



# 供应链优化

## 理论、方法及应用

柏庆国 徐健腾◎编著



123456

清华大学出版社  
北京

## 内 容 简 介

本书以培养学生的建模能力与分析能力为指导思想,注重引导学生理解和掌握本领域的重要理论与方法。全书从基本概念、基本模型、求解方法及算例分析等方面,系统介绍供应链优化的理论、方法与应用。在内容编排上,本书力求清晰阐明相关概念与方法的经济内涵,结合各类模型的构建过程及其在实际中的应用场景,做到理论与实践相统一。

作为一部兼具研究性与教学性的专业教材,本书既可作为高等院校管理科学与工程专业本科生与研究生的核心教材,也可供相关领域研究人员与从业人员学习与参考。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。举报:010-62782989, beiqinquan@tup.tsinghua.edu.cn。

### 图书在版编目(CIP)数据

供应链优化:理论、方法及应用 / 柏庆国,徐健腾编著. -- 北京:清华大学出版社, 2026. 6. -- ISBN 978-7-302-71787-4

I. F252.1

中国国家版本馆 CIP 数据核字第 2026S12Z87 号

责任编辑:陈凯仁

封面设计:傅瑞学

责任校对:王淑云

责任印制:刘海龙

出版发行:清华大学出版社

网 址: <https://www.tup.com.cn>, <https://www.wqxuetang.com>

地 址:北京清华大学学研大厦 A 座 邮 编:100084

社 总 机:010-83470000 邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质量反馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

印 装 者:三河市春园印刷有限公司

经 销:全国新华书店

开 本:185mm×260mm 印 张:15.5

字 数:374千字

版 次:2026年6月第1版

印 次:2026年6月第1次印刷

定 价:62.00元

---

产品编号:107077-01

# 前言

在当今全球化商业环境中,供应链已成为企业运营与发展的核心脉络。从原材料的源头采购,历经复杂的生产加工流程,到最终产品送达全球各地消费者手中,供应链涵盖从供应商到客户的全流程体系,其高效运作直接关乎企业的竞争力与盈利能力。与此同时,随着市场竞争的加剧、消费者需求的多样化以及新兴技术的不断涌现,供应链的复杂性与日俱增,如何对其进行科学管理与优化,成为企业界与学术界共同关注的焦点。

供应链建模与优化的重要性在当下尤为凸显。从企业层面来看,科学的建模能够将复杂的供应链系统转化为可量化、可分析的模型,帮助管理者精准把握各环节的内在联系与运行规律,从而摆脱经验决策的局限性。优化方法则能在成本、效率、服务等多重目标之间找到最佳平衡点,即通过合理规划库存水平,可降低资金占用与仓储成本;通过供应链协同优化,能够实现上下游企业的资源共享与产能匹配;通过协调生产与销售计划,可提升客户满意度与市场响应速度。在市场波动日益频繁的背景下,建模优化更是企业抵御风险的重要工具。

本书旨在系统阐述供应链建模与优化的理论及方法,为读者提供全面且深入的知识体系。通过对供应链各环节进行抽象建模,我们能够精准地描述其运作机制,剖析其中的关键问题;优化方法则帮助企业在诸多约束条件下,找到最优的决策方案,实现成本的有效控制、效率的显著提升以及服务质量的优化。在供应链建模部分,我们从基础的概念入手,逐步深入复杂的模型构建,涵盖结构模型,用于清晰展示供应链各节点间的组织关系与连接方式,帮助读者识别关键节点与风险传导路径;流程模型,聚焦业务流程的标准化与效率提升,助力企业发现流程中的瓶颈并进行优化;优化模型,以最优化理论与方法为核心,解决协调优化、库存管理、低碳运营、风险管理以及信息共享等实际决策难题;案例模型,通过各种算例,对不同情境下的供应链运行结果进行测试,帮助企业提前应对各种不确定性。

本书的编写汇聚了众多供应链领域专家学者的智慧与经验以及作者近年来在该领域的科研成果。本书共分8章。第1章是基础理论部分,主要介绍供应链的特征、概念以及建模优化的方法。第2章讨论供应链库存优化模型以及求解方法。第3章聚焦供应链协调策略以及优化方法。第4章重点探讨易变质产品供应链的建模与优化方法。第5章和第6章则从可持续的视角重点研究闭环供应链和低碳供应链的建模与优化方法。第7章主要介绍考虑企业决策者风险态度的供应链及模型构建。第8章重点描述信息不对称下的供应链建模及优化方法。各章节针对所构建的模型,均配套设计相应的算例,通过具体数据与计算过程的演示,帮助读者直观理解模型的应用逻辑与求解步骤。

本书编写过程中,借鉴了许多国内外的最新研究成果,相关文献已尽可能详尽地收录入参考文献,在此谨向这些为供应链领域研究做出贡献的学者们致以诚挚谢意。还要感谢刘

世豪、李福宇、姜嫚、苏迪、李姿璇、王可婷、谢羨芳、库颖、崔儒雨等团队成员，他们在资料收集、整理以及文稿的编排等方面做了大量的工作。此外，众多科研合作者包括烟台大学经济管理学院的马雪丽教授、苏州工学院商学院的陈玉玉副教授以及北京物资学院物流学院的祁琪副教授等对本书提供了诸多帮助与支持，在此一并感谢。

本书的编写得到了山东省泰山学者建设工程专项基金项目(项目编号：tsqn202103063、tsqn201812061)和山东省研究生教育教学改革研究项目(项目编号：SDYJSJGC2024072)的资助，对此表示衷心的感谢。

鉴于供应链建模与优化领域的复杂性与前沿性，本书对相关理论方法的探索仍显粗浅，加之作者学识所限，书中难免存在疏漏与不妥之处，恳请各位读者不吝赐教、批评指正。

作者

2025年11月



# 目录

<b>第 1 章 供应链管理基本概念</b> .....	1
1.1 供应链的结构和特征 .....	1
1.1.1 供应链的概念.....	1
1.1.2 供应链的基本结构.....	2
1.1.3 供应链的特征.....	4
1.2 供应链管理的概念 .....	5
1.3 供应链优化的主要方法 .....	8
1.3.1 最优化理论与方法.....	8
1.3.2 博弈理论 .....	16
1.4 思考题.....	27
<b>第 2 章 供应链库存模型构建与优化</b> .....	28
2.1 供应链库存的基本概念.....	28
2.1.1 库存的基本概念 .....	28
2.1.2 供应链库存管理策略 .....	29
2.2 库存的基本模型.....	32
2.2.1 随机需求库存问题的基本模型 .....	32
2.2.2 确定性均匀需求库存问题的基本模型 .....	37
2.2.3 确定性离散需求库存问题的基本模型 .....	46
2.3 易损耗产品的库存优化模型.....	48
2.3.1 生鲜品的报童模型 .....	48
2.3.2 易腐品的 EOQ 模型 .....	49
2.3.3 易腐品的动态批量模型 .....	53
2.4 思考题.....	55
<b>第 3 章 供应链协调模型构建与优化</b> .....	56
3.1 需求变异放大现象及产生的原因.....	56
3.2 缓解需求变异放大效应的方法.....	58
3.3 供应链协调优化方法.....	60
3.3.1 供应链契约的概念 .....	60

3.3.2	供应链契约参数 .....	60
3.3.3	常见的供应链契约 .....	63
3.4	思考题 .....	70
<b>第4章</b>	<b>易变质产品供应链模型构建与优化 .....</b>	<b>71</b>
4.1	易变质产品供应链的基本概念与意义 .....	71
4.1.1	基本概念 .....	71
4.1.2	发展易变质产品供应链的意义 .....	72
4.2	易变质产品供应链运营的设备与模式 .....	73
4.2.1	易变质产品供应链的主要设备与设施 .....	73
4.2.2	易变质产品供应链运营模式 .....	74
4.2.3	易变质产品供应链的范围及关键因素 .....	75
4.3	易变质产品供应链优化的基本模型 .....	79
4.3.1	有限计划期内易变质产品两货栈库存优化模型 .....	79
4.3.2	二级易变质产品供应链的联合订购策略 .....	87
4.4	易变质产品供应链的协调优化模型 .....	93
4.4.1	连续性需求下易变质产品的供应链运营模型 .....	93
4.4.2	随机需求下生鲜农产品供应链优化模型 .....	98
4.5	思考题 .....	104
<b>第5章</b>	<b>闭环供应链模型构建与优化 .....</b>	<b>105</b>
5.1	再制造的定义和特点 .....	105
5.1.1	定义 .....	105
5.1.2	过程 .....	106
5.1.3	特点 .....	107
5.1.4	意义 .....	108
5.2	闭环供应链 .....	108
5.2.1	闭环供应链概述 .....	108
5.2.2	闭环供应链回收模式 .....	109
5.2.3	闭环供应链管理 .....	115
5.3	再制造企业运营优化模型 .....	117
5.3.1	确定性需求下的再制造系统优化模型 .....	117
5.3.2	随机需求下的再制造优化模型 .....	124
5.4	闭环供应链优化模型 .....	135
5.4.1	集中回收的供应链 .....	136
5.4.2	制造商负责回收的闭环供应链 .....	137
5.4.3	零售商负责回收的闭环供应链 .....	138
5.4.4	第三方负责回收的闭环供应链 .....	139
5.5	思考题 .....	140

<b>第 6 章 低碳供应链模型构建与优化</b> .....	141
6.1 低碳减排的背景 .....	141
6.1.1 全球气候变暖现状.....	141
6.1.2 国际社会的关注与行动.....	142
6.1.3 传统供应链碳排放情况.....	142
6.1.4 消费者环保意识的觉醒.....	143
6.2 低碳减排的方式 .....	143
6.2.1 政府引领与激励.....	144
6.2.2 企业技术创新与绿色管理.....	144
6.2.3 社会公众参与.....	145
6.3 供应链的低碳减排 .....	146
6.3.1 采购环节.....	146
6.3.2 生产环节.....	146
6.3.3 运输环节.....	147
6.3.4 仓储环节.....	147
6.3.5 销售环节.....	147
6.4 低碳供应链的建模与优化 .....	148
6.4.1 碳政策下 EOQ 模型构建与优化 .....	148
6.4.2 碳政策下报童模型与优化.....	153
6.4.3 低碳减排下供应链运营与优化.....	156
6.5 思考题 .....	171
<b>第 7 章 考虑风险态度的供应链模型构建与优化</b> .....	172
7.1 供应链风险 .....	172
7.1.1 风险的概念.....	172
7.1.2 供应链风险的概念.....	173
7.1.3 供应链风险的特点.....	173
7.1.4 供应链风险管理.....	175
7.2 供应链风险分类与测量方法 .....	175
7.2.1 供应链风险类型.....	175
7.2.2 供应链风险的识别和评估.....	176
7.2.3 供应链风险的测量方法.....	177
7.3 考虑风险态度的供应链优化模型 .....	178
7.3.1 考虑风险态度的供应链企业优化模型.....	178
7.3.2 风险厌恶型的供应链运营优化模型.....	190
7.4 思考题 .....	210

第 8 章 信息不对称下的供应链模型构建与优化	211
8.1 不对称信息理论	211
8.1.1 不对称信息理论提出的背景	211
8.1.2 不对称信息理论的问题与作用	212
8.2 供应链管理中的信息不对称	213
8.2.1 供应链管理中信息的分类	213
8.2.2 供应链管理中信息的特点	214
8.2.3 信息不对称对供应链管理产生的影响	214
8.2.4 解决供应链管理中信息不对称的案例	215
8.3 信息不对称下的供应链优化模型	215
8.3.1 企业社会责任投入成本信息不对称下的供应链优化模型	215
8.3.2 营销成本信息不对称下的闭环供应链优化模型	219
8.3.3 营销成本信息不对称下的双渠道绿色供应链优化模型	224
8.3.4 需求信息不对称下的供应链优化模型	231
8.4 思考题	236
参考文献	237

# 第 1 章



## 供应链管理基本概念

### 1.1 供应链的结构和特征

交换过程是社会经济活动的基础。无论是在工业化社会还是非工业化社会,可供产品和需求产品之间存在的数量、类型、供应时间的差异和不平衡决定了产品交换的基础。任何一件产品都是从生产地点运到消费地点,以实现该产品的交换。当生产者和消费者之间发生许多交换时,那些将产品或服务带到市场上的企业所组成的序列就成为供应链(supply chain)、需求链或价值链(value chain)。

现代企业竞争理论的一个显著变化就是,企业竞争不再是简单地以自治实体的方式进行竞争,而是参与到供应链中,以供应链的方式进行竞争。供应链的概念经历了一个发展过程。早期的观点认为,供应链是制造企业中的一个内部过程,它是指把从企业外部采购的原材料和零部件,通过生产和销售活动,再传递到零售商和用户的一个过程。这一概念将供应链局限在企业的内部操作层上,注重企业自身的资源利用。

#### 1.1.1 供应链的概念

供应链源于价值链的概念。在“现代管理学之父”彼得·德鲁克(Peter F. Drucker)提出的经济链的基础上,1985年,美国学者迈克尔·波特(Michael Porter)在《竞争优势》一书中提出了价值链。最初,波特所指的价值链主要强调将单个企业内部价值活动联系为一个整体。1992年,约翰·尚克(John K. Shank)和维贾伊·戈文达拉扬(Vijay Govindarajan)将价值链的概念进行了扩充。他们认为价值链包含从原材料延伸到最终用户的一系列价值创造活动。1996年,詹姆斯·沃麦克(James Womack)和丹尼尔·琼斯(Daniel Jones)在《精益思想》一书中将价值链概念进一步拓展为价值流,将企业内部以及企业与供应商、用户之间的信息沟通形成的信息流包含在内。1996年,贝恩德·朔尔茨-瑞特(Bernd Scholz-Reiter)在整合了上述价值链和价值流思想的基础上,首次明确提出了供应链的定义:“运作实体的网络,通过这一网络组织将产品或服务传递到特定的用户市场,即由用户需求开始,贯通从产品设计到原材料供应、生产、批发、零售等过程,把产品送到最终用户的各项业务活动。”1998年,美国物流管理协会(Council of Logistics Management, CLM)将物流定义为供

应链活动的一部分,这成为物流管理向供应链管理发展的开端。

我国在 2021 年发布实施的《物流术语》(GB/T 18354—2021)国家标准中对供应链的定义是生产及流通过程中,围绕核心企业的核心产品或服务,由所涉及的原材料供应商、制造商、分销商、零售商直到最终用户等形成的网链结构。

近年来,供应链的概念更加强调围绕核心企业的网链关系,即核心企业与供应商、供应商的供应商的一切向前关系,与用户、用户的用户的一切向后的关系。《国务院办公厅关于积极推进供应链创新与应用的指导意见》(国办发〔2017〕84号)对供应链的定义为“以客户需求为导向,以提高质量和效率为目标,以整合资源为手段,实现产品设计、采购、生产、销售、服务等全过程高效协同的组织形态”。我国知名学者马士华在其编著的《供应链管理》教材中将供应链定义为“围绕核心企业,通过对工作流(work flow)、信息流(information flow)、物流(physical flow)、资金流(funds flow)的协调与控制,从采购原材料开始,制成半成品及最终产品,最后由销售网络把产品送到用户手中的将供应商、制造商、分销商、零售商,直至最终用户连成一个整体的功能网链结构”。其他一些学者(徐贤浩和马士华,2011;刘伟华等,2022;刘宝红,2023)也从不同角度对供应链和供应链管理进行了界定。

### 1.1.2 供应链的基本结构

苹果公司一直以高品控在业界著称。究其原因就是通过强大的产业链控制能力,对生产的各个环节都进行严密的监控和检查,保证消费者最终享受到最好的硬件和服务体验。目前,约半数 iPhone 是在郑州制造的。图 1-1 所示为苹果公司供应链的基本结构。接下来,将详细阐述 iPhone 从在郑州制造到出现在用户手中的“旅程”。

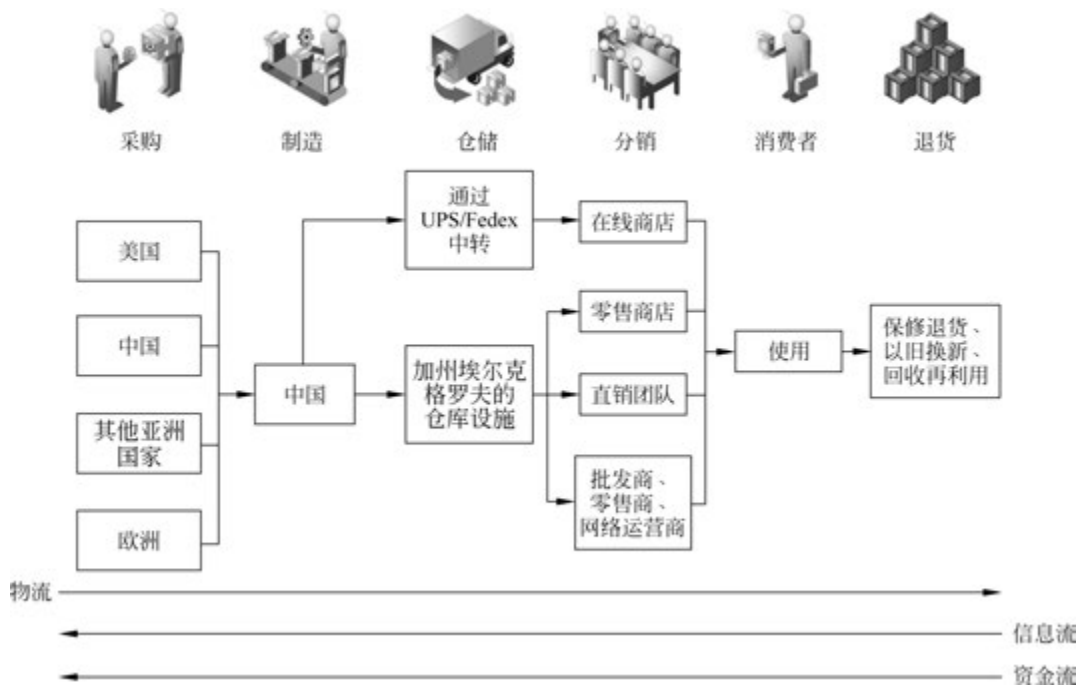


图 1-1 苹果公司供应链的基本结构图

## 1. 生产组装

每部 iPhone 有来自超过 200 多个供应商的零部件,包括内存芯片、调制解调器、相机模块等。苹果公司向全球供应商订购零部件,并整体供应给其组装工厂,如我国的富士康。当制造 iPhone 的零件都汇聚到富士康时,接下来便是组装了。例如,富士康在郑州的工厂占地 5.7km<sup>2</sup>,约有员工 35 万人,共有 94 条生产线,一天就能生产 50 万部 iPhone,大致相当于每分钟生产 350 部。这些 iPhone 通常大约经过 400 道工序,包括抛光、焊接、钻孔和上螺丝才可完成组装。从生产线下下来后,iPhone 将被放置在一个漂亮的白色纸板盒中,经过包装被放到木托盘中,然后用推车推出去,装载至等待装货的卡车上。

## 2. 海关报关

iPhone 需要通过海关,办理海关业务的地方就在富士康工厂的外面。刚组装好的 iPhone 被运送到距离工厂大门几百米的地方。在那里,修建了一处大型海关设施,海关位于一个保税区内,这使得苹果公司能更方便地向我国消费者销售 iPhone。

作为 iPhone 生产终点的我国,也是苹果公司全球税务战略的起点。在郑州,通常是在海关,富士康把 iPhone 出售给苹果公司,然后苹果公司把它们出售给其位于全球各地的分部。这一过程主要通过电子方式完成,使得苹果公司能把一部分利润划归爱尔兰分部,享受到税收优惠。这一系统在我国并非唯一的。完成海关程序后,此时的 iPhone 才算是真正“出生”了。但面向不同市场的 iPhone,会踏上不同的“销售”之路。

## 3. 出口再进口

对于出口到其他国家的 iPhone,会被运送到机场,然后通过波音 747 运输到其他国家的苹果零售店等待销售。

对于面向国内市场的 iPhone,海关工作人员会利用一个电子系统先将货物标记为“出口”,再重新标记为“进口”。一旦被宣布为进口产品,海关会在进口价格的基础上征收 17% 的增值税。随后,这些 iPhone 便获准运往国内各地。叠加汇率的波动等因素,从而导致我国市场上 iPhone 售价比美国高出约 20%。

在我国大陆地区销售的 iPhone 通常会被装载在大型卡车上,经过 18h 长途跋涉后到达上海,苹果公司在上海建有全国性配送中心。一辆大型卡车可以运载 3.6 万部 iPhone。由于每辆卡车运载的 iPhone 价值约 2 700 万美元(约合 1.9 亿元),它们通常配备有监控摄像头,有时甚至配备有保安。

iPhone 离开富士康郑州工厂后,平均需要两天就能到达 590 英里(950km)外的上海商店,而到达 6 300 英里(10 139km)外的旧金山商店,平均需要 3 天时间。

## 4. 销售和售后

即使 iPhone 到达了用户的手中,一次销售仍然没有完成。苹果公司还需要为用户手中的 iPhone 准备一系列的配套售后服务。

从上述案例可以看出,供应链的主要结构包括以下 5 部分。

(1) 供应商:向下游企业供应各种所需资源的企业。其供应的资源包括原材料、设备、

能源、商品和服务等。

(2) 制造商：负责产品制造的企业。其职能包括产品开发、生产和售后服务等。

(3) 分销商：在贸易中获得商品所有权的中间商。其承担购买商品所有权并转售的职能，需要承担风险，并且具有价格决定权。

(4) 零售商：将最终产品直接销售给终端用户的中间商。其承担组织商品、储存商品、承担风险和服务的职能，是分销渠道的最后环节。

(5) 用户：产品和服务的最终使用者。其作为整条供应链的唯一收入来源，是供应链的最后环节。

此外，供应链还包含3大流：物流、信息流和资金流。

(1) 物流。物流主要是物资(商品)的流通过程，既包括原材料、中间品和成品从供应商通过制造商、分销商、零售商等送到用户的过程，也包括用户的退货、维修等活动。因此，物流的方向是双向的。物流理论涉及如何在物资流通过程中用最短的时间以低成本对原材料、中间品和成品进行交付。

