-s) + cs$, получаем

$$s = S - \sqrt{gS/c}$$

5.2. Хранение продукта при естественной убыли

Предположим, что спрос на предмет хранения имеет постоянную интенсивность λ , скорость восполнения составляет μ , а скорость естественной убыли в каждый данный момент пропорциональна (с коэффициентом γ) наличному запасу. Динамика изменения запаса за один полный цикл длительности T показана на рис 5.4

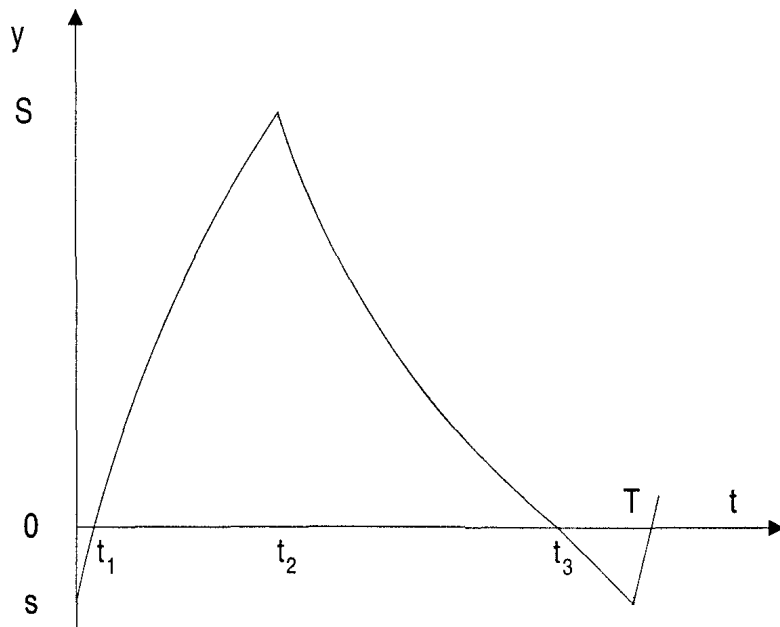


Рис 5.4 Динамика запаса при естественной убыли

и описывается дифференциальными уравнениями

$$\frac{dy(t)}{dt} = \begin{cases} \mu - \lambda, & 0 \leq t < t_1, \\ \mu - \lambda - \gamma y, & t_1 \leq t < t_2, \\ -\lambda - \gamma y, & t_2 \leq t < t_3, \\ -\lambda, & t_3 \leq t < T \end{cases}$$

с решениями

$$y(t) = \begin{cases} s + (\mu - \lambda)t, & 0 \leq t < t_1, \\ C_1 e^{-\gamma(t-t_1)} + \frac{\mu - \lambda}{\gamma}, & t_1 \leq t < t_2, \\ C_2 e^{-\gamma(t-t_2)} - \frac{\lambda}{\gamma}, & t_2 \leq t < t_3, \\ -\lambda(t - t_3), & t_3 \leq t < T. \end{cases}$$

Из условий $y(0) = s$ (уровень начала восполнения) и $y(t_2) = S$ (максимальный запас) определяем постоянные интегрирования. Окончательно

$$y(t) = \begin{cases} s + (\mu - \lambda)t, & 0 \leq t < t_1, \\ [1 - e^{-\gamma(t-t_1)}] (\mu - \lambda) / \gamma, & t_1 \leq t < t_2, \\ (S + \lambda/\gamma) e^{-\gamma(t-t_2)} - \lambda/\gamma, & t_2 \leq t < t_3, \\ -\lambda(t - t_3), & t_3 \leq t < T. \end{cases}$$

Вычислим затраты L_T за цикл. Очевидно, затраты на хранение

$$L_T^{(h)} = h \int_{t_1}^{t_3} y(t) dt, \quad (5.2.1)$$

а сумма штрафов

$$L_T^{(d)} = -d \left[\int_0^{t_1} y(t) dt + \int_{t_3}^T y(t) dt \right].$$

Потери от убыли можно определить по разности между поставленным за цикл количеством продукта и его фактическим потреблением, так что

$$L_T^{(\gamma)} = c(\mu t_2 - \lambda T).$$

Добавим к этим затратам расходы на организацию цикла g и перейдем к затратам в единицу времени. Исключая слагаемое $-c\lambda T/T = -c\lambda$ как

не зависящее от выбора управляющих параметров, приходим к

$$\tilde{L} = \left\{ g + h \int_{t_1}^{t_3} y(t) dt - d \left[\int_0^{t_1} y(t) dt + \int_{t_3}^T y(t) dt \right] + c\mu t_2 \right\} / T. \quad (5.2.2)$$

Вычислим входящие в это выражение интегралы. Из условий $y(t_1) = 0$, $y(t_2) = S$, $y(t_3) = 0$, $y(T) = s$ можно получить временные параметры

$$\begin{aligned} t_1 &= s/(\lambda - \mu), \\ t_2 &= t_1 - \frac{1}{\gamma} \ln \left(1 - \frac{\gamma S}{\mu - \lambda} \right), \\ t_3 &= t_2 + \frac{1}{\gamma} \ln \left(1 + \frac{\gamma S}{\lambda} \right), \\ T &= t_3 + s/\lambda. \end{aligned} \quad (5.2.3)$$

Разобьем интеграл в (5.2.1) на две части и найдем их величину. Первая часть

$$\begin{aligned} \int_{t_1}^{t_2} y(t) dt &= \frac{\mu - \lambda}{\gamma} \int_0^{t_2 - t_1} (1 - e^{-\gamma t}) dt = \frac{\mu - \lambda}{\gamma} \left(t + \frac{1}{\gamma} e^{-\gamma t} \right) \Big|_0^{t_2 - t_1} \\ &= \frac{\mu - \lambda}{\gamma} \left[-\frac{1}{\gamma} \ln \left(1 - \frac{\gamma S}{\mu - \lambda} \right) - \frac{S}{\mu - \lambda} \right]. \end{aligned}$$

Аналогично

$$\int_{t_2}^{t_3} y(t) dt = \frac{1}{\gamma} \left[S - \frac{\lambda}{\gamma} \ln \left(1 + \frac{\gamma S}{\lambda} \right) \right].$$

Наконец, из элементарных соображений

$$\int_0^{t_1} y(t) dt = -\frac{s^2}{2(\mu - \lambda)}, \quad \int_{t_3}^T y(t) dt = -\frac{s^2}{2\lambda}.$$

Из (5.2.3) можно получить явное выражение длины цикла

$$T = \frac{\mu s}{\lambda(\lambda - \mu)} + \frac{1}{\gamma} \ln \frac{1 + \gamma S/\lambda}{1 - \gamma S/(\mu - \lambda)}.$$

Подстановка полученных выражений в (5.2.2) дает искомую функцию затрат.

Представляют интерес частные случаи рассмотренной модели. При бесконечной скорости восполнения (поставки с вышестоящего склада) в начальный момент времени уровень запаса скачком поднимается от s до S . Дальнейшее его изменение происходит так же, как и на интервалах $[t_2, t_3]$ и $[t_3, T)$ в общем случае. Функция затрат принимает вид

$$L = \frac{1}{T} \left[h \int_{t_2}^{t_3} y(t) dt + d \int_{t_3}^T y(t) dt + g + c(S - s - \lambda T) \right],$$

где

$$T = -\frac{s}{\lambda} + \frac{1}{\gamma} \ln(1 + \gamma S/\lambda),$$

а $t_1 = t_2 = 0$

Высокая цена штрафов, как и в предыдущем разделе, приводит к недопущению отрицательных уровней, т.е. $s = 0$. В этом случае

$$L = \frac{1}{T} \left[h \int_{t_1}^{t_3} y(t) dt + g + c\mu(t_2 - t_1) - \lambda T \right],$$

где

$$T = \frac{1}{\gamma} \ln \frac{1 + \gamma S/\lambda}{1 - \gamma S/(\mu - \lambda)},$$

а $t_1 = 0$, $t_3 = T$

Наконец, комбинируя особенности обоих случаев, получаем модель с неотрицательным уровнем запасов и мгновенной поставкой, для которой $t_1 = t_2 = 0$, $t_3 = T$, затраты

$$L = \frac{1}{T} \left[h \int_{t_2}^{t_3} y(t) dt + g + c(S - \lambda T) \right],$$

а длина цикла

$$T = \frac{1}{\gamma} \ln(1 + \gamma S/\lambda).$$

Минимизацию затрат, получаемых по формулам этого раздела, приходится проводить численно. В качестве разумных начальных значений для S , T и s можно принять оптимальные параметры модели без естественной убыли

5.3. Стареющий продукт

К стареющим продуктам относятся пища, лекарства, фотоматериалы, консервированная кровь, а также модные и иные товары с высоким темпом морального старения (например, персональные компьютеры, аксессуары к ним, программное обеспечение). Управление запасами подобных товаров реализуется при следующих допущениях:

- рассчитываются периодические стратегии при заданной длине периода T ;
- заявка на восполнение подается в начале периода;
- весь прибывший продукт считается «новым»,
- отпуск партий делается по принципу «первый пришел — первый вышел»;
- спрос в последовательные периоды — независимая случайная переменная с общим распределением;
- неудовлетворенный спрос задалживается;
- продукт, не востребованный в течение m периодов, теряет годность;
- в целевую функцию включаются расходы на хранение, дефицит и списание устаревших партий.

Состояние запаса в моделях этого типа представляется вектором из $m-1$ компоненты (наличие продукта с распределением по оставшимся срокам годности в периодах). Оптимальная стратегия сложным образом зависит от всех его компонент. Приближенный метод решения этой задачи рассмотрен в работе [182].

5.4. Объем заказа с учетом инфляции

В этом разделе мы дадим обзор результатов [118] с исправленными итоговыми формулами.

Предположение о постоянном темпе инфляции приводит к представлению реальной стоимости в виде

$$b(t) = b_0 e^{kt}.$$

Будем считать, что по такому же закону меняются цена заказа $g(t)$ и цена закупки $u(t)$. Интенсивность спроса λ примем постоянной. Целью считается максимизация дохода к концу горизонта планирования H .

Обозначим $v(t)$ цену продажи товара в момент t , а выручку — через

$$R = \int_0^H v(t) \lambda dt.$$

Чистый доход равен выручке минус затраты. Товар покупается в кредит с постоянной процентной ставкой r , кредит погашается после продажи товара. Если инфляция одинакова для продажной цены и себестоимости, то выручка

$$R = \int_0^t v(t) \lambda dt = \frac{\lambda v_0}{k} (e^{kH} - 1)$$

и не зависит от частоты заказов. Положим $i = H/T$ целым. Затраты на поставки

$$G = g(0) + g(T) + g(2T) + \dots + g[(i-1)T] = g_0 (e^{kH} - 1) / (e^{kT} - 1).$$

Стоимость закупок

$$C = \lambda T \{u(0) + u(T) + u(2T) + \dots + u[(i-1)T]\} = \lambda T u_0 (e^{kH} - 1) / (e^{kT} - 1).$$

Стоимость хранения

$$\sum_{j=0}^{i-1} \int_0^T \lambda h(jT, jT+t) dt = \frac{\lambda r T^2}{2} u_0 (e^{kH} - 1) / (e^{kT} - 1).$$

Общие затраты

$$L(H, T) = \left(g_0 + u_0 \lambda T + \frac{u_0 \lambda r T^2}{2} \right) \frac{e^{kH} - 1}{e^{kT} - 1}. \quad (5.4.1)$$

Требуется найти значение T , при котором они минимальны (выручка фиксирована).

Временно откажемся от требования кратности H/T и ограничимся квадратичным разложением $e^{kT} - 1 \approx kT + k^2 T^2 / 2$. Тогда

$$L(H, T) = \left(g_0 + u_0 \lambda T + \frac{u_0 \lambda r T^2}{2} \right) \frac{e^{kH} - 1}{kT + k^2 T^2 / 2}.$$

Оптимальное решение

$$T^* = \frac{1}{u_0 \lambda (r - k)} \left[g_0 k + \sqrt{g_0^2 k^2 + 2g_0 u_0 \lambda (r - k)} \right].$$

В отсутствие инфляции ($k = 0$) при процентной ставке r получаем классическую формулу

$$T^* = \sqrt{\frac{2g_0}{\lambda u_0 r}}.$$

Обычны две возможности назначения цены с процентной надбавкой к стоимости закупки и с абсолютной надбавкой

Процентная надбавка. Если $u(t)$ — цена закупки детали, то продажная цена на интервале $[t, t + T]$

$$u(t, t + T) = (1 + m)u(t),$$

где m — наценка Общая выручка

$$R = \sum_{j=0}^{i-1} \lambda T u(jT) (1 + m).$$

Если закупочные цены подвержены инфляции, так что $u(jT) = u_0 e^{kjT}$, то общая выручка

$$R = u_0 \lambda T (1 + m) \left[\frac{e^{kH} - 1}{e^{kT} - 1} \right].$$

Чистый доход

$$R(L, T) = \left(u_0 \lambda T (1 + m) - g_0 - u_0 \lambda T - \frac{r u_0 \lambda T^2}{2} \right) \left(\frac{e^{kH} - 1}{e^{kT} - 1} \right).$$

Оптимальная периодичность

$$T^* = \frac{1}{u_0 \lambda (r + mk)} \left[-g_0 k + \sqrt{g_0^2 k^2 + 2g_0 u_0 \lambda (r + mk)} \right].$$

Абсолютная надбавка В этом случае

$$v(t, t + T) = u(t) + p.$$

Общая выручка

$$R = u_0 \lambda T \left(\frac{e^{kH} - 1}{e^{kT} - 1} \right) + \lambda H p.$$

Чистый доход

$$R(H, T) = \lambda H p - \left(g_0 + \frac{r u_0 \lambda T^2}{2} \right) \left(\frac{e^{kH} - 1}{e^{kT} - 1} \right).$$

Оптимальный интервал

$$T^* = \frac{1}{u_0 \lambda r} \left[g_0 k + \sqrt{g_0^2 k^2 + 2g_0 u_0 \lambda r} \right].$$

5.5. Детерминированный нестационарный спрос

5.5.1. Постановка задачи

Модели разд. 5.1 идеализируют действительность, предполагая интенсивность спроса λ постоянной. Часто спрос задается как последовательность рассчитанных, например, в соответствии с известной на весь период nT производственной программой величин суммарного потребления $\{x_k\}$, $k = \overline{1, n}$, в смежные интервалы времени длины T . Пронумеруем интервалы и введем для них обозначения:

z_k — остаток от $(k - 1)$ -го периода,

x_k — спрос в k -й период,

S_k — запас, создаваемый на k -й период, $S_k \geq x_k$,

$h_k(S_k - x_k)$ — расходы на хранение избыточного запаса в k -й период,

$c_k(S_k - z_k)$ — расходы на доведение запаса до величины S_k .

Суммарные затраты на снабжение подсчитываются по формуле

$$L_{nT} = \sum_{k=1}^n [c_k(S_k - z_k) + h_k(S_k - x_k)]. \quad (5.5.1)$$

5.5.2. Динамическое программирование как метод решения

Для минимизации L воспользуемся методом динамического программирования. Будем последовательно минимизировать затраты за 1, 2 и т.д. интервалов на основе принципа оптимальности Р. Беллмана:

для аддитивной целевой функции решения на все оставшиеся интервалы должны составлять оптимальное поведение относительно состояния, полученного в результате предыдущего решения, независимо от ранее принятых решений и начального состояния.

Минимальные затраты за последний период

$$L_T(z_n) = \min_{S_n \geq z_n} [c_n(S_n - z_n) + h_n(S_n - x_n)].$$

Затраты за два последних периода

$$L_{2T}(z_{n-1}) = \min_{S_{n-1} \geq z_{n-1}} [c_{n-1}(S_{n-1} - z_{n-1}) + h_{n-1}(S_{n-1} - x_{n-1}) + L_T(S_{n-1} - x_{n-1})].$$

Вообще для $k = 2, 3, \dots, n$

$$L_{kT}(z_{n+1-k}) = \min_{S_{n+1-k} \geq z_{n+1-k}} [c_{n+1-k}(S_{n+1-k} - z_{n+1-k}) + h_{n+1-k}(S_{n+1-k} - x_{n+1-k}) + L_{(k-1)T}(S_{n+1-k} - x_{n+1-k})].$$

В процессе минимизации затрат для поиска $\{S_k\}$ необходимо использовать свойства функций $\{c_k\}$ и $\{h_k\}$. В типичном для практики случае, когда $\{c_k\}$ и $\{h_k\}$ — возрастающие функции своих аргументов, равные нулю при нулевом аргументе, оптимальный запас для последнего периода

$$S_n = \begin{cases} x_n & \text{при } z_n \leq x_n, \\ z_n & \text{при } z_n > x_n. \end{cases}$$

Отсюда легко найти $L_T(z_n)$

Можно показать, что для любого периода оптимальная стратегия имеет вид

$$S_k^* = \begin{cases} S_k & \text{при } z_k \leq S_k, \\ z_k & \text{при } z_k > S_k, \end{cases}$$

причем функция $L_{kT}(x_{n+1-k})$ достигает минимума при $z_{n+1-k} = S_k$.

Полученные соотношения позволяют построить простой алгоритм численного решения задачи. Установим диапазоны изменения для S_k и z_k . Очевидно, что $x_k \leq S_k \leq \sum_{i=k}^n x_i$ (запас на каждый период должен быть не меньше спроса в этом периоде, но не больше спроса за все оставшееся время). Из этого требования вытекает и условие для остатков: $0 \leq z_k \leq \sum_{i=k}^n x_i$. Разобьем эти диапазоны на интервалы с некоторым шагом Δ и для всех допустимых $\{z_k\}$ рассчитаем оптимальные S_k и $L_T(z_k)$. Далее, численно минимизируя суммы

$$c_{n-1}(S_{n-1} - z_{n-1}) + h_{n-1}(S_{n-1} - x_{n-1}) + L_T(S_{n-1} - x_{n-1})$$

по S_{n-1} с использованием полученной на предыдущем шаге таблицы $L_T(z_n)$, получим новую таблицу функции $L_{2T}(z_{n-1})$ и связанную с ней $S_{n-1}(z_{n-1})$, и т.д. Результатом очередного шага явятся, наконец, таблицы $L_{(n-1)T}(z_2)$ и $S_2(z_2)$. Минимум суммы

$$c_1(S_1 - z_1) + h_1(S_1 - x_1) + L_{(n-1)T}(S_1 - x_1)$$

даст нам итоговое значение затрат $L_{nT}(z_1)$ и укажет оптимальный запас в первом периоде S_1 . Теперь, просматривая таблицы $\{S_k(z)\}$ в обратном порядке, легко получить оптимальную последовательность нормативных запасов: аргумент z_k , по которому выбирается S_k для $k \geq 2$, равен $S_{k-1} - x_{k-1}$.

Заметим, что на каждом этапе минимизации необходимо иметь только одну таблицу $L_{kT}(z_{n+1-k})$, полученную на предыдущем шаге, ранее найденные $\{L_{(k-1)T}\}$ более не нужны, а таблицы $\{S_k^*\}$ используются только на заключительном этапе вычислений. Это позволяет легко разместить все требуемые результаты в оперативной памяти ЭВМ.

С общематематических позиций динамическое программирование решает задачу рекурсивно — как набор подзадач для ограничений вида $C \leq \bar{C}$ (например, по стоимости запасов) и $k \leq n$, причем последняя задача ($k = 1$) имеет тривиальное решение. Важные особенности метода:

- Функции затрат $\{L_i\}$ не обязаны быть дифференцируемыми и могут задаваться таблично.
- Гарантируется получение глобального минимума, причем наличие локальных минимумов не создает никаких трудностей.
- Дополнительные ограничения только облегчают получение решения, поскольку сужают пространство поиска.

- После того как получено решение для \bar{C} , его легко получить для любого $C \leq \bar{C}$, выполняя только обратный ход алгоритма.
- Взяв \bar{C} с некоторым запасом, можно определить чувствительность решения и целесообразность ослабления ограничения. Выделив один из этапов и раскручивая оставшиеся части во встречных направлениях, можно оценить влияние на решение параметров выделенного этапа.
- Используя часть полученных таблиц, можно решить задачу с меньшим числом периодов.
- Метод может быть обобщен на многоресурсные задачи (в частности, с предварительным распределением затрат по уровням).

Динамическое программирование применяется и в других задачах многоэтапного распределения ресурсов (например, между отдельными номенклатурами, складами и т.п.). Абстрактный термин «ресурс» может относиться к деньгам, весу, объему. Заданный ресурс оптимальным образом распределяется между указанным числом «активностей» (например, общий запас — по иерархическим уровням системы). Предполагается, что эффекты всех назначений могут быть измерены некоторой общей мерой, а общий эффект измеряется их суммой (произведением). Если фиксирована общая сумма, то задача одномерная, если фиксированы суммы по уровням системы — многомерная. Если мерой эффективности служит число случаев дефицита на нижнем уровне или ожидаемые задержки поставок, то это аддитивные монотонно убывающие функции уровней запаса. Эффективность метода зависит от порядка, в котором рассматриваются «активности».

Схема динамического программирования может быть легко переформулирована для задач максимизации.

Динамическое программирование непосредственно ориентировано на решение дискретных задач; однако его можно использовать и для задач, в которых все переменные непрерывны. В этих случаях непрерывная область пространства решений дискретизируется и отыскивается оптимальное управление. Затем в его окрестности используется более мелкая сетка, и т.п. Непрерывные функции заменяются аппроксимациями по ряду дискретных точек. Доказанная вогнутость или выпуклость функций дохода (затрат) существенно ограничивают перебор.

5.6. Случайный спрос, периодическая стратегия

В § 4.3 [56] приведены результаты классических исследований по типам оптимальных периодических стратегий при вероятностном спросе и показано, что при весьма общих допущениях (в частности, для линейных функций затрат) оптимальна стратегия с критическим уровнем. Рассмотрим типичные задачи этого класса.

5.6.1. Непрерывный спрос

Однопериодная задача со случайным спросом и мгновенной поставкой часто называется «задачей газетчика» (спрашивается, сколько газет должен взять распространитель, чтобы минимизировать убытки от возврата непроданных экземпляров и упущенной выгоды при неудовлетворенном спросе). Столь же популярным ее приложением является задача о новогодних елках. В [89, с. 252] она иллюстрируется также на продаже радиоприемников (стоимость каждого 40 долл.; неудовлетворенная заявка обходится магазину в 20 долл. из-за потери предпочтения плюс потеря 25 долл. валового дохода; стоимость подачи заказа оценивается в 3 долл.). Наконец, приведем из того же источника «хлебную» задачу: в универсаме необходимо определить, сколько хлеба нужно заказать на день. Спрос за день можно считать случайной величиной, подчиненной нормальному закону со средним 300 и среднеквадратическим отклонением 50. Один батон продается за 25 центов, себестоимость для магазина 19 центов, весь непроданный хлеб сбывается на следующий день по цене 15 центов за штуку (оптимально $S = 313$, доход $L^* = 16.07$; при заказе по средней потребности $S = 300$, $L^* = 16.01$). Заметим, что довольно заметное отклонение решения от оптимума очень слабо повлияло на ожидаемые затраты.

«Задача газетчика» хорошо описывает специфику торговли остро модными и сезонными товарами. Более серьезные ее применения могут быть связаны с поставками запасных частей и дополнительного оборудования к уникальным комплексам на плановый срок эксплуатации (турбогенераторам, физическим установкам, химическим реакторам, прокатным станам, крупным кораблям, стационарным суперЭВМ). Дело в том, что вместе с основным оборудованием запчасти могут быть поставлены по относительно низкой цене, а позже — лишь по гораздо более высокой (из-за необходимости повторного запуска производства). К то-

му же, очень дорого может обойтись простой оборудования в ожидании поставки. Наконец, к этому классу относятся проблемы снабжения объектов с сезонным завозом (арктические, антарктические и высокогорные гляциологические и метеорологические станции, маяки) и разовая закупка топлива на отопительный сезон при недостоверном прогнозе погоды.

Оптимальные решения. При подсчете затрат по состоянию запаса к концу периода целевая функция

$$L(y) = h \int_0^S (S-x)f(x) dx + d \int_S^\infty (x-S)f(x) dx + c(S-z). \quad (5.6.1)$$

Из условия

$$\begin{aligned} \frac{dL}{dS} &= h \int_0^S f(x) dx - d \int_S^\infty f(x) dx + c \\ &= hF(S) - d[1 - F(S)] + c = 0 \end{aligned}$$

для нахождения оптимального уровня запаса получаем уравнение

$$F(S) = \frac{d-c}{d+h}, \quad (5.6.2)$$

где $F(\cdot)$ — интегральная функция распределения спроса за время T .

Приведем несколько конкретизаций этого результата для частных видов распределения спроса. Для *показательного* закона формула (5.6.2) сводится к

$$1 - \exp(-S^*/\bar{x}) = \frac{d-c}{d+h},$$

так что

$$S^* = \bar{x} \ln \frac{h+d}{h+c}. \quad (5.6.3)$$

Здесь $\bar{x} = \lambda T$ — средний спрос за период.

Для *распределения Релея* имеем

$$1 - \exp[-S^2/(2M^2)] = \frac{d-c}{d+h}$$

и

$$S^* = M \sqrt{2 \ln \frac{h+d}{h+c}}. \quad (5.6.4)$$

Наконец, для *распределения Вейбулла* из уравнения

$$1 - \exp(-S^k/W) = \frac{d-c}{d+h}$$

получаем

$$S^* = \left(W \ln \frac{h+d}{h+c} \right)^{1/k} \quad (5.6.5)$$

Минимальные затраты. В работах по управлению запасами обычно основное внимание уделяется расчету оптимальных стратегий управления запасами. Между тем, не менее важна оценка издержек, связанных с их реализацией: только на ее основе можно определить объем средств, ассигнуемых на операционные затраты в системах снабжения. Сравнение же минимальных издержек со средними фактическими позволяет оценить качество применяемых методов управления запасами и сделать вывод о необходимости их корректировки. Для многих задач исследования операций (в частности, управления запасами) характерно очень медленное изменение показателя качества в районе оптимума. Это позволяет корректировать оптимальные решения с учетом удобства их реализации: например, периодичность поставок приблизить к квартальной или их объем — к транзитной норме.

Оценка минимальных затрат может быть получена прямой постановкой в функцию затрат оптимального решения. Однако такой путь требует большого объема вычислений. Ниже рассматривается общий метод непосредственного определения минимальных затрат в периодической модели управления запасами со случайным спросом и его конкретизации для некоторых частных распределений спроса за период

Перепишем формулу (5.6.1) в виде

$$\begin{aligned} L(S) &= hS \int_0^S f(x) dx - h \int_0^S x f(x) dx \\ &\quad - d \cdot S \int_S^\infty f(x) dx + d \int_S^\infty x f(x) dx + c(S-z). \end{aligned}$$

Заметим, что $\int_0^S f(x) dx = F(S)$ и $\int_S^\infty f(x) dx = 1 - F(S)$. Кроме того,

$$\begin{aligned} \int_S^\infty xf(x) dx &= \int_0^\infty xf(x) dx - \int_0^S xf(x) dx \\ &= \bar{x} - \int_0^S xf(x) dx, \end{aligned}$$

где \bar{x} — математическое ожидание спроса за период. Теперь

$$L(S) = (d+h)SF(S) - (d+h) \int_0^S xf(x) dx + c(S-z) + d(\bar{x}-S). \quad (5.6.6)$$

При выборе $S = S^*$ согласно (5.6.2) имеем $(d+h)F(S^*) = d-c$. Подставляя этот результат в (5.6.6), находим простое выражение для минимальных затрат

$$L^* = L(S^*) = d\bar{x} - cz - (d+h) \int_0^{S^*} xf(x) dx. \quad (5.6.7)$$

В отличие от (5.6.1), здесь необходимо вычислять только один интеграл. Поскольку оптимальный запас однозначно определяется из уравнения (5.6.2), минимальные затраты зависят только от начального запаса z .

Для некоторых частных распределений возможны дальнейшие упрощения. Пусть, например, спрос за период имеет *показательное* распределение со средним \bar{x} . Тогда

$$\int_0^{S^*} xf(x) dx = \int_0^{S^*} (x/\bar{x})e^{-x/\bar{x}} dx = \bar{x}[1 - e^{-S^*/\bar{x}}(S^*/\bar{x} + 1)].$$

Для этого распределения

$$S^* = \bar{x} \ln \frac{d+h}{c+h}, \quad (5.6.8)$$

так что

$$e^{-S^*/\bar{x}} = \exp\left(-\ln \frac{d+h}{c+h}\right) = \frac{c+h}{d+h}.$$

Подставляя результат в (5.6.7), находим

$$L^* = \bar{x} \left[(c+h) \left(1 + \ln \frac{d+h}{c+h} \right) - h \right] - cz.$$

Очень простой результат (при известном S^*) можно вывести для *гамма-распределения* с плотностью

$$f(x) = \frac{\lambda(\lambda x)^{k-1}}{\Gamma(k)} e^{-\lambda x}.$$

В этом случае $\bar{x} = k/\lambda$, а

$$\begin{aligned} \int_0^S xf(x) dx &= \int_0^S x \frac{\lambda(\lambda x)^{k-1}}{\Gamma(k)} e^{-\lambda x} dx = \frac{1}{\lambda\Gamma(k)} \gamma(k+1, \lambda S) \\ &= \frac{1}{\lambda\Gamma(k)} \left[k\gamma(k, \lambda S) - (\lambda S)^k e^{-\lambda S} \right] \\ &= \frac{k}{\lambda} \left[\frac{\gamma(k, \lambda S)}{\Gamma(k)} - \frac{(\lambda S)^k e^{-\lambda S}}{\Gamma(k+1)} \right]. \end{aligned}$$

Здесь было использовано известное рекуррентное соотношение для неполной гамма-функции. Нормированная неполная гамма-функция в данном случае равна оптимальной вероятности обеспеченности спроса, и

$$L^* = c(\bar{x} - z) + \frac{d+h}{\lambda\Gamma(k)} \cdot (\lambda S^*)^k e^{-\lambda S^*}.$$

Исключив $\lambda = k/\bar{x}$, получаем

$$L^* = \bar{x} \left[c + \frac{d+h}{\Gamma(k+1)} \cdot (kS^*/\bar{x})^k e^{-kS^*/\bar{x}} \right] - cz.$$

Параметр k выражается через среднее и дисперсию D спроса за период: $k = (\bar{x})^2/D$.

Для спроса, подчиненного закону *Релея* с плотностью

$$f(x) = \frac{x}{M^2} e^{-x^2/2M^2},$$

где M — наиболее вероятное значение (мода) спроса, средний спрос $\bar{x} = M\sqrt{\pi/2}$. Кроме того, здесь

$$\int_0^S xf(x) dx = \int_0^S \frac{x^2}{M^2} e^{-x^2/2M^2} dx = \int_0^S e^{-x^2/2M^2} dx - S e^{-S^2/2M^2}.$$

Последний интеграл выражается через функцию Лапласа $\Phi(\cdot)$, а экспонента — через интегральную функцию распределения закона Релея, так что

$$\int_0^{S^*} xf(x) dx = M\sqrt{\pi/2}\Phi\left(\frac{S^*}{M\sqrt{2}}\right) - S^*[1 - F(S^*)]$$

Но $M\sqrt{\pi/2} = \bar{x}$, а выражение $1 - F(S)$ при оптимальном выборе $S = S^*$ сводится к $(c+h)/(d+h)$. Поэтому

$$L^* = \bar{x} \left[d - (d+h)\Phi\left(\frac{S^*}{M\sqrt{2}}\right) \right] - cz + (c+h)S^*.$$

В эту формулу можно подставить явное выражение для оптимального запаса при релейском распределении спроса

$$S^* = M\sqrt{2 \ln[(d+h)/(c+h)]}.$$

Окончательно

$$L^* = \bar{x} \left[d - (d+h)\Phi\left(\sqrt{\ln \frac{d+h}{c+h}}\right) + 2(c+h)\sqrt{\frac{1}{\pi} \frac{d+h}{c+h}} \right] - cz$$

5.6.2. Учет погрешностей в параметрах

Для статической вероятностной модели функция (5.6.1) затрат за период легко преобразуется к виду

$$L = c(S-z) + S(d+h) \int_0^S f(x) dx - d(S-\bar{x}) - (d+h) \int_0^S xf(x) dx. \quad (5.6.9)$$

В частности, при оптимальном выборе нормативного запаса по формуле (5.6.2)

$$\begin{aligned} L^* &= c(S^* - z) + S^*(d+h) \frac{d-c}{d+h} - d(S^* - \bar{x}) \\ &- (d+h) \int_0^{S^*} xf(x) dx \\ &= d\bar{x} - cz - (d+h) \int_0^{S^*} xf(x) dx. \end{aligned} \quad (5.6.10)$$

Подстановка в (5.6.1) $S = S^* + \Delta$ дает

$$\begin{aligned} L &= c\Delta + (S^* + \Delta)(d+h) \int_0^{S^* + \Delta} f(x) dx - d\Delta \\ &- (d+h) \int_0^{S^* + \Delta} xf(x) dx + c(S^* - z) - d(S^* - \bar{x}). \end{aligned} \quad (5.6.11)$$

Воспользовавшись теоремой о среднем, приближенно представим входящие в последнюю формулу интегралы в виде

$$\int_0^{S^* + \Delta} f(x) dx \approx \int_0^{S^*} f(x) dx + \Delta \cdot f(S^* + \Delta/2)$$

и

$$\int_0^{S^* + \Delta} xf(x) dx \approx \int_0^{S^*} xf(x) dx + \Delta(S^* + \Delta/2) \cdot f(S^* + \Delta/2).$$

Осуществив в (5.6.11) эти замены и вычтя из $L(S^* + \Delta)$ величину $L(S^*)$, получаем приращение затрат за период

$$\Delta L \approx \frac{d+h}{2} \Delta^2 \cdot f(S^* + \Delta/2). \quad (5.6.12)$$

Поскольку в типичной области высоких вероятностей обеспечения спроса $S^* \gg \Delta$, зависимость $\Delta L(\Delta)$ с достаточной для практики точностью можно считать квадратичной.

Установим зависимость ΔS от погрешности в ценах c, h и d , для чего применим логарифмическое дифференцирование к форму-

ле (5.6.2):

$$\ln \int_0^{S^*} f(x) dx = \ln(d-c) - \ln(d+h),$$

$$\frac{f(S^*)}{\int_0^{S^*} f(x) dx} \Delta S = \frac{\Delta d}{d-c} - \frac{\Delta c}{d-c} - \frac{\Delta d}{d+h} - \frac{\Delta h}{d+h}.$$

Подставляя оптимальную вероятность обеспечения спроса и группируя слагаемые, получаем

$$\Delta S = \frac{1}{f(S^*)(d+h)^2} [(c+h)\Delta_d - (d+h)\Delta_c - (d-c)\Delta_h].$$

Действия в случае неизвестных знаков погрешностей аналогичны предыдущему случаю

5.6.3. Дискретный спрос

При дискретном спросе оптимальный запас S^* выбирается из условий

$$\begin{aligned} L(S^* + 1) - L(S^*) &\geq 0, \\ L(S^* - 1) - L(S^*) &\geq 0. \end{aligned} \quad (5.6.13)$$

Построим целевую функцию по ожидаемым избытку и недостатку

$$L(S) = h \sum_{x=0}^{S-1} (S-x)p(x) + d \sum_{x=S+1}^{\infty} (x-S)p(x) + c(S-z).$$

Соответственно

$$L(S+1) = h \sum_{x=0}^S (S+1-x)p(x) + d \sum_{x=S+2}^{\infty} (x-S-1)p(x) + c(S+1-z).$$

Приращение затрат при увеличении запаса относительно оптимума должно быть неотрицательным:

$$\Delta(S^*) = L(S^* + 1) - L(S^*) = (h+d) \sum_{x=0}^{S^*} p(x) - d + c \geq 0,$$

или

$$\sum_{x=0}^{S^*} p(x) \geq \frac{d-c}{d+h}.$$

Очевидно, что знак аналогичного приращения при увеличении запаса от $S^* - 1$ до S^* должен быть обратным

$$\sum_{x=0}^{S^*-1} p(x) \leq \frac{d-c}{d+h}.$$

Итак, условие оптимальности запаса S^* есть

$$\sum_{x=0}^{S^*-1} p(x) \leq \frac{d-c}{d+h} \leq \sum_{x=0}^{S^*} p(x). \quad (5.6.14)$$

Это условие полностью аналогично полученному ранее для непрерывного случая. Поэтому мы в дальнейшем будем рассматривать только один из вариантов вывода. Попутно отметим, что принцип проверки смены знака функции при единичном приращении аргумента используется во многих других экономических задачах и называется *маргинальным анализом*.

5.6.4. «Импульсные» расходы

При исчислении расходов по *средневзвешенным* запасу и дефициту в случае $x \leq S$ средний положительный запас $\bar{y}_+ = [S + (S-x)]/2 = S - x/2$, а время наличия запаса $T_+ = T$, соответственно средний отрицательный запас $\bar{y}_- = 0$ и время существования дефицита $T_- = 0$. Если спрос превысил запас, то $\bar{y}_+ = S/2$, $\bar{y}_- = (x-S)/2$, а части периода с положительным и отрицательным запасами соответственно будут равны $T_+ = TS/x$; $T_- = (1-S/x)T$. Ожидаемое значение затрат должно подсчитываться согласно

$$\begin{aligned} L(S) &= h \int_0^S (S-x/2)f(x) dx + h \int_S^{\infty} \frac{S^2}{2x} f(x) dx \\ &+ d \int_S^{\infty} \frac{(x-S)^2}{2x} f(x) dx + c(S-z). \end{aligned}$$

Дифференцируя по S , имеем

$$\begin{aligned} \frac{dL}{dS} &= h \int_0^S f(x) dx + h \int_S^\infty \frac{S}{x} f(x) dx - d \int_S^\infty \frac{x-S}{x} f(x) dx + c \\ &= (d+h) \left[\int_0^S f(x) dx + S \int_S^\infty \frac{f(x)}{x} dx \right] - d + c \\ &= (d+h) \left[F(S) + S \int_S^\infty \frac{f(x)}{x} dx \right] - d + c. \end{aligned}$$

Условием минимума $L(S)$ является равенство

$$F(S) + S \int_S^\infty \frac{f(x)}{x} dx = \frac{d-c}{d+h}. \quad (5.6.15)$$

Поскольку

$$S \int_S^\infty \frac{f(x)}{x} dx = \int_S^\infty \frac{Sf(x)}{x} dx < S \int_S^\infty f(x) dx,$$

при $S \rightarrow \infty$ предел левой части равенства (5.6.15) — нуль. Но $F(\infty) = 1$; следовательно, решение уравнения (5.6.15) существует, если

$$\lim_{S \rightarrow 0} S \int_S^\infty \frac{f(x)}{x} dx \leq \frac{d-c}{d+h}.$$

Остается показать его единственность. Обозначим левую часть (5.6.15) через $Q(S)$ и найдем ее производную.

$$\frac{dQ}{dS} = f(S) + \int_S^\infty \frac{f(x)}{x} dx - S \frac{f(S)}{S} = \int_S^\infty \frac{f(x)}{x} dx > 0$$

при всех S . Следовательно, функция $Q(S)$ монотонна, что и доказывает требуемый результат.

5.6.5. Штраф по вероятности дефицита

Определим функцию затрат

$$L(S) = h \int_0^S (S-x)f(x)dx + d \int_S^\infty f(x) dx + c(S-z), \quad (5.6.16)$$

те со штрафом d , выплачиваемым в случае хотя бы одной недостачи независимо от величины дефицита. Наивыгоднейший уровень запаса дается решением уравнения

$$L'(S) = h \int_0^S f(x) dx - df(S) + c = 0,$$

или

$$\int_0^S f(x) dx = \frac{df(S) - c}{h}. \quad (5.6.17)$$

Очевидно, $L'(\infty) = c + h > 0$, и для существования решения его достаточно выполнения в некотором диапазоне значений S условия

$$\int_0^S f(x) dx \leq \frac{df(S) - c}{h}$$

Уравнение (5.6.17) для унимодальных распределений либо имеет не менее двух корней, либо не имеет их совсем. В случае двух корней минимум расходов соответствует большему корню, поскольку именно в этой точке знак производной меняется с минуса на плюс.

5.6.6. Штраф по времени дефицита

При спросе x , превысившем нормативный запас, доля времени существования дефицита равна $1 - S/x$. Функция затрат может быть записана как

$$L(S) = h \int_0^S (S-x)f(x)dx + d \int_S^\infty (1-S/x)f(x) dx + c(S-z). \quad (5.6.18)$$

Приравняем нулю ее производную по запасу

$$c + h \int_0^S f(x) dx - d \int_S^\infty \frac{f(x)}{x} dx = 0.$$

Левая часть этого уравнения монотонно возрастает по S . При $S \rightarrow \infty$, очевидно, имеем $c + h > 0$. Для существования экстремумов необходимо выполнение $c < d \int_0^\infty [f(x)/x] dx$

Рассмотрим более подробно случай пуассоновского распределения спроса. Функция затрат будет иметь вид, аналогичный (5.6.18), с заменой интегрирования по x суммированием. Найдем плотность $\psi_S(\tau)$ распределения времени дефицита. Распределение времени наступления k -го события пуассоновского потока подчинено закону Эрланга k -го порядка. Дефицит начинается при израсходовании всего запаса S и еще одной единицы, так что

$$\psi_S(\tau) = \frac{\lambda[(\lambda(T-\tau))^S]}{S!} e^{-\lambda(T-\tau)}. \quad (5.6.19)$$

Вычислим среднее время существования дефицита

$$\begin{aligned} \bar{\tau}_S &= \int_0^T \frac{\lambda\tau[\lambda(T-\tau)]^S}{S!} e^{-\lambda(T-\tau)} d\tau \\ &= \int_0^T \frac{[\lambda(T-\tau)]^S}{S!} e^{-\lambda(T-\tau)} (\lambda\tau - \lambda T + \lambda T) d\tau \\ &= - \int_0^T \frac{[\lambda(T-\tau)]^{S+1}}{S!} e^{-\lambda(T-\tau)} d\tau + \lambda T \int_0^T \frac{[\lambda(T-\tau)]^S}{S!} e^{-\lambda(T-\tau)} d\tau. \end{aligned}$$

Произведя замену переменных $\lambda(T-\tau) = u$, получаем

$$\bar{\tau}_S = \frac{1}{S!} \left\{ - \frac{1}{\lambda} \int_0^{\lambda T} u^{S+1} e^{-u} du + T \int_0^{\lambda T} u^S e^{-u} du \right\}.$$

Заключенные в фигурные скобки интегралы являются табличными — см. [11], формула (567.9). В частности, первый из них

$$\int_0^{\lambda T} u^{S+1} e^{-u} du = -e^{-\lambda T} \sum_{k=0}^{S+1} \frac{(S+1)!(\lambda T)^k}{k!} e^{-\lambda T} + (S+1)!.$$

Записав аналогичное выражение для второго и подставив их в формулу для $\bar{\tau}_S$, имеем

$$\bar{\tau}_S = \frac{1}{\lambda} e^{-\lambda T} (S+1) \sum_{k=0}^{S+1} \frac{(\lambda T)^k}{k!} - \frac{1}{\lambda} (S+1)! - T e^{-\lambda T} \sum_{k=0}^S \frac{(\lambda T)^k}{k!} + T.$$

После несложных преобразований приходим к

$$\bar{\tau}_S = e^{-\lambda T} \left[\left(\frac{S+1}{\lambda} - T \right) \sum_{k=0}^S \frac{(\lambda T)^k}{k!} + \frac{S+1}{\lambda} \frac{(\lambda T)^{S+1}}{(S+1)!} \right] - \left(\frac{S+1}{\lambda} - T \right).$$

Оптимальный S вновь должен выбираться из условий (5.6.13). Конкретизируем их для данного случая. При замене S на $S+1$ дополнительные ожидаемые расходы на поставки, хранение и выплату постоянной составляющей штрафов составят

$$\begin{aligned} \Delta_1 &= c + h \sum_{k=0}^S \frac{(\lambda T)^k}{k!} e^{-\lambda T} - d \frac{(\lambda T)^{S+1}}{(S+1)!} e^{-\lambda T} \\ &= c + e^{-\lambda T} \left[h \sum_{k=0}^S \frac{(\lambda T)^k}{k!} - d \frac{(\lambda T)^{S+1}}{(S+1)!} \right]. \end{aligned}$$

Приращение повременных штрафов

$$\begin{aligned} \Delta_2 &= d(\bar{\tau}_{S+1} - \bar{\tau}_S) = d \left\{ - \frac{S+2}{\lambda} + \frac{S+1}{\lambda} \right. \\ &+ e^{-\lambda T} \left[\left(\frac{S+2}{\lambda} - T \right) \sum_{k=0}^{S+1} \frac{(\lambda T)^k}{k!} + \frac{S+2}{\lambda} \frac{(\lambda T)^{S+2}}{(S+2)!} \right. \\ &\left. \left. - \left(\frac{S+1}{\lambda} - T \right) \sum_{k=0}^S \frac{(\lambda T)^k}{k!} - \frac{S+1}{\lambda} \frac{(\lambda T)^{S+1}}{(S+1)!} \right] \right\} \\ &= - \frac{d}{\lambda} \left[1 - e^{-\lambda T} \sum_{k=0}^{S+1} \frac{(\lambda T)^k}{k!} \right]. \end{aligned}$$

Таким образом, первое из условий (5.6.13) преобразуется в

$$\begin{aligned} \Delta L &= -\frac{d}{\lambda} + c + e^{-\lambda T} \left[h \sum_{k=0}^S \frac{(\lambda T)^k}{k!} - d \frac{(\lambda T)^{S+1}}{(S+1)!} + \frac{d}{\lambda} \sum_{k=0}^{S+1} \frac{(\lambda T)^k}{k!} \right] \\ &= c - \frac{d}{\lambda} + e^{-\lambda T} \left[\left(\frac{d}{\lambda} + h \right) \sum_{k=0}^S \frac{(\lambda T)^k}{k!} + \left(\frac{d}{\lambda} - d \right) \frac{(\lambda T)^{S+1}}{(S+1)!} \right] \geq 0. \end{aligned}$$

Второе условие (5.6.13) может быть получено из него заменой S на $S-1$ и обращением знака неравенства

5.6.7. Многопериодные обобщения

Приведенные выше формулы относятся к статической модели с минимизацией затрат за период. Однако доказано, что они верны и для бесконечношаговой модели с дисконтированным накоплением затрат и переносом возникающего дефицита в очередной период — при умножении цены закупки c на множитель $1-\alpha$. В динамических задачах с конечным числом шагов оптимальные уровни запасов будут находиться между решениями для статического и динамического случаев. Точнее,

$$S_1 \leq S_2 \leq \dots \leq S_\infty.$$

Те же соображения применимы к периодическим стратегиям с задержкой поставок, если эта задержка постоянна и равна kT . В этом случае наличный запас рассматривается как сумма физического наличия и всех ранее сделанных, но еще не выполненных заказов, а распределение спроса заменяется на $(k+1)$ -кратную свертку распределения однопериодного спроса (см. [56] и имеющиеся там дополнительные ссылки).

5.7. Неизвестная интенсивность спроса

Довольно часто вид распределения отказов устанавливается из теоретических соображений, а область вероятных значений параметра может быть указана в результате анализа опыта работы аналогичных устройств. В этом случае управление запасами должно проводиться на основе статистических данных, которые по мере своего накопления позволяют уточнять параметры распределения спроса и пересчитывать критические числа стратегий управления. Нормативный запас

определяется как то его значение, при котором максимум риска минимален (минимаксная стратегия), или как величина его, минимизирующая математическое ожидание функции риска по всем возможным значениям параметра с учетом их апостериорных вероятностей (байесовская стратегия). Покажем эту методику на примере задачи управления запасами со случайным спросом, подчиненным распределению Пуассона с параметром (средним значением спроса за период) $x \in [a, b]$.

Подставим в условие оптимальности (11.2.3) левую границу нашего интервала a и найдем значение запаса y_1 , оптимальное при $x_0 = a$. Строгое выполнение этих неравенств означает, что запас y_1 оптимален в диапазоне $a \leq x \leq x_1$, где x_1 есть корень уравнения

$$\sum_{k=0}^{x_1} \frac{x^k}{k!} e^{-x} = \frac{d - c(1-\alpha)}{d+h}. \quad (5.7.1)$$

Для сохранения (11.2.3) при $x > x_1$ придется увеличить число слагаемых в сумме (для начала на единицу). Зафиксировав $y_2 = y_1 + 1$, найдем соответствующее ему граничное значение спроса x_2 и т.д. — до тех пор, пока не окажется, что очередное значение $x_n \geq b$. В этом случае примем $x_n = b$. В результате область возможных значений параметра оказывается разбитой на n непересекающихся отрезков, в каждом из которых оптимален вполне определенный y . Таким образом, в отыскании точечной оценки для x нет необходимости: достаточно указать лишь интервал $X_i = [x_{i-1}, x_i]$, к которому этот параметр принадлежит. Примем, что в каждом из интервалов значения x распределены равномерно. При этом математическое ожидание затрат за период, связанное с выбором решения y при $x \in X_i$, равно

$$\begin{aligned} L(y) &= \frac{1}{x_i - x_{i-1}} \int_{x_{i-1}}^{x_i} \left[c(1-\alpha)y + h \sum_{k=0}^y (y-k) \frac{x^k}{k!} e^{-x} \right. \\ &\quad \left. + d \sum_{k=y+1}^{\infty} (k-y) \frac{x^k}{k!} e^{-x} \right] dx \\ &= \frac{1}{x_i - x_{i-1}} \left\{ [c(1-\alpha) - d]xy + d \frac{x^2}{2} \right. \\ &\quad \left. - \frac{d+h}{2} e^{-x} \sum_{k=0}^y (y-k)(y-k+1) \frac{x^k}{k!} \right\} \Big|_{x_{i-1}}^{x_i}. \end{aligned}$$

Слагаемое $d \cdot x^2/2$ в фигурных скобках не зависит от выбора y , а потому при расчете стратегий и риска может быть опущено.

Матрица риска

	1	2	3	4	5	6	7	8	9	10	11
1	0.0	1.2	11.6	25.7	41.2	57.10	73.1	89.1	105.1	121.1	137.1
2	7.5	0.0	5.2	17.0	31.5	47.03	62.9	78.9	94.9	110.9	126.9
3	30.1	6.5	0.0	4.8	15.9	29.96	45.3	61.1	77.0	93.0	109.0
4	66.0	26.7	5.8	0.0	4.5	14.98	28.6	43.7	59.3	75.2	91.2
5	111.9	59.5	24.1	5.3	0.0	4.18	14.2	27.4	42.3	57.8	73.6
6	164.9	102.6	54.5	22.1	4.9	0.00	3.9	13.6	26.4	41.0	56.4
7	222.1	153.4	95.1	50.4	20.6	4.58	0.0	3.8	13.0	25.5	39.8
8	282.5	209.5	143.8	88.9	47.0	19.24	4.3	0.0	3.6	12.5	24.7
9	344.9	269.3	198.5	135.6	83.6	44.26	18.1	4.1	0.0	3.4	12.0
10	408.5	331.4	257.3	188.7	128.4	78.99	41.8	17.2	3.9	0.0	3.3
11	457.5	379.7	303.9	232.1	166.9	110.87	66.0	33.1	12.2	1.8	0.0

Поочередно подставляя в выражение для $L(y)$ полученные ранее значения $\{y_j\}$, $j = \overline{1, n}$, можно получить ожидаемые затраты $L_i(y_j)$, связанные с выбором каждого из них при $x \in X_i$. Риск же, связанный с принятием решения y_j вместо y_i , равен $L_{i,j}^o = L_i(y_j) - L_i(y_i)$. По смыслу величин $\{L_{i,j}^o\}$ образованная ими матрица риска L^o состоит из неотрицательных элементов, а на главной диагонали имеет нули. Вектор ожидаемого риска

$$R(m) = L^o \times P(m),$$

где $P(m)$ — вектор апостериорных вероятностей $X \in X_i$ на m -м шаге процесса; его минимальная составляющая указывает номер y_i , оптимального на данном шаге. Априорные вероятности пересчитываются по формуле Байеса, в данном случае принимающей вид

$$P_i(m) = \frac{(x_i - x_{i-1}) \prod_{k=1}^{m-1} \Pr\{x_k | x \in X_i\}}{\sum_{l=0}^n (x_l - x_{l-1}) \prod_{k=1}^{m-1} \Pr\{x_k | x \in X_l\}},$$

в которой

$$\Pr\{x_k | x \in X_i\} = \frac{1}{x_i - x_{i-1}} \int_{x_{i-1}}^{x_i} \frac{x^k}{k!} e^{-x} dx = \frac{1}{x_i - x_{i-1}} \left(e^{-x} \sum_{k=0}^x \frac{x^k}{k!} \right) \Big|_{x_{i-1}}^{x_i}.$$

Описанный алгоритм был реализован при следующих исходных данных:

$$a = 2, \quad b = 10, \quad d = 80, \quad c = 20, \quad h = 15, \quad \alpha = 0.95.$$

Весь диапазон x оказался разбит на 11 интервалов с координатами точек деления 2.00, 2.14, 2.90, 3.70, 4.49, 5.31, 6.14, 6.98, 7.83, 8.68, 9.55, 10.00, что дало для априорных вероятностей принадлежности параметра к интервалам значения 0.0175, 0.0947, 0.0984, 0.1007, 0.1024, 0.1040, 0.1051, 0.1061, 0.1070, 0.1077, 0.0564 соответственно. Матрица риска приведена в табл. 5.1, а минимальная составляющая вектора ожидаемого риска $r_1 = 30.24$. Обращает на себя внимание заметно более быстрое возрастание риска при уменьшении запаса относительно оптимального, чем при его увеличении. В [56, разд. 8.1] этот пример рассматривается как основа для имитационного эксперимента — со случайно выбранным средним спросом, генерацией соответствующих реализаций, пересчетом вероятностей «интервальных» гипотез и вектора ожидаемого риска.

5.8. Расчет пороговых стратегий

При отсутствии регламентированной периодичности поставок оказывается целесообразным применение стратегий с критическими уровнями — прежде всего вида (s, q) : при снижении запаса до уровня s заказывается партия объема q . Будем считать, что штраф во всех случаях пропорционален ожидаемому дефициту к концу цикла (перед прибытием поставки)

$$\bar{D} = \int_s^{\infty} (x - s) f(x) dx,$$

где $f(x)$ — плотность распределения спроса за время задержки поставки. Определим также среднее значение положительного остатка к концу цикла

$$\bar{R} = \int_0^s (s - x) f(x) dx.$$

Заметим на будущее, что одна из этих величин легко выражается через другую:

$$\begin{aligned}\int_s^{\infty} (x-s)f(x) dx &= \int_s^{\infty} xf(x) dx - s \int_s^{\infty} f(x) dx \\ &= \bar{x} - \int_0^s xf(x) dx - s \left(1 - \int_0^s f(x) dx\right) \\ &= \bar{x} - s + \int_0^s (s-x)f(x) dx.\end{aligned}$$

Здесь \bar{x} — средний спрос за время задержки поставок. Итак,

$$\bar{D} = \bar{x} - s + \bar{R}, \quad \bar{R} = \bar{D} + s - \bar{x}. \quad (5.8.1)$$

Для вычисления начальных приближений (s_0, q_0) к оптимальным параметрам следует:

- 1) минимизируя затраты за время задержки поставки, определить пороговый уровень s_0 согласно (5.6.2) при $c = 0$ и пересчитанной на время $\bar{\tau}$ цене хранения;
- 2) вычислить соответствующие минимальные затраты L за время задержки по формуле (5.6.7);
- 3) полагая цену заказа равной $g + L$, по формуле Уилсона (5.1.8) для максимального запаса вычислить величину S ;
- 4) принять $q_0 = S - s_0$.

Варианты расчетной схемы определяются возможностью переноса дефицита в следующий цикл.

