随机服务系统的理论与实务

周玮民 著

科学出版社

随机服务系统的理论与实务

周玮民 著

科学出版社

北京

内 容 简 介

在人类活动中，服务系统的理论有着广泛的应用. 本书所讨论的是: 因顾客对服务要求的随机性而引发的系统行为与绩效的变化. 对系统而言, 顾客人数变化是一个生灭过程的结果. 由顾客的观点来看, 他经历的却是一个等待过程. 因此系统的行为与绩效主要是以系统中顾客拥挤的程度以及他们花费在系统上的时间来表示.

利用随机过程的数学模型对系统进行分析时, 顾客不同的服务要求与系统操作的规则就成为模型的假设条件. 为此, 作者对多年来从事的许多实际案例进行整理与资料分析, 以验证假设条件的合理性. 书中提供的实例包括: 公路交通、紧急救援、计算机、网络通信、生产、库存、维修、搬运系统、在制品存放空间设置以及生产线规划等.

随机过程作为数学的一支, 有些读者可能对它较为生疏. 因此本书尽量避免繁琐的数学推演, 而代之以直观方式来阐述概念. 学过微积分与概率论的读者即可阅读本书, 解读其中的论述, 并获取相应的资料.

本书的最终目的是希望读者能够通过研读, 提高解决实际问题的能力. 所以在面对难解的数学模型时, 也提倡近似解法, 以求在合理的时间内得到合理的解. 对于一些非基础的但是有价值的材料, 则以提纲的方式编入书末的练习与讨论一章中. 读者可以此作为练习材料, 并从中获得更多的知识.

图书在版编目(CIP)数据

随机服务系统的理论与实务/周玮民著. —北京：科学出版社, 2016.10

ISBN 978-7-03-050027-4

Ⅰ. ①随… Ⅱ. ①周… Ⅲ. ①排队论—研究 Ⅳ. ①O226

中国版本图书馆 CIP 数据核字(2016) 第 232493 号

责任编辑: 李 欣 赵彦超 / 责任校对: 彭 涛
责任印制: 张 伟 / 封面设计: 陈 敬

科学出版社 出版
北京东黄城根北街 16 号
邮政编码: 100717
http://www.sciencep.com

北京教图印刷有限公司 印刷
科学出版社发行 各地新华书店经销
*
2016 年 10 月第 一 版 开本: 720 × 1000 B5
2016 年 10 月第一次印刷 印张: 11 1/4
字数: 213 000

定价: 68.00 元

(如有印装质量问题, 我社负责调换)

前　言

1981 年 7 月我受当时的中国企业管理学会邀请, 在清华大学讲授运筹学的组合最优化、随机过程与排队论. 次年 7 月我又接受联合国的资助, 由机械工业部安排在陕西机械学院专讲排队论. 两次经过上海时, 我都见了上海科学研究所的朋友, 其中刘吉先生鼓励我写一本这方面的专著, 后来又耐心地帮着找出版社. 先后经历了四年, 书才有机会出版. 当时的中文书稿都是手写, 熟悉英文者又不多, 出版前也未作校正, 书中谬误较多. 另一方面, 我自己也不够成熟, 认识浅薄, 从学校出来就一直在 IBM 公司的研究部门工作, 对服务系统的实务工作也仅限于大型计算机与网络的绩效分析与优化, 书中提供的材料多半缺少自己的见解. 1982 年年底, 由于个人志趣而转向实务工作, 故请调生产部门, 从事物流改进与生产整合, 以后又走上生产策略、运营管理, 乃至企业过程、企业文化与组织发展的道路. 其间我搜集并分析过许多资料, 并从解决实际问题的过程中, 逐渐增长了见识. 离开企业界后, 我去了台湾任教, 又三次重新讲授这门课, 并添加了实务上的案例. 离校返回美国后, 因为校方一时没有续教的老师, 于是又在隔年夏天把这门课改为“随机服务系统”, 尽量去除较深的数学部分, 并加入更多的实际资料分析与实例. 对学生的成绩考核, 则是以他们专题调研报告 (诸如: 邮局服务、医院挂号系统、校园小吃部座位安排、校园网络登录系统、模拟法与等候线理论比较等课题) 成绩为准. 本书就在最后这本教材的基础上发展而来.

当某些特定对象 (顾客) 进入一个组织 (服务系统), 在经历接纳、处理、留置、释放过程后离去, 这就可视为一个“排队-服务”现象. 研究这种现象的学问称为“排队论”(queuing theory), 或“等候线原理”(theory of waiting line). 也许为了强调应用, 有时也被称为“随机服务系统”(stochastic service system). 实则几乎所有的现存的书籍都着眼于理论介绍, 很少涉及实际资料. 本书的写作方式是企图从现实的观察来启发对理论的探讨, 然后再反过来把理论应用于实务. 从教学的过程中我还认识到:

(1) 以直观所获得的概念去解说问题的本质与解法是帮助学生学习的有效办法.