(2) 信息流。信息流是商品、交易等信息的传递过程。信息是在供应商与用户之间双向流动的。信息流是一种虚拟形态，在由供应商流向用户的过程中，包括生产能力信息、促销计划和交付时间表等；在由用户流向供应商的过程中，包括订单、销售情况、库存和质量信息等。

(3) 资金流。资金流是现金、所有权、发票等在供应链的循环过程，因此资金流的方向也是双向的。一方面，建立完善的经营体系，必须确保资金的及时回收，此时资金流以现金、信用卡支付等形式从用户通过零售商、分销商、制造商等流向供应商；另一方面，资金流还包括所有权和发票的转移，这种情况下资金流的方向是从供应商到用户的。

### 1.1.3 供应链的特征

从供应链的结构模型可以看出，供应链是一个网链结构，由核心企业、供应商、供应商的供应商、分销商、分销商的分销商以及用户组成。一个企业是一个节点，节点企业和节点企业之间是一种需求与供应的关系。一般来说，供应链主要有以下特征。

(1) 系统性。供应链本身是一个整体合作、协调一致的系统，它由多个合作者如链条般环环相扣，参与者为了一个共同的目标，协调运作，紧密配合。供应链是一个大的系统，每个供应链成员企业都是“链”中的一个环节。

(2) 复杂性。供应链是一个复杂的网络，这个网络是由具有不同冲突目标的成员和组织构成的。尤其当供应链是跨国、跨地区和跨行业的组合时，各国国情、制度、法律、文化、环境、习俗等方面存在差异，经济发达程度、物流基础设施以及管理水平、技术能力等也有很大的不同，而供应链的操作又必须保证其目的的准确性、行动的快速反应性和服务的高水准，因而导致供应链高度复杂的特点。同时，根据供应链节点企业组成的跨度(层次)不同，可以有生产型、加工型和服务型等，也可以有核心层和非核心层等，即供应链往往是由多个、多类型甚至多国企业构成的，因此，其结构远比单一企业复杂。

(3) 动态性。供应链系统会随时间发生变化。因为在供应链上，即使能够较准确地预测需求，计划过程也需要考虑在一段时间内季节波动、趋势、广告和促销、竞争者的策略等因素引起的需求和成本参数的变化。这些随时间而变化的需求和成本参数使确定最有效的供

应链变得很困难,这种动态性给管理者带来了挑战。同时,供应链中的企业都是在众多企业中筛选出来的合作伙伴,合作关系并非固定的,需要根据企业战略和市场需求的变化,实时动态地更新,这也使得供应链具有明显的动态性。

(4) 用户需求响应性。供应链的形成、存在、重构,都是基于最终用户需求而发生的,并且在供应链的运作过程中,用户的需求拉动是供应链中信息流、物流、资金流运作的驱动力。因此,准确、及时、有效地收集用户需求信息,并快速、动态、高质量地满足用户需求,是供应链管理的主要目标之一。

(5) 交叉性。供应链节点企业可以既是这个供应链的成员,同时又是另一个供应链的成员,众多的供应链形成交叉结构,增加了协调管理的难度。

由此可见,供应链是一个复杂的大系统。面对如此复杂的系统,必须认清不同情况下供应链系统的特征,这样才能有目的地选择适合本企业的运作模型,才能有针对性地选择最适宜的管理模式。

## 1.2 供应链管理的概念

供应链是一个有组织的网链结构,其运作效率和收益的最大化,必须通过供应链管理来完成。供应链管理可追溯到 20 世纪 80 年代,在全球化进程明显加快、科学技术快速迭代发展的背景下,市场竞争的主体已由企业之间转向供应链之间。下面介绍以 ZARA(飒拉)为代表的极速供应链及其管理模式。

ZARA 是 1975 年设立于西班牙、隶属 Inditex 集团旗下的一个子公司,既是服装品牌,也是专营 ZARA 品牌服装的连锁零售企业。ZARA 是全球排名第三、西班牙排名第一的服装商,在 87 个国家开设了超过 2 000 多家的服装连锁店。ZARA 的战略定位是为顾客提供“买得起的快速时装”。为了保证这一战略的实现,ZARA 采取了“快速、少量、多款”的商业模式,以实现快速设计、快速生产、快速出售、快速更新的市场目标。与此同时,ZARA 还构建了独具特色的供应链体系和管理流程,以确保市场目标有效完成。

ZARA 的供应链主要由 4 个关键环节组成,即产品开发、生产制造、物流配送以及专卖店直销。在 4 个环节中,“快速”这一战略核心自始至终都被清晰、明确地贯彻执行,最终形成了具有强大竞争优势的极速供应链系统,创造出令同行惊叹的速度:①每年提供 12 000 种不同的款式供顾客选择。②从设计理念到成品上架仅需要十几天。我国服装业该过程一般为 6~9 个月,国际名牌一般为 120 天,而 ZARA 最短只有 7 天,一般为 12 天。③所有专卖店能做到商品每周更新两次,每隔 3 周所有商品全部换新。④其是全球唯一一家能够在 15 天内将成衣配送到全球 850 多家门店的时装公司。

ZARA 的极速表现很大程度上要归功于其独特的供应链管理,即强大的供应链管理系统及其信息技术(information technology,IT)系统的支撑,这使其成为全球服装行业中响应速度与弹性管理的标杆企业,主要体现在 4 个方面。

### 1. 产品设计

ZARA 有近 400 名设计师,这些设计师是典型的“空中飞人”,他们经常穿梭于各种时装发布会或者出入各种时尚场所。通常,一些顶级品牌的最新设计刚摆上柜台,ZARA 就

会迅速发布和这些设计高度相似的时装。这样的设计方式能保证 ZARA 紧跟时尚潮流。据悉,在欧洲,ZARA 每年都需要投入数千万欧元的预算,用于处理与各大奢侈品牌之间的设计争议与法律纠纷。尽管如此,ZARA 依然坚持其快速跟进潮流的设计模式。从商业角度看,这预示了该模式所能带来的巨大收益,很可能远远超过其为解决法律问题所付出的成本。

ZARA 安排专职员工不间断地收集来自世界各地有关顾客需求及需求变化的信息,然后将这些信息汇总到总部的数据库。利用信息技术,将收集的信息进行有效处理。在此环节,所有的时尚信息都被清晰界定、分门别类后,存储于总部数据库的各个模块中。然后,这个数据库又与其原材料仓储数据库相连接。

这样,一方面,ZARA 的设计团队能以最快的速度,将时尚信息数据库中标准化的、归档保存的各种时尚信息调出来,然后进行快速、恰当的调整和组合,最后拟订新的设计方案,同时制定清晰的裁剪生产指令;另一方面,设计师还可以参考产品信息和原料库存信息,尽量设计现有原材料可以完成的服装款式,有效压缩小企业生产成本。通过数据的标准化,ZARA 的设计师可以相对轻松地掌握数以千计的布料、各种规格的装饰品、设计清单和库存商品信息,顺利完成任意一款服装的设计。可见,这种标准化的信息系统,是保证 ZARA 设计团队的工作迅速、有效地进行和每年推出大量不同的时尚设计款式的有力支撑。

## 2. 生产制造

ZARA 的运营经费几乎全部投在了工厂设备的扩充和改善上。位于西班牙加利西亚拉科鲁尼亚省的仓库,是一栋 4 层楼高、占地 691 万平方英尺(1 平方英尺 $\approx$ 0.0929 平方米)的超大型建筑物,其面积相当于 90 个足球场,而此座仓库连接着 14 座工厂,仓库内有机器人,可以 24h 随时待命进行压模、制布、染料,几乎不存在时间与人的浪费。因为 ZARA 知道,唯有自行生产制造,才能分秒不差地控制流程,并跟上时尚潮流的更新速度。

一旦设计团队确定某个方案可以进行生产,设计师就会通过计算机设计系统对颜色和材质进行优化。若需要在 ZARA 自己的工厂中生产,他们就会直接把各种规格传输给工厂中的剪裁等设备。在这里,被裁剪后的面料上就已经有了标准化的条形码,这种条形码会贯穿产品从生产、配送、运输至门店的全过程。这样,在整个生产直到销售的过程中,都使用着统一标准的条形码识别系统,从而保证了 ZARA 衣物在整个过程中能够流畅、快速地进行流通。世界各地连锁店的订单,经合理评估后传到工厂,将库存量降到最低。ZARA 的库存量为 15%~20%,比其他服饰连锁企业的 40% 低很多。

另外,在服装设计之余,参与设计的采购专家与市场专家就已经共同完成了该款服装的定价工作,这一价格当然也是参照数据库中类似产品在市场中的价格信息来确定的。然后,定好的价格就被换算成多国的货币额,并与服装的条形码一起印在商品标价牌上,并在生产之初就已经附着在服装上了。因此,新款服装在生产出来之后无须再定价和贴标签,运到世界各地的专卖店之后,就可以直接上架销售。

## 3. 物流配送

在配送环节中,ZARA 首先建立了非常先进的分销设施,在巨大的货物配送中心下面,长约 20km 的地下传送带将成衣从 ZARA 的工厂传送至此。为了确保每一笔订单准时到

达目的地,ZARA 没有采用浪费时间的人工分拣方法,而是采用了光学读取工具进行货品分拣。例如,成衣上的条形码是标准化的,因此,能以最快的效率、持续不断地读取条形码,这样,每小时能挑选并分拣超过 60 000 件成衣。其次,在运输环节,ZARA 利用各个门店与配送中心的信息互联,并采取类似“公共汽车式”的配送模式,通过优化配送路径,尽量缩短运输路径和空车状态的时长,将运输成本降至最低。

ZARA 公司使用企业资源计划(enterprise resource planning,ERP)系统对企业供、产、销各个环节实施有效管理与控制。它还拥有一个完整的电子商务系统和快捷的物流配送系统。特别是 ZARA 下属的各个服装生产工厂,都可通过各自的渠道向集团公司的大物流配送中心直递产品,然后由配送中心每周两次向全球 1 000 多个分仓发货,以完成服装产品的终端销售。

ZARA 在法国、德国、意大利、西班牙等欧盟国家以卡车运送为主,平均 48h 即可运达连锁店,在这些地区的销售占总销售量的 70%。剩下 30%的销售量,则以空运的方式送到日本、美国、东欧等较远的国家和地区。

#### 4. 专卖店直销

ZARA 的零售只设专卖店,不开展特许经营。在门店每天上报总部的数据中,不仅包括订单、销量和销售走势等量化指标,也包括顾客反馈、流行信息等“软”数据,而且,“软”数据甚至细化到风格、颜色、材质以及可能的价格等。为了方便每位门店经理及时地向总部传送这些信息,ZARA 还专门为每位店长配备了特制的手提数据传输设备。

除此之外,设计团队也实时与全球各地的零售店店长进行电话会议,通过各地的销售状况与顾客反馈,灵活地调整商品的设计方向,满足客人的多变口味,而且在顾客购买的同时,店员已经将商品特征以及顾客数据输入计算机,由网络将数据送回 ZARA 总部,设计团队就能掌握各种精确的销售分析数据与顾客喜好,再结合自身的专业判断,来决定下一批产品的设计风格取向与生产数量。

从 ZARA 的实际运营可以看出,供应链只是供应链管理的客体,供应链管理是一种集成的管理思想和方法,它执行供应链中从供应商到最终用户的物流的计划和控制等职能。表 1-1 所示为几个常见的关于供应链管理的定义。

表 1-1 供应链管理的定义

定 义	提出者及提出时间
供应链管理是为了生产和提供最终产品,包括从供应商的供应商到客户的客户的一切活动	国际供应链协会(Supply Chain Council,SCC),1996
供应链管理是从供应链整体目标出发,对供应链中采购、生产、销售各环节的商流、物流、信息流及资金流进行统一计划、组织、协调、控制的活动和过程	中华人民共和国国家标准《物流术语》(GB/T 18354—2021),2021
供应链管理即对供应链上所有企业的采购、生产、销售环节的商流、物流、信息流、资金流活动进行的协调、整合、优化和控制等	克里斯托弗,2019
供应链管理是使以核心企业为中心的供应链运作达到最优化,以最低的成本,令供应链从采购开始到满足最终顾客的所有过程,包括工作流、物流、资金流和信息流等均高效率地运转,把合适的产品以合理的价格及时准确地送到消费者手上	马士华和林勇,2020

综上所述,供应链管理是通过对供应商、制造商、分销商、零售商以及最终用户进行全要素、全过程的集成化管理,实现物流、信息流和资金流的高效运作,从而实现供应链的全局最优管理模式。

## 1.3 供应链优化的主要方法

凡事追求尽善尽美是人的天性,因而人们在解决实际问题的时候往往需要寻求最优的方案,尤其在运营管理领域,优化更是基石,所以最优化理论与方法在运营管理领域得到了广泛应用。

### 1.3.1 最优化理论与方法

#### 1. 基本概念

从数学意义上说,最优化方法是一种求极值的方法,即在一组约束为等式或不等式的条件下,使系统的目标函数取得极值,即最大值或最小值。从经济意义上看,是在一定的人力、物力和财力资源条件下,使经济效益(如产值、利润)达到最大,或者在完成规定的生产或经济任务下,使投入的物力和财力等资源为最少。

用最优化方法解决实际问题,一般要经过下列步骤:①提出最优化问题,收集有关数据和资料;②建立最优化问题的数学模型,确定变量,列出目标函数和约束条件;③分析模型特征,选择合适的最优化方法;④求解模型,一般通过编制程序,用计算机求最优解;⑤最优解的检验和实施。上述5个步骤中的工作相互支撑和相互制约,在实践中常常反复交叉进行。

#### 2. 模型的基本要素和分类

最优化问题一般由3个要素构成。①决策变量,表示在最优化问题中要求解的变量;②目标函数,表示需要最大化或最小化的表达式;③约束条件,表示决策变量需要满足的等式条件或不等式条件。

为了便于利用最优化方法分析问题,通常根据最优化问题数学模型的特点对其进行分类,从而根据其特性设计相应的求解算法。

(1) 无约束最优化问题与约束最优化问题。无约束最优化问题本质上是求一个函数的极值问题,即  $\min f(x)$ , 其中  $x \in \mathbb{R}^n$  称为决策变量,  $f(x) \in \mathbb{R}$  称为目标函数。上述问题的解称为最优解,记为  $x^*$ , 该点的函数值称为最优值。

如果极值问题受到了某些条件的限制,该极值问题就成为约束最优化问题,如:

$$\begin{aligned} \min f(x) \\ \text{s. t. } c_i(x) = 0, \quad i \in \varepsilon \\ c_i(x) \geq 0, \quad i \in \gamma \end{aligned} \quad (1-1)$$

式中,包含等式约束和不等式约束, $\varepsilon$  和  $\gamma$  为约束的指标集合。

(2) 连续最优化问题与离散最优化问题。这一分类方法主要根据决策变量的取值进行划分。

$$\text{目标函数: } f(x) = (x - 1.4)^2 + 1; \text{ 约束条件: } x \in \mathbb{R} / x \in \mathbb{Z} \quad (1-2)$$

(3) 随机最优化问题与确定性最优化问题。随机优化是指目标函数或约束条件中涉及随机变量而具有不确定性的问题。若优化问题中不含有随机变量,则为确定性最优化问题。

(4) 线性规划问题与非线性规划问题。若目标函数和约束条件均为决策变量的线性函数,这样的问题为线性规划问题。而目标函数和约束条件中至少存在一个关于决策变量的非线性函数,则为非线性规划问题。

(5) 凸优化问题与非凸优化问题。凸优化问题是指目标函数和约束条件都是凸函数的优化问题。如果其中有一个或者两个都不是凸函数,那么相应的最小化问题是非凸优化问题。因为凸优化问题的任意局部最优解都是全局最优解,所以,相应的算法设计以及理论分析比非凸优化问题简单得多。

如果  $C(x \in C)$  为一凸集,  $f: C \rightarrow \mathbb{R}$  和  $c_i: C \rightarrow \mathbb{R}$  为凸函数,则下式定义的优化模型为一凸优化模型:

$$\begin{aligned} \min_x f(x) \\ \text{s. t. } c_i(x) \leq 0, \quad i \in [n] \end{aligned} \quad (1-3)$$

因此,凸优化指的是在凸集上对目标凸函数取最小值,约束条件是具有上限限制的凸函数集。

### 3. 模型最优解和最优值的求解方法

首先给出最优化模型解的相关定义。设最优化问题为  $(P) \min_{x \in D} f(x), D = \{x \mid g_i(x) \geq 0, i = 1, 2, \dots, m\}$ 。

**定义 1.1** 如果存在  $x^* \in D$  使得  $f(x^*) = \min_{x \in D} f(x)$ , 即  $\exists x^* \in D$ , 使得对  $\forall x \in D$  有  $f(x) \geq f(x^*)$ , 则称  $x^*$  为问题(P)的全局最优解,  $f(x^*)$  为全局最优值。

在定义中,如果当  $\forall x \in D$  且  $x \neq x^*$  时,恒有  $f(x) > f(x^*)$ , 则称  $x^*$  为问题(P)的严格全局最优解,  $f(x^*)$  为严格全局最优值。

**定义 1.2** 如果存在  $x^* \in D$  及  $\delta > 0$ , 使得当  $x \in D \cap N_\delta(x^*)$  时恒有  $f(x) \geq f(x^*)$ , 则称  $x^*$  为问题(P)的局部最优解,  $f(x^*)$  为局部最优值。

这里的  $N_\delta(x^*) = \{x \mid \|x - x^*\| < \delta\}$  为  $x^*$  的  $\delta$  邻域。范数  $\|\cdot\|$  指的是  $\|x\| = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2}$ 。

同样,在定义中,如果当  $x \neq x^*$  时可将“ $\geq$ ”改为“ $>$ ”, 则称  $x^*$  为问题(P)的严格局部最优解,  $f(x^*)$  为严格局部最优值。

根据有无约束以及约束特征,求解最优化模型的主要方法包括微分学中的极值法、拉格朗日(Lagrange)乘数法以及 KKT(Karush-Kuhn-Tucker)条件等。

#### 1) 极值法

广义的极值法是求解无约束优化问题的主要方法。该方法主要依托导数(或偏导数)判断函数的单调性,从而找到可能的极值点。

对于一元函数的极值问题,有如下结论。

**定理 1.1(必要条件)** 设函数  $f(x)$  在  $x_0$  处可导,并且在  $x_0$  处取得极值,则  $f'(x_0) = 0$ 。

**定理 1.2** (第一充分条件) 设函数  $f(x)$  在  $x_0$  处连续, 并且在  $x_0$  的某去心邻域  $\dot{U}(x_0, \delta)$  内可导。

(1) 若  $x \in (x_0 - \delta, x_0)$  时,  $f'(x) > 0$ , 而  $x \in (x_0, x_0 + \delta)$  时,  $f'(x) < 0$ , 则  $f(x)$  在  $x_0$  处取得极大值。

(2) 若  $x \in (x_0 - \delta, x_0)$  时,  $f'(x) < 0$ , 而  $x \in (x_0, x_0 + \delta)$  时,  $f'(x) > 0$ , 则  $f(x)$  在  $x_0$  处取得极小值。

(3) 若  $x \in \dot{U}(x_0, \delta)$  时,  $f'(x)$  的符号保持不变, 则  $f(x)$  在  $x_0$  处没有极值。

当函数在驻点处的二阶导数存在且不为零时, 也可以通过在驻点处的二阶导数的符号来判定在该驻点取极大值还是极小值。

**定理 1.3** (第二充分条件) 设函数  $f(x)$  在  $x_0$  处具有二阶导数且  $f'(x_0) = 0$ ,  $f''(x_0) \neq 0$ , 则:

(1) 当  $f''(x_0) < 0$  时, 函数  $f(x)$  在  $x_0$  处取得极大值。

(2) 当  $f''(x_0) > 0$  时, 函数  $f(x)$  在  $x_0$  处取得极小值。

对于二元函数的极值问题, 有如下结论。

**定理 1.4** (必要条件) 设函数  $f(x, y)$  在  $(x_0, y_0)$  处可导, 并且在  $(x_0, y_0)$  处取得极值, 则

$$\frac{\partial f}{\partial x}(x_0, y_0) = 0, \quad \frac{\partial f}{\partial y}(x_0, y_0) = 0 \quad (1-4)$$

**定理 1.5** (第一充分条件) 设函数  $f(x, y)$  在  $(x_0, y_0)$  处连续, 并且在  $(x_0, y_0)$  的某去心邻域  $\dot{U}((x_0, y_0), \delta)$  内可导。

(1) 若对  $(x, y) \in \dot{U}[(x_0, y_0), \delta]$  中任意点, 当  $x$  增大或  $y$  增大时  $f(x, y)$  均比  $f(x_0, y_0)$  小, 即  $f(x_0, y_0)$  是该邻域内的极大值。

(2) 若对  $(x, y) \in \dot{U}[(x_0, y_0), \delta]$  中任意点, 当  $x$  增大或  $y$  增大时  $f(x, y)$  均比  $f(x_0, y_0)$  大, 即  $f(x_0, y_0)$  是该邻域内的极小值。

(3) 若在邻域  $\dot{U}[(x_0, y_0), \delta]$  内  $f(x, y)$  没有满足以上条件, 则  $f(x, y)$  在  $(x_0, y_0)$  处没有极值。

**定理 1.6** (第二充分条件) 设函数  $f(x, y)$  在  $(x_0, y_0)$  处具有二阶偏导数, 如果  $f(x, y)$  的一阶导数在某个点  $(x_0, y_0)$  为零, 可以通过计算二阶偏导数来进一步判断。定义海塞 (Hessian) 矩阵为

$$\mathbf{H} = \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix} \quad (1-5)$$

并计算海塞矩阵的行列式  $D = f_{xx}f_{yy} - (f_{xy})^2$ 。

海塞矩阵判别准则如下。

(1) 当海塞矩阵的行列式  $D > 0$  且  $f_{xx}(x_0, y_0) > 0$  时,  $f(x, y)$  在  $(x_0, y_0)$  取得极小值。

(2) 当海塞矩阵的行列式  $D > 0$  且  $f_{xx}(x_0, y_0) < 0$  时,  $f(x, y)$  在  $(x_0, y_0)$  取得极大值。