5.8.1. Модель без переноса дефицита

В этом варианте предполагается, что образовавшийся к концу цикла дефицит ликвидируется посредством экстренных поставок, так что запас непосредственно перед поставкой заведомо неотрицателен. Оплата хранения среднего запаса в единицу времени составляет $h(q/2 + R) = h(q/2 + s - \bar{x} + \bar{D})$. Кроме того, в цикле выплачиваются стоимость заказа g и штраф $d\bar{D}$, а средняя длина цикла равна $q/\lambda + \bar{\tau}$. Затраты в единицу времени

$$L = h[q/2 + s - \bar{x} + \bar{D}] + \frac{1}{q/\lambda + \bar{\tau}}[g + d\bar{D}]$$

$$= h(q/2 + s - \lambda\bar{\tau}) + \frac{g}{q/\lambda + \bar{\tau}} + \left(h + \frac{d}{q/\lambda + \bar{\tau}}\right)\bar{D}.$$

Условия оптимальности параметров

$$L'_q = h/2 - g/[\lambda(q/\lambda + \bar{\tau})^2] - \frac{d}{\lambda(q/\lambda + \bar{\tau})^2} \int_s^{\infty} (x-s)f(x) dx = 0,$$

$$L'_s = h - \left[h + \frac{d}{q/\lambda + \bar{\tau}}\right] \int_s^{\infty} f(x) dx = 0$$

можно привести к более удобному виду, допускающему итерационный расчет

$$\begin{aligned}\int_s^{\infty} f(x) dx &= \frac{h}{h + d/(q/\lambda + \bar{\tau})}, \\ q &= \sqrt{\frac{2\lambda}{h} \left[g + d \int_s^{\infty} (x-s)f(x) dx \right]} - \lambda\bar{\tau}.\end{aligned}$$

5.8.2. Модель с переносом дефицита

Здесь ограничение на неотрицательность остатка к концу цикла снимается. В среднем он равен $s - \lambda\bar{\tau}$, так что затраты в единицу времени

$$L = h(q/2 + s - \lambda\bar{\tau}) + \frac{1}{q/\lambda + \bar{\tau}}(g + d\bar{D}).$$

Условия оптимальности параметров принимают вид

$$L'_q = h/2 - \frac{1}{\lambda(q/\lambda + \bar{\tau})^2} \left[g + \int_s^{\infty} (x-s)f(x) dx \right] = 0,$$

$$L'_s = h - \left(\frac{d}{q/\lambda + \bar{\tau}} \right) \int_s^{\infty} f(x) dx = 0.$$

В этой модели s определяется из условия

$$\int_s^{\infty} f(x) dx = \frac{h(q/\lambda + \bar{\tau})}{d},$$

а формула для вычисления q остается прежней

Ниже рассматриваются два частных варианта этой модели — при непрерывном показательном распределенном спросе и при пуассоновском спросе за гамма-распределенное время задержки

5.8.3. Показательно распределенный спрос

В этом случае $\int_s^{\infty} f(x) dx = e^{-\mu s}$, а $\int_s^{\infty} (x - s)f(x) dx = e^{-\mu s} / \mu$

Подставляя эти результаты в формулы для оптимальных параметров с одновременной заменой $\mu = 1/(\lambda\bar{\tau})$, получаем

$$s = \lambda\bar{\tau} \ln \frac{d}{h(q/\lambda + \bar{\tau})},$$

$$q = \sqrt{\frac{2\lambda}{h} \left(g + d\lambda\bar{\tau}e^{-s/(\lambda\bar{\tau})} \right)}.$$

5.8.4. Гамма-распределенная задержка

При пуассоновском потоке заявок и задержке поставок, подчиненной гамма-распределению с параметрами r и μ , вероятности спроса на j единиц вычисляются согласно (3.6.4). Для показательного закона ($r = 1$) результат упрощается

$$p_j = \frac{\mu}{\lambda + \mu} \left(\frac{\lambda}{\lambda + \mu} \right)^j, \quad j = 0, 1, \dots$$

Здесь можно получить явную формулу для вероятности дефицита при пороге заказа s

$$p_-(s) = \frac{\mu}{\lambda + \mu} \sum_{j=s+1}^{\infty} \left(\frac{\lambda}{\lambda + \mu} \right)^j = \left(\frac{\lambda}{\lambda + \mu} \right)^{s+1},$$

а математическое ожидание дефицита

$$y_-(s) = \sum_{j=s+1}^{\infty} (j - s)p_j = \frac{\lambda}{\mu} \left(\frac{\lambda}{\lambda + \mu} \right)^s.$$

Далее, временно полагая s и q вещественными, можно подставить эти выражения в условия оптимальности вместо соответствующих интегралов

Окончательный выбор s и q должен проводиться с учетом их целочисленности. В задачах с учетом штрафов за недостачу при уменьшении запаса функция затрат возрастает значительно быстрее, чем при его увеличении. Поэтому вместо перебора четырех точек с координатами (s, q) , ближайших к найденной, можно применять упрощенное правило s округлять в большую сторону, а q — до ближайшего целого.

5.9. Комбинированные стратегии

Комбинация правил заказа периодических и пороговых стратегий дает стратегию типа (T, s, S) — уровень запаса x контролируется через интервалы T , и при обнаружении остатка $y \leq s$ заказывается партия $S - y$. Это правило сочетает удобства календарного планирования с достаточными возможностями оптимизации. Данный случай сводится к рассмотренному выше, если считать момент пересечения критического уровня распределенным равномерно на интервале $[0, T]$ и взять в качестве распределения перескока соответственно усредненное распределение спроса.

Глава 6

Многономенклатурные задачи

6.1. Классификация номенклатур

Система снабжения, как правило, включает в себя запасы различных предметов с количеством их видов от нескольких единиц до сотен тысяч наименований. Согласно [190], в некоторых военных организациях учитывается свыше 600 тыс позиций, в торговле — до 100 тыс, на среднем предприятии порядка 10 тыс. Дополнительно увеличивают число учитываемых номенклатур различия

- по весу, цвету, размеру, фасону,
- по номиналу (емкости, мощности, сопротивления, быстродействия),
- по виду и кратности упаковки, и т д

При большом числе номенклатур реализация сколько-нибудь сложных методик управления запасами оказывается невозможной, а стоимость информационной системы может перекрыть возможную экономию. Поэтому в зависимости, главным образом, от затрат на снабжение по каждой номенклатуре их делят на группы, подход к которым осуществляется дифференцированно («схема ABC»)

Первым шагом оптимизации УЗ является ранжировка списка номенклатур в порядке убывания стоимостного спроса. По мнению специ-

алистов, строгая оптимизация должна проводиться лишь по группе А из 5–10% номенклатур, суммарный спрос на которые в стоимостном исчислении составляет до 65% от общего. Для группы В (около 25% по составу и 30% по стоимостному спросу) допустимо применение простейших расчетных методов. Для всех остальных (С) возобновление запасов организуется из соображений практического удобства или по стабильным нормам. Попавшие сюда дорогостоящие предметы с крайне низким спросом хранятся на складе высшего звена системы снабжения, их запас пополняется при возникновении каждой потребности.

В [193, с. 191–195] рассматривается обобщение этого подхода с дополнительным учетом критичности и трудности ликвидации дефицита по каждой номенклатуре: I — критическая, III — безразличная, II — промежуточный случай. Итоговая классификация выглядела следующим образом:

AA — {AI, AII, BI},

BB — {AIII, BII, CI},

CC — {BIII, CII, CIII}.

В книге [190] каждой группе ABC-классификации посвящены отдельные главы. К *группе А* отнесены изделия с большим стоимостным спросом, а также некоторые другие — с учетом дополнительных соображений важности, комплектности, трудности организации поставки. Отмечается желательность аналитического прогнозирования спроса. Для деталей с высоким и низким спросом внутри группы используются разные подходы. Уровень запасов контролируется при каждом сообщении (реализуются пороговые правила, обычно $q = 1$). При низком спросе ремонт и замена планируются при проведении регламента. Учет штрафов считается предпочтительнее работы с уровнем обслуживания.

В особо сложных случаях для оценки стратегий управления запасами по номенклатурам группы А рекомендуется имитационное моделирование. К таким ситуациям относятся:

- зависимый спрос;
- совместный заказ со скидками;
- эшелонированные системы;
- переходные (нестационарные) режимы работы;

- планирование последовательности обработки;
- агрегированное планирование производства и снабжения.

В [56] дан обзор выполненных в 1960-х гг. работ фирмы RAND Corporation по моделированию систем снабжения американских ВВС (в том числе баз межконтинентальных ракет). Сейчас эти работы интересны лишь с исторической точки зрения и как пример сочетания военной научно-исследовательской игры с обучением линейного персонала. По их итогам был принят ряд важных и весьма эффективных организационных решений.

В упомянутых и многих других работах по имитационному моделированию подчеркивается, однако, его серьезный недостаток: тестируется только запланированный набор решений, среди которых наилучшее может отсутствовать. Добавим к этому низкую (без использования методов понижения дисперсии [71]) относительную точность определения вероятностей редких событий, которыми в данном случае являются особенно интересующие нас случаи дефицита запаса.

К планированию управления запасами изделий *группы В* применяются такие методы, как задача газетчика и расчет пороговых стратегий при случайном спросе. Могут рассматриваться задачи с совместными ограничениями (на множители Лагранжа). Множители Лагранжа интерпретируются как цены дефицитного ресурса. При решении задач этого типа рекомендуется всегда выполнять расчет с ослабленными ограничениями — для оценки его влияния на целевую функцию и качество снабжения. Такой расчет может служить основанием для пересмотра ограничений: в меньшую сторону для использования дефицитных ресурсов в других целях, в большую — при выявлении возможности заметного выигрыша ценой незначительного ослабления ограничений.

Для изделий *группы С* решения принимаются по совокупности. Из этой группы могут быть переведены в *В* критически важные, трудно добываемые, предметы особого интереса начальника [190]. Информация о спросе регистрируется укрупненно (напр., измеряется в коробках). Соответственно может быть сформулировано и правило восполнения: выдавать заказ, когда будет вскрыта последняя коробка.

По группе *С* обычно создают большой страховой запас, так как это обходится недорого. Но спрос должен быть обеспечен в комплекте. Поскольку этих номенклатур очень много, вероятность дефицита хотя бы

по одной велика. для примера укажем, что при $n = 20$ и надежности обеспечения по каждой 0.95 она составит $1 - 0.95^{20} = 0.64$

Особая проблема в этой группе — как поступать с очень редко требуемыми изделиями (согласно [190, с. 432], процент деталей, не истребованных за год ни разу, составляет от 16 до 47) Здесь рассматриваются следующие альтернативы.

- вернуть поставщикам (со скидкой),
- передать в высшее звено;
- продавать в нагрузку или давать в виде премии за покупку;
- использовать на благотворительные цели;
- устроить аукцион;
- продать в утиль для переработки.

Ниже рассматриваются алгоритмы управления запасами для номенклатур, достойных серьезного внимания экономиста и исследователя. Задачу иногда удается расчленить на ряд однопродуктовых. Но в тех случаях, когда допускается совмещение заказов по нескольким номенклатурам (общий поставщик), штрафы исчисляются с учетом обеспечения спроса по группе номенклатур (общий штраф) или же имеются общие ограничения, оптимизация должна проводиться для *группы* номенклатур, объединяемых одним или несколькими из перечисленных факторов.

6.2. Детерминированный спрос

Предположим, что стоимость организации поставки партии из n номенклатур от одного поставщика можно представить в форме

$$g(n) = g_0 + \sum_{i=1}^n g_i, \quad (6.2.1)$$

где g_0 — цена акта заказа. При высокой цене штрафа и независимых заказах с оптимальной для каждой номенклатуры периодичностью согласно (5.1.10) суммарные расходы в единицу времени составят

$$L = \sum_{i=1}^N \sqrt{2(g_0 + g_i)\lambda_i h_i}. \quad (6.2.2)$$

При одновременном заказе всех N номенклатур его периодичность будет в общем случае отличаться от оптимальной периодичности по каждой из компонент при независимом снабжении, что приведет к некоторому увеличению расходов. С другой стороны, это позволит сэкономить на заказах. Суммарные затраты в единицу времени при данной организации поставок должны подсчитываться по формуле

$$L = \frac{1}{2}T \sum_{i=1}^N \lambda_i h_i + \frac{1}{T} \sum_{i=0}^N g_i$$

и при оптимальном выборе T составят

$$L = \sqrt{2 \sum_{i=0}^N g_i} \sqrt{\sum_{i=1}^N \lambda_i h_i}. \quad (6.2.3)$$

Сравнение формул (6.2.2) и (6.2.3) указывает на то, что совместная оптимизация может дать некоторую экономию.

Достоинства обоих рассмотренных подходов (независимой и полностью совмещенной оптимизации) соединяют поставки по системе *кратных периодов*. При этом отдельные номенклатуры со сходными значениями стоимостного спроса объединяются в совместно заказываемые группы, что позволяет получить малое расхождение между групповым и индивидуальным оптимумами. За счет же кратности периодов снабжения по группам удается достичь частого совмещения заказов. Такая стратегия была впервые предложена автором в [52] (1966 г.) и многократно переоткрывалась в дальнейшем [104, 105, 106, 109, 122, 134, 136, 146, 158, 163, 166, 170, 175, 203].

Поставим задачу о расчете оптимальной системы кратных периодов, в которой по крайней мере одна из номенклатур заказывается в каждом *базисном* периоде T . Обозначим:

$\langle k \rangle$ — множество номенклатур с периодичностью поставок kT ,
 $k = 1, 2, \dots$;

n_k — число элементов такого множества.

Для некоторых k соответствующие множества могут быть и пустыми.

Расходы на снабжение i -й номенклатурой в единицу времени при $k_i \geq 2$ составляют

$$L_i = \frac{1}{2} \lambda_i h_i k_i T + \frac{g_i}{k_i T},$$

а для номенклатур первого множества

$$L_i = \frac{1}{2} \lambda_i h_i T + \frac{g_i}{T} + \frac{g_0}{n_1 T}$$

(здесь стоимость заказа, не зависящая от числа номенклатур, равномерно разложена по номенклатурам первого множества). Необходимо таким образом выбрать базисный период T и так произвести разбивку всех номенклатур на упомянутые множества, чтобы сумма

$$L(T) = \frac{1}{2}T \sum_{i=1}^N \lambda_i h_i k_i(T) + \frac{1}{T} \left(g_0 + \sum_{i=1}^N \frac{g_i}{k_i(T)} \right) \quad (6.2.4)$$

была минимальной.

Найдем производные функции $L(T)$ в интервалах постоянства группировок, т.е. при фиксированных $\{k_i\}$:

$$\frac{dL}{dT} = \frac{1}{2} \sum_{i=1}^N \lambda_i h_i k_i - \frac{1}{T^2} \left(g_0 + \sum_{i=1}^N \frac{g_i}{k_i} \right),$$

$$\frac{d^2L}{dT^2} = \frac{2}{T^3} \left(g_0 + \sum_{i=1}^N \frac{g_i}{k_i} \right).$$

Вторая производная на этих интервалах всегда положительна. Рассмотрим поведение dL/dT в крайних областях полуоси $[0, \infty)$. При достаточно большом T все номенклатуры окажутся в множестве $\langle 1 \rangle$ и производная будет равна

$$\frac{dL}{dT} = \frac{1}{2} \sum_{i=1}^N \lambda_i h_i - \frac{1}{T^2} \sum_{i=0}^N g_i,$$

откуда

$$\lim_{T \rightarrow \infty} dL/dT = \frac{1}{2} \sum_{i=1}^N \lambda_i h_i > 0.$$

С другой стороны, при очень малых T каждое множество при неравных $\{\lambda_i h_i\}$ сведется к одному элементу. Тогда главную часть расходов в единицу времени составит g_0/T ; соответственно $dL/dT = -g_0/T^2$ и $\lim_{T \rightarrow 0} (dL/dT) = -\infty$.

Найдем теперь границы постоянства группировок. Коэффициенты при T и $1/T$, входящие в (6.2.4), меняются при перегруппировке номенклатур, связанной с переходом номенклатур множества $\langle k \rangle$ в множество $\langle k-1 \rangle$ при увеличении T . Очевидным условием такого перехода i -й номенклатуры является неравенство

$$\left(\frac{1}{2} \lambda_i h_i k_i T + \frac{g_i}{k_i T} \right) - \left[\frac{1}{2} \lambda_i h_i (k_i - 1) T + \frac{g_i}{(k_i - 1) T} \right] \geq 0,$$

а момент перехода определяется равенством уменьшаемого и вычитаемого, т.е.

$$T_{\text{п}i} = \sqrt{\frac{2g_i}{\lambda_i h_i k_i (k_i - 1)}} \quad (6.2.5)$$

(индекс «п» означает *правую* границу интервала).

Расположим все номенклатуры в порядке возрастания отношений $\{g_i/(\lambda_i h_i)\}$ и перенумеруем их заново. Тогда условие (6.2.5) будет выполнено прежде всего для первого элемента множества $\langle k \rangle$, имеющего наименьшую в данном множестве величину упомянутого отношения. Пусть в точке $T_{\text{п}r}$ номенклатура r перешла из множества $\langle k \rangle$ в множество $\langle k-1 \rangle$. Функция $L(T)$, очевидно, непрерывна. Скачок производной в точке перегруппировки $T_{\text{п}r}$ составит

$$\begin{aligned} \Delta &= \frac{1}{2} \lambda_r h_r [(k_r - 1) - k_r] - \frac{g_r}{T_{\text{п}r}^2} \left[\frac{1}{k_r - 1} - \frac{1}{k_r} \right] \\ &= -\frac{g_r}{T_{\text{п}r}^2} \frac{1}{k_r (k_r - 1)} - \frac{1}{2} \lambda_r h_r. \end{aligned}$$

Но согласно (6.2.5)

$$T_{\text{п}r} = \sqrt{\frac{2g_r}{\lambda_r h_r k_r (k_r - 1)}}.$$

Следовательно,

$$\Delta = -\frac{g_r \lambda_r h_r k_r (k_r - 1)}{2g_r k_r (k_r - 1)} - \frac{1}{2} \lambda_r h_r = -\lambda_r h_r.$$

Таким образом, функция $L(T)$ непрерывна, кусочно-выпукла и имеет отрицательные скачки производной на правых границах интервалов

постоянства группировок. Очевидно, что минимум функции $L(T)$, задаваемой правой частью (6.2.4), равен наименьшему из минимумов ее, достигаемых при постоянной группировке номенклатур. Наивыгоднейший период при постоянной группировке определяется приравнением нулю производной dL/dT

$$T^* = \sqrt{2 \left(\sum_{i=1}^N \frac{g_i}{k_i} + g_0 \right) / \sum_{i=1}^N \lambda_i h_i k_i}. \quad (6.2.6)$$

Ему соответствуют минимальные затраты в единицу времени

$$L^* = \sqrt{2 \left(\sum_{i=1}^N \frac{g_i}{k_i} + g_0 \right) \sum_{i=1}^N \lambda_i h_i k_i}. \quad (6.2.7)$$

Обозначим числитель и знаменатель подкоренного выражения в (6.2.6) через G и H соответственно и найдем их приращения при переходе номенклатуры r из множества $\langle k_r \rangle$ в множество $\langle k_r - 1 \rangle$, т.е. за правой границей интервала

$$\begin{aligned} \Delta G_{\text{п}r} &= 2g_r \left(\frac{1}{k_r - 1} - \frac{1}{k_r} \right) = \frac{2g_r}{k_r (k_r - 1)}, \\ \Delta H_{\text{п}r} &= \lambda_r h_r (k_r - 1) - \lambda_r h_r k_r = -\lambda_r h_r. \end{aligned} \quad (6.2.8)$$

Поскольку в новых обозначениях

$$T^* = \sqrt{G/H} \quad (6.2.9)$$

и

$$L^* = \sqrt{GH}, \quad (6.2.10)$$

уменьшение суммарных расходов после такого перехода возможно лишь при условии $\Delta G/G < |\Delta H|/H$, или $\Delta G/|\Delta H| < G/H$. Извлекая из обеих частей последнего равенства квадратные корни и подставляя значения ΔG и ΔH , убеждаемся в его эквивалентности условию

$$T_{\text{п}r} < T^*. \quad (6.2.11)$$

Как видно из формулы (6.2.5), величина $T_{\text{п}r}$ быстро возрастает с уменьшением k_r , тогда как T^* увеличивается очень медленно. Таким

образом, выполнение условия $T_{nr} \geq T^*$ для всех r говорит о нецелесообразности дальнейшего увеличения T и связанного с ним уменьшения коэффициентов $\{k_r\}$, так как минимумы L^* будут возрастать

Аналогичные рассуждения могут быть проведены и для возрастающих коэффициентов $\{k_r\}$ при движении по оси T в сторону уменьшения базисного периода. Расчетное соотношение для *левой* границы интервала постоянства группировок

$$T_{lr} = \sqrt{\frac{2g_r}{\lambda_r h_r k_r (k_r + 1)}} \quad (6.2.12)$$

определяется *последним* элементом группы и получается из (6.2.5) заменой k на $k+1$. Формулы для расчета приращений ΔG и ΔH принимают вид

$$\begin{aligned} \Delta G_{lr} &= -\frac{2\gamma g}{k_r(k_r + 1)}, \\ \Delta H_{lr} &= \lambda_r h_r. \end{aligned} \quad (6.2.13)$$

Условием целесообразности уменьшения T с одновременным увеличением коэффициентов $\{k_r\}$ становятся неравенства

$$T_{lr} > T^*, \quad r = \overline{1, N}$$

Приведенные соображения позволяют утверждать, что графики функций $L(T)$ и $L'(T)$ имеют вид, показанный на рис. 6.1 (штриховыми линиями показаны границы постоянства группировок, пунктиром — локальные оптимумы). Они дают основания предложить следующий алгоритм планирования многономенклатурных поставок

- 1) Упорядочить номенклатуры по возрастанию отношений $\{g_i/(\lambda_i h_i)\}$ и произвести их перенумерацию
- 2) Выбрать начальное приближение для базисного периода T_0
- 3) Рассчитать для него оптимальный набор коэффициентов $\{k_i^0\}$
- 4) Рассчитать значения G, H для этих коэффициентов по формулам (6.1.6) и (6.1.7). Вычислить T^* и L^*
- 5) По формуле (6.2.5) рассчитать *правые* границы интервала постоянства группировок $\{T_{pi}(k_i^0)\}$, по (6.1.12) — *левые*
- 6) Установить в нуль указатель key направления изменения T

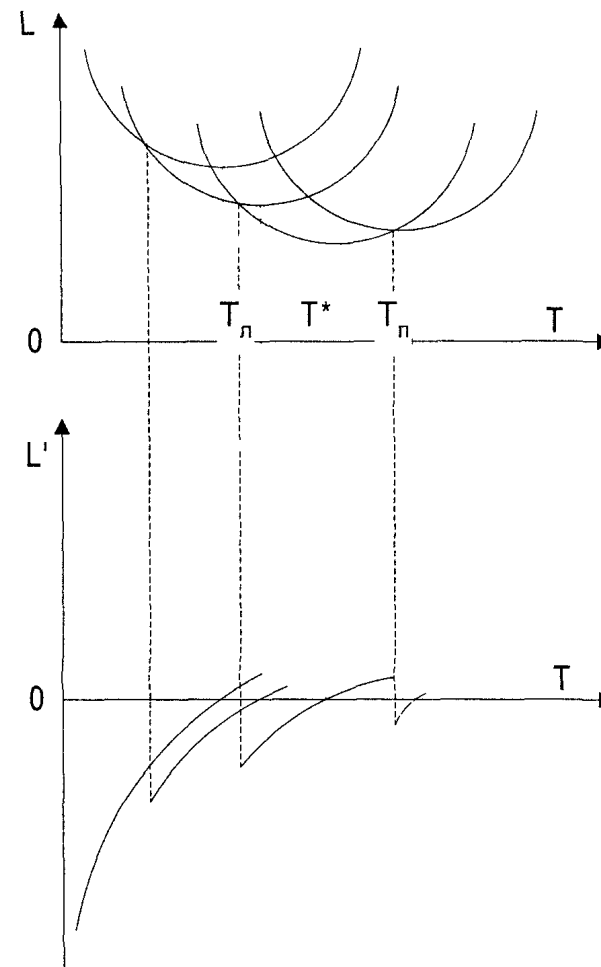


Рис. 6.1. Функция затрат и ее производная

- 7) Если существует такое i , что $T_{pi} < T^*$, положить $key = 1$, иначе, если существует такое i , что $T_{li} > T^*$, положить $key = -1$
- 8) Пока $key = -1$ установить $key = 0$. Рассчитать ΔG и ΔH по всем номенклатурам, для которых $\{T_{li}\} > T^*$, найти $\Sigma \Delta G$ и $\Sigma \Delta H$, заменить для всех этих номенклатур k_i на $k_i + 1$ и рассчитать новые $\{T_{li}\}$. Вычислить новые значения $G = G + \Delta G$, $H = H + \Delta H$ и пересчитать по ним T^* и L^* . Если корректировки были, установить $key = -1$ и вернуться на начало пункта

- 9) Пока $key = 1$ установить $key = 0$. Рассчитать ΔG и ΔH по всем номенклатурам, для которых $\{T_{Pi}\} < T^*$, найти $\Sigma \Delta G$ и $\Sigma \Delta H$, заменить для всех этих номенклатур k_i на $k_i - 1$ и рассчитать новые $\{T_{Pi}\}$. Вычислить новые значения $G := G + \Delta G$, $H = H + \Delta H$ и пересчитать по ним T^* и L^* . Если корректировки были, установить $key = 1$ и вернуться на начало пункта.
- 10) Восстановить первоначальную нумерацию и выдать T^* , L^* и набор $\{k_i\}$.
- 11) Конец алгоритма.

Процесс увеличения (уменьшения) T заканчивается за конечное число шагов, так как в первом случае с переходом всех номенклатур в множество $\langle 1 \rangle$ оказывается $\{T_{Pi}\} = \infty$, тогда как T^* имеет конечное значение $\sqrt{2 \sum_{i=0}^N g_i / \sum_{i=1}^N \lambda_i h_i}$. При уменьшении же T границы $\{T_{Pi}\}$, в знаменатели которых входят члены порядка k , уменьшаются быстрее, чем T^* , и условие $\{T_{Pi}\} \leq T^*$ в конце концов будет выполнено для всех i .

Остается рассмотреть вопрос о выборе начального значения T_0 и соответствующей ему группировке номенклатур. В качестве T_0 разумно принять

$$T = \sqrt{2(g_0 + g_1) / (\lambda_1 h_1)}, \quad (6.2.14)$$

т.е. наивыгоднейший период заказа для первой номенклатуры — с минимальной величиной отношения $g_i / (\lambda_i h_i)$ — при независимой оптимизации. Основанием для такого решения является очевидная необходимость достаточной близости итогового T^* к оптимальной периодичности наиболее часто заказываемой номенклатуры.

В процессе начальной группировки номенклатур каждой из них должно быть поставлено в соответствие такое k_i , при котором расходы на ее поставки будут минимальны. По условию, $k_1 = 1$. При возрастании $\{g_i / \lambda_i h_i\}$ по i , что обеспечивается предварительным упорядочением, индивидуальные оптимальные периоды и, следовательно, $\{k_i\}$ будут возрастать по i . Это позволяет не проверять i -ю номенклатуру на значения k_i , меньшие k_{i-1} . Увеличение же k_i целесообразно при

$$\left(\frac{\lambda_i h_i k T + \frac{g_i}{k T}}{2} \right) - \left[\frac{\lambda_i h_i (k+1) T + \frac{g_i}{(k+1) T}}{2} \right] \geq 0,$$

или

$$\lambda_i h_i \leq \frac{2g_i}{k(k+1)T^2} = R_k(T). \quad (6.2.15)$$

Подставляя значение T_0 , можно получить

$$R_k(T_0) = \frac{g_i \lambda_i h_i}{(g_0 + g_i) k(k+1)}. \quad (6.2.16)$$

Для i -й номенклатуры оптимальным является наименьшее k , при котором $\lambda_i h_i \geq 2g_i / [k(k+1)T^2]$.

В [56] приведен пример реализации описанного алгоритма для реальных исходных данных на примере снабжения ЭВМ первого поколения лампами и полупроводниковыми диодами с пошаговым выводом промежуточных результатов. Наиболее интересно сопоставление суммарных затрат:

- исходная организация поставок — 52.50;
- раздельная оптимизация — 23.70;
- объединение поставок по всем номенклатурам — 14.58;
- система кратных периодов при группировке по T_0 — 12.81;
- она же при группировке по оптимальному периоду — 12.74.

В [77] сообщается, что эти модели широко применялись в Белоруссии при формировании оптимальных поставочных комплектов строительной и электротехнической продукции на базу Белстройснабсбыта. Ожидаемые затраты при раздельной оптимизации составляли 84%, при совместной — 60%, а при системе кратных периодов лишь 42% от стартовых.

Об улучшении начальных приближений. В [158] предложено выбирать T_0 по множеству номенклатур, заказываемых в каждом периоде. Пусть выполнено упорядочение по убыванию отношений $\{g_i / (h_i \lambda_i)\}$ и первые j номенклатур заказываются одновременно. Оптимальная периодичность при этом определится согласно

$$T_0 = \sqrt{2 \sum_{i=0}^j g_i / \sum_{i=1}^j h_i \lambda_i}. \quad (6.2.17)$$

Присоединение к этой же группе следующей номенклатуры невыгодно, если для нее при указанном T_0 двойная кратность предпочтительнее одинарной, т.е.

$$\frac{g_{j+1}}{T_0} + \frac{h_{j+1}\lambda_{j+1}}{2}T_0 > \frac{g_{j+1}}{2T_0} + h_{j+1}\lambda_{j+1}T_0,$$

или

$$g_{j+1} > h_{j+1}\lambda_{j+1}T_0^2.$$

Подставляя сюда вышеуказанное значение T_0 , приходим к условию прекращения накопления $\langle 1 \rangle$ -группы

$$\frac{g_{j+1}}{h_{j+1}\lambda_{j+1}} > 2 \sum_{i=0}^j g_i / \sum_{i=1}^j h_i \lambda_i. \quad (6.2.18)$$

Проверка начинается со второй номенклатуры, начальные значения сумм в правой части $S_1 = g_0 + g_1$, $S_2 = h_1 \lambda_1$. Суммы пересчитываются рекуррентно. При выполнении условия (6.2.18) для j -й номенклатуры для всех последующих $i > j$ вычисляется оптимальная периодичность T_i и по отношению T_i/T_0 — начальная кратность. Далее в соответствии со схемой основного алгоритма определяется необходимость уточнения T_0 и соответствующего пересчета кратностей.

Реализующая этот алгоритм программа на Фортране 90 (вместе с внутренней процедурой упорядочения) дана в Приложении А. В табл. 6.1 приводятся исходные данные и результаты расчета системы кратных периодов снабжения для 19 номенклатур. Коэффициент $g_0 = 1.33$ задан в тексте программы, значения $\{g_i\} = 0.29$ вводились из файла и для экономии места в таблицу не включены. Стоимости хранения и интенсивности спроса (два левых столбца) упорядочены в соответствии с исходной нумерацией номенклатур. Далее следуют новая нумерация, ее соответствие старой и результаты счета — критерий упорядочения, стартовые кратности и границы интервалов постоянства группировок, оптимальные кратности.

Затраты при независимой оптимизации составляют 28.623 и для дальнейших сопоставлений приняты за 100%. При одновременных поставках (период равен 0.677) они равны 20.200 (70.58%). Начальный этап алгоритма приводит к затратам 16.367, или 57.18%. Минимум затрат 16.356 (57.14%) был достигнут на третьем шаге основного алгоритма при движении влево по оси T и соответствовал базисному периоду 0.416.

Т а б л и ц а 6.1

Расчет системы кратных периодов

h	λ	i_{new}	i_{old}	$g/(h * \lambda)$	k	T_n	T_p	k^*
0.900e-2	0.975e2	1	15	.183e-1	1	135	∞	1
0.300e-1	0.196e3	2	2	.493e-1	1	222	∞	1
0.100e-1	0.382e2	3	13	.847e-1	1	291	∞	1
0.150e-1	0.126e2	4	1	.330e+0	2	.332	.575	2
0.420e-1	0.740e1	5	11	.351e+0	2	.342	.593	2
0.100e-1	0.650e1	6	8	.365e+0	2	.349	.604	2
0.800e-2	0.340e1	7	9	.546e+0	2	.427	.739	3
0.100e-1	0.795e2	8	3	.759e+0	3	.356	.503	3
0.150e-1	0.354e2	9	5	.933e+0	3	.394	.558	3
0.100e-1	0.115e2	10	4	.153e+1	4	.392	.506	4
0.140e-0	0.590e1	11	14	.175e+1	4	.418	.540	5
0.110e-1	0.640e1	12	10	.252e+1	5	.410	.502	5
0.500e-1	0.685e2	13	16	.293e+1	5	.442	.541	6
0.200e-1	0.830e1	14	17	.300e+1	5	.447	.548	6
0.800e-1	0.198e3	15	12	.412e+1	6	.443	.524	7
0.200e-2	0.495e2	16	19	.426e+1	6	.451	.533	7
0.200e-2	0.483e2	17	6	.446e+1	7	.399	.461	7
0.600e-2	0.106e2	18	18	.456e+1	7	.404	.466	7
0.200e-2	0.340e2	19	7	.107e+2	10	.440	.487	11

Таким образом, уже начальная группировка дала решение, практически совпадающее с оптимальным по значению целевой функции.

Описанное выше «улучшенное» приближение было связано с затратами 16.928 и для достижения минимума (того же) потребовало двух дополнительных шагов. Это не исключает возможного преимущества «улучшенного» начального приближения при других исходных данных.

Аргументами в пользу объединения заказов являются также оптовые скидки и возможное требование поставщика о стоимости каждого заказа не ниже некоторого порога. В любом случае начинать следует с реализации вышеописанного алгоритма. Если стоимость заказа по $\langle 1 \rangle$ -множеству превышает порог скидки, то расчет следует повторить с учетом скидок и на этом вычисления закончить. Иначе необходимо подсчитать стоимости заказов для каждого периода из набора в K периодов, где K — наименьшее общее кратное чисел $\{k_i\}$, определить

потери от упущенных скидок в единицу времени и отнести их к общим затратам на организацию поставок. Если потери существенны, необходима модификация алгоритма с целью максимизации полученных скидок [190, с. 504–505].

6.3. Взаимозаменяемые продукты

Рассмотрим двухпродуктовую модель с возможной заменой второго продукта первым. Предполагается, что на замену единицы второго продукта требуется k единиц первого и дополнительные затраты c . На рис. 6.2 представлены синхронизированные графики изменения запасов первого и второго типов для одного цикла работы системы.

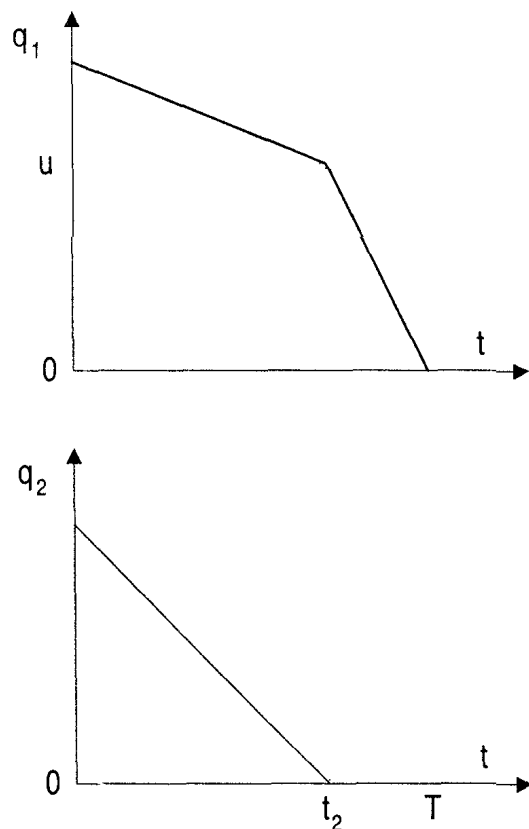


Рис. 6.2. К модели с заменой продукта

Затраты за цикл и длительность цикла удобно выразить через объем поставки q_2 и остаток u запаса по первому продукту, обеспечивающий покрытие спроса по обоим продуктам после исчерпания запаса по второму. Прежде всего отметим, что $t_2 = q_2/\lambda_2$. Далее,

$$q_1 = u + q_2\lambda_1/\lambda_2.$$

Расходы на хранение за цикл пропорциональны «импульсу» соответствующего запаса. Для *второго* продукта на участке $[0, t_2]$ средний запас равен $q_2/2$, а «импульс запаса» — $q_2^2/(2\lambda_2)$. На заключительной фазе цикла запас и импульс равны нулю. По *первому* продукту на участке $[0, t_2]$ средний запас равен u плюс половина спроса за это время, что приводит к импульсу $[u + \lambda_1 q_2/(2\lambda_2)]q_2/\lambda_2$. На участке $[t_2, T]$ запас u расходуется за время $u/(\lambda_1 + k\lambda_2)$, а средний запас равен $u/2$. Импульс запаса равен $u^2/[2(\lambda_1 + k\lambda_2)]$.

В каждом цикле выплачиваются также стоимость выполнения g и стоимость компенсации дефицита второго продукта соответствующим объемом первого по цене c . Этот объем равен пересчитанной к первому продукту потребности во втором в заключительной фазе цикла, т.е. $k\lambda_2 u/(\lambda_1 + k\lambda_2)$. Таким образом, суммарные затраты за цикл

$$\begin{aligned} L_T &= g + c u k \lambda_2 / (\lambda_1 + k \lambda_2) + h_2 q_2^2 / (2 \lambda_2) \\ &+ h_1 \left[\left(u + \frac{\lambda_1 q_2}{2 \lambda_2} \right) \frac{q_2}{\lambda_2} + \frac{u^2}{2(\lambda_1 + k \lambda_2)} \right] \\ &= g + u \left(\frac{c k \lambda_2}{\lambda_1 + k \lambda_2} + \frac{h_1 q_2}{\lambda_2} \right) \\ &+ \frac{q_2^2}{2 \lambda_2^2} (h_1 \lambda_1 + h_2 \lambda_2) + \frac{h_1 u^2}{2(\lambda_1 + k \lambda_2)}. \end{aligned}$$

Для перехода к затратам в единицу времени эта сумма должна быть разделена на длительность цикла

$$T = q_2/\lambda_2 + u/(\lambda_1 + k\lambda_2).$$

Оптимальные u и q_2 находятся численными методами. Чтобы получить начальные приближения, запишем функцию затрат за период T по второй номенклатуре. В нее войдут расходы:

- на содержание среднего запаса $q_2/2$ в течение q_2/λ_2 — итого $h_2 q_2^2/(2\lambda_2)$;

- на покрытие избыточного спроса по второй номенклатуре запасом по первой — всего $ck\lambda_2(T - q_2/\lambda_2) = ck(\lambda_2 T - q_2)$,
- на хранение всего этого запаса в течение q_2/λ_2 и в среднем его половины — на интервале длины $(T - q_2/\lambda_2)$, итого

$$h_1 \left[k\lambda_2(T - q_2/\lambda_2) \frac{q_2}{\lambda_2} + \frac{k\lambda_2}{2}(T - q_2/\lambda_2)^2 \right] = \frac{h_1 k}{2\lambda_2} [(\lambda_2 T)^2 - q_2^2].$$

Суммируя найденные компоненты, получаем общие затраты за цикл по второй номенклатуре

$$L_2 = \frac{h_2 q_2^2}{2\lambda_2} + ck(\lambda_2 T - q_2) + \frac{h_1 k}{2\lambda_2} [(\lambda_2 T)^2 - q_2^2].$$

Дифференцируя по q_2 , имеем условие его оптимальности

$$\frac{\partial L_2}{\partial q_2} = \frac{h_2 q_2}{\lambda_2} - ck - \frac{h_1 k q_2}{\lambda_2} = 0,$$

откуда

$$q_2 = \frac{ck\lambda_2}{h_2 - kh_1}.$$

Далее находим $t_2 = q_2/\lambda_2$ и вычисляем $u = (\lambda_1 + k\lambda_2)(T - t_2)$.

6.4. Планирование запасов при ограничениях

6.4.1. Метод множителей Лагранжа

Как было показано Куном и Таккером, оптимальное решение и связанный с ним вектор множителей образуют седловую точку функции Лагранжа. В частности, проблема оптимального комплектования наборов запчастей сводится к поиску максимума скалярной целевой функции

$$\max_{X \in S} \Phi(X), \quad G(X) \leq \bar{C}, \quad X = \{x_i\}, \quad x_i = 0, 1, 2, \dots, \quad i = \overline{1, I}. \quad (6.4.1)$$

Целевая функция обычно монотонна и выпукла, аргументы — целочисленные неотрицательные, вектор-функция $G(X)$ состоит из линейных компонент (стоимость, вес, объем). Чаще всего ограничение одно

Существенный элемент метода — замена задачи (6.4.1) на максимизацию вспомогательной функции

$$\max_{X \in S} [\Phi(X) - \theta G(X)]. \quad (6.4.2)$$

Здесь θ — множитель Лагранжа (вещественный, неотрицательный). В случае m ограничений решается задача максимизации

$$\Phi(X) - \sum_{i=1}^m \theta_i G_i(X).$$

6.4.2. Многономенклатурные закупки

Совместное ограничение чаще всего имеет вид

$$C = k \sum_j q_j c_j,$$

где k — коэффициент одновременности (практически 1/2), q_j — объем партии; c_j — стоимость закупки единицы товара. Цену хранения представим в виде $h_j = f c_j$. Для «бездефицитной» модели с мгновенным исполнением функция затрат

$$L = \sum_j \left(\frac{\lambda_j g_j}{q_j} + f \frac{q_j c_j}{2} \right) + z \left(C - k \sum_j q_j c_j \right),$$

где C — допустимые капиталовложения и z — множитель Лагранжа. Оптимальное решение при известном значении z

$$q_j = \sqrt{\frac{2\lambda_j g_j}{c_j (f - 2kz)}}.$$

Следовательно, ограничение можно переписать в виде

$$k \sum_j \sqrt{\frac{2\lambda_j c_j g_j}{f - 2kz}} = C.$$

Полагая $\sum_j \sqrt{2\lambda_j c_j g_j} = U$, преобразуем его в $U/\sqrt{f - 2kz} = C/k$ и после несложных преобразований находим

$$z^* = [f - (kU/C)^2]/(2k).$$

Более общие (и более реалистические) постановки задачи с ограничениями типа неравенств приводят к проблеме математического программирования, обычно решаемой численно с применением различных модификаций градиентного спуска

6.5. Периодические поставки, вероятностный спрос

6.5.1. Заданная периодичность

В данном разделе будем считать длину периода фиксированной (что избавляет нас от учета фиксированных составляющих стоимости заказов), а все стоимостные параметры и распределения спроса — заданными в расчете на этот период. Затраты на хранение естественно принять пропорциональными остаткам к концу периода

Имея в виду комплектное обеспечение спроса, следует исчислять штраф на основе распределения максимума дефицита

$$F(u) = \prod_i F_i(S_i + u),$$

где $\{S_i\}$ — запланированные уровни запасов. Однако при этом получить простое решение не удастся. Поэтому будем считать штраф пропорциональным максимуму взвешенного ожидания дефицита, т.е. исчислять его как

$$\max_i d_i \int_{S_i}^{\infty} (x - S_i) f_i(x) dx \quad (6.5.1)$$

С учетом затрат на доведение исходных уровней запаса $\{z_i\}$ до расчетных минимизации подлежит сумма затрат за период

$$L(\vec{S}) = \sum_{i=1}^N \left[c_i (S_i - z_i) + h_i \int_0^{S_i} (S_i - x) f_i(x) dx \right] + \max_i d_i \int_{S_i}^{\infty} (x - S_i) f_i(x) dx. \quad (6.5.2)$$

Из (6.5.2) видно, что транспортные затраты и издержки на хранение могут быть сокращены без какого-либо увеличения штрафа —

выравниванием взвешенных ожидаемых дефицитов через уменьшение запасов. Тогда для всех i можно записать

$$d_i \int_{S_i}^{\infty} (x - S_i) f_i(x) dx = d_1 \int_{S_1}^{\infty} (x - S_1) f_1(x) dx. \quad (6.5.3)$$

Уравнения (6.5.3) указывают на связь между дифференциалами $\{S_i\}$ в оптимальной точке вида

$$d_i \left(\int_{S_i}^{\infty} f_i(x) dx \right) dS_i = d_1 \left(\int_{S_1}^{\infty} f_1(x) dx \right) dS_1,$$

откуда

$$\frac{dS_i}{dS_1} = \frac{d_1}{d_i} \int_{S_1}^{\infty} f_1(x) dx \left(\int_{S_i}^{\infty} f_i(x) dx \right)^{-1}. \quad (6.5.4)$$

Для нахождения оптимального S_1 продифференцируем (6.5.2) по S_1 и приравняем производную нулю:

$$\sum_{i=1}^N \frac{d_1}{d_i} \left[c_i + h_i \int_0^{S_i} f_i(x) dx \right] \frac{\int_{S_1}^{\infty} f_1(x) dx}{\int_{S_i}^{\infty} f_i(x) dx} - d_1 \int_{S_1}^{\infty} f_1(x) dx = 0.$$

Таким образом, оптимальный набор $\{S_i\}$ для функции затрат (6.5.2) дается решением системы уравнений

$$\sum_{i=1}^N \frac{1}{d_i} \left[(c_i + h_i) \left(\int_{S_i}^{\infty} f_i(x) dx \right)^{-1} - h_i \right] - 1 = 0, \\ d_i \int_{S_i}^{\infty} (x - S_i) f_i(x) dx - d_1 \int_{S_1}^{\infty} (x - S_1) f_1(x) dx = 0, \quad i = \overline{2, N}. \quad (6.5.5)$$

Необходимым условием существования решения этой системы в области неотрицательных $\{S_i\}$ является $\sum_{i=1}^N c_i/d_i < 1$ — оно получается из первого уравнения системы (6.5.5) при всех $S_i = 0$.

Решение подобных систем возможно только численно. В [56, §6.3] приводится пример расчета запаса по трем номенклатурам со спросом за период, распределенным по закону Релея, и начальными $\{S_i\}$, найденными независимой оптимизацией. Решение было получено методом Ньютона. При большом числе номенклатур удобнее, однако, свести задачу к одномерному поиску S_1 , добиваясь выполнения первого из уравнений (6.5.5) и на каждом шаге выравнивая штрафы.

Увеличение запаса по некоторым номенклатурам относительно текущего наличия может оказаться нецелесообразным. Критерий этого — неравенство

$$d_i \int_0^{\infty} x f_i(x) dx \leq d_1 \int_{S_1}^{\infty} (x - S_1) f_1(x) dx,$$

которое можно использовать и на промежуточных этапах алгоритма минимизации затрат. Соответствующие индексы из дальнейших вычислений исключаются.

В заключение данного раздела приведем варианты основной системы уравнений для других способов построения функции затрат. При расходах на хранение и штрафы по «импульсам» положительно-го остатка и дефицита, т.е. для целевой функции

$$L(\vec{S}) = \sum_{i=1}^N \left\{ c_i (S_i - z_i) + h_i \left[\int_0^{S_i} (S_i - x/2) f_i(x) dx + \frac{S_i^2}{2} \int_{S_i}^{\infty} \frac{1}{x} f_i(x) dx \right] \right. \\ \left. + \frac{1}{2} \max_i d_i \int_{S_i}^{\infty} \frac{(x - S_i)^2}{x} f_i(x) dx, \right.$$

оптимальные решения должны удовлетворять системе уравнений

$$\sum_{i=1}^N \frac{1}{d_i} \left[\frac{c_i + h_i}{\int_{S_i}^{\infty} (1 - S_i/x) f_i(x) dx} - h_i \right] - 1 = 0, \\ d_i \int_{S_i}^{\infty} \frac{(x - S_i)^2}{x} f_i(x) dx - d_1 \int_{S_1}^{\infty} \frac{(x - S_1)^2}{x} f_1(x) dx = 0, \quad i = \overline{2, N}.$$

Наконец, при исчислении затрат на хранение по избыточным запасам, а штрафов по взвешенным вероятностям недостачи имеем

$$L(\vec{S}) = \sum_{i=1}^N \left\{ c_i (S_i - z_i) + h_i \left[\int_0^{S_i} (S_i - x) f_i(x) dx \right] + \max_i d_i \int_{S_i}^{\infty} f_i(x) dx \right.$$

и условия оптимальности

$$\sum_{i=1}^N \frac{c_i + h_i \left(1 - \int_{S_i}^{\infty} f_i(x) dx \right)}{d_i f_i(S_i)} - 1 = 0, \\ d_i \int_{S_i}^{\infty} f_i(x) dx - d_1 \int_{S_1}^{\infty} f_1(x) dx = 0, \quad i = \overline{2, N}.$$

Начальные приближения во всех случаях естественно определить независимой минимизацией затрат по отдельным номенклатурам

Изложенные результаты применяются к статической модели. Однако для вариантов задачи с переносом дефицита они применимы и в многопериодной схеме — с уже обсуждавшейся заменой c на $c(1 - \alpha)$

6.5.2. Выбираемая периодичность

В разд 6.2 было показано, что при детерминированном спросе наивыгоднейшие периоды снабжения для различных номенклатур не совпадают. Можно полагать, что в случае вероятностного спроса наилучший результат также достигается при организации восполнения по системе кратных периодов. Описанный ниже *приближенный* метод заключается в расчленении задачи на условно независимые этапы:

- 1) Случайный спрос по всем номенклатурам заменяется детерминированным с той же средней интенсивностью. Далее с помощью алгоритма разд 6.2 рассчитывается система кратных периодов снабжения, т.е. оптимальный базисный период T и набор чисел $\{k_i\}$. Полученные периоды считаются окончательными и дальнейшей корректировке не подвергаются. Среднее значение и дисперсия спроса принимаются пропорциональными $\{k_i T\}$, после чего выбираются аппроксимирующие распределения

2) Составляется функция дополнительных затрат, возникающих из-за случайных отклонений спроса от его среднего значения, за период KT , где K есть общее наименьшее кратное чисел $\{k_i\}$. Для этого

- вычисляются затраты на хранение избытков для i -й номенклатуры за период $k_i T$ и умножаются на K/k_i — число периодов снабжения по данной номенклатуре, уложившихся в отрезок KT ;
- эти затраты суммируются по всем номенклатурам,
- вычисляются ожидаемые штрафы как произведение на K штрафов по первой номенклатуре за базовый период;
- результаты суммируются и делятся на KT .

В итоге приходим к целевой функции

$$L = \sum_{i=1}^N h_i \int_0^{S_i} (S_i - x) f_i(x) dx + \frac{1}{T} d_1 \int_{S_1}^{\infty} (x - S_1) f_1(x) dx, \quad (6.5.6)$$

минимизация которой проводится уже рассмотренным способом

Второй этап алгоритма может быть реализован и с другими вариантами построения вероятностной составляющей функции затрат. Напомним, что распределения спроса и цены хранения в этой формуле берутся для интервалов $\{k_i T\}$

6.6. Зонная стратегия

В разд. 5.8 был описан приближенный способ расчета оптимальных параметров стратегии (s, q) для однородной задачи и малой задержки поставок. Для многономенклатурной проблемы эта стратегия неудобна тем, что она из-за одновременного достижения критических уровней по разным предметам снабжения не допускает совмещения заказов. В «зонном» варианте пороговой стратегии заказ на восполнение выдается, если хотя бы для одной номенклатуры j выполняется условие $z_j \leq s'_j$, а объемы заказа определяются для всех номенклатур i согласно

$$v_i = \begin{cases} S_i - z_i, & \text{если } z_i \leq s''_i; \\ 0 & \text{иначе} \end{cases}$$

Таким образом, для каждой номенклатуры существует зона заказа $[s'_i, s''_i]$ и верхний порог запаса S_i .

При большом числе планируемых номенклатур каждая из них, как правило, будет заказываться при снижении запаса z_i до s''_i (цена заказа g_i) и лишь в редких случаях — при снижении z_i до s'_i (цена заказа $g_0 + g_i$). Отсюда вытекает следующий приближенный алгоритм расчета оптимальной зонной стратегии для $i = \overline{1, N}$:

- 1) Решением соответствующей системы уравнений из разд. 5.8 с дополнительным индексом номенклатуры i и при цене заказа g_i определить оптимальные пороги $\{s''_i\}$ и объемы заказов $\{q''_i\}$.
- 2) Вычислить оптимальные запасы по формуле

$$S_i = s''_i + q''_i. \quad (6.6.1)$$

- 3) Согласно методике для однородного случая определить нижние пороги $\{s'_i\}$ и объемы поставок $\{q'_i\}$ при цене заказа $G_i = g_0 + g_i$ и ограничении

$$s'_i + q'_i = S_i. \quad (6.6.2)$$

Способ расчета s''_i, q''_i и S_i дополнительных пояснений не требует.

Для нахождения s'_i запишем функцию затрат в единицу времени

$$L = h(q'/2 + s' - \lambda \bar{r}) + \frac{\lambda}{q'} [G + \pi(s')], \quad (6.6.3)$$

где через $\pi(s')$ обозначено математическое ожидание штрафа за цикл при пороговом запасе s' , а индексы i для краткости опущены. Подставив в это выражение найденное из (6.6.2) значение q , получаем

$$L = h \left(\frac{S + s'}{2} - \lambda \bar{r} \right) + \frac{\lambda}{S - s'} [G + \pi(s')]. \quad (6.6.4)$$

Поскольку максимальный запас S уже известен, величина затрат оказывается функцией только одной переменной s' . Дифференцируя правую часть уравнения (6.6.4) по s' , имеем

$$\frac{dL}{ds'} = \frac{h}{2} + \frac{\lambda}{(S - s')^2} [\pi'(s')(S - s') + G + \pi(s')].$$

Приравняв результат нулю, получаем уравнение для оптимального s'

$$-\frac{h}{2\lambda}(S-s')^2 = \pi'(s')(S-s') + G + \pi(s'). \quad (6.6.5)$$

Дальнейшие преобразования этого уравнения проводятся для конкретных вариантов определения $\pi(s')$ с целью получения итерационной формулы вида $s' = \varphi(s')$. В качестве начального приближения разумно выбирать $s'_0 = s''$.

6.6.1. Показательно распределенный спрос

Для показательно распределенного спроса со средним значением за время задержки поставки $\bar{x} = 1/\mu$ и штрафом по ожидаемому дефициту можно получить выражения

$$\begin{aligned} \pi(s) &= d\bar{x}e^{-\mu s}, \\ \pi'(s) &= -de^{-\mu s}. \end{aligned}$$

Подстановка их в уравнение (6.6.5) для нижнего порога дает

$$-\frac{h}{2\lambda}(S-s')^2 = -de^{-\mu s'}(S-s') + G + d\bar{x}e^{-\mu s'},$$

или

$$de^{-\mu s'}[\bar{x} - (S-s')] = -G - \frac{h}{2\lambda}(S-s')^2.$$

Отсюда с учетом $\bar{x} = 1/\mu$ выводим итерационную формулу

$$s' = \bar{x} \ln \frac{d \cdot (S - \bar{x} - s')}{G + \frac{h}{2\lambda}(S - s')^2}. \quad (6.6.6)$$

6.6.2. Дискретно распределенный спрос

В рассмотренном выше (разд. 3.6.3) случае простейшего потока заявок и показательно распределенной задержки поставок при исчислении штрафа по ожидаемому дефициту уравнение (6.6.5) обращается в

$$-\frac{h}{2\lambda}(S-s')^2 = d \frac{\lambda}{\mu} \left(\frac{\lambda}{\lambda+\mu} \right)^{s'} \ln \frac{\lambda}{\lambda+\mu} \times (S-s') + G + d \frac{\lambda}{\mu} \left(\frac{\lambda}{\lambda+\mu} \right)^{s'}.$$

После элементарных преобразований получаем расчетную формулу

$$s' = \frac{\ln \frac{d \frac{\lambda}{\mu} [(S-s') \ln(1+\mu/\lambda) - 1]}{\frac{h}{2\lambda}(S-s')^2 + G}}{\ln(1+\mu/\lambda)}. \quad (6.6.7)$$

Применим эту методику к задаче расчета организации поставок по 10 номенклатурам при цене штрафа 500 руб. за недостачу единицы запаса в год, средней задержке поставок $\tau = 0.2$ года и стоимости поставок $g_0 = 100$ руб. Остальные исходные данные сведены в табл. 6.2. Все цены выражены в рублях, цены хранения и интенсивности заявок даны в расчете на год

Т а б л и ц а 6.2

К расчету зонной стратегии

i	λ_i	h_i	g_i	s'_i	s''_i	S_i
1	20	10	10	15	19	28
2	30	5	5	25	33	44
3	15	8	6	11	15	22
4	12	5	4	10	13	20
5	10	4	5	9	12	19
6	8	5	6	7	9	15
7	6	8	4	5	7	11
8	4	50	5	2	2	5
9	20	6	4	16	22	30
10	15	8	10	12	15	23

В эту же таблицу включены оптимальные уровни $\{s'_i, s''_i, S_i\}$. Они вполне соответствуют интуитивным представлениям о росте соответствующих порогов при увеличении интенсивности отказов и стоимости заказа и уменьшении — с ростом цены хранения (при прочих равных условиях). Программа расчета представлена в Приложении Б.