(2) 中国学生在课堂上普遍不会质疑教学内容, 对演绎法运用纯熟, 然而对观察现象以寻求客观规律的归纳法却较为生疏.

(3) 只有在符合实际的假设条件下找到有效的解决办法时所建立的数学模型才是对学生日后有用的材料. 这些经历就构成了本书的基本观点与内容, 尤其是在阐述理论时, 尽量避免繁琐的数学推演, 而诉诸易解的基本原理 (rationale) 与直观. 我从事实务工作的经验也不断地证实了“概念无直观容易落入空洞而难以应用”.

“排队-服务”现象在自然界和人类社会广泛存在. 譬如: 客户进入银行办理业务后离去, 工件投入生产线成为产品, 发出订货单到货物收验完毕, 一个地区野生动物的出生与死亡, 资料送入资料库后又被移出等, 都是顺着一个由接纳到释放的流程. 因此, 可以从广阔的角度来看待服务系统及其行为. 本书提供的例子包括不同方面的应用: 如公路交通、卫星通信、区域网络、物料搬运系统、生产线、设备维修、物料存量管理、紧急救援等.

为了能够建立一个优良的服务系统, 并使之有效运转, 必须先要清楚地了解在不同条件下, 系统行为以及状态的变化, 并以此作为设置服务系统和运作规则的依据. 第 1 章讨论什么是系统行为以及衡量其绩效尺度的基本概念, 以此作为进入以后各章节理论分析与应用的引导.

认识任何事物最直接的方式就是进行观察. 透过对服务系统行为的观察, 了解顾客的服务需求, 以及系统所能提供的服务水平 (例如: 为满足需求, 顾客所经历的等待时间). 所谓的需求可分两方面来看: 其一是需求提出时刻, 另一则是所需服务时间. 从逻辑来说, 当不同需求提出时刻较密集, 或者各自服务时间较长, 那么顾客等待时间也会相对较长. 然而不那么明显的现象是: 因为随机性而使得需求时刻或服务时间有较大的变易时 (粗浅地说, 就是忽长忽短), 也会导致较长的等待时间. 需求发生过程以及服务时间分布是决定服务系统行为的两个主要因素. 从课本上或教室里学习服务系统理论者往往偏执于数学解法, 而对此二者的合理性涉及不足. 因此有关这方面的讨论, 专门写入第 2 章, 其中引用的现象 (如稀有事件、衰率变化)、数据 (如人工操作时间)、统计资料 (如设备修整时间分布) 都来自于对现实世界的观察.

然而仅凭观察得到的资料, 无法从逻辑上厘清顾客需求与服务水平的量化关系. 要解决这个问题, 就需要利用数学模型来进行分析. 所谓的数学模型就是一套能够把顾客需求和系统服务能量联系到系统绩效的“需求-服务-绩效”的数学方程式. 建立一个合理的模型有三个相互关联的先决条件: (i) 从实际的问题转换为数学问题 (模型的结构) 的抽象过程, (ii) 必要的假设 (如顾客到达的随机过程和服务时间的分布) 与参数 (如顾客到达率和服务率等), (iii) 数学问题的解法. 因为没有解答的模型显然是毫无用处的. 为此之故, 第二个条件必须同时照顾到系统模型与假设的合理性以及数学上处理的难度 (tractability). 第 3、4 章讨论的简单服务系统模型就是为了能够在写出“需求-服务-绩效”方程式后, 很容易找到数学的解. 一般初级运筹学所介绍的多属这类模型. 也由于假设过于理想化, 可以应用的范围也相对受到限制. 虽然如此, 从定性的 (qualitative) 角度来看, 以及在后面章节处理较复杂问题而论及近似法时, 简单模型的结果仍具相当价值.

第 5、6 章对常用服务系统模型作了较多的讨论.“常用”的说法源于一个事实: 以泊松到达过程为假设的数学模型, 能符合许多现实的案例. 第 2 章的前半部分从不同角度对此进行了详细的论述. 到目前为止, 在泊松假设条件下, 对于多个服务

台的问题仍然无解, 因此在 5.10 节提供了一个近似解法. 此外, 在优先权 (排队等候者先后接受服务的优先顺序) 方面, 也着重地讨论了常见的优先排队规则以及它们的应用. 在这些规则中, "(服务时间) 短者优先" 是降低平均延误时间的有效措施. 书中介绍的"循环占用" (round robin) 和"反馈占用" (feedback) 基本上都是为了遵循此原则.

第 7 章论述的复杂系统模型主要是指: 求解的运算程序过于繁杂的模型, 包括任意到达过程与任意服务时间分布的服务站, 以及网络服务系统. 前者的讨论侧重于寻求平均延误时间的上、下限和高负荷状态下的近似解, 后者专注于有乘积形式解的网络的讨论.