(3) 当海塞矩阵的行列式  $D < 0$  时,  $f(x, y)$  在  $(x_0, y_0)$  处是一个鞍点, 即不取得极值。

(4) 当海塞矩阵的行列式  $D=0$  时,无法通过海塞矩阵确定该点的极值性质,需要进一步分析。

对于多元函数的极值问题,有如下结论:多元函数的导数法涉及偏导数和梯度等概念,是求解多变量函数极值问题的核心工具。对于一个多元函数  $f(x_1, x_2, \dots, x_n)$ ,导数法主要通过分析函数的一阶偏导数和二阶偏导数(或海塞矩阵)来确定其极值点。

此外,对于无约束优化问题,其他的一些分析方法还有梯度下降法、最速下降法、牛顿法等。

## 2) 拉格朗日乘数法

含等式约束的优化问题,通常使用拉格朗日乘数法求得最优值,即通过引入拉格朗日乘数(Lagrange multiplier)将等式约束和优化问题整合,构造新的函数表达式,再对各个变量求导使其为零,求解出候选最优值。

考虑如下含约束条件的优化问题:

$$\begin{aligned} \max f(x_1, x_2, \dots, x_n) \\ \text{s. t. } g_i(x_1, x_2, \dots, x_n) = 0, \quad i = 1, 2, \dots, m \end{aligned} \quad (1-6)$$

式中,  $f(x_1, x_2, \dots, x_n)$  为要最大化(或最小化)的目标函数;  $g_i(x_1, x_2, \dots, x_n) = 0$  为约束条件。拉格朗日乘数法的基本思想是将约束条件引入目标函数中,构造一个新的函数(称为拉格朗日函数),然后通过求解这个新函数的极值点得到原问题的最优值。

具体地,定义拉格朗日函数为

$$L(x_1, x_2, \dots, x_n, \lambda_1, \lambda_2, \dots, \lambda_m) = f(x_1, x_2, \dots, x_n) + \sum_{i=1}^m \lambda_i g_i(x_1, x_2, \dots, x_n) \quad (1-7)$$

式中,  $\lambda_i$  为拉格朗日乘数,用来表示约束条件对目标函数的影响程度。这样,原来的含约束条件的优化问题就被转化为一个无约束的优化问题,即求解拉格朗日函数  $L(x_1, x_2, \dots, x_n, \lambda_1, \lambda_2, \dots, \lambda_m)$  的极值点。

求解拉格朗日函数的极值点的具体步骤如下。

(1) 对所有的自变量  $x_i$  和拉格朗日乘数  $\lambda_i$  分别计算拉格朗日函数的偏导数,即  $\frac{\partial L}{\partial x_i}$  和  $\frac{\partial L}{\partial \lambda_i}$ 。

(2) 令上述偏导数等于零,即  $\frac{\partial L}{\partial x_i} = 0$  和  $\frac{\partial L}{\partial \lambda_i} = 0$ 。

(3) 解方程组,找到所有满足条件的解。

(4) 将解代入原目标函数中,计算对应的函数值。

(5) 找出使得目标函数取得最大值(或最小值)的解,即为含约束条件的优化问题的解。

拉格朗日乘数法的优点在于,它能够将有约束条件的优化问题转化为一个无约束的优化问题,从而方便了问题的求解。然而,对于复杂的约束条件和目标函数,可能会存在多个极值点,因此需要结合实际条件进一步分析和验证。

## 3) KKT 条件

含不等式约束的优化问题通常使用 KKT 条件进行求解。KKT 条件为解决含等式和

不等式约束的优化问题提供了必要的条件。在处理含不等式约束的优化问题时, KKT 条件提供了关于最优解的必要条件, 主要包括: ①梯度条件。在极值点处, 目标函数的梯度可以表示为约束条件的梯度的线性组合。②互补松弛性。对于不等式约束, 如果某个约束在最优解处是活动的(即等号成立), 那么该约束的拉格朗日乘数(权值)应该为零。这表明, 只有在最优解位于不等式约束的边界上时, 对应的拉格朗日乘数才不为零。③可行性。从可行域中的每一个点出发, 沿着某个方向前进一点点, 如果仍然保持在可行域内或者偏离可行域的程度非常小(即偏移量是行进距离的高阶无穷小量), 那么这个方向为可行方向。KKT 条件确保了最优解不仅满足目标函数的优化条件, 还满足所有约束条件的可行性。

综上, KKT 条件为解决非线性优化问题提供了一种系统的方法(步骤如图 1-2 所示), 尤其是在处理不等式约束时。它不仅为理论分析提供了基础, 也为实际应用中的优化算法设计提供了指导。

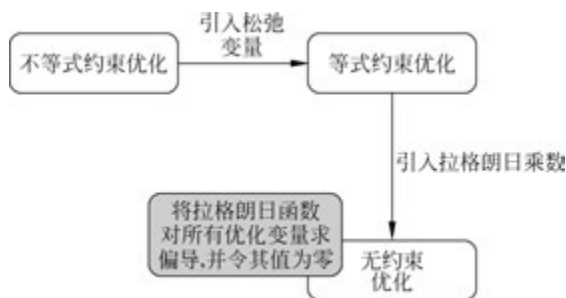


图 1-2 从不等式约束优化到无约束优化

考虑一个如下包含等式约束和不等式约束的优化问题:

$$\begin{aligned} \min f(x) \\ \text{s. t. } g_i(x) \leq 0, \quad i = 1, 2, \dots, m \\ h_j(x) = 0, \quad j = 1, 2, \dots, n \end{aligned} \quad (1-8)$$

式中,  $f, h_j: \mathbb{R}^n \rightarrow \mathbb{R}$  均为连续可微的函数。令  $\mathcal{X} = \{x \in \mathbb{R}^n \mid g_i(x) \leq 0, h_j(x) = 0\}$ 。假设  $x^*$  是该优化问题的最优点(即局部最小点), 则有参数  $\lambda_i \geq 0, \eta_j \in \mathbb{R}$ , 并且在  $x^*$  处满足线性独立约束条件(linear independent constraint qualification, LICQ), 使得 KKT 条件为

$$\begin{cases} \nabla f(x^*) + \sum_{i=1}^m \lambda_i \nabla g_i(x^*) + \sum_{j=1}^n \eta_j \nabla h_j(x^*) = 0 \\ \lambda_i g_i(x^*) = 0, i = 1, 2, \dots, m \\ \lambda_i \geq 0, i = 1, 2, \dots, m \\ g_i(x^*) \leq 0, i = 1, 2, \dots, m \\ h_j(x^*) = 0, j = 1, 2, \dots, n \end{cases} \quad (1-9)$$

**【例 1-1】** 利用 KKT 方法求解如下优化问题的最优解。

$$\begin{aligned} \min f(x) &= \left(x_1 - \frac{4}{9}\right)^2 + (x_2 - 2)^2 \\ \text{s. t. } &\begin{cases} -x_1^2 + x_2 \geq 0 \\ x_1 + x_2 \leq 6 \\ x_1, x_2 \geq 0 \end{cases} \end{aligned}$$

解 首先,将约束条件统一为“ $\leq$ ”的形式:

$$\begin{cases} h_1(x) = x_1^2 - x_2 \leq 0 \\ h_2(x) = x_1 + x_2 - 6 \leq 0 \\ g_1(x) = -x_1 \leq 0 \\ g_2(x) = -x_2 \leq 0 \end{cases}$$

校验其是否满足正则(regularity)条件:

$$\nabla h_1(x) = \begin{bmatrix} 2x_1 \\ -1 \end{bmatrix}, \quad \nabla h_2(x) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \nabla g_1(x) = \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \quad \nabla g_2(x) = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$$

上述结果线性无关满足正则条件。

然后,将目标函数和约束条件转换成拉格朗日函数:

$$\begin{aligned} L(x, \lambda_1, \lambda_2, \mu_1, \mu_2) &= f(x) + \lambda_1 h_1(x) + \lambda_2 h_2(x) + \mu_1 g_1(x) + \mu_2 g_2(x) \\ &= \left(x_1 - \frac{4}{9}\right)^2 + (x_2 - 2)^2 + \lambda_1(x_1^2 - x_2) + \lambda_2(x_1 + x_2 - 6) - \\ &\quad \mu_1 x_1 - \mu_2 x_2 \end{aligned}$$

再通过 KKT 条件建立方程组:

$$\nabla L_x = \left( \frac{\partial L}{\partial x_1}, \frac{\partial L}{\partial x_2} \right) = \left[ 2\left(x_1 - \frac{4}{9}\right) + 2\lambda_1 x_1 + \lambda_2 - \mu_1, 2(x_2 - 2) - \lambda_1 + \lambda_2 - \mu_2 \right] = (0, 0)$$

$$\begin{cases} 2\left(x_1 - \frac{4}{9}\right) + 2\lambda_1 x_1 + \lambda_2 - \mu_1 = 0 \\ 2(x_2 - 2) - \lambda_1 + \lambda_2 - \mu_2 = 0 \\ h_1(x) = x_1^2 - x_2 \leq 0 \\ h_2(x) = x_1 + x_2 - 6 \leq 0 \\ g_1(x) = -x_1 \leq 0 \\ g_2(x) = -x_2 \leq 0 \\ \lambda_1(x_1^2 - x_2) = 0 \\ \lambda_2(x_1 + x_2 - 6) = 0 \\ -\mu_1 x_1 = 0 \\ -\mu_2 x_2 = 0 \\ \lambda_1, \lambda_2 \geq 0 \\ \mu_1, \mu_2 \geq 0 \end{cases}$$

下面结合互补松弛定理分情形讨论。

(1) 情形 1:  $\lambda_1 > 0, \lambda_2 > 0$ 。由互补松弛定理可得

$$\begin{cases} \lambda_1 > 0 \Rightarrow x_1^2 - x_2 = 0 \\ \lambda_2 > 0 \Rightarrow x_1 + x_2 - 6 = 0 \end{cases}$$

可解得

$$\begin{cases} x_1 = 2 \\ x_2 = 4 \end{cases} \quad \text{或} \quad \begin{cases} x_1 = -3 \\ x_2 = 9 \end{cases}$$

由于  $\begin{cases} x_1 = -3 \\ x_2 = 9 \end{cases}$  不满足约束条件  $\begin{cases} g_1(x) = -x_1 \leq 0 \\ g_2(x) = -x_2 \leq 0 \end{cases}$ , 故舍去; 将  $\begin{cases} x_1 = 2 \\ x_2 = 4 \end{cases}$  代入

$$\begin{cases} 2\left(x_1 - \frac{4}{9}\right) + 2\lambda_1 x_1 + \lambda_2 - \mu_1 = 0 \\ 2(x_2 - 2) - \lambda_1 + \lambda_2 - \mu_2 = 0 \\ -\mu_1 x_1 = 0 \\ -\mu_2 x_2 = 0 \end{cases} \quad \text{中, 计算拉格朗日乘数, 并验证其非负性。可解得拉格朗日乘数为}$$

日乘数为

$$\begin{cases} \lambda_1 = \frac{8}{45}, \lambda_2 = -\frac{172}{45} < 0 \\ \mu_1 = 0, \mu_2 = 0 \end{cases}$$

上述数值不满足拉格朗日乘数的非负性, 故该点不是优化问题的 KKT 点。因此, 在此情形下, 无法找出最优解。

(2) 情形 2:  $\lambda_1 > 0, \lambda_2 = 0$ 。由互补松弛定理可得

$$\lambda_1 > 0 \Rightarrow x_1^2 - x_2 = 0$$

此时, 梯度条件为

$$\begin{cases} 2\left(x_1 - \frac{4}{9}\right) + 2\lambda_1 x_1 - \mu_1 = 0 \\ 2(x_2 - 2) - \lambda_1 - \mu_2 = 0 \end{cases}$$

若  $x_1 > 0, x_2 > 0$ , 由  $\begin{cases} -\mu_1 x_1 = 0 \\ -\mu_2 x_2 = 0 \end{cases}$  可得  $\mu_1 = \mu_2 = 0$ 。代入梯度条件得

$$\begin{cases} 2\left(x_1 - \frac{4}{9}\right) + 2\lambda_1 x_1 = 0 \\ 2(x_2 - 2) - \lambda_1 = 0 \end{cases}$$

联立  $x_1^2 - x_2 = 0$  可得

$$18x_1^3 - 27x_1 - 4 = 0$$

求解该三次方程, 并保证方程解的非负性, 可得  $x_1 \approx 1.3$ 。此时,  $x_2 = x_1^2 = 1.69$ , 将其代入 KKT 条件中, 计算拉格朗日乘数, 并验证其非负性。可解得拉格朗日乘数为

$$\begin{cases} \lambda_1 = -0.62, \lambda_2 = 0 \\ \mu_1 = 0, \mu_2 = 0 \end{cases}$$

上述数值不满足拉格朗日乘数的非负性, 故该点不是优化问题的 KKT 点。若  $x_1 = 0$ , 则  $x_2 = x_1^2 = 0$ , 将其代入 KKT 条件中并进行验证:

$$\begin{cases} x_1 + x_2 - 6 < 0 \\ x_1^2 - x_2 = 0 \\ \mu_1 = -\frac{8}{9} \\ \lambda_1 = \mu_2 + 4 \end{cases}$$

上述数值不满足拉格朗日乘数的非负性,故该点不是优化问题的 KKT 点。若  $x_2=0$ ,则  $x_2=x_1^2=0$ ,将其代入 KKT 条件中与上述情况相同,故该点不是优化问题的 KKT 点。因此,在此情形下,无法找出最优解。

(3) 情形 3:  $\lambda_1=0, \lambda_2>0$ 。由互补松弛定理可得

$$\lambda_2 > 0 \Rightarrow x_1 + x_2 - 6 = 0$$

此时,梯度条件为

$$\begin{cases} 2\left(x_1 - \frac{4}{9}\right) + \lambda_2 - \mu_1 = 0 \\ 2(x_2 - 2) + \lambda_2 - \mu_2 = 0 \end{cases}$$

若  $x_1>0, x_2>0$ ,由  $\begin{cases} -\mu_1 x_1 = 0 \\ -\mu_2 x_2 = 0 \end{cases}$  可得  $\mu_1 = \mu_2 = 0$ 。代入梯度条件得

$$\begin{cases} 2\left(x_1 - \frac{4}{9}\right) + \lambda_2 = 0 \\ 2(x_2 - 2) + \lambda_2 = 0 \end{cases}$$

求解上述方程组可得  $x_2 = x_1 + \frac{14}{9}$ 。结合  $x_1 + x_2 - 6 = 0$ ,解得

$$\begin{cases} x_1 = \frac{20}{9} \\ x_2 = \frac{34}{9} \end{cases}$$

将上述数值代入 KKT 条件中,计算拉格朗日乘数,并验证其非负性,可得

$$\lambda_2 = -\frac{32}{9} < 0$$

这表明上述求解结果不满足拉格朗日乘数的非负性,故该点不是优化问题的 KKT 点。若  $x_1=0$ ,则  $\mu_1 = -\frac{8}{9}$ ,不满足拉格朗日乘数的非负性。若  $x_1>0, x_2=0$ ,由  $-\mu_1 x_1 = 0$  可得  $\mu_1 = 0$ ,并且由  $x_1 + x_2 - 6 = 0$  可得  $x_1 = 6$ 。将  $x_1, x_2$  代入 KKT 条件中,计算拉格朗日乘数,并验证其非负性,可得

$$\lambda_2 = -\frac{100}{9} < 0$$

显然,上述计算结果不满足拉格朗日乘数的非负性,故该点不是优化问题的 KKT 点。因此,在此情形下,不存在最优解。

(4) 情形 4:  $\lambda_1=0, \lambda_2=0$ 。由互补松弛定理可得

$$\begin{cases} \lambda_1 = 0 \Rightarrow x_1^2 - x_2 < 0 \\ \lambda_2 = 0 \Rightarrow x_1 + x_2 - 6 < 0 \end{cases}$$

从而可知  $x_2>0$ 。由  $-\mu_2 x_2 = 0$  可得  $\mu_2 = 0$ ,由梯度条件  $2(x_2 - 2) - \mu_2 = 0$  可得

$$x_2 = 2$$

若  $x_1=0$ , 则  $\mu_1=-\frac{8}{9}$ , 不满足拉格朗日乘数的非负性。若  $x_1>0$ , 由  $-\mu_1 x_1=0$  可得  $\mu_1=0$ , 由梯度条件  $2\left(x_1-\frac{4}{9}\right)-\mu_1=0$  可得

$$x_1=\frac{4}{9}$$

将上述计算结果代入后可知其满足 KKT 条件, 故  $\left(\frac{4}{9}, 2\right)$  为极值点, 此时  $f(x)$  的极小值为

$$f(x)=\left(\frac{4}{9}-\frac{4}{9}\right)^2+(2-2)^2=0$$

## 1.3.2 博弈理论

### 1. 基本概念

博弈论(game theory)是研究多个个体或团队之间在特定条件制约下的对局中利用相关方的策略, 而实施对应策略的学科。其有时又称对策论, 或者赛局理论, 是研究具有斗争或竞争性质现象的理论和方法, 是应用数学的一个分支, 也是现代数学的一个新分支与运筹学的一个重要学科。目前, 博弈论在生物学、经济学、国际关系学、计算机科学、政治学、军事战略学和其他很多学科都有广泛的应用, 主要研究公式化的激励结构(游戏或者博弈)间的相互作用。

下面介绍博弈论中的一些主要概念。

① 局中人(players): 在一场竞赛或博弈中, 每一个有决策权的参与者称为一个局中人。只有两个局中人的博弈现象称为“两人博弈”, 而多于两个局中人的博弈称为“多人博弈”。

② 策略(strategies): 一局博弈中, 每个局中人都选择实际可行的完整的行动方案, 即方案不是某阶段的行动方案, 而是指导整个行动的一个方案, 一个局中人的一个可行的贯穿全局筹划的一个行动方案, 称为这个局中人的一个策略。如果在一局博弈中, 所有局中人总共有有限个策略, 则称为“有限博弈”, 否则称为“无限博弈”。

③ 得失(payoffs): 一局博弈结局时的结果称为得失。每个局中人在一局博弈结束时的得失, 不仅与该局中人自身所选择的策略有关, 而且与全局中人所取定的一组策略有关。所以, 一局博弈结束时每个局中人的“得失”是全体局中人所取定的一组策略的函数, 通常称为支付(payoff)函数。

④ 次序(orders): 各博弈方的决策有先后之分, 并且一个博弈方要做不止一次的决策选择, 于是就出现了次序问题; 若其他要素相同次序不同, 博弈也就不同。

⑤ 博弈涉及均衡: 均衡是平衡的意思, 在经济学中, 均衡即指相关量处于稳定值。例如, 在供求关系中, 某一商品市场如果在某一价格下, 想以此价格买此商品的人均能买到, 而想卖的人均能卖出, 此时就说该商品的供求达到了均衡。

结合先前对博弈论基本概念的梳理,借鉴当前关于博弈论方法的总结和介绍<sup>①</sup>,下面将系统介绍几种经典博弈模型及其分析过程。

## 2. 斯塔克尔伯格博弈

斯塔克尔伯格博弈(Stackelberg game)是一个两阶段的完全信息动态博弈,博弈的时序是序贯的。该模型基于产量展开博弈,主要思想是:领导者(leader)和追随者(followers)均根据对方的策略选择自身最优策略,以实现自身利益最大化,从而达到纳什均衡。在该博弈模型中,先做出决策的一方称为领导者,在领导者做出决策后,剩余的球员观测到其决策并据此进行决策(称为追随者),双方策略共同构成纳什均衡。

在一般的斯塔克尔伯格模型中,厂商之间存在行动次序的区别。产量决策遵循如下次序:领导者厂商先确定一个产量,跟随者厂商观察到该产量后,再据此决定自身产量。要注意的是,领导者厂商在决定产量时,充分了解跟随者厂商的行动逻辑,即知晓其反应函数。因此,领导者厂商能够预判自身产量决策对跟随者厂商的影响。正是基于这种影响的考量,领导者厂商的产量决策以跟随者厂商的反应函数为约束条件,追求自身利润最大化。在斯塔克尔伯格模型中,领导型厂商的决策无须依赖自身的反应函数。上述逻辑可通过数学语言形式化描述。具体来说,设市场需求价格函数为

$$D = D(p_1 + p_2) = a - b(p_1 + p_2) \quad (1-10)$$

式中, $p_1$ 和 $p_2$ 分别为两个企业的销售价格。假设两个企业的成本函数相同,都为 $C = cp$ ,即

$$\max p_2[a - b(p_1 + p_2)] - cp_2 \quad (1-11)$$

在企业2对任意给定产量做出反应后,企业1的最优产量模型为

$$\begin{aligned} \max p_1[a - b(p_1 + p_2)] - cp_1 \\ \text{s. t. } p_2 = g(p_1) \end{aligned} \quad (1-12)$$

因此斯塔克尔伯格模型先求解如下的优化模型:

$$\max p_2[a - b(p_1 + p_2)] - cp_2 \quad (1-13)$$

得到 $p_2 = g(p_1)$ 。

再求解如下的优化模型:

$$\begin{aligned} \max p_1[a - b(p_1 + p_2)] - cp_1 \\ \text{s. t. } p_2 = g(p_1) \end{aligned} \quad (1-14)$$

得到 $p_1$ ,代入 $p_2 = g(p_1)$ ,得到 $p_2$ ,最终得到斯塔克尔伯格均衡时的 $(p_1, p_2)$ 。

**【例 1-2】** 假设市场需求价格函数为 $D = 61.2 - 10 \cdot (p_1 + p_2)$ ,两个企业的成本函数都为 $C = 1.2p$ ,求斯塔克尔伯格均衡时两个企业的产量(企业1为领导者,企业2为跟随者)。

这个例子其实就是先求解如下的优化模型:

$$\max p_2[61.2 - 10 \cdot (p_1 + p_2)] - 1.2p_2$$

<sup>①</sup> 博客园. 博弈论——古诺博弈及其变形模型(十八)[EB/OL]. (2023-11-16) <https://www.cnblogs.com/haohai9309/p/17833572.html>.