Согласно [190], уменьшение затрат в примерах использования зонных пороговых стратегий достигало 15–20%.

При исчислении штрафа по вероятности дефицита, выполняя соответствующие подстановки в (6.6.5), находим

$$-\frac{h}{2\lambda}(S-s')^2 = d \left(\frac{\lambda}{\lambda+\mu} \right)^{s'+1} \ln \frac{\lambda}{\lambda+\mu} \times (S-s') + G + d \left(\frac{\lambda}{\lambda+\mu} \right)^{s'+1},$$

откуда следует

$$s' = \frac{\ln \frac{d[(S - s') \ln(1 + \mu/\lambda) - 1]}{\frac{h}{2\lambda}(S - s')^2 + G}}{\ln(1 + \mu/\lambda)} - 1. \quad (6.6.8)$$

Порядок расчета s''_i и q''_i не меняется.

Глава 7

Системы снабжения

7.1. Общие положения

Рассмотренные выше случаи относились к управлению запасами на изолированном складе (точнее, тройке «потребитель — склад — поставщик»). Однако часто встречаются ситуации, в которых отдельные или все элементы этой тройки являются множествами: обеспечение *группы* потребителей, через *систему складов*, от *нескольких* поставщиков

Реальная система снабжения, как правило, включает в себя целую сеть связанных между собою складов нескольких ступеней, имеющих различный объем и специализацию хранения. Условия работы отдельных звеньев ее оказываются зависящими от состояния запасов в других звеньях и от структуры системы в целом. Типична *иерархическая* система снабжения, в которой продукт, проходя путь от внешнего поставщика к потребителю, может накапливаться на нескольких складах (например, склады предприятия-изготовителя, оптовой базы, магазина). В *централизованной* системе для формирования заказов на поставку в каждый склад допускается использование информации о состоянии запасов на любых складах, входящих в систему, эта информация должна поступать в единый центр принятия решений, где и формируются заказы поставщикам. В иерархической системе верхние уровни осуществляют управляющие воздействия более редко, а сами воздействия оказывают влияние на более длительные сроки

Оптимизация работы системы складов требует более широкого подхода — поиска минимума затрат с точки зрения системы *в целом*. Потери от пролеживания продуктов труда несравнимо меньше, чем потери от дефицита, которые приводят к простоям оборудования, рабочей силы и сокращению выпуска продукции. Поэтому предприятия идут на превышение производственных запасов над общественно необходимым уровнем. Этот запас не имеет достаточной маневренности и распылен по десяткам тысяч предприятий. При любых изменениях потребности часть его становится ненужной предприятию, возрастают убытки от пролеживания. Введением общего страхового запаса вместо индивидуальных для групп складов и объединением поставок нескольким потребителям вплоть до низшего звена (потребителей) удастся добиться дополнительного снижения затрат по сравнению с описанными выше независимыми решениями. Величина экономии определяется масштабами системы и ролью случайных факторов и возрастает с ростом числа складов и вариации спроса. Кроме того, она существенно зависит от организации системы снабжения и способа размещения резервного запаса.

По всем разумным показателям (среднему числу простоев, среднему времени простоя, потерям производительности из-за простоев и т.п.) централизованное обслуживание обладает преимуществом. При неучете информационных аспектов процесса децентрализация может только ухудшить качество работы, поскольку она ограничивает класс допустимых стратегий. Вообще же должны учитываться расходы на обработку информации и повышение быстродействия (свойство прямых связей). Удобно считать оптимальным такое правило принятия решений в каждом локальном центре, при котором минимизируются затраты на хранение запаса только в ассоциированных с этим центром складах (при фиксированной гарантии удовлетворения внешнего для этих складов спроса).

Поскольку в условиях крупномасштабного производства или большой системы снабжения имеется много переменных и управляющих воздействий, можно предположить, что система управления будет строиться с применением принципов декомпозиции и координации. По одним каналам система будет полностью автоматической, по другим — выступать в роли советчика, по третьим ограничится сбором и упорядочением информации. Координирующую роль играют ограничения и иерархия целевых функций. В типичном случае целевые функции верхних звеньев формируют ограничения для нижних.

7.2. Факты и тенденции

Основанием для анализа систем снабжения является колоссальная стоимость содержащихся в них запасов — см., например, заимствованную из [27] табл. 7.1.

Т а б л и ц а 7.1

Оборотные средства в запасах СССР, млрд. руб.

Отрасли хозяйства	1966 г.	1968 г.	1970 г.	1972 г.	1974 г.
Промышленность	47.06	56.35	63.28	69.18	77.51
Сельское хозяйство	11.72	14.40	18.50	21.42	25.38
Транспорт и связь	1.65	1.93	2.56	2.82	3.14
Строительство	5.24	8.13	10.69	18.35	28.04
Заготовки	9.23	10.14	9.20	9.06	11.74
Снабжение и сбыт	5.52	8.30	9.31	10.38	11.55
Торговля	29.00	37.42	43.74	49.14	54.81
Прочие	3.33	5.10	6.08	6.83	8.84
Итого	112.75	141.77	163.36	187.17	221.02

Традиционной и во многих отношениях предпочтительной формой снабжения считались прямые связи поставщик — потребитель («транзитные поставки»). Прямые связи

- ускоряют реакцию изготовителей на новые требования к качеству и ассортименту продукции,
- обеспечивают ритмичность поставок,
- снижают число переналадок оборудования и брак,
- повышают качество конечной продукции. дают экономию сырья и материалов.

Однако чрезмерное использование транзита приводит к завышению потребности тех заказчиков, у которых суммарная потребность в планируемом периоде меньше транзитной нормы, что вызывает искусственный дефицит. Заметим, что грузоподъемность железнодорожных вагонов за последние годы возросла в 6 раз (с 20 до 129 тонн), контейнеров — в 4 раза (с 5 до 20 тонн), автотранспорта в 2.5 раза, речных судов в 5 раз (с 1 до 5 тыс. тонн). Кроме того, заказчики транзитных поставок из-за

ограниченности оборотных средств могут остаться без достаточного запаса по некоторым другим номенклатурам. На динамике запасов прямые связи сказываются отрицательно (велики транзитные нормы → редки поставки → велики максимальный и средний уровни запаса).

За складскую форму снабжения можно привести следующие аргументы:

- совмещаются поставки для групп потребителей;
- уменьшается время оперативной поставки для ликвидации дефицита;
- в связи с увеличением оборота с каждым поставщиком продукция приобретает не по розничным, а по оптовым ценам (возможно, с дополнительными скидками);
- уменьшается совокупный страховой запас;
- снимается проблема транзитной нормы поставки;
- относительно дешево выполняется дополнительная обработка продукции с учетом нужд конкретного потребителя (нарезка и раскрой металлопродукции и древесины, профилирование, перфорирование, сварка, окраска, нанесение виниловых покрытий, придание квадратной или шестигранной формы и т.п.).

Для иллюстрации последнего пункта этого перечня укажем, что удельный вес продукции, подвергающейся подобной обработке, в США составляет 75% (стоимость услуг порядка 18 млрд. долларов), в ФРГ около 30%.

Отмеченные обстоятельства определили долгосрочную тенденцию увеличения абсолютного объема и доли складских поставок и соответствующую реорганизацию систем снабжения. Структура оптового товарооборота средств производства на 1970 г., по данным [81], показана в табл. 7.2.

В те годы в стране действовало несколько складских систем материально-технического снабжения. Еще в 1961 г. было организовано Всесоюзное объединение «Сельхозтехника» по продаже минеральных удобрений, сельскохозяйственной техники, запасных частей и ремонту этой техники. В 1965 г. был создан Госнаб СССР, который впервые

Т а б л и ц а 7.2

Товарооборот средств производства

Организационная форма	Всего, млрд. руб.	Из них складской	
		в млрд. руб.	в %
Госнаб СССР	132.2	25.4	19.2
Ведомственные системы	70.3	9.2	13.1
«Сельхозтехника»	14.0	11.2	80.1
Местные системы	9.6	2.3	24.0
Итого	226.1	48.1	21.3

в истории народного хозяйства объединил в самостоятельную отрасль снабжение и сбыт для обеспечения материальными ресурсами потребителей независимо от их ведомственной принадлежности. Органы Госнаба обеспечивали около 60% общего товарооборота средств производства. Кроме того, были сохранены системы снабжения предприятий и организаций на транспорте, в связи, энергетике, газовой промышленности и ряде других министерств, а также при исполкомах местных Советов депутатов трудящихся.

Всего в 1985 г. на складах предприятий и систем снабжения находилось оборотных средств в запасах на сумму свыше 300 млрд. руб. За 1970–1986 гг. валовый общественный продукт увеличился вдвое, а объем запасов — в 2.83 раза. В строительстве за те же годы объем строительно-монтажных работ возрос в 1.53 раза, а оборотные средства в запасах — в 7.6 раза. От перемещения в систему Госнаба запасов на 12.4 млрд. руб. ожидалось снижение совокупного запаса средств производства на 28.4 млрд. руб. [28].

В 1980-х гг. основные фонды складского хозяйства снабженческо-сбытовых органов оценивались в 55 млрд. руб. Общая стоимость только тарного хозяйства составляла 10 млрд. руб.: свыше 1 млрд. ящиков, около 60 млн. бочек и барабанов, 1 млрд. мешков.

Совершенствование размещения запасов с необходимостью предполагает перспективное планирование производственной и складской структуры системы, совершенствование прямых долговременных хозяйственных связей. Любые изменения структуры совокупного запаса не могут быть осуществлены без предварительного изменения отраслевой структуры вложений. Так, увеличение доли складского снабжения

и, следовательно, запаса в снабженческо-сбытовом звене прежде всего требует дополнительных капитальных вложений в новое строительство или в расширение складских помещений, увеличение парка подъемно-транспортных механизмов, строительство подъездных путей, оснащение оргтехникой, количественное и качественное развитие тарного хозяйства. В связи с последним направлением отметим, что

- себестоимость картонного ящика в 2.2 раза меньше, чем деревянного,
- 1 т картона сберегает 14 кубометров круглого леса,
- в сложенном виде картонные ящики занимают вдесятеро меньше места, чем порожние деревянные той же емкости

Как наиболее прогрессивные варианты складской формы снабжения отметим гарантированное снабжение и оптовую торговлю *Гарантированное снабжение* — это обеспечение поставок по согласованным графикам. Снабженческо-сбытовой организации передается до 50% производственных запасов, что требует высокой степени доверия и взаимной ответственности сторон. В 1980–1985 гг. объем таких поставок составлял около 10 млрд. руб., а число охваченных ими предприятий увеличилось с 7 тыс. до 8600. В большинстве районов страны оплата услуг по гарантированному снабжению составляла 0.2% от стоимости поставок или оформлялась отдельным договором на основе согласованной калькуляции [28]. По [18], в Чехословакии при комплексном обслуживании база устанавливала наценку в размере 5–6% вместо обычных четырех

Экономичность гарантированного снабжения привела к тому, что в Госснаб от 22 министерств поступили предложения принять на снабжение их предприятия. В Ленинграде все НИИ и КБ, а также ряд предприятий были переведены на гарантированное комплексное снабжение. В результате на 1 июня 1970 г. сверхнормативные запасы продукции по объединению им. Свердлова снизились на 25%, по заводу им. Котлякова на 39%, по «Электросиле» — на 61%.

При внедрении складской формы снабжения полномочия и ответственность передавались все ниже. Это вызывало противодействие работников среднего звена, которые боялись потери контроля над ресурсами, увеличения объема и ответственности работы. Дальнейшим шагом явилась *оптовая торговля*, поставившая заслон волюнтаристским решениям в распределении материальных ресурсов. Работники снабжения не были заинтересованы в организации оптовой торговли, мелкие отгрузки в которой требовали почти тех же затрат на оформление, что и при

крупных партиях. Эта незаинтересованность преодолевалась взиманием дополнительной наценки за услуги мелкооптовой торговли. Подчеркнем, что запасы в розничной торговле утрачивают свою маневренность, а расходы на их хранение выше (в 1980-е гг. в розничном звене они составляли 3.27%, а в оптовом 2.32%). Кроме того, на оптовых базах возможны более полная загрузка персонала, его специализация, механизация учетных и иных операций.

В основных направлениях экономического и социального развития СССР на 1986–1990 гг. и на период до 2000 года была поставлена задача «развивать и совершенствовать общегосударственную систему материально-технического снабжения, повысить ответственность Госснаба СССР, органов снабжения за соблюдение дисциплины поставок, обеспечение экономного и рационального использования топлива, сырья, материалов, вторичных ресурсов», перейти к плановому распределению оборудования, материалов и полуфабрикатов путем оптовой торговли. Введение в 11-й пятилетке в органах Госснаба пакетно-контейнерного способа перевозки с учетом централизованной доставки грузов позволило повысить производительность труда на погрузо-разгрузочных работах в 4–5 раз; увеличить пропускную способность грузовых фронтов вдвое; сократить простой железнодорожного транспорта в 4–5 раз, а автомобильного — в 5–6 раз.

Переход к рыночной экономике повышает актуальность проблем оптовой торговли, учета скидок, ускорения оборачиваемости запасов и надежности поставок, не отменяя целесообразности вышеперечисленных мер для хозяйственных структур общероссийского и регионального масштаба. Инфляционные ожидания являются аргументом в пользу увеличения запасов.

7.3. Детерминированный спрос

Попытки построения и строгого анализа моделей ветвящихся систем снабжения наталкиваются на значительные аналитические трудности, а полученные результаты основаны на тех или иных допущениях. Таким образом, повышение точности описания реальной ситуации сопровождается снижением математической точности получаемых результатов. Простота модели позволяет не только заметно продвинуться в ее

аналитическом или численном исследовании, но и существенно ослабить требования к информационному обеспечению решения практических задач. Качественное понимание характера протекающих процессов может стать хорошим ориентиром при решении более сложных задач. Разумеется, наиболее существенные факторы должны быть учтены. С другой стороны, отметим, что на современном этапе развития вычислительной техники применительно к «непрограммирующим пользователям» требование простоты расчетных методик может быть ослаблено (но это предполагает централизованную профессиональную разработку математического обеспечения систем управления запасами).

В разных звеньях сложной складской системы возможно использование разных стратегий.

К синтезу систем снабжения возможны различные подходы. Прежде всего отметим, что относительно просто решается проблема локализации ее звеньев. Положение центрального склада может быть известно заранее по внешним условиям. То же иногда можно сказать о промежуточных складах, которые естественно привязываются к узлам транспортной сети. Если такой привязки нет, то для определения места промежуточных складов следует провести окружности с центрами в местах скопления потребителей и радиусом, определяемым временными ограничениями (задержка поставок). Промежуточные склады размещаются в зоне их пересечения. Одновременно определится и число промежуточных складов (по количеству зон пересечения).

Описание техники экономической оценки вариантов структуры мы начнем с простейшей модели иерархической системы снабжения при детерминированном спросе. Экономический эффект объединения складов в систему в этом случае создается совмещением поставок для групп потребителей, в результате чего уменьшается число прямых связей потребителей с поставщиками и сокращается общее число перевозок в системе.

Оценим величину этого выигрыша для двухкаскадной системы, состоящей из m низовых складов, одного центрального и n поставщиков. Для простоты и наглядности предположим, что каждая из номенклатур производится только одним из поставщиков. Пусть, далее,

M_i — множество номенклатур, поставляемых i -м поставщиком, $i = \overline{1, n}$,

Q_j — множество номенклатур, потребляемых j -м низовым складом, $j = \overline{1, m}$,

λ_{rj} — спрос на номенклатуру r по j -му низовому складу, $r \in Q_j$,

g_{ij} — постоянная составляющая стоимости перевозок по маршруту $i-j$,

g_i — постоянная составляющая стоимости перевозок от i -го поставщика, $i = \overline{1, n}$, к центральному складу,

g_{0j} — постоянная составляющая стоимости перевозок от центрального склада к низовому с номером j , $j = \overline{1, m}$,

h_r — стоимость хранения единицы номенклатуры r в единицу времени (независимо от места хранения).

Считая долю γ стоимости заказа, приходящуюся на поставку каждой дополнительной номенклатуры, не зависящей от r , i и j , для затрат на поставки в условиях *прямых* связей $i-j$ согласно разд. 6.2 имеем

$$L_1 = \sum_{j=1}^m \sum_{i=1}^n \sqrt{2g_{ij} \left(\gamma \sum_{r \in Q_j \cap M_i} 1/k_{rj} + 1 \right) \sum_{r \in Q_j \cap M_i} \lambda_{rj} h_r k_{rj}}. \quad (7.3.1)$$

Здесь суммирование под корнем происходит по пересечению множеств Q_j и M_i , т.е. по тем номенклатурам, которые производятся i -м поставщиком и потребляются j -м складом, а коэффициенты кратности периодов поставок $\{k_{rj}\}$ определяются по методике разд. 6.2.

При поставках через *центральный* склад производится независимая минимизация затрат на обоих уровнях системы. Расходы на обеспечение низовых звеньев могут быть вычислены как

$$L_2 = \sum_{j=1}^m \sqrt{2g_{0j} \left(\gamma \sum_{r \in Q_j} 1/k_{rj} + 1 \right) \sum_{r \in Q_j} \lambda_{rj} h_r k_{rj}}. \quad (7.3.2)$$

Центральный склад суммирует спрос низовых звеньев по каждой из номенклатур. Его затраты на обеспечение этого спроса

$$L_3 = \sum_{i=1}^n \sqrt{2g_i \left(\gamma \sum_{r \in M_i} 1/k_r + 1 \right) \sum_{r \in M_i} h_r k_r \sum_{j=1}^m \lambda_{rj}}. \quad (7.3.3)$$

При такой организации поставок суммарные затраты $L = L_2 + L_3$ должны сравниваться с L_1 . Кроме того, следует учесть дополнительный эффект: в складской форме поставок перевозки более крупных партий могут осуществляться более дешевым транспортом (например, речным или железнодорожным вместо автомобильного), возможно удешевление хранения и организации отгрузок. Такие сопоставления должны проводиться в процессе проектирования систем снабжения для определения целесообразности промежуточных звеньев.

Обобщения предложенной расчетной схемы на случаи переменных $\{\gamma_{ij}\}$, зависящих от места хранения цен $\{h_{rj}\}$, а также большее число каскадов вполне очевидны. Сразу же укажем, что стоимость хранения у поставщиков определяется себестоимостью произведенной продукции, на оптовых базах — оптовой ценой, а на нижнем уровне системы — розничной. Это соображение становится дополнительным аргументом за «сдвиг» запасов в сторону высших звеньев.

Следует отметить, что в системах снабжения, построенных по второму варианту (через промежуточные склады), расчет для высших звеньев указывает лишь *среднюю* периодичность заказов по каждой номенклатуре. Для обеспечения потребности низовых звеньев, периодичность и объемы поставок по которым будут известны точно, придется внести некоторую неравномерность в объемы поставок, поступающих в центральный склад: поправка должна быть равна разности между фактическим объемом поставок в низовые склады за интервал $k_r T$, определяемый совмещением календарных графиков обеспечения низовых складов, и их средним объемом $k_r T \sum_{j=1}^m \lambda_{rj}$. Относительная величина этой поправки должна убывать с ростом числа низовых складов.

Пример. Оценим сравнительные затраты по вариантам организации снабжения для обсуждавшейся выше системы в предположении, что каждый поставщик производит только одну номенклатуру, а каждый низовой склад потребляет их все. В этом случае индекс r может быть отождествлен с i , а расчетные формулы упрощаются и принимают вид

$$L_1 = \sum_{j=1}^m \sum_{i=1}^n \sqrt{2g_{ij}(1+\gamma)\lambda_{ij}h_i} \quad (7.3.4)$$

(независимые поставки по всем номенклатурам и низовым складам),

$$L_2 = \sum_{j=1}^m \sqrt{2g_{0j} \left(\gamma \sum_{i=1}^n 1/k_{ij} + 1 \right) \sum_{i=1}^n \lambda_{ij} h_i k_{ij}} \quad (7.3.5)$$

(многономенклатурные поставки из центрального склада в низовые),

$$L_3 = \sum_{i=1}^n \sqrt{2g_i(1+\gamma)h_i \sum_{j=1}^m \lambda_{ij}} \quad (7.3.6)$$

(независимые поставки по каждой из номенклатур в центральный склад). Пусть $\gamma = 0.20$, стоимости хранения $\{h_i\}$ и спрос $\{\lambda_{ij}\}$ заданы таблицей 7.3, а стоимости поставок пропорциональны геометрическим расстояниям между соответствующими пунктами и заданы табл. 7.4.

Т а б л и ц а 7.3

Спрос и стоимости хранения

i	Низовые склады								Центр	h_i
	1	2	3	4	5	6	7	8		
1	100	80	40	50	60	30	40	50	450	0.08
2	350	140	270	20	50	40	200	160	1210	0.40
3	240	60	40	30	70	60	120	50	670	0.15
4	180	70	150	20	80	130	250	80	960	0.20
5	80	300	120	250	140	200	20	300	1410	0.25
6	20	50	90	80	200	100	40	120	700	0.15

Т а б л и ц а 7.4

Стоимости перевозок

Поставщик	Получатель								
	1	2	3	4	5	6	7	8	центр
1	1.51	1.82	2.05	1.49	2.08	1.83	2.26	2.41	1.88
2	1.50	1.80	2.11	1.33	2.01	1.61	2.05	2.28	1.74
3	1.91	2.04	2.40	1.55	2.07	1.50	1.74	2.10	1.77
4	2.46	2.25	1.90	2.75	2.22	2.80	2.62	2.26	2.50
5	2.20	1.92	1.60	2.45	1.77	2.32	2.05	1.68	2.08
6	2.33	2.08	2.09	2.26	1.80	1.92	1.49	1.43	1.90
Центр	0.44	0.32	0.64	0.41	0.32	0.28	0.44	0.54	0.00

Подстановка указанных величин дает для прямых связей поставщик — низовой склад $L_1 = 471$.

Стоимость поставок в центральный склад составляет $L_3 = 175.5$. Применяя методику разд. 6.2 к расчету оптимальных периодичностей поставок из центрального склада в низовые, получаем $L_{21} = 21.1$, $L_{22} = 15.2$, $L_{23} = 22.6$, $L_{24} = 12.9$, $L_{25} = 12.8$, $L_{26} = 11.9$, $L_{27} = 17.1$, $L_{28} = 20.6$, откуда $L_2 = 134.2$. Общая стоимость организации снабжения по складскому варианту составляет 309.7, т.е. оказалась на треть ниже, чем для транзитных поставок.

Некоторое упрощение изложенной методики путем замены системы кратных периодов снабжения *общим* периодом по всем номенклатурам, поставки которых можно объединить, согласно (6.2.3) приводит к формулам

$$L_1 = \sum_{j=1}^m \sum_{i=1}^n \sqrt{2g_{ij}(1 + \gamma|Q_j \cap M_i|)} \sum_{r \in Q_j \cap M_i} \lambda_{rj} h_r, \quad (7.3.7)$$

$$L_2 = \sum_{j=1}^m \sqrt{2g_{0j}(1 + \gamma|Q_j|)} \sum_{r \in Q_j} \lambda_{rj} h_r, \quad (7.3.8)$$

$$L_3 = \sum_{i=1}^n \sqrt{2g_i(1 + \gamma|M_i|)} \sum_{r \in M_i} h_r \sum_{j=1}^m \lambda_{rj}, \quad (7.3.9)$$

в которых под $|M|$ подразумевается число элементов множества M . Результаты разд. 6.2 позволяют предполагать сравнительно близкое со-

впадение затрат с полученными по исходным зависимостям, что подтверждается расчетом для приведенного выше примера. В указанном случае формула для L_2 преобразуется в

$$L_2 = \sum_{j=1}^m \sqrt{2g_{0j}(1 + \gamma n) \sum_{i=1}^n \lambda_{ij} h_i}, \quad (7.3.10)$$

а остальные сохраняются. Подсчет затрат на снабжение низовых складов дает $L_{21} = 21.6$, $L_{22} = 15.3$, $L_{23} = 23.0$, $L_{24} = 13.1$, $L_{25} = 12.7$, $L_{26} = 12.1$, $L_{27} = 17.7$, $L_{28} = 20.9$, так что $L_2 = 136.4$. Общая стоимость снабжения по второму варианту составит 311.9, что всего на 0.7% больше полученной точным способом.

Окончательное решение о выборе варианта организации системы складов должно приниматься с учетом стоимости самой системы — точнее, с учетом тех затрат, которые *не включены* в стоимости хранения и поставок, найденные согласно рекомендациям главы 2.

7.4. Синтез системы при детерминированном спросе

Результаты данного раздела применимы в тех случаях, когда необходимо снабжать обширный район с приблизительно равномерным распределением однородных потребителей, т.е. с постоянным спросом с единицы площади. Учитывая, что расходы на поставки через кратные периоды можно удовлетворительно аппроксимировать затратами в модели с общим периодом, мы будем задавать плотность суммарного стоимостного спроса в виде $\lambda h = \sum_r \lambda_r h_r$ на единицу площади.

Предположим, что

- склады всех рангов (ступеней) равномерно распределены в районе дислокации потребителей;
- все склады одного ранга обслуживают одинаковые площади (районы);
- расходы $g = a + bd$ на поставку партии из одного склада в другой являются линейной функцией расстояния d между ними и не зависят от объема перевозок;

- все районы обслуживания аппроксимируются кругами.

Для определенности рассмотрим симметричную систему, состоящую из одного центрального склада (III), n_2 складов второго ранга (II) и $n_1 \cdot n_2$ складов первого ранга (I). Пусть центральный склад снабжается с расстояния l внешним поставщиком, а площадь обслуживаемого им района равна S . Если S — площадь круга, то его радиус $r = \sqrt{S/\pi}$. Среднее расстояние точки внутри круга от его центра

$$\bar{\rho} = \frac{1}{\pi R^2} \int_0^R \rho \cdot 2\pi\rho d\rho = \frac{2}{3}R.$$

Следовательно, среднее расстояние от центрального склада до склада II

$$d_{3-2} = \frac{2}{3} \sqrt{\frac{S}{\pi}}.$$

Таким же в среднем будет и расстояние d_{3-1} . Среднее расстояние от склада II до склада I

$$d_{2-1} = \frac{2}{3} \sqrt{\frac{S}{n_2 \cdot \pi}}.$$

Пусть интервал между поставками из промышленности равен T . За это время средние затраты на хранение составят $\lambda h S T^2 / 2$ независимо от распределения запаса внутри системы. Каждая поставка в склад III связана с фиксированным расходом $a + bl$. Затраты в единицу времени составят

$$L = \frac{\lambda h S T}{2} + \frac{a + bl}{T}. \quad (7.4.1)$$

Распределение прибывшего пополнения должно минимизировать расходы на перевозки внутри системы в единицу времени. Следовательно, последние должны осуществляться возможно реже и во всяком случае с интервалом не менее T . Верхняя граница этого интервала определяется необходимостью отправлять партию хотя бы одному потребителю при каждом пополнении и, следовательно, равна $n_2 T$ для складов II и $n_1 n_2 T$ для склада I. Конкретные условия могут потребовать уменьшения этих интервалов.

Рассмотрим три возможных способа распределения пополнения:

- при каждой поставке удовлетворять только одного из потребителей (поочередно);

- распределять каждое пополнение поровну между всеми потребителями
- при каждой поставке поочередно удовлетворять группы потребителей

Третий вариант в зависимости от числа групп ν образует различные промежуточные случаи между двумя первыми. Для определенности мы будем считать $\nu = 2$. Поскольку спрос в системе должен удовлетворяться полностью, часть расходов пропорциональная объему поставок для каждого потребителя, изменена быть не может. Однако рациональная организация перевозок позволяет сократить их число и уменьшить среднюю протяженность. Обозначим через g_Σ сумму всех фиксированных транспортных расходов при данной организации перевозок. Тогда оптимальная периодичность поставок и соответствующий ей минимум расходов на снабжение могут быть найдены по формулам

$$T^* = \sqrt{2g_\Sigma / (S\lambda h)}, \quad (7.4.2)$$

$$L^* = \sqrt{2g_\Sigma S\lambda h} \quad (7.4.3)$$

Все величины входящие в эти формулы кроме g_Σ заданы. Следовательно, суммарные расходы пропорциональны $\sqrt{g_\Sigma}$ и для выбора организации снабжения достаточно сравнить значения g_Σ полученные при различных ее вариантах. Мы найдем g_Σ для всех перечисленных выше вариантов в двух- и трехкаскадной системе, а также условия, при которых введение промежуточных складов приводит к уменьшению расходов. В ходе рассуждений для упрощения выкладок примем $n_1 = n_2 = n$ что приближенно выполняется в реальных системах снабжения.

Поочередное снабжение

В течение одного периода поставка, прибывшая в центральный склад, целиком переадресуется одному из складов уровня II, последний же направляет ее в какой-нибудь из подчиненных ему низовых складов. Первая из этих перевозок связана с затратами $a + \frac{2}{3}b\sqrt{S/\pi}$, а вторая — с затратами $a + \frac{2}{3}b\sqrt{S/(n\pi)}$. С учетом стоимости внешней поставки $a + bl$ имеем

$$g_\Sigma = 3a + b \left[l + \frac{2}{3} \sqrt{\frac{S}{\pi}} \left(1 + \frac{1}{\sqrt{n}} \right) \right]$$

При непосредственных поставках из центрального склада в низовые получим

$$g_{\Sigma} = 2a + b \left(l + \frac{2}{3} \sqrt{\frac{S}{\pi}} \right)$$

Очевидно, что при поочередной поставке промежуточные склады не нужны

Равномерное распределение поставок

В этом случае перевозок первого типа будет n , а второго типа — n^2 . Суммирование расходов дает

$$g_{\Sigma} = a(1 + n + n^2) + b \left[l + \frac{2}{3} n \sqrt{\frac{S}{\pi}} + \frac{2}{3} n^2 \sqrt{\frac{S}{n\pi}} \right],$$

или

$$g_{\Sigma} = a(1 + n + n^2) + b \left[l + \frac{2}{3} n(1 + \sqrt{n}) \sqrt{\frac{S}{\pi}} \right] \quad (7.4.4)$$

Непосредственные поставки в низовые склады дают для g_{Σ} формулу

$$g_{\Sigma} = a(1 + n^2) + b \left[l + \frac{2}{3} n^2 \sqrt{\frac{S}{\pi}} \right] \quad (7.4.5)$$

Вычислим разницу между этими оценками g_{Σ} и приравняем ее нулю

$$\Delta g = an + \frac{2}{3} b \frac{\sqrt{S}}{\pi} (n + n\sqrt{n} - n^2) = 0,$$

или

$$\frac{2}{3} b \sqrt{\frac{S}{\pi}} (n - \sqrt{n}) - \left(a + \frac{2}{3} b \sqrt{\frac{S}{\pi}} \right) = 0$$

Заменив \sqrt{n} на m , приходим к квадратному уравнению

$$m^2 - m - \frac{3a + 2b\sqrt{S/\pi}}{2b\sqrt{S/\pi}} = 0$$

Смысл имеет только положительный корень этого уравнения

$$m_1 = \frac{1}{2} + \sqrt{\frac{1}{4} + \left(1 + \frac{3a}{2b\sqrt{S/\pi}} \right)} = \frac{1}{2} + \sqrt{\frac{5}{4} + \frac{3a}{2b\sqrt{S/\pi}}}$$

При $m > m_1$ введение промежуточных складов является оправданным (если не учитывать стоимость их эксплуатации). Фактически при разработке системы нам задается только общее число n^2 низовых складов. Следовательно, вопрос о введении промежуточной ступени возникает при выполнении неравенства

$$n^2 > \left(\frac{1}{2} + \sqrt{\frac{5}{4} + \frac{3a\sqrt{\pi}}{2b\sqrt{S}}} \right)^4. \quad (7.4.6)$$

Равномерное распределение поставки между половиной потребителей каждого склада

Такая организация поставок вызывает за цикл T следующие затраты:

$$\frac{n}{2} \left(a + \frac{2}{3} b \frac{\sqrt{S}}{\pi} \right)$$

на поставки с центрального склада в половину промежуточных;

$$\frac{n^2}{4} \left(a + \frac{2}{3} b \sqrt{\frac{S}{n\pi}} \right)$$

на поставки в низовые звенья. В итоге

$$g_{\Sigma} = a \left(1 + \frac{n}{2} + \frac{n^2}{4} \right) + b \left[l + \frac{1}{3} \left(n + \frac{n\sqrt{n}}{2} \right) \sqrt{\frac{S}{\pi}} \right]. \quad (7.4.7)$$

При непосредственной поставке в низовые склады число перевозок составит $n^2/2$ по цене $a + \frac{2}{3} b \sqrt{S/\pi}$, откуда

$$g_{\Sigma} = a \left(1 + \frac{n^2}{2} \right) + b \left[l + \frac{1}{3} n^2 \sqrt{\frac{S}{\pi}} \right]. \quad (7.4.8)$$

Теперь

$$\Delta g = n \left[\frac{a}{2} \left(1 - \frac{n}{2} \right) + \frac{b\sqrt{S}}{3\pi} \left(1 + \frac{\sqrt{n}}{2} - n \right) \right].$$

Для определения предельно допустимого числа складов в двухкаскадной системе имеем квадратное уравнение

$$m^2 \left(\frac{a}{4} + \frac{b\sqrt{S}}{3\pi} \right) - m \frac{b\sqrt{S}}{6\pi} - \left(\frac{a}{2} + \frac{b\sqrt{S}}{3\pi} \right) = 0,$$

решение которого дается формулой

$$m_1 = \frac{b\sqrt{S/\pi} + \sqrt{17b^2S/\pi + 18a^2 + 36ab\sqrt{S/\pi}}}{3a + 4b\sqrt{S/\pi}}. \quad (7.4.9)$$

Условие для n очевидно.

Поочередные поставки в промежуточные звенья и равномерное распределение между низшими складами

При этом в каждом цикле снабжения производится одна поставка в промежуточное звено и n поставок — в низшее. Очевидно, что

$$g_\Sigma = a(1 + 2n) + b \left[l + \frac{2}{3}(1 + \sqrt{n})\sqrt{\frac{S}{\pi}} \right]. \quad (7.4.10)$$

По смыслу этого варианта число каскадов не может быть меньше трех.

Рассмотренные примеры не исчерпывают всех форм организации снабжения в трехкаскадной системе, но иллюстрируют методы их анализа. Выбор варианта должен проводиться на основе вычисления суммарных затрат L^* для значений g_Σ , соответствующих сравниваемым вариантам. При этом трехкаскадный вариант выгоднее двухкаскадного лишь в том случае, когда расходы L^* в единицу времени при заданном n^2 для трехкаскадной модели меньше затрат в двухкаскадной на величину, превышающую расходы по содержанию складов промежуточной ступени. Описанная методика обобщается на любое число каскадов.

Если спрос на предметы снабжения подвержен значительным колебаниям, данная методика должна быть дополнена обсуждаемыми ниже соображениями по экономии от централизации резервов.

7.5. Децентрализованная система при вероятностном спросе

Достоинством систем снабжения с вероятностным спросом является возможность объединения резервов отдельных складов в целях покрытия дефицита на некоторых из них. Отказ от минимизации расходов

на каждом складе *порознь* позволяет ценой малого увеличения штрафов по относительно хорошо обеспеченным складам добиться существенного уменьшения *суммы* штрафов по системе *в целом*. Такая возможность, в свою очередь, позволяет снизить общий запас. Оценим это снижение количественно.

Пусть дана система из n складов с одинаковыми условиями хранения, распределениями спроса и ценами штрафов. Тогда при расчете затрат за период по математическим ожиданиям остатка и дефицита оптимальный запас на каждом складе, взятый отдельно, определится из уравнения (5.6.2) и может быть представлен в виде

$$S^* = \bar{x} + k\sigma_x,$$

где множитель k полностью определяется видом распределения и отношением $(d - c)/(d + h)$. Таким образом, суммарный запас в системе окажется равным $n(\bar{x} + k\sigma_x)$.

Если считать возможным идеальное (бесплатное) перераспределение, то функция затрат по системе в целом запишется как

$$L_\Sigma = c(S_\Sigma - z_\Sigma) + h \int_0^{S_\Sigma} (S_\Sigma - x)f_\Sigma(x) dx + d \int_{S_\Sigma}^{\infty} (x - S_\Sigma)f_\Sigma(x) dx$$

(индексом Σ отмечены спрос, запас и остаток по системе в целом), а ее минимальное значение доставляется решением уравнения

$$F_\Sigma(S^*) = (d - c)/(d + h). \quad (7.5.1)$$

Распределения спроса на отдельных складах обычно хорошо аппроксимируются нормальным. Распределение суммарного спроса уже при небольшом n оказывается близким к нормальному независимо от распределений составляющих. Следовательно,

$$S_\Sigma^* = \bar{x}_\Sigma + k\sigma_\Sigma,$$

где множитель k остается прежним. Очевидно, $\bar{x}_\Sigma = n\bar{x}$ и $\sigma_\Sigma^2 = n\sigma_x^2$, а достигаемое снижение запаса

$$\Delta S = k(n\sigma_x - \sqrt{n}\sigma_x) = k\sigma_x(n - \sqrt{n}).$$

Добавим к этому, что расчет по суммарным характеристикам обеспечивает оптимальную вероятность p^* отсутствия дефицита по системе в

целом, тогда как раздельный дает ту же вероятность по каждому складу *порознь* и, следовательно, по системе в целом лишь $(p^*)^n$. Значит, требования к «системной» вероятности и соответственно системный запас можно снизить дополнительно. При значительном n выигрыш от уменьшения запаса может перекрыть ожидаемые затраты на перераспределение.

Проиллюстрируем эти эффекты на числах. Функция распределения нормального закона

$$p = \text{Pr}\{X < y\} = \frac{1}{2} \left[\Phi \left(\frac{y - \bar{x}}{\sigma\sqrt{2}} \right) + 1 \right],$$

где $\Phi(\cdot)$ — функция Лапласа. Таким образом,

$$\Phi \left(\frac{y - \bar{x}}{\sigma\sqrt{2}} \right) = 2p - 1,$$

откуда следует

$$y = \bar{x} + \sigma\sqrt{2}\Phi^{-1}(2p - 1) = \bar{x} + k(p)\sigma.$$

Выигрыш в величине запаса составит

$$\Delta y = (n \cdot k(p) - \sqrt{n} \cdot k(p^n))\sigma.$$

Для определенности возьмем $p = 0.95$, $n = 10$. Тогда $p^{10} = 0.59874$, и

$$\Delta y = (10 \cdot 1.64485 - 0.250087\sqrt{10})\sigma = 15.6577\sigma.$$

Применительно к децентрализованной системе имеет смысл рассматривать следующие задачи:

- 1) организация восполнения суммарного запаса в системе;
- 2) оптимальное перераспределение наличного запаса;
- 3) оптимальное распределение прибывшей партии между складами.

Рассмотрим способы их решения в зависимости от числа номенклатур в системе.

7.5.1. Однородный запас

При заданной периодичности снабжения оптимальный суммарный запас может быть найден согласно (7.5.1) с подстановкой средней цены штрафа \bar{d} . Если же период является свободной переменной, то для нахождения уровня заказа s_Σ и объема поставки q_Σ можно воспользоваться методами разд. 5.8 с подстановкой суммарных характеристик спроса и средней цены штрафа.

При постановке задачи о *распределении* запасов, ввиду приблизительно одинаковых условий хранения в системе, сумма затрат на хранение считается не зависящей от принятого решения и исключается из рассмотрения, так что минимизации подлежит сумма затрат на штрафы и распределение (перераспределение) запаса. Введем обозначения:

c_{ij} — цена единичной перевозки между складами $i-j$, $i, j = \overline{1, n}$;

q_{ij} — объем перевозок между этими складами;

z_j — наличный запас на j -м складе;

d_j — цена штрафа на складе j ;

M^- — множество складов, получающих запас при перераспределении;

M^+ — множество складов, отдающих запас при перераспределении.

Тогда ожидаемые затраты за время до очередной поставки составят

$$\begin{aligned} L = & \sum_{j \in M^-} d_j \int_{z_j + \sum_{i \in M^+} q_{ij}}^{\infty} \left(x - z_j - \sum_{i \in M^+} q_{ij} \right) f_j(x) dx \\ & + \sum_{i \in M^+} d_i \int_{z_i - \sum_{j \in M^-} q_{ij}}^{\infty} \left(x - z_i - \sum_{j \in M^-} q_{ij} \right) f_i(x) dx \quad (7.5.2) \\ & + \sum_{j \in M^-} \sum_{i \in M^+} c_{ij} q_{ij} \end{aligned}$$

(плотности и цены даны в расчете на время ожидания поставки). В работе С. Аллена [102] показано, что для выпуклых функций затрат, к которым принадлежит и (7.5.2), в оптимальном плане перераспределения либо $\partial L_i / \partial q_{ij} = 0$, либо $\partial L_i / \partial q_{ij} > 0$ и $q_{ij} = 0$. Выполняя

дифференцирование, убеждаемся, что

$$\frac{\partial L_i}{\partial q_{ij}} = -d_j \int_{z_j + \sum_{i \in M^+} q_{ij}}^{\infty} f_j(x) dx + d_i \int_{z_i - \sum_{j \in M^-} q_{ij}}^{\infty} f_i(x) dx + c_{ij}.$$

Нижние пределы интегрирования в последнем уравнении представляют собой оптимальные запасы *после* перераспределения, так что условие оптимальности перевозки q_{ij} может быть записано в форме

$$d_i \int_{S_i^*}^{\infty} f_i(x) dx - d_j \int_{S_j^*}^{\infty} f_j(x) dx + c_{ij} = 0 \quad (7.5.3)$$

(i — поставщик, j — получатель).

Для разделения складов на множества M^+ и M^- можно, временно отвлекаясь от транспортных расходов, минимизировать сумму штрафов в системе

$$L_d = \sum_{i=1}^n d_i \int_{S_i}^{\infty} (x - S_i) f_i(x) dx \quad (7.5.4)$$

при условии, что суммарный запас в системе остается неизменным. Дифференцируя (7.5.4) по $\{S_i\}$, приходим к системе уравнений для условно оптимальных запасов

$$\begin{aligned} d_i \int_{\tilde{S}_i}^{\infty} f_i(x) dx &= d_1 \int_{\tilde{S}_1}^{\infty} f_1(x) dx, & i = \overline{2, n}, \\ \sum_{i=1}^n \tilde{S}_i &= \sum_{i=1}^n z_i. \end{aligned} \quad (7.5.5)$$

Искать корни такой системы удобно, задаваясь набором значений одного из $\{\tilde{S}_i\}$ (для определенности положим, что это \tilde{S}_1). Независимым решением первой группы уравнений (7.5.5) можно получить $\{\tilde{S}_i(\tilde{S}_1)\}$ для $i = \overline{2, n}$, а затем и их сумму. Решая относительно \tilde{S}_1 уравнение

$$\sum_{i=1}^n \tilde{S}_i(\tilde{S}_1) = \sum_{i=1}^n z_i,$$

находим \tilde{S}_1 и, выравнивая взвешенные вероятности дефицита, рассчитываем остальные $\{\tilde{S}_i\}$. Очевидно,

$$\begin{aligned} i \in M^+, & \text{ если } \tilde{S}_i < z_i, \\ i \in M^-, & \text{ если } \tilde{S}_i > z_i. \end{aligned} \quad (7.5.6)$$

При $\tilde{S}_i = z_i$ склад в перераспределении не участвует.

Рассмотрим динамику изменения левой части (7.5.7) в процессе перераспределения. Поскольку запас у поставщика снижается, а у получателя растет, эта функция будет возрастать. Но ее значение в оптимальной точке равно нулю. Следовательно, при исходном состоянии запасов должно выполняться условие

$$d_i \int_{z_i}^{\infty} f_i(x) dx - d_j \int_{z_j}^{\infty} f_j(x) dx + c_{ij} < 0 \quad (7.5.7)$$

для всех пар складов, участвующих в перераспределении. Проверка данных соотношений позволяет сформировать множества $\{U_i\}$ получателей склада i и множества $\{V_j\}$ поставщиков склада j и уточнить списки получателей и поставщиков в смысле их сокращения.

Нетрудно показать, что в процессе перераспределения новые направления целесообразных перевозок возникать не могут. В самом деле, пусть склад i — поставщик, j — получатель, а склад k в перераспределении в начальный момент времени не участвует. Это означает, что для $S_i = z_i$, $S_j = z_j$

$$\begin{aligned} d_i \int_{S_i}^{\infty} f_i(x) dx - d_k \int_{z_k}^{\infty} f_k(x) dx + c_{ik} &\geq 0, \\ d_k \int_{z_k}^{\infty} f_k(x) dx - d_j \int_{S_j}^{\infty} f_j(x) dx + c_{kj} &\geq 0. \end{aligned}$$

Но в ходе перераспределения $\int_{S_i}^{\infty} f_i(x) dx$ растет, а $\int_{S_j}^{\infty} f_j(x) dx$ уменьшается. Следовательно, списки поставщиков и получателей могут только сокращаться.

Выпишем уравнения (7.5.7) для всех пар перевозок, целесообразных при исходном состоянии запаса, и просуммируем их квадраты по

каждому поставщику и получателю. В результате имеем систему уравнений

$$\begin{aligned} u_i \left(d_i \int_{S_i}^{\infty} f_i(x) dx \right)^2 - \sum_{j \in U_i} \left(d_j \int_{S_j}^{\infty} f_j(x) dx - c_{ij} \right)^2 &= 0, \quad i \in M^+, \\ v_j \left(d_j \int_{S_j}^{\infty} f_j(x) dx \right)^2 - \sum_{i \in V_j} \left(d_i \int_{S_i}^{\infty} f_i(x) dx + c_{ij} \right)^2 &= 0, \quad j \in M^-. \end{aligned} \quad (7.5.8)$$

Здесь через u_i и v_j обозначены мощности множеств U_i и V_j соответственно.

Решение этой системы для определения оптимальных запасов может проводиться одним из численных методов — например, по Ньютону или методом скорейшего спуска. На каждом шаге необходимы:

- контроль неотрицательности запасов у поставщиков;
- при необходимости — изменение шага;
- исключение перевозок, ставших нерациональными;
- корректировка состава множеств U, V, M^+ и M^- .

Расчет $\{S_i^*\}$ заканчивается, когда множество M^- окажется пустым (одновременно опустеют и все остальные множества).

Сопоставляя объемы $\{S_i^*\}$ запасов на каждом складе после перераспределения с начальными запасами, можно определить суммарный объем поставок, полученных (отправленных) данным складом, и, решая транспортную задачу, найти наиболее экономичный план перевозок, реализующий окончательное распределение $\{S_i^*\}$. Общая схема алгоритма:

- 1) Решая систему (7.5.5), получить условно оптимальные запасы $\{\tilde{S}_i\}$.
- 2) Сформировать исходные множества поставщиков и получателей согласно (7.5.6).
- 3) Рассчитать произведения $d_i \int_{z_i}^{\infty}$ для всех $i \in M^+ \cup M^-$.
- 4) Для всех $j \in M^-$ принять V_j пустым, $v_j = 0$.
- 5) Для всех $i \in M^+$ выполнить следующие операции:

а) положить U_i пустым, $u_i = 0$;

б) для всех $j \in M^-$, удовлетворяющих условию (7.5.7):

- занести j в множество U_i и заменить u_i на $u_i + 1$,
- занести i во множество V_j и заменить v_j на $v_j + 1$.

- 6) Для всех $i \in M^+$, если $u_i = 0$, вычеркнуть склад i из числа поставщиков.
- 7) Для всех $j \in M^-$, если $v_j = 0$, вычеркнуть склад j из числа получателей.
- 8) Решая систему уравнений (7.5.8), найти оптимальные запасы $\{S_i^*\}$.
- 9) Сформировать окончательные множества поставщиков и получателей, для чего для всех i :
 - вычислить $q_i = S_i^* - z_i$;
 - если $q_i > 0$, отнести этот склад к получателям с общей потребностью q_i , иначе к поставщикам с наличием $-q_i$.
- 10) Решить транспортную задачу по наиболее экономичной перевозке груза $\sum_{i \in M^-} q_i$ от поставщиков к получателям.
- 11) Выдать в качестве результата объемы и направления перевозок.
- 12) Конец алгоритма.

Решение оказывается значительно проще при распределении внешнего пополнения. Предположим, что вся партия q прибывает на узловую станцию железной дороги и оттуда распределяется по складам в количествах $\{q_j\}$ с транспортными расходами $\{c_j q_j\}$. Тогда функция затрат примет вид

$$L = \sum_{j=1}^n \left(d_j \int_{z_j+q_j}^{\infty} (x - z_j - q_j) f_j(x) dx + c_j q_j \right) \quad (7.5.9)$$

при условии $\sum_{j=1}^n q_j = q$. Метод неопределенных множителей Лагранжа приводит к системе уравнений

$$\begin{aligned} \int_{z_j+q_j}^{\infty} f_j(x) dx &= (c_j + \xi)/d_j, \quad j = \overline{1, n}, \\ \sum_{j=1}^n q_j - q &= 0. \end{aligned} \quad (7.5.10)$$

Решение этой системы удобно получить с помощью независимого определения $(z_j + q_j)$ по обратным функциям законов распределения спроса для фиксированного ξ из диапазона $0 < \xi < \min_j \{d_j - c_j\}$. Повторив расчет для различных значений ξ , можно аппроксимировать зависимость $\sum_{j=1}^n q_j(\xi)$, получить обращающее сумму в q значение ξ и, вернувшись к первой группе уравнений (7.5.10), найти искомые $\{q_j\}$.

7.5.2. Многономенклатурный запас

Многономенклатурная задача также состоит из двух частей: снабжение системы в целом и распределение запасов. Как уже отмечалось, для ориентировочных расчетов затрат допустимо планирование поставок в систему по *общему* периоду снабжения. Следовательно, оптимальная периодичность снабжения

$$T^* = \sqrt{2g_{\Sigma} / \sum_{r=1}^N \lambda_r h_r}, \quad (7.5.11)$$

а объем поставок по номенклатуре r

$$q_r = \lambda_r T^*. \quad (7.5.12)$$

Задача распределения запаса в многономенклатурной системе ставится в двух вариантах. Если штрафы по отдельным номенклатурам на каждом складе суммируются, то решение может быть получено N -кратным (по числу номенклатур) независимым применением методики предыдущего подраздела. При выплате штрафов на каждом складе по максимуму взвешенного дефицита необходимо свести к минимуму эти максимумы, причем необходимость перераспределения будет определяться состоянием запаса на данном складе по всем номенклатурам. В таких случаях в системе всегда оказывается одна наиболее дефицитная («ключевая») номенклатура, наличный запас по которой *полностью* определяет минимальную сумму штрафов. В результате задача разбивается на три этапа:

- 1) выделить ключевую номенклатуру;
- 2) оптимальным образом распределить запас по ней;

- 3) распределить ресурсы по остальным номенклатурам таким образом, чтобы штрафы по ним на всех складах не превышали штрафов по ключевой номенклатуре, а транспортные расходы были минимальны.

Наиболее вероятным кандидатом в «ключевые» следует считать ту номенклатуру, для которой максимальна взвешенная вероятность недовыполнения по системе в целом, т.е. достигается $\max_r \bar{d}_r \int_{z_{\Sigma,r}}^{\infty} f_{\Sigma,r}(x) dx$, где \bar{d}_r — среднее значение d_{jr} по всем j . Схема вычислительного процесса выглядит следующим образом:

- 1) Рассчитать суммарные запасы $\{z_{\Sigma,r}\}$ по всем номенклатурам.
- 2) Вычислить взвешенные вероятности $\bar{d}_r \int_{z_{\Sigma,r}}^{\infty} f_{\Sigma,r}(x) dx$ и выбрать номенклатуру r_0 , для которой это произведение максимально.
- 3) Для номенклатуры r_0 применением методики предыдущего раздела рассчитать по всем складам j нормативные запасы $\{S_{jr_0}^*\}$
- 4) Решить уравнения

$$d_{jr} \int_{\tilde{S}_{jr}}^{\infty} (x - \tilde{S}_{jr}) f_{jr}(x) dx = d_{jr_0} \int_{S_{jr_0}^*}^{\infty} (x - S_{jr_0}^*) f_{jr_0}(x) dx \quad (7.5.13)$$

относительно запасов $\{\tilde{S}_{jr}\}$, $j = \overline{1, n}$, $r = \overline{1, N}$.

- 5) Вычислить для всех номенклатур r разности $q_r = \sum_{j=1}^n \tilde{S}_{jr} - \sum_{j=1}^n z_{jr}$ между запасом, необходимым для выравнивания штрафов под r_0 , и имеющимся в наличии
- 6) Если между этими разностями нет положительных, перейти к этапу 8.
- 7) Выбрать номенклатуру r_1 , для которой q_r максимально. Заменить r_0 на r_1 . Перейти к этапу 3.
- 8) Рассчитать разности $q_{jr} = z_{jr} - S_{jr}^*$ по всем складам и номенклатурам.
- 9) Сформировать значения избытков по всем j и r .

$$\Delta_{jr}^+ = \begin{cases} q_{jr}, & \text{если } q_{jr} > 0; \\ 0 & \text{иначе.} \end{cases}$$

10) Сформировать значения дефицитов по всем j и r .

$$\Delta_{jr}^- = \begin{cases} -q_{jr}, & \text{если } q_{jr} < 0; \\ 0 & \text{иначе} \end{cases}$$

11) Занести $\{\Delta_{jr}^-\}$ и $\{\Delta_{jr}^+\}$ в таблицу постановки транспортной задачи.

12) Приступить к выбору объемов перевозок $\{q_{ijr}\}$ по минимуму транспортных затрат (в процессе минимизации затрат мы будем использовать обычно выполняемые на практике условия $c_{ijr} = c_{ij} \cdot c_r$), для чего.

- Выбрать $\min_{i,j} c_{ij}$
- Для данной пары складов вычислить $q_{ijr} = \min\{\Delta_{jr}^-, \Delta_{ir}^+\}$ и скорректировать остатки путем их уменьшения на q_{ijr} . Провести эту операцию для всех r . Если на складе j все $\{\Delta_{jr}^-\}$ оказались равными нулю, исключить склад j из числа получателей. Если на складе i все $\{\Delta_{jr}^+\}$ оказались нулевыми, вычеркнуть склад i из числа поставщиков.
- Для той же пары складов вычислить объемы перевозок во встречном направлении $q_{jir} = \min\{\Delta_{jr}^+, \Delta_{ir}^-\}$, скорректировать остатки и списки получателей и поставщиков.
- Проверить, остался ли хоть один получатель (иначе говоря, есть ли положительный дефицит Δ_{jr}^-). Если остался, найти минимальный элемент в оставшейся части таблицы. Перейти к шагу б).

13) Конец алгоритма.