最后, 在第 8 章里介绍了两则案例, 以此帮助说明理论如何运用到实际问题. 另一方面, 鉴于许多实务工作的效果与效率有时不尽理想, 所以在该章的最后一节对此作了简短的评论, 希望对从事设计、规划或分析服务系统者有所助益.

在学术或实务工作活动中, 曾经数次听到过同一种意见: 以为有了"仿真法"就没必要再学习"排队论". 其实这可能是由于对两者都认识不足而引起的偏见. 因此在附录 I: "仿真法与随机服务系统"作了一些澄清. 另外, 由于在实务工作上, 指数服务时间的例子实在不多, 就以"指数服务时间的服务系统"为题, 把一般教科书里讨论的"$G/M/k$ 队列"写成附录 II. 此外许多较有价值的材料以及概念, 因篇幅关系, 就以提纲挈领的方式写入"练习与讨论"里. 读者可以依照各题的指引, 将其当成功课自习以增加这方面的知识.

在读研究所时, 我曾上过加州大学伯克利分校纽厄尔教授 (Gordon F. Newell) 的课, 也私下和他有过接触, 受他影响, 在日后的工作中, 总把解决问题当作第一要务. 更因为从学校毕业后, 大部分时间都在企业界工作, 尤其在离开研究单位后, 对发表论文的意愿越来越淡薄. 本书也有些过去未曾发表 (也因此从未被审稿) 的内容, 主要部分包括: 2.1 节"到达时间是均匀分布时, 到达间隔即为指数变数"的证明, 5.6 节"$M/H/1$ 队长分布的迭代计算法", 以及 5.10 节"$M/G/k$ 队列的近似解". 图 5.12 ~ 图 5.19 中各例的近似解与仿真的运算得到褚修玮先生的协助, 他是我在台湾教书时的学生. 在写作过程中曾得到美国加州圣荷西州立大学曹孝先教授 (Jacob Tsai) 和南开大学王谦教授的帮助. 原先只准备写 7 章, 也是在一次和王老师交谈后, 才加写了第 8 章, 在此表示感谢.

书稿完成后, 作过三次校读, 但是不能保证没有谬误. 若能得到读者的指正, 将是我的荣幸. 如有赐教, 可以通过电子信箱联络: weminchow@yahoo.com.

周玮民

2016 年 4 月 10 日

美国加州洛斯阿图 (Los Altos, California, USA)

目　　录

第1章　基本概念

“服务”作为一种行为, 通常会牵连至少两方, 其中一方为了满足其他各方的需要, 而采取某种措施或提供资料. 这种行为可以是商品、劳务、专业技能、设施、信息、知识等项目的任何组合. 服务的对象就称为“顾客”.

“服务系统”就是进行这些措施与资料的组织体. 该系统可以只是一部特定功能的机器 (以制作的工件为其服务对象), 一条生产线, 或者是一团队从事设备安装、保养维修、财务管理、商业交易等, 也可以是医院、交通设施、学校以及其他公共服务系统. 系统内提供服务的单位称为“服务台” (server), 如生产系统中的操作员或医院的病床.

当顾客的需求具有随机性时, 服务系统就称为“随机服务系统”. 一般来说, 随机性表现在 (i) 需求的发生以及 (ii) 服务量上. 这二者分别以“到达时刻” (到达系统的时刻) 与“服务时间”来表示. 当系统尚未满足顾客需求时, 就可视为该顾客在“等候”完成服务. 为了方便起见, 可以认定一个顾客对应一个需求. 等候完成服务的顾客可视为排队的“队列”, 顾客人数就是“队列长度”.

有效地设计、规划、运作一个服务系统的依据在于系统的成本与绩效 (performance). 通常二者关系被“(边际) 报酬递减律” (law of diminishing return) 所规定, 如图 1.1 所示, 为一增凹函数曲线 (increasing concave function). 决策者可依此决定适当的投入成本.

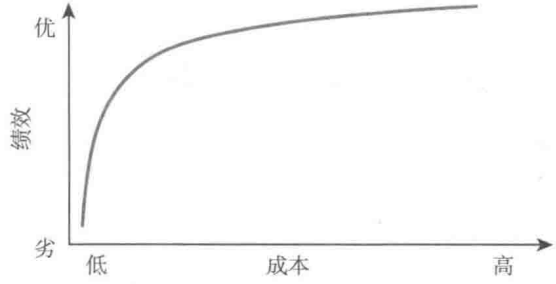


图 1.1　成本与绩效

成本的项目可以包括购置、安装、租赁、运作、维修、管理以及外包等费用. 计算方法以会计学为基础. 细节不在本书讨论范围, 读者可参考财务管理或工程经济书籍. 本书的讨论将集中绩效方面. 首先的问题: 什么是随机服务系统的绩效?

1.1　随机服务系统绩效的量度

这里介绍五类基本量度 (measure) 的