得到  $p_2 = \frac{60 - 10p_1}{20}$ 。然后求解下列优化模型：

$$\max p_1 [61.2 - 10 \cdot (p_1 + p_2)] - 1.2p_1, \quad \text{s. t.} \quad p_2 = \frac{60 - 10p_1}{20}$$

最终得到结果为  $p_1 = 3, p_2 = 1.5$ 。

从上述例子可以看出，求解斯塔克尔伯格模型通常需要确定领导者和追随者的最佳响应策略，即使得各自收益最大化的策略。一种常用的求解方法是使用逆推法(backward induction)。

(1) 追随者的最佳响应：给定领导者的策略  $s_L$ ，追随者选择使得自己的收益最大化的策略  $s_F^*(s_L)$ 。

(2) 领导者的最佳响应：在追随者选择了策略  $s_F^*(s_L)$ 后，领导者选择使得自己的收益最大化的策略  $s_L^*$ 。

(3) 确定斯塔克尔伯格均衡：将追随者的最佳响应  $s_F^*(s_L)$ 和领导者的最佳响应  $s_L^*$ 结合起来，构成斯塔克尔伯格均衡。

### 3. 纳什均衡

首先引入一个博弈论中经典的例子。

**【例 1-3】(囚徒困境)** 两个小偷 A 和 B 共同作案，因非法侵入他人住宅而被警察抓获。警方将两人分别置于两个不同的房间内进行审讯。

如果一个犯罪嫌疑人坦白了罪行，交出了赃物，于是证据确凿，两人都被判有罪。

如果另一个犯罪嫌疑人也做了坦白，则两人各被判刑两年。

如果另一个犯罪嫌疑人没有坦白而是抵赖，则以妨碍公务罪(因已有证据表明其有罪)再加刑一年，而坦白者有功被减刑两年，立即释放。

如果两人都抵赖，则警方因证据不足不能判两人的偷窃罪，但可以非法侵入他人住宅的罪名将两人各判入狱一年。

囚徒困境可以用一个支付矩阵来表示，其中每个囚徒都有两个策略，矩阵的每个单元格表示根据双方选择的策略组合，各自的收益(刑期，以数值大小表示服刑时长)如表 1-2 所示。

表 1-2 囚徒困境支付矩阵

选 择	小偷 B 坦白	小偷 B 不坦白
小偷 A 坦白	(2, 2)	(0, 3)
小偷 A 不坦白	(3, 0)	(1, 1)

从个人理性的角度出发，每个囚徒都会思考对方可能的选择，并据此来决定自己的最优策略。不难发现，无论对方选择什么，坦白总是看起来对自己更有利。

(1) 如果预期对方不坦白，自己坦白能得到自由。

(2) 如果预期对方坦白，自己也坦白能获得较短的刑期。

因此，逻辑推理表明两人都会选择坦白，尽管如果他们都不坦白，总体上会有更好的结果(每人只需服刑一年，而非两年)。

在上述例子中，一个人的最优选择并不依赖于他人的选择。这样的最优策略，称为占优

策略(dominant strategy)。由所有参与人的占优策略构成的策略组合称为占优策略均衡(dominant-strategy equilibrium)。从其中一个罪犯的角度来看,无论是对方选择认罪还是不认罪,都是自己选择认罪的策略对自己最优,尽管双方都不认罪的做法对两个人总体来说是最优解,这里体现了个人理性和集体理性的冲突。这也表明,当个体在决策时仅关注自身利益最大化,而不考虑其行为对群体或他人的影响时,可能会导致群体利益受损。

在囚徒困境的例子中,当两个罪犯都选择坦白,此时两个罪犯之间的策略互动达到一种相对静止的状态。这一状态在博弈论中称为均衡。简单来说,博弈论中的均衡就是参与者策略互动的稳定点。

纳什均衡是较为著名的均衡概念之一,由约翰·纳什(John Nash)在20世纪提出。它描述了在博弈中各参与者选择策略的一种稳定状态,即每个参与者的策略都是对其他参与者策略选择的最佳响应。纳什均衡是一个重要的博弈解概念,有助于理解博弈过程中的策略选择和结果分析。

**定义 1.3** 在一个博弈中,如果每个参与者都选择了一个策略,并且没有一个参与者可以通过单独改变自己的策略来提高自己的支付,则此时达到了纳什均衡。

更具体地说,对于一个有限的博弈,假设有 $n$ 个参与者,每个参与者有一个策略集合,记为 $S_i$ ,其中 $i=1,2,\dots,n$ 。如果对于每个参与者 $i$ 的策略组合 $s_i^* \in S_i$ ,都有其他参与者选择的策略组合 $s_{-i}^*$ 使得参与者 $i$ 的支付最大化,则该策略组合 $(s_1^*, s_2^*, \dots, s_n^*)$ 就是一个纳什均衡。

**性质 1.1** ①存在性。纳什证明了在有限博弈中至少存在一个纳什均衡。但在无限博弈中,纳什均衡不一定存在。②唯一性。纳什均衡可能存在多个,也可能只有一个。对于有限的零和博弈(zero-sum games),纳什均衡是唯一的。③稳定性。在纳什均衡下,每个参与者的策略都是对其他参与者策略选择的最佳响应,因此该策略组合是一个稳定状态,没有参与者有动机单方面改变自己的策略。

#### 4. 古诺模型

古诺模型又称古诺双寡头模型(Cournot duopoly model)。它是早期的寡头模型,由法国经济学家古诺于1838年提出。该模型是纳什均衡应用的最早版本,通常被视为寡头理论分析的出发点。

古诺模型对寡头垄断市场的基本假设为所有厂商均知晓市场销量、价格及利润信息,因此该模型属于完全信息静态博弈,具体包括如下内容。

- (1) 假定市场为双寡头结构,两家厂商生产同质产品,并且均以利润最大化为目标。
- (2) 双寡头之间进行产量竞争而非价格竞争,产品价格由两者的总产量共同决定。
- (3) 边际成本为常数。
- (4) 面临相同的线性市场需求曲线。
- (5) 双寡头之间无勾结行为,决策相互独立。

首先对上述经典古诺模型进行建模求解。

##### 1) 经典古诺模型求解

设市场反需求函数为 $P=a-bQ$ , $Q=Q_1+Q_2$ ,总成本 $TC_i=cQ_i$ ,边际成本 $MC_i=c$ , $i=1,2$ ,求两家企业的利润最大化。

根据经典古诺模型中的参数假设,寡头1的利润表示为

$$\begin{aligned}\pi_1(Q_1) &= TR_1 - TC_1 = PQ_1 - cQ_1 = [a - b(Q_1 + Q_2)]Q_1 - cQ_1 \\ &= (a - c)Q_1 - bQ_1^2 - bQ_1Q_2 \quad (\text{等利润线})\end{aligned}\quad (1-15)$$

使得利润最大化的一阶条件为

$$\frac{d\pi_1}{dQ_1} = (a - c) - 2bQ_1 - bQ_2 = 0 \quad (1-16)$$

从而得  $Q_1 = \frac{(a-c) - bQ_2}{2b}$ 。该公式表达式是寡头 1 的反应函数, 对应的曲线为反应曲线。

同理, 可得寡头 2 的反应函数为

$$Q_2 = \frac{(a-c) - bQ_1}{2b} \quad (1-17)$$

联合  $Q_1$  和  $Q_2$  对应的反应曲线, 可得

$$Q_1^* = \frac{a-c}{3b} = \frac{1}{2+1} \frac{a-c}{b}, \quad Q_2^* = \frac{a-c}{3b} = \frac{1}{2+1} \frac{a-c}{b} \Rightarrow Q^* = \frac{2(a-c)}{3b} \quad (1-18)$$

$$P^* = a - bQ_1^* = a - b \frac{2(a-c)}{3b} = a - \frac{2}{3}(a-c) \quad (1-19)$$

$$\pi_1^* = (P^* - c) \times Q_1^* = \frac{1}{9} \frac{(a-c)^2}{b} = \frac{1}{(2+1)^2} \frac{(a-c)^2}{b} \quad (1-20)$$

同理, 有

$$\pi_2^* = (P^* - c) \cdot Q_2^* = \frac{1}{9} \frac{(a-c)^2}{b} = \frac{1}{(2+1)^2} \frac{(a-c)^2}{b} \Rightarrow \pi^* = \frac{2}{9} \frac{(a-c)^2}{b} \quad (1-21)$$

## 2) $n$ 家厂商的古诺模型求解

假定古诺模型中有  $n$  个企业, 令  $Q_i$  代表企业  $i$  的产量, 并且  $Q = Q_1 + Q_2 + \dots + Q_n$  表示市场总产量,  $p$  表示市场出清价格, 并假设反需求函数由  $P(Q) = a - bQ$  给出, 设企业  $i$  生产出  $Q_i$  的总成本  $C_i(Q_i) = cQ_i$ , 即没有固定成本, 并且边际成本为常数  $c$ , 这里设  $c < a$ 。根据古诺的假定, 企业同时就产量进行决策。试写出该博弈的模型, 并求出该博弈的纳什均衡; 当  $n$  趋向于无穷时, 将会发生什么情况?

对于第  $i$  个企业, 其目标为最大化自己的利益。若达到纳什均衡, 则

$$\max \pi_i = \max_{Q_i \geq 0} [P(Q) - c]Q_i = \max_{Q_i \geq 0} (a - bQ_i - bQ_{-i} - c)Q_i \quad (1-22)$$

一阶条件为

$$Q_i^* = \frac{1}{2b}(a - bQ_{-i} - c)$$

由于  $Q^* = Q_i^* + Q_{-i}^*$ , 根据对称性, 所有的  $Q_i^*$  都相等, 即  $Q^* = nQ_i^*$ ,  $Q_{-i}^* = (n-1)Q_i^*$ , 得

$$\begin{aligned}Q_i^* &= \frac{a-c}{(n+1)b}, \quad Q^* = n \frac{a-c}{(n+1)b} \\ \pi_i &= \frac{1}{(n+1)^2} \frac{(a-c)^2}{b}\end{aligned}\quad (1-23)$$

当  $n \rightarrow \infty$  时,

$$\lim_{n \rightarrow \infty} \pi_i = \lim_{n \rightarrow \infty} \frac{1}{(n+1)^2} \frac{(a-c)^2}{b} = 0 \quad (1-24)$$

即当企业数量无穷多时,  $P \rightarrow c$ , 即市场出清价格趋向于边际成本, 此时市场趋向于完全竞争市场。根据上面的推导可得最终结果如下:

$$Q_1^* = Q_2^* = \dots = Q_n^* = \frac{1}{n+1} \frac{a-c}{b} = \frac{a-c}{(n+1)b} \quad (1-25)$$

$$\pi_1^* = \pi_2^* = \dots = \pi_n^* = \frac{1}{(n+1)^2} \frac{(a-c)^2}{b} \quad (1-26)$$

上述古诺模型均属于完全信息静态博弈框架, 是经济学中最经典的形式。其核心假设包括厂商相互知晓对方的产量与成本且市场价格统一, 因此博弈双方的收益情况属于共同知识, 不存在信息隐瞒。然而, 在现实经济中, 无论是竞争还是合作的厂商之间, 为追求自身利益, 往往会将生产销售的相关信息作为商业秘密加以保护(这种保密行为是否符合利润最大化目标仍须进一步分析), 导致其他厂商难以获取真实信息。因此, 完全信息静态博弈古诺模型的假设与现实存在一定偏差, 在现实寡头垄断市场的产量博弈中, 各博弈方的收益往往难以成为共同知识。例如, 在双寡头厂商模型中, 若一方不了解另一方的生产成本, 便无法准确预判对方在各种产量组合下的收益。基于此, 有必要深入讨论不完全信息静态博弈古诺模型。

### 3) 不完全信息静态博弈古诺模型求解

假设  $P(Q) = \alpha - Q$ , 厂商 1 的平均成本  $C$  是共同知识, 而厂商 2 的平均成本有两种类型, 即高成本  $C_H$  或低成本  $C_L$ 。厂商 2 知道自己是哪种成本, 而厂商 1 不知道, 但知道  $C_H$  的概率为  $\theta$ ,  $C_L$  的概率为  $(1-\theta)$ 。由已知条件可求得

$$\operatorname{argmax}_{q_1 \geq 0} \{ \theta [P(q_1 + q_L) - C] q_1 + (1-\theta) [P(q_1 + q_H) - C] q_1 \} \quad (1-27)$$

$$\operatorname{argmax}_{q_L \geq 0} \{ [P(q_1 + q_L) - C_L] q_L \} \quad (1-28)$$

$$\operatorname{argmax}_{q_H \geq 0} \{ [P(q_1 + q_H) - C_H] q_H \} \quad (1-29)$$

$$q_1^* = \frac{1}{3} [\alpha - 2C + \theta C_L + (1-\theta) C_H] \quad (1-30)$$

$$q_L^* = \frac{1}{3} (\alpha - 2C_L + C) - \frac{1}{6} (1-\theta) (C_H - C_L) \quad (1-31)$$

$$q_H^* = \frac{1}{3} (\alpha - 2C_H + C) + \frac{1}{6} \theta (C_H - C_L) \quad (1-32)$$

通过对比结果可以发现, 相较于完全信息, 当企业生产成本较高时, 会选择生产更多; 而生产成本较低时, 会选择生产更少。不完全信息使得企业 2 具备一定的优势, 企业 1 最终会选择适中的产量。

**【例 1-4】** 市场中存在两个生产同质商品的企业(企业 1, 企业 2), 其产量分别为  $q_1$ 、 $q_2$ , 市场上的总产量为  $Q = q_1 + q_2$ , 市场价格为  $P = 16 - Q$  ( $Q \leq 16$ ), 企业 2 的边际生产成本为 4, 即  $C(q_2) = 4q_2$ , 而企业 1 的边际成本为私有信息, 但已知其概率分布: 企业 1 以  $1/3$  的概率为低边际成本  $C_L = 2$ ,  $C(q_1; L) = 2q_1$ , 以  $2/3$  的概率为高边际成本  $C_H = 6$ ,  $C(q_1; H) = 6q_1$ 。给定企业 2 的策略也就是产量  $q_2$ , 记参与人 1 的最优反应为  $(q_1(L), q_1(H))$ 。试求解两个企业的最优产量。

**解** 若为低边际成本, 给定企业 2 的策略, 企业 1 的最优反应为

$$\max_{q_1(L)} [16 - q_1(L) - q_2] q_1(L) - 2q_1(L)$$

求最值得到最优反应为

$$q_1(L) = \frac{14 - q_2}{2}$$

若为高边际成本, 给定企业 2 的策略, 企业 1 的最优反应为

$$\max_{q_1(H)} [16 - q_1(H) - q_2]q_1(H) - 6q_1(H)$$

求最值得到最优反应为

$$q_1(H) = \frac{10 - q_2}{2}$$

给定参与人 1 的策略  $[q_1(L), q_1(H)]$ , 参与人 2 的最优反应为  $q_2$ , 有

$$\max_{q_2} \frac{1}{3} \{ [16 - q_1(L) - q_2]q_2 - 4q_2 \} + \frac{2}{3} \{ [16 - q_1(H) - q_2]q_2 - 4q_2 \}$$

求最值得到最优反应为

$$q_2 = \frac{12 - \frac{q_1(L) + 2q_1(H)}{3}}{2}$$

则得到不完全信息古诺模型的最优反应函数的策略组合

$$\{ [q_1^*(L), q_1^*(H)], q_2^* \} \quad (1-33)$$

将上述得到的 3 个最优解联立, 就可以得到  $q_2 = \frac{38}{9}$ , 进而得出策略组合的其他两个解。

## 5. 伯川德模型

伯川德(Bertrand)模型由法国经济学家约瑟夫·伯川德(Joseph Bertrand)于 1883 年提出, 是分析寡头垄断市场中企业价格竞争的经典模型。与古诺模型(以产量为决策变量的产量竞争模型)不同, 伯川德模型以价格为核心决策变量, 聚焦寡头垄断市场的价格竞争行为。在现实竞争中, 定价是企业更基础的战略决策, 消费者对企业产品的需求规模往往直接取决于其定价策略。尤其在寡头垄断市场中, 企业定价策略的差异对产品需求的影响更为显著。因此, 伯川德模型为解析寡头垄断企业价格竞争的行为特征及其市场影响提供了重要的理论框架。

在伯川德模型中, 关键假设如下。

(1) 产品同质性: 所有厂商生产的产品同质无差异, 消费者购买决策仅取决于价格, 不考虑品牌、服务等非价格因素。

(2) 非合作性: 厂商间无任何形式的合作或串通行为。

(3) 价格决策: 厂商独立制定产品价格, 以自身利润最大化为目标。

(4) 完全信息: 厂商知晓所有竞争对手的价格决策、成本结构及市场需求曲线。

在伯川德模型中, 由于产品同质无差异, 消费者会严格选择价格最低的厂商购买。因此, 厂商会陷入价格竞争, 通过持续降价争夺市场份额。当价格降至边际成本水平时, 厂商若进一步降价将导致亏损, 并且无法通过降价获得额外销量(因价格已无优势), 此时市场达到纳什均衡。下面对伯川德模型进行建模求解。

假设厂商 1 生产产品 1, 厂商 2 生产产品 2, 其中, 两个厂商无固定成本, 边际生产成本分别为  $c_1$  和  $c_2$ 。两个厂商销售产品的价格分别为  $P_1$  和  $P_2$ 。 $a_1$  和  $a_2$  分别表示产品 1、产

品 2 的潜在市场规模。用  $b_1$  和  $b_2$  表示产品 1、产品 2 的价格弹性系数,用  $d_1$  和  $d_2$  表示两个产品的替代系数。此时,厂商 1 和厂商 2 的需求函数分别表示为

$$q_1 = q_1(P_1, P_2) = a_1 - b_1 P_1 + d_1 P_2 \quad (1-34)$$

$$q_2 = q_2(P_1, P_2) = a_2 - b_2 P_2 + d_2 P_1 \quad (1-35)$$

在该博弈中,两博弈方的决策变量为产品价格,因此各自的策略空间为  $s_1 = [0, P_{1\max}]$  和  $s_2 = [0, P_{2\max}]$ ,其中,  $P_{1\max}$  和  $P_{2\max}$  是厂商 1 和厂商 2 还能卖出产品的最高价格。两博弈方的收益是各自的利润,即销售收益减去成本,它们都是双方价格的函数:

$$\pi_1 = \pi_1(P_1, P_2) = P_1 q_1 - c_1 q_1 = (P_1 - c_1)(a_1 - b_1 P_1 + d_1 P_2) \quad (1-36)$$

$$\pi_2 = \pi_2(P_1, P_2) = P_2 q_2 - c_2 q_2 = (P_2 - c_2)(a_2 - b_2 P_2 + d_2 P_1) \quad (1-37)$$

用反应函数法分析这个博弈。对上述得益函数求偏导,并且偏导为 0 时存在最大值:

$$\frac{\partial \pi_1}{\partial P_1} = -2b_1 P_1 + c_1 b_1 + a_1 + d_1 P_2 \quad (1-38)$$

$$\frac{\partial \pi_2}{\partial P_2} = -2b_2 P_2 + c_2 b_2 + a_2 + d_2 P_1 \quad (1-39)$$

令  $\frac{\partial \pi_1}{\partial P_1} = 0, \frac{\partial \pi_2}{\partial P_2} = 0$ , 得到两个厂商的反应函数为

$$P_1 = R_1(P_2) = \frac{1}{2b_1}(c_1 b_1 + a_1 + d_1 P_2) \quad (1-40)$$

$$P_2 = R_2(P_1) = \frac{1}{2b_2}(c_2 b_2 + a_2 + d_2 P_1) \quad (1-41)$$

回顾一下在反应函数部分的介绍,该博弈的纳什均衡是两条反应函数对应图像的交点  $(P_1^*, P_2^*)$ , 并且这个交点需要满足:

$$P_1^* = \frac{1}{2b_1}(c_1 b_1 + a_1 + d_1 P_2^*) \quad (1-42)$$

$$P_2^* = \frac{1}{2b_2}(c_2 b_2 + a_2 + d_2 P_1^*) \quad (1-43)$$

解上述二元一次方程组,得

$$P_1^* = \frac{d_1(a_2 + b_2 c_2) + 2b_2(a_1 + c_1 b_1)}{4b_1 b_2 - d_1 d_2} \quad (1-44)$$

$$P_2^* = \frac{d_2(a_1 + c_1 b_1) + 2b_1(a_2 + b_2 c_2)}{4b_1 b_2 - d_1 d_2} \quad (1-45)$$

由上述推导,可以知道  $(P_1^*, P_2^*)$  为该博弈的唯一纳什均衡。将  $P_1^*, P_2^*$  代入收益函数中,可以求得两个厂商的均衡收益,这里就不再赘述了,有兴趣的读者可以自行代入计算。

需要说明的是,伯川德模型作为典型的价格竞争模型,假设寡头厂商生产同质产品,导致市场呈现“低价者赢得全部市场,高价者失去全部市场”的特征。因此,寡头间会通过持续削价争夺市场,直至价格降至各自的边际成本水平。此时进一步降价将导致亏损,价格竞争终止。由此,根据伯川德均衡可以得出两个关键结论,即寡头垄断市场的均衡价格等于边际成本,并且长期经济利润为零。这一结论与“寡头垄断应存在超额利润”的直觉相悖,故被称为伯川德悖论。

显然,若企业生产的产品存在差异(非完全同质),则可通过差异化竞争规避直接价格竞

争。这意味着,破解该悖论的路径之一是引入产品差异化,当产品存在差异时,其替代弹性不再为无限大。此时消费者对不同企业的产品存在偏好差异,价格不再是唯一决策变量,均衡价格也因此高于边际成本。破解伯川德悖论涉及的产品差异形式多样,包括包装、品牌、质量等非空间特征。此外,一种特殊的差异形式是空间差异(spatial differences),对应的经典模型为霍特林模型(Hotelling model)。在霍特林模型中,空间差异体现为企业的地理位置分布差异,如两家店铺可能位于同一直线街道的不同位置(如两端或中间),也可能处于非直线分布的区域。

## 6. 霍特林模型

在霍特林模型中,产品同质,但存在空间位置差异,不同位置的消费者须支付不同的运输成本。因此,消费者的决策须综合考虑两个因素,即产品价格与运输成本(或称旅行成本)。这一逻辑不难理解:假设一条街道的街头与街尾各有一家冰激凌店,并且两家店的产品口味完全相同。此时,你在选择购买时,会同时权衡两家店的价格及往返的时间成本(或路程代价)。若这条街道上均匀分布着住户,那么其他位置的消费者会如何选择呢?