Заметим, что предложенный в данной задаче многомерный аналог метода минимального элемента нахождения опорного плана транспортной задачи линейного программирования не требует исчерпания излишков у всех поставщиков; необходимо лишь покрытие дефицитов (при этом по ключевой номенклатуре одновременно закрываются и излишки).

Совершенно аналогично решается и задача о распределении прибывшего в систему пополнения по нескольким номенклатурам: предварительно выявляется самая дефицитная номенклатура, и для нее решается система уравнений (7.5.10). Далее независимым решением уравнений (7.5.13) определяются необходимые запасы, подсчитывается

их сумма и сравнивается с располагаемым резервом. Окончательно выделив ключевую номенклатуру, вновь решаем систему уравнений (7.5.10) и назначаем объем поставок q_{jr} по всем j, r до уравнивания штрафов. Излишняя в указанном смысле часть поставок направляется по наиболее дешевым маршрутам с учетом наличия свободного места на складах.

Задачи со способом расчета штрафов, отличным от максимального взвешенного дефицита, могут быть решены по аналогичной схеме.

Небольшая модификация предложенных алгоритмов позволяет получить приближенную методику решения задачи с учетом фиксированных составляющих $\{g_{ij}\}$ по маршруту $i \rightarrow j$. Для этого необходимо всякий раз сопоставлять выигрыш от уменьшения штрафов с транспортными расходами. В частности, для однородного случая это уменьшение штрафа за вычетом расходов, пропорциональных объему перевозки, составит

$$\Delta_{ij} = d_i \int_{S_i^* + q_{ij}}^{\infty} (x - S_i^* - q_{ij}) f_i(x) dx + d_j \int_{S_j^* - q_{ij}}^{\infty} (x - S_j^* + q_{ij}) f_j(x) dx - d_i \int_{S_i^*}^{\infty} (x - S_i^*) f_i(x) dx - d_j \int_{S_j^*}^{\infty} (x - S_j^*) f_j(x) dx - c_{ij} q_{ij}.$$

Разбив первые два интеграла на два слагаемых, можно показать, что

$$\Delta_{ij} = - d_i \int_{S_i^*}^{S_i^* + q_{ij}} (x - S_i^*) f_i(x) dx + d_j \int_{S_j^* - q_{ij}}^{S_j^*} (x - S_j^*) f_j(x) dx + q_{ij} \left[d_j \int_{S_j^* - q_{ij}}^{\infty} f_j(x) dx - d_i \int_{S_i^* + q_{ij}}^{\infty} f_i(x) dx - c_{ij} \right].$$

Для малых объемов перевозок $\{q_{ij}\}$ содержимое квадратных скобок можно приравнять нулю. Применив теорему о среднем, имеем

$$\Delta_{ij} \approx \frac{q_{ij}^2}{2} [d_j f_j(S_j^* - q_{ij}/2) - d_i f_i(S_i^* + q_{ij}/2)]. \quad (7.5.14)$$

Если $\Delta_{ij} > g_{ij}$, такая перевозка является оправданной. В противном случае она должна быть исключена. Проверив все возможные перевозки и запретив невыгодные, необходимо вновь вернуться к алгоритму разд. 7.5.1 и т.д. — до тех пор, пока не обнаружится, что все оставшиеся перевозки выгодны. Последний полученный план перевозок считается окончательным.

7.6. Линейные многокаскадные системы снабжения

Под линейной системой снабжения понимается цепь складов, из которых каждый последующий удовлетворяет заявки предыдущего, причем любой поставщик имеет только одного потребителя. Такая схема, разумеется, лишь весьма приближенно описывает реальные эшелонированные системы. Тем не менее она заслуживает внимания, поскольку при относительной простоте модели сохраняет качественное своеобразие эшелонированных систем.

Большинство работ по управлению запасами предполагает независимость задержки выполнения требований от объема этих требований. Такое допущение неверно для многокаскадных систем, где время задержки будет определяться текущими уровнями запаса в различных ее звеньях. Эта задержка, вообще говоря, будет состоять из случайного числа вероятностных слагаемых. Кроме того, ограниченность запасов в высших звеньях системы может привести к неполному удовлетворению требований низшего звена. При этом условие оптимальности объема поставки, полученное методами предшествующих глав, окажется нарушенным, и расходы в нижнем звене отклонятся от своего минимума. Оптимальная стратегия в эшелонированной системе должна строиться с учетом такой возможности.

Для иллюстрации методов исследования линейных систем рассмотрим статическую задачу из трехзвенной цепи складов 1, 2 и 3, хранящих продукты производственного процесса (1 — окончательный продукт, 2 — промежуточный, 3 — сырье)¹. Спрос на конечный продукт за период T , на который создается запас, имеет известную плотность распределения $f(x)$. Будем считать, что при дефиците на складе i выплачивается штраф по цене d_i на единицу дефицита и неудовлетворенный спрос передается в следующее звено. Положим далее, что запас во всех звеньях исчисляется в расчете на единицу конечного продукта. Естественно считать, что цены хранения $\{h_i\}$ возрастают с повышением готовности продукта, что делает невыгодным хранение только предметов непосредственного спроса. Расходы на хранение примем пропорциональными остатку продукта на каждом складе. Необходимо найти такие величины запасов $\{S_i\}$, при которых общие расходы L на хранение и

штраф будут минимальны.

В зависимости от величины спроса расчетные формулы будут различны. Именно:

$$\begin{aligned} L_1 &= h_1(S_1 - x) + h_2S_2 + h_3S_3, & 0 < x \leq S_1, \\ L_2 &= d_1(x - S_1) + h_2(S_1 + S_2 - x) + h_3S_3, & S_1 < x \leq S_1 + S_2, \\ L_3 &= d_1(x - S_1) + d_2(x - S_1 - S_2) + h_3(S_1 + S_2 + S_3 - x), & S_1 + S_2 < x \leq S_1 + S_2 + S_3, \\ L_4 &= d_1(x - S_1) + d_2(x - S_1 - S_2) + d_3(x - S_1 - S_2 - S_3), & x > S_1 + S_2 + S_3 \end{aligned}$$

Просуммировав $\{L_i\}$ с учетом вероятности соответствующих событий, после перегруппировки членов получаем математическое ожидание затрат:

$$\begin{aligned} L &= (h_1S_1 + h_2S_2 + h_3S_3) \int_0^{S_1} f(x) dx - h_1 \int_0^{S_1} xf(x) dx \\ &+ [h_2(S_1 + S_2) + h_3S_3] \int_{S_1}^{S_1+S_2} f(x) dx - h_2 \int_{S_1}^{S_1+S_2} xf(x) dx \\ &+ h_3(S_1 + S_2 + S_3) \int_{S_1+S_2}^{S_1+S_2+S_3} f(x) dx - h_3 \int_{S_1+S_2}^{S_1+S_2+S_3} xf(x) dx \\ &- d_1S_1 \left[1 - \int_0^{S_1} f(x) dx \right] - d_2(S_1 + S_2) \left[1 - \int_0^{S_1+S_2} f(x) dx \right] \\ &- d_3(S_1 + S_2 + S_3) \left[1 - \int_0^{S_1+S_2+S_3} f(x) dx \right] \\ &+ d_1 \int_{S_1}^{\infty} xf(x) dx + d_2 \int_{S_1+S_2}^{\infty} xf(x) dx + d_3 \int_{S_1+S_2+S_3}^{\infty} xf(x) dx. \end{aligned}$$

Вычислим частные производные $\partial L / \partial S_i$, $i = \overline{1, 3}$, и приравняем их нулю. Для упрощения положим $\int_0^{S_1} f(x) dx = \pi_1$, $\int_{S_1}^{S_1+S_2} f(x) dx = \pi_2$.

¹Очевидно применение этой схемы к задачам теории надежности при расчете числа элементов, находящихся в горячем, теплом и холодном резервах.

$\int_{S_1+S_2}^{S_1+S_2+S_3} f(x) dx = \pi_3$. В результате приходим к системе уравнений, линейных относительно $\{\pi_i\}$:

$$\begin{aligned} \pi_1(d_1 + d_2 + d_3 + h_1) + \pi_2(d_2 + d_3 + h_2) + \pi_3(d_3 + h_3) &= \sum_{i=1}^3 d_i, \\ \pi_1(d_2 + d_3 + h_2) + \pi_2(d_2 + d_3 + h_2) + \pi_3(d_3 + h_3) &= \sum_{i=2}^3 d_i, \\ \pi_1(d_3 + h_3) + \pi_2(d_3 + h_3) + \pi_3(d_3 + h_3) &= d_3. \end{aligned}$$

Определитель этой системы

$$\Delta = (d_2 + h_2 - h_3)(d_1 + h_1 - h_2)(d_3 + h_3). \quad (7.6.1)$$

По условию, $h_1 > h_2 > h_3$. Кроме того, цены штрафа выше стоимостей хранения. Следовательно, $\Delta > 0$, и система имеет единственное решение. Находим остальные определители:

$$\begin{aligned} \Delta_1 &= d_1(d_3 + h_3)(d_2 + h_2 - h_3), \\ \Delta_2 &= (d_3 + h_3)[d_1(h_3 - h_2) + d_2(h_1 - h_2)], \\ \Delta_3 &= (d_1 + h_1 - h_2)[d_3(h_2 - h_3) - d_2h_3]. \end{aligned}$$

Таким образом,

$$\begin{aligned} \pi_1 &= \frac{d_1}{d_1 + h_1 - h_2}, \\ \pi_2 &= \frac{d_1(h_3 - h_2) + d_2(h_1 - h_2)}{(d_2 + h_2 - h_3)(d_1 + h_1 - h_2)}, \\ \pi_3 &= \frac{d_3(h_2 - h_3) - d_2h_3}{(d_2 + h_2 - h_3)(d_3 + h_3)}. \end{aligned} \quad (7.6.2)$$

Введем теперь новые переменные

$$\begin{aligned} \pi_1^0 &= \pi_1, \\ \pi_2^0 &= \pi_1 + \pi_2, \\ \pi_3^0 &= \pi_1 + \pi_2 + \pi_3. \end{aligned} \quad (7.6.3)$$

Подставляя в (7.6.3) ранее найденные π_1, π_2 и π_3 , имеем

$$\pi_1^0 = \frac{d_1}{d_1 + h_1 - h_2},$$

$$\begin{aligned} \pi_2^0 &= \frac{d_2}{d_2 + h_2 - h_3}, \\ \pi_3^0 &= \frac{d_3}{d_3 + h_3}. \end{aligned} \quad (7.6.4)$$

С помощью (7.6.4) по таблицам интегрального закона распределения $F(x)$ спроса на конечный продукт или непосредственным вычислением можно найти $S_1, S_1 + S_2$ и $S_1 + S_2 + S_3$, а затем и сами искомые переменные.

До сих пор мы считали все величины $\{\pi_i\}$ положительными. Получим необходимые и достаточные условия справедливости этого предположения. Анализ системы (7.6.2) с учетом исходных допущений показывает, что

- 1) π_1 всегда больше нуля;
- 2) $\pi_2 > 0$, если $d_1(h_2 - h_3) < d_2(h_1 - h_2)$, или

$$\frac{d_1}{h_1 - h_2} < \frac{d_2}{h_2 - h_3}; \quad (7.6.5)$$

- 3) $\pi_3 > 0$, если $d_3(h_2 - h_3) > d_2h_3$, или

$$\frac{d_2}{h_2 - h_3} < \frac{d_3}{h_3}. \quad (7.6.6)$$

Следовательно, все $\{\pi_i\}$ положительны, когда

$$\frac{d_1}{h_1 - h_2} < \frac{d_2}{h_2 - h_3} < \frac{d_3}{h_3}. \quad (7.6.7)$$

Общий алгоритм расчета трехзвенной линейной системы:

- 1) Вычислить π_1 согласно первому из уравнений (7.6.2).
- 2) Если условие (7.6.5) выполнено, вычислить π_2 по второму из уравнений (7.6.2), иначе положить $\pi_2 = 0$.
- 3) Если условие (7.6.6) ложно, положить $\pi_3 = 0$. Иначе
 - если $\pi_2 > 0$, вычислить π_3 по третьему из уравнений (7.6.2),
 - иначе положить $\pi_3 = d_3/(d_3 + h_3) - \pi_1$ (это третье уравнение исходной системы с подстановкой в него $\pi_2 = 0$).

- 4) Согласно (7.6.3) рассчитать $\{\pi_i^{(0)}\}$.
- 5) Найти соответствующие им квантили распределений.
- 6) Конец алгоритма.

7.7. АСУ снабжением: опыт и проблемы

Процесс получения информации при ручной обработке занимает 40–70% времени оперативных работников системы снабжения. Преимущества научного управления запасами в системах снабжения могут быть реализованы только с помощью соответствующих автоматизированных систем управления (АСУ). В экономически развитых странах это было осознано достаточно давно. В подтверждение напомним читателю расшифровку наименования крупнейшей корпорации IBM — *International Business Machines*, а также дату создания языка Cobol (*Commercial Oriented Business Language*) — 1960 г. В Японии по состоянию на апрель 1970 г из 6700 ЭВМ использовались в торговле 1116.

Внедрению АСУ снабжением в СССР (начиная с 1965 г.) был присущ ряд просчетов стратегического и тактического характера [81]. Основы неправильной стратегии закладывались на высшем уровне тогдашнего руководства, отсталость которого в научной подготовке порождала непонимание ситуации в целом, потенциальных возможностей АСУ и их границ, условий успешного применения. Руководители ориентировались на автоматизацию решения традиционных задач («большой арифмометр») при сохранении неизменными структуры организации и форм управления, считали необходимым «для надежности» дублировать машинный счет ручным. Опасение за престиж и должности и желание показаться на гребне научно-технической революции вынуждало форсировать разработки АСУ снабжением. Между тем проектирование АСУ связано с совместной работой специалистов по вычислительной технике, математиков и экономистов, только достижение взаимопонимания между которыми требовало 3–4 лет. Но за это время требовалось выдать готовую АСУ!! Разработчики недооценили сложность и новизну проблемы, не провели поисковые исследования, не учли зарубежный опыт. Мало внимания уделялось централизации и типизации разработки программ, средствам и технологии настройки на конкретные условия, унификации документации, номенклаторов продукции, сквозной интеграции систем. Все это обнаруживалось на завершающих этапах и требовало глобальных переделок. Естественно, что при

этом могли быть реализованы только *фрагменты* необходимого целого, которые делаются легче и быстрее (учетно-статистические задачи).

Среднее звено управленцев болезненно воспринимало сужение сферы своей деятельности, резкий рост требований к качеству информации и ее анализу. Требовалось определить состав требуемых данных, их структуру и способ представления, разработать набор классификаторов (продукция; поставщики и потребители; министерства и ведомства; услуги и работы; их исполнители; основные фонды, здания и сооружения; транзитные нормы; тарифные справочники; виды тары и т.д.). Надлежало собрать и систематизировать собственно данные — размещение объектов системы, спрос, тарифы, прямые связи; продумать технические средства и организационные формы их обновления и пересчета «по факту». Изданное по этому поводу во исполнение поручения Президиума Совета Министров постановление Госплана от 10 марта 1971 г. своевременно выполнено не было. Документы оказались разнотипными по отраслям, часть информации — выводимой из других сведений и потому избыточной. В то же время выяснились и упущения, многого (в особенности для решения задач оптимизации) недоставало. Информация оказалась не приспособленной для передачи по каналам связи.

Позже Госнаб провел работу по унификации [81, с. 170], снизившую число форм с 1015 до 41 — но только для верхних уровней системы управления снабжением. Внизу так и остался разнобой, отчасти порождавшийся разнотипностью устройств подготовки первичной информации. Явно недостаточное внимание уделялось проверке правильности ввода и выполненных вычислений (контрольным кодированием и суммированием).

Материальная заинтересованность во внедрении АСУ снабжением на всех уровнях отсутствовала. Не было своевременно осознано (и разъяснено старым кадрам), что ЭВМ надо использовать не *вместо* человека, а *вместе* с ним (но обученным). Ситуация многократно усугубилась чудовищной стоимостью, низкой надежностью и «недружественностью к пользователю» ЭВМ первых поколений. Машины класса «Минск» последовательных серий не были программно совместимыми. Программы для экономических задач — все многократного использования — не комплексовались в цепочки по исходным данным. Не была выдержана общая система программирования.

Тем не менее был приобретен и бесспорно положительный опыт. Удалось отработать сжатие информации, планирование и учет в укрупненной номенклатуре (в верхних звеньях). Было выяснено, что для

достоверного планирования практически достаточен детальный обсчет крупных потребителей (все тот же принцип 20/80 %). Более 90% эффекта, полученного в 8-й пятилетке в организациях Госснаба, было достигнуто на нетрадиционных оптимальных задачах (а их было менее 10% от общего числа автоматизированных). Например, решение задачи по перераспределению производства лаков и красок по 8 заводам снизило затраты на их изготовление на 700 тыс. руб. в год. Перепланирование загрузки трубных станов позволило дополнительно произвести 200 тыс. тонн дефицитных труб. Полномасштабное внедрение оптимальных планов перевозок дало выигрыш в 170 млн. рублей. Кстати заметим, что от комплексного внедрения научных достижений в управление американцы ожидали ежегодного дохода в 40 млрд. долл. По данным Стэнфордского университета, от оснащения АСУ 5000 складов США ожидалась ежегодная экономия в 250 млн. долл.

С учетом накопленного опыта в 9-й пятилетке объем капиталовложений в АСУ был увеличен в 6 раз.

Таким образом, накануне перестройки в СССР сложились уникально благоприятные условия для реального перехода к научному управлению запасами:

- сосредоточение в руках государства экономических и административных рычагов управления экономикой;
- осознание как верхами, так и низами актуальности рационального управления запасами;
- достаточный теоретический задел по управлению запасами;
- широкое внедрение теории запасов в учебный процесс подготовки экономистов и организаторов производства;
- большой (пусть неудачный, но правильно осмысленный) опыт разработки АСУ снабжением;
- появление высокопроизводительных, надежных и недорогих ПЭВМ и универсального программного обеспечения их (офисные системы, системы управления базами данных).

Все эти факторы имеют место и сейчас — у «естественных монополистов» типа Газпрома, РАО ЕС, министерств путей сообщения, обороны и т.п. Однако методы ТУЗ актуальны и в рыночной экономике — от упомянутых гигантов до промышленного или агропредприятия и даже

рыночного ларька. Кстати, приятно отметить, что каждое приобретение в ларьке сейчас фиксируется продавцом с точным указанием купленного — несомненно, в целях оптимального управления запасами (в зависимости от образования — в бытовом или научном смысле).

Часть III

**Проблема
запасных частей**

Глава 8

Снабжение запасными частями

8.1. Специфика проблемы

Поломки случаются на самых современных предприятиях и электростанциях, при эксплуатации наиболее совершенных потребительских продуктов (автомобили, видеотехника, персональные ЭВМ), в супернадежной авиации и в системах военного назначения, опирающихся на новейшие технологии. Какова бы ни была причина отказа, для потребителей очень важен коэффициент готовности системы, зависящий от возможности замены отказавшего узла. Ведь большой ущерб может быть причинен даже за малое время. Очень существенна деловая репутация в сильно конкурентном бизнесе (ПЭВМ), в особенности поддержание гарантированного обслуживания. Примеры критических систем: управление воздушным движением, электронные СМИ, системы поддержания жизнедеятельности в госпиталях, системы военного назначения. Ключевым элементом поддержания готовности сложных технических систем является ЗИП.

ЗИПом обычно называют комплект запасных элементов, предназначенных для поддержания техники в работоспособном состоянии в течение расчетного срока ее эксплуатации. В ЗИП часто включают специальный инструмент, принадлежности и нестандартные материалы.

Одиночный комплект ЗИП предназначен для текущего ремонта

аппаратуры путем замены неисправных элементов, а также для проведения профилактических работ и настройки аппаратуры в эксплуатационных условиях. Одиночный ЗИП прилагается одному определенному изделию.

Групповой комплект ЗИП служит для пополнения одиночного ЗИПа и обеспечения техники теми элементами, которых нет в одиночном ЗИПе. Групповой комплект прилагается группе из n агрегатов. Как правило, комплекты ЗИПа поставляются вместе с изделиями заводом-изготовителем. В процессе эксплуатации аппаратуры комплекты ЗИПа пополняются: одиночный — за счет группового, а групповой — со специального склада или из органа снабжения. Такие элементы назовем ЗИП-россыпью. Пополнение ЗИПа может проводиться сразу после изъятия из него элемента для установки в аппаратуру или через определенные интервалы времени. Нетрудно видеть, что восполнение ЗИПа аналогично процессам, происходящим в системах снабжения пирамидальной структуры.

Стоимость ЗИПа обычно приводится к стоимости одиночного комплекта добавлением к ней n -й части стоимости группового.

Изготовители машин стараются оптимизировать номенклатуру и состав ЗИП, включая в ассортимент преимущественно узлы и агрегаты и уменьшая набор стандартных деталей (крепеж, лампочки и т.п.), доступных из других источников. Унификация узлов, как правило, существует в пределах номенклатуры одного изготовителя, но встречается и на уровне субпоставщиков. Например, английский филиал «Форда» выпускает лишь четыре базовых модели с большим числом вариантов комплектации из типовых узлов. Болты, гайки, шланги, подшипники качения, свечи — унифицируются по машиностроению в целом.

Ремонтируемые узлы технически сложных изделий обычно заменяются на взятые из ЗИПа, а отказавшие поступают в ремонт и пополняют ЗИП. Соответствующие органы должны входить в структуру системы поддержания готовности. Управление запасами с учетом восстановления и задержек — более сложная задача. Поскольку стоимость восстанавливаемых узлов может в десятки раз превышать стоимость расходных, проблемы данного вида вызывают особый интерес.

Как справедливо отмечали классики теории надежности Б.В. Гнеденко, Ю.К. Беляев и И.Н. Коваленко в «Итогах науки» за 1966 год, «методы расчета количества запасных частей, необходимого для нормальной эксплуатации оборудования, находятся в крайне запущенном состоянии». В особенности это относится к математико-экономическим

методам. К сожалению, с тех пор ни в теории, ни в ее использовании значительного прогресса не наблюдалось.

Запасные части как товар имеют сложную техническую и коммерческую специфику. Каждый автомобиль, комбайн и т.п. собирается из тысяч деталей, и в период эксплуатации может потребоваться замена практически любой из них. Характерным примером статической задачи является проблема закупки запасных частей вместе с каким-либо уникальным и дорогостоящим оборудованием (атомным ледоколом, прокатным станом, турбогенератором большой мощности, обрабатывающим центром). Вместе с основным оборудованием запчасти могут быть поставлены по относительно низкой цене, а позже — по гораздо более высокой (из-за необходимости повторного запуска производства). Типичным примером является задача из [89, с. 361] относительно детали рулевого механизма танкера: начальная цена 5 тыс. долл., цена дефицита 25 тыс. Еще один пример из той же книги [89, с. 344–346] связан с запасными частями для самолета: каждая запасная часть стоит 2 тыс. долл., утилизация неиспользованных — по 100 долл., ликвидация каждого случая дефицита обходится в 13 тыс. долл. Добавим, что в обоих случаях штраф исчислялся только по стоимости ликвидации дефицита, и к нему следует добавить ущерб от простоя оборудования в ожидании поставки. Высокие «штрафы» часто приводили к завышению запасов, особенно в системах военного назначения. На пятом симпозиуме США по надежности отмечалось большое накопление запасных элементов на складах Министерства обороны. Стоимость этих элементов достигла 21 млрд. долл. и ежегодно увеличивалась на 2 млрд.

Актуальность проблемы в век информатики возросла в связи с тем, что в качестве сменных элементов электронной аппаратуры все чаще используются микросхемы, модули, субланели и т.п., стоимость которых значительно выше цены отдельных схемных элементов (резисторов, конденсаторов, ламп и полупроводников) недавнего прошлого:

«В качестве одной из серьезных вычислительных проблем можно назвать проблему организации в нашей стране системы обеспечения типовыми запасными элементами парка ЭВМ Единой системы "Ряд". Учитывая, что все модели ЕС ЭВМ имеют единую техническую базу, можно легко понять, насколько существенный выигрыш может дать система централизованного обеспечения всех вычислительных машин

запасными элементами. Создание научно обоснованной централизованной системы снабжения потребует привлечения всего арсенала уже разработанных математических моделей теории управления запасами, а также проведения новых исследований»

(из предисловия И. А. Ушакова к [50]).

Важнейшими особенностями систем снабжения запчастями являются:

- глобальный характер подобных систем, сложность их структуры и функционирования,
- включение в систему функций и органов ремонта вышедших из строя деталей и узлов;
- колоссальное число планируемых номенклатур;
- высокая цена хранения и «штрафов» за дефицит;
- большая роль фактора взаимозаменяемости.

Эти особенности обусловили изменение методов работы с запчастями и проведение крупных научных исследований в области организации и специализации производства, маркетинга, управления запасами их. К определению товарной политики относятся разработка методов поставки на рынок (источники, периодичность, объемы и сроки), упаковки и партионности (комплекты), закупки, складирования и т.п. Важность и сложность (теоретическая и практическая) этих проблем привели к тому, что все крупнейшие производители выделили производство и сбыт запасных частей в самостоятельные службы.

Организация технического обслуживания и ремонта — важнейшая обязанность производителя технически сложного товара и критический элемент маркетинга. Разумеется, техническое обслуживание должно быть рентабельно для обеих сторон. Практически ценовая политика выстраивается так, что возможная дополнительная прибыль вдвое превышает доход от продаж [17].

Строго говоря, задачи выбора номенклатуры запасных элементов и оптимального назначения периодичности и объема поставок для обеспечения непрерывного функционирования системы следует решать совместно. Ниже описана согласно [17] структура систем обеспечения

запчастями импортных автомобилей. Такие системы весьма развиты, достаточно типичны и интересны широкому кругу лиц.

Мастерские импортера не являются главными потребителями запасных частей; тем не менее их доля в сбыте запасных частей значительна. Для дилеров продажа через их мастерские является основным методом сбыта запасных частей. Накладные расходы мастерских очень велики вследствие использования дорогостоящего оборудования. Для улучшения их окупаемости организаторы автосервиса стремятся увеличить объем продаж запасных частей. Мастерские снабжаются в кратчайшие сроки непосредственно со склада; механиков заинтересовывают в продаже запчастей и вменяют им в обязанность не просто заменить отказавшие детали, но и указать клиенту на те детали, которые пора заменить.

Специалисты, занимающиеся техническим обслуживанием оборудования, являются основным источником идей по совершенствованию производимых фирмой изделий.

8.2. Системы обеспечения запчастями

8.2.1. Структура

В работах по теории снабжения запасными частями обычно изучаются следующие варианты: один склад, линейная система, иерархическая система. Многоуровневые системы дают шанс повысить эффективность и экономичность запасов. Из них, как правило, рассматривается двухуровневая — «базы» на нижнем уровне и «депо» на верхнем. Учитывая краткость и наглядность этих терминов, а также обычность первого из них для военных приложений, мы будем в дальнейшем использовать именно их. Увеличение числа уровней не вызывает принципиальных трудностей.

Для описания структуры системы применяются теоретико-множественные, графовые и комбинированные обозначения. В графовых схемах удобно определять звенья-предшественники. С точки зрения программирования расчетных методик значительные преимущества дают списковые структуры данных.

Горизонтальный обмен между звеньями одного иерархического эшелона обычно не рассматривается. Поставки из внешнего источника поступают только на верхний уровень. На нижних уровнях могут быть

«фиктивные» (shadow — теневые) склады, не хранящие запас физически и организующие удовлетворение заявок по почте.

Системный подход требует постоянно иметь в виду не максимальную экономию на отдельных складах или по каждой номенклатуре, но конечную цель. В связи с этим концепция цены недостачи представляется почти неизбежной. Ее можно попытаться обойти через уровень обслуживания *с задержкой, не превышающей допуска*. Возможные варианты целевой функции:

- суммарный ожидаемый простой потребителей, вызванный отсутствием запаса;
- ожидаемое число отложенных на нижнем уровне поставок;
- ожидаемое число находящихся в готовности конечных потребителей.

В военных приложениях часто используются коэффициент готовности потребителей или ожидаемое число недостач на нижнем уровне (базы), непосредственно связанные с их оперативными возможностями. Цели оптимизации на разных уровнях могут быть частично противоречивы. Это естественно и необходимо для успешной работы системы в целом.

Существенно новым элементом рассматриваемых в этой части книги задач является включение ремонтных звеньев. Ремонтные услуги могут быть на всех уровнях. Иногда различаются (по месту проведения, задержкам и, разумеется, условиям оплаты) гарантийное и послегарантийное обслуживание. Обычно ремонтный орган (РО) связан со своим складом и РО высшего уровня. Отремонтированные агрегаты остаются на складе своего уровня.

Новыми аспектами иерархических систем являются следующие проблемы:

- связанность низовых баз общей возможностью дефицита на верхнем уровне;
- выбор в случае дефицита (кому отказать или как распределить дефицит);
- вертикальная синхронизация периодических стратегий (согласование периодичностей поставок для разных эшелонов системы);
- общая структура ремонта (суммарная загрузка; на каком уровне восстанавливать; кому адресовать восстановленный агрегат; вводить ли приоритетное восстановление).

Еще один новый аспект — древовидная структура самих запчастей (стойки, модули, ТЭЗы, микросхемы). Листьями дерева являются неремонтируемые (расходуемые) детали, запас которых возобновляется от поставщиков через эшелонированную систему складов. Пара (i, k) обозначает узел i на k -м уровне иерархии изделия. Эта структура может быть наложена на иерархию системы материально-технического обеспечения. Ремонт связывается с определенным уровнем структуры изделия и эшелонем системы.

Системы этого типа сложны для анализа, но существует общее убеждение в их предпочтительности. Одна из причин — невозможность организовать ремонт во всех точках потребления. Вторая — глобальный масштаб некоторых организаций. Третья заключается в уменьшении страховых резервов и возможности поставить их на службу большому числу потребителей.

Возникающие трудности анализа можно разделить на концептуальные и расчетные. Концептуальные порождаются сложностью системы и множеством ее функций (поставки, хранение, транспорт, ремонт) в их взаимосвязи, а также ее вхождением в суперсистему. Нелегко провести декомпозицию задачи: финансы, начальные запасы, стратегия поставок, распределение, перераспределение и т.д. Поэтому отдельные подзадачи приходится решать последовательно, а весь их комплекс — с применением итераций.

Расчетные трудности порождаются большой размерностью задач (десятки тысяч номенклатур, сотни потребителей), а также сложностью получения распределения спроса в высших звеньях. Если, например, имеется J складов, то точное выражение для плотности распределения ненулевого спроса в высшем звене будет иметь $2^J - 1$ составляющих. Для вполне реального значения $J = 100$ это 10^{30} .

8.2.2. Организация поставок

Источниками поставок запасных частей к импортным автомобилям (напомним, что мы рассматриваем эту ситуацию как типовую) служат:

- склады при заводах — изготовителях машин (для деталей невысокого спроса собственного производства);

- центральные склады в стране экспортера (для деталей, получаемых от субпоставщиков, и деталей собственного производства с высоким спросом);
- филиалы центральных складов в стране экспортера, специализированные на хранении деталей определенной номенклатуры

Жесткая конкуренция фирм-производителей, импортеров и местных производителей обеспечила высокую плотность сбытовой сети и облегчила покупателям приобретение деталей, но уменьшила объем сбыта отдельного продавца и его прибыль.

Крупные экспортеры имеют промежуточные склады не только в своей, но и в других странах. Импортеру обычно поручается продажа машин и запасных частей от его имени и за его счет (чтобы экспортер не облагался налогами в стране импортера, не нес ответственности по претензиям покупателей и убытков из-за их несостоятельности). Экспортер отвечает только за количество, качество и сроки поставки и часто соглашается на оплату по мере реализации. Возврат непроданных деталей обычно ограничивается 5–15% годового объема продаж (как правило — с уценкой).

Сети распределения с оптовыми промежуточными базами считаются более прогрессивными. На розничных складах хранятся только детали группы А, на оптовых — А и В, на заводских — С. При транзитной форме поставок везде хранится запас всех категорий. Значительно уменьшается страховой запас. На складах импортеров обычно хранится 20% номенклатуры, на промежуточных — до 50% и на центральных — до 80%. Остальные 20% деталей низкого спроса поставляются по мере необходимости со складов заводов-изготовителей.

Сбытовая сеть многих крупных компаний строится таким образом, чтобы гарантировать получение клиентом деталей, относящихся к группе А, немедленно, а к группам В и С — не позднее суток после получения заказа. Для обеспечения подобного обслуживания на складах дилера хранятся одно-двухмесячные запасы деталей высокого спроса, на складах импортеров — трех-четырёхмесячные запасы деталей высокого и постоянного спроса, на промежуточных — то же для нерегулярного спроса. Необходимо учитывать, что АВС-классификация относительна и зависит от уровня склада.

Некоторые компании применяют такую структуру, при которой запасы на складах рассматриваются как совокупный запас и могут перераспределяться по решению руководства (склады компании «Форд» в

Англии, Бельгии и Дании) Компания «Джон Дир», имеющая в Канаде центральный склад запасных частей и несколько филиалов, ввела в 1973 г. единую систему FLAH (Fast Locating and Handling) управления запасами на всех складах в стране. Центральный склад располагает информацией о запасах во всех филиалах. Теперь 95% срочных заказов дилеров выполняются ближайшим филиалом, 3% — его соседом и лишь 2% — центральным складом. Аналогичная система создана в компании «Интернэшнл Харвестер».

Уровень удовлетворения спроса редко превышает 90–95%. Детали нерегулярного спроса дают малый процент выручки и хранятся по соображениям престижа и сохранения рынка. Часто дилерам предписывается содержание запаса деталей определенной номенклатуры или на определенную сумму. Распределяя между дилерами всю номенклатуру запчастей, включая детали редкого спроса, импортеры имеют возможность сократить запасы на своих складах за счет неходовой номенклатуры. Наиболее часто в торговле запасными частями в звене импортер–экспортер применяется система с фиксированным интервалом заказа (для планирования производства или массовых закупок) и с фиксированным объемом — в звене дилер–импортер. Основным планированием, как правило, является трехмесячное. Заявленное на первый месяц рассматривается как жесткий заказ, на второй и третий — как прогноз спроса (скользящее планирование). Для запчастей к машинам, снятым с производства, применяются полугодовые заказы. Допуск (1–5%) обычно делается в сторону увеличения, чтобы избежать риска неудовлетворенного спроса.

Склады «Форда» в Финляндии и Норвегии направляют заказы на детали высокого спроса еженедельно, постоянного — ежемесячно, нерегулярного — ежеквартально. Склады запасных частей агентов компаний «Фиат» и «Тойота» в этих же странах размещают заказы на детали высокого спроса два раза в месяц. Экспортеры рекламируют поставку запасных частей по заказам агентов в течение 1–3 месяцев. На практике в эти сроки отгружается не все (по объему и наименованию) заказанное. Недопоставка составляет от 5% («Фольксваген», «Форд») до 30% («Фиат»).

Срочные заказы на детали выполняются экспортерами в сроки от нескольких часов до 3–15 дней с момента получения заявки по телексу. Срок определяется местонахождением источника поставки. Авиаотгрузка при таких поставках оплачивается импортерами. Крупные экспортеры, осуществляющие большое количество срочных поставок, часто

закljučают договоры с авиакомпаниями на резервирование грузовых мест в самолетах. Например, «Интернэшнл Харвестер» (США) по договору с авиакомпанией имеет право отгружать каждые сутки 1 т запасных частей ночным самолетом в Париж. В Канаде законодательные акты регламентируют максимальные сроки поставки запчастей фермерам (72 часа в период сезонных работ и 2 недели в остальное время).

При отгрузке крупных партий на большие расстояния предпочтителен морской, речной или железнодорожный транспорт как наиболее дешевый. Для отгрузки небольших партий (до 20 т) применяются автомобильные перевозки. В Европе имеется несколько фирм, занимающихся автоперевозками запчастей. Достоинствами таких перевозок являются: маневренность; повышающее сохранность отсутствие перевалок; экономия на tare; более высокая скорость. Автомобильные перевозки практически всегда используются в сочетании с другими видами на конечных участках маршрута.

Большинство экспортеров упаковывают мелкие запасные части в стандартные картонные коробки. Тем самым заранее устанавливается минимальный размер заказываемой партии деталей. Другим вариантом является упаковка необходимого для одной машины комплекта деталей в одну коробку и поставка деталей только комплектом (например, поршни — по 4 и 6 штук). Третий вариант — упаковка в одну коробку ремонтных комплектов для карбюраторов, бензонасосов, генераторов, а также комплекты прокладок, сальников и т.д. Запасные детали большого размера, но малого веса (детали кузова) могут не иметь индивидуальной упаковки и отгружаются в картонных коробках, ящиках и контейнерах с сепарацией. Крупные тяжелые узлы, как правило, крепятся на салазках и укладываются в контейнеры или автофургоны с предохранительными распорками. Прецизионные поверхности узлов защищаются специальными прокладками или толстыми пленками. Компания «Фольксваген» организовала на заводских складах и складах европейских импортеров единую систему оборотных складских контейнеров. «Харвестер» использует авиаконтейнеры.

8.3. Военные системы снабжения

Военные системы снабжения относятся к числу наиболее масштабных и аккумулируют в себе все типичные проблемы: глобальные

масштабы, громадное число номенклатур (до 600 тыс.), дороговизну запчастей, большую роль процессов ремонта. По [184], 65% всех вложений в запасы ВВС США — ремонтируемые, их стоимость превышает 10 млрд. долл. В ВМФ доля ремонтируемых деталей составляет 58%.

Использование математико-экономических методов в организации снабжения вооруженных сил имеет давние традиции, в особенности на флоте (увы, не нашем ¹), о чем свидетельствует выходящий с середины 1950-х гг. на высоком теоретическом уровне журнал «Naval Research Logistics Quarterly». Тем не менее значительная часть военных запасов все еще управляется по правилам, не имеющим теоретического обоснования [124]. Много лет существует проблема оценки *оперативной* значимости запасов (традиционные показатели обеспеченности не выводили на этот уровень). Стоимостный эквивалент невыполненного требования на практике не использовался. В практической работе столкнулись два критерия — «хорошего бизнеса» и оперативной готовности. В 1971–72 гг. за 4 года только 10–15% позиций корабельного ЗИПа имели ненулевой спрос. Это привело к уменьшению вложений в запчасти — и снизило оперативную готовность, в особенности старых кораблей.

В августе 1971 г. командующий морскими операциями (Chief of Naval Operations) выпустил директиву, в которой предписал разработать автоматизированную модель, связывающую затраты на снабжение (запасы, ремонтные возможности) с боевой готовностью флота. Был учрежден комитет из представителей соответствующих служб. В 1975 г. было приказано развить и детализировать эти исследования, вследствие чего была выполнена серия работ по математическому обеспечению снабжения фрегатов и эсминцев. В рамках этих работ Кларк [124] предложил модель Logistics Support Economic Evaluation (LSEE), в основу которой были положены иерархические схемы как системы снабжения, так и самого оборудования. Предполагалось, что ожидаемое время дополнительных задержек зависит от наличия запаса и ремонтных возможностей (в том числе в высших звеньях). Задачей считалось обеспечение максимальной готовности при выделенных затратах или двойственная ей — минимум затрат для обеспечения заданной готовности. Она решалась для соединения из 10 эсминцев и береговой базы по 243 типам оборудования. На предварительных этапах возникало много традиционных трудностей с номенклаторами и исходными данными. В частности, реальные уровни разборки не согласовывались с номенклатурой запасов.

¹Первая публикация об управлении запасами в ВМФ СССР принадлежит автору этой книги — см. Морской Сборник, 1964, №10.

Примеры трехуровневых структурированных номенклатур: корабельная помпа (34 детали), компьютер (280 деталей). Важную роль сыграла автоматизация обработки статистической информации из учетных файлов. Стали учитываться назначение оборудования, возможность многоцелевого использования деталей и агрегатов.

Для сравнения с моделью фактически наблюдавшейся готовности было использовано ретроспективное моделирование. Расчет выполнялся снизу вверх (по уровням системы и сборок) с итеративным пересчетом задержек в верхних уровнях в зависимости от характера ремонта и запасов в них. В конце концов по каждому типу конечного оборудования на основе анализа стоимость/эффективность определялся оптимум. Итог расчета представлял собой зависимость максимально достижимой боеготовности от бюджетных ограничений.

Расчеты показали, что на затраты и готовность кораблей сильнее всего влияет изменение запасов сборок высшего уровня. Суммарные затраты были снижены на 40%. Непосредственно на кораблях было предложено хранить 63% номенклатур вместо 10% по старой схеме.

Те же проблемы возникали (и примерно так же решались) в ВВС США (Air Force Logistics Command [129]) Представляют интерес основные группы позиций, по которым производился расчет:

- Конечные изделия (самолеты, грузовики, наземное заправочное и испытательное оборудование).
- Ремонтируемые изделия (навигационные компьютеры, насосы, радио, генераторы) — 170 тыс. видов изделий стоимостью 10 млрд. долл.
- Двигатели (в принципе это особо важные и дорогостоящие ремонтируемые части, выделенные отдельно).
- Расходуемые детали (500 тыс. видов, 2 млрд. долл.).
- Горюче-смазочные материалы.
- Вооружение (бомбы, ракеты, зажигательные средства).
- Предметы обеспечения личного состава (одежда, пища, медикаменты).

Обычно рассматривалась стратегия поставок $(S - 1, S)$. В системе восстановления предполагалось (как в обсуждаемой ниже схеме METRIC) бесконечное число каналов и соответственно — отсутствие

очередей Модель была применена в ВВС для расчета организации снабжения истребителей-бомбардировщиков F-15 [184].

Авторы цитированных работ отмечают, что обновление и рост надежности техники, обучение ремонтного персонала, совершенствование контрольного оборудования требуют периодического (через 4–5 лет) пересчета стратегий управления запасами.

8.4. Экономические параметры

Выпуск деталей «впрок» связан со значительными убытками от их физического хранения и омертвления капитала. Эти реальные обстоятельства используются в качестве щита продуцентами, которые считают, что «продать машину значит посадить дерево, продавать запчасти — снимать урожай с этого дерева» Цены на запчасти растут быстрее, чем на машины. Отчасти это объясняется ростом надежности техники, вследствие чего уменьшается суммарный спрос и удорожается одиночное изделие. С другой стороны, некоторые фирмы пытаются компенсировать снижение цен на новые машины ростом стоимости запчастей после завоевания рынка В итоге стоимость машины, собранной из запчастей, оказывается в 2–3 раза выше стоимости новой. Противостоят этой тенденции лишь конкуренция и рынок подержанных запасных частей.

Постоянное обновление техники и широкая продажа подержанных машин обусловили необходимость наличия и, возможно, изготовления запчастей и после снятия машины с производства. «Даймлер-Бенц» гарантирует 15-летнее снабжение. Фирмы либо создают запасы на такой срок, либо сохраняют технологическую документацию и оснастку, пока имеется более или менее устойчивый спрос. Экспортеры учитывают даже потребности любителей старых машин, в том числе довоенных. Продуценты заинтересованы в таких ситуациях как в лучшей рекламе. Детали несерийного производства считаются дороже серийных в 4–5 раз. Наблюдается значительное расхождение в ценах на аналогичную продукцию. Цены на запчасти к старым машинам растут наиболее быстро. Это объясняется издержками хранения, уменьшением объемов выпуска и стремлением оказать давление на потребителя («купи, наконец, новую машину!»). Мелкие фирмы, а иногда и поставщики («Форд», «Вольво», «СААБ», «Крайслер», «БМВ») практикуют восстановление изношенных деталей (двигателей, коробок передач, водяных и масляных насосов и т.п.).

Запчасти делятся на особо конкурирующие (от многочисленных поставщиков), конкурирующие (изготавливаются на специализированных заводах), неконкурирующие (только от изготовителя машины и его субподрядчиков). В связи с обострением конкуренции покупатель убеждает приобретать только оригинальные запасные части, информирует об их преимуществах и об отрицательных последствиях применения запчастей, предлагаемых конкурентами.

В общем цены на запчасти слабо связаны с издержками производства и определяются рыночными соображениями. Дилерам дают возможность иметь прибыль на уровне средней по району или чуть выше, остальное забирает экспортер. Розничные цены обычно намного выше (в Европе — до двух раз, в африканских странах в 6 раз и более), чем в стране продуцента, и не только из-за налогов, транспортировки и таможенных пошлин. В цены запасных частей, не подверженных конкуренции, вводят большую долю прибыли для компенсации мало доходной продажи конкурирующих деталей.

Существующие нормы предусматривают оценку запасов в автоцентрах и на станциях техобслуживания в розничных ценах, а на складах более высокого уровня в существенно более низких оптовых ценах.

Приведем издержки английских дилеров на хранение автотракторных запчастей (в процентах от закупочной цены) [16]:

Всего потерь	5.5
В том числе	
неликвиды	1.0
повреждения при обработке	1.0
хищения	1.5
порча при хранении	0.8
устаревание	1.2
Стоимость найма помещений	12.0
Амортизация инвентаря и оборудования	7.5
Страхование оборудования, имущества и запасов	2.0
Заработная плата персонала	10.5
Итого	45.0

Эти оценки существенно отличаются от применявшихся в системах военного назначения. В табл. 8.1 приводятся данные по американским

ВВС для низшего звена системы снабжения (база) и окружного склада из статьи Петерсена и Джислера [189].

Наиболее полный анализ цен дается в работе [162], посвященной исследованию системы снабжения ВВС Канады, где предложено рассчитывать годовые расходы, связанные с обеспечением потребителей одной номенклатурой, по формуле

$$L = c_1 T + c_2 h \bar{y} + c_3 n_3 + c_4 n_4 + c_5 n_5 + c_6 n_6.$$

Входящие в нее переменные имеют следующий смысл:

T — время хранения номенклатуры;

\bar{y} — средний за T уровень ее запаса;

n_3 — число требований на пополнение в год;

n_4 — число требований на склад высшей ступени в год;

n_5 — число экстренных требований в год;

n_6 — число изменений состояния запаса в год.

Соответственно задается размерность коэффициентов $\{c_i\}$.

Т а б л и ц а 8.1

Стоимостные факторы в ВВС США

Слагаемые функции затрат		Цены, \$	
		для склада	для базы
Оформление заявки (на одну заявку)	обычную	0.72	1.03
	срочную	3.88	4.93
Обработка сообщения		0.98	2.62
Хранение и обработка (за тонну)		13.84	7.50
Упаковка и погрузка (за тонну)		14.18	14.18
Транспортные расходы	обычные ж/д	0.05	
	экспресс	0.19	
	самолет	0.50	

Смысл и величины цен $c_1 - c_6$ даются табл. 8.2. Поучительно сопоставить их значения с приведенными в главе 2 по перечню статей и численным значениям. Подчеркнем уже отмечавшиеся и типичные для военных систем несообразно низкие оценки ущерба от омертвления капитала, а также потери от морального старения (военная техника обычно относится к быстро стареющим высокотехнологичным изделиям). Использование для оптимизации этих цен наверняка приведет к завышенным запасам. Этот эффект отчасти компенсируется невысокой ценой «штрафов», сводящихся здесь к стоимости экстренной поставки.

Т а б л и ц а 8.2

Цены в снабжении ВВС Канады

Обозн.	Расходы	Компоненты	Цена, \$	Из расчета
c_1	Администр.	Взятие на хранение	0.25	В год
		Проверка счета, осмотр	0.25	
		Итого	0.50	
c_2	Пропорц.	Процент на капитал	0.04	На \$
		Собственно хранение	0.04	На \$
		Моральное старение	0.06	На \$
		Итого	0.14	
c_3	Нормальная поставка	Подг. перфокарты	0.05	На заявку
		Подготовка наряда	0.68	
		Обработка на складе	0.74	
		Оформление приема	0.40	
		Итого	1.80	
c_4	Цена отказа	Подготовка заявки	0.20	На заявку
		Обработка заявки	0.63	
		Канторские расходы	0.74	
		Экспедит. документы	0.25	
		Доп. трансп. расходы	0.40	
		Оформление приема	0.40	
Итого	2.50			
c_5	Срочн. треб.		0.70	На поставку
c_6	Корр. учета		0.10	На корр.

8.5. Спрос на запчасти

8.5.1. Общие соображения

Спрос на дорогие детали с низким спросом и восстанавливаемый ЗИП всегда рассматривается как дискретный. Зависимый спрос на взаимозаменяемые и на комплектующие детали (с положительной и отрицательной корреляцией соответственно) часто встречается на практике и редко — в теории.

Необходимо тщательное прогнозирование спроса на запчасти для всего жизненного цикла системы (марки автомобиля или самолета, вооружения). При этом приходится отвечать на стандартные вопросы из следующего перечня:

- на каких данных основывать расчет потребности;
- как часто изготавливать (заказывать);
- сколько стоит заплатить;
- как быть с новыми продуктами;
- как учесть изменение популяции потребителей (например, в случае запчастей к устаревающей технике число инсталляций растет, интенсивность отказов тоже; с другой стороны, часть потребителей переходит на новую технику).

Структура спроса на запчасти непосредственно на агрегате (автомобиле) определяется стратегией его технической эксплуатации: «по ресурсу» или «по состоянию». В первом случае деталь подлежит замене после наработки расчетного количества часов или километров пробега, во втором — при обнаружении критического состояния ее в момент проведения очередного регламентного цикла. Здесь для диагностики без разборки требуется дорогостоящее оборудование. Разумеется, *отказавшая* деталь в обоих случаях заменяется немедленно.

Спрос, порождаемый регламентом, можно выразить как детерминированную составляющую. Практически большее значение имеет случайный спрос.

Прогноз можно делать с учетом условий эксплуатации — температуры, влажности, квалификации персонала, фактической наработки и т.д., а также недостоверного знания интенсивности отказов. В среднем срок службы автомобилей и автотракторной техники в тропиках составляет 3–5 лет, в Европе 8–11. Соответственно и расход запчастей

в тропиках выше в 2–3 раза. По опыту среднегодовая стоимость покупаемых запчастей в год для автомобиля равна 2% от его стоимости, для колесного трактора 3–5%. Каждое наименование имеет свои сроки службы и специфические причины спроса и выступает как отдельный товар. Однако в машинах имеются и такие детали, которые практически не изнашиваются; потребность в их замене вызывается случайными или аварийными поломками.

Ассортимент запасных частей даже к одной модели машин постоянно меняется из-за модификаций агрегатов и отдельных деталей, вызываемых опытом эксплуатации или постепенным переходом к новым или модифицированным моделям машин. Изменения в номенклатуре запасных частей требуют четкого информирования оптовых и розничных покупателей об этих изменениях, о степени взаимозаменяемости новых и старых деталей, о правилах монтажа и технического обслуживания новых узлов. Обычные сброшюрованные каталоги заменяются каталогами со сменными листами. Затем появились каталоги на микрофишах (США, Англия, ФРГ, Франция). В наши дни просмотреть каталоги и прайс-листы можно через Интернет.

Сокращение объема сбыта в некоторых районах может быть результатом их низкого потенциала, падения спроса, изменения в мотивах и привычках покупателей, неудачной рекламы, острой конкуренции, плохой работы различных звеньев сбыта.

Прогнозы общего объема сбыта делятся на краткосрочные, среднесрочные и долгосрочные. *Краткосрочные* составляются на финансовый или календарный год и используются в качестве основы для планирования потребностей в денежных средствах, товарах и рабочей силе в течение года. Разработав программу сбыта по всем районам рынка, фирма составляет программу заказов и складирования на весь год, чтобы быть готовой к сезонным изменениям продаж в каждом районе. *Среднесрочные* прогнозы охватывают период от двух до пяти лет и обычно являются экстраполяцией существующих тенденций на будущее с учетом предполагаемых изменений в парке машин, срока службы и объема продаж ранее выпущенных марок, конъюнктуры рынка и т.п.

Прогноз делается на основе прошлого товарооборота, анализа тенденций и циклов, корреляционного анализа. Ожидается корреляция сбыта машин и запчастей (с соответствующим временным лагом, переменным по номенклатурам). По возможности выявляют множественную корреляцию. Источниками информации об *общих экономических усло-*

виях могут быть официальные публикации, банковские бюллетени и т.п. Известно, например, что энергетический кризис 1973–1974 гг. вызвал сокращение продаж автомобилей, но положительно отразился на продаже запчастей по следующим причинам:

- у клиентов не стало денег на новые автомобили,
- ввиду нехватки топлива водители лучше следили за машинами;
- по той же причине возросла доля коротких поездок, связанных с наибольшим износом деталей из-за частых разгонов и торможений

В общем случае рассчитывается сумма спроса по плановым заменам и стохастического — по отказам. Имеются разные подходы к расчету последнего в зависимости от типа оборудования: электронное, электромеханическое, механическое, гидравлическое. Для электроники типичен пуассоновский спрос, независимый от других компонент. Разные интенсивности отказов для рабочего режима, ожидания и пребывания на полке умножаются на соответствующие количества. Обычно отказами хранящихся деталей пренебрегают. Фактические интенсивности изменяются в диапазоне от 10^{-2} до 10^{-8} 1/час. Спрос часто не коррелирован с длительностью работы под нагрузкой. Спорят о том, нужно ли при прогнозе спроса в одном месте учитывать данные о других. В статистическом смысле это желательно (уменьшается разброс итоговых оценок); но одновременно затрудняется выявление местной специфики.

Необходимо отметить существенные успехи в повышении точности расчетов износостойкости и усталостной прочности деталей, что позволяет лучше прогнозировать динамику спроса на запчасти.

Запасы некоторых деталей высокого спроса имеют оборот до 8–10 в год, но средний оборот в 3–5 раз меньше в связи с необходимостью хранения деталей с низким (нерегулярным) спросом — см табл. 8.3 по данным 1972 г., [16].

Т а б л и ц а 8.3

Бездействующие запасы

Страна	«Форд»	«Фольксваген»	«Фиат»
Швеция	1/3.5	10/24	2/8
Финляндия	1/4	8.8/28	2/7.6

Здесь первая цифра показывает процент бездействующих запасов по стоимости, а вторая — по номенклатуре.

Некоторые импортеры сокращают мертвые и бездействующие запасы путем распределения их по складам дилеров и при появлении заявки быстро разыскивают деталь. Детали машин, снятых с производства, стараются заменить деталями относительно недавнего выпуска, если конструкция допускает замену с незначительной подгонкой. Чаще всего так поступают при ремонте подержанных автомобилей с целью последующей перепродажи. Для ликвидации мертвых и бездействующих запасов организуются специальные распродажи и аукционы, предоставляются повышенные скидки.

Статистика спроса дилеров существенно отличается от первичной: она укрупнена, сглажена и отстает от фактической картины. Ее важнейшей особенностью является *групповой* спрос, который предъявляется пачками случайного объема в промежуточный или центральный склад с установленной периодичностью. На этом фоне могут реализовываться (за счет клиента) экстренные поставки со складов высших уровней.

8.5.2. Распределение спроса для резервированных систем

Резервирование технических систем фактически эквивалентно введению в них встроенного запаса, который расходуется и время от времени пополняется в процессе эксплуатации. Внешним проявлением выхода из строя резервных элементов с точки зрения теории запасов является спрос на запасные элементы, наблюдаемый только в моменты контроля исправности системы. Характеристики спроса определяются тремя факторами:

- типом резерва («холодный» или «горячий»),
- глубиной контроля (полный или частичный),
- организацией контроля (периодический или непрерывный),

которые могут сочетаться в различных комбинациях. Рассмотрим влияние каждого из них порознь.

Тип резерва определяет фактическую (в отличие от внешне проявляемого спроса) скорость расхода резервных элементов. При *холодном* резерве эта скорость постоянна, определяется только числом работающих элементов и не зависит от начального состояния системы. Для экс-

понижения вероятности наработки на отказ вероятности снижения числа резервных элементов на x единиц за время t

$$p_x = \frac{(\lambda t)^x}{x!} e^{-\lambda t}, \quad x = \overline{0, k+1}, \quad (8.5.1)$$

где k — кратность резервирования.