现在把这个问题抽象成一个数学问题。假设有一个长度为 1 的线性城市,消费者均匀地分布在  $[0,1]$  区间内。此时有两个商店出售同质的产品,商品同质表示就商品本身而言,能够带给消费者的效用(utility)是相同的。假设商店 1、2 出售的产品的效用均为  $U$ ,并且效用  $U >$  旅行成本(travel cost) + 价格(price)。每个商店单位产品的成本为  $c$ ,消费者购买商品的旅行成本为  $t$ 。根据空间差异的不同,霍特林模型有如下 3 种情形。

### 1) 情形 1: 两家商店分别位于街道两端

假设两个商店分别位于街道的两端,商店 1 位于最左端,商店 2 位于最右端,即商店 1 在  $x=0$  处,商店 2 在  $x=1$  处,如图 1-3 所示。住在  $x$  处的消费者,如果要去商店 1 消费,要花费  $tx$  的旅行成本;在商店 2 消费,要花费  $t(1-x)$  的旅行成本。假设所有消费者均购买一单位的产品,商店 1 的价格为  $p_1$ ,商店 2 的价格为  $p_2$ 。用  $D_i(p_1, p_2)$ ,  $i=1,2$  表示两个商店的需求函数。在情形 1 下,消费者们从商品上获得的最终效用为“商品效用 - 旅行成本 - 价格”,即

$$U_{\text{final}} = U - \text{travel cost} - \text{price} \quad (1-46)$$



图 1-3 商店位置示意图

因此,消费者选择商店 1 或者商店 2 的临界位置为消费者去两家店最终获得的效用相同,即

$$U_{\text{final}}^1 = U_{\text{final}}^2 \quad (1-47)$$

$$U - tx^* - p_1 = U - t(1 - x^*) - p_2$$

求解上述公式可得

$$x^* = \frac{p_2 - p_1 + t}{2t} \quad (1-48)$$

所以,位于  $x^*$  左边的( $x \in [0, x^*)$ )的消费者都会去商店 1 购买;位于  $x^*$  右边的( $x \in (x^*, 1]$ )的消费者都会去商店 2 购买;位于  $x^*$  处的消费者无论去哪个店最终效用都一样,因此该位置称为无差别消费者位置。图 1-4 所示为无差别消费者所处的位置。



图 1-4 无差别消费者位置图

此时,由于消费者服从 $[0,1]$ 的均匀分布,两个商店的需求函数分别为

$$D_1(p_1, p_2) = \int_0^{x^*} 1 dx = x^* = \frac{p_2 - p_1 + t}{2t} \quad (1-49)$$

$$D_2(p_1, p_2) = \int_{x^*}^1 1 dx = 1 - x^* = \frac{p_1 - p_2 + t}{2t} \quad (1-50)$$

从上述需求函数可以知道:① $D_1(p_1, p_2)$ 关于 $p_1$ 递减,关于 $p_2$ 递增,也就是说当商店1的价格变高时,商店1的需求量就下降;当商店2的价格变高时,商店1的需求量就上升。 $D_2(p_1, p_2)$ 是关于 $p_2$ 递减,关于 $p_1$ 递增的,解释也类似。② $p_1 = p_2$ 时,无差别消费者位置位于中间( $x=1/2$ )处,此时商店1、2平分整个市场份额;当 $p_1 > p_2$ 时,无差别消费者位置位于 $x=1/2$ 的左边;当 $p_1 < p_2$ 时,无差别消费者位置位于 $x=1/2$ 的右边。

此时,两个商店的利润函数分别为

$$\pi_1(p_1, p_2) = (p_1 - c)D_1(p_1, p_2) = \frac{1}{2t}(p_1 - c)(p_2 - p_1 + t) \quad (1-51)$$

$$\pi_2(p_1, p_2) = (p_2 - c)D_2(p_1, p_2) = \frac{1}{2t}(p_2 - c)(p_1 - p_2 + t) \quad (1-52)$$

对上述利润函数求导并令其等于0,可得

$$\frac{\partial \pi_1(p_1, p_2)}{\partial p_1} = \frac{1}{2t}(p_2 - 2p_1 + t + c) = 0 \quad (1-53)$$

$$\frac{\partial \pi_2(p_1, p_2)}{\partial p_2} = \frac{1}{2t}(p_1 - 2p_2 + t + c) = 0 \quad (1-54)$$

求解上述两个方程,解得博弈的均衡价格为

$$p_1^* = p_2^* = c + t \quad (1-55)$$

两家店的均衡利润则表示为

$$\pi_1^* = \pi_2^* = \frac{t}{2} \quad (1-56)$$

需要说明的是,消费者的位置差异在此被转化为产品差异,而这种差异具体体现为消费者的旅行成本,即购买产品的交通代价。旅行成本越高,产品的差异化程度越大,均衡价格随之提高,均衡利润也相应增加。其逻辑在于,随着旅行成本上升,不同商店产品的替代性降低,每个商店对周边消费者的垄断势力增强,商店间的竞争程度随之减弱;同时,消费者的价格敏感度降低,每个商店的最优定价也因此更接近垄断价格水平。

## 2) 情形 2: 两家商店在同一个位置

在情形1中,假定两个商店分别位于街道的两端,事实上,均衡结果对于商店的位置是很敏感的。考虑一个极端的情况,假定两个商店位于同一个位置 $x$ ,此时对于所有位置的消费者,他们到两家店的旅行成本都相同,所以对于任意一个消费者,不管他住哪,他去两家店的旅行成本都相同,也就是两家店的空间差异不存在了(在同一个位置当然空间位置没有差

异)。此时,由于出售的是同质的产品,消费者又变成了只关心价格,该模型的唯一均衡就是伯川德均衡,即

$$\begin{aligned} p_1^* &= p_2^* = c \\ \pi_1^* &= \pi_2^* = 0 \end{aligned} \quad (1-57)$$

当两家商店位于同一位置时,不同位置的消费者之间依然存在旅行成本差异:距离商店越远,旅行成本越高;距离越近,旅行成本越低。但从单个消费者的视角来看,由于两家店位置相同、不存在空间差异,其前往任意一家店的旅行成本是相同的。

### 3) 情形 3: 两家商店可以位于街道任意位置

在此情形下,讨论商店位于任何位置的情况,这是更一般的情况,根据这种情况的结论也可以推导出上述两种情形的结果。假定商店 1 位于  $a \geq 0$  处,商店 2 位于  $1-b$  ( $b \geq 0$ ) 处,如图 1-5 所示。为不失一般性,假定  $1-(a+b) \geq 0$  (即商店 1 位于商店 2 的左边)。如果旅行成本为二次形式,即旅行成本为  $td^2$ ,这里  $d$  是消费者到商店的距离。同样,假设无差别消费者位置为  $x$ , $x$  左边的消费者都在商店 1 消费, $x$  右边的消费者都在商店 2 消费。

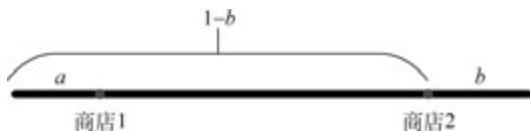


图 1-5 商店位置示意图

类似情形 1 的讨论,同样可以根据最终效用来计算  $x^*$  :

$$U_{\text{final}}^1 = U_{\text{final}}^2 \quad (1-58)$$

$$U - t(x^* - a)^2 - p_1 = U - t[(1-b) - x^*]^2 - p_2 \quad (1-59)$$

求解上述方程,解得

$$x^* = \frac{1-b+a}{2} + \frac{p_2 - p_1}{2t(1-a-b)} \quad (1-60)$$

即  $x < x^*$  的消费者都在商店 1 购买, $x > x^*$  的消费者都在商店 2 购买, $x = x^*$  处为无差别消费者位置。

因此,可以得到需求函数分别为

$$D_1(p_1, p_2) = \int_0^{x^*} 1 dx = x^* = \frac{1-b+a}{2} + \frac{p_2 - p_1}{2t(1-a-b)} \quad (1-61)$$

$$D_2(p_1, p_2) = \int_{x^*}^1 1 dx = 1 - x^* = \frac{1+b-a}{2} + \frac{p_1 - p_2}{2t(1-a-b)} \quad (1-62)$$

化简上述两个需求函数表达式,进一步得到

$$D_1(p_1, p_2) = a + \frac{1-b-a}{2} + \frac{p_2 - p_1}{2t(1-a-b)} \quad (1-63)$$

$$D_2(p_1, p_2) = b + \frac{1-b-a}{2} + \frac{p_1 - p_2}{2t(1-a-b)} \quad (1-64)$$

在上述需求函数中,第 1 项是商店自己的“地盘”(a 是住在商店 1 左边的消费者,b 是住在商店 2 右边的消费者);第 2 项是位于两商店之间的消费者中靠近自己的一半;第 3 项代表

需求对价格差异的敏感度。图 1-6 所示为无差别消费者位置图。

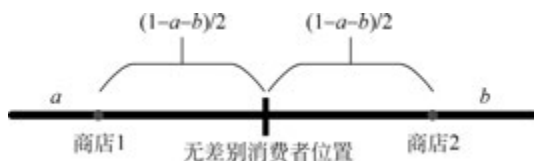


图 1-6 无差别消费者位置图

在此情形下,两个商店的利润函数表示为

$$\pi_1(p_1, p_2) = (p_1 - c)D_1(p_1, p_2) = (p_1 - c) \left[ a + \frac{1-b-a}{2} + \frac{p_2 - p_1}{2t(1-a-b)} \right] \quad (1-65)$$

$$\pi_2(p_1, p_2) = (p_2 - c)D_2(p_1, p_2) = (p_2 - c) \left[ b + \frac{1-b-a}{2} + \frac{p_1 - p_2}{2t(1-a-b)} \right] \quad (1-66)$$

将两个利润函数分别对  $p_1, p_2$  求导并令其等于 0, 联合求解  $\frac{\partial \pi_1(p_1, p_2)}{\partial p_1} = 0$  和  $\frac{\partial \pi_2(p_1, p_2)}{\partial p_2} = 0$  得

$$p_2 - 2p_1 = -t(1-b+a)(1-a-b) - c \quad (1-67)$$

$$p_1 - 2p_2 = -t(1+b-a)(1-a-b) - c \quad (1-68)$$

利用上述两个方程可得

$$p_1^* = c + \frac{1}{3}t(1-a-b)(3+a-b) \quad (1-69)$$

$$p_2^* = c + \frac{1}{3}t(1-a-b)(3+b-a) \quad (1-70)$$

根据这个一般模型的结论,可以看出,当  $a=b=0$  时,此时两个商店分别在街道两端,对应的结果即情形 1 的计算结果。当  $a=1-b$ ,此时两个商店在同一个位置,对应的结果即情形 2 的计算结果。

## 1.4 思考题

1. 供应链的定义是什么?
2. 供应链的结构特征有哪些?
3. ZARA 供应链管理的特点有哪些?
4. 写出下述优化问题的 KKT 条件。

$$\begin{aligned} & \min f(x) \\ & \text{s. t. } g_i(x) \leq 0, \quad i=1,2,\dots,m \\ & \quad h_j(x) = 0, \quad j=1,2,\dots,n \end{aligned}$$

5. 简述斯塔克尔伯格博弈的决策顺序。

# 第 2 章



## 供应链库存模型构建与优化

### 2.1 供应链库存的基本概念

#### 2.1.1 库存的基本概念

库存,译自英语中的“inventory”,是指以支持生产、维护、操作和客户服务为目的而存储的各种物料,包括原材料和在制品、维修件和生产消耗品、成品和备件等。狭义的库存定义为仓库里存放的东西。广义的库存则是指具有经济价值的任何物品的停滞与储藏,是供将来使用的所有闲置资源。一般情况下,设置库存是为了防止物资短缺,就像水库里储水一样,以便保证能向外界源源不断地提供产品。在当今快节奏、高竞争的商业环境中,库存已成为供应链企业运营的核心。有效的库存不仅关系到企业的成本效益,更直接影响到客户满意度和市场响应速度。尽管库存是出于种种经济考虑而存在的,但其也是一种无奈之举。因为库存的存在是由于人们无法预测未来的需求变化,才不得已采用以应付外界变化的手段。

长期以来,库存作为企业生产和销售的物资保障服务环节,在企业的经营中占有重要地位。企业持有一定的库存,有助于保证生产正常、连续、稳定进行,也有助于保质、保量地满足客户需求,维护企业声誉,巩固市场的占有率。在任何一条供应链中,零售商有自己的库存,批发商有自己的库存,供应商也有自己的库存,供应链各个环节都有自己的库存及其控制策略。因此,库存在供应链中以多种形式存在。

(1) 循环库存:这种库存是由企业的订货政策决定的,其规模受订货数量和订货频率的共同影响。它体现了企业为满足日常需求而进行的常规补货活动。

(2) 安全库存:又称缓冲库存,它的存在是为了应对供需过程中的不确定性。安全库存是确保能够持续满足客户需求的关键,同时直接关联到企业服务水平的高低。

(3) 预期库存:是指基于已知需求而提前准备的库存,如针对计划内的需求、新产品上市、促销活动或季节性商品的需求。预期库存的管理以需求为导向,目的是在需求实现时能够迅速响应。

(4) 移动库存:指的是在运输过程中的库存,它处于供应商和客户之间的移动状态,并

且每批货物都能被单独识别和追踪。

(5) 投机库存: 这种库存是出于投资目的而建立的, 常见于企业为未来生产所需而提前购买的原材料。投机库存以供应为导向, 企业通过预判市场变化来决定其规模。

## 2.1.2 供应链库存管理策略

在供应链管理中, 库存往往被视为独立于其他环节的元素, 每个环节各自为政, 采取不同的库存控制策略。这种分散化的做法不可避免地导致需求信息的扭曲, 即所谓的需求放大现象, 也就是供应链中的牛鞭效应, 这增加了供应商的供应和库存风险。

然而, 近年来一些新的供应链库存管理方法应运而生, 如供应商管理库存 (vendor managed inventory, VMI)、客户管理库存 (customer managed inventory, CMI) 和联合库存管理 (jointly managed inventory, JMI) 等。这些策略打破了传统的各自为政的模式, 通过集成化的管理思想, 有效控制了供应链中的库存风险, 适应了市场变化的要求, 代表了库存管理的新趋势。

### 1. 供应商管理库存策略

1987年, 沃尔玛与宝洁公司围绕帮宝适 (Pampers) 婴儿纸尿裤的存货管理进行合作。由于宝洁作为帮宝适产品的制造商和供应商, 与作为零售商的沃尔玛相比, 更了解、也更有经验管理帮宝适, 沃尔玛提出由宝洁向沃尔玛建议进货时间和数量, 经同意后执行。经过一段时间的合作, 又免去了建议手续, 由宝洁在认为需要时直接送货。沃尔玛将存货交给供应商管理, 取得了双赢结果。宝洁提供了额外服务, 但它成了沃尔玛的首选供应商, 在沃尔玛零售店拥有了数量更多、位置更好的货柜; 而且沃尔玛与它共享销售信息, 使宝洁降低了顾客需求的不确定性, 可以更好地管理帮宝适的制造和销售业务, 还可以更好地计划送货, 降低相关成本。沃尔玛则大幅提高了存货管理效率, 降低了各零售点中帮宝适的存货水平, 又减少了缺货现象的发生。这次沃尔玛和宝洁公司实施的 VMI 战略, 明显改善了前者的准时交货和后者的销售, 库存周转率和顾客满意度均得到大幅度提高。随后, 凯马特 (Kmart)、戴尔 (Dell)、惠普 (HP) 等公司也采用 VMI 来降低库存成本。VMI 在欧美国家迅速发展并走向成熟。

传统供应链上的各节点企业对库存管理各自为政, 导致库存需求预测放大及库存费用居高不下。这时候仅仅依靠企业内部决策降低成本已经很困难, 需要从企业间的协同上寻找新的利润点。VMI 通过企业间的合作减少了中间环节, 从而降低了成本, 其实质上就是供货方代替用户管理库存, 库存的职能由供应商负责。

VMI 的主要思想就是实施供应厂商一体化。在这种方式下, 供应链的上游企业不再是被动地按照下游订单发货或补充订货, 而是根据自己对众多下游要货方需求的整体把握, 主动安排一种更合理的发货方式, 既满足下游要货方的需求, 又使自己的库存管理和补充订货策略更合理, 从而使供应链上供需双方的成本降低。这种库存管理策略打破了传统的各自为政的库存管理模式, 体现了供应链的集成化管理思想, 可以适应市场变化的要求, 是一种新的有代表性的库存管理思想。

## 2. 客户管理库存

相对于 VMI, CMI 是另外一种和它相对的库存控制方式, 配送系统中很多人认为, 按照与消费市场的接近程度, 零售商在配送系统中最接近消费者, 在了解消费者的消费习惯方面最有发言权, 因此应该是最核心的一环, 库存自然应归零售商管理。持这种观点的人认为, 配送系统中离消费市场越远的成员就越不能准确地预测消费者需求的变化。

雀巢公司(以下简称雀巢)是世界上最大的食品公司, 总部位于瑞士沃韦市, 目前在全球范围内拥有百余家子公司、500 多家工厂, 员工总数约有 27 万名。其产品行销 80 多个国家。家乐福集团(以下简称家乐福)是世界第二大连锁销售集团, 于 1959 年在法国成立, 全球有 9 061 家门店, 24 万名员工。雀巢从 2009 年 10 月开始, 积极与家乐福合作, 建立 CMI 示范计划的整体运作机制, 总目标是增加商品的供应率、降低家乐福库存天数、缩短订货提前期以及降低双方物流作业成本。具体指标包括雀巢对家乐福物流中心的产品到达率达 90%、家乐福物流中心对零售店面的产品到达率达 95%、家乐福物流中心库存天数下降至预设标准以及家乐福对雀巢的建议订货单修改率下降至 10% 等具体指标。另外, 雀巢希望将新建的模式扩展至其他销售渠道上加以运用, 以加强掌控能力并获得更大规模效益, 家乐福也会与更多的重点供应商进行相关合作。

在信息系统的投入上, 家乐福主要是在电子数据交换(electronic data interchange, EDI)系统建设上的花费。雀巢除 EDI 系统建设外, 还引进了一套 CMI 系统。在系统建设方面, 雀巢与家乐福双方均采用 EDI 网络的形式来进行资料传输, 在雀巢公司的 CMI 系统部分, 则是采取外购产品的方式来建设。雀巢在家乐福、法国及其他国家雀巢的建议下, 在充分考虑系统需求特性后, 最后选用了 Infule 的企业级补货系统(enterprise warehouse and replenishment, EWR)产品。

经过近一年的推进实施, 雀巢和家乐福整个 CMI 运作方式逐渐形成了如下 5 个步骤的运作模式: ①每日 9:30 以前, 家乐福用 EDI 方式将结余库存与出货资料等信息传送给雀巢。②9:30—10:30, 雀巢将收到的资料合并至 EWR 的销售资料库中, 并产生预估的补货需求, 系统将预估的需求量写入后端的 BPCS ERP 系统, 依据实际库存量计算出可行的订货量, 产生建议订单。③10:30 前雀巢以 EDI 方式传送建议订单给家乐福。④10:30—11:00 家乐福在确认订单并进行必要的修改后回传至雀巢公司。⑤11:00—11:30 雀巢依据确认后的订单进行拣货与出货。

除建立一套 CMI 系统外, 双方在具体目标方面也取得显著效果。雀巢对家乐福物流中心的产品到货率由原来的 80% 提升到 95%; 家乐福物流中心对零售店面的产品到货率也由 70% 左右提升为 90% 左右, 而且仍在继续改善中; 库存天数由原来的 25 天左右下降至目标值以下; 订单修改率也由原来的 60%~70% 下降为 10% 以下。

## 3. 联合库存管理

JMI 是介于供应商管理库存和客户管理之间的一种库存管理方式, 顾名思义, 就是由供应商与客户共同管理库存, 进行库存决策。它结合了对产品的制造更为熟悉的生产商或供应商以及掌握消费市场信息、能对消费者消费习惯做出更快更准反应的零售商双方的优点, 因此能更准确地对供应和销售做出判断。在配送系统的上游, 通过销售点提供的信息和零

售商提供的库存状况,供应商能够更加灵敏地掌握消费市场变化,销售点汇总信息使整个系统都能灵活应对市场趋势;在系统另一端,销售点通过整个系统的可视性可以更加准确地控制资金的投入和库存水平。通过在配送系统成员中减少系统库存、增加系统的灵敏度,提升配送响应速度与资源利用效率。由于减少了需求的不确定性和应对突发事件所产生的高成本,整个系统都可以从中获益。在JMI环境下,零售商可以从供应商那里得到最新的商品信息以及库存控制相关参数的指导或建议,但由于是独立的组织,零售商同样需要制定自己的库存决策。

武汉钢铁江北集团公司(以下简称江北公司)作为武钢集团有限公司(以下简称武钢)的全资子公司,是武钢高端钢材深加工基地。其冷弯型钢厂生产的型材处于国内行业领先地位。作为武钢大力发展钢铁相关产业“两化融合”重点工程,武汉钢铁工程技术集团于2015年年底为江北公司启动了办公自动化(office automation, OA)、ERP系统等一系列信息化项目。在项目实施过程中,发现钢材深加工企业与一贯作业钢厂的库存与物流管理存在差异,所以在系统设计上结合JMI模式进行了联合库存与物流管理。

目前钢材深加工行业属于买方市场,尤其是一些大客户一般会比较强势,江北公司冷弯型钢厂作为钢材供应商,会尽量满足这些客户的特殊要求,如成品产出后马上发货给客户,每个月客户按用量跟江北公司结算。之前为江北公司设计的ERP系统采用的是一种比较精细化的管理方式,实现了每一张发票都能与实物发货明细一一对应,可具体到每张订单下的存货数据。精细化成品库存与发货结算管理和客户的特殊需求产生了矛盾,客户要求不按江北公司提供的发货明细资料进行结算,并且如果客户收货称量超出磅差规定范围,客户不会以称量为准,最后只会按品种规格的使用量来结算,于是就产生了如下问题:①结算数据与发货实物数据无法对应;②结算称量与发出称量有差异,每单需要进行差异调整;③存货账务混乱,无法对应实际消耗。

根据JMI的理念,冷弯型钢产品供应链进一步集成。将冷弯型钢厂视为供应链的上游协调管理中心,将冷弯型钢作为原料进行生产,其他产品的客户视为供应链的下游协调管理中心,这两者同时参与,共同制订企业库存计划,共同合理地分配利益、分担风险,建立互利的关系。

江北公司冷弯型钢厂目前已经与部分战略客户从以下4个方面建立了协调管理机制。

- (1) 供需双方本着互利互惠的原则,树立共同的合作目标;
- (2) 建立供需双方JMI的协调管控方法;
- (3) 建立供需双方良好的信息沟通渠道,如收货通知,在库及使用情况信息及时传递与共享;
- (4) 建立供需双方利益分配机制和激励监督机制。

随着江北公司冷弯型厂业务信息量的不断膨胀,靠着落后的手工方式对信息进行管理已经是很难想象的事情了。只有利用先进的信息管理技术,建立适应当前形势的信息系统才能对信息进行有效的管理和使用。

目前江北公司冷弯型厂与其战略客户建立了JMI信息系统,主要用于对库存情况进行计算机实时控制和分析,以实现库存的有效管理,为双方的决策提供依据。基于JMI与物流系统功能架构如图2-1所示。

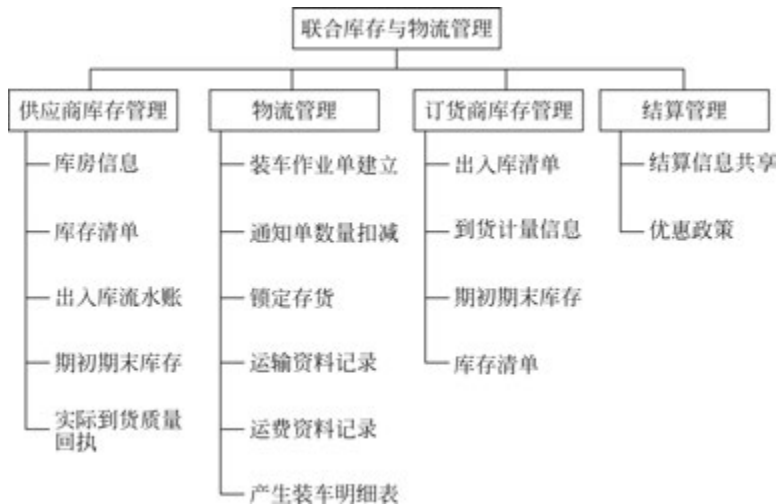


图 2-1 基于 JMI 与物流系统功能架构图

## 2.2 库存的基本模型

在库存理论中,一般根据物品需求的重复程度可分为单周期需求问题和多周期需求问题。单周期需求问题又称一次性订购问题,这种需求具有偶发性且物品生命周期短的特征,因而很少重复订货,如电子产品、时尚女装等。多周期需求问题是指需求在长时间内反复发生,库存需要不断补充,这种需求现象在实际中也比较多见。基于这两类问题建立的模型分别为报童模型、经济订货量(economic order quantity,EOQ)模型以及动态批量模型,本节重点围绕这3类模型进行介绍。关于库存问题的更多优化模型请参考当前一些经典教材与著作如《库存管理》(赵晓波等,2008)与《库存控制理论与方法》(周永务等,2009)。

### 2.2.1 随机需求库存问题的基本模型

#### 1. 经典的报童模型

报童问题,又称单周期随机库存问题,是运营管理中的经典库存决策问题,在优化库存、降低企业成本和提高供应链效率等方面具有重要的意义。在经典的报童问题中,考虑了某一零售商需要订购数量为 $Q$ 的短生命周期产品销售给顾客,用 $c$ 和 $p$ 分别表示产品的订购成本和销售价格。由于顾客的需求 $D$ 是随机的,分别用 $f(x)$ 和 $F(x)$ 来表示随机需求 $D$ 的概率密度函数和分布函数。假设零售商没有容量限制和供应时间延迟,来自顾客的订单能够在订货期的开始瞬间被满足。此时,商品的销售一般发生在订货期的最后。当订货量 $Q$ 大于需求 $D$ 时,零售商处产生了库存,这些产品会以残值价格 $v$ 销售;否则,出现缺货且单位缺货损失成本为 $s$ 。对于零售商来讲,需要确定最优的订货量使得其利润最大。

零售商的利润函数可表示为

$$\Pi(Q) = p \min\{Q, D\} + v(Q - D)^+ - s(D - Q)^+ - cQ \quad (2-1)$$

式中,  $\min\{Q, D\} = Q - (Q - D)^+$ ;  $(D - Q)^+ = D - Q + (Q - D)^+$ 。因此,  $\Pi(Q)$  可以化简为

$$\Pi(Q) = (p + s - c)Q - (p - v + s)(Q - D)^+ - sD \quad (2-2)$$

由于需求  $D$  是随机的, 故零售商的期望利润为

$$E[\Pi(Q)] = (p + s - c)Q - (p - v + s) \int_{Q > D} (Q - x) dF(x) - su \quad (2-3)$$

求解  $\frac{\partial E[\Pi(Q)]}{\partial Q} = 0$ , 可得最佳订货量为

$$F(Q) = \frac{p + s - c}{p + s - v} \quad (2-4)$$

**【例 2-1】** 考虑某企业的订购决策问题。该企业以单位成本  $c=20$  订购某一种季节性产品, 然后将其以价格  $p=100$  销售给顾客。此时, 顾客的需求  $D$  为随机变量, 但已知此随机变量服从均匀分布, 即  $D \sim U(0, 100)$ , 不考虑企业的缺货损失和残值成本。试以该企业期望利润最大化为目标构建模型, 并求得企业的最佳订货量。

**解** 由于  $D$  服从均匀分布  $U(0, 100)$ , 概率密度函数为  $f(D) = \frac{1}{100}$ , 因此, 可以求得  $E[\min(Q, D)] = Q - \frac{1}{200}Q^2$ 。此时, 期望利润函数为  $E[\pi(Q)] = (p - c)Q - \frac{p}{200}Q^2$ 。对  $E[\pi(Q)]$  关于  $Q$  求导并令其等于 0, 得出最佳订货量  $Q^*$  满足  $(p - c) - \frac{p}{100}Q^* = 0$  以及  $Q^* = \frac{100(p - c)}{p}$ 。代入已知参数  $p=100, c=20$ , 得  $Q^* = 80$ , 即企业的最佳订货量为 80 单位。

## 2. 价格影响需求的报童模型

在经典的报童模型中, 假设了销售价格为外生变量。而在实际中, 销售价格是影响市场需求的主要因素。现在考虑一家企业, 该公司具有某一单一产品的库存, 市场需求随机, 并受到价格的影响, 即需求函数与价格相关, 企业的目标是同时确定库存量  $q$  和销售价格  $p$ , 以最大化预期利润。需求中的随机性与价格相关, 相关性可以通过加法或乘法两种方式表示。具体来说, 需求可以定义为加法形式  $D(p, \epsilon) = y(p) + \epsilon$ , 或乘法形式  $D(p, \epsilon) = y(p)\epsilon$ , 其中,  $y(p)$  是一个递减函数, 表示需求与价格之间的依赖关系, 而  $\epsilon$  是一个定义在区间  $[A, B]$  上的随机变量。

在加法形式下令  $y(p) = a - bp$  ( $a > 0, b > 0$ ), 而在乘法形式下令  $y(p) = ap^{-b}$  ( $a > 0, b > 1$ )。这两种形式的  $y(p)$  在经济学文献中都非常常见, 前者表示线性需求曲线, 后者表示等弹性需求曲线。需求曲线的形状是确定性的, 表示市场规模的缩放参数是随机的。为了保证在某些价格范围内需求是正值, 在加法形式下需要添加约束  $A > -a$ , 在乘法形式下需要添加  $A > 0$ 。

然而, 从实际角度来看, 如果  $a$  相对于随机变量  $\epsilon$  的方差足够大, 那么无界概率分布 (如正态分布) 可以对需求函数提供适当的近似。为了表示更为一般的情况, 用  $F(\cdot)$  表示随机变量  $\epsilon$  的累积分布函数 (cumulative distribution function, CDF), 用  $f(\cdot)$  表示概率密度函数 (probability density function, PDF)。同样, 定义  $\mu$  和  $\sigma$  分别表示随机变量  $\epsilon$  的均值和标

准差。

### 1) 加法需求形式

在加法需求形式中,需求函数为  $D(p, \epsilon) = y(p) + \epsilon$ , 其中,  $y(p) = a - bp$ 。在开始时, 企业以单位成本  $c$  采购  $q$  个单位的库存。如果在销售期内的需求不超过  $q$ , 则收入为  $pD(p, \epsilon)$ , 剩余的  $q - D(p, \epsilon)$  单位库存将以单位成本  $h$  处理掉。需要注意的是,  $h$  可能为负值 ( $h \geq -c$ ), 在这种情况下,  $h$  表示每单位的回收价值。

另外, 如果需求超过  $q$ , 则收入为  $pq$ , 并且每个  $D(p, \epsilon) - q$  的短缺部分将以每单位罚金成本  $s$  来计算。因此, 该周期内的利润  $\Pi(q, p)$  是销售收入与总成本之差, 其计算公式如下:

$$\Pi(q, p) = \begin{cases} pD(p, \epsilon) - cq - h[q - D(p, \epsilon)], & D(p, \epsilon) \leq q \\ pq - cq - s[D(p, \epsilon) - q], & D(p, \epsilon) > q \end{cases} \quad (2-5)$$

在上述公式中, 当需求  $D(p, \epsilon)$  小于或等于库存量  $q$  时, 利润通过销售收入减去采购成本和剩余库存的处理成本计算得出; 当需求超过库存量  $q$  时, 利润则通过全部销售收入减去采购成本和短缺罚金计算得出。

通过将  $D(p, \epsilon) = y(p) + \epsilon$  代入之前的利润函数, 并定义  $z = q - y(p)$ , 可以得到一个简便的利润函数表达式, 如下:

$$\Pi(z, p) = \begin{cases} p[y(p) + \epsilon] - c[y(p) + z] - h[z - \epsilon], & \epsilon \leq z \\ p[y(p) + z] - c[y(p) + z] - s[\epsilon - z], & \epsilon > z \end{cases} \quad (2-6)$$

这种变量替换提供了库存决策的一种替代解释, 即如果选择的  $z$  大于随机变量  $\epsilon$  的实际值, 就会产生剩余库存; 而如果选择的  $z$  小于  $\epsilon$  的实际值, 就会发生短缺。因此, 相应的最优库存和定价策略为存储  $q^* = y(p^*) + z^*$  个单位的产品, 并以单位价格  $p^*$  进行销售, 其中  $z^*$  和  $p^*$  使得预期利润最大化。

最终, 预期利润为

$$E[\Pi(z, p)] = \int_A^z \{p[y(p) + u] - h[z - u]\} f(u) du + \int_z^B \{p[y(p) + z] - s[u - z]\} f(u) du - c[y(p) + z] \quad (2-7)$$

定义

$$\Delta(z) = \int_A^z (z - u) f(u) du, \quad \Theta(z) = \int_z^B (u - z) f(u) du$$

因此, 预期利润可以写为

$$E[\Pi(z, p)] = \Psi(p) - L(z, p) \quad (2-8)$$

式中,

$$\begin{aligned} \Psi(p) &\equiv (p - c)[y(p) + \mu] \\ L(z, p) &= (c + h)\Delta(z) + (p + s - c)\Theta(z) \end{aligned}$$

目标函数是最大化预期利润, 即

$$\max_{z, p} E[\Pi(z, p)] \quad (2-9)$$

接下来, 对  $E[\Pi(z, p)]$  关于  $z$  和  $p$  求解一阶和二阶偏导数, 即得

$$\frac{\partial E[\Pi(z, p)]}{\partial z} = -(c + h) + (p + s + h)[1 - F(z)] \quad (2-10)$$

$$\frac{\partial^2 E[\Pi(z, p)]}{\partial z^2} = -(p + s + h)f(z) \quad (2-11)$$

$$\frac{\partial E[\Pi(z, p)]}{\partial p} = 2b(p^0 - p) - \Theta(z) \quad (2-12)$$

$$\frac{\partial^2 E[\Pi(z, p)]}{\partial p^2} = -2b \quad (2-13)$$

其中,  $p^0$  表示最优的无风险价格, 是使得  $\Psi(p)$  最大化的价格, 其定义为

$$p^0 = \frac{a + bc + \mu}{2b} \quad (2-14)$$

上述偏导数用来描述如何通过调整库存量  $z$  和价格  $p$  来最大化预期利润  $E[\Pi(z, p)]$ 。式(2-10)为对  $z$  的一阶偏导, 描述了利润对库存调整的敏感性; 式(2-11)为对  $z$  的二阶偏导, 用来判断  $z$  的选择是否使利润最大化(通常通过检验其为负来确认凹性); 而式(2-12)为对  $p$  的一阶偏导, 用于确定最优价格的方向。

从式(2-11)可以看出, 对于给定的价格  $p$ , 预期利润  $E[\Pi(z, p)]$  关于  $z$  是凹函数。因此, 可以通过先求解最优的  $z$ , 将其作为  $p$  的函数, 再回代至  $E[\Pi(z, p)]$  中, 从而将优化问题转化为关于单变量  $p$  的优化问题。类似地, 从式(2-13)可以看出, 对于给定的  $z$ , 预期利润  $E[\Pi(z, p)]$  关于  $p$  也是凹函数。

由式(2-12)和式(2-13)可得, 对于给定的  $z$ , 最优价格  $p^*$  唯一地由  $z$  的函数确定:

$$p^* = p(z) = p^0 - \frac{\Theta(z)}{2b} \quad (2-15)$$

由上式可以得出, 在加法需求形式下, 单周期的最优库存和定价策略是存储  $q^* = y(p^*) + z^*$  个单位, 并以单位价格  $p^*$  进行销售, 其中,  $p^*$  已经由上式确定, 而  $z^*$  根据以下条件确定。

(1) 如果  $F(\cdot)$  是任意分布函数, 那么可以利用穷举搜索法在区间  $[A, B]$  中确定最优的  $z^*$ 。

(2) 如果  $F(\cdot)$  是一个满足以下条件的分布函数: 在  $A \leq z \leq B$  区间内,  $2r(z)^2 + dr(z)/dz > 0$ , 式中,  $r(z) = f(z)/1 - F(z)$  表示风险率(hazard rate), 则  $z^*$  是在区间  $[A, B]$  中满足  $\frac{dE[\Pi(z, p(z))]}{dz} = 0$  的最大值。

(3) 如果条件(2)成立, 并且  $a - b(c - 2s) + A > 0$ , 那么在区间  $[A, B]$  中  $z^*$  是唯一一个满足  $\frac{dE[\Pi(z, p(z))]}{dz} = 0$  的值。

**【例 2-2】** 某企业生产一种季节性产品, 在销售季开始前须决定采购库存量  $q$  和销售价格  $p$ 。市场需求由加法需求函数表示, 即  $D(p, \epsilon) = y(p) + \epsilon$ 。其中,  $y(p)$  表示市场需求随价格的变化函数,  $y(p) = 100 - 2p$ 。 $\epsilon$  为市场需求的随机波动, 服从均匀分布  $\epsilon \sim U[-10, 10]$ 。企业以单位成本  $c = 5$  采购库存  $q$ 。在销售期结束时, 如果产品产生剩余, 则以单位处理成本  $h - 1$  处理, 当缺货时企业需要额外支付单位惩罚成本  $s = 3$ 。试确定最佳订货量。

**解** 该企业的利润函数为

$$\Pi(z, p) = \begin{cases} p(100 - 2p + \epsilon) - 5(100 - 2p + z) - (z - \epsilon), & \epsilon \leq z \\ p[y(p) + z] - 5[y(p) + z] - 3(\epsilon - z), & \epsilon > z \end{cases}$$

式中,  $z$  为  $\epsilon$  的真实值。此时, 预期利润为

$$E[\Pi(z, p)] = \Psi(p) - L(z, p)$$

其中,

$$\begin{aligned}\Psi(p) &= (p - c)[y(p) + \mu] = (p - 5)y(p) \\ L(z, p) &= \frac{3 \times (z^2 + 20z + 100)}{20} + (p - 2) \frac{z^2 - 20z + 100}{40}\end{aligned}$$

对  $E[\Pi(z, p)]$  关于  $z$  和  $p$  求偏导得

$$\begin{aligned}\frac{\partial E[\Pi(z, p)]}{\partial z} &= -6 + (p + 4) \frac{10 - z}{20} \\ \frac{\partial E[\Pi(z, p)]}{\partial p} &= 4 \times (27.5 - p) - \frac{z^2 - 20z + 100}{40}\end{aligned}$$

联立求得  $27 < p^* < 28$ 。用数值逼近可得  $p \approx 27.7$ , 进而得到  $z^* = 6.21$ 。因此, 最优库存  $q^* = y(p^*) + z^* = 100 - 55.4 + 6.21 = 50.81$ 。

## 2) 乘法需求形式

在乘法需求形式中, 需求函数为  $D(p, \epsilon) = y(p)\epsilon$ , 其中,  $y(p) = ap^{-b}$ 。对于这一部分, 类似的结果也适用于  $y(p) = ae^{-b}$  的情况, 这在经济学文献中是另一种常见的需求表示形式。通过将  $D(p, \epsilon) = y(p)\epsilon$  和  $z = q/y(p)$  代入, 可以方便地将单周期利润函数写为

$$\Pi(z, p) = \begin{cases} py(p)\epsilon - cy(p)z - hy(p)[z - \epsilon], & \epsilon \leq z \\ py(p)z - cy(p)z - sy(p)[\epsilon - z], & \epsilon > z \end{cases} \quad (2-16)$$

在上述表达式中, 当随机变量  $\epsilon$  小于或等于给定的  $z$  时, 利润由需求量乘以价格减去库存成本和处理剩余库存成本计算得出; 当  $\epsilon$  大于  $z$  时, 利润由最大库存量乘以价格减去库存成本和短缺成本计算得出。

对于给定的  $z$ , 最优价格  $p^*$  可以唯一地表示为  $z$  的函数:

$$p^* = p(z) = p^0 + \frac{b}{b-1} \left[ \frac{(c+h)\Lambda(z) + s\Theta(z)}{\mu - \Theta(z)} \right] \quad (2-17)$$

其中,

$$p^0 = \frac{bc}{b-1}$$

在这里,  $p^0$  是一个基础价格, 而其余部分的表达式相当于根据给定的  $z$  来调整最终的最优价格  $p^*$ , 其中,  $\Lambda(z)$  和  $\Theta(z)$  是与需求分布相关的函数。

在乘法需求形式下, 单周期的最优库存和定价策略为: 存储  $q^* = y(p^*)z^*$  个单位, 并以单位价格  $p^*$  进行销售; 其中,  $z^*$  根据以下条件确定。

(1) 如果  $F(\cdot)$  是任意的分布函数, 那么可以通过在区间  $[A, B]$  中对所有  $z$  值进行穷尽搜索来确定最优的  $z^*$ 。

(2) 如果  $F(\cdot)$  是一个满足以下条件的分布函数: 在  $A \leq z \leq B$  的区间内, 并且  $b > 2$ , 有  $2r(z)z^2 + dr(z)/dz > 0$ , 其中  $r(z)$  表示危险率, 则在区间  $[A, B]$  中  $z^*$  是唯一满足  $\frac{dE[\Pi(z, p(z))]}{dz} = 0$  的值。

**【例 2-3】** 继续考虑例 2-2, 当市场需求由乘法需求函数表示时,  $D(p, \epsilon) = y(p)\epsilon$ , 其中,  $y(p) = 200p^{-1.5}$ 。试确定最优的库存量  $q^*$  和价格  $p^*$ 。

解 利用上述具体函数,可以求得

$$\frac{\partial E[\Pi(z, p)]}{\partial z} = y(p)[-6 + (p + 4)(1.5 - z)]$$

$$p^*(z) = 15 + 3 \times \frac{4.5z^2 - 7.5z + 4.125}{-0.5z^2 + 1.5z - 0.125}$$

令  $\frac{\partial E[\Pi(z, p)]}{\partial z} = 0$ , 联立求得  $z^* \approx 1.14 \in [0.5, 1.5]$ 。通过计算可得,  $p^* \approx 12.67$  和  $q^* = y(p^*)z^* \approx 5.06$ 。

## 2.2.2 确定性均匀需求库存问题的基本模型

### 1. EOQ 模型

EOQ 模型产生于 1913 年,该模型研究了如何从经济的角度确定最优库存数量,从根本上改变了人们对库存决策问题的传统认识,是对库存理论研究的一个重要突破。

EOQ 模型的假设条件如下。

- (1) 市场需求率为恒定,即单位时间的需求率  $\lambda$  为常值。
- (2) 补充能力无限大,即无论补充多少货物都能立即实现。
- (3) 补货提前期为零,即发出补货请求后产品立即会得到补充。
- (4) 不允许缺货,即当库存水平降到 0 时,就应进行产品补充。

在以上假设下,EOQ 模型如图 2-2 所示。在订货周期的开始时刻,订货量  $Q$  使得系统的库存水平达到最大,然后库存水平以固定需求率  $\lambda$  减少。当库存水平降到 0 时,下一个订货批量  $Q$  马上补充到位,如此重复循环下去。

假设系统运行无限长的时间,库存量随时间的变化呈现如图 2-2 所示的形态,每次订货批量为  $Q$ ,然后库存量按照需求率为  $\lambda$ ,即斜率为  $-\lambda$  的直线方式下降,当库存量降到 0 时,下一个批量  $Q$  马上补充到位,如此重复循环下去,每个循环周期实际上就是一个订货周期。

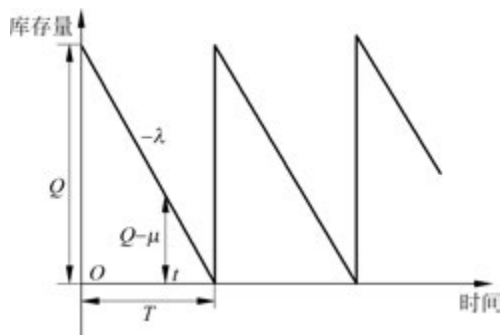


图 2-2 基本 EOQ 的库存量变化

记  $v(t, Q)$  为系统运行到时刻  $t$  时的总成本,定义长期运行下单位时间的总成本为

$$C(Q) = \lim_{t \rightarrow \infty} \frac{v(t, Q)}{t}$$

从理论上可以很容易证得上式即

$$C(Q) = \frac{v(T, Q)}{T} \quad (2-18)$$

式中,  $T$  为一个批量  $Q$  所维持的时间长度, 即订货周期, 如图 2-2 所示。因此, 系统长期运行下单位时间的总成本就可以在图 2-2 中一个三角形的范围内进行分析。在  $[0, T)$  时间区间内, 系统所发生的费用如下。