В случае *горячего* резерва удельная интенсивность отказов выше, причем весь резерв также подвержен отказам. Распределение числа отказавших элементов

$$p_x = \binom{k+1}{x} e^{-(k+1-x)\lambda t} (1 - e^{-\lambda t})^x, \quad x = \overline{0, k+1}. \quad (8.5.2)$$

Эти вероятности удобно рассчитывать рекуррентно по соотношениям

$$p_0 = e^{-(k+1)\lambda t},$$

$$p_x = p_{x-1} \frac{k+2-x}{x} (e^{\lambda t} - 1), \quad x = \overline{1, k+1}.$$

Глубина контроля определяет его полноту, которую мы будем рассматривать в двух вариантах. Назовем *полным* контролем поэлементную проверку системы, а *частичным* — проверку на функционирование системы в целом. При полном контроле выявляемый в момент проверки спрос равен числу отказавших элементов. При частичном спрос равен нулю, если система в целом исправна, и числу отказавших элементов — в противном случае.

Организация контроля. Непрерывный контроль приводит к тому, что момент предъявления спроса на запасной элемент (несколько элементов) совпадает с моментом отказа очередного элемента или системы в целом. При периодическом контроле спрос накапливается в течение периода и предъявляется в момент очередной проверки.

Комплексные варианты. Сочетание перечисленных факторов порождает специфические способы расчета спроса. При непрерывном и полном контроле поток требований совпадает с потоком отказов: для холодного резерва это будет пуассоновский поток с интенсивностью λ , а для горячего — с интенсивностью $(k+1)\lambda$. Ограничение полноты контроля при сохранении его непрерывности существенно меняет структуру потока требований на восполнение. При холодном резерве поток прореживается и становится эрланговским $(k+1)$ -го порядка с групповыми требованиями объема $(k+1)$. При горячем резерве имеем рекуррентный поток с временем между последовательными групповыми

требованиями, определяемым временем расходования $(k+1)$ элемента в «схеме гибели». В этом случае плотность распределения времени между последовательными требованиями есть

$$f(t) = -\frac{d}{dt} p_{k+1}(t) = -\frac{d}{dt} [(1 - e^{-\lambda t})^{k+1}]$$

$$= (k+1)\lambda (1 - e^{-\lambda t})^k e^{-\lambda t}$$

(в предположении, что $1/\lambda$ значительно больше времени замены)

При полном периодическом контроле спрос за период t в точности равен расходу элементов за этот срок и для холодного резерва распределен согласно (8.5.1), а для горячего — в соответствии с (8.5.2).

Наиболее сложен для расчета случай неполного периодического контроля. Здесь расход элементов подсчитывается как в предыдущем варианте, но восполнение производится лишь при снижении резерва до уровня, обнаруживаемого системой контроля. Естественно, что вероятность этого события зависит от начального запаса. Распределение спроса за период, следовательно, будет определяться стационарным распределением начального запаса. Обозначим

S — максимальный запас,

s — пороговый уровень, достижение которого обнаруживается системой контроля,

p_z — стационарная вероятность начального запаса $z = \overline{0, S}$,

$\pi_{i,j}$ — вероятность снижения запаса за период от уровня i до уровня j .

Запас z к началу очередного периода для $z = \overline{s+1, S}$ может образоваться в двух случаях:

- 1) к началу предыдущего периода был запас, меньший или равный s , так что запас восполнили до уровня S ; в течение периода запас изменился с уровня S до $S - z$, т.е. спрос составил z единиц;
- 2) к началу предыдущего периода был запас $x > s$, так что восполнение не производилось; в течение периода запас изменился с x до z .

Для $z = \overline{0, s}$ первый случай аналогичен, а второй допускает начальный запас $x = s + 1, \overline{S}$

Суммируя вероятности соответствующих событий и считая распределение вероятностей начальных состояний не зависящим от номера периода, приходим к системе уравнений

$$\begin{aligned} p_z &= \pi_{S, S-z} \sum_{x=0}^s p_x + \sum_{x=z}^S \pi_{x,z} p_x, & z = \overline{s+1, S}, \\ p_z &= \pi_{S, S-z} \sum_{x=0}^s p_x + \sum_{x=s+1}^S \pi_{x,z} p_x, & z = \overline{0, s}. \end{aligned} \quad (8.5.3)$$

В частности, при холодном резерве $\pi_{x,z} = \pi(x-z)$ (вероятность перехода определяется разностью индексов) Полученные системы уравнений линейны относительно $\{p_z\}$ и в практических ситуациях имеют малую размерность Распределение спроса за период имеет следующий вид:

$$\begin{aligned} Q_0 &= \sum_{x=s+1}^S p_x, \\ Q_z &= p_{S-z}, & z = \overline{S-s, S}, \\ Q_z &= 0, & z = \overline{1, S-s-1}. \end{aligned} \quad (8.5.4)$$

Приведем подсчитанные по (8.5.4) средние значения спроса и его дисперсии при $s = 2, S = 5$ (табл. 8.4)

Т а б л и ц а 8.4

Характеристики распределения спроса

λT	0.2	0.3	0.5	0.7	1.0	1.5	2.0
Среднее	0.200	0.299	0.496	0.688	0.959	1.345	1.655
Дисперсия	0.586	0.867	1.398	1.878	2.488	3.206	3.628

При пуассоновском спросе среднее и дисперсия совпадают и равны λT (первая строка). Как видно из таблицы, средний спрос за период совпадает со средним значением пуассоновского спроса лишь при малых значениях упомянутого произведения, а дисперсия во всем диапазоне существенно превышает дисперсию пуассоновского спроса Таким образом, пренебрежение полнотой контроля может привести к грубым ошибкам в оценке возможного спроса за период

«Отказовый» спрос наблюдается и на фоне периодического контроля

8.6. Расчет распределения времени ремонта

Время ремонта — сумма длительностей диагностики, доставки в ремонт, собственно ремонта и доставки восстановленного агрегата на склад Чистая длительность ремонта зависит не только от вида детали и поломки, но и от квалификации персонала и оснащения ремонтного органа Процесс моделируется посредством систем массового обслуживания с числом каналов $1, n, \infty$. Здесь рассматриваются только два фактора: различные длительности ремонта (и их вероятности) и периодичность возобновления ремонта. Оптимизация может заключаться в балансировании между большим запасом и малым временем восстановления (решается вопрос, во что предпочтительнее вложить деньги: в оборотный запас или аппаратное оснащение, количество и квалификацию ремонтников).

Полное время ремонта определяется, в частности, организацией рабочего цикла Чистая длительность собственно ремонта в обычно используемых методиках предполагается распределенной показательной, однако современное состояние теории очередей позволяет рассчитывать временные задержки при более общих допущениях — см главу 3 Особо подчеркнем необходимость и возможность представления ремонтных органов высших звеньев как *сетей обслуживания*.

8.6.1. О распределении длительности диагностики

В технических системах дежурные смены локализуют неисправности с точностью до блоков, содержащих десятки элементов. Подозрительные блоки заменяются исправными из ЗИПа и поступают в группу ремонта, где происходит поиск отказавшего элемента и его замена Для расчета оптимального ЗИПа необходимо знать распределение времени восстановления блока, однако статистических данных для его построения, как правило, имеется мало. В этой ситуации естественно возникает вопрос об априорном построении упомянутых распределений Мы рассмотрим его решение при поиске неисправного элемента непосредственным перебором. Обозначим

$w(t), f(t), g(t)$ — плотности распределения времени ремонта блока, проверки исправного элемента и проверки неисправного элемента с последующей заменой соответственно;

w_k, f_k, g_k — моменты этих распределений k -го порядка, $k = 1, 2, \dots$;

$\omega(s), \varphi(s), \gamma(s)$ — ПЛС тех же распределений.

Будем считать блок состоящим из N равнонадежных элементов, из которых отказал один и только один. Перебирая порядковые номера отказавших элементов, получаем для ПЛС распределения времени восстановления блока выражение

$$\omega(s) = \frac{1}{N} \sum_{i=1}^N \gamma(s) \varphi^{i-1}(s) = \frac{\gamma(s) \varphi^N(s) - 1}{N \varphi(s) - 1}. \quad (8.6.1)$$

Применив обратное ПЛС к правой части (8.6.1), можно найти искомое распределение $w(t)$. Более выгодно, однако, рассчитать моменты $\{w_k\}$ этого распределения и по ним подобрать удобную для расчетов аппроксимацию плотности $w(t)$. Представим левую часть (8.6.1) в виде

$$\omega(s) = \int_0^{\infty} e^{-st} w(t) dt = \int_0^{\infty} \left(1 - st + \frac{s^2}{2!} t^2 - \frac{s^3}{3!} t^3 + \dots \right) w(t) dt.$$

Интегрируя почленно, убеждаемся, что

$$\omega(s) = 1 - sw_1 + \frac{s^2}{2!} w_2 - \frac{s^3}{3!} w_3 + \dots$$

Аналогичным образом могут быть записаны $\varphi(s)$ и $\gamma(s)$. Подставим эти результаты в (8.6.1) и избавимся от знаменателя. Приравнявая коэффициенты при одинаковых степенях s в левой и правой частях полученного уравнения, получаем выражения для искомых моментов.

$$\begin{aligned} w_1 &= \frac{N-1}{2} f_1 + g_1, \\ w_2 &= (N-1) \left[\frac{f_2}{2} + f_1 g_1 + \frac{N-2}{3} f_1^2 \right] + g_2, \\ w_3 &= (N-1) \left[\frac{f_3}{2} + \frac{3}{2} (f_1 g_2 + f_2 g_1) + (N-2) (f_1^2 g_1 + f_1 f_2) \right. \\ &\quad \left. + \frac{(N-2)(N-3)}{4} f_1^3 \right] + g_3. \end{aligned}$$

С их помощью можно найти дисперсию:

$$D_w = D_g + \frac{N-1}{2} D_f + \frac{N^2-1}{12} f_1^2.$$

Из общих формул при больших N следуют асимптотические оценки

$$w_1 \approx \frac{N}{2} f_1; \quad w_2 \approx \frac{N^2}{3} f_1^2 \quad w_3 \approx \frac{N^3}{4} f_1^3 \quad (8.6.2)$$

Рассмотрим возможность аппроксимации $w(t)$ при большом N каким-либо удобным в расчетах и легко подбираемым по методу моментов распределением. Проще всего воспользоваться гамма-распределением

$$w(t) = \frac{\mu(\mu t)^{r-1}}{\Gamma(r)} e^{-\mu t},$$

параметры которого выражаются через среднее значение и дисперсию по формулам $\mu = w_1/D_w$, $r = w_1^2/D_w$. Но $D_w = w_2 - w_1^2 = (N f_1^2)/12$. Значит, $\mu = 6/(N f_1)$, а $r = 3$. Следовательно, аппроксимирующим по двум моментам является распределение Эрланга третьего порядка. Третий момент этого распределения $w_3 = 3 \cdot 4 \cdot 5/\mu^3 = (N f_1)^3/36$ всего на 10% отличается от найденного из системы (8.6.2). Значит, эрланговское распределение третьего порядка в рассмотренном предельном случае практически обеспечивает выравнивание трех моментов искомого распределения.

8.6.2. Учет организации рабочего цикла

В некоторых задачах расчета восстанавливаемого ЗИП приходится делать поправку на разницу в режимах работы источников заявок органа восстановления, которая может быть весьма существенной. Так, в вычислительных центрах ремонт типовых элементов замены (ТЭЗов) обычно производится в течение одной смены, тогда как машина (мощная стационарная ЭВМ) работает две или три

Укажем способ расчета моментов необходимых распределений. Далее будут использоваться следующие обозначения

t_c — продолжительность ремонтной смены,

$F(t)$ — распределение чистого времени обслуживания,

$F_c(t)$ — распределение времени обслуживания с учетом сменности,

$\psi(\theta)$ — распределение момента прибытия заявки,

T — длительность цикла

Заметим, что из цикла полностью исключается время, в течение которого не работают ни система, ни ремонтный орган, так что при двухсменной работе системы можно считать $t_c = 8$, $T = 16$ часам.

Приход заявки в свободную систему. Рассмотрим три варианта обслуживания:

А — начатое и завершённое в течение рабочей смены,

Б — заявка поступила в перерыв,

В — ремонт разорван междусменным простоем.

В случае А моменты распределения ремонта с учетом сменности и соответствующие моменты распределения чистого ремонта совпадают

$$f_k^A = f_k, \quad k = 1, 2, \dots$$

В случае Б к чистому времени ремонта добавляется случайное время ожидания начала смены, равномерно распределенное на интервале $[0, T - t_c]$. Моменты распределения времени ожидания

$$f_k^{ож} = \frac{(T - t_c)^k}{k + 1}, \quad k = 1, 2, \dots,$$

а моменты свертки распределений ожидания и чистого ремонта

$$f_k^B = \sum_{j=0}^k \binom{k}{j} \frac{(T - t_c)^j}{j + 1} f_{k-j}, \quad k = 1, 2, \dots$$

В случае В время ожидания не случайно и в точности равно $T - t_c$, так что

$$f_k^B = \sum_{j=0}^k \binom{k}{j} (T - t_c)^j f_{k-j}, \quad k = 1, 2, \dots$$

Найдем вероятности соответствующих событий в предположении, что вероятность превышения чистой длительностью ремонта продолжительности смены пренебрежимо мала. Прежде всего,

$$\Pr\{A/A \cup B\} = \int_0^{t_c} \frac{1}{t_c} F(t_c - \theta) d\theta = \frac{1}{t_c} \int_0^{t_c} \int_0^{t_c - \theta} f(\tau) d\tau d\theta$$

$$\begin{aligned} &= \frac{1}{t_c} \int_0^{t_c} \left(\int_0^{t_c - \tau} d\theta \right) f(\tau) d\tau = \frac{1}{t_c} \int_0^{t_c} (t_c - \tau) f(\tau) d\tau \\ &\approx 1 - f_1/t_c. \end{aligned}$$

Совокупная вероятность вариантов А и В составляет t_c/T , а вероятность варианта Б дополняет ее до единицы. Следовательно,

$$P_A = \frac{t_c}{T} \left(1 - \frac{f_1}{t_c} \right) = \frac{t_c - f_1}{T}; \quad P_B = \frac{T - t_c}{T}; \quad P_B = \frac{t_c}{T} f_1/t_c = f_1/T.$$

С учетом этих вероятностей получаем моменты распределения времени обслуживания для заявки, поступающей в свободную систему:

$$\nu_k'' = \frac{1}{T} \left[(t_c - f_1) f_k + \sum_{j=0}^k \binom{k}{j} (T - t_c)^j f_{k-j} \left(\frac{T - t_c}{j + 1} + f_1 \right) \right], \quad k = 1, 2, \dots \quad (8.6.3)$$

Приход заявки в занятую систему. Для занятой системы начало обслуживания очередной заявки совпадает с окончанием обслуживания предыдущей и обязательно приходится на рабочую смену. Длительность ожидания начала обслуживания подчинена известному из теории восстановления распределению «перескока» с дополнительной функцией распределения

$$\bar{\Phi}(x) = \frac{1}{f_1^\Sigma} \int_x^\infty \bar{F}_\Sigma(u) du$$

и моментами

$$\varphi_k = \frac{f_{k+1}^\Sigma}{(k + 1) f_1^\Sigma}, \quad k = 1, 2, \dots$$

Индексы Σ относятся к средневзвешенным моментам суммарного потока.

Если начатое обслуживание заявки i -го типа успевает завершиться до конца рабочей смены (случай А), то моменты обслуживания с учетом сменности

$$f_{i,k}^A = \sum_{j=0}^k \binom{k}{j} \varphi_j f_{i,k-j}, \quad k = 1, 2, \dots$$

Вероятность этой ситуации

$$\pi_i = \int_0^{t_c} \varphi(x) F_i(t_c - x) dx = \frac{1}{f_1^{\Sigma}} \int_0^{t_c} \bar{F}_{\Sigma}(x) F_i(t_c - x) dx.$$

В противном случае (В) к времени обслуживания добавляется между-сменный простой:

$$f_{i,k}^B = \sum_{j=0}^k \binom{k}{j} f_{i,k-j}^A (T - t_c)^j, \quad k = 1, 2, \dots$$

Моменты распределения времени обслуживания для заявки, приходящей в занятую систему, суть

$$\nu'_{i,k} = \pi_i f_{i,k}^A + (1 - \pi_i) f_{i,k}^B, \quad k = 1, 2, \dots$$

Обслуживание произвольного требования. Получим распределение времени обслуживания для произвольного требования, считая ремонтный орган однолинейной системой массового обслуживания. Известно, что вероятность занятости такой системы

$$\rho = \sum_i \lambda_i \nu_{i,1}.$$

Очевидно, результирующие моменты

$$\nu_{i,k} = \rho \nu'_{i,k} + (1 - \rho) \nu''_{i,k}. \quad (8.6.4)$$

Умножая это уравнение при $k = 1$ на λ_i и суммируя по всем i , приходим к

$$\sum_i \lambda_i \nu_{i,1} = \rho \sum_i \lambda_i \nu'_{i,1} + (1 - \rho) \sum_i \lambda_i \nu''_{i,1}.$$

Но левая его часть равна ρ . Следовательно, коэффициент загрузки

$$\rho = \frac{\sum_i \lambda_i \nu''_{i,1}}{1 - \sum_i \lambda_i \nu'_{i,1} + \sum_i \lambda_i \nu''_{i,1}}. \quad (8.6.5)$$

Теперь можно вычислить искомые моменты согласно (8.6.4).

Обобщение на случай выходных дней. При непрерывной работе системы и наличии выходных дней у ремонтного органа (четырёхсменный режим) изложенная выше методика нуждается в дополнительном уточнении. Пусть производственный цикл системы ремонта содержит M периодов длины T , из них m выходных помещены в конце цикла (типичный случай $T = 1$ сут., $M = 7, m = 2$). Тогда для элемента, поступающего в свободную систему, моменты распределения времени ремонта с вероятностью $(M - m - 1)/M$ будут задаваться ранее полученным соотношением (8.6.3), а с вероятностью $(m + 1)/M$ — им же с заменой T на $(m + 1)T$. Таким образом, для четырёхсменного режима результирующие моменты подсчитываются согласно

$$\nu''_{k,i} = \frac{M - m - 1}{M} \nu''_{k,i}(T) + \frac{m + 1}{M} \nu''_{k,i}((m + 1)T).$$

Аналогично обстоит дело и для занятой системы ремонта.

8.7. Политика заказов и процедуры управления

Основные типы стратегий управления запасами обсуждались в главе 1. Здесь мы обсудим некоторые их особенности применительно к иерархическим системам снабжения. Прежде всего отметим, что в практически используемых методиках обычно ограничиваются разумными аппроксимациями, которые можно реально обсчитывать.

«Запрос-заказ» $(S - 1, S)$ фактически транслирует спрос уровнем выше, где он суммируется. При пуассоновских первичных потоках результирующий поток также оказывается пуассоновским с суммарной интенсивностью во всех вышестоящих звеньях. Это центральная (очень простая и вполне обоснованная) методика для дорогостоящих узлов с низким спросом. Названная стратегия количеством заказов не управляет — оно полностью определяется фактическим спросом. Ее дальнейшие варианты: расходные поставлять сразу или периодически; ремонтируемые чинить сразу или в соответствии с рабочим циклом.

Смешанная стратегия на нижних уровнях системы использует правило $(S - 1, S)$, а в высших звеньях (s, S) . Она может применяться как к расходным, так и к ремонтируемым изделиям. Пуассоновский первичный спрос сохраняется до первого уровня с правилом (s, S) включительно.

Общая пороговая (s, S) с непрерывным просмотром на всех уровнях, кроме нижнего, предполагает групповые заявки (на уровнях выше второго — случайного объема, поскольку групповые заявки порождают случайный уровень запаса в момент заказа восполнения). В практических методиках частичные поставки обычно исключаются. Все длительности поставок принимаются постоянными. Дефицит считается накапливаемым. Эта стратегия рекомендуется для расходуемых запчастей. Для восстанавливаемых она целесообразна при большом удалении ремонтного органа и высокой цене транспортировки (в системах глобального масштаба).

Практически и с точки зрения удобства анализа предпочтительна унификация правил заказа (хотя бы по эшелонам системы снабжения).

8.8. Параметры моделей

Перечислим параметры моделей систем обеспечения запчастями с их обозначениями.

Детали. w_i — вес, v_i — объем, h_i — цена хранения, d_i — цена недостачи, λ_i — интенсивность отказов (в нестационарном случае задается распределением Вейбулла), μ_i — средняя скорость ремонта (могут быть заданы пара $\{p_{in}, T_{in}\}$ — вероятность и длительность ремонта типа n).

Склады характеризуются следующим набором параметров:

- Число номенклатур I , по эшелонам $I_{m+1} > I_m$. При составлении списка сначала перечисляют детали, хранящиеся на всех уровнях; затем без первого, без первых двух и т.д.
- Ограничения по весу \bar{W}_{jm} , по объему \bar{V}_{jm} , по затратам \bar{C}_{jm} .
- Интенсивность спроса λ_{ijm} на складе j уровня m .
- Ожидаемое время задержки поставки L и его дисперсия D с той же индексацией (ijm) .
- Среднее время восстановления T_{ijm} .
- Задержка экстренной поставки (emergency provision) E_{ijm} .
- Вероятность r_{ijm} ремонтируемости i в звене (jm) .
- s_{ijm} — уровень заказа, S_{ijm} — верхний уровень, T_{im} — периодичность просмотра.

Система в целом определяется числом уровней, количеством складов на каждом, поставщиком каждого (его уровнем и номером). Бюджетные ограничения обозначаются C . Они могут быть детализованы по эшелонам или номенклатурам. Для отдельных складов, эшелонов системы или номенклатур могут быть заданы ограничения снизу на качество обеспечения.

Окружение системы задается количеством потребителей N_{jm} , интенсивностью их работы, рабочими температурами и т.д., дополнительными параметрами (например, уровнями разборки изделий). Если деталь входит в несколько сборок, то возникают дополнительные трудности.

8.9. Проблема штрафов для военных систем

Обычный подход к расчету штрафов связан с оценкой снижения коэффициента готовности обеспечиваемой техники. Среднее время восстановления изделия определяется как средневзвешенное по всем типовым отказам (с учетом ожидаемой частоты их). Тогда установившееся значения коэффициента готовности аппаратуры можно представить в виде

$$K = \frac{T_0}{T_0 + T_r} \frac{T_0 + T_r}{T_0 + T_r + T_z} = K_1 K_2,$$

где T_0 , T_r , T_z — средние длительности безотказной работы, чистой длительности ремонта и задержки ремонта из-за нехватки ЗИПа соответственно, K_1 дает коэффициент готовности при неограниченном ЗИПе, а K_2 — его снижение из-за ограниченности последнего.

В ответственных случаях расчет штрафов из-за дефицита запасных частей усложняется разнообразием последствий дефицита и требует специфического подхода к каждой конкретной ситуации. Проблема особенно остро ставится для систем военного назначения — вплоть до принципиального отрицания такой возможности и, следовательно, применения теории запасов в военном деле вообще. Эта ситуация вполне вписывается в классическую схему «Кто хочет работать, ищет возможности, кто не хочет — причины». В порядке «поиска возможностей» мы рассмотрим две задачи о расчете штрафов старую (авиация) и сравнительно новую (ракетные комплексы).

8.9.1. Авиация

Простейшее решение задачи сводится к стоимости аренды нового самолета взамен отказавшего. Например, в [89, с. 255] на указанной основе недельный простой самолета из-за отсутствия запасной части оценивается в 10 тыс. долл.

Более развитый подход к этой ситуации — также из опыта снабжения ВВС — имеется у Карра [165]. В этой работе предлагалось считать «полезность» детали функцией двух переменных: влияния неисправности данной детали на боеспособность самолета и возможности компенсировать ее отсутствие. Каждый из этих факторов имел пять градаций, например, первый из них

- 1) самолет не может летать, и при перебазировании эскадрильи его придется уничтожить,
- 2) самолет может летать, но только при абсолютной необходимости и с большим риском для экипажа,
- 3) самолет может летать, но его боевое использование допустимо лишь в критической оперативной обстановке,
- 4) самолет может летать, но его боевые возможности несколько ограничены,
- 5) отсутствие данной детали практически не влияет на боеспособность самолета.

Комбинируя эти оценки с характеристиками детали, ее относили к одному из пяти классов важности (табл. 8.5)

Т а б л и ц а 8.5

Разбивка деталей на классы важности

Факторы		Влияние на боеспособность		
		сильное	умеренное	слабое
Возможность замены	отсутствует	I	II	V
	имеется	III	IV	

Оценка каждого фактора, определяющего важность, производилась путем обработки суждений техников — специалистов по конкретным самолетным системам. Для первого класса величину штрафа предлагалось

принять равной стоимости возмещения (аренда самолета того же класса) — 1000 долл. в день, для второго — порядка 300 долл., пятого — 10 долл. Для деталей третьего и четвертого классов штраф определялся суммарными затратами на необходимые материалы, рабочую силу и прикрытие на время ремонта и мог достигать 100 долл.

В результате был предложен новый табель снабжения, оказавшийся весьма экономичным.

При стохастическом спросе с функцией распределения, определенной для бесконечно больших значений аргумента, при любом конечном запасе существует положительная вероятность недостачи, и принятие цены штрафа $d = \infty$ привело бы к бесконечно большим ожидаемым расходам. В подобных случаях часто вводят ограничение по вероятности недостачи. Необходимость экономии на запасах приводит к тому, что вероятность недостачи оказывается в точности равной граничному значению (по крайней мере для непрерывно распределенного спроса). Таким образом, задание вероятности обеспечения полностью определяет решение задачи. Следует, однако, заметить, что обоснование этой цифры невозможно без хотя бы приближенной оценки экономических последствий недостачи, т.е. того же самого штрафа. Сложность этой задачи ни в коей мере не оправдывает субъективного назначения соответствующих вероятностей. Другим возможным подходом является минимизация вероятности или ожидаемого объема дефицита при ограниченных расходах на снабжение.

8.9.2. Ракетные комплексы

Военные системы снабжения всегда были склонны к раздуванию запасов. Как отмечалось в [89, с. 485],

«основная задача командования военной базы состоит в том, чтобы все самолеты были готовы к полетам (корабли — к походам). За неготовность больно бьют, за избыточные запасы мягко критикуют. В результате на базе будет скапливаться намного больше запасных частей и материалов, чем на самом деле необходимо».

Далее (с. 487) в той же книге утверждается, что

«в военных системах снабжения почти невозможно проводить оценки в терминах стоимости, так как многие виды

потерь непосредственно измерить нельзя (например, штраф за неготовность межконтинентальной баллистической ракеты из-за отсутствия какой-либо запасной части)».

Рассматривая это заявление как вызов, попытаемся наметить схему приближенного расчета цены штрафа в упомянутой ситуации на основе информации из хорошо известных специалистам и далеко не новых *открытых* источников [1, с. 323–325], [6, с. 329–365], [82].

Для начала запишем дифференциальные уравнения Ланчестера, определяющие непрерывную модель динамики ракетной дуэли, при следующих базовых предположениях:

- Каждая боевая единица Красных может вести огонь по каждой боевой единице Синих, и наоборот
- Огонь является прицельным (по конкретной единице); одним выстрелом может быть поражена только одна единица
- Эффективная скорострельность $\lambda^* = \lambda \cdot p$, где p — вероятность поражения цели. Скорострельность λ учитывает время расчета прицельных данных, прицеливание, перезарядание и т.п.
- Обстрелу с одинаковой вероятностью подвергается любая из еще не пораженных единиц, после поражения единицы огонь по ней прекращается и немедленно переносится на еще не пораженные единицы.
- Пораженная единица прекращает стрельбу и в дальнейших боевых действиях не участвует.

Модель с переносом огня. Здесь обстреливаются только непораженные цели (иначе говоря, имеется полная информация о состоянии противной стороны). Модель описывается уравнениями

$$\begin{aligned} dm_1(t)/dt &= -\lambda_2 p_2 m_2(t), \\ dm_2(t)/dt &= -\lambda_1 p_1 m_1(t) \end{aligned} \quad (8.9.1)$$

с решением [6]

$$\begin{aligned} m_1(t) &= m_1(0) \cosh \sqrt{\lambda_1 \lambda_2} t - m_2(0) \sqrt{\lambda_2 \lambda_1} \sinh \sqrt{\lambda_1 \lambda_2} t, \\ m_2(t) &= m_2(0) \cosh \sqrt{\lambda_1 \lambda_2} t - m_1(0) \sqrt{\lambda_1 \lambda_2} \sinh \sqrt{\lambda_1 \lambda_2} t. \end{aligned} \quad (8.9.2)$$

Эти уравнения справедливы при не слишком малых значениях численностей сторон. Количество сохранившихся единиц более сильной стороны

$$n = n_0 \sqrt{1 - \Phi' / \Phi''},$$

где Φ' — потенциал проигравшей (Φ'' — выигравшей) стороны. Для рассматриваемой модели боя (квадратичной) потенциал вычисляется как

$$\Phi = p \lambda n_0^2.$$

Модель без переноса огня (линейная). Здесь

$$\begin{aligned} dm_1(t)/dt &= -\frac{\lambda_2 p_2}{m_1(0)} m_1(t) m_2(t), \\ dm_2(t)/dt &= -\frac{\lambda_1 p_1}{m_2(0)} m_1(t) m_2(t). \end{aligned} \quad (8.9.3)$$

Решение имеет вид

$$\begin{aligned} m_1(t) &= m_1(0) \frac{u_2 - u_1}{u_2 e^{-(u_2 - u_1)t} - u_1}, \\ m_2(t) &= m_2(0) \frac{u_1 - u_2}{u_1 e^{-(u_1 - u_2)t} - u_2}. \end{aligned} \quad (8.9.4)$$

Вспомогательные константы $u_1 = \lambda_1 m_1(0) / m_2(0)$, $u_2 = \lambda_2 m_2(0) / m_1(0)$. Для линейной модели

$$n = n_0 (1 - \Phi_m / \Phi_n),$$

а потенциал пропорционален *первой* степени числа боевых единиц:

$$\Phi = p \lambda n_0.$$

Итоговая оценка. Поскольку исход ракетной дуэли определяется потенциалом сторон, естественно оценивать ущерб от дефицита некоторой запчасты формулой вида

$$h = C \Delta \Phi,$$

где C — приведенные годовые затраты на создание и поддержание упомянутого потенциала, а $\Delta \Phi$ — его снижение из-за дефицита (с учетом времени τ существования последнего).

Для заинтересованных в конкретных числовых оценках приведем разумные исходные данные. Согласно [82, с. 55], годовые затраты на стратегические ядерные силы США в восьмидесятые годы оценивались в 50 млрд. долл. Снижение потенциала для модели с перенацеливанием

в линейном приближении можно оценить, пользуясь производными

$$\partial\Phi/\partial p = \lambda n^2,$$

$$\partial\Phi/\partial \lambda = pn^2,$$

$$\partial\Phi/\partial n = 2p\lambda n.$$

Выбор производной определяется последствиями дефицита конкретной детали:

- уменьшение числа стартующих ракет — для отказов, совершенно исключающих запуск (в частности, групповой);
- уменьшение скорострельности — для системы прицеливания (например, в случае подвижных стартов);
- уменьшение вероятности поражения цели — если дефицит уменьшает мощность взрыва q или увеличивает среднеквадратическое отклонение точки падения боевой части σ .

Для подсчета эффектов двух последних видов следует воспользоваться формулой вероятности поражения точечной цели из [82, с. 32–34] и результатами ее дифференцирования:

$$p = 1 - \exp\left(-\frac{kq^{2/3}}{2\sigma^2}\right),$$

$$\partial p/\partial q = \frac{k}{3q^{1/3}\sigma^2} \exp\left(-\frac{kq^{2/3}}{2\sigma^2}\right),$$

$$\partial p/\partial \sigma = -\frac{kq^{2/3}}{\sigma^3} \exp\left(-\frac{kq^{2/3}}{2\sigma^2}\right).$$

Согласно данным той же книги [82, с. 10–11], для ракет Минитмен-3 тротильный эквивалент боеголовки $q=0.2-0.4$ Мт, предельное отклонение точки падения (2.7σ) при стрельбе на максимальную дальность составляет 0.5 км, откуда $\sigma=180$ м. Коэффициент прочности цели k определяется как расстояние в метрах от центра взрыва мощностью 1 Мт, при котором во фронте ударной волны создается избыточное давление в 7 МПа (10 кгс/см^2).

Обобщения и ограничения. Подобным же образом могут быть учтены последствия отказа в средствах преодоления противоракетной обороны противника.

Модель с учетом времени полета снаряда и запаздыванием в переносе огня позволяет учесть задержки в системе раннего предупреждения о ракетном нападении противника и системе космического наблюдения за его ракетно-ядерными силами. Эта модель определяется уравнениями

$$\frac{dm_1(t)}{dt} = -p_2\lambda_2(t-\tau_2)m_2(t-\tau_2)\frac{m_1(t)}{m_1(t-\theta_1)},$$

$$\frac{dm_2(t)}{dt} = -p_1\lambda_1(t-\tau_1)m_1(t-\tau_1)\frac{m_2(t)}{m_2(t-\theta_2)}.$$

Здесь τ_i, θ_i — соответственно время полета и запаздывание в перенацеливании для i -й стороны.

Описанные подходы являются сугубо приближенными хотя бы потому, что «ракетная дуэль» не является непрерывным процессом и состоит в обмене *дискретными* ударами. Кроме того, не учитывается различие целей в зависимости от стратегической обстановки, так как различным формам удара соответствуют в общем случае разные цели. Если, например, в совокупность целей *упреждающего* удара ракетные базы противника включаются в качестве первоочередных, то для *ответного* удара их включение в перечень целей может оказаться необоснованным, поскольку ракеты с этих баз к моменту нанесения ответного удара уже стартуют [82, с. 23]. Тем не менее открывается возможность получить хоть какие-то количественные оценки, имеющие безусловно положительную корреляцию с реальностью и позволяющие, в частности, определить относительную важность отдельных подсистем ракетного вооружения. Напомним кстати, что для выбора оптимального запаса цена штрафа может быть задана с большой относительной погрешностью.

Глава 9

Локальная задача о восстанавливаемом ЗИПе

9.1. Предварительные соображения

Одной из важнейших задач эксплуатации сложной техники является расчет восстанавливаемого ЗИПа. Как правило, экспресс-диагностика состояния сложной техники проводится с точностью до входящего в нее блока (узла, агрегата). С целью минимизации простоя комплекса неисправный блок немедленно заменяется взятым из ЗИПа, а сам направляется в ремонтный орган. Отремонтированный агрегат пополняет собою ЗИП. Подобная схема широко применяется в ситуациях, где эксплуатируется большое число однородных агрегатов: типовые элементы замены (ТЭЗ) мощных вычислительных машин, самолетные двигатели в крупном авиапредприятии, диагностическая и физиотерапевтическая аппаратура в больницах и госпиталях, оборудование электровозов и систем управления движением на железных дорогах и в метрополитене, кассовые аппараты в больших универмагах. Растет ее популярность и в современных системах бытового сервиса, которые на время ремонта отказавшего сложного устройства (персонального компьютера, видеомагнитофона, телевизора, аудио-системы, холодильника, СВЧ-печи, автомобиля и т.п.) предоставляют пользователю эквивалентную замену.

Возможны случаи, когда недопоставка приводит к замедлению производства или снижению ожидаемого объема его. Обслуживание мощной стационарной ЭВМ коллективного пользования (в современной терминологии — mainframe) является типичной ситуацией именно этого рода. Практически при выходе из строя какого-либо ТЭЗа его немедленно заменяют исправным из ЗИПа, а отказавший поступает в ремонт, после завершения которого пополняет ЗИП. Недопоставка элементов приводит к тому, что запас резервных ТЭЗов начинает уменьшаться, а вероятность остановки машины из-за их отсутствия — возрастать. Таким образом, здесь в качестве штрафа в единицу времени целесообразно принять произведение стоимости машинного часа на прирост стационарной вероятности недостачи ТЭЗов.

Недостаточность ЗИПа увеличивает простои пользователей, и система поддержки эксплуатации (для упрощения терминологии мы сохраним за ней наименование системы снабжения) по справедливости должна компенсировать им материальный и моральный ущерб, выплачивая «штрафы». С другой стороны, избыточный ЗИП омертвляет значительные денежные средства и требует расходов на хранение, которые в связи со сложностью входящих в него агрегатов могут быть существенно больше расходов на хранение обычных предметов снабжения. Таким образом, весьма актуальна задача об оптимальном выборе объема восстанавливаемого ЗИПа. Специфика этой задачи отражается в особой структуре целевой функции и в том, что ожидаемое снижение объема ЗИП приходится рассчитывать методами теории массового обслуживания.

Сразу же отметим, что агрегатная система ремонта должна постоянно подпитываться расходными деталями и материалами, организация поставок которых рассчитывается обычными методами теории запасов. Укажем также, что в контур задачи может включаться и оптимизация восстанавливающей системы: логично поставить вопрос о выборе такой производительности восстанавливающих устройств и их количества, при которых минимальны суммарные затраты на упомянутую систему и на ЗИП. В этом случае ЗИП оптимизируется при фиксированных параметрах восстанавливающей системы, а глобальная оптимизация достигается направленным перебором значений последних, выбираемых из дискретных рядов возможностей (числа «каналов» восстановления, типов диагностических и технологических устройств, сменности работы).

При решении задачи о восстанавливаемом ЗИП исследователь имеет дело с комплексом «снабжаемый агрегат — ЗИП — ремонтный

орган» С точки зрения восполнения ЗИПа, временем задержки поставок является не чистое время восстановления, а *полное* время нахождения элемента в системе восстановления, совпадающее с длительностью собственно восстановления лишь при отсутствии очереди. Разрешение существования очереди является экономически необходимым условием достаточной загрузки системы ремонта, так что при планировании восстанавливаемого ЗИПа пренебрегать возможностью очереди в системе ремонта нельзя. Это обстоятельство определяет необходимость совместного рассмотрения систем «запас» и «ремонт» в следующем порядке

- 1) расчет вероятностных характеристик очереди на ремонт,
- 2) расчет распределения времени восстановления элемента с учетом очереди,
- 3) расчет вероятностей наличия запаса,
- 4) выбор оптимального запаса

Кроме того, в задачах о восстанавливаемом ЗИПе

- плата за хранение исчисляется по максимальному комплекту ЗИП, а не по числу исправных элементов,
- сумма штрафа пропорциональна времени простоя обеспечиваемого агрегата,
- простой агрегата фиксируется при снижении запаса до отрицательной величины,
- при достаточно большом дефиците (в частном случае — сразу при его возникновении) поток отказов может прерываться

Классическим подходом к применению ТМО в управлении запасами является рассмотрение дискретных единиц последнего как каналов обслуживания. Истребование изделия рассматривается как занятие канала, при стратегии $(S-1, S)$ каждая заявка вызывает заказ на восстановление (ремонт), длительность (задержка) которого интерпретируется как время обслуживания. Этот подход дает основание использовать для расчета средней задержки восполнения формулу Литтла (отношение среднего числа «занятых каналов» к интенсивности входящего потока)

Употребительны два варианта моделей с неограниченной очередью (допускается дефицит) и с отказами. Примером последней может

служить $M/M/1/S$. Здесь в состоянии «0» запас равен S . В состоянии « $S+1$ » отказавшую деталь заменить нечем, и система простаивает. Вероятность простоя

$$p_{S+1} = \rho^{S+1}(1 - \rho^{S+2})/(1 - \rho), \quad \rho = \lambda/\mu.$$

Модели различаются также типами входящего потока и распределений длительности обслуживания. Недостатком этого подхода является трудность расчета многоканальных СМО с немарковским обслуживанием (см. главу 3).

Особым случаем является часто используемая без явных ссылок модель $M/G/\infty$, где очередь и отказы не возникают вследствие бесконечного числа каналов.

9.2. Основная модель

Будем считать, что штраф за недостачу ЗИПа пропорционален среднему числу пользователей, простаивающих в единицу времени. Очевидно, вложенные в ЗИП средства омертвлены независимо от того, исправны его элементы или находятся в ремонте. Таким образом, функция затрат в единицу времени, связанная с ЗИПом объема s , принимает вид

$$L(s) = hs + d \sum_{k=s+1}^{\infty} (k-s)p_k, \quad (9.2.1)$$

где h — цена хранения; d — цена штрафа, p_k — стационарная вероятность нахождения в ремонте k агрегатов. Общие условия (5.6.13) оптимальности дискретного запаса s^* с помощью (9.2.1) преобразуются в неравенства

$$\begin{aligned} h - d \sum_{k=s^*+1}^{\infty} p_k &\geq 0, \\ d \sum_{k=s^*}^{\infty} p_k - h &\geq 0. \end{aligned}$$

Их можно объединить в

$$\sum_{k=s^*+1}^{\infty} p_k \leq h/d \leq \sum_{k=s^*}^{\infty} p_k. \quad (9.2.2)$$

Если штраф определяется самим фактом недостачи и не зависит от ее объема (это бывает при обеспечении одной системы, нормальная работа которой при недостатке ЗИПа не гарантируется), целевую функцию следует записать в виде

$$L(s) = hs + d \sum_{k=s+1}^{\infty} p_k. \quad (9.2.3)$$

Здесь условием оптимальности будет

$$p_{s^*+1} \leq h/d \leq p_{s^*}. \quad (9.2.4)$$

Неравенства (9.2.2) и (9.2.4) были получены безотносительно к организации ремонта и характеристикам потока заявок и в этом смысле являются универсальными. Конкретное распределение $\{p_k\}$ вычисляется методами теории массового обслуживания — см. главу 3

При простейших допущениях (система $M/M/1$) удается получить явные версии условий (9.2.2) и (9.2.4). В этом случае при коэффициенте загрузки системы восстановления $\rho = \lambda/\mu$ стационарные вероятности $\{p_k\}$ подсчитываются согласно

$$p_k = (1 - \rho)\rho^k, \quad k = 0, 1, \dots \quad (9.2.5)$$

Соответственно при штрафе по ожидаемому дефициту функция затрат

$$L(s) = hs + d\rho^{s+1}/(1 - \rho), \quad (9.2.6)$$

а оптимальный объем ЗИПа дается формулой

$$s^* = \left\lfloor \frac{\ln(d/h)}{\ln(1/\rho)} \right\rfloor. \quad (9.2.7)$$

При штрафе по вероятности недостачи имеем функцию затрат

$$L(s) = hs + d\rho^{s+1}, \quad (9.2.8)$$

а оптимальный ЗИП

$$s^* = \left\lfloor \frac{\ln(d(1 - \rho)/h)}{\ln(1/\rho)} \right\rfloor. \quad (9.2.9)$$

В обоих случаях берется целая часть взятых в скобки отношений логарифмов. Логарифмы могут быть взяты по любому основанию.

Пример. Предположим, что требуется оценить экономичность оборотного запаса $s=10$ штук восстанавливаемых узлов авиационного двигателя при стоимости детали 600 долл., среднем времени восстановления $\mu^{-1}=12$ дней и интенсивности потока отказов $\lambda = 0.02$ в день. Стоимость годового хранения примем $h = 150$ долл. Штраф будем исчислять по ожидаемой недостаче. В [165] удельный «штраф» за простой самолета определен как стоимость аренды нового в сумме 1 тыс. долл. в день, или 365 тыс. долл. в год. Примем $d=350$ тыс. долл. Тогда годовые ожидаемые затраты при плановом запасе 10 штук

$$\begin{aligned} L(10) &= 150 \cdot 10 + 350 \cdot 10^3 \frac{(0.02 \cdot 12)^{11}}{1 - 0.02 \cdot 12} \\ &= 1500 + 0.35 \cdot 10^6 \frac{1.52 \cdot 10^{-7}}{0.76} = 1500.07 \text{ долл.} \end{aligned}$$

При тех же данных оптимальный запас составит целую часть от

$$\tilde{s} = \frac{\lg(350 \cdot 10^3/150)}{\lg(0.0833/0.02)} = \frac{\lg 2330}{\lg 4.16} = 3.367/0.620 = 5.44,$$

т.е. 5 штук, а ожидаемые затраты в 1.79 раза меньше.

$$\begin{aligned} L(5) &= 150 \cdot 5 + 350 \cdot 10^3 \frac{(0.02 \cdot 12)^6}{1 - 0.02 \cdot 12} \\ &= 750 + 0.35 \cdot 10^6 \frac{1.91 \cdot 10^{-4}}{0.76} = 750 + 88 = 838 \text{ долл.} \end{aligned}$$

Следует подчеркнуть, что распределение числа заявок в системе массового обслуживания даже при равенстве средних значений чрезвычайно чувствительно к виду распределения длительности обслуживания. На рис. 9.1 и 9.2 представлены графики распределения и ДФР числа заявок в системе $M/G/1$ для загрузки $\rho = 0.7$ и гамма-распределения длительности обслуживания при указанных на кривых коэффициентах вариации. Случай $v = 0$ соответствует регулярному обслуживанию

Графики демонстрируют большой разброс определяющих выбор оптимального запаса точек пересечения кривых с горизонталью, проведенной на уровне отношения h/d . Это указывает на обязательность учета дисперсии распределений длительности обслуживания. Разброс быстро возрастает с уменьшением упомянутого отношения.

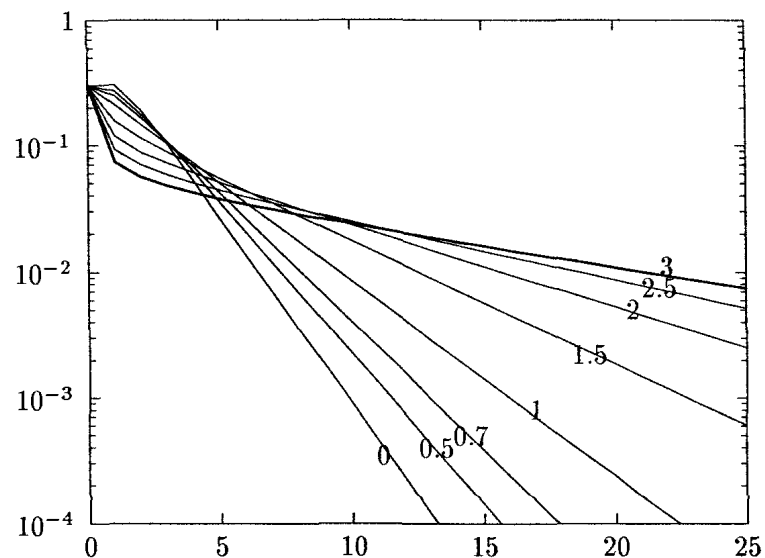


Рис. 9.1 Распределение числа заявок в системе $M/G/1$

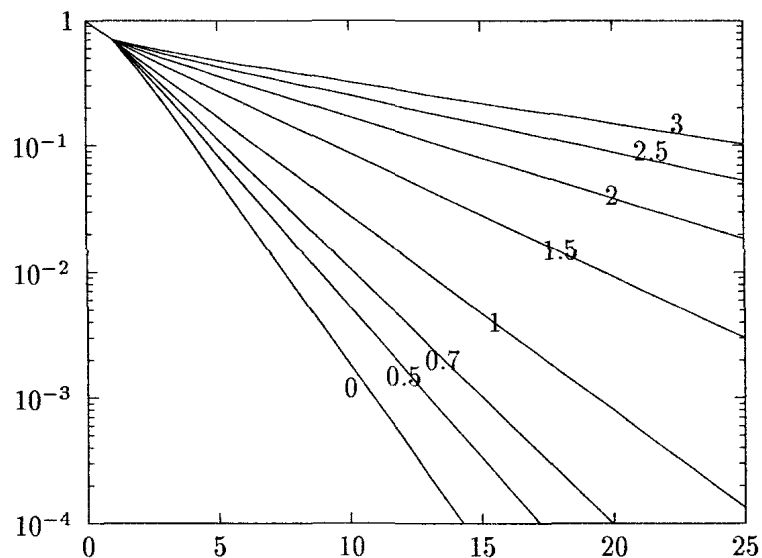


Рис. 9.2. Распределение сумм вероятностей в системе $M/G/1$

9.3. Многолинейная марковская система

Естественно считать $s > n$. В этой области согласно (3.12.2) вероятность дефицита

$$\varphi^{(1)}(s) = p_n \sum_{i=s+1}^{\infty} \rho^{i-n} = p_n \frac{\rho^{s+1-n}}{1-\rho}.$$

Следовательно, функция затрат

$$L(s) = hy + dp_n \frac{\rho^{s+1-n}}{1-\rho}.$$

Приращение вероятности дефицита в случае замены s на $s+1$

$$\Delta\varphi^{(1)}(s) = \frac{p_n}{1-\rho} [\rho^{s+2-n} - \rho^{s+1-n}] = -p_n \rho^{s+1-n}. \quad (9.3.1)$$

Значит, оптимальный запас при известном ρ должен определяться из неравенств

$$p_n \rho^{s+1-n} \leq h/d \leq p_n \rho^{s-n},$$

откуда

$$s_n^{(1)} = n + \left\lfloor \frac{\ln(dp_n/h)}{\ln(1/\rho)} \right\rfloor.$$

При исчислении штрафов по вероятности дефицита последняя в случае $s > n$ вычисляется согласно

$$\varphi^{(2)}(s) = p_n \frac{\rho^{s+1-n}}{(1-\rho)^2}.$$

Следовательно, функция затрат

$$L(s) = hy + dp_n \frac{\rho^{s+1-n}}{(1-\rho)^2}.$$

Приращение ожидаемого дефицита

$$\Delta\varphi^{(2)}(s) = \frac{p_n}{(1-\rho)^2} \rho^{s+1-n} (\rho - 1) = -\frac{p_n}{1-\rho} \rho^{s+1-n}.$$

Оптимальный запас вычисляется по формуле

$$s_n^{(2)} = n + \left\lfloor \frac{\ln \frac{dp_n}{h(1-\rho)}}{\ln(1/\rho)} \right\rfloor.$$

9.4. Учет дополнительных задержек

При заметной удаленности ремонтного органа следует учитывать дополнительное снижение объема ЗИПа за время T доставки агрегата в ремонт и обратно. При простейшем потоке заявок это распределение для фиксированного T подчинено закону Пуассона

$$a_j = \frac{(\lambda T)^j}{j!} e^{-\lambda T}, \quad j = 0, 1, \dots,$$

а для случайного с плотностью распределения $v(t)$ считается как

$$a_j = \int_0^{\infty} \frac{(\lambda t)^j}{j!} e^{-\lambda t} v(t) dt, \quad j = 0, 1, \dots$$

Вопрос об эффективном расчете этих вероятностей рассматривался в разд. 3.6.3.

Результирующее распределение снижения ЗИПа получается сверткой распределений числа агрегатов, находящихся в транспортировке и непосредственно в ремонте. Эта рекомендация относится и ко всем моделям с восстанавливаемым ЗИПом, которые рассматриваются ниже.

9.5. Восстановление в сети

Восстановление отказавшего агрегата может потребовать ряда операций, выполняемых на специализированном оборудовании. При этом возможно совмещение ремонта нескольких агрегатов, проходящих различные фазы обслуживания. В таких случаях приходится иметь дело с *сетью обслуживания*, а снижение ЗИПа определяется числом заявок, находящихся в сети (включая очереди к ее узлам). Это обстоятельство, не меняя общего подхода и условий оптимальности, требует подстановки в последние распределения числа заявок в сети. Сети обслуживания рассчитываются методом потокэквивалентной декомпозиции (см. разд. 3.15), в котором после балансировки потоков выполняется независимый расчет распределений числа заявок в узлах сети. Эти распределения нужно последовательно свернуть по узлам — см. формулу (3.4.9).

9.6. Оптимизация ЗИП по затратам за время эксплуатации

Выше обсуждалась минимизация затрат в единицу времени. При малом сроке службы техники приходится принимать в расчет и их распределение по годам. Наиболее обоснованным показателем является взвешенная сумма затрат за весь срок эксплуатации

$$L_{\Sigma} = \sum_{i=0}^N \alpha^i L_i, \quad (9.6.1)$$

где L_0 — затраты на создание ЗИП, а $\{L_i\}$, $i = \overline{1, N}$ — затраты в соответствующие годы эксплуатации. Полагая последние постоянными и вычисляя их по формулам предыдущего раздела, убеждаемся, что затраты и оптимальные решения можно определять по тем же формулам с заменами

$$\begin{aligned} \tilde{h} &= c + \frac{\alpha h}{1 - \alpha} (1 - \alpha^N), \\ \tilde{d} &= \frac{\alpha d}{1 - \alpha} (1 - \alpha^N). \end{aligned}$$

Здесь c — стоимость предварительной закупки одного элемента ЗИПа.

9.7. Восстановление с порогом включения и разогревом

В ряде систем массового обслуживания, в особенности при относительно малой загрузке, оказывается целесообразным введение «порога включения» — обслуживание начинается при скоплении в системе $r > 1$ заявок и заканчивается при полном освобождении системы. Такой режим увеличивает как период непрерывной занятости, так и время, в течение которого обслуживание не ведется. Это на относительно длительные периоды позволяет переводить автоматическую аппаратуру — в целях экономии ресурса и электроэнергии — в облегченный (дежурный) режим. В системах с участием человека появляется возможность полностью выключать значительную часть техники и переводить оператора на решение других задач.

Очевидно, что обслуживание первой заявки в этих условиях (а практически даже при $r = 1$) сопряжено с выполнением некоторых дополнительных операций («разогревом» системы в прямом и/или переносном смысле) и в среднем будет продолжаться дольше, чем обслуживание прочих заявок, причем изменение показателей оперативности обслуживания в общем случае не может быть сведено к постоянной задержке.

Ниже предлагается методика расчета системы $M/G/1$, позволяющая учитывать обе упомянутых особенности алгоритма ее функционирования.

9.7.1. Вложенная цепь Маркова

Аналогично разд. 3.13 выберем моменты регенерации процесса. Снова обозначим q_j вероятность прибытия j новых требований за время обслуживания с распределением $B(t)$. Эти вероятности определяются формулой (3.6.3), и в главе 3 описан способ их эффективного вычисления. Аналогично определим вероятности $\{q_j^*\}$ для распределения $B^*(t)$ обслуживания с разогревом. Для установившегося режима застать в системе $k < r-1$ требований можно, если в ней в предыдущий момент регенерации находилось $j = \overline{1, k+1}$ заявок, одна обслужилась и пришло еще $k+1-j \geq 0$ заявок. После разогрева вложенная цепь Маркова может оказаться только в состояниях $r-1, r, \dots$, причем до начала разогрева следует дождаться прибытия r заявок. Следовательно, финальные (предельные при устремлении номера момента регенерации к бесконечности) вероятности состояний вложенной цепи связаны системой уравнений

$$\pi_k = \sum_{j=1}^{k+1} \pi_j q_{k+1-j}, \quad k = \overline{0, r-2},$$

$$\pi_k = \sum_{j=1}^{k+1} \pi_j q_{k+1-j} + \pi_0 q_{k-r+1}^*, \quad k = r-1, r, r+1, \dots$$

Из этих формул выводим рекуррентные выражения для расчета вероятностей:

$$\pi_k = \left(\pi_{k-1} - \sum_{j=1}^{k-1} \pi_j q_{k-j} \right) / q_0, \quad k = \overline{1, r-1},$$

$$\pi_k = \left(\pi_{k-1} - \pi_0 q_{k-r}^* - \sum_{j=1}^{k-1} \pi_j q_{k-j} \right) / q_{k,0}, \quad (9.7.1)$$

$$k = r, r+1, \dots$$

9.7.2. Переход к стационарным вероятностям

Обозначим

α — среднюю частоту немарковских (по завершению обслуживания) переходов между состояниями системы,

b — среднюю длительность обслуживания,

b^* — среднюю длительность обслуживания с разогревом.