(1) 订货启动费用:  $K$ 。

(2) 订货可变费用。单位货物的进货价格或生产费用即订货单价为  $c$ , 当订货批量为  $Q$  时, 订货可变费用为  $cQ$ 。

(3) 持货成本。单位货物持有单位时间的成本即持货成本系数为  $h$ , 由图 2-2 可知, 在时刻  $t$  的库存量为  $Q - \lambda t$ , 则在  $[0, T)$  时间区间内的持货成本为

$$\int_0^T h(Q - \lambda t) dt = hQT - \frac{1}{2}h\lambda T^2 \quad (2-19)$$

因为  $T = \frac{Q}{\lambda}$ , 代入上式可得持货成本等于  $\frac{hQ^2}{2\lambda}$ 。

由此, 可得出  $[0, T)$  时间区间内系统的总成本为

$$v(T, Q) = K + cQ + \frac{hQ^2}{2\lambda} \quad (2-20)$$

故系统长期运行下单位时间的总成本为

$$C(Q) = \frac{1}{T} \left( K + cQ + \frac{hQ^2}{2\lambda} \right) = \frac{\lambda}{Q}K + \lambda c + \frac{Q}{2}h \quad (2-21)$$

式中,  $\frac{\lambda}{Q}K$  相当于单位时间上分摊到的订货启动费用  $K$  的量;  $\lambda c$  相当于单位时间花费在订货可变费用上的量;  $\frac{Q}{2}h$  相当于单位时间上的持货成本 ( $\frac{Q}{2}$  正好等于图 2-2 中三角形区域的平均持货量或称为单位时间的持货量)。

因此, 订货批量的决策就变成一元函数的极值问题, 对其求一阶导数有

$$\frac{d}{dQ}C(Q) = -\frac{\lambda}{Q^2}K + \frac{1}{2}h$$

令上式等于零可得

$$Q^* = \sqrt{\frac{2\lambda K}{h}} \quad (2-22)$$

容易验证  $C(Q)$  关于  $Q$  的二阶导数  $\frac{d^2}{dQ^2}C(Q) > 0$ , 因此,  $Q^*$  即为  $C(Q)$  的极小值点且为全

局最小值点,  $Q^* = \sqrt{\frac{2\lambda K}{h}}$  也被称为基本 EOQ。

在基本 EOQ 的运作下, 可得系统长期运行下单位时间的总成本为

$$C(Q^*) = \sqrt{2\lambda Kh} + \lambda c \quad (2-23)$$

此外, 在基本 EOQ 的运作下, 每次订货的时间间隔也被称为基本经济订货周期, 计算公式如下:

$$T^* = \frac{Q^*}{\lambda} = \sqrt{\frac{2K}{\lambda h}} \quad (2-24)$$

**【例 2-4】** 某电子产品零售商,在市场需求恒定的情况下,月需求为 2 000 件,进货单价  $c=100$  元/件,补货启动费用  $K=3\,000$  元/次,假设该零售商要求达到年度回报率为 25%。试求该零售商的基本经济补货批量。

**解** 首先将所有参量的量纲换算到以年度为时间单位时的量,本例中要将每月的需求量换算为每年的需求量,由于 1 年=12 个月,则年度需求量  $\lambda=24\,000$  件/年,持货成本系数为  $h=100 \times 0.25=25$  (元/(件·年))。经过计算,可得基本经济补货批量为  $Q^*=2\,400$  件。基本经济补货周期则为  $T^*=0.1$  年=1.2 月。从而,零售商的年度总成本为  $C(Q^*)=2\,460\,000$  元/年。

## 2. 允许缺货的 EOQ 模型

在实际企业运营中,往往也存在这样的一种情况,即只要不影响到企业信誉,可以允许“缺货”现象存在。允许缺货比不允许缺货时的库存量要小,并且保管费要少。但是,由于缺货要支付一定的缺货损失费,当缺货损失费太大而使缺货不合算时,该模型就退化成本不允许缺货的 EOQ 模型。

对于允许缺货的 EOQ 模型,在 EOQ 模型假设条件下,还须作如下假设: ① 允许缺货,并采取缺货回补的方式; ② 每单位缺货单位时间所造成的惩罚成本(即缺货惩罚成本)为  $s$ 。

如图 2-3 所示,在一个订货周期内,最大缺货量为  $B$ ,最高库存量为  $Q-B$ ,累积库存量对应于时间轴以上的三角形的面积为  $\frac{(Q-B)^2}{2\lambda}$ ,累积缺货量对应于时间轴以下的三角形的面积为  $\frac{B^2}{2\lambda}$ 。

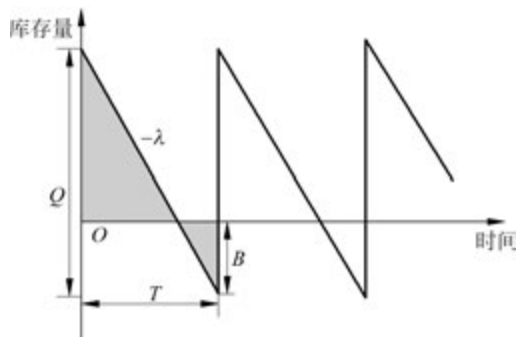


图 2-3 允许缺货的 EOQ 模型的库存量变化

由上可知,在缺货回补的处理方式下,有两个变量需要进行优化: 一是订货批量  $Q$ ,二是最大缺货量  $B$ 。这里最大缺货量  $B$  相当于订货时机,即当缺货量达到  $B$  时立即进货物补充,并且订货批量为  $Q$ 。

与基本 EOQ 中的总成本函数相比,本模型还要增加缺货惩罚成本  $s$ ,则系统长期运行下单位时间的总成本为

$$\begin{aligned}
 C(B, Q) &= \frac{1}{T} \left[ K + cQ + h \frac{(Q-B)^2}{2\lambda} + s \frac{B^2}{2\lambda} \right] \\
 &= \frac{\lambda}{Q} K + \lambda c + h \frac{(Q-B)^2}{2Q} + s \frac{B^2}{2Q}
 \end{aligned} \tag{2-25}$$

上式为二元函数的极值问题,先对  $B$  求偏导数,有

$$\frac{\partial}{\partial B} C(B, Q) = -h \frac{Q-B}{Q} + s \frac{B}{Q}$$

令其等于 0,可得

$$\frac{B}{Q} = \frac{h}{h+s} \tag{2-26}$$

上式说明,在缺货回补经济补货批量中,最大缺货量与补货批量的比值是一个常数。将上述结果代入  $C(Q)$  中,则有

$$C(Q) = \frac{\lambda}{Q} K + \lambda c + \frac{Q}{2} \frac{hs}{h+s} \tag{2-27}$$

类比基本 EOQ 中的总成本函数,允许缺货的 EOQ 模型的补货批量的决策即式(2-27)所示的一元函数的极值问题,对其求一阶导数得

$$\frac{d}{dQ} C(Q) = -\frac{\lambda}{Q^2} K + \frac{1}{2} \frac{hs}{h+s}$$

令其等于 0,得 EOQ 为

$$Q^* = \sqrt{\frac{2\lambda K}{h} \frac{h+s}{s}} \tag{2-28}$$

则对应的经济订货周期为

$$T^* = \sqrt{\frac{2K}{\lambda h} \frac{h+s}{s}} \tag{2-29}$$

进一步可得最优的最大缺货量为

$$B^* = \sqrt{\frac{2\lambda Kh}{(h+s)s}} \tag{2-30}$$

系统长期运行下单位时间的总成本为

$$C(Q^*) = \sqrt{2\lambda Kh \frac{s}{h+s}} + \lambda c \tag{2-31}$$

其中,系统长期运行下单位时间的运作成本为

$$C'(Q^*) = \sqrt{2\lambda Kh \frac{s}{h+s}} \tag{2-32}$$

如果令  $s \rightarrow \infty$ ,意味着不允许缺货,上述所有结果就与基本 EOQ 中的结果一致。

**【例 2-5】** 继续考虑例 2-4,当零售商面临缺货时,需要承担额外的缺货成本,假设缺货惩罚成本 25 元/(件·年)。计算此时的最优订货批量和年度运作成本。

**解** 当考虑惩罚成本  $s = 25$  元/(件·年)时,将相关参数代入式(2-28)可得 EOQ 为  $Q^* = 3\,394$  件。此时,对应的经济订货周期为  $T^* = 1.697$  月。进一步,可以得到最大缺货量为  $B^* = 1\,697$  件。因此,零售商的年度总成本为  $C(Q^*) = 2\,442\,426.4$  元。

### 3. 经济生产批量模型

EOQ 模型假设库存的补货是瞬时到货,但这种假设不符合企业生产的实际情况,显然是一种理想状态。一般来说,在进行某种产品生产时,成品是逐渐生产出来的。也就是说,当生产率大于需求率时,库存是逐渐增加的,不是一瞬间补充的。当库存水平达到一定状态时,就停止生产,然后当前的库存量以固定需求率减少。当库存水平减少到 0 时,下一个生产补货周期开始。这一问题对应的就是经济生产批量(economic production quantity)模型。

现在假设补货能力为  $\mu$ , 并且  $\mu$  为常数,量纲为每单位时间能完成的补充量。当补货批量为  $Q$  时,要经历  $\frac{Q}{\mu}$  时间才最终完成补充过程,而在这段时间,市场需求保持不变,即单位时间的需求量为  $\lambda$ 。因此,要想系统能正常运行,下列条件必须得到满足:

$$\lambda \leq \mu$$

否则,系统将发生缺货,并且最终缺货量将达到无穷大。

在进行货物补充过程中,库存量的增加速率是  $\mu - \lambda$ , 如图 2-4 所示。在补充时间  $\frac{Q}{\mu}$  结束时,库存量达到最高,为  $\frac{Q}{\mu}(\mu - \lambda)$ , 然后库存量按照斜率为  $-\lambda$  的直线下降,当达到 0 时,下一补充循环再开始。显然,一个补货周期是  $T = \frac{Q}{\lambda}$ , 同时,在一个补货周期内系统的累积库存量即三角形的面积  $\frac{1}{2} \frac{Q}{\lambda} \frac{Q}{\mu} (\mu - \lambda)$ 。

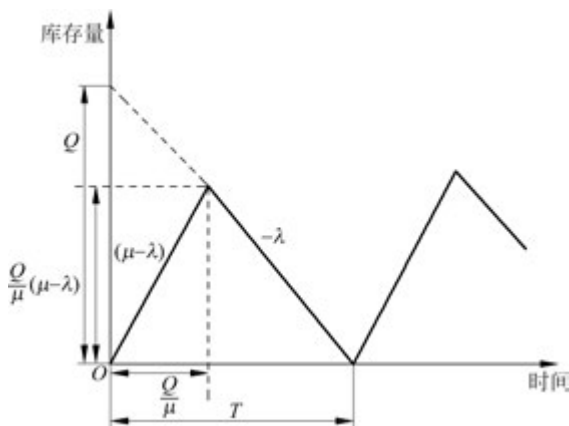


图 2-4 有限补充能力经济补货批量的库存量变化

通过上面的分析,可得系统长期运行下单位时间的总成本为

$$C(Q) = \frac{1}{T} \left[ K + cQ + h \frac{1}{2} \frac{Q}{\lambda} \frac{Q}{\mu} (\mu - \lambda) \right] = \frac{\lambda}{Q} K + \lambda c + \frac{hQ}{2\mu} (\mu - \lambda) \quad (2-33)$$

使上式最小的极值点为

$$Q^* = \sqrt{\frac{2\lambda K}{h \left(1 - \frac{\lambda}{\mu}\right)}} \quad (2-34)$$

上式被称为有限补充能力经济补货批量,对应的经济补货周期为

$$T^* = \frac{\sqrt{2K}}{\sqrt{\lambda h \left(1 - \frac{\lambda}{\mu}\right)}} \quad (2-35)$$

将  $T^*$  代入  $C(Q)$ , 可得系统长期运行下单位时间的总成本为

$$C(Q^*) = \sqrt{2\lambda K h \left(1 - \frac{\lambda}{\mu}\right)} + \lambda c \quad (2-36)$$

其中,系统长期运行下单位时间的运作成本为

$$C'(Q^*) = \sqrt{2\lambda K h \left(1 - \frac{\lambda}{\mu}\right)} \quad (2-37)$$

**【例 2-6】** 继续考虑例 2-4, 在例 2-4 中假设订货量不受限制, 现在假设补货能力有限, 最大补货能力为  $\mu = 30\,000$  (件/年)。试计算此时的最优订货批量。

**解** 当  $\mu = 30\,000$  (件/年) 时, 将  $\mu$  代入 EOQ 公式得  $Q^* = 5\,367$  件。此时, 对应的经济订货周期为  $T^* = 2.68$  月; 该电子零售商的年度总成本为  $C(Q^*) = 2\,426\,832.82$  元。

#### 4. 有数量折扣的 EOQ 模型

在前面几节的讨论中, 假设订货单价  $c$  是固定不变的。而在现实中, 有时会出现订货价格与订货批量相关联的情况, 订货批量越多, 价格就越优惠。那么在价格具有按量折扣的情况下, 如何确定 EOQ 呢? 根据不同的行业、不同的商品, 价格的折扣形式可能会有所不同。比较常见的有两种形式: 一是全量折扣, 即对所订货物执行同一价格, 该价格与订货批量有关; 二是增量折扣, 即将所订货物分成多部分, 每一部分会有不同的价格。

##### 1) 全量价格折扣 EOQ 模型

假设持货成本系数  $h$  主要由资金占用成本  $Ic$  决定, 在全量折扣条件下, 当订货单价为  $c_i$  时, 持货成本系数  $h_i = Ic_i$ , 显然,  $h_1 > h_2 > \dots > h_n$ 。

若用  $c(Q)$  表示订货批量为  $Q$  时的订货单价, 则在全量折扣条件下, 系统长期运行下单位时间的总成本为

$$C(Q) = \frac{\lambda}{Q}K + \lambda c(Q) + \frac{Q}{2}Ic(Q) \quad (2-38)$$

这与基本 EOQ 的区别只在于将单价  $c$  由常数改为随订货批量  $Q$  而变的量。

虽然订货单价  $c$  是随订货批量  $Q$  而变的量, 但其为分段常数的形式, 对应于每个价格, 上式为基本 EOQ 模型。因此, 可写出以下系列的成本方程:

$$\begin{aligned} C_1(Q) &= \frac{\lambda}{Q}K + \lambda c_1 + \frac{Q}{2}Ic_1 \\ C_2(Q) &= \frac{\lambda}{Q}K + \lambda c_2 + \frac{Q}{2}Ic_2 \\ &\vdots \\ C_n(Q) &= \frac{\lambda}{Q}K + \lambda c_n + \frac{Q}{2}Ic_n \end{aligned}$$

将以上方程画在坐标图上, 如图 2-5 所示, 可以发现它们具有以下特点。

(1)  $c_i$  是关于  $i$  有序的, 即  $c_1 > c_2 > \dots > c_n$ , 因此, 对于同一  $Q$  点,  $C_i(Q)$  也是关于  $i$  有

序的,即  $C_1(Q) > C_2(Q) > \dots > C_n(Q)$ 。

(2) 对应于第  $i$  个订货单价  $c_i$ , 其 EOQ 为

$$Q_i^* = \sqrt{\frac{2\lambda K}{Ic_i}} \quad (2-39)$$

因此,  $Q_i^*$  也是关于  $i$  有序的, 即  $Q_1^* < Q_2^* < \dots < Q_n^*$ 。

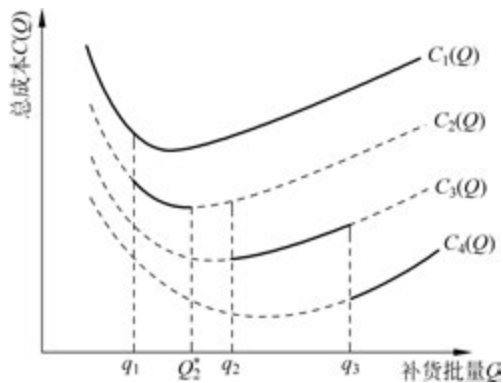


图 2-5 全量折扣的费用函数

以上信息对于分析全量折扣条件下的 EOQ 很有帮助, 下面给出具体的求解步骤。

(1) 计算所有的 EOQ, 即  $Q_1^*, Q_2^*, \dots, Q_n^*$ 。

(2) 找出最大的有效的 EOQ。所谓有效的 EOQ, 是指在所有的  $Q_1^*, Q_2^*, \dots, Q_n^*$  点中, 满足  $q_{i-1} \leq Q_i^* < q_i$  的点, 记其中最大的点为  $Q_j^*$ 。简单地说, 有效的 EOQ 就是起作用的 EOQ。就如图 2-5 所示的情形, 第一个 EOQ 没有处在有效的区间  $[0, q_1)$  内, 第二个 EOQ 处在有效的区间  $[q_1, q_2)$  内, 后面的 EOQ 也没有处在有效的区间内。因此, 最大的有效的 EOQ 是  $Q_2^*$ 。

(3) 计算所有的  $C_{j+1}(Q_j^*), C_{j+2}(q_j), \dots, C_n(q_{n-1})$  的值。在步骤(2)中找出了最大的有效的 EOQ(即  $Q_j^*$ ), 但它不一定是本问题的最优解。如图 2-5 所示, 除  $Q_2^*$  外,  $q_2, q_3$  也为潜在的最优点, 因为它们成本值  $C_3(q_2), C_4(q_3)$  有可能比  $C_j(Q_j^*)$  小。至于比  $Q_j^*$  小的点可以不用再考虑了, 因为在比  $Q_j^*$  小的所有左边的有效区间上, 其成本值一定比  $C_j(Q_j^*)$  大。

(4) 比较  $C_j(Q_j^*), C_{j+1}(q_j), \dots, C_n(q_{n-1})$  的大小, 其中最小的即为最优 EOQ 的单位时间的总成本。

**【例 2-7】** 现有一产品, 该产品的需求率为 500 件/年, 补货启动费用为 40 元/次, 资本年回报率为 20%, 并且不允许缺货。假设补货批量  $Q$  为  $0 \leq Q < 100$  时, 补货单价为 10 元/件; 当  $Q \geq 100$  时, 补货单价可享受 3% 的全量价格折扣。试计算 EOQ。

**解** 当  $0 \leq Q < 100$  时, 进货单价  $c_1 = 10$  元/件; 当  $Q \geq 100$  时,  $c_2 = 9.7$  元/件。因此计算得各个区间的 EOQ 为  $Q_1^* \approx 141$  件和  $Q_2^* \approx 144$  件。此时, 对应的总成本为  $C_1(Q_1^*) = 5282.84$  元和  $C_2(Q_2^*) = 5128.57$  元。由于  $C_2(Q_2^*) < C_1(Q_1^*)$ , 因此最优订货批量为 144 件, 总成本为 5128.57 元。

## 2) 增量价格折扣 EOQ 模型

除全量折扣外, 另一个常见的折扣形式就是增量折扣, 即将所订货物分成多部分, 每一

部分会有不同的价格。举例来说,订货价格与订货批量呈现如下关系。当订货批量处在  $0 \leq Q < q_1$  时,订货单价为  $c_1$ ; 当订货批量处在  $q_1 \leq Q < q_2$  时,  $0 \sim q_1$  部分的订货单价为  $c_1$ , 剩下部分的订货单价为  $c_2$ ; 当订货批量处在  $q_2 \leq Q < q_3$  时,  $0 \sim q_1$  部分的订货单价为  $c_1$ ,  $q_1 \sim q_2$  部分的订货单价为  $c_2$ , 剩下部分的订货单价为  $c_3$ ; 以此类推。假设有  $n$  个价格区间, 并且满足  $c_1 > c_2 > \dots > c_n$ 。

图 2-6 所示为订货可变费用与订货批量之间的关系, 在各个转折点处它们都是连续的。因此, 增量折扣形式可以避免全量折扣形式中的不合理之处。令  $r_1 = q_1, r_2 = q_2 - q_1, \dots, r_{n-1} = q_{n-1} - q_{n-2}$ 。

在进行最优订货批量分析时, 需要确定订货单价。在增量折扣的形式下, 不同部分货物的订货单价不一样, 订货批量与“等价”的订货单价之间的关系如图 2-6 所示。当订货批量处在  $0 \leq Q < q_1$  时, 订货可变费用为  $c_1 Q$ , 等价的订货单价为  $\frac{c_1 Q}{Q} = c_1$ 。当订货批量处在  $q_1 \leq Q < q_2$  时, 订货可变费用为  $c_1 r_1 + c_2 (Q - r_1)$ , 则等价的订货单价为  $\frac{c_1 r_1 + c_2 (Q - r_1)}{Q}$ ; 当订货批量处在  $q_2 \leq Q < q_3$  时, 订货可变费用为  $c_1 r_1 + c_2 r_2 + c_3 (Q - r_1 - r_2)$ , 则等价的订货单价为  $\frac{c_1 r_1 + c_2 r_2 + c_3 (Q - r_1 - r_2)}{Q}$ 。

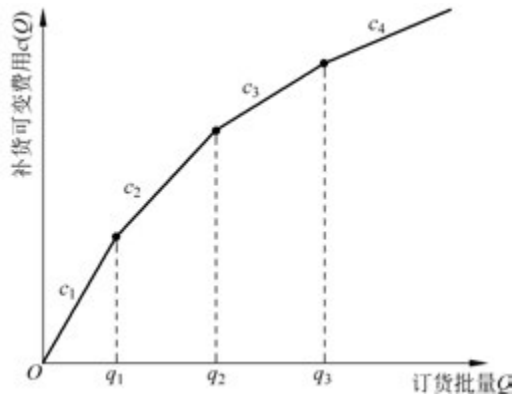


图 2-6 增量折扣的订货可变费用与订货批量的关系

若用  $c(Q)$  表示订货批量为  $Q$  时的等价订货单价, 一般地, 可将上述关系表示为以下形式:

$$c(Q) = \begin{cases} c_1, & 0 \leq Q < q_1 \\ \frac{c_1 r_1 + c_2 (Q - r_1)}{Q}, & q_1 \leq Q < q_2 \\ \frac{\sum_{i=1}^{n-1} c_i r_i + c_n (Q - \sum_{i=1}^{n-1} r_i)}{Q}, & q_n \leq Q \end{cases} \quad (2-40)$$