Средний интервал между немарковскими переходами

$$\alpha^{-1} = (r/\lambda + b^*)\pi_0 + b(1 - \pi_0). \quad (9.7.2)$$

Свяжем немарковские переходы с моментами регенерации вложенной цепи, для которых рассчитываются $\{\pi_k\}$. Параметр λ_k марковских переходов (прибытие новых заявок) при наличии в системе ровно k заявок может быть отнесен к произвольному моменту и должен связываться со стационарной вероятностью p_k . Баланс переходов между смежными состояниями системы в стационарном режиме требует равенства

$$\lambda p_k = \alpha \pi_k, \quad k = 1, 2, \dots \quad (9.7.3)$$

Частоты включения и выключения режима обслуживания равны:

$$\lambda p_{r-1}^* = \alpha \pi_0.$$

Поскольку вероятности $\{p_k^*\}$ всех r состояний с отключенным обслуживанием равны между собой, его можно заменить на

$$\lambda p_0^* = \alpha \pi_0. \quad (9.7.4)$$

Определим суммарную вероятность свободного состояния p^* . Она равна отношению средней длительности свободного состояния r/λ к сумме этой длительности и средней длины T_B периода непрерывной занятости (ПНЗ). В среднюю длину ПНЗ входят:

- средняя длительность разогрева b^* ;
- длительность обслуживания остальных $r - 1$ заявок, формирующих порог включения — в среднем $(r - 1)b$;
- длительность обслуживания вновь прибывших заявок — в среднем $b\lambda T_B$.

Итак, $T_B = b^* + b(r - 1 + \lambda T_B)$, откуда

$$T_B = \frac{b^* + (r - 1)b}{1 - \lambda b}. \quad (9.7.5)$$

Общая вероятность свободного состояния

$$p^* = \frac{r/\lambda}{r/\lambda + T_B} = \frac{r}{r + \lambda T_B},$$

а p_0^* — в r раз меньше.

Подставляя полученные выражения в (9.7.4), приходим к уравнению

$$\pi_0 = \frac{\lambda}{r + \lambda \frac{b^* + (r - 1)b}{1 - \lambda b}} \left[\left(\frac{r}{\lambda} + b^* \right) \pi_0 + (1 - \pi_0)b \right]$$

с решением

$$\pi_0 = \frac{1 - \lambda b}{r + \lambda(b^* - b)}.$$

Подстановка его в (9.7.2) дает $\alpha = \lambda$. Таким образом, рассматриваемая нами система, как и ее простейший аналог — обычная $M/G/1$, при сделанном выборе моментов регенерации процесса представима вложенной цепью Маркова.

Итак, расчет стационарных вероятностей состояний для нашей системы должен выполняться в следующем порядке:

- 1) Вычислить среднюю длину периода непрерывной занятости согласно (9.7.5).
- 2) Положить $p_0^* = \pi_0 = 1/(r + \lambda T_B)$.
- 3) По формулам (9.7.2) найти вероятности состояний для режима со включенным обслуживанием.
- 4) К найденным вероятностям для $k = \overline{1, r - 1}$ добавить p_0^* .

9.7.3. Распределение времени пребывания

Используя закон сохранения стационарной очереди, можно показать (см. разд. 3.11), что ПЛС $\nu(s)$ распределения времени пребывания заявки в системе с дисциплиной обслуживания FCFS связано с производящей функцией $P(z)$ распределения числа заявок в системе формулой

$$\nu(s) = P(1 - s/\lambda). \quad (9.7.6)$$

Умножая строки системы уравнений (9.7.2) на соответствующие степени z и выполняя суммирование, для производящей функции вероятностей состояний со включенным обслуживанием получаем уравнение

$$\tilde{P}(z) = p_0^* \frac{z^r Q^*(z) - Q(z)}{z - Q(z)}.$$

Здесь $Q(z)$ и $Q^*(z)$ — производящие функции вероятностей $\{q_k\}$ для обычного режима обслуживания и разогрева соответственно. Заметим, что

$$\begin{aligned} Q(z) &= \sum_{k=0}^{\infty} z^k \int_0^{\infty} \frac{(\lambda t)^k}{k!} e^{-\lambda t} dB(t) = \int_0^{\infty} \left[\sum_{k=0}^{\infty} \frac{(z\lambda t)^k}{k!} \right] e^{-\lambda t} dB(t) \\ &= \int_0^{\infty} e^{-\lambda t(1-z)} dB(t) = \beta(\lambda(1-z)), \end{aligned}$$

т.е. сводится к ПЛС с указанным параметром от распределения длительности обслуживания.

С учетом вероятностей дополнительных состояний с отключенным обслуживанием производящая функция распределения числа заявок

$$P(z) = \tilde{P}(z) + p_0^* \sum_{j=1}^{r-1} z^j = \tilde{P}(z) + p_0^* \frac{z^r - z}{z - 1}.$$

Подставляя эти результаты в (9.7.6), получаем обобщение формулы Полячека-Хинчина:

$$\nu(s) = p_0^* \left\{ \frac{(1 - s/\lambda)^r \beta^*(s) - \beta(s)}{1 - s/\lambda - \beta(s)} - \frac{\lambda}{s} [(1 - s/\lambda)^r - 1 + s/\lambda] \right\}.$$

Моменты соответствующего распределения можно найти дифференцированием $\nu(s)$ в нуле — аналитическим или численным.

Самостоятельный интерес представляют частные случаи — порог включения и разогрев порознь. В первом случае $b = b^*$, $\beta^*(\cdot) = \beta(\cdot)$ и подставляемая в выражение для p^* средняя длина ПНЗ

$$T_B = rb/(1 - \lambda b).$$

Во втором случае ($r = 1$) ПЛС распределения пребывания сводится к

$$\nu(s) = p_0^* \frac{(1 - s/\lambda)\beta^*(s) - \beta(s)}{1 - s/\lambda - \beta(s)},$$

причем $T_B = b^*/(1 - \lambda b)$, $p_0^* = 1/(1 + \lambda T_B)$.

9.8. Замкнутая система восстановления

Полученные выше результаты относились к ситуации, когда интенсивность λ_k потока заявок на восстановление не зависит от числа k находящихся в ремонтном органе необслуженных заявок. В противном случае говорят о *замкнутых* системах обслуживания. При ограниченном числе R источников заявок обычно считают, что $\lambda_k = \lambda(R - k)$. Методы расчета марковских систем подобного вида хорошо известны (формулы Энгсета). Рассчитывать немарковские системы значительно сложнее. Особенно труден анализ системы, где интенсивность отказов зависит от объема ЗИПа s (запас s рассматривается как холодный резерв, не подверженный отказам). Между тем этот случай достаточно типичен. Если считать, что в рабочей системе установлены R источников заявок, то интенсивность отказов будет оставаться постоянной и равной λR , пока в системе восстановления не скопится $k > s$ заявок. Тогда интенсивность потока заявок начнет убывать по закону $\lambda_k = \lambda[R - (k - s)]$. Методика расчета подобной СМО вида $M/G/1/(R + s)$ была предложена автором в статье [65], оказалась весьма громоздкой и к тому же неприменимой для многоканальных систем восстановления. Однако аппроксимационные методы, описанные в главе 3, без труда обобщаются и на этот случай. Здесь мы отметим особенности его реализации:

- 1) Число ярусов диаграммы равно $R + s + 1$.
- 2) Интенсивность потока заявок зависит от номера яруса.
- 3) Система уравнений баланса вероятностей микросостояний решается только методом итераций.

- 4) Условие баланса заявок, используемое для расчета вероятности свободного состояния для разомкнутых систем, должно быть заменено требованием нормировки вероятностей.

В отличие от ранее рассмотренных ситуаций, редуцированные формы условия оптимальности объема ЗИПа здесь неприменимы, и задачу приходится решать перебором в разумном диапазоне, непосредственно сравнивая значения ожидаемых затрат. При этом для каждого рассматриваемого s систему восстановления придется рассчитывать заново.

Найдем способ расчета стационарных вероятностей состояний *одноканальной* системы с указанной зависимостью интенсивности потока от числа заявок в ней и произвольным распределением длительности обслуживания $B(t)$.

9.8.1. Расчет вложенной цепи Маркова

Вложенную цепь Маркова для интересующего нас процесса мы вновь будем строить аналогично разд. 3.13. Обозначим $q_{k,j}$ вероятность прибытия j новых требований за время обслуживания *при начальном состоянии* k . В нашем случае финальные вероятности состояний вложенной цепи связаны системой уравнений

$$\pi_k = \sum_{j=1}^{k+1} \pi_j q_{j,k+1-j} + \pi_0 q_{0,k}, \quad k = \overline{0, R + s - 2}. \quad (9.8.1)$$

Из этих формул выводим рекуррентные выражения для расчета вероятностей:

$$\pi_k = \left(\pi_{k-1} - \pi_0 q_{0,k-1} - \sum_{j=1}^{k-1} \pi_j q_{j,k-j} \right) / q_{k,0}, \quad k = \overline{0, R + s - 1}, \quad (9.8.2)$$

$$\sum_{k=0}^{R+s-1} \pi_k = 1.$$

9.8.2. Расчет стационарных вероятностей

Вновь обозначим через α среднюю частоту немарковских переходов между состояниями системы. Средний интервал между такими

переходами

$$\alpha^{-1} = b(1 - \pi_0) + (1/\lambda_0 + b)\pi_0 = b + \pi_0/\lambda_0. \quad (9.8.3)$$

Из условия (9.7.2) для стационарных вероятностей следует

$$p_k = \frac{\pi_k}{\lambda_k} \frac{\lambda_0}{\lambda_0 b + \pi_0}, \quad k = \overline{0, s+R-1},$$

$$p_{R+y} = 1 - \frac{\lambda_0}{\lambda_0 b + \pi_0} \sum_{k=0}^{s+R-1} \frac{\pi_k}{\lambda_k}. \quad (9.8.4)$$

9.8.3. Вероятности перехода

Ключом к реализации описанного алгоритма является расчет входящих в (9.8.2) вероятностей $\{q_{k,j}\}$ прибытия ровно j заявок при начальном состоянии k . Введем вспомогательные вероятности

$$\varphi_j(n) = \int_0^\infty \frac{(n\lambda t)^j}{j!} e^{-n\lambda t} dB(t) \quad (9.8.5)$$

прибытия за время обслуживания $B(t)$ ровно j заявок пуассоновского потока интенсивности $n\lambda$. В наиболее простом случае $k+j \leq s$

$$q_{k,j} = \varphi_j(R), \quad k \leq s, \quad j = \overline{0, s-k}.$$

Для случая $k \geq s$ потенциальных источников заявок будет $R - (k - s)$, причем каждый из них независимо от остальных за время t даст заявку с вероятностью $1 - e^{-\lambda t}$ и не даст — с вероятностью $e^{-\lambda t}$. Распределение числа поступивших за t заявок будет биномиальным, и

$$q_{k,j} = \int_0^\infty \binom{R - (k - s)}{j} (1 - e^{-\lambda t})^j e^{-\lambda t [R - (k - s) - j]} dB(t)$$

$$= \int_0^\infty \binom{R - k + s}{j} \sum_{i=0}^j \binom{j}{i} (-1)^i e^{-i\lambda t} e^{-\lambda t [R - (k - s) - j]} dB(t)$$

$$= \binom{R - k + s}{j} \sum_{i=0}^j \binom{j}{i} (-1)^i \int_0^\infty e^{-\lambda [R + s + i - (k + j)]t} dB(t)$$

$$= \binom{R - k + s}{j} \sum_{i=0}^j \binom{j}{i} (-1)^i \varphi_0(R + s + i - (k + j)).$$

Наиболее сложен для расчета вариант, когда $k < s$, а $j > s - k$. В этом случае новый индекс $k + j > s$ и первые $m = s - k$ заявок при постоянной интенсивности потока λR прибывают за время, подчиненное распределению Эрланга порядка m . После этого происходит выборка $n = k + j - s$ элементов из R оставшихся. Для упрощения обозначений введем вспомогательные коэффициенты

$$\tilde{q}_{m,n} = \int_0^\infty dB(t) \int_0^t \frac{\lambda R [\lambda R (t - \tau)]^{m-1}}{(m-1)!} e^{-\lambda R (T - \tau)}$$

$$\times \binom{R}{n} (1 - e^{-\lambda \tau})^n e^{-\lambda \tau (R - n)} d\tau.$$

«Выборочное» подвыражение можно записать в форме

$$\binom{R}{n} (1 - e^{-\lambda \tau})^n e^{-\lambda \tau (R - n)} = \binom{R}{n} \sum_{i=0}^n \binom{n}{i} (-1)^i e^{-\lambda \tau (R - n + i)}.$$

Теперь внутренний интеграл из формулы (9.8.6) преобразуется в

$$I_1 = \binom{R}{n} \sum_{i=0}^n \binom{n}{i} (-1)^i \frac{(\lambda R)^m}{(m-1)!} \int_0^t (t - \tau)^{m-1} e^{-\lambda [R - (n-i)\tau]} d\tau$$

$$= e^{-\lambda R t} \frac{(\lambda R)^m}{(m-1)!} \binom{R}{n} \sum_{i=0}^n \binom{n}{i} (-1)^i \int_0^t (t - \tau)^{m-1} e^{(n-i)\lambda \tau} d\tau.$$

При $i < n$ его можно свести к

$$I_2 = \int_0^t \sum_{l=0}^{m-1} \binom{m-1}{l} t^l \tau^{m-1-l} (-1)^{m-l-1} e^{\lambda(n-i)\tau} d\tau$$

$$= \sum_{l=0}^{m-1} \binom{m-1}{l} t^l (-1)^{m-l-1} \int_0^t \tau^{m-1-l} e^{\lambda(n-i)\tau} d\tau.$$

Выполним подстановку $\lambda(n-i)\tau = u$. Согласно формуле (5.67.9) из [11], интеграл можно записать в виде

$$I_3 = \frac{1}{[\lambda(n-i)]^{m-l}} \int_0^{\lambda(n-i)t} u^{m-l-1} e^u du$$

$$= \frac{(-1)^{m-l-1} (m-l-1)!}{[\lambda(n-i)]^{m-l}} \left[e^{\lambda(n-i)t} \sum_{r=0}^{m-l-1} \frac{[-\lambda(n-i)t]^r}{r!} - 1 \right].$$

Возвращаясь к I_2 , имеем

$$I_2 = \sum_{l=0}^{m-1} \binom{m-1}{l} \frac{(m-l-1)!}{[\lambda(n-i)]^{m-l}} \left[e^{\lambda(n-i)t} \sum_{r=0}^{m-l-1} \frac{[-\lambda(n-i)t]^r t^{r+l}}{r!} - t^l \right].$$

Следует иметь в виду, что при $i = n$

$$I_2 = \int_0^t (t-\tau)^{m-1} d\tau = t^m/m.$$

С использованием этих результатов имеем

$$\begin{aligned} \tilde{q}_{m,n} &= \int_0^\infty dB(t) e^{-R\lambda t} \binom{R}{n} \frac{(\lambda R)^m}{(m-1)!} \left\{ (-1)^n \frac{t^m}{m} + \sum_{i=0}^{n-1} \binom{n}{i} (-1)^i \right. \\ &\quad \left. \sum_{l=0}^{m-1} \frac{(m-l-1)!}{[\lambda(n-i)]^{m-l}} \left[e^{\lambda(n-i)t} \sum_{r=0}^{m-l-1} \frac{[-\lambda(n-i)t]^r}{r!} t^{r+l} - t^l \right] \right\} \\ &= \binom{R}{n} \frac{(\lambda R)^m}{m!} (-1)^n \int_0^\infty t^m e^{-R\lambda t} dB(t) \\ &\quad + \binom{R}{n} \frac{(\lambda R)^m}{(m-1)!} \sum_{i=0}^{n-1} \binom{n}{i} (-1)^i \sum_{l=0}^{m-1} \frac{(m-l-1)!}{[\lambda(n-i)]^l} \\ &\quad \times \int_0^\infty \left\{ e^{-\lambda t(R-n+i)} \sum_{r=0}^{m-l-1} \frac{[-\lambda(n-i)t]^r}{r!} t^{r+l} - t^l e^{-R\lambda t} \right\} dB(t). \end{aligned}$$

Первая строка результата сводится к

$$\binom{R}{n} (-1)^n \int_0^\infty \frac{(R\lambda t)^m}{m!} e^{-R\lambda t} dB(t) = \binom{R}{n} (-1)^n \varphi_m(R),$$

а последний интеграл

$$I_4 = \sum_{r=0}^{m-l-1} \frac{[-\lambda(n-i)]^r (r+l)!}{r! [\lambda(R-n+i)]^{r+l}} \int_0^\infty \frac{[\lambda(R-n+i)t]^{r+l}}{(r+l)!} e^{-\lambda(R-n+i)t} dB(t)$$

$$= \frac{l!}{(R\lambda)^l} \int_0^\infty \frac{(R\lambda t)^l}{l!} e^{-R\lambda t} dB(t)$$

$$= \frac{1}{[\lambda(R-n+i)]^l} \sum_{r=0}^{m-l-1} \frac{(r+l)!}{r!} \left(\frac{i-n}{R+i-n} \right)^l \varphi_{r+l}(R-n+i) - \frac{l!}{(R\lambda)^l} \varphi_l(R).$$

Окончательно

$$\begin{aligned} \tilde{q}_{m,n} &= \binom{R}{n} (-1)^n \varphi_m(R) + \binom{R}{n} \frac{(\lambda R)^m}{(m-1)!} \sum_{i=0}^{n-1} \binom{n}{i} (-1)^i \\ &\quad \times \sum_{l=0}^{m-1} \left\{ \frac{(m-l-1)!}{[\lambda(n-i)]^l} \frac{1}{[\lambda(R-n+i)]^l} \right. \\ &\quad \times \sum_{r=0}^{m-l-1} \frac{(r+l)!}{r!} \left(\frac{i-n}{R+i-n} \right)^l \varphi_{r+l}(R-n+i) \\ &\quad \left. - \frac{l!}{(R\lambda)^l} \varphi_l(R) \right\}. \end{aligned}$$

Возвращаясь к первоначальной постановке задачи, для «пограничного» диапазона $k < s$, $j = s - k + 1$, $R - k$ имеем

$$q_{k,j} = \tilde{q}_{y-k, k+j-y}.$$

Теперь можно найти вероятности вложенной цепи и далее — стационарные вероятности состояний

Поскольку распределение последних зависит от s , здесь уже нельзя воспользоваться условиями оптимальности вида (9.2.2) или (9.2.2) приходится непосредственно сопоставлять значения целевой функции $L(s)$. В связи с этим имеет смысл наиболее трудоемкую часть алгоритма — расчет $\{\tilde{q}_{m,n}\}$ — выполнить для значений m и n , обеспечивающих реализацию основного алгоритма во всем диапазоне проверяемых значений s .

9.9. Учет погрешности исходных данных

Наибольший практический интерес вызывает задача расчета ЗИПа к проектируемой технике, когда низкая точность исходных данных (в

особенности по времени восстановления элементов ЗИП) оправдывает применение простейших методов расчета. Более того, эта точность может быть настолько низкой, что потребует учета дополнительных затрат, возникающих из-за незнания истинных параметров модели.

Подобные ситуации рассматривались в разд. 5.7 в связи с планированием невосстанавливаемого запаса. Применим тот же подход к задаче о восстанавливаемом ЗИП при априорно неизвестных интенсивности отказов λ и восстановлений μ . Как было показано выше, оптимальный запас зависит от названных параметров только через их отношение $\rho = \lambda/\mu$. С помощью предварительных просчетов по аналогии или методом экспертных оценок можно установить лишь диапазон изменения $0 < a \leq \rho \leq b < 1$, причем ввиду недостаточности информации ρ приходится считать случайной величиной, распределенной на $[a, b]$ равномерно. Попутно заметим, что равномерность распределения не является ограничением обсуждаемого метода. Это положение используется только при выводе формулы для расчета затрат и ее конкретизаций. При наличии другого распределения оно вводится в расчетную схему аналогичным образом.

Запишем целевую функцию затрат в общем виде как

$$L = hs + d\varphi(s), \quad (9.9.1)$$

где $\varphi(s)$ — основание для начисления штрафа. Необходимым условием оптимальности запаса s является одновременное выполнение неравенств

$$\begin{aligned} L(s+1) - L(s) &= h + d\Delta\varphi(s) \geq 0, \\ L(s) - L(s-1) &= h + d\Delta\varphi(s-1) \leq 0, \end{aligned}$$

или

$$\Delta\varphi(s-1) \leq -h/d \leq \Delta\varphi(s). \quad (9.9.2)$$

Расчет начинается с определения минимального запаса s_{\min} при $\rho_0 = a$ и максимального s_{\max} — при $\rho = b$. Очевидно, отрезок $[a, b]$ может быть разбит на $s_{\max} - s_{\min} + 1$ непересекающихся интервалов, в каждом из которых оптимально свое значение запаса. Найдем координату точки деления ρ_1 . На границе смежных крайних слева интервалов запасы s_{\min} и $s_{\min} + 1$ должны доставлять равные значения правой части (9.9.1), откуда следует

$$h + \Delta\varphi(s_{\min}, \rho_1) = 0. \quad (9.9.3)$$

Решая это уравнение относительно ρ_1 , можно найти первую границу между интервалами постоянства оптимального запаса. Далее запас увеличивается на единицу и находится граница следующего интервала и т.д. — до $s_{\max} - 1$. Правой границей последнего интервала служит $\rho = b$. Теперь нетрудно сформировать массивы $\rho[0 : m]$ точек деления, $p[1 : m]$ вероятностей $\rho \in [\rho_{i-1}, \rho_i]$ и оптимальных запасов $s^*[1 : m]$. Усредненное по интервалу значение ожидаемых затрат

$$\bar{L}_i(s) = \int_{\rho_{i-1}}^{\rho_i} L(s, \rho) f(\rho) d\rho = hs + \frac{d}{\rho_i - \rho_{i-1}} \int_{\rho_{i-1}}^{\rho_i} \varphi(s, \rho) d\rho. \quad (9.9.4)$$

Придавая переменной s значения оптимальных запасов для различных интервалов, можно сформировать матрицу затрат $\bar{L} = \{\bar{L}_i(s_j)\}$. Теперь можно получить вектор ожидаемых затрат

$$R = \bar{L} \times P \quad (9.9.5)$$

и по номеру его минимальной компоненты выбрать оптимальный запас.

Другим критерием может служить минимум ожидаемого риска. Вычитая из элементов каждой строки матрицы \bar{L} диагональные, получаем матрицу ожидаемого риска \bar{L}^0 и по номеру минимальной компоненты вектора ожидаемого риска выбираем наименее рискованное значение запаса. Нетрудно показать, что этот выбор, вообще говоря, не совпадает с выбором по минимуму затрат. В самом деле, i -я составляющая вектора ожидаемого риска

$$r_i^0 = \sum_{j=1}^m \bar{L}_{ij}^0 P_j = \sum_{j=1}^m (\bar{L}_{ij} - \bar{L}_{ii}) P_j = \sum_{j=1}^m \bar{L}_{ij} P_j - \bar{L}_{ii} \sum_{j=1}^m P_j = r_i - \bar{L}_{ii},$$

где r_i — одноименная компонента вектора ожидаемых затрат. Поскольку разность между r_i и r_i^0 зависит от i , положения минимальных компонент могут не совпадать. Очевидная монотонность возрастания \bar{L}_{ii} по i позволяет указать и направление возможного смещения: выбор по минимуму риска всегда приводит к большим или равным запасам, чем выбор по минимуму затрат.

Перейдем к выбору расчетных соотношений для основных типов исследуемых моделей.

9.9.1. Однолинейная система со штрафом по вероятности дефицита

Воспользовавшись формулами разд. 9.2, для названного случая имеем

$$\bar{L}_i(s) = hs + \frac{d}{\rho_i - \rho_{i-1}} \int_{\rho_{i-1}}^{\rho_i} \rho^{s+1} d\rho = hs + \frac{\rho_i^{s+2} - \rho_{i-1}^{s+2}}{(s+2)(\rho_i - \rho_{i-1})}. \quad (9.9.6)$$

Приращение вероятности дефицита

$$\Delta\varphi(s) = \rho_i^{s+2} - \rho_i^{s+1}, \quad (9.9.7)$$

что позволяет переписать уравнение для граничного коэффициента загрузки в форме $h - d\rho^{s+1}(1 - \rho) = 0$ и искать его решение по итерационной формуле

$$\rho = \left[\frac{h}{d(1 - \rho)} \right]^{1/(s+1)}. \quad (9.9.8)$$

В качестве начального значения может использоваться ранее найденная левая граница того же интервала (она же — правая граница предыдущего).

9.9.2. Однолинейная система со штрафом по ожидаемому дефициту

В данном случае

$$\bar{L}_i(s) = hs + \frac{d}{\rho_i - \rho_{i-1}} \int_{\rho_{i-1}}^{\rho_i} \frac{\rho^{s+1}}{1 - \rho} d\rho.$$

Но

$$\frac{\rho^{s+1}}{1 - \rho} = \frac{\rho^{s+1} - 1 + 1}{1 - \rho} = \frac{1}{1 - \rho} - \frac{\rho^{s+1} - 1}{\rho - 1} = \frac{1}{1 - \rho} - \sum_{k=0}^s \rho^k.$$

Интегрируя это выражение почленно, получаем

$$\bar{L}_i(s) = hs + \frac{d}{\rho_i - \rho_{i-1}} \left[\ln(1 - \rho) + \sum_{k=1}^{s+1} \frac{\rho^k}{k} \right] \Big|_{\rho_{i-1}}^{\rho_i}. \quad (9.9.9)$$

Приращение вероятности дефицита

$$\Delta\varphi(s) = -\rho_i^{s+1}, \quad (9.9.10)$$

граничный коэффициент загрузки здесь выражается явно:

$$\rho = (h/d)^{1/(s+1)}. \quad (9.9.11)$$

9.9.3. Многоканальные марковские системы

Для таких систем ожидаемые затраты при $\rho \in [\rho_{i-1}, \rho_i]$ в связи со сложной зависимостью p_n от ρ находятся численным интегрированием.

На основании формулы (9.3.1) границы интервала разбиения при штрафе по вероятности дефицита определяются из уравнения

$$h - dp_n \rho^{s+1-n} = 0,$$

или

$$\rho = \left(\frac{h}{dp_n} \right)^{1/(s+1-n)}$$

При штрафе по ожидаемому дефициту используется уравнение

$$\rho = \left[\frac{h(1 - \rho)}{dp_n} \right]^{1/(s+1-n)}$$

Отметим, что в выражение для p_n (см. систему (3.12.2)) входит искомый параметр ρ . Это определяет итерационный характер нахождения граничного значения для обоих вариантов исчисления штрафов.

9.9.4. Пример

Рассмотрим реализацию методики применительно к однолинейной системе восстановления со штрафом, пропорциональным вероятности дефицита. Пусть $\rho \in [0.3, 0.7]$, цена хранения составляет $h = 10$ руб., $d = 7000$ руб. Применение (9.2.9) позволило установить, что $s_{\min} = 5$, а $s_{\max} = 14$ штук. В соответствии с этим диапазон значений коэффициента загрузки оказался разбит на 10 интервалов. С помощью (9.9.8) установлены границы интервалов: 0.3000, 0.3617, 0.4245, 0.4783, 0.5245, 0.5644, 0.5990, 0.6292, 0.6558, 0.6793, 0.7000. По формуле (9.9.6) была построена матрица затрат $\{\bar{L}_i(s_j)\}$, общий элемент

которой соответствует ожидаемым затратам при выборе $s = s_i$ (из диапазона от 5 до 14 штук), если истинный коэффициент загрузки принадлежит к интервалу, в котором оптимален запас s_j . Элементы матрицы приведены ниже:

63.1	64.2	71.3	80.4	90.1	100.0	110.0	120.0	130.0	140.0
76.6	70.6	74.2	81.7	97.7	100.3	110.1	120.0	130.0	140.0
110.2	87.4	82.5	85.7	92.6	101.2	110.5	120.2	130.1	140.0
162.4	116.6	98.5	94.4	97.2	103.7	111.9	120.9	130.5	140.2
233.5	160.2	124.7	109.9	106.3	108.9	114.9	122.7	131.5	140.8
322.4	218.7	162.5	134.0	121.4	118.4	120.7	126.2	133.6	142.1
426.7	291.6	212.4	167.7	143.9	133.2	130.4	132.6	137.7	144.8
543.6	377.4	274.1	211.3	174.5	154.4	145.0	142.5	144.5	149.3
670.4	474.4	346.8	265.0	213.6	182.6	165.2	156.9	154.7	156.5
804.0	580.2	428.9	327.7	260.9	217.9	191.4	176.2	168.8	166.8

Вычитание из каждой строки матрицы \bar{L} ее элемента, стоящего на главной диагонали, дает матрицу риска \bar{L}^0 . Сформировав 10-мерный вектор P априорных вероятностей, пропорциональных длинам интервалов разбиения, получаем вектор ожидаемого риска $R^0 = \bar{L}^0 \times P$. Его составляющие: 24.38, 20.62, 18.47, 23.20, 36.16, 58.10, 89.21, 129.23, 177.63, 233.41. Минимальная составляющая $r^0 = 18.47$ указывает на целесообразность выбора s^* , связанного с третьим интервалом, т.е. $s^* = 7$ штук. Ожидаемые затраты могут быть получены умножением вектор-столбца матрицы затрат, соответствующего $s = s^*$, на вектор-строку вероятностей P и в данном случае составляют 148.47 руб

Как и в разд. 5.7, в каждой строке матрицы затрат (и соответственно матрицы риска) при уменьшении запаса относительно оптимального затраты растут гораздо быстрее, чем при его увеличении. Это обстоятельство может рассматриваться как количественное «обоснование» известной тенденции скорее преувеличивать запасы, чем преуменьшать их.

9.9.5. Интервальная оценка частных параметров

В некоторых случаях интервальная оценка задается не для $\rho = \lambda/\mu$, а для параметров λ и μ порознь. Можно показать, что в предположении о равномерности распределений λ и μ плотность отношения

λ/μ должна вычисляться по формуле

$$f(\rho) = \begin{cases} 0, & \text{если } \rho \notin [c_0, c_3], \\ b[\mu_{\max}^2 - (\lambda_{\min}/\rho)^2], & c_0 \leq \rho \leq c_1, \\ b[\mu_{\max}^2 - \mu_{\min}^2], & c_1 \leq \rho \leq c_2, \\ b[(\lambda_{\max}/\rho)^2 - \mu_{\min}^2], & c_2 \leq \rho \leq c_3. \end{cases} \quad (9.9.12)$$

Здесь принято

$$\begin{aligned} c_0 &= \frac{\lambda_{\min}}{\mu_{\max}}; & c_1 &= \min \left\{ \frac{\lambda_{\min}}{\mu_{\min}}, \frac{\lambda_{\max}}{\mu_{\max}} \right\}; \\ c_2 &= \max \left\{ \frac{\lambda_{\min}}{\mu_{\min}}, \frac{\lambda_{\max}}{\mu_{\max}} \right\}; & c_3 &= \frac{\lambda_{\max}}{\mu_{\min}} < 1; \\ b &= [2(\lambda_{\max} - \lambda_{\min})(\mu_{\max} - \mu_{\min})]^{-1}. \end{aligned}$$

Из общей схемы легко выводятся частные случаи $\mu = \text{const}$ и/или $\lambda = \text{const}$. Первый из них отдельного рассмотрения не требует, поскольку ρ оказывается равномерно распределенным на интервале $\left[\frac{\lambda_{\min}}{\mu}, \frac{\lambda_{\max}}{\mu} \right]$ и мы вновь находимся в условиях применения основной расчетной схемы алгоритма. Во втором случае уравнение (9.9.12) сводится к

$$f(\rho) = \frac{\lambda}{(\mu_{\max} - \mu_{\min})\rho^2}, \quad \frac{\lambda}{\mu_{\max}} \leq \rho \leq \frac{\lambda}{\mu_{\min}}. \quad (9.9.13)$$

В любом варианте общая идея метода и алгоритм разбиения диапазона ρ остаются прежними — меняется только выражение для ожидаемых затрат на интервалах. Практически

$$\bar{L}(s) = \int_{\rho_{i-1}}^{\rho_i} L(s, \rho) d\rho$$

в целях унификации программ целесообразно находить численным интегрированием, а расчет $f(\rho)$ оформлять в виде сменной процедуры.

Несколько заключительных замечаний по проблеме учета погрешностей.

- 1) Описанный подход может быть применен и при интервальной оценке «нормы штрафа» d/h (на практике h устанавливается довольно точно, а ошибка возникает из-за неопределенности в цене штрафа d). Сразу же отметим, что величина оптимального запаса гораздо менее чувствительна к колебаниям нормы штрафа, чем к изменению коэффициента загрузки системы восстановления.

- 2) При рассмотрении многопериодных задач возможен пересчет вероятностей принадлежности интересующего нас отношения на основе апостериорной информации (по формуле Байеса)

9.10. Многономенклатурная задача

Методы предыдущего раздела в принципе могут быть применены для определения состава многономенклатурного восстанавливаемого ЗИПа порознь. Задача становится существенно неоднородной при выполнении хотя бы одного из следующих условий

- элементы ЗИПа необходимы для обеспечения нормальной работы одного и того же комплекса (общая цель),
- восстановление элементов ЗИПа выполняется в одном и том же ремонтном производстве в порядке общей очереди.

9.10.1. Общая цель

При выполнении первого условия функция затрат записывается как

$$\begin{aligned} L(\vec{s}) &= \sum_{i=1}^N h_i s_i + d \left(1 - \prod_{i=1}^N \sum_{k=0}^{s_i} p_{ik} \right) \\ &= \sum_{i=1}^N h_i s_i + d \left[1 - \prod_{i=1}^N \left(1 - \sum_{k=s_i+1}^{\infty} p_{ik} \right) \right]. \end{aligned}$$

Для простоты рассуждений введем среднюю по всем номенклатурам вероятность дефицита

$$\delta = \frac{1}{N} \sum_{i=1}^N \sum_{k=s_i+1}^{\infty} p_{ik} = \frac{1}{N} \sum_{i=1}^N \delta_i. \quad (9.10.1)$$

Тогда

$$\begin{aligned} L(\vec{s}) &\approx \sum_{i=1}^N h_i s_i + d [1 - (1 - \delta)^N] \\ &= \sum_{i=1}^N h_i s_i + d \left[N\delta - \frac{N(N-1)}{2!} \delta^2 + \dots \right]. \end{aligned} \quad (9.10.2)$$

Установим, при каком соотношении между N и δ можно ограничиться первым членом выражения, заключенного в квадратные

скобки. Погрешность суммы знакопередающегося ряда не превосходит модуля первого отброшенного члена. Обозначим относительную погрешность этой суммы через ε . Тогда

$$\varepsilon \leq \frac{N(N-1)}{2! N \delta} \delta^2 = \frac{N-1}{2} \delta.$$

Сразу же отметим, что фактическая относительная погрешность суммарных затрат будет меньше ε , поскольку в затраты, определяемые правой частью (9.10.2), входят и расходы на хранение элементов ЗИПа. Таким образом, достаточно проверить условие

$$\frac{N-1}{2} \delta \leq \varepsilon_{\text{доп}}. \quad (9.10.3)$$

При выполнении этого ограничения можно с приемлемой точностью переписать целевую функцию (9.10.2) в виде

$$L(\vec{s}) \approx \sum_{i=1}^N h_i s_i + d \cdot N \delta = \sum_{i=1}^N (h_i s_i + d \delta_i), \quad (9.10.4)$$

т.е. свести ожидаемые затраты к сумме функций вида (11.2.3) по каждой номенклатуре порознь. Это означает, что для выбора оптимального запаса можно независимо применять неравенства (11.2.3) по каждой номенклатуре порознь — разумеется, с учетом того, является ли восстановление запаса общим или отдельным.

Для проверки допустимости декомпозиции функции затрат по критерию (9.10.3) необходимо знать среднюю вероятность дефицита δ . Попытаемся исключить этот показатель на этапе предварительной оценки. В марковских однолинейных системах вероятности $\{p_k\}$ образуют геометрическую прогрессию со знаменателем ρ (коэффициентом загрузки), так что

$$\delta = p_{s^*+1} / (1 - \rho).$$

Согласно (9.2.4), $p_{s^*+1} \leq h/d$. Следовательно, условие (9.10.3) можно привести к виду

$$\frac{N-1}{2(1-\rho)} \cdot \frac{h}{d} \leq \varepsilon_{\text{доп}}. \quad (9.10.5)$$

Если предварительная оценка ε , полученная с помощью левой части (9.10.5), дала результат, близкий к $\varepsilon_{\text{доп}}$, то после выполнения расчетов целесообразно вычислить среднее значение фактических вероятностей дефицита согласно (9.10.5), а затем проверить допустимость декомпозиции целевой функции посредством (9.10.3).

9.10.2. Общее восстановление

При общем ремонтном органе восстановление заявок одних типов будет задерживаться из-за его занятости восстановлением других. Расчет n -канальной СМО с беспriorитетным обслуживанием неоднородного потока заявок может быть выполнен методами, описанными в главе 3, по суммарной интенсивности входящего потока $\Lambda = \sum_i \lambda_i$ и средневзвешенным моментам распределения длительности обслуживания $B_k = \sum_i \lambda_i b_{i,k} / \Lambda$, $k = 1, 2, \dots$.

Пусть найдено распределение $\{\pi_k\}$ числа «обобщенных» заявок в системе, а нас интересуют вероятности $\{p_k\}$ наличия в ней ровно k заявок выделенного i -го типа. Определим вероятности принадлежности наугад выбранной заявки к этому типу для выбора из очереди, $u = \lambda_i / \Lambda$, а из проходящих обслуживание — как $v = \lambda_i b_{i,1} / \sum_j \lambda_j b_{j,1}$, а также дополнителные к ним \bar{u}, \bar{v} . Вероятность иметь k заявок меченого типа в очереди

$$p'_k = \sum_{m=k}^{\infty} \binom{m}{k} u^k \bar{u}^{m-k} \pi_{n+m}, \quad k = 0, 1, \dots$$

Вероятность иметь k заявок меченого типа в каналах обслуживания

$$p''_k = \sum_{m=k}^n \binom{m}{k} v^k \bar{v}^{m-k} \pi_m, \quad k = \overline{0, n}.$$

Искомое результирующее распределение задается сверткой этих частных распределений:

$$p_k = \sum_{m=0}^n p''_m p'_{k-m}, \quad k = 0, 1, \dots$$

Возможна *приоритетная* организация ремонта — как правило, без прерывания начатого обслуживания. Теория систем с приоритетным обслуживанием практически разработана только для одноканальных систем с простейшим потоком заявок и ориентирована на анализ оперативных характеристик СМО — расчет распределения времени пребывания заявок каждого вида в системе. В разд. 3.17 описан пакет программ, позволяющих, в частности, вычислять моменты этих распределений. По моментам можно построить аппроксимацию плотности распределения времени пребывания.

Для одноканальных систем распределение числа заявок, пришедших в систему за время пребывания заявки в ней, совпадает с распределением числа заявок в системе (закон сохранения стационарной очереди). Опуская индекс типа заявки, требуемое распределение можно записать в форме

$$a_j = \int_0^{\infty} \frac{(\lambda t)^j}{j!} e^{-\lambda t} v(t) dt, \quad j = 0, 1, \dots$$

Вопрос об эффективном расчете этих вероятностей рассматривался в разд. 3.6.3.

Описанный подход (в обеих формах — как для приоритетного, так и для беспriorитетного восстановления) полезен и при анализе централизованного ремонта однотипных агрегатов, когда «неоднородность» состоит в их принадлежности разным владельцам.

Сети обслуживания обычно рассчитываются методом потокоэквивалентной декомпозиции (см. разд. 3.15), в котором после балансировки потоков выполняется независимый расчет распределений числа заявок в узлах сети. Описанным здесь способом можно получить распределения числа заявок выделенного типа в них, а затем выполнить свертку по узлам. Для экономии памяти свертку можно выполнять последовательно — по мере расчета узлов.

9.11. Частично восстанавливаемый ЗИП

Отказы, наблюдаемые в сложной технике, по условиям восстановления элементов можно разбить на три группы:

- 1) восстановление в порядке рекламаций;
- 2) восстановление в условиях эксплуатации;
- 3) восстановление невозможно.

Обозначим интенсивности соответствующих отказов через λ_1, λ_2 и λ_3 , а восстановлений — через μ_1 и μ_2 . Тогда стационарное распределение числа элементов в системах восстановления порознь можно получить методами ТМО (при простейших допущениях согласно (9.2.5)), а

общее распределение R — их сверткой. Распределение числа невосстанавливаемых отказов $H(T)$ за время T можно считать пуассоновским со средним $\lambda_3 T$. Распределение окончательного снижения запаса $W(T) = R * H(T)$ — еще одна свертка.

Теперь получим функцию затрат. Естественно определить сумму штрафов пропорционально ожидаемому числу недостач к концу периода. Плату за хранение следует исчислять по остатку, включая в него элементы, находящиеся в системе восстановления. Следовательно, общие затраты за период T составят

$$L = c(s - z) + h \sum_{k=0}^{s-1} (s - k) H_k(T) + d \sum_{k=s+1}^{\infty} (k - s) W_k(T). \quad (9.11.1)$$

Применив условия оптимальности типа (5.6.13), убеждаемся, что для оптимального запаса s^* должны выполняться неравенства

$$\begin{aligned} c + h \sum_{k=0}^{s^*} H_k(T) - d \sum_{k=s^*+1}^{\infty} W_k(T) &\geq 0, \\ c + h \sum_{k=0}^{s^*-1} H_k(T) - d \sum_{k=s^*}^{\infty} W_k(T) &\leq 0. \end{aligned} \quad (9.11.2)$$

Очевиден способ расчета s^* : задавшись $s_0 = 1$ и последовательно увеличивая его, вычислять левую часть второго неравенства (9.11.2). Первый же s , для которого результат окажется положительным, и будет оптимальным.

Глава 10

«Метрический» анализ систем

Сложность систем управления запасами восстанавливаемых частей и разнообразие возникающих в связи с ними математических задач вполне объясняют отсутствие единой теории их расчета. Различные аспекты работы таких систем исследуются методами теории массового обслуживания, теории восстановления, цепей Маркова, но главным образом — посредством рассмотренной в предыдущих главах элементарной теории запасов. В оставшихся главах мы рассмотрим, как комбинируются эти методы в реально применявшихся методиках для сложных систем. Типичная последовательность этапов расчета такова:

- Определить спрос на каждой базе (в верхних эшелонах он зависит от структуры системы и стратегий УЗ).
- Определить задержки восполнения.
- Рассчитать вероятности состояний системы.
- Выбрать и рассчитать через вероятности состояний показатель эффективности.

10.1. Специфика спроса в иерархии

10.1.1. Особенности генерации заявок

Спрос в высших звеньях в основном формируется спросом нижестоящих звеньев. Спрос непосредственно в высший уровень может быть:

- при наличии «фиктивных» складов — заказы по почте адресуются прямо вверх;
- по результатам дефектации ремонтируемого узла — на редко используемые детали, хранимые только на складах высших уровней

Введем единообразные обозначения: r_{ijm} — вероятности того, что деталь i будет ремонтировать в узле j уровня m ; γ_i — вероятность ремонта i -детали на втором эшелоне. Регулируя значения параметров, можно переходить от восстанавливаемых деталей к расходным. Структура типичной модели иерархической двухуровневой системы поддержания готовности с ремонтными органами (РО) показана на рис. 10.1. Для упрощения на нем показаны одна база и движение одной номенклатуры.

Интенсивность спроса, вообще говоря, зависит не только от числа «конечных потребителей», но и от того, сколько из них находится в действующем состоянии. Если их сравнительно немного, то математической моделью процесса восстановления становятся более сложные для расчета замкнутые СМО.

Обычно предполагается, что внешний спрос в разных пунктах имеет общий тип распределения (отличаются параметры).

Опустим для простоты индекс типа заявок. Каждая j -я база, $j = \overline{1, J}$, пытается поддержать уровень запаса S_j (возможно, нулевой). Спрос возникает только при отказе у потребителя и (в зависимости от наличия) либо удовлетворяется немедленно, либо порождает заявку в депо. Задержка поставки может зависеть от типа детали, типа поломки, типа ремонта, запаса и времени восстановления в депо и т.д. Задержки предполагаются взаимно независимыми с произвольной ФР $v_j(t)$ и средним τ_j . Спрос — пуассоновский (обычный или составной). Для пуассоновского процесса времена между событиями распределены экспоненциально с параметром λ_j , а число заявок за время τ

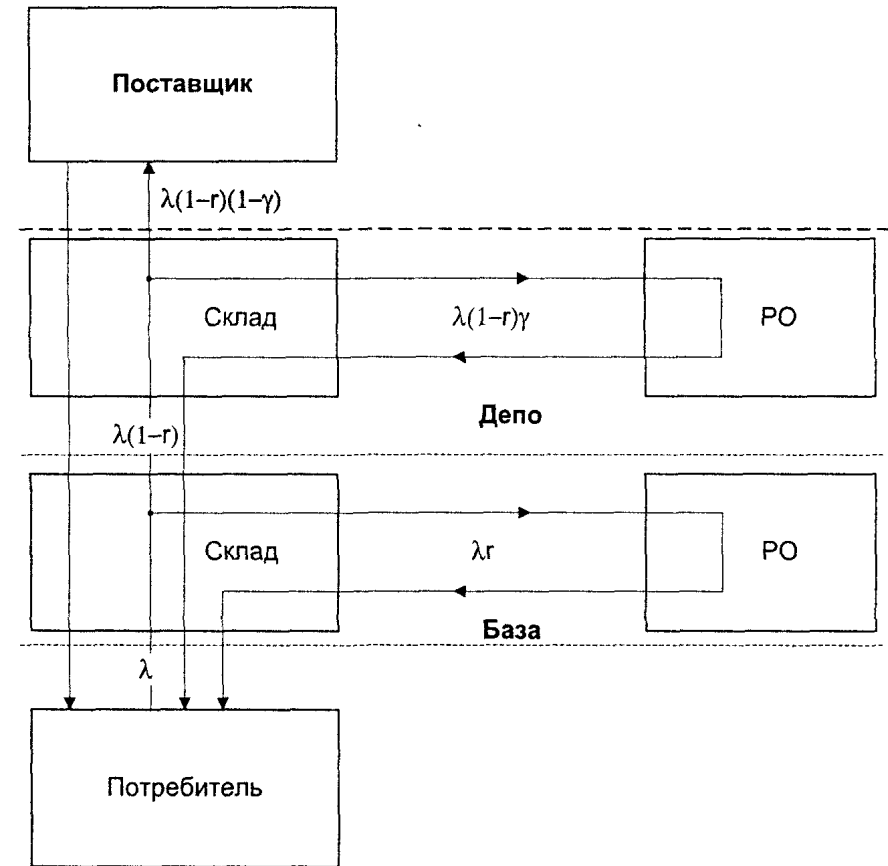


Рис. 10.1. Двухуровневая система с восстанавливаемым ЗИПом имеет пуассоновское распределение

$$p(n) = [(\lambda\tau)^n / n!] \exp(-\lambda\tau) \quad (10.1.1)$$

с коэффициентом вариации, строго равным единице. В случае простого потока (заявки единичного объема) так же распределен и суммарный спрос.

10.1.2. Составной пуассоновский поток

При составном потоке имеется простейший поток заявок, причем объем заявки j задается некоторым дискретным распределением $\{h_j\}$

независимо от других заявок. Предполагается, что этот объем равен по крайней мере единице. Если же $h_0 > 0$, то считается

$$\begin{aligned} h'_0 &= 0, \\ h'_j &= h_j/[1 - h_0], \\ \lambda' &= \lambda[1 - h_j]. \end{aligned} \quad (10.1.2)$$

Вероятность отсутствия спроса равна вероятности отсутствия требований, так что

$$p_0 = e^{-\lambda\tau}. \quad (10.1.3)$$

Для составного пуассоновского процесса средний спрос за время τ

$$\theta = \lambda\tau\bar{h}, \quad (10.1.4)$$

а коэффициент вариации больше единицы.

При «геометрическом» распределении пачки

$$h_j = (1 - \rho)\rho^{j-1}, \quad j = 1, 2, \dots \quad (10.1.5)$$

Средний объем пачки

$$\bar{h} = 1/(1 - \rho);$$

соответственно $\rho = 1 - 1/\bar{h}$.

Для логарифмического распределения

$$h(j) = \left(\frac{q-1}{q}\right)^j \frac{1}{j \ln q}, \quad j = 1, 2, \dots, \quad (10.1.6)$$

где $q > 1$ — параметр распределения. Средний объем пачки $\bar{h} = (q - 1)/\ln q$.

Пусть n — суммарный объем заявок составного пуассоновского потока за время τ . Тогда

$$p_n = \sum_{i=0}^n \frac{(\lambda\tau)^i}{i!} e^{-\lambda\tau} h^{i*}(n). \quad (10.1.7)$$

Здесь пуассоновская вероятность появления ровно i пачек умножается на $h^{i*}(n)$, т.е. i -кратную свертку $h(n)$ (вероятность равенства числу n их суммарного объема).

Обозначим через

$$g(z) = \sum_{i=1}^{\infty} z^i h_i$$

производящую функцию (ПФ) распределения объема пачки. Поскольку ПФ свертки равна произведению ПФ составляющих, из (10.1.7) следует выражение для ПФ распределения суммарного объема заявок составного пуассоновского потока $\lambda\tau$

$$P(z) = \sum_{n=0}^{\infty} \frac{(\lambda\tau)^n}{n!} e^{-\lambda\tau} g^n(z) = e^{-\lambda\tau(1-g(z))}. \quad (10.1.8)$$

Пусть, например, размер требования подчиняется модифицированному геометрическому распределению

$$h_j = p(1-p)^{j-1} = (1-\nu)\nu^{j-1}, \quad j = 1, 2, \dots$$

($\nu = 1 - p$, требование должно быть как минимум единичного объема). Производящая функция объема требования

$$\pi(z) = \frac{z(1-\nu)}{1-\nu z}. \quad (10.1.9)$$

Тогда ПФ составного пуассоновского распределения

$$P(z) = \exp(-\lambda\tau\{1 - [z(1-\nu)]/(1-\nu z)\}) = \exp[-\lambda\tau(1-z)/(1-\nu z)].$$

Можно показать (см. [89]), что вероятность поступления n заявок

$$p_n = \nu^n e^{-\lambda\tau} \sum_{i=1}^n \frac{1}{i!} \binom{n-1}{i-1} \left[\left(\frac{1-\nu}{\nu} \right) \lambda\tau \right]^i, \quad n = 1, 2, \dots$$

Для случая логарифмического распределения пачки В. Феллер вывел отрицательно-биномиальное распределение спроса:

$$p_n = \binom{k+n-1}{n} \frac{1}{q^k} (1-1/q)^n, \quad n = 0, 1, \dots, \quad (10.1.10)$$

где $k = \lambda\tau/\ln q$ — натуральное число. Эти вероятности легко считаются рекуррентно:

$$\begin{aligned} p_0 &= \frac{1}{q^k}, \\ p_n &= p_{n-1} \frac{k+n-1}{n} (1-1/q), \quad n = 1, 2, \dots \end{aligned}$$

При спросе на пачки постоянного объема требуется только перемасштабирование задачи. При случайном объеме пачки заявка на поставку будет выдана при *пересечении* текущим запасом уровня заказа s . Фактически модель ведет себя так, как если бы к случайному спросу за время задержки поставок прибавилась величина «перескока». Соответственно распределение спроса за время задержки должно быть заменено его сверткой с распределением перескока. Последнее хорошо изучено в теории восстановления, совпадает с «остаточным» распределением и легко может быть получено из распределения объема пачки. Простейшим способом учета перескока является добавление к нижнему порогу s средней величины перескока $r = b_2/(2b_1)$, где $\{b_i\}$ — соответствующие моменты распределения объема пачки.

Другой особенностью, порождаемой неординарным спросом, является возможность применения двухуровневой («двухбункерной») стратегии (s, S) , когда при снижении запаса y до критического уровня s или ниже заказывается партия объема $S - y \geq q$. В этом случае:

- при накоплении дефицита остаток к концу цикла будет равен нижнему порогу минус «свернутый» спрос, а начальный запас — верхнему уровню минус чистый спрос за время задержки;
- при исключении реального дефицита неотрицательный остаток вычисляется по «свернутому» распределению;
- ожидаемый дефицит, используемый при вычислении штрафа, а также вероятность дефицита в условиях оптимальности рассчитываются для «свернутого» распределения.

При реализации этой стратегии с переносом дефицита текущий запас всегда оценивается в сумме с ранее сделанными, но еще не выполненными заказами.

10.2. Показатели эффективности

Имея вероятности $\{p_n\}$, легко рассчитать требуемые характеристики (коэффициент готовности R , коэффициент обеспеченности, вероятность немедленной поставки и т.д.) как функции от нормативного запаса S_j . Коэффициент обеспеченности определяется как доля заявок, немедленно покрываемых наличным запасом.

Дефициты (отложенные заказы, backorders) случаются, если в процессе выполнения находятся более чем S заявок. Ожидаемое число дефицитов

$$B(S_j) = \sum_{i=S_j}^{\infty} (i - S_j) p(i|\lambda_j \tau_j). \quad (10.2.1)$$

В частности,

$$B(0) = \sum_{i=0}^{\infty} i p(i|\lambda_j \tau_j) = \lambda_j \tau_j \bar{h}_j = \theta_j. \quad (10.2.2)$$

Среднее число немедленно обслуживаемых заявок

$$\begin{aligned} U(S_j) &= \sum_{i=1}^{S_j} i p(i|\lambda_j \tau_j) + S_j \sum_{i=S_j+1}^{\infty} p(i|\lambda_j \tau_j) \\ &= \sum_{i=1}^{\infty} i p(i|\lambda_j \tau_j) - \sum_{i=S_j}^{\infty} (i - S_j) p(i|\lambda_j \tau_j) \\ &= \lambda_j \tau_j \bar{h}_j - B(S_j). \end{aligned} \quad (10.2.3)$$

Согласно последнему уравнению, $U(S_j)$ можно определить как ожидаемое число деталей в процессе выполнения минус ожидаемое число отложенных поставок.

Уровень обслуживания по детали i

$$R(S_i) = U(S_i)/U(0) = [\lambda_i \tau_i \bar{h}_i - B(S_i)]/\lambda_i \tau_i \bar{h}_i = 1 - B(S_i)/B(0). \quad (10.2.4)$$

Уровень обслуживания — удобная мера для многоуровневых систем. Для системы в целом

$$R = 1 - \sum_{i=1}^I \sum_{j=1}^J B(S_{ij}|\lambda_{ij} \tau_{ij}) / \sum_{i=1}^I \sum_{j=1}^J B(0|\lambda_{ij} \tau_{ij}^0). \quad (10.2.5)$$

Ожидаемое число отложенных заявок в (10.2.5) намеренно записано как функция уровней запаса S_{ij} и произведения $\lambda_{ij} \tau_{ij}$; набор задержек τ_{ij} учитывает влияние многоэтапной структуры; τ_{ij}^0 означает среднее время выполнения по детали i на базе j , когда $S_{ij} = 0$.

Знаменатель в (10.2.5) не зависит от переменных S_{ij} , так что оптимизацию можно провести, минимизируя общее число случаев дефицита на нижнем уровне (без депо):

$$B = \sum_{i=1}^I \sum_{j=1}^J \sum_{n=S_{ij}}^{\infty} (n - S_{ij}) p(n|\lambda_{ij} \tau_{ij}). \quad (10.2.6)$$

Однако при анализе *затрат* процессы в депо учитываться должны.

10.3. Задержки

Теперь пора вернуться к многоуровневой системе и посмотреть, как запасы в ней влияют на задержки $\{\tau_j\}$. Проверка ремонтпригодности отказавшей детали делается только на нижнем уровне. Предполагается, что деталь (в этом разделе мы будем опускать индекс ее типа) ремонтируема на уровне базы с вероятностью r_j , в депо с вероятностью $(1 - r_j)\gamma$, восстановлению не подлежит с вероятностью $(1 - r_j)(1 - \gamma)$. После проверки ремонтируемая деталь отсылается на соответствующий уровень. Среднее время на ремонт, включая перевозку в оба конца, есть T_j . При непригодности детали депо немедленно подает заявку на восполнение запаса извне. Задержка зависит от состояния запаса в депо. Если запас имеется, то она в среднем составляет L_j . Иначе возникает дополнительная задержка, расчет среднего значения которой Δ описан ниже.

Среднее время восполнения

$$\tau_j = r_j T_j + (1 - r_j)(L_j + \Delta). \quad (10.3.1)$$

Для расчета дополнительной задержки Δ в высшем звене мы должны определить стационарное распределение состояния запаса в депо. Спрос в депо от базы создается деталями, не ремонтируемыми на этой базе. Интенсивность его есть $(1 - r_j)\lambda_j$. Соответственно в депо интенсивность заявок

$$\lambda_0 = \sum_{j=1}^J (1 - r_j)\lambda_j. \quad (10.3.2)$$

Общий спрос в депо — обобщенный пуассоновский со средней интенсивностью

$$\theta_0 = \sum_{j=1}^J (1 - r_j)\bar{h}_j\lambda_j = \sum_{j=1}^J (1 - r_j)\theta_j. \quad (10.3.3)$$

В депо имеется S_0 запасных частей. В запасе учитываются также детали, находящиеся в ремонте или в пути. Не подлежащая восстановлению деталь немедленно заказывается внешнему поставщику. Средняя задержка внешней поставки равна L_0 . Итак, средняя задержка пополнения для депо

$$\tau_0 = \gamma T_0 + (1 - \gamma)L_0. \quad (10.3.4)$$

Аналогично вышеприведенным результатам для базы j число деталей в цикле пополнения для депо также имеет обобщенное пуассоновское распределение $p(n|\lambda_0\tau_0)$ со средним $\lambda_0\tau_0$. Ожидаемое число отложенных заказов в депо

$$B(S_0|\lambda_0\tau_0) = \sum_{n=0}^{\infty} (n - S_0)p(n|\lambda_0\tau_0). \quad (10.3.5)$$

Это число не должно смешиваться с задержанными на нижнем уровне из (10.2.1) и не входит в целевую функцию (10.2.6). Средняя задержка

$$\Delta = B(S_0|\lambda_0\tau_0)/\theta_0 \quad (10.3.6)$$

варьируется от нуля (при бесконечном запасе) до полного времени внешнего пополнения τ_0 и может быть представлена как произведение

$$\Delta = \delta(S_0)\tau_0. \quad (10.3.7)$$

Очевидно,

$$\delta(S_0) = B(S_0|\lambda_0\tau_0)/\theta_0\tau_0 = B(S_0|\lambda_0\tau_0)/B(0|\lambda_0\tau_0). \quad (10.3.8)$$

Это доля заявок, для которых выполнение задерживается.

10.4. Основы METRIC

METRIC расшифровывается как MultiEchelon Technique for Recoverable Item Control (эшелонированное управление восстанавливаемыми запчастями). Эта методика была разработана в исследовательской группе RAND Corporation для командования тыла ВВС США [201] и использовалась как заказчиком, так и во многих других областях.

Система считается двухуровневой. На нижнем уровне предполагается стратегия восполнения $(S - 1, S)$ и ограниченные ремонтные возможности. На каждой базе входной спрос — составной пуассоновский. Дефицит накапливается. Показателем эффективности служит сумма ожидаемых дефицитов по всем складам. Для характеристики ремонтируемости деталей используются обозначения LRU/SRU (Line/Shop Replaceable Units). Горизонтальный обмен между базами отсутствует. Средние времена ремонта на базе T_{ij} , доставки из депо L_{ij} , ремонта в депо T_{i0} считаются известными константами.

В верхнем звене все детали ремонтируемы ($\gamma = 1$). Спрос — составной (логарифмический) пуассоновский процесс типа (10.2.6). Неявно предполагается отсутствие очередей на восстановление.

10.5. Базовая METRIC

В схеме базовой METRIC

- спрос прогнозировался по характеристикам надежности и с учетом фактических данных;
- ремонтируемые детали имели индивидуальные цены дефицита d_{ij} .

Она позволяла оценивать уже обсуждавшиеся показатели — число задержек и общие вложения в запас для двухэшелонной системы. Решались также задачи распределения фиксированного запаса по каждой номенклатуре: минимизировать число задержек на *базах*

$$\varphi(N) = \sum_{j=1}^J B_j(\{S_{ij}\}) \quad (10.5.1)$$

(см. (10.2.6)) при ограничении

$$\sum_{j=1}^J S_{ij} = \bar{S}_i. \quad (10.5.2)$$

Напомним, что термин «база» имеет не только снабженческий, но и оперативный аспект.

В алгоритме METRIC сначала вычисляется ожидаемое число задержанных поставок для депо по (10.3.5). Для этого используются $\tau_{i0} = T_{i0}$ из (10.3.4), $\lambda_{i0} = \sum_{j=1}^J \lambda_{ij}(1 - r_{ij})$ из (10.3.2) и $p(n|\lambda_{i0}\tau_{i0})$ из (10.1.10). Затем можно рассчитать время дополнительной задержки Δ по (10.3.6) и средние времена восполнения (10.3.1). Наконец, согласно (10.2.6) вычисляется ожидаемое число задержек по всем деталям и базам.

Управление запасами расходуемых деталей в базовой модели METRIC не рассматривалось.

10.6. MOD-METRIC

Мукштадт [179] модифицировал METRIC применительно к структурированным (multi-indent) изделиям — с учетом логических отношений между сборками. Рассматривались два уровня: конечное *изделие*

(мы будем использовать этот термин, столь популярный в отечественной оборонной промышленности) и составная часть (*модуль*). Модульность ускоряет ремонт изделия в целом, который сводится к диагностике на уровне модулей и замене отказавшего, и удешевляет соответствующее оборудование. С другой стороны, ремонт модулей связан с большими задержками, требует специализированного оборудования и квалифицированного персонала. Сосредоточение ремонта модулей со всех баз на уровне депо может вызвать очереди на ремонт, которые в «метрических» моделях обычно не учитываются.

Простой в MOD-METRIC исчисляются по задержке *изделий*. Дефицит модулей оказывает косвенное влияние, задерживая ремонт изделий. По отношению к базовой METRIC в данной схеме сделаны два упрощения:

- Влияние нехватки любых деталей одинаково.
- Спрос — *простой* пуассоновский.

Рассмотрим расчетную схему MOD-METRIC по [179], опуская индекс детали. Изделие считаем собранным из N модулей. Интенсивность отказов *изделия* на базе j равна λ_j , запас S_j . Требуется рассчитать ожидаемое число задержек в системе

$$B(S_1, S_2, \dots, S_J) = \sum_{j=1}^J \sum_{\xi=S}^{\infty} (\xi - S_j) p(\xi|\lambda_j \tau_j). \quad (10.6.1)$$

Используя (10.3.1)–(10.3.4) при $\gamma = 1$ (ремонтируется все), получаем

$$\tau_j = r_j T_j + (1 - r_j)[L_j + \delta(S_0)T_0], \quad (10.6.2)$$

где параметры определяются аналогично METRIC. Время ремонта изделия T_j зависит, однако, от наличия модулей. Предполагается, что ремонт при наличии модулей займет T_j' , иначе будет отложен. Пусть Δ_j — ожидаемая задержка для базы j . Тогда

$$\Delta_j = \sum_{n=1}^N q_{nj} \Delta_{nj}, \quad (10.6.3)$$

$$T_j = T_j' + \Delta_j, \quad j = \overline{1, J}. \quad (10.6.4)$$

Здесь q_{ij} — вероятность того, что ремонта требует модуль i . Интенсивность отказов изделий на базе j

$$\lambda_j = \sum_{i=1}^N \lambda_{ij}. \quad (10.6.5)$$

Следовательно,

$$q_{ij} = \lambda_{ij}/\lambda_j, \quad j = \overline{1, J}, \quad i = \overline{1, N}. \quad (10.6.6)$$

Ожидаемые задержки рассчитываются как в METRIC-модели:

$$\Delta_{ij} = (1/\lambda_{ij}) \left[\sum_{\xi=S_{ij}}^{\infty} (\xi - S_{ij}) p(\xi | \lambda_{ij} \tau_{ij}) \right], \quad (10.6.7)$$

$$j = \overline{1, J}, \quad i = \overline{1, N}.$$

Здесь

$$\delta(S_{i0}) = \frac{1}{\lambda_{i0} T_{i0}} \left[\sum_{\xi=S_{i0}}^{\infty} (\xi - S_{i0}) p(\xi | \lambda_{i0} T_{i0}) \right],$$

$$\tau_{ij} = r_{ij} T_{ij} + (1 - r_{ij}) [L_{ij} + \delta(S_{i0}) T_{i0}],$$

$$\lambda_{i0} = \sum_{j=1}^J \lambda_{ij} (1 - r_{ij}), \quad j = \overline{1, J}, \quad i = \overline{1, N}.$$

Обозначим c_0 стоимость закупки изделия и $\{c_i\}$ — модулей. Общая стоимость

$$C = \sum_{j=1}^J \left(c_0 S_j + \sum_{i=1}^N c_i S_{ij} \right) + c_0 S_0 + \sum_{i=1}^N c_i S_{i0}. \quad (10.6.8)$$

Мукштадт предложил приближенный алгоритм выбора $\{S_{ij}\}$, минимизирующий ожидаемое число недостатков (10.6.1) при заданном C . Близкие к оптимальным назначения для изделия, полученные при различных значениях C , используются как входные данные на втором шаге алгоритма, который применяет маргинальный анализ для оптимального распределения бюджета между изделиями и модулями.

10.7. METRIC реального времени

Эта модель (сокращенно — RTM) предложена Миллером [177]. Здесь акцент делается на распределении: на какую базу послать следующее отремонтированное изделие из депо. Спрос предполагается пуассоновским с интенсивностями $\{\lambda_j\}$, времена восстановления на базах распределены экспоненциально с параметрами $\{\mu_j\}$, дефициты выражаются через отрицательные значения текущего запаса y_j . Целью является минимизация общего числа дефицитов за T дней

$$B = \sum_{j=1}^J T^{-1} \int_0^T [y_j(t)]^- dt. \quad (10.7.1)$$

Здесь $[y_j(t)]^-$ — абсолютная величина дефицита.

Модель базируется на эвристических правилах. Решение принимается в момент t_0 на основе прогноза, что случится за следующие T_j дней (время доставки из депо на базу). На решение влияет «depo reluctance» DR — степень нежелания депо расставаться с имеемым запасом. Предполагается, что

$$DR = \alpha \exp[-\beta y_0(t_0)]. \quad (10.7.2)$$

Экспериментально подобраны $\alpha = 1.5$, $\beta = 2$.

Потребность базы B_j определяется:

- временем транспортировки T_j (константа);
- ежедневным спросом λ_j ;
- долей r_j заявок, ремонтируемых на базе.

Соответствующее описание состояния системы включает в себя:

- наличный запас на базе z_j ;
- число n_{sj} изделий в пути из депо на базу;
- число n_{rj} находящихся в ремонте на базе изделий.

Если сумма $z_j(t_0) + n_{sj}(t_0) + n_{rj}(t_0) < 0$, потребность базы считается бесконечной. Резон прост: посланная на такую базу деталь достоверно уменьшит число случаев дефицита на единицу. Обычно эта сумма неотрицательна, и потребность базы определяется как ожидаемое число дефицитов B_j на момент $t_0 + T_j$. Для расчета B_j должно быть

известно распределение $z_j(t_0 + T_j)$. Это (дискретное) распределение в модели RTM аппроксимируется нормальным.

Первый шаг процедуры — вычисление среднего и дисперсии наличного запаса через интервал T_j :

$$\begin{aligned} z_j &= M[z_j(t_0 + T_j)] = z_j(t_0) + n_{sj}(t_0) + n_{rj}(t_0)Y_0 - (1 - r_j)\lambda_j T_j - r_j \lambda_j Y_j / \mu_j, \\ \sigma_j^2 &= D[z_j(t_0 + T_j)] = n_{rj}(t_0)Y_j(1 - Y_j) + (1 - r_j)\lambda_j T_j + r_j \lambda_j Y_j / \mu_j. \end{aligned} \quad (10.7.3)$$

Здесь $Y_j \doteq 1 - \exp(-\mu_j T_j)$ означает вероятность завершения начатого ремонта через T_j . Множитель $r_j \lambda_j Y_j / \mu_j$ отражает уменьшение запаса вследствие устраняемых на базе поломок за время T_j . Он не равен $r_j \lambda_j Y_j$, поскольку некоторые поломки к моменту T_j будут устранены. Выражение для него следует из

$$\int_0^{T_j} r_j \lambda_j \exp[-(T_j - w)\mu_j] dw = \frac{r_j \lambda_j}{\mu_j} \exp[-(T_j - w)\mu_j] \Big|_0^{T_j} = r_j \lambda_j Y_j / \mu_j.$$

Второй шаг состоит в расчете ожидаемого числа дефицитов. В случае нормального распределения со средним ξ и дисперсией σ^2 (индекс базы временно опущен) плотность распределения наличного запаса

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}[(y - \xi)/\sigma]^2\right\}, \quad (10.7.4)$$

ожидаемое число дефицитов по определению есть

$$B = - \int_{-\infty}^0 y f(y) dy. \quad (10.7.5)$$

Подставляя $w = (y - \xi)/\sigma$ и $dw = \sigma dy$, имеем

$$\begin{aligned} B &= -\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-\xi/\sigma} \exp(-w^2/2)(\sigma w + \xi) dw \\ &= -\frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{-\xi/\sigma} \exp(-w^2/2) w dw - \frac{\xi}{\sqrt{2\pi}} \int_{-\infty}^{-\xi/\sigma} \exp(-w^2/2) dw \\ &= \frac{\sigma}{\sqrt{2\pi}} \exp[-(\xi/\sigma)^2/2] - \frac{\xi}{\sqrt{2\pi}} \int_{-\infty}^{-\xi/\sigma} \exp(-w^2/2) dw = \sigma F(\xi/\sigma), \end{aligned} \quad (10.7.6)$$

где

$$F(y) = \frac{1}{\sqrt{2\pi}} \exp(-y^2/2) - \frac{y}{\sqrt{2\pi}} \int_{-\infty}^y \exp(-w^2/2) dw.$$

Ожидаемое число дефицитов на базе j

$$B_j = \sigma_j F(y_j), \quad (10.7.7)$$

где $y_j = \xi_j / \sigma_j$.

После вычисления $\{B_j\}$ можно принять решение. Если «нежелание» депо $DR > B_j$ для всех j , то поставок не будет. В противном случае изделие будет отправлено на базу с наибольшим B_j . Имитационное моделирование показало, что стратегия RTM снижает ожидаемое число дефицитов на 30% для деталей с высоким спросом (более одной в месяц) без значительных ухудшений для редко требуемых. Среднее уменьшение числа дефицитов составило 20%

Поскольку эта модель требует наличия в депо полной информации о состоянии запасов на момент принятия решения, ее реализация подразумевает наличие в системе снабжения централизованной компьютерной системы.

10.8. Двухэшелонная система с экстренными поставками

Рассмотрим однородную задачу в прежних обозначениях. Поток считаем простым пуассоновским. Каждая база хранит S_j изделий. При появлении заявки и при наличии изделие поставляется потребителю и делается заказ в депо. Депо выполняет заказ, если есть чем, через случайное время t_j с конечным средним τ_j . Пополнение запаса в депо (индекс «0») организуется аналогично и также в схеме $(S - 1, S)$.

Если запас на базе исчерпан, в депо выдается срочный заказ. Если нужной детали нет и там, делается внешний заказ. Пусть E_j — срок экстренной поставки в j из депо и $E_j + E_0$ — внешней. Наша цель — рассчитать среднее время поставки изделия t_j . Определим событие G_j как исчерпание запаса на базе j . Тогда

$$t_j = E_j \Pr[G_j | \bar{G}_0] \Pr[\bar{G}_0] + (E_j + E_0) \Pr[G_j | G_0] \Pr[G_0]. \quad (10.8.1)$$

Все упомянутые вероятности — эрланговские и определяются по правилу

$$\text{erl}(n|\alpha) = p(n|\alpha) / \sum_{\xi=0}^n p(\xi|\alpha) = (\alpha^n/n!) / \sum_{\xi=0}^n \alpha^\xi/\xi! \quad (10.8.2)$$

с параметрами

$$\begin{aligned} \Pr[G_j|\bar{G}_0] &= \text{erl}(S_j|\alpha'_j), & \Pr[G_j|G_0] &= \text{erl}(S_j|\alpha''_j), \\ \Pr[G_0] &= \text{erl}(S_0|\alpha_0), & \Pr[\bar{G}_0] &= 1 - \Pr(G_0). \end{aligned} \quad (10.8.3)$$

Здесь

$$\alpha'_j = \lambda_j \tau_j, \quad \alpha''_j = \lambda_j (\tau_j + \tau_0), \quad \alpha_0 = \tau_0 \sum_{j=1}^J \lambda_j.$$

Рассчитаем среднее время недостачи по базе j . Введем условные вероятности появления на базе j заявки на изделие n

$$q_{nj} = \lambda_{nj} / \sum_{i=1}^I \lambda_{ij}. \quad (10.8.4)$$

Тогда

$$t_j = \sum_{i=1}^I q_{ij} t_{ij}, \quad (10.8.5)$$

где t_{ij} определяются согласно (10.8.1)–(10.8.4) с дополнительным индексом i .

Задержки для системы в целом вычисляются как их взвешенная сумма. Веса можно назначить как важности баз или условные вероятности дефицита на базе j при условии дефицита в системе. Они определяются аналогично $\{q_{ij}\}$.

10.9. Трехуровневая METRIC

Эта модель легко обобщается и на $M > 3$ уровней. В ней рассматриваются только полностью восстанавливаемые изделия, но это допущение не является критическим. Обычный пуассоновский спрос

предполагается только для простоты и наглядности. Цель — расчет ожидаемого числа дефицитов на уровне баз.

Отведем индекс $k = 0$ для высшего уровня, $k = \overline{1, K}$ для второго и $j = \overline{K+1, J}$ — для нижнего. Обозначим S_{ij} , T_{ij} нормативный запас и время выполнения для предмета i в пункте j , $i = \overline{1, I}$, $j = \overline{0, J}$. Состояние запаса будем определять как наличие плюс ранее сделанный, но еще не выполненный заказ минус дефициты. Стратегией выполнения везде считаем $(S-1, S)$, а времена восстановления — независимыми случайными величинами.

Первичный спрос возникает на нижнем уровне с интенсивностями $\{\lambda_{ij}\}$. Как и раньше, отказавшее изделие с вероятностью r_{ij} может быть направлено на местный ремонт, с вероятностью r_{ijk} — в узел k второго уровня и с r_{ij0} — в высшее звено. Если отказавшая деталь ремонтируется вверх, делается заказ на восполнение из соответствующего склада. Если есть чем восполнить, имеет место случайная задержка со средним L_{ijk} , иначе возникает дополнительная задержка.

Спрос на *втором* уровне — также простой пуассоновский с интенсивностью $\lambda_{ik} = \sum_{j=K+1}^J r_{ijk} \lambda_{ij}$. Отказавшая деталь ремонтпригодна в узле k с вероятностью r_{ik} , а с дополнительной к ней вероятностью отправляется на третий уровень. Одновременно выдается заказ на восполнение с центрального склада. Ситуация там аналогична рассмотренной выше.

Простой пуассоновский процесс на *третьем* уровне имеет интенсивность

$$\lambda_{i0} = \sum_{j=K+1}^J r_{ij0} \lambda_{ij} + \sum_{k=1}^K (1 - r_{ik}) \lambda_{ik}. \quad (10.9.1)$$

Все заявки, прибывшие на третий уровень, будут на нем ремонтироваться. Ожидаемое число задержек на верхнем уровне по номенклатуре i

$$B(S_{i0}, \lambda_{i0} T_{i0}) = \sum_{n=S_{i0}}^{\infty} (n - S_{i0}) p(n|\lambda_{i0} T_{i0}),$$

где $p(n|x)$ — пуассоновская вероятность при среднем x . На основании формулы Литтла среднее время задержки в центральном звене

$$\Delta_{i0} = B(S_{i0}, \lambda_{i0} T_{i0}) / \lambda_{i0}.$$

Для складов с индексами $k = \overline{1, K}$ оно должно добавляться ко времени выполнения L_{ik0} . Следовательно, среднее время выполнения

$$\tau_{ik} = r_{ik}T_{ik} + (1 - r_{ik})(L_{ik} + \Delta_{i0}), \quad i = \overline{1, I}, \quad k = \overline{1, K}. \quad (10.9.2)$$

Ожидаемое число дефицитов

$$B(S_{ik}, \lambda_{ik}\tau_{ik}) = \sum_{n=S_{ik}}^{\infty} (n - S_{ik})p(n|\lambda_{ik}\tau_{ik}).$$

Среднее время задержки при пополнении со склада k из-за отсутствия запаса

$$\Delta_{ik} = B(S_{ik}, \lambda_{ik}\tau_{ik})/\lambda_{ik}.$$

Наконец, на нижнем уровне среднее время выполнения

$$\tau_{ij} = r_{ij}T_{ij} + r_{ijk}(L_{ijk} + \Delta_{ik}) + r_{ij0}(L_{ij0} + \Delta_{i0}), \quad (10.9.3)$$

и количество дефицитов

$$B(S_{ij}, \lambda_{ij}\tau_{ij}) = \sum_{n=S_{ij}}^{\infty} (n - S_{ij})p(n|\lambda_{ij}\tau_{ij}). \quad (10.9.4)$$

Общее число дефицитов на нижнем уровне

$$\bar{B} = \sum_{i=1}^I \sum_{j=K+1}^J \sum_{n=S_{ij}}^{\infty} (n - S_{ij})p(n|\lambda_{ij}\tau_{ij}). \quad (10.9.5)$$

10.10. Децентрализованная система

Если задержки поставок из-за нехватки в высшем уровне относительно малы, ими при расчете количества дефицитов можно пренебречь и система может анализироваться как децентрализованная. Эта аппроксимация очень полезна для анализа альтернатив и позволяет определить баланс между вложениями в ремонт и запасы.

Дадим индекс $j = 1, 2, \dots, J$ складам с ремонтными возможностями. Взвешенная сумма дефицитов по всем номенклатурам и складам

$$\bar{B} = \sum_{i=1}^I \sum_{j=1}^J d_{ij} B(S_{ij}, \lambda_{ij}\tau_{ij}) = \sum_{i=1}^I \sum_{j=1}^J d_{ij} \sum_{n=S_{ij}}^{\infty} (n - S_{ij})p(n|\lambda_{ij}\tau_{ij}), \quad (10.10.1)$$

где d_{ij} — весовые коэффициенты. Цель проектирования — получить минимум \bar{B} по $\{S_{ij}\}$ при заданном бюджетном ограничении.

С формальной точки зрения восполнение с высшего уровня может рассматриваться как вид ремонта. Время τ_{ij} выполнения для изделия i в пункте j является средневзвешенным по дефектам N типов, имеющих вероятность r_{ij}^n и среднюю длительность T_{ij}^n . Среднее время

$$\tau_{ij} = \sum_{n=1}^N r_{ij}^n T_{ij}^n. \quad (10.10.2)$$

Управляемыми параметрами считаются времена ремонта T_{ij}^n . Более короткое время требует больших вложений в ремонт — измерительную аппаратуру, инструменты, персонал. Предлагались убывающие зависимости $c_n^R(T_{ij}^n)$ затрат от среднего времени восстановления вида $a_{ij}^n/(T_{ij}^n - b_{ij}^n)$ и $a_{ij}^n \exp[-(T_{ij}^n - b_{ij}^n)/c_{ij}]$. Стоимость содержания запаса обычно считается линейной функцией:

$$c^s(S_{ij}) = c_{ij}S_{ij}.$$

Необходимо так распределить затраты на запас и на ремонтные возможности, чтобы

$$\sum_{i=1}^I \sum_{j=1}^J c_{ij}S_{ij} + \sum_{i=1}^I \sum_{j=1}^J \sum_{n=1}^N c_n^R(T_{ij}^n) \leq \bar{C}. \quad (10.10.3)$$

Портеус и Лэнсдаун [192] предложили алгоритм, решающий сначала подпроблему. Прежде всего определяются положительные числа $T_{ij}^1, T_{ij}^2, \dots, T_{ij}^N$ для минимизации

$$c^R(\tau_{ij}) = \sum_{n=1}^N c_n^R(T_{ij}^n)$$

при ограничении

$$\sum_{n=1}^N r_{ij}^n T_{ij}^n = \tau_{ij}. \quad (10.10.4)$$

Это делается несколько раз для различных значений τ_{ij} , для каждого изделия и пункта. Оптимальные значения $c^R(\tau_{ij})$ включаются в

(10.10.3); затем решается задача минимизации (10.10.1) при ограничении

$$\sum_{i=1}^I \sum_{j=1}^J c_{ij} S_{ij} + \sum_{i=1}^I \sum_{j=1}^J \sum_{n=1}^N c_n^R(\tau_{ij}) \leq \bar{C}. \quad (10.10.5)$$

В обоих случаях целевая функция и ограничения — сепарабельные функции, так что для решения задачи применим метод Лагранжа.

10.11. Стратегии, отличные от $(S-1, S)$

Естественным обобщением рассмотренных подходов является комбинация стратегий УЗ, различающихся по уровням системы. Наиболее типичен случай использования $(S-1, S)$ на базах и (s, S) в депо. Обзор литературы и ссылки на первоисточники по этой теме имеются в [191]. Мы приведем некоторые результаты для (s, S) , (s, Q) и (T, S) . Как отмечалось выше, (s, S) и (s, Q) совпадают только для простого пуассоновского спроса — с единичными заявками. Это предположение не соответствует случаю деталей с высоким спросом и, очень часто, верхним эшелонам.

10.11.1. Спрос в верхнем эшелоне

Ограничимся двухэшелонной системой с депо и несколькими базами без непосредственного обмена между ними. Спрос вверх статистически независим. Запас считаем однородным, индекс типа детали опускаем.

Случай $(S-1, S)$ и пуассоновского (простого или обобщенного) внизу рассматривался в разд. 3.1.1. Спрос наверху — также пуассоновский с суммарным средним. Для остальных случаев предлагаются только аппроксимации, тестированные на имитационных моделях.

Распределение спроса в депо зависит от временного интервала — времени выполнения или периодичности контроля уровня. Чем длиннее интервал, тем точнее прогноз и тем ближе он к суммарному спросу (отличия могут быть в любую сторону: спрос, не приведший к критическому уровню, наверх не передается; с другой стороны, одна дополнительная заявка может породить групповую вверх).

Первый метод предполагает, что спрос $X_j(t)$ за выбранный интервал удовлетворительно описывается нормальным распределением с

параметрами ξ_j, σ_j^2 . Политикой возобновления запаса считается непрерывный просмотр, т.е. (s_j, Q_j) .

Обозначим X_0, X_{0j} соответственно спрос в депо и на базе j за время выполнения запаса в депо. Общий спрос в депо — свертка спроса с баз, так что

$$X_0 = \sum_{j=1}^J X_{0j}. \quad (10.11.1)$$

Для любого временного интервала X_{0j} может быть только кратным объемам Q_j заявок с базы j . Число N_j заказов, поступивших с базы j , зависит от детерминированных переменных s_j, Q_j, z_{j0} и случайного спроса X_j . Ясно, что

$$N_j = \lfloor (s_j + Q_j - z_{j0} + X_j) / Q_j \rfloor, \quad j = \overline{1, J}, \quad (10.11.2)$$

где $\lfloor z \rfloor$ означает наибольшее целое, не превышающее z . Тогда

$$X_{0j} = N_j Q_j = \lfloor (s_j + Q_j - z_{j0} + X_j) / Q_j \rfloor Q_j, \quad j = \overline{1, J}. \quad (10.11.3)$$

Итак, спрос X_{0j} имеет дискретное распределение. Вероятность появления ровно k заказов

$$\begin{aligned} p_{jk} &= \Pr[X_{0j} = kQ_j] = \Pr[k \leq (s_j + Q_j - z_{j0} + X_j) / Q_j \leq k+1] \\ &= \Pr[k-1 \leq (X_j - z_{j0} + s_j) / Q_j \leq k]. \end{aligned} \quad (10.11.4)$$

Полагая $a(j, k) = kQ_j + z_{j0} - s_j$, имеем

$$\begin{aligned} p_{jk} &= \int_{a(j, k-1)}^{a(j, k)} \psi_j(y) dy, \\ p_{j0} &= \int_{-\infty}^{a(j, 0)} \psi_j(y) dy. \end{aligned}$$

10.11.2. Нормальная аппроксимация

Считая плотность распределения спроса ψ_j нормальной с вышеуказанными параметрами и полагая $h_j = s_j - z_{j0} + \xi_j$, получаем

$$p_{jk} = \Pr \left[\frac{(k-1) - h_j / Q_j}{\sigma_j / Q_j} \leq \frac{d_j - h_j}{\sigma_j} \leq \frac{k - h_j / Q_j}{\sigma_j / Q_j} \right]$$

$$= F\left(\frac{k - h_j/Q_j}{\sigma_j/Q_j}\right) - F\left(\frac{k - 1 - h_j/Q_j}{\sigma_j/Q_j}\right), \quad k = 1, 2, \dots,$$

$$p_{j0} = F\left(\frac{-h_j/Q_j}{\sigma_j/Q_j}\right) = F(-h_j/\sigma_j). \quad (10.11.5)$$

Здесь

$$F(y) = \int_{-\infty}^y f(t) dt, \quad f(t) = \frac{1}{\sqrt{2\pi}} \exp(-t^2/2).$$

Имея вероятности $\{p_{jk}\}$, легко получить математическое ожидание и дисперсию спроса с базы j и, наконец,

$$M[X_0] = \sum_{j=1}^J M[X_{0j}], \quad (10.11.6)$$

$$D[X_0] = \sum_{j=1}^J D[X_{0j}].$$

Этот метод имеет два недостатка:

- для больших h_j/Q_j и J велика трудоемкость счета;
- требуются ξ_j и σ_j^2 для каждой базы.

10.12. Стратегия с непрерывным просмотром

Вернемся к двухэшелонной модели с простым пуассоновским спросом $\{\lambda_j\}$ на базах. Предположим, что базы используют стратегию с непрерывным просмотром (s_j, Q_j) , из-за единичного объема заявок идентичную (s_j, S_j) . Интервалы между заказами с базы j образуют рекуррентный поток событий $\{N_j(t)\}$, $j = \overline{1, J}$, где $N_j(t)$ — число заказов на отрезке $[0, t]$. Спрос в депо образуется наложением J независимых процессов с общим числом заявок в депо

$$N_0(t) = \sum_{j=1}^J N_j(t). \quad (10.12.1)$$

Без потери общности можно предположить, что каждая база поместила заказ в момент $t = 0$. Времена T между заказами с базы имеют

эрланговские распределения порядка Q_j с функцией распределения

$$G_j(t) = 1 - \exp(-\lambda_j T) \sum_{k=0}^{Q_j-1} (\lambda_j T)^k / k! \quad (10.12.2)$$

со средним Q_j/λ_j .

При большом J и однородных базах с $\lambda_j = \lambda_B$ получим $\lambda_0 = J\lambda_B$. Времена между заказами для этой «объединенной» базы имеют распределения Эрланга порядка Q_B со средним Q_B/λ_0 .

В депо часто применяют *периодический* контроль запаса. Например, в Eastman Kodak Company [196] в центре применяется стратегия (T, Q) , на региональных базах (s, Q) . Поставка с фабрики в центр выполняется через L_0 , распределение спроса в центре за $T_0 + L_0$ считается нормальным. Уровень S_0 рассчитывается по заданному «центральному» уровню обслуживания D_0 как решение уравнения

$$1 - D_0 = \int_{S_0}^{\infty} (a - S_0) f_A(a) da / M[z_0(T_0)], \quad (10.12.3)$$

где $M[z_0(T_0)]$ означает ожидаемый спрос в центре за интервал T_0 между просмотрами.

Объемы заказов Q_j на местах рассчитываются по стандартной формуле Уилсона. Точка заказа s_j находится из уравнения

$$1 - D_j = \int_{s_j}^{\infty} (b - s_j) f_j(b) db / Q_j, \quad (10.12.4)$$

распределение спроса берется $f_j(b)$ за время L_j и считается нормальным.

Фактический уровень обслуживания для региональных центров уменьшается, поскольку $D_0 < 100\%$. По определению это

$$C_j = 1 - M[B_j]/\xi_j, \quad (10.12.5)$$

где $M[B_j]$ — ожидаемое число задержек за год и ξ_j — ожидаемый годовой спрос. Розенбаум различает два случая в зависимости от наличного запаса в центре Z_0 в момент прибытия «местной» заявки:

$$M[B_j] = M[B_j | Z_0 > 0] \Pr[Z_0 > 0] + M[B_j | Z_0 = 0] \Pr[Z_0 = 0].$$

Для первого члена

$$M[B_j|Z_0 > 0] = (1 - D_j)\xi_j, \quad \Pr\{Z_0 > 0\} = D_0.$$

Второй член есть ожидаемое число задержек, случающихся в год за время, когда в депо нет запаса.

Глава 11

Оптимизация систем обеспечения запчастями

Важнейшая из проблем в этой области — оптимизация уровней запасов. Она интересна как с математической точки зрения (многомерная, нелинейная, целочисленная), так и практически. Основным показателем для ремонтируемых изделий служит ожидаемое число дефицитов при ограничениях на бюджет. По расходуемым (consumable) деталям часто минимизируется ожидаемое время простоя. На базах предполагается правило заказов $(S - 1, S)$.

11.1. Задача о рюкзаке

11.1.1. Постановка задачи

Установившееся в литературе и вынесенное в заголовок раздела название описываемого ниже метода объясняется очевидной его применимостью при укладке альпинистского рюкзака (по убывающей полезности предметов при ограничении на суммарный вес). Классическую (но не единственную) область его использования в прикладной математике образуют задачи теории надежности резервируемых систем, в первую очередь комплектование ЗИПа.

Расчет невозполняемого ЗИПа обычно формулируется как статическая многономенклатурная задача с ограничениями. Типичный ее

пример приводится в [89 с 346] нужно определить количество и типы запасного оборудования для атомной подводной лодки и ее ракет которые следует взять на борт на время патрулирования. Предполагается что для времени патрулирования известно распределение объема требований на запасное оборудование. Объем отсеков для хранения запасного оборудования строго ограничен¹. По такому же принципу можно комплектовать груз для космического корабля, здесь более существенным представляется ограничение на суммарный вес.

Многие задачи теории управления запасами после чисто терминологической модификации сводятся к задачам оптимального резервирования. К показателям надежности относятся *вероятность безотказной работы* системы в течение заданного промежутка времени и *коэффициент готовности* — вероятность застать систему работоспособной в случайный момент времени. Задачи оптимального резервирования возникают тогда, когда существуют определенные ограничения на затрачиваемые для повышения надежности средства. Очевидно, затраты того или иного вида ресурсов определяются количеством резервных элементов и как правило растут линейно по числу элементов.

Прямая задача требует найти такое количество резервных элементов, чтобы требуемый показатель надежности системы обеспечивался при минимальных суммарных затратах на все резервные элементы.

В *обратной задаче* отыскивается такое количество резервных элементов для каждой подсистемы, чтобы при заданных допустимых затратах на систему в целом обеспечивался максимально возможный показатель надежности системы. Обратная задача может решаться и при нескольких ограничениях.

Идея решения упомянутых задач оптимального резервирования легко уясняется из графического представления двумерной проблемы. Построим координатную сетку с целочисленными узлами (n_1, n_2) , задающими кратность резервирования первой и второй подсистемы — рис 11.1. На этой сетке можно провести пунктирные ломаные линии которые отграничивают внизу слева области с суммарными затратами, не превышающими значения $C = C_1 < C_2 < C_3$. Поскольку уменьшение любой из координат выбранной точки приводит к уменьшению

¹Надо иметь в виду, что если остались свободными два кубометра, это не означает, что удастся разместить там агрегат объемом 17 кубометра. Существенно соответствие формы свободного пространства форме детали.

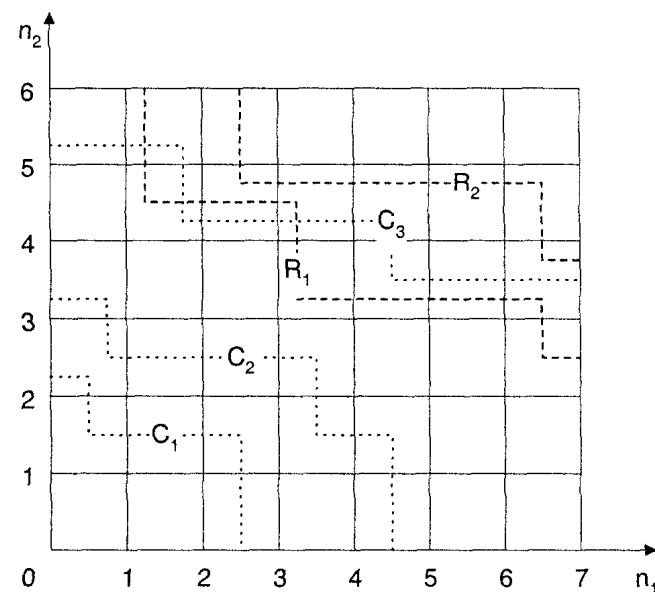


Рис. 11.1. К задаче оптимального резервирования

показателя надежности, решение обратной задачи дается одной из внутренних точек упомянутой C -допустимой области, прилегающих к границе $C = C_0$ (заданный объем дефицитного ресурса).

С другой стороны, подсчитав для некоторого множества точек обеспечиваемый ими показатель надежности R , можно провести аналогичные границы расположенных сверху-справа R -допустимых областей $R_1 < R_2 < \dots$ (штриховые линии). Решение прямой задачи содержится среди точек, прилегающих сверху и справа к границе $R = R_0$. Определенные таким образом точки называются *доминирующими*, поскольку никакие другие не могут быть оптимальными решениями. При ограничениях $R \geq R_1$ и $C \leq C_3$ единственным допустимым решением является точка $(4,4)$.

Запишем показатель *ненадежности* системы в виде

$$Q(S) \approx \sum_{i=1}^N Q_i(s_i), \quad (11.1.1)$$

где

$$Q_i(s_i) = q_i^{s_i+1} \quad (11.1.2)$$

(используется нагруженный резерв). Будем искать

$$\min_S \sum_{i=1}^N c_i s_i$$

при ограничении

$$Q(S) = \sum_{i=1}^N q_i^{s_i+1} = Q_0.$$

Запишем вспомогательную функцию Лагранжа

$$F(S) = C(S) + \varphi Q(S) = \sum_{i=1}^N c_i s_i + \varphi \sum_{i=1}^N q_i^{s_i+1}.$$

Дифференцируя ее по s_i , получаем условия

$$c_i + \varphi (\ln q_i) q_i^{s_i+1} = 0,$$

откуда

$$q_i^{s_i+1} = -c_i / (\varphi \ln q_i). \quad (11.1.3)$$

Положим $\alpha_i = -c_i / \ln q_i$, перепишем (11.1.3) в форме

$$q_i^{s_i} = \frac{\alpha_i}{\varphi q_i} \quad (11.1.4)$$

и подставим (11.1.3) в уравнение связи:

$$\frac{1}{\varphi} \sum_{i=1}^N \alpha_i = Q_0.$$

Теперь ясно, что

$$\varphi = \frac{1}{Q_0} \sum_{i=1}^N \alpha_i.$$

В свою очередь, из (11.1.4) следует

$$s_i = \frac{1}{\ln q_i} \ln \frac{\alpha_i}{\varphi \ln q_i}.$$

Окончательно

$$s_i = \frac{1}{\ln q_i} \ln \left(\frac{Q_0}{q_i} \alpha_i / \sum_{j=1}^N \alpha_j \right). \quad (11.1.5)$$

Поскольку, вообще говоря, получаются нецелые значения $\{s_i\}$, требуется перебор 2^N точек — с округлением координат в меньшую и большую стороны. Из них исключаются точки с недопустимым коэффициентом ненадежности, а из оставшихся выбирается связанная с минимальными затратами. Подчеркнем, что истинно оптимальной точки среди полученных округлением координат может и не оказаться. Это относится ко всем решениям целочисленных задач, отыскиваемым округлением вещественных решений.

Сформулируем теперь обратную задачу оптимального резервирования. Будем искать

$$\min_S Q(S) = \min_S \sum_{i=1}^N q_i^{s_i+1}$$

при ограничении

$$C(S) = \sum_{i=1}^N c_i s_i = C_0.$$

Запишем вспомогательную функцию Лагранжа

$$F(S) = Q(S) + \psi C(S) = \sum_{i=1}^N q_i^{s_i+1} + \psi \sum_{i=1}^N c_i s_i.$$

Дифференцируя ее по s_i , получаем условия

$$\psi c_i + (\ln q_i) q_i^{s_i+1} = 0,$$

откуда

$$q_i^{s_i} = -\frac{\psi c_i}{q_i \ln q_i} = \frac{\psi \alpha_i}{q_i}, \quad (11.1.6)$$

где α_i имеет прежнее значение. Теперь

$$s_i = \frac{\ln(\psi \alpha_i / q_i)}{\ln q_i} = \frac{\ln \psi}{\ln q_i} + \frac{\ln(\alpha_i / q_i)}{\ln q_i}. \quad (11.1.7)$$

Подставим результат в уравнение связи:

$$\ln \psi \sum_{i=1}^N \frac{c_i}{\ln q_i} + \sum_{i=1}^N \frac{c_i \ln(\alpha_i / q_i)}{\ln q_i} = C_0.$$

Оно преобразуется в

$$-\ln \psi \sum_{i=1}^N \alpha_i - \sum_{i=1}^N \alpha_i \ln(\alpha_i/q_i) = C_0,$$

откуда

$$\ln \psi = - \left(\sum_{i=1}^N \alpha_i \ln(\alpha_i/q_i) + C_0 \right) / \sum_{i=1}^N \alpha_i.$$

Подставив результат в (11.1.7), находим

$$s_i = - \frac{1}{\ln q_i} \left[\left(\sum_{i=1}^N \alpha_i \ln(\alpha_i/q_i) + C_0 \right) / \left(\sum_{i=1}^N \alpha_i \right) + \ln(\alpha_i/q_i) \right]. \quad (11.1.8)$$

К этому решению можно сделать замечания, аналогичные предыдущему случаю.

11.1.2. Основная расчетная схема

Метод состоит в *последовательном* наращивании решения и позволяет решать как прямую, так и обратную задачу (разница заключается только в критерии завершения).

Определим результирующий показатель надежности

$$R = \prod_{i=1}^N R_i$$

и будем на очередном шаге увеличивать запас по той i^* -й подсистеме, для которой достигается максимум относительного удельного приращения

$$\gamma_i = \frac{R_i(s_i + 1) - R_i(s_i)}{c_i R_i(s_i)}. \quad (11.1.9)$$

Процесс решения прямой задачи останавливается, когда достигнута требуемая надежность. При решении обратной задачи после каждого увеличения запаса корректируется остаток ресурса $z = z - c_{i^*}$, и очередное приращение выбирается только на множестве номенклатур $I(z)$, для которых $c_i \leq z$. Процесс завершается, когда $I(z)$ оказывается пустым.

На заключительных шагах, когда достигнута надежность достаточно высока ($R_i \approx 1$), вместо (11.1.9) применяется отношение

$$\gamma_i = \frac{Q_i(s_i) - Q_i(s_i + 1)}{c_i R_i(s_i)} \approx \frac{Q_i(s_i)}{c_i}. \quad (11.1.10)$$

11.1.3. Ограничения и обобщения метода

Описанный метод (максимума удельных приращений) не гарантирует получение строго оптимального решения, поскольку остановка процесса возможна при положительном остатке z . В таких случаях можно вернуться назад на один шаг, выбрать приращение по другой допустимой позиции (с меньшим значением c_i) и продолжить процесс. Перебор вариантов существенно увеличивает трудоемкость алгоритма, в особенности если начать ветвление с отката на большее число шагов. «Свертка» последних шагов может быть реализована методом динамического программирования.

Может оказаться, что в прямой задаче достижения заданной надежности в результате последнего шага требуемый показатель был превышен. В этом случае можно получить более дешевое решение задачи: выбрать из тех позиций, по которым

$$R(S)[Q_i(s_i) - Q_i(s_i + 1)]/R_i(s_i) \geq R_0$$

или

$$\frac{Q_i(s_i) - Q_i(s_i + 1)}{R_i(s_i)} \geq R_0/R(S),$$

имеющую минимальное значение c_i . Здесь также возможен многошаговый откат.

Большое число шагов, необходимое для получения решения, порождает желание начать процесс с достаточно близкой к решению точки. Рассмотрим сначала прямую задачу — достижение надежности R_0 при минимальных затратах. Будем считать, что

$$Q_0 \approx \sum_{i=1}^N Q_i(s_i) \ll 1.$$

С другой стороны, все относительные приращения

$$\gamma_i \approx \frac{Q_i(s_i)}{c_i} \approx u$$

независимо от i , т.е. достаточно близки друг к другу (по абсолютной величине). Тогда

$$Q_0 = \sum_{i=1}^N Q_i(s_i) \approx u \sum_{i=1}^N c_i$$

и

$$u = Q_0 / \sum_{i=1}^N c_i.$$

Определив u , последовательно увеличиваем $\{s_i\}$, пока $\{\gamma_i(s_i)\} < u$. Затем переходим к основному алгоритму. Можно сразу получить приближенное решение прямой задачи, решив уравнения

$$s_i = Q_i^{-1} \left(Q_0 c_i / \sum_{i=1}^N c_i \right).$$

При комплектовании ЗИПа расчетную *ненадежность* обеспечения можно распределять между номенклатурами пропорционально единичным стоимостям — см. [25, с. 152].

Для решения обратной задачи вычисляется (см. [86, с. 109])

$$\ln u = [C_0 - \sum_{i=1}^N c_i (\ln c_i / \ln q_i - 1)] / \sum_{i=1}^N \alpha_i.$$

Затем находится u и используется так же, как для прямой задачи

В случае *нескольких* ограничений обычно решается обобщенная обратная задача (максимум надежности при выполнении всех ограничений). Решение отыскивается по максимуму удельных приращений с контролем на каждом шаге всех ограничений. Метод реализуется в следующих вариантах:

- 1) *Последовательная* по всем ресурсам оптимизация (если первым нарушено ограничение именно по текущему ресурсу или по всем сразу, решение оптимально).
- 2) *Управляемая смена* «ведущего ресурса» — в качестве очередного рассматривается ресурс, по которому первым нарушено ограничение, пока цикл не замкнется. Максимумы надежности запоминаются и выбирается наибольший.
- 3) *Отражающий экран* При нарушении ограничения по j -му фактору из набора исключаются несколько «наиболее весомых» по этому фактору элементов, и процесс последовательной оптимизации продолжается с полученной новой исходной точки.

- 4) *Поочередная оптимизация*. На каждом шаге меняется ведущий ресурс.
- 5) *«Диагональная» оптимизация*. На каждом шаге выбирается ведущим тот ресурс j , по которому достигает максимума «недорасход» $\delta C_j = \Delta C_j / \sum_{i=1}^N c_{ij}$. Возможны и другие варианты взвешивания недорасходов $\{\delta C_j\}$.

11.2. «Рюкзак» и двухуровневая система

Опишем применение метода максимума удельных приращений к симметричной системе из J одинаковых баз и одного депо. В качестве целевой функции выберем суммарный простой пользователей:

$$t = \sum_{i=1}^I q_i t_i(S_{i1}, S_{i0}). \quad (11.2.1)$$

Простой i -го пользователя можно подсчитать согласно

$$t_i = [1 - \text{erl}(S_{i0}|\alpha_0)] \text{erl}(S_{i1}|\alpha'_i) T_1 + \text{erl}(S_{i0}|\alpha_i^0) \text{erl}(S_{i1}|\alpha''_i) (T_0 + T_1), \quad (11.2.2)$$

где

$$\alpha'_i = \lambda_i \tau_{i1}, \quad \alpha''_i = \lambda_i (\tau_{i1} + \tau_{i0}), \quad \alpha_i^0 = \lambda_{i0} \tau_{i0} = J \lambda_i \tau_{i0}, \quad (11.2.3)$$

а T_0, T_1 — сроки экстренных поставок для соответствующих уровней (обычные задержки обозначены через τ).

При увеличении S_{i1} на единицу приращение целевой функции

$$\delta t = q_i [t_{si}(S_{i1} + 1, S_{i0}) - t_{si}(S_{i1}, S_{i0})], \quad (11.2.4)$$

а затраты возрастают на Jc_i . При аналогичном увеличении запаса в депо

$$\delta t_s = q_i [t_{si}(S_{i1}, S_{i0} + 1) - t_{si}(S_{i1}, S_{i0})], \quad (11.2.5)$$

при дополнительных затратах c_i . Значения функции erl легко вычисляются рекуррентно. Приведем соответствующий фрагмент Фортран-программы:

```
b=1.0; c=1.0; n=0
erl(0)=1.0
```

```

do n=1, nmax
  b=b*a/n; c=c+b
  erl(n)=b/c
end do

```

Теперь можно применить к поставленной задаче алгоритм предыдущего раздела.

Разумеется, во всех случаях будут дополнительные задержки на испытательном оборудовании (в том числе из-за его неготовности), очереди на ремонт и транспортировку и т.п.

Аналогично можно решать несимметричную задачу.

Алгоритм не гарантирует оптимальности решения исходной задачи, но для полученного C^* не существует решения с лучшим значением целевой функции. Очень легко исследовать чувствительность к ограничению — достаточно его несколько ослабить и посмотреть получаемые значения целевой функции для нужных контрольных точек.

Метод максимума удельных приращений можно применить для трехуровневой системы — например, посредством декомпозиции задачи. На *нижнем* уровне можно минимизировать затраты при готовности пользователей (достаточность запаса по всем номенклатурам) не ниже заданной. Для единственного *депо* будем искать минимум затрат на запас в нем при ограничении на ожидаемые задержки из-за дефицита (запасы на первом уровне уже фиксированы). Дефицит в депо покрывается посредством экстренных поставок. На *третьем уровне* (центральный склад) максимизируется готовность пользователей при ограничении на стоимость запасов в нем. Расчет готовности выполняется как на первом уровне, запас и спрос по всем номенклатурам считаются суммарными.

11.3. Лагранжевы методы для систем снабжения

11.3.1. Постановка задачи для двухуровневой системы

Структуру системы образуют базы и депо, все с ремонтными возможностями. Отказавшее изделие немедленно заменяется на запасное и поступает в ремонт — в зависимости от его сложности местный или

«центральный». Горизонтальные связи в системе отсутствуют. Задача: найти минимум

$$B = \sum_{i=1}^I \sum_{j=1}^J B_{ij}(S_{ij}, S_{i0}), \quad \sum_{j=0}^J \sum_{i=1}^I c_i S_{ij} \leq \bar{C}, \quad S_{ij} = 0, 1, 2, \dots \quad (11.3.1)$$

Ожидаемое число дефицитов по детали i на базе j

$$B_{ij}(S_{ij}, S_{i0}) = \sum_{n=S_{ij}+1}^{\infty} (n - S_{ij}) p(n | \lambda_{ij} \tau_{ij}). \quad (11.3.2)$$

При этом $\tau_{ij} = \tau_{ij}(S_{i0})$. Точнее,

$$\tau_{ij} = r_{ij} T_{ij} + (1 - r_{ij}) [L_{j0} + \delta(S_{i0})]. \quad (11.3.3)$$

В этой формуле

T_{ij} — среднее время ремонта детали i на базе j ,

r_{ij} — вероятность ремонта детали i на базе j ,

L_{j0} — среднее время доставки детали с базы j в депо.

Дополнительно отметим равенства для средней задержки в депо

$$\delta(S_{i0}) T_{ij} = \frac{1}{\lambda_{i0}} \sum_{n=S_{i0}+1}^{\infty} (n - S_{i0}) p(n | \lambda_{i0} T_{i0}) \quad (11.3.4)$$

и интенсивности потока заявок на ремонт в депо

$$\lambda_{i0} = \sum_{j=1}^J (1 - r_{ij}) \lambda_{ij}. \quad (11.3.5)$$

11.3.2. Процедура Фокса и Лэнди

Эта процедура [142] адаптирует к поставленной задаче метод множителей Лагранжа. Разыскиваются $\{S_{ij}\}$, которые минимизируют лагранжиан

$$L = \sum_{j=1}^J \sum_{i=1}^I B_{ij}(S_{ij}, S_{i0}) + \theta \sum_{j=0}^J \sum_{i=1}^I c_i S_{ij}. \quad (11.3.6)$$

При разработке алгоритма использованы сепарабельность целевой функции по i , строгое убывание $B_{ij}(S_{ij}, S_{i0})$ по обоим своим аргументам, дискретная выпуклость и рекурсивность по первому аргументу.

Вследствие сепарабельности по i оптимальное решение задачи (11.3.6) находится последовательным решением I подзадач вида

$$W_i(\theta) = \min_{S_{ij}, S_{i0}} \left\{ \sum_{j=1}^J B_{ij}(S_{ij}, S_{i0}) + \theta \sum_{j=0}^J c_i S_{ij} \right\} \quad (11.3.7)$$

с учетом требования целочисленности. Разделяя далее переменные S_{i0} и S_{ij} , перепишем эту задачу как

$$W_i(\theta) = \min_{S_{i0}} \left\{ \theta c_i S_{i0} + \min_{S_{ij}} \left[\sum_{j=1}^J B_{ij}(S_{ij}, S_{i0}) + \theta c_i S_{ij} \right] \right\} \quad (11.3.8)$$

(при вычислении внутреннего минимума переменная S_{i0} считается фиксированной). Обозначим содержимое фигурных скобок

$$Z_i(S_{i0}, \theta) = \theta c_i S_{i0} + \min_{S_{ij}} \left[\sum_{j=1}^J B_{ij}(S_{ij}, S_{i0}) + \theta c_i S_{ij} \right]. \quad (11.3.9)$$

Теперь подзадача (11.3.8) может быть поставлена как нахождение для всех i

$$W_i(\theta) = \min_{S_{i0}} Z_i(S_{i0}, \theta). \quad (11.3.10)$$

Введем оператор минимума в (11.3.9) под знак суммы:

$$Z_i(S_{i0}, \theta) = \theta c_i S_{i0} + \sum_{j=1}^J \min_{S_{ij}} \left[B_{ij}(S_{ij}, S_{i0}) + \theta c_i S_{ij} \right]. \quad (11.3.11)$$

Чтобы определить $Z_i(S_{i0}, \theta)$, мы должны для каждого S_{i0} найти

$$\min_{S_{ij}} \left[B_{ij}(S_{ij}, S_{i0}) + \theta c_i S_{ij} \right]. \quad (11.3.12)$$

Так как эта функция монотонно убывает, дискретно выпукла и рекурсивна по аргументу S_{ij} , нахождение минимума в (11.3.12) сводится к нахождению первого целого S_{ij} , начиная с нуля, для которого выполняется

$$B_{ij}(S_{ij}, S_{i0}) - B_{ij}(S_{ij} + 1, S_{i0}) \leq \theta c_i. \quad (11.3.13)$$

Фокс и Лэнди предложили выбирать значения множителя Лагранжа из диапазона $[\theta_0, \theta_M]$, разбив его на M равных частей. Построение этого интервала — главная проблема при реализации алгоритма. В примере из [191] первоначальный диапазон множителя Лагранжа охватывал четыре порядка, а в дальнейшем суживался до двух порядков. В процессе поиска диапазон делился на 64 равных отрезка.

11.3.3. Половинное деление

При поиске множителя Лагранжа можно использовать метод бисекции (половинного деления): на каждом шаге определять такой интервал (θ_L, θ_U) , чтобы внутри него происходила смена знака функции

$$\sum_{i=1}^I \sum_{j=0}^J c_i S_{ij}(\theta) - \bar{C}.$$

Затем интервал делится пополам и определяется, в какой из половин меняется знак. После этого процедура применяется к ней. Условием завершения процесса служит близость к нулю модуля вышеприведенной разности. В [191] отмечается хорошая сходимость метода при допуске в 0.5% бюджета (с учетом погрешности исходных данных вряд ли стоит стремиться к большей точности) требовалось не более 10 итераций.