在增量折扣条件下, 系统长期运行下单位时间的总成本为

$$C(Q) = \frac{\lambda}{Q} K + \lambda c(Q) + \frac{Q}{2} I c(Q) \quad (2-41)$$

对应于第  $j$  个价格区间, 上式改写为

$$C_j(Q) = \frac{\lambda}{Q}K + \lambda c_j(Q) + \frac{Q}{2}Ic_j(Q) \quad (2-42)$$

将订货单价  $c_j(Q) = \frac{\sum_{i=1}^{j-1} c_i r_i + c_j(Q - \sum_{i=1}^{j-1} r_i)}{Q}$  代入上式中, 经整理后可得

$$C_j(Q) = \frac{\lambda \left[ K + \sum_{i=1}^{j-1} (c_i - c_j) r_i \right]}{Q} + \lambda \left[ c_j + \frac{I \left( \sum_{i=1}^{j-1} (c_i - c_j) r_i \right)}{2\lambda} \right] + \frac{Q}{2} Ic_j \quad (2-43)$$

上式  $C(Q)$  与基本 EOQ 模型具有相同的形式。因此, 最小化的极值点为

$$Q_j^* = \sqrt{\frac{2\lambda \left( K + \sum_{i=1}^{j-1} (c_i - c_j) r_i \right)}{Ic_j}} \quad (2-44)$$

由于  $c_j$  是关于  $j$  有序的, 即  $c_1 > c_2 > \dots > c_n$ , 因此上式表明  $Q_j^*$  也是关于  $j$  有序的, 即  $Q_1^* < Q_2^* < \dots < Q_n^*$ 。但对于同一  $Q$  点, 虽然第 1 项和第 2 项是关于  $j$  增加的, 但第 3 项是关于  $j$  减少的, 因此,  $C_j(Q)$  不一定是关于  $j$  有序的。

图 2-7 所示为在增量折扣形式下成本与订货批量之间的一般关系。除上述性质外, 成本函数还有一个重要性质, 就是在每个价格分段点处, 成本函数是连续的, 即下述关系成立:

$$C_j(q_j) = C_{j+1}(q_j) \quad (2-45)$$

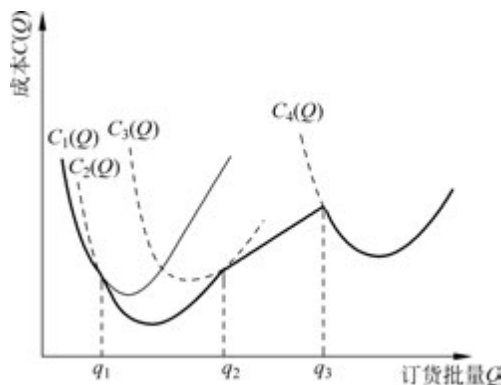


图 2-7 增量折扣的费用函数

以上性质表明, 在  $n$  个极小值点  $Q_1^*, Q_2^*, \dots, Q_n^*$  中, 至少有一个是落入有效区间内的, 而不包含极小值点的有效区间, 肯定不存在全局最小点。如图 2-7 所示, 曲线  $C_1(Q)$  和  $C_5(Q)$  的有效区间未包含它们各自的极小值点  $Q_1^*$  和  $Q_3^*$ , 全局最小点就肯定不在这两个区间内, 而曲线  $C_2(Q)$  和  $C_4(Q)$  的有效区间包含了它们各自的极小值点  $Q_2^*$  和  $Q_4^*$ , 则全局最小点就一定是在它们中的一个或多个。由上面的分析可知, 增量折扣条件下的 EOQ 可通过以下步骤求得。

(1) 根据式  $Q_j^*$ , 计算所有的极小值点  $Q_1^*, Q_2^*, \dots, Q_n^*$ 。

(2) 根据价格所对应的区间  $[0, q_1), [q_1, q_2), \dots, [q_{n-1}, \infty)$ , 判断有效的极小值点。

(3) 在所有有效的极小值点中, 成本最小的点就是全局最小点, 该点即为增量折扣条件下的 EOQ。

**【例 2-8】** 继续考虑例 2-7, 现假设订货单价和订货批量有以下关系, 假设补货批量  $Q$  为  $0 \leq Q < 100$  时, 补货单价为 10 元/件; 当  $Q \geq 100$  时, 补货单价可享受 3% 的增量价格折扣。试计算 EOQ。

**解** 当  $0 \leq Q < 100$  时, 进货单价  $c_1 = 10$  元/件; 当  $Q \geq 100$  时,  $Q \geq 100$  的部分, 进货单价  $c_2 = 9.7$  元/件。由题意得  $r_1 = 100$ , 因此计算各个区间的 EOQ 为  $Q_1^* \approx 141$  件和  $Q_2^* \approx 190$  件。此时, 计算总成本得  $C_1(Q_1^*) = 5\,282.84$  元和  $C_2(Q_2^*) = 5\,221.51$  元。由于  $C_2(Q_2^*) < C_1(Q_1^*)$ , 因此最优订货批量为 190 件, 总成本为 5 221.51 元。

### 2.2.3 确定性离散需求库存问题的基本模型

以上讨论的都是库存管理中均匀需求的订货批量问题。然而, 在现实生产中遇到的更多是离散需求问题, 即需求率随时间的变化而变化。如表 2-1 所示的例子, 全年 12 个月的需求都是已知的, 但每个月的需求量不同, 在这种情况下, 如何确定使得库存总费用最低的订货批量, 则是解决确定性离散需求库存问题的关键所在。由于需求率随时间离散变化, 所以不能简单地像处理 EOQ 模型那样, 按平均成本和库存量来解决离散需求的订货批量问题。

表 2-1 某产品月需求量

月份	1	2	3	4	5	6	7	8	9	10	11	12
需求量/箱	100	52	48	92	128	80	90	110	120	130	80	120

针对上述问题, W-W 法是一种具有代表性的求出最优订货策略的方法。W-W 法的出发点是在总费用最低的情况下, 确定各个时期的订货批量。如表 2-1 所示, 管理者可以分月订货, 每次只订本月需要量, 则全年要订货 12 次。W-W 法假设当月消耗掉的产品不计算库存保管费, 所以只发生订货费用; 管理者也可以一次订购满足数月需求的物品, 这样全年订货次数少了, 订货费用就降低了, 但当月未消耗完的产品又增加了库存保管费。W-W 法就是在两者之间寻求平衡, 找出最优订货量。相关符号及含义如下。

$(r_1, r_2, \dots, r_n)$  为  $n$  个时段的需求量。

$(y_1, y_2, \dots, y_n)$  为  $n$  个时段的订货量。

$x_t = \sum (y_j - r_j)$  为时段  $t$  上的期末库存量。

$h(x_t)$  为时段  $t$  上的库存保管费用。

$c(y_t)$  为时段  $t$  上的订货费用。

$F(t)$  为时段  $t$  上的总费用。

目标函数为

$$\min \sum_{t=1}^n [c(y_t) + h(x_t)] \quad (2-46)$$

$$\text{s. t. } \begin{cases} x_t = \sum_{j=1}^t (y_j - r_j) \\ x_t \geq 0 \\ x_0 = 0, t = 1, 2, \dots, n \end{cases} \quad (2-47)$$

这里不加证明地直接引用 W-W 法的最优订货策略。当满足下列条件时,在时段  $t$  发生一次批量为  $y_t$  的订货,即

$$y_t \cdot x_{t-1} = 0 \quad (t=1,2,\dots,n) \quad (2-48)$$

因为  $y_t$  不为 0,否则没有订货量,这就意味着只有  $x_{t-1}=0$  且  $y_t>0$ ,即当时段  $t$  的期初库存为零时,才在第  $t$  期发生订货。

**【例 2-9】** 一公司全年 12 个月对某产品的需求如表 2-2 所示。已知订货费用  $S=54$  元/次,  $I=0.02$  元/月,该产品的单位成本  $C=20$  元/箱。试给出经济订购批量策略。

表 2-2 某公司产品需求量

月份	1	2	3	4	5	6	7	8	9	10	11	12	合计
需求量/箱	10	62	12	130	154	129	88	52	124	160	238	41	1200

**解** 设  $F(t)$  表示时段  $t$  上的总费用;期初库存等于 0;本期消耗掉的库存不计算库存保管费用。假定 1 月份的期初库存等于 0,本月需要 10 箱的货物,首先在月初订购 10 箱产品,则 1 月份的总费用  $F(1)$  只与订货费用有关,即  $F(1)=0+S=54$  元。

满足 2 月份需求的方案有两种:方案 1,1 月初订 10 箱,1 月份用完;2 月初订 62 箱,2 月份用完。方案 2,1 月初订 72 箱(1 月份的 10 箱和 2 月份的 62 箱)。此时,对应两种方案的总费用分别如下:

方案 1 的总费用:  $C_1 = F(1) + S = 54 + 54 = 108$  元

方案 2 的总费用:  $C_2 = 54 + 62 \times 0.02 \times 20 \times 1 = 78.60$  元

因为  $C_2 < C_1$ ,最佳选择应是方案 2,即一次订购满足 1—2 月的需求量,此时  $F(2) = 78.60$  元。

满足 3 月份需求的方案有 3 种:方案 1,2 月底 3 月初订 12 箱,3 月份用完;方案 2,2 月初订 74 箱(2 月份的 62 箱和 3 月份的 12 箱);方案 3,1 月初订 84 箱(1 月份的 10 箱、2 月份的 62 箱、3 月份的 12 箱)。此时,对应 3 种方案的总费用分别如下:

方案 1 的总费用:  $C_1 = F(2) + S = 78.60 + 54 = 132.6$  元

方案 2 的总费用:  $C_2 = 54 + 54 + (12 \times 0.02 \times 20 \times 1) = 112.80$  元

方案 3 的总费用:  $C_3 = 54 + (62 \times 0.02 \times 20 \times 1) + (12 \times 0.02 \times 20 \times 2) = 88.40$  元

因为  $C_3 < C_2 < C_1$ ,最佳选择应该是方案 3,即一次订购满足 1—3 月的需求量,此时的  $F(3) = 88.40$  元。如此下去,直到 12 个月全部计算完毕,如表 2-3 所示,最终得到最优的订货批量。

表 2-3 12 个月计算结果

单位:箱

月份	1	2	3	4	5	6	7	8	9	10	11	12	合计
期初库存	0	74	12	0	0	129	0	52	0	0	0	41	308
进货量	84	—	—	130	283	—	140	—	124	160	279	—	1200
需求量	10	62	12	130	154	129	88	52	124	160	238	41	1200
期末库存	74	12	0	0	129	0	52	0	0	0	41	0	308

## 2.3 易损耗产品的库存优化模型

### 2.3.1 生鲜品的报童模型

生鲜品的生产具有较强的地域特征,因此往往需要通过长途运输销售到远方市场。生鲜品容易发生变质的特性使得它们在远程运输途中的保鲜成为一个很大的难题。肖勇波等(2008)是国内较早研究随机需求下生鲜品报童问题的学者。本节的模型分析主要借鉴他们的结论。具体问题描述如下:在销售周期初期,某一生产商生产一种高度易变质的时鲜产品,产品的单位生产成本记为 $c_m$ ,数量记为 $q$ 。产品因为其自身的特性,具有一定的保鲜期(记为 $\tau$ ),时间超过保鲜期后产品将开始变质。生产完毕后,生产商借助某种运输工具将产品运送到远方批发市场,运输的费用和风险都由生产商自行承担,记单位产品运输费用为 $c_t$ 。为不失一般性,不考虑运输过程中可能存在的固定费用。因为天气、路况、设备故障、调度调整等方面的原因,产品运送到目的市场的时间 $T$ 是不确定的,但是根据以往经验,其分布是已知的。为便于处理,假设 $T$ 在区间 $[a, b]$ 上连续分布( $a < \tau < b$ ),其累积分布函数和概率密度函数分别为 $G(t)$ 和 $g(t)$ 。运输过程会给产品带来以下两方面的影响。

(1)“量”的影响:运输过程中,难免有部分产品会发生损坏,而且运输经历的时间越长,损坏的产品将越多。故定义一个与运输时间相关的损坏因子 $\sigma(t) \in [0, 1]$ ,即运输时间为 $t$ 时,将有 $q \times \sigma(t)$ 数量的产品损坏,因此产品的有效供给为 $q[1 - \sigma(t)] = q \cdot m(t)$ 。

(2)“质”的影响:运输经历的时间越长,产品也变得越不新鲜。定义 $\theta(t)$ 为新鲜度因子 $[0 \leq \theta(t) \leq 1]$ ,它是运输时间的分段连续减函数,产品新鲜度将体现在最终零售市场的需求上。在同样的零售价格下,产品新鲜度越高,则市场需求越大。

到达批发市场后,生产商根据产品的状态(剩余的有效供给量和产品的新鲜度)确定一个合适的零售价格 $p$ 在市场上进行销售。这里,为了简化模型,假设剩余产品的残值为零,也不再考虑缺货。此时,制造商需要确定最优的生产量和销售价格使得期望利润最大。

作为一类高度易变质的产品,生鲜品具有自身的特殊性,即其变质不仅会导致产品有效供给的降低(数量损耗),而且会导致产品新鲜程度的降低,从而直接影响到市场的需求。因此,对于该类产品的市场需求,采用乘积形式的市场需求函数(即市场需求对价格的弹性是常数):

$$D(p, t) = y_0 p^{-k} \theta(t) \cdot \epsilon, \quad k > 1 \quad (2-49)$$

式中, $\epsilon$ 为一个均值为1的连续分布的随机因子,其概率密度函数和累积分布函数分别为 $f(x)$ 和 $F(x)$ ;  $y_0$ 是对市场规模的一个度量; $k$ 为需求对零售价格的弹性指数,重点考察需求对价格敏感的情形,因此假定 $k > 1$ 。通过考虑产品新鲜度 $\theta(t)$ 对需求的影响,上述需求函数实际上是对传统乘积形式需求的扩展。

对于上述问题,在集权情形下,当初始发货量为 $q_c$ 、运输时间为 $t$ 时,最优零售定价为 $p_c^* = \left( \frac{z_0 y_0 \theta(t)}{q_c m(t)} \right)^{1/k}$ 。于是,总期望利润关于初始发货量 $q_c$ 的函数为

$$\begin{aligned} \Pi_c(q_c) &= E_t \{ p_c^*, E_\epsilon \{ \min[q_c m(t), D(p_c^*, t)] \} \} - q_c(c_c + c_m) \\ &= \frac{k}{k-1} K_0 (z_0 y_0)^{\frac{1}{k}} [1 - F(z_0)] q_c^{1-\frac{1}{k}} - q_c(c_t + c_m) \end{aligned} \quad (2-50)$$

则集权情形下的最优发货量决策为

$$q_c^* = z_0 y_0 \left[ K_0 \times \frac{1 - F(z_0)}{c_m + c_t} \right]^k$$

由  $\Pi_c(q_c)$  的表达式可得,  $\Pi_c(q_c)$  对  $q_c$  的二阶导数为负, 因此  $\Pi_c(q_c)$  是关于  $q_c$  的严格凹函数。于是, 规划问题  $\max[\Pi_c(q_c)]$  有且仅有一个最大值点, 其极值点  $q_c^*$  由一阶条件取得。令

$$\frac{d\Pi_c(q_c)}{dq_c} = k_0 (z_0 y_0)^{\frac{1}{k}} [1 - F(z_0)] q_c^{-\frac{1}{k}} - (c_t + c_m) = 0$$

将  $q_c^*$  代入  $\Pi_c(q_c)$  的表达式, 即得供应链的最大整体期望利润为

$$\Pi_c^* = \Pi_c(q_c^*) = \frac{c_m + c_t}{k - 1} z_0 y_0 \left[ K_0 \times \frac{1 - F(z_0)}{c_m + c_t} \right]^k \quad (2-51)$$

**【例 2-10】** 假设某生鲜品生产商种植一种高度易变质的时鲜水果, 准备通过长途运输将其销售到远方的批发市场。单位生产成本  $c_m = 10$  元, 单位运输费用  $c_t = 5$  元。由于水果易腐, 产品的保质期  $\tau = 3$  天, 运输时间  $T$  受天气、路况等因素影响, 在区间  $[2, 4]$  天内均匀分布, 即  $T \sim U(2, 4)$ 。运输过程中, 水果会因时间延长而发生损耗和新鲜度下降。运输过程中的损坏函数为  $\sigma(t) = 0.1(t - 2)$ , 即运输时间每超过两天, 损耗比例增加 10%。新鲜度函数为  $\theta(t) = 1 - 0.2(t - 2)$ , 即运输时间每超过两天, 新鲜度下降 20%, 并且  $0 \leq \theta(t) \leq 1$ 。此外, 市场需求受价格影响, 需求函数为  $D(p, t) = 1000 p^{-2} \theta(t) \epsilon$ , 式中,  $\epsilon$  为一个均值为 1 的随机因子, 服从均匀分布  $\epsilon \sim U(0, 2)$ 。假设剩余产品无残值, 不考虑缺货成本。试求最优初始发货量  $q_c^*$  和最大整体期望利润  $\Pi_c^*$ 。

**解** 将上述参数代入理论分析结果, 可以得出, 当初始生产量为  $q_c$ , 运输时间为  $t$  时, 最优零售价格为  $p_c^* = \left[ \frac{z_0 1000 \theta(t)}{q_c m(t)} \right]^{1/2}$ , 式中,  $z_0$  为随机因子  $\epsilon$  的某个临界值。此时, 总期望利润函数为  $\Pi_c(q_c) = \frac{k}{k-1} K_0 (z_0 y_0)^{\frac{1}{k}} [1 - F(z_0)] q_c^{1-\frac{1}{k}} - q_c (c_t + c_m)$ , 式中,  $K_0 \approx 7.18$ 。

利用  $z_0 = F^{-1} \left[ 1 - \frac{c_m + c_t}{p_c^2 m(t)} \right]$ , 通过迭代计算, 在均值点  $z_0 = 1$  的条件下可求得  $F(1) = 0.5$ 。将其代入期望函数可得  $\Pi_c(q_c) = 227.03 q_c^{1/2} - 15 q_c$ 。利用极值条件可求得  $q_c^* = (7.57)^2$ , 从而最大期望利润为  $\Pi_c(57.3) = 859.32$ 。

### 2.3.2 易腐品的 EOQ 模型

经典 EOQ 模型假定了物品在存储过程中不发生变质现象, 因而没有考虑变质对库存系统订货策略的影响。然而物品在储存过程中发生变质却是库存管理实践中常见的现象, 如食品、水果、蔬菜的变质, 挥发性物品的挥发, 放射性物质的衰变等, 这就是说经典的 EOQ 模型无法适用于易腐品的存储问题。下面分别来讨论易腐品在不允许缺货和允许缺货时的 EOQ 模型。

#### 1. 不允许缺货的易腐品的 EOQ 模型

为建立模型的需要, 假定该物品的变质率为常数  $\alpha$ , 即单位货物每单位时间腐烂  $\alpha$  量。

单位时间需求量为  $D$ , 物品的单位进价为  $c$ , 单位物品单位时间的库存保管费为  $h$ 。每次订货的固定费用订货费为  $A$ , 相邻两次订货时间即订货周期的长度为  $T$ , 如果在  $t$  时刻的库存量为  $I(t)$ , 则单位时间腐烂的货物量为  $\alpha I(t)$ 。由于货物腐烂的量与库存量有关, 可建立以下微分方程:

$$\frac{d}{dt}I(t) = -\alpha I(t) - D \quad (2-52)$$

式中, 左端表示在  $t$  时刻库存量为  $I(t)$  时所产生的增量变化(实际上是减少); 右端表示在  $t$  时刻引起库存量的变化有两个因素: ①货物的腐烂; ②货物被市场的需求消耗。

假设  $t=0$  时刻正好是货物补充的时刻, 订货批量为  $Q$ , 利用这一边界条件, 可得到微分方程的解是

$$I(t) = \left(Q + \frac{D}{\alpha}\right) e^{-\alpha t} - \frac{D}{\alpha} \quad (2-53)$$

当订货批量为  $Q$  时, 对应的订货周期为  $T$ , 则在  $T$  时刻库存量下降到零, 同时马上补充一个批量  $Q$ , 库存状态的变化如图 2-8 所示。在  $T$  时刻库存量为零, 即  $I(T)=0$ , 代入式(2-53)中, 可得

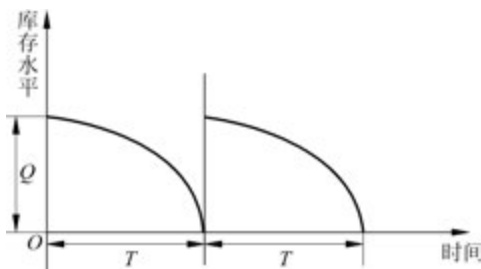


图 2-8 连续腐烂货物的库存状态变化

$$Q = \frac{D}{\alpha}(e^{\alpha T} - 1) \quad (2-54)$$

系统的平均库存与一个订货周期内的平均库存相等, 而一个订货周期内的平均库存是图 2-8 中  $I(t)$  曲线下方的积分面积除以订货周期的长度, 即

$$\begin{aligned} \bar{I} &= \frac{1}{T} \int_0^T I(t) dt = \frac{1}{T} \int_0^T \left[ \left(Q + \frac{D}{\alpha}\right) e^{-\alpha t} - \frac{D}{\alpha} \right] dt \\ &= \frac{D}{\alpha^2 T} (e^{\alpha T} - \alpha T - 1) \end{aligned} \quad (2-55)$$

通过上面的分析, 系统长期运行下单位时间的总成本为

$$\begin{aligned} \text{ATC}(T) &= \frac{1}{T}(A + Qc) + \bar{I}h \\ &= \frac{A}{T} + \frac{D}{\alpha T}(e^{\alpha T} - 1)c + \frac{D}{\alpha^2 T}(e^{\alpha T} - \alpha T - 1)h \end{aligned} \quad (2-56)$$

将上式对  $T$  求一阶导数并令其等于零, 有

$$(\alpha T - 1)e^{\alpha T} = \frac{\alpha^2 K}{D(h + \alpha c)} - 1 \quad (2-57)$$