Главный недостаток метода бисекции в том, что требуется переисчислять запас для всех баз по всем изделиям на каждом шаге. Для большого числа номенклатур нужно применять другой критерий завершения и ограничивать число итераций.

Процедура Фокса — Лэнди и бисекция требуют хороших оценок диапазона для множителя Лагранжа

11.3.4. Метод Мушштадта

Мушштадт [180] предложил решать исходную задачу (11.3.1) с помощью двух аппроксимаций: для запасов в депо и множителя Лагранжа. Метод опирается на тот факт, что число дефицитов в каждом пункте в практически интересных ситуациях хорошо приближается экспоненциальной функцией максимального запаса S_{ij} .

Ожидаемое число «низовых» дефицитов по детали i может быть

вычислено как функция суммарного запаса этих деталей $N_i = \sum_{j=0}^J S_{ij}$:

$$B_i(N_i) = \sum_{j=1}^J \sum_{n=S_{i,j}+1}^{\infty} (n - S_{ij}) p[n|\lambda_{ij}\tau_{ij}(S_{i0})]. \quad (11.3.14)$$

В целях аппроксимации $B_i(N_i)$ экспонентой

$$B_i(N_i) = a_i \exp(-b_i N_i) \quad (11.3.15)$$

предлагается для всех i решить задачу вычисления

$$Z_i(N_i) = \min_{\{S_{i,j}\}} \sum_{j=1}^J \sum_{n=S_{i,j}+1}^{\infty} (n - S_{ij}) p[n|\lambda_{ij}\tau_{ij}(S_{i0})] \quad (11.3.16)$$

при дополнительном условии

$$\sum_{j=1}^J S_{ij} = N_i, \quad S_{ij} = \overline{0, N_i}.$$

Пусть она решена для K наборов $\{N_i^k\}$ и получены оптимальные значения $Z_i^k = Z_i(N_i^k)$, $k = \overline{1, K}$. Коэффициенты $\{a_i, b_i\}$ получаются как параметры регрессии, именно

$$a_i = \exp \left\{ \frac{\sum (N_i^k)^2 \sum \ln Z_i^k - \sum N_i^k \sum N_i^k \ln Z_i^k}{K \sum (N_i^k)^2 - (\sum N_i^k)^2} \right\},$$

$$b_i = - \frac{K \sum N_i^k \ln Z_i^k - \sum N_i^k \sum \ln Z_i^k}{K \sum (N_i^k)^2 - (\sum N_i^k)^2}.$$

Здесь все суммы берутся по k .

Оптимальные решения задачи (11.3.16) получаются решением $N_i + 1$ задач минимизации

$$\hat{Z}_i(N_i, S_{i0}) = \min_{\{S_{i,j}\}} \sum_{j=1}^J \sum_{n=S_{i,j}+1}^{\infty} (n - S_{ij}) p[n|\lambda_{ij}\tau_{ij}(S_{i0})] \quad (11.3.17)$$

при ограничениях

$$\sum_{j=1}^J S_{ij} = N_i - S_{i0}, \quad S_{ij} = 0, 1, \dots, \quad S_{i0} = \overline{0, N_i}.$$

Тогда

$$Z_i(N_i) = \min_{\{S_{i0}\}} \hat{Z}_i(N_i, S_{i0}). \quad (11.3.18)$$

Эта задача решается «рюкзачным» методом: начиная с $S_{ij} = 0$, $j = \overline{1, J}$, увеличивают на единицу те значения S_{ij} , которые дают наибольшее уменьшение ожидаемого числа дефицитов, пока не будет достигнуто ограничение. Хотя алгоритм прост и эффективен, он должен быть применен к каждой номенклатуре для всех возможных значений S_{i0} , т.е. $\sum_{i=1}^I (N_i + 1)$ раз, и с точки зрения вычислений задача нетривиальна. Значительная экономия достигается, если мы оценим оптимальное значение S_{i0}^* . Тогда вместо $(N_i + 1)$ раз задачу (11.3.17) нужно решить лишь несколько раз в окрестности S_{i0}^* .

Мукштадт показал, что оптимальный запас в депо может быть хорошо аппроксимирован выражением

$$\hat{S}_{i0} = - \frac{1}{b_0} \ln \left(\frac{1}{a_0 b_0} \right), \quad (11.3.19)$$

где a_0, b_0 — параметры экспоненты, аппроксимирующей ожидаемое число недостатков в депо. Эти параметры также определяются посредством регрессионного анализа, основанного на данных S_{i0} и $B_0(S_{i0}|\lambda_{i0}\tau_{i0})$. Полезность аппроксимации возрастает при увеличении оптимальных значений S_{i0} .

Определение параметров $\{a_i, b_i\}$, $i = \overline{1, I}$, представляет первую фазу предлагаемых аппроксимационных алгоритмов. В первом алгоритме она сопровождается оценкой оптимального значения множителя Лагранжа

$$\hat{\theta} = \exp \left(\frac{\alpha - \bar{C}}{\beta} \right), \quad (11.3.20)$$

где

$$\alpha = \sum_{i=1}^I c_i d_i / b_i, \quad \beta = \sum_{i=1}^I c_i / b_i, \quad d_i = \ln(a_i b_i / c_i). \quad (11.3.21)$$

После этого поставленная проблема решается описанными выше методами.

Важнейшее достоинство этого подхода — не требуется знать диапазон для множителя Лагранжа. Время счета оказывается много меньше. Уменьшаются также число проходов и объем хранимых данных.

В сравнении с другими процедурами, описанными здесь, второй алгоритм показывает значительные преимущества в определении уровня запаса $\{N_i\}$. Первая фаза его сопровождается вычислением

$$N'_i = \frac{d_i - (\alpha - \bar{C})/\beta}{b_i}, \quad i = \overline{1, I}, \quad (11.3.22)$$

и их округлением. Отрицательные значения заменяются нулями. Соответствующие индексы суммирования i в дальнейшем пропускаются. Затем рассчитываем новые значения для α, β и снова применяем аппроксимацию (11.3.22).

11.4. Сатисфакционный подход

Сатисфакционным называется не претендующее на оптимальность, но приемлемое решение. Приемлемость обычно обеспечить намного легче, поскольку снимается проблема сопоставления ущерба от дефицита и затрат на создание и поддержание запаса. Это особенно существенно для реальных задач высокой размерности. Получение приемлемого исходного решения является начальным шагом многих методов оптимизации.

Для примера вновь рассмотрим симметричную задачу с J периферийными базами. Показатель эффективности — ожидаемое время простоя из-за дефицита — считается заданным. Математически этот показатель — сепарабельная функция своих аргументов. Заданное значение

$$\bar{t} = \sum_{i=1}^I q_i t_i \quad (11.4.1)$$

надо распределить по номенклатурам. Обычно используемые способы распределения: одинаковые задержки, равный вклад, с учетом стоимостей (допустить большие задержки для дорогих деталей). Для равного вклада

$$\bar{t}_i = \bar{t}_I / q_i, \quad i = \overline{1, I}. \quad (11.4.2)$$

Для последнего способа, полагая относительную стоимость $\sigma_i = c_i / \sum_{i=1}^I c_i$, принимаем

$$\bar{t}_i = \bar{t} \sigma_i / q_i, \quad i = \overline{1, I}. \quad (11.4.3)$$

Вспомним теперь, что

$$t_i(S_{i1}, S_{i0}) = T_1 \text{erl}(S_{i1} | \alpha'_i) [1 - \text{erl}(S_{i0} | \alpha_{i0})] + (T_0 + T_1) \text{erl}(S_{i1} | \alpha''_i) \text{erl}(S_{i0} | \alpha_{i0}), \quad (11.4.4)$$

и проанализируем это выражение. Имеются несколько характерных случаев.

1. Если для некоторого i выбрано $S_{i1} = S_{i0} = 0$, то вероятности задержек

$$\text{erl}(S_{i1} | \alpha'_i) = \text{erlm}(S_{i1} | \alpha''_i) = \text{erl}(S_{i0} | \alpha_{i0}) = 1, \quad (11.4.5)$$

и

$$t_i = T_0 + T_1. \quad (11.4.6)$$

Значит, если оказалось $\bar{t}_i \geq T_0 + T_1$, то запастись данной деталью не нужно — задержка и так окажется допустимой.

2. Если $S_{i1} = 0$ и $S_{i0} \rightarrow \infty$, получаем

$$\text{erl}(S_{i1} | \alpha'_i) = \text{erl}(S_{i1} | \alpha''_i) = 1, \quad \text{erl}(S_{i0} | \alpha_{i0}) = 0 \quad (11.4.7)$$

и

$$t_i = T_1. \quad (11.4.8)$$

Итак, если $T_1 < \bar{t}_i \leq T_0 + T_1$, деталь нужно хранить только в депо.

3. Если $S_{i0} = 0$, то

$$\text{erl}(S_{i0} | \alpha_{i0}) = 1, \quad (11.4.9)$$

и

$$t_i = (T_0 + T_1) \text{erl}(S_{i1} | \alpha''_i). \quad (11.4.10)$$

4. Из уравнения (11.4.4) получаем

$$\text{erl}(S_{i0} | \alpha_{i0}) = \frac{t_{si} - T_1 \text{erl}(S_{i1} | \alpha'_i)}{T_0 \text{erl}(S_{i1} | \alpha''_i) + T_1 \text{erl}(S_{i1} | \alpha''_i) - \text{erl}(S_{i1} | \alpha'_i)}. \quad (11.4.11)$$

Как мы видим из (11.4.11), должно выполняться

$$t_{si} \geq T_1 \text{erl}(S_{i1} | \alpha'_i). \quad (11.4.12)$$

Значит, при $\bar{t}_{si} < T_1$ должно быть назначено неотрицательное число деталей на периферийный склад. Если $\bar{t}_{si} \geq (T_0 + T_1) \text{erl}(S_{i1} | \alpha'_i)$, деталь

не должна храниться в депо. При невыполнении этого условия запас должен быть и на базе, и в депо.

Теперь можно сформулировать общий алгоритм:

Шаг 1. Ввод I, J, \bar{t} , плановых задержек P_0 и P_1 , T_0, T_1 , $\{c_i, \lambda_i\}$, $i = \overline{1, I}$.

Инициализация: $S_{ij} = 0$ для всех i, j .

Шаг 2. Распределить \bar{t} по номенклатурам одним из рассмотренных выше способов. Положить $i = 1$.

Шаг 3. Если $\bar{t}_i \geq T_0 + T_1$, перейти к шагу 6.

Шаг 4. Если $T_1 < \bar{t}_i \leq T_0 + T_1$, увеличивать число деталей в центральном складе, пока не будет выполнено $\text{erl}(S_{i0}|\alpha_{i0}) \leq (\bar{t}_i - T_1)/T_0$. Перейти к шагу 6.

Шаг 5. Если $\bar{t}_i \leq T_1$, увеличивать число S_{i1} деталей на базе до выполнения условия $\text{erl}(S_{i1}|\alpha'_i) \leq \bar{t}_{si}/T_1$. Если

$$\bar{t}_i \geq (T_0 + T_1) \text{erl}(S_{i1}|\alpha''_i),$$

перейти к шагу 6, иначе увеличивать запас в депо S_{i0} , пока не будет выполнено

$$\text{erl}(S_{i0}|\alpha_{i0}) \leq \frac{t_i - T_1 \text{erl}(S_{i1}|\alpha'_i)}{T_0 \text{erl}(S_{i1}|\alpha'_i) + T_1 \text{erl}(S_{i1}|\alpha''_i) - \text{erl}(S_{i1}|\alpha'_i)}.$$

Шаг 6. Если $i < I$, увеличить его на единицу и перейти к шагу 3.

Шаг 7. Закончить процедуру.

В *несимметричном* случае задержки распределяются по одному из обсуждавшихся принципов между базами, алгоритм последовательно применяется к каждой из них. Наконец, для депо назначаются максимальные по базам значения $\{S_{i0}\}$ для всех i .

Аналогичный подход можно применить и для большего числа уровней. Напомним, что он не гарантирует получения строго оптимальных решений.

Заключение

Сделанный выше обзор моделей и методов теории управления запасами, разумеется, не является исчерпывающим, но дает достаточно ясное представление о них и о смежных областях знания как из точных наук, так и из экономических. То же относится к списку литературы. Кстати, поиск в ней готового ответа для нестандартной ситуации вряд ли будет оправданным, и автор настоятельно призывает читателя к комбинированию предложенных подходов и к самостоятельным разработкам в этой увлекательной, содержательной в научном смысле, вполне доступной современному менеджеру и исключительно ценной практически науке. Содержание книги должно убедить читателя в четырех перечисленных тезисах и послужить базой для полезных аналогий на этапах экономического анализа, сбора и обработки исходных данных, построения содержательной и математической моделей, исследования и оптимизации последней и реализации результатов.

Отрезвление, наступившее в обществе после псевдолиберальных реформ, неизбежно приведет (и уже приводит) к активизации в экономике роли государства. В этом контексте задача управления запасами *снова* ставится более широко [27]:

- как рационально распределить запасы между всеми участниками производства в народном хозяйстве (термин, обозначающий то, во что оно преобразовалось, автору неизвестен);
- каковы необходимые размеры запасов в стоимостных и натуральных показателях в целом по народному хозяйству;
- каковы экономическая природа, состав, структура и уровень этих затрат;
- какой организационно-экономический механизм обеспечит практические возможности рационализировать структуру запасов, и т.д.

Такие вопросы соответствуют общегосударственному и региональному уровням рассмотрения проблемы управления запасами в экономических структурах соответствующего масштаба. В их решении основную роль играют балансовые соотношения. Однако всегда (независимо от их решения) возникает вопрос, когда и на какую величину пополнять запас у каждого конкретного участника процесса. Макро- и микроуровни рассмотрения проблемы управления запасами не следует противопоставлять друг другу: один другого не заменяет, а дополняет. В современных экономических условиях актуальность проблем второго уровня многократно возросла; именно они и рассматривались в этой книге.

Литература

- [1] *Абчук В.А., Матвейчук Ф.А., Томашевский Л.П.* Справочник по исследованию операций — М.: Воениздат МО, 1979. — 368 с.
- [2] *Белкин Н.В., Шахова Г.Л.* Математические методы и вычислительная техника в управлении запасами на предприятиях Франции. — М.: Госкомитет по снабжению, 1969 — 41 с
- [3] *Белоголовцев С.Д.* Оптимальное управление запасами при двух уровнях спроса со случайной доставкой заказа // Математические методы решения экономических задач. — М.: Наука, 1972. — С. 172-177.
- [4] *Беляев Ю.А.* Дефицит, рынок и управление запасами. — М.: Ун-т дружбы народов, 1991 — 228 с.
- [5] *Букан Дж., Кенигсберг Э.* Научное управление запасами /Пер. с англ. — М.: Наука, 1967.
- [6] *Вентцель Е.С.* Исследование операций. — М.: Сов Радио, 1972. — 551 с.
- [7] *Голдобина Н.Н.* Управление запасами средств производства: учеб. пособие — Л.: ЛФЭИ, 1991. — 71 с.
- [8] *Голенко Д.И. и др.* Моделирование в технико-экономических системах (управление запасами). — Л.: Изд-во Ленингр ун-та, 1975. — 197 с.
- [9] *Громенко В.М.* Применение методов теории управления запасами в экономических задачах — М.: Мос. ин-т управления, 1981. — 58 с.

- [10] *Грызанов Ю.П., Файницкий А.И.* Управление товарными запасами в торговле. — М., 1975.
- [11] *Двайт Г.Б.* Таблицы интегралов и другие математические формулы /Пер. с англ. — М.: Наука, 1964. — 228 с.
- [12] *Добрынин П.Ф.* Планирование и финансирование поставок оборудования. — М.: Финансы, 1974. — 128 с.
- [13] *Зеваков А.М.* Методические основы решения задач по управлению запасами. учеб. пособие. — Караганда, 1989. — 98 с.
- [14] *Зермати П.* Практика управления товарными запасами /Пер. с фр. — М.: Экономика, 1982. — 112 с.
- [15] *Ивашкин В.А., Волгин В.В., Миронов В.М.* Организация и техника торговли запасными частями. Опыт зарубежных фирм. — М., 1980.
- [16] *Ивашкин В.А., Волгин В.В., Миронов В.М.* Организация и техника торговли запасными частями. — М.: Экономика, 1984. — 97 с.
- [17] *Ивашкин В.А.* Организация технического обслуживания машин, оборудования и приборов: учеб. пособие. — М.: Высш. коммерч. шк., 1990.
- [18] *Инютина К.В.* Нормирование производственных запасов с применением математико-статистических методов. — М.: Статистика, 1969. — 112 с.
- [19] *Инютина К.В., Куровский В.Н.* Модели задач планирования производства и материально-технического снабжения в АСУП. М.: Статистика, 1975. — 134 с.
- [20] *Инютина К.В.* Повышение надежности и качества снабжения. — Л.: Изд-во Ленингр. ун-та, 1983. — 240 с.
- [21] *Инютина К.В.* Совершенствование планирования и организация материально-технического снабжения производственных объединений. — Л.: Машиностроение, 1986. — 246 с.

- [22] *Казанер Д., Моулер К., Нэш С.* Численные методы и математическое обеспечение /Пер. с англ. — М.: Мир, 1998. — 575 с.
- [23] *Кендалл М.Дж., Стьюарт А.* Теория распределений /Пер. с англ. — М.: Наука, 1966. — 587 с.
- [24] *Кудрявцев Б.М. и др.* Модели управления запасами. — М.: Ин-т управления им. С. Орджоникидзе, 1987. — 52 с.
- [25] *Кульбак Л.И.* Основы расчета обеспечения электронной аппаратуры запасными элементами. — М.: Сов. Радио, 1970. — 208 с.
- [26] *Лагуткин В.М.* Автоматизированные системы управления материально-техническим снабжением. — М.: Экономика, 1975. — 271 с.
- [27] *Лагуткин В.М., Соколов Р.Г.* Оптимизация запасов средств производства. — М.: Мысль, 1977. — 245 с.
- [28] *Лагуткин В.М., Соколов Р.Г.* Комплексное снабжение: проблемы, перспективы. — М.: Экономика, 1989. — 160 с.
- [29] *Ледин М.И.* Управление запасами (экономико-математические методы). — М.: Знание, 1978. — 64 с.
- [30] *Лотоцкий В.А., Мандель А.С.* Модели и методы управления запасами. — М.: Наука, 1991. — 188 с.
- [31] *Лукасин С.* Оптимальное планирование запасов в условиях стохастического спроса //Автоматика и телемеханика. 1975 №9.
- [32] *Меткалф М., Рид Дж.* Описание языка программирования Фортран-90 /Пер. с англ. — М.: Мир, 1995. — 302 с.
- [33] *Микитьянц С.Р.* Модели процессов материально-технического снабжения. — Л.: Изд-во Ленингр. ун-та, 1974. — 99 с.
- [34] *Микитьянц С.Р., Голдобина Н.Н.* Применение математических методов в управлении запасами. — Л.: ЛФЭИ, 1982. — 69 с.

- [35] *Михно М.К., Лобанова Л.С.* Моделирование систем снабжения. — Киев: Наукова думка, 1976 — 163 с.
- [36] Модели управления запасами: учеб пособие. — М.: Мос. ин-т управления, 1987 — 52 с.
- [37] *Никитин В.П.* Теория управления запасами: учеб. пособие. Рига: Латвийский гос ун-т, 1973. — 118 с.
- [38] Организация производства на промышленных предприятиях США /Пер. с англ. — М.: ИЛ, 1960–1961.
- [39] Организация и планирование материально-технического снабжения и сбыта в народном хозяйстве: учеб пособие для системы повышения квалификации — М.: Металлургия, 1977 — 383 с
- [40] *Первозванский А.А.* Математические методы в управлении производством. — М.: Наука, 1975. — 615 с.
- [41] *Первозванская Т.Н., Первозванский А.А.* Элементы теории управления запасами: учеб. пособие. — Л.: Изд-во Ленингр. ун-та, 1983. — 109 с
- [42] *Плоткин Б.К.* Экономико-математические методы и материально-техническое снабжение. сб. задач. — Л.: ЛФЭИ, 1984 — 73 с
- [43] *Плоткин Б.К.* Экономико-математические методы и модели в управлении материальными ресурсами: учеб пособие. — СПб. Ун-т экономики и финансов, 1992 — 63 с
- [44] Проблемы управления запасами /Докл. II симпозиума по управлению запасами. — М.: ЦЭМИ, 1972. — 292 с.
- [45] *Проценко О.Д., Рексин В.Д.* Управление материальными запасами — М., 1968. — 45 с.
- [46] *Проценко О.Д. и др.* Комплекс экономико-математических моделей управления складскими поставками /Докл. от СССР на IX Международном симпозиуме. 1975. 14 с.

- [47] *Проценко О.Д.* Организация планирования и управление материально-техническим снабжением. — М.: АНХ, 1982. — 86 с.
- [48] *Проценко О.Д.* Оперативное регулирование поставок продукции производственно-технического назначения. — М.: Экономика, 1985. — 175 с.
- [49] *Проценко О.Д. и др.* Перестройка материально-технического снабжения в условиях ускоренного развития народного хозяйства: учеб. пособие. — М.: АНХ, 1989. — 75 с.
- [50] *Рубальский Г.Б.* Управление запасами при случайном спросе (модели с непрерывным временем). — М.: Сов. Радио, 1977. — 160 с.
- [51] *Рубальский Г.Б.* Вероятностные и вычислительные методы оптимального управления запасами. — М.: Знание, 1987. — 115 с.
- [52] *Рыжиков Ю.И.* Многономенклатурная задача об управлении запасами. //Изв. АН СССР. Техн. кибернетика. 1966. №4. С. 20–31.
- [53] *Рыжиков Ю.И.* Планирование многономенклатурных запасов при вероятностном спросе //Изв. АН СССР. Техн. кибернетика, 1968. №2. С. 31–37.
- [54] *Рыжиков Ю.И.* О задаче управления запасами при неизвестном параметре распределения спроса //Кибернетика АН УССР. 1968. №4. С. 85–90.
- [55] *Рыжиков Ю.И.* Многономенклатурная задача о восстанавливаемом ЗИП //Изв. АН СССР. Техн. кибернетика. 1969. №6. С. 58–68.
- [56] *Рыжиков Ю.И.* Управление запасами. — М.: Наука, 1969. — 344 с.
- [57] *Рыжиков Ю.И.* О методах расчета оптимального набора запасных частей //Экономика и математические методы. 1971. т.VII. Вып.2. С. 207–212.

- [58] *Рыжиков Ю.И.* К расчету пороговых стратегий управления запасами //Кибернетика АН УССР. 1972 №4. С. 50–55.
- [59] *Рыжиков Ю.И.* Расчет восстанавливаемого набора запасных частей с учетом погрешности исходных данных //Экономика и математические методы. 1973. №5. С. 976–982.
- [60] *Рыжиков Ю.И.* Расчет системы массового обслуживания с порогом включения //Проблемы управления и теории информации. 1974. №1. С. 85–97.
- [61] *Рыжиков Ю.И., Смагин В.А.* Расчет распределения времени ремонта типового элемента замены на базе интегральных схем // Автоматика и вычислительная техника. 1974. №4. С. 31–33.
- [62] *Рыжиков Ю.И.* Расчет системы массового обслуживания с порогом включения и разогревом //Иzv. АН СССР, Техн. кибернетика. 1974. №6. С. 125–131.
- [63] *Рыжиков Ю.И.* К расчету многономенклатурного набора восстанавливаемых частей //Экономика и математические методы. 1976. №1. С. 166–170.
- [64] *Рыжиков Ю.И.* Комплекс программ для расчета систем массового обслуживания повышенной сложности //Программирование. 1978. №4. С. 87–91.
- [65] *Рыжиков Ю.И.* Об оптимальном оборотном запасе для замкнутых систем восстановления //Экономика и математические методы. 1979. №3. С. 616–620.
- [66] *Рыжиков Ю.И.* Машинные методы расчета систем массового обслуживания: учеб. пособие. — Л.: Воен. инж.-космический ин-т им.А.Ф.Можайского, 1979. — 177 с.
- [67] *Рыжиков Ю.И.* Алгоритм расчета многоканальной системы с эрланговским обслуживанием //Автоматика и телемеханика. 1980. №5. С. 30–37.
- [68] *Рыжиков Ю.И., Хомоненко А.Д.* Итеративный расчет многоканальных систем с произвольным распределением времени

обслуживания //Проблемы управления и теории информации 1980. №3. С. 32–38.

- [69] *Рыжиков Ю.И., Демиденко Ю.А.* Определение моментов распределения времени пребывания заявки в вычислительной сети //Автоматика и вычислительная техника. 1988. №1. С. 24–27.
- [70] *Рыжиков Ю.И., Хомоненко А.Д.* Расчет разомкнутых немарковских сетей с преобразованием потоков //Автоматика и вычислительная техника. 1989. №3. С. 15–24.
- [71] *Рыжиков Ю.И.* Имитационное моделирование систем массового обслуживания: учеб. пособие. — Л.: Воен. инж.-космический ин-т им.А.Ф.Можайского, 1991. — 111 с.
- [72] *Рыжиков Ю.И.* Три метода расчета временных характеристик разомкнутых систем массового обслуживания //Автоматика и телемеханика. 1993. №3. С. 127–133.
- [73] *Рыжиков Ю.И.* Пакет прикладных программ для анализа систем и сетей обслуживания //Тез. докл. 47-й НТК СПб Гос. ун-та телекоммуникаций. — СПб, 1994. — С. 80.
- [74] *Рыжиков Ю.И.* Руководство по расчету систем с очередями СПб: Воен. инж.-космический ин-т им.А.Ф.Можайского, 1995. 73 с.
- [75] *Рыжиков Ю.И.* Программирование на Фортране PowerStation для инженеров. — СПб.: КОРОНА принт, 1999. — 160 с.
- [76] *Рыжиков Ю.И.* Решение научно-технических задач на персональном компьютере. — СПб.: КОРОНА принт, 2000. — 272 с.
- [77] *Сакович В.А.* Модели управления запасами. — Минск: Наука и техника, 1986. — 319 с.
- [78] *Сакович В.А.* Управление комплектными поставками. — Минск: Вышэйшая школа, 1989. — 175 с.
- [79] *Саульев В.К., Лавренченко А.С.* Математическая теория оптимального управления запасами: конспект лекций. — М.: МАИ, 1974. — 93 с.

- [80] *Сизоненко П.З.* Методы оптимального управления запасами: учеб. пособие. — Одесса, 1979. — 61 с.
- [81] *Соколов Р.Г., Демьянов А.А.* Системный анализ и управление снабжением. — Л.: Стройиздат, 1975. — 216 с.
- [82] Технические основы эффективности ракетных систем. /Под ред. Е.Б. Волкова. — М.: Машиностроение, 1990. — 256 с.
- [83] *Тренина С.Л.* Обзор зарубежных экономико-математических методов управления запасами. — М.: Госкомитет по снабжению, 1968. — 58 с.
- [84] *Тюрин Ю.Н., Макаров А.А.* Анализ данных на компьютере. М.: ИНФРА-М, Финансы и статистика, 1995. — 384 с.
- [85] *Уайт О.У.* Управление производством и материальными запасами /Пер. с англ. — М.: Прогресс, 1978. — 304 с.
- [86] *Ушаков И.А.* Методы решения простейших задач оптимального резервирования. — М. Сов. Радио, 1969. — 176 с.
- [87] *Феклисов Г.И.* Математическое обеспечение систем управления запасами. — М.: Статистика, 1977. — 112 с.
- [88] *Хачатрян В.И.* Математические методы управления запасами: текст лекций. — М.: МЭСИ, 1983. — 56 с.
- [89] *Хедли Дж., Уайтин Г.* Анализ систем управления запасами /Пер. с англ. — М.: Наука-Физмат, 1969. — 512 с.
- [90] *Хемминг Р.В.* Численные методы для научных работников и инженеров /Пер. с англ. — М.: Наука, 1968. — 400 с.
- [91] *Хруцкий Е.А.* Оптимизация хозяйственных связей и материальных запасов. — М.: Экономика, 1977. — 263 с.
- [92] *Хруцкий Е.А.* Проблема эффективности принятия решений: на примере материально-технического снабжения. — М.: Наука, 1983. — 303 с.
- [93] *Хэнсменн Ф.* Применение математических методов в управлении производством и запасами /Пер. с англ. — М.: Прогресс, 1966.

- [94] *Чуев Ю.В.* Исследование операций в военном деле. — М.: Воениздат, 1970. — 256 с.
- [95] *Штундюк В.Д.* Управление запасами в новых условиях хозяйствования. — М.: Знание, 1990. — 62 с.
- [96] Экономика, разработка и использование программного обеспечения: учеб. для вузов — М.: Финансы и статистика, 1995. 286 с.
- [97] Экономико-математические методы в снабжении /Под ред. В.М. Лагуткина. — М.: Экономика, 1971. — 367 с.
- [98] *Юсупов Р.М., Заболотский В.П.* Научно-методологические основы информатизации. — СПб: Наука, 2000. — 455 с
- [99] *Abboud N.E.* The markovian two-echelon repairable items provisioning problems //J. of Opns. Res. Soc — 1996. — no 2. — P. 284-296
- [100] *Aggarwal S.C.* A review of current inventory theory and its applications // Int. J. Product Res. — 1974. — P. 443-482.
- [101] *Aggarwal V.* A closed-form approach to multi-item inventory grouping //Naval Research Logistics Quaterly. — 1983. — P. 471-485.
- [102] *Allen S.G.* Redistribution of total stock over several user locations //Naval Research Logistics Quaterly. — 1958. — P. 337-345.
- [103] *Allen S.G., d'Esopo D.A.* An ordering policy for repairable stock item //Operations Research. — 1968. — no. 3. — P. 669-674.
- [104] *Andress F.M., Emmons H.* On the optimal packaging frequency of products jointly replenished //Management Science — 1976. — P 1165-1166.
- [105] *Anly S.* Multi-item replenishment and storage problem! heuristic and bounds //Operations Research. — 1991. — no. 2. — P. 233-243.
- [106] *Anly S., Federgruen A.* Capacitated two-stage multi-item production-inventory model with joint setup costs //Operations Research. — 1991. — no. 3. — P 443-455.

- [107] *Baker K.R.* On the relation of the stockout probability to the optimum reorder mechanism in complex production-inventory systems // *Operations Research*. — 1970. — no. 3. — P. 335–340.
- [108] *Baskett F., Chandy K.M., Muntz R.R., Palacios J.G.* Open, closed, and mixed networks of queuing with different classes of customers // *J of the ACM*. — 1975. — v.22, no.2. — P. 248–260.
- [109] *Bastian M.* Joint replenishment in multi-item inventory systems // *J. of the Operat. Res. Soc.* — 1986. — no. 12. — P. 1113–1120.
- [110] *Bemeimans R.* The capacity aspect of inventories // *Lecture notes in economics and mathematical systems*. — 1986. — v. 267. — 165 pp.
- [111] *Boness A., Schwartz A.* A cost-benefit analysis of military aircraft replacement policies // *Naval Research Logistics Quarterly*. — 1969. — no. 2 — P. 237–258.
- [112] *Bradford J.W., Sugrue P.K.* Inventory rotation policies for slow moving parts // *Naval Research Logistics*. — 1991. — no. 1. — P. 87–106.
- [113] *Bregman R.L., Ritzman L.P., Krajewski L.J.* A heuristic for the control of inventory in a multi-echelon environment with transportation costs and capacity limitations // *J. of the Operat. Res. Soc.* — 1990. — no. 9. — P. 809–820.
- [114] *Brenner J.L.* Stock control in a many depot system // *Naval Research Logistics Quarterly* — 1969. — no. 3 — P. 359–380.
- [115] *Brumelle S.L.* A generalization of $L = \lambda W$ to moments of queue length and waiting times // *Operations Research*. — 1972. — v. 20, no.6. — P. 1127–1136.
- [116] *Burgin T.A.* Time to sell a fixed quantity of stock depleted by distributed demand // *Operat. Res. Quarterly*. — 1969. — no. 4. — P. 421–428.
- [117] *Burgin T.A.* The Gamma distribution and inventory control // *Operat. Res. Quarterly*. — 1975. — no. 3. — P. 507–525.

- [118] *Buzacott J.A.* Economic order quantities with inflation // *Operat Res. Quarterly* — 1975 — no. 3. — P. 553–558.
- [119] *Caie J.P., Maxwell W.L.* Hierarchical machine load planning // *In [200]*, P. 111–126.
- [120] *Cantley M.F.* A note on the efficient solution of the reorder quantity equation // *Operat. Res. Quarterly*. — 1975 — no. 2 P. 445–447.
- [121] *Chakravarty A.K.* Multi-item inventory aggregation into groups // *J. of the Operat. Res. Soc.* — 1981. P. 19–26.
- [122] *Chakravarty A.K.* An optimal heuristic for coordinated multi-item inventory replenishment // *J. of the Operat. Res. Soc.* — 1985. P 1027–1039.
- [123] *Chen J.F., Wilhelm W.E.* Optimizing the allocation of components to kits in small-lot, multiechelon assembly systems // *Naval Research Logistics* — 1994 — no. 2. — P 229–256
- [124] *Clark A.J.* Experiences with multi-indentured, multi-echelon inventory model // *In [200]*, P. 299–329
- [125] *Cohen M.A., Kleindorfer P.R. et al.* Multi-item service constrained (s,S) policies for spare parts logistics systems // *Naval Research Logistics* — 1992. — no. 4. — P 561–578
- [126] *Colomb P.* Gestion des stocks et des pièces de recharge. — Paris: Eyrolles, 1978
- [127] *Croston J.D.* Forecasting and stock control for intermittent demands // *Operat. Res. Quarterly* — 1972. — no. 3. — P. 289–304.
- [128] *Dave U.* On reducing excessive stock in the order-level lot-size system for decaying inventories // *In [193]*. — P. 175–180.
- [129] *Demmy W.S., Pressuti V.J.* Multi-echelon inventory theory in the AFLC // *Там же*, P. 279–298.
- [130] *Deuermeyer B.L., Schwartz L.B.* A model for the analysis of system service level in warehouse-retailer distribution systems: the identical retailer case // *In [200]*. — P. 163–193.

- [131] *Donaldson W.A.* The allocation of inventory items to lot size/reorder level (Q,r) and periodic review (T,Z) control systems //Operat. Res. Quarterly. — 1974. — no 3. — P 481-486.
- [132] *Drezner, Gilnam, Pasternac B.A.* An EOQ model with substitutions between products //J. of Operat. Res Soc — 1995. — no 7 — P. 887-891.
- [133] *Ehrhardt R.A., Schultz C.R., Wagner H.M.* (s,S) policies for a wholesale inventory system //In [200]. — P 145-161.
- [134] *Eijs van, M.J.G.* A note on the joint inventory replenishment problem under constant demand //J. of Operat. Res. Soc. — 1993. — no. 2. — P 185-193
- [135] *Eppen G., Schrage L.* Centralized ordering policies in a multi-warehouse system with lead times and random demand // In [200]. — P 51-67.
- [136] *Erenguk S.S., Mercan H.M.* A multifamily dynamic lot-sizing model with coordinated replenishment //Naval Research Logistics Quarterly — 1990. — no. 4. — P. 539-558.
- [137] *Evans R.D.* Geometric distribution in some two dimensional queuing systems //Operations Research. — 1967. — v.15, no 5. — P. 830-846.
- [138] *Falkner C.H.* Jointly optimal inventory and maintenance policies for stochastically failed equipment //Operations Research. — 1968. — no. 3 — P 587-601.
- [139] *Falkner C.H.* Optimal spares for stochastically failing equipment //Naval Research Logistics Quarterly. — 1969 — no. 3 — P. 287-297.
- [140] *Feeney G.J., Sherbrooke C.C.* The (s-1,s) inventory policy under compound Poisson demand // Management Science. — 1966 — P. 391-411
- [141] *Fisher W.W.* Issues and models in maintenance systems incorporating cannibalization: a review //INFOR. — 1990. — no. 1 — P. 67-88.
- [142] *Fox B.L., Landi D.M.* Searching for the multiplier in one constraint optimization problem // Operations Research. — 1970. — P. 253-262.
- [143] *Frieten R.* Integrating key suppliers — essential part of a just-to-time concept //Progress in Inventory Research. — 1989. — P. 185-190.
- [144] *Ghosal A.* Some aspects of queuing and storage systems. — Berlin: Springer, 1970. — 93 pp.
- [145] *Goyal A.K.* Stochastic behaviour of a complex system with bulk failure and priority repairs //Revue Française d'Informatique et de Recherche Opérationnelle. — 1971. — V-1. — P. 51-56.
- [146] *Goyal S.K.* Determination of economic packaging frequency for items jointly replenished //Management Science. — 1973. — P. 232-235.
- [147] *Goyal S.K.* Determination of optimum packaging frequency of items jointly replenished //Management Science. — 1974. — P. 436-443.
- [148] *Goyal S.K.* Economic packaging frequency of jointly replenished items having interdependent packaging set-up cost under constant demand //Operational Research Quarterly. — 1976. — no. 1. — P. 111-118.
- [149] *Goyal S.K.* A simple method of determining nearly optimum order quantities for a multi-product single-supplier problem //The Economics and Management of Inventories. — Budapest: Akad. Kiadó, 1981. — P. 365-372.
- [150] *Goyal S.K.* Determining the optimal production-packaging policy for jointly replenished items //Progress in Inventory Research. — 1989. — P. 339-342.
- [151] *Goyal S.K., Satir A.T.* Joint replenishment inventory control: deterministic and stochastic models //European Journal of Operational Research. — 1989. — P. 2-13.
- [152] *Graves S.C.* Multi-stage lot-sizing: an iterative procedure //In [200], P. 95-110.

- [153] *Gross D., Pinkus C.E., Soland R.M.* Designing a multi-product, multi-echelon inventory system //Operations Research. — 1971. — P. 11–50.
- [154] *Gross D., Harris C.M.* On one-for-one ordering inventory policies with state-dependent leadtimes //Operations Research. — 1971. — P. 735–760.
- [155] *Gross D., Pinkus C.E., Soland R.M.* Designing a multi-product, multi-echelon inventory system //In [200]. — P. 11–50.
- [156] *Gupta A.* Approximate solution of a single-base multi-indentured repairable item inventory system //J. of Operat. Res. Soc. — 1992. — P. 701–710.
- [157] *Gupta A.* The (Q,r) inventory system with an reliable supplier //INFOR. — 1996. — no. 2. — P. 59–76.
- [158] *Hariga M.* Two new heuristic procedures for the joint replenishment algorithm //J. of Operat. Res. Soc. — 1994. — no. 4. — P. 463–471.
- [159] *Henery R.J.* Inventory replenishment policy for cyclic demand //J. of Operat. Res. Soc. — 1990. — no. 7. — P. 639–643.
- [160] *Hochberg M.* Generalized multicomponent systems under cannibalization //Naval Research Logistics Quarterly. — 1973. no. 4. — P. 589–606.
- [161] *Inventory models* /Ed. by A. Chican. — Akadémiai kiadó, Budapest, 1990. — 419 pp.
- [162] *Johnson J.W.* On stock selection at spare parts stores sections //Naval Research Logistics Quarterly. — 1962. — no. 1. — P. 49–60.
- [163] *Joneja D.* The joint replenishment problem: new heuristics and worst case performance bounds //Operations Research — 1990. — no. 4. — P. 619–623
- [164] *Kalmon B.* A decomposition algorithm for arborescence inventory systems //J. of Operat. Res. Soc. — 1972. — no. 4. — P. 860–874.
- [165] *Karr H.W.* A method of estimating spare part essentiality //Naval Research Logistics Quarterly. — 1958. — no. 1.
- [166] *Kaspi M.* Determination of economic packaging set-up frequency of jointly replenished items //J. of Operat. Res. Soc. — 1991. — no. 2. — P. 165–168.
- [167] *Kelle P.* Optimal service levels in multi-item inventory systems //Progress in Inventory Research. — 1989. — P. 375–380.
- [168] *Kok de, A.G.* Production — inventory control models, approximations, and algorithms — Amsterdam, 1987. — 214 pp.
- [169] *Kono H., Nakamura Z.* The balanced lot size for a single-machine multi-product lot shedding problem //J. of Operat. Res. Soc. of Japan. — 1990. — no. 2. — P. 119–138.
- [170] *Kumar S., Arora S.* Optimal ordering policy for a multi-item, single-supplier system with constant demand rates //J. of Operat. Res. Soc. — 1990. — no. 4. — P. 345–349.
- [171] *Love R.F.* A two-station stochastic inventory model with exact methods of computing optimal policies //Naval Research Logistics Quarterly. — 1967. — P. 185–217.
- [172] *Mabin V.J.* A practical near-optimal order quantity method //Progress in Inventory Research. — 1989. — P. 381–386.
- [173] *Malec H.A., Steinhorn D.* A new technique for depot and sub-depot spares // IEEE Trans. Reliability, v. R-29, 1980, P. 381–386.
- [174] *Mandal B.N., Phaujdar S.* An inventory model for deteriorating items and stock-dependent consumption rate //J. of Operat. Res. Soc. — 1989. — no. 5. — P. 483–488.
- [175] *Mariani L., Nicoletti B.* Optimization of deterministic multiproduct inventory model with joint replenishment //Management Science. — 1973. — no. 3. — P. 349–362.
- [176] *Muller B.L.* A multi-item inventory model with joint backorder criterion //Operations Research. — 1971. — P. 1467–1476.

- [177] *Muller B.L.* Dispatching from depot repair in a recoverable item inventory system: on the optimality of a heuristic rule // *Management Science*. — 1974. — P. 316-325.
- [178] *Muller B.L., Modares-Yazdi M.* The distribution of recoverable inventory items from a repair center when the number of consumption centers is large // *Naval Research Logistics Quarterly*. — 1978. — P. 597-604
- [179] *Muckstadt J.A.* A model for a multi-item, multi-echelon, multi-indenture inventory system // *Management Science*. — 1973. — P. 472-481.
- [180] *Muckstadt J.A.* Some approximations in multi-item, multi-echelon inventory systems for recoverable items // *Naval Research Logistics Quarterly*. — 1978. — P. 377-394.
- [181] *Muckstadt J.A.* A three-echelon, multi-item model for recoverable items // *Naval Research Logistics Quarterly*. — 1979. — P. 199-221.
- [182] *Nahmias S.* Myopic approximations for the perishable inventory problems // *Management Science*, v. 22, 1976. — No.9. — P. 1002-1008.
- [183] *Nahmias S.* Inventory models. // In *Encyclopedia of Computer Science and Technology*. — N.Y.: M.Dekker, 1978.
- [184] *Nahmias S.* Managing repairable item inventory system: a review // In [200]. — P. 253-277.
- [185] *Nahmias S., Smith S.A., Mamer J.* A two echelon inventory system with emergency orders // 29-th Internat. Symp. on Inventory. — Budapest, 1982.
- [186] *Okyano L.Y., Yeh M.C., Wu K.* Mixture inventory model with backorders and lost sales for variable lead time // *J. of Operat. Res. Soc.* — 1996. — no. 4. — P. 829-832.
- [187] *Park K.S., Kim D.H.* Congruential inventory model for two-echelon distribution system // *J. of Operat. Res. Soc.* — 1987. — no. 7. — P. 643-650.

- [188] *Pasternack B.A., Diezner Z.* Optimal inventory policies for substitutable commodities with stochastic demand // *Naval Research Logistics*. — 1991. — no. 2. — P. 221-240
- [189] *Petersen J.W., Gusler M.A.* The costs of alternative air base stocking and requisitioning policies // *Naval Research Logistics Quarterly*. — 1955 — no. 1-2. — P. 69-82
- [190] *Petersen R., Silver E.A.* Decision systems for inventory management and production planning. — N.Y.: Wiley, 1979.
- [191] *Petrović R., Šenborn A., Vujošević M.* Hierarchical spare parts inventory systems. — Amsterdam etc: Elsevier, 1986. — 296 pp
- [192] *Porteus E., Lansdowne Z.* Optimal design of multi-item, multi-location, multi-repair type repair and supply system // *Naval Research Logistics Quarterly*. — 1974. — P. 213-238.
- [193] Progress in inventory research. // *Proc. of the 4-th Internat. Symp Budapest: Akad. Kiadó, 1989. — 446 pp.*
- [194] *Raafat F.* Survey of literature on continuously deteriorating inventory models // *J. of Operat. Res. Soc.* — 1991. — no. 1. — P. 27-38.
- [195] *Robinson L.W.* Optimal and approximate policies in multiperiod multilocation inventory models with transshipment // *Operations Research*. — 1990. — no. 2. — P. 278-295.
- [196] *Rosenbaum B.A.* Service level relationship in a multi-echelon inventory system // *Management Science*. — 1981. — P. 926-945.
- [197] *Rutenberg D.P.* Design commonality to reduce multi-item inventory: optimal depth of a product line // *Operations Research*. — 1971. — P. 491-509.
- [198] *Sand G.* Predicting demand on the secondary echelon: a case study // In [200] — P. 209-223.
- [199] *Sarker, Farga* An optimal batch size for a production system operating under a fixed-quantity periodic delivery policy // *J. of Operat. Res. Soc.* — 1991. — no. 8. — P. 891-900.

- [200] *Schwartz L.B. (ed.) Multi-level production/inventory control systems: theory and practice. //Studies in the Management Science, v. 16. — North-Holland, 1981.*
- [201] *Sherbrooke C.C. METRIC: a multi-echelon technique for recoverable item control //Operations Research. — 1968. — P. 122–141.*
- [202] *Sherbrooke C.C. An evaluator for the number of operationally ready aircrafts in a multi-level supply system // Operations Research. — 1972. — P. 261–280.*
- [203] *Silver E.A. A simple method of determining order quantities in joint replenishments under deterministic demand // Management Science. — 1976. — P. 1351–1361.*
- [204] *Silver E.A. Operations research in inventory management: a review and critique // Operations Research. — 1981. — P. 628–645.*
- [205] *Simon R.M. Stationary properties of a two-echelon inventory for low demand items // Operations Research. — 1971. — P. 761–773.*
- [206] *Song J.-S., Zipkin P.H. Evaluation of base-stock policies in multiechelon inventory systems with state-dependent demands. Part 1: state-dependent policies //Naval Research Logistics. — 1992. — no. 5. — P. 715–728.*
- [207] *Srinivasan V.S. First emptyness in the spare parts problem for repairable components //Operations Research. — 1968. — no. 2. — P. 407–415.*
- [208] *Sworder D.D., Kazangy T. Optimal control repair and inventory strategies for a linear stochastic systems //IEEE Trans. Syst., Man and Cybern. — 1972. — no. 2. — P. 342–347.*
- [209] *Takahashi Y., Takami Y. A numerical method for the steady-state probabilities of a GI/G/c queuing system in a general class //J. of the Operat. res. soc. of Japan. — 1976. — v.19, no.2. — P. 147–157.*

- [210] *Tan F K. Optimal policies for a multi-echelon inventory problem with periodic ordering //Management Science — 1974 — no 7 — P 1104–1112*
- [211] *Tanaka T, Phan Thien Thach, Suzuki S. Two nonconvex procedures for the problem of determining an economic ordering policy for jointly replenished items //J of the Operat Res Soc of Japan — 1991 — no 2 — P 109–124*
- [212] *Thiuz J M. Analysis of Lagrangian decomposition for the multi item capacitated lot-sizing problem //INFOR — 1991 — no 4 — P 271–283*
- [213] *Tsao Che S A Bayesian approach to estimating decision parameters in a replacement inventory system //Operational Research Quarterly — 1971 — no 4 — P 363–374*
- [214] *Van Beek P. An application of the logistic density on a stochastic continuous review stock control model //Zeitschrift für Operations Research — 1978 — P 165–173*
- [215] *Van Beek P. Modelling and analysis of multi-echelon inventory system //European J of Operations Research — 1981 — P 175–183*
- [216] *Wright G.P. Optimal ordering policies for inventory systems with emergency ordering //Operational Research Quarterly — 1969 — no 1 — P 111–124*
- [217] *Zheng Y S, Zipkin P. A queuing model to analyse the value of centralized inventory information //Operations Research — 1990 — no 2 — P 296–307*

Приложения

A

Расчет системы кратных периодов

```
program multidet
  integer          i,j,kk,key,n
  real             gg,ll,lhk,tt,r
  real, allocatable :: h(:),lam(:),g(:),glh(:)
  real, allocatable :: hs(:),ls(:),gs(:),left(:),right(:)
  integer, allocatable :: k(:),y(:)

  n=19
  allocate (h(n), g(n), gs(0:n), lam(n), glh(n))
  allocate (hs(n), ls(n), k(n), y(n), left(n), right(n))

  gs(0)=1.33
  open (3, file='multidet.dat')
  do i=1,n
    read (3,'(3e14.5)') g(i),h(i),lam(i)
  end do
  do i=1,n; glh(i)=g(i)/(lam(i)*h(i)); end do
  call chanlin(n,glh,y) !! Массив glh упорядочен по возрастанию
                       !! В массиве y --- соответствующие номера

  do i=1,n
    j=y(i); hs(i)=h(j); ls(i)=lam(j); gs(i)=g(j);
```

```

end do

!! Независимая оптимизация
ll=0
do i=1,n
  ll=ll+sqrt(2*(gs(0)+gs(i))*ls(i)*hs(i))
end do
print *, ' Независимая оптимизация, ll = ',ll

print *, ' Общий период'
gg=gs(0); lhk=0
do i=1,n
  gg=gg+gs(i); lhk=lhk+ls(i)*hs(i)
end do
gg=2*gg
tt=sqrt(gg/lhk); ll=sqrt(gg*lhk)
write (*,'(a,e12.5)') ' Минимальные затраты = ', ll
write (*,'(a,e12.5)') ' Базисный период      = ', tt

!! Начальное приближение
tt=sqrt(2*(gs(0)+gs(1))/(ls(1)*hs(1)))

!! Начальная группировка
k(1)=1; kk=1
r=2*gs(1)/(kk*(kk+1)*tt**2)
do i=2,n
  do while (ls(i)*hs(i)<=r)
    kk=kk+1; r=2*gs(i)/(kk*(kk+1)*tt**2)
  end do
  k(i)=kk
end do

gg=gs(0); lhk=0
do i=1,n
  gg=gg+gs(i)/k(i); lhk=lhk+ls(i)*hs(i)*k(i)
end do
gg=2*gg
tt=sqrt(gg/lhk); ll=sqrt(gg*lhk)

```

```

!! Правые границы интервала постоянства группировок
do i=1,n
  if (k(i)==1) then
    right(i)=1e20
  else
    right(i)=sqrt(2*gs(i)/(ls(i)*hs(i)*(k(i)*(k(i)-1))))
  end if
end do

!! Левые границы интервала постоянства группировок
do i=1,n
  left(i)=sqrt(2*gs(i)/(ls(i)*hs(i)*(k(i)*(k(i)+1))))
end do

print *, ' Начальное состояние'
write (*,'(a,e12.5)') ' Минимальные затраты = ', ll
write (*,'(a,e12.5)') ' Базисный период      = ', tt
print *, ' inew iold g/(h*lam) k left right'
do i=1,n
  j=y(1);
  write(*,'(2i6,e12.3,i6,2e12.3)') &
    i,j,glh(i),k(i),left(i),right(i)
end do

!! Выбираем направление движения
key=0
if (any(right<tt)) then
  key=1
else
  if (any(left>tt)) key=-1
end if

if (key==1) then
  print *, 'Пошли вправо'
  do while (key==1)
    key=0
    do i=1,n
      if (right(i)<=tt) then
        key=1

```

```

        gg=gg+2*gs(i)/(k(i)*(k(i)-1))
        lhk=lhs-ls(i)*hs(i)
        k(i)=k(i)-1
        right(i)=sqrt(2*gs(i)/ &
            (ls(i)*hs(i)*k(i)*(k(i)-1)))
    end if
end do
tt=sqrt(gg/lhk); ll=sqrt(gg*lhk)
print *, ' tt = ',tt, ' ll = ',ll
end do
else
print *, 'Пошли влево'
do while (key== -1)
    key=0
    do i=1,n
        if (left(i)>=tt) then
            key=-1
            gg=gg-2*gs(i)/(k(i)*(k(i)+1))
            lhk=lhs+ls(i)*hs(i)
            k(i)=k(i)+1
            left(i)=sqrt(2*gs(i)/ &
                (ls(i)*hs(i)*k(i)*(k(i)+1)))
        end if
    end do
    tt=sqrt(gg/lhk); ll=sqrt(gg*lhk)
    print *, ' tt = ',tt, ' ll = ',ll
end do
end if

write (*,'(a,e12.5)') ' Минимальные затраты = ', ll
write (*,'(a,e12.5)') ' Базисный период      = ', tt
kk=0
do i=1,n
    if (k(i)/=kk) then
        kk=k(i)
        write (*,'(//,a,i3)') 'Коэффициент кратности = ',k(i)
    end if
    write (*,'(i6\)'') y(i)
end do

```

```

deallocate (h,hs,g,gs,lam,ls,glh,left,right,k,y)
contains
    subroutine chanlin(n,x,y)
!!      Элементы x будут упорядочены по возрастанию
!!      В массиве y - соответствующие номера старым

    integer, intent(in)      :: n
    real,    intent(inout)  :: x(:)
    integer, intent(out)    :: y(:)
    integer                                i,j,k,l
    real                                    p

        do i=1,n; y(i)=i; end do
        do i=1,n
            p=x(i); k=i; l=y(i)
            do j=i+1,n
                if (x(j)<p) then
                    p=x(j); k=j; l=y(k)
                end if
            end do
            x(k)=x(i); x(i)=p;
            y(k)=y(i); y(i)=l
        end do
    end subroutine chanlin
end program

```

В строке с заголовками граф результирующей таблицы для размещения ее по ширине страницы часть пробелов опущена

Б

Расчет зонной стратегии

```
program zone
  integer                i,n
  real                   d, hh,gg,ll,tau
  real                   eps, qq, qq2, sss,ss1,ss2,ss3
  integer, allocatable  :: s1(:), s2(:), ss(:)
  integer, allocatable  :: h(:), lam(:), g(:)

  n=10
  allocate (h(10), g(0:10), lam(10), s1(10), s2(10), ss(10))

  data d, eps, tau /500, 1e-5, 0.2/
  open (3, file='zone.dat')
  rewind (3)
  read (3,'(11i5)') g
  read (3,'(10i5)') h
  read (3,'(10i5)') lam
  do i=1,n
    gg=g(i); hh=h(i); ll=lam(i)
    qq2=0; qq=1e0
! Вычисляем q''
    do while(abs(qq-qq2)>eps)
      qq=qq2
      qq2=ll*gg/hh+(qq+ll*tau)*(1e0+ll*tau)
      qq2=sqrt(2*qq2)-ll*tau
```

```
    end do
    qq2=floor(qq2+0.5) !! Ближайшее целое
!! Вычисляем второй порог
    gg=g(0)+g(i)
    ss2=log(d*ll/(hh*(qq2+ll*tau)))/log(1+1.0/(ll*tau))-1.0
    ss2=ceiling(ss2) !! Ближайшее большее целое
!! Верхний уровень
    sss=ss2+qq2

    ss3=ss2; ss1=0
!! Вычисляем первый порог
    do while(abs(ss1-ss3)>eps)
      ss1=ss3
      ss3=log(d*ll*tau*((sss-ss1)*log(1+1.0/(ll*tau))-1.0)/ &
              (hh/(2*ll)*(sss-ss1)**2+gg))/ &
              log(1+1.0/(ll*tau))
    end do
    ss1=ceiling(ss3) !! Ближайшее большее целое
    s1(i)=ss1; s2(i)=ss2; ss(i)=sss
  end do
  write (*, '(7a5)') 'i', 'lam', 'h', 'g', 's1', 's2', 'ss'
  do i=1,n
    write (*,10) i,lam(i), h(i), g(i), s1(i), s2(i), ss(i)
  end do
10 format (7i5)
  deallocate (h, g, lam, s1, s2, ss)
  close (3)
end program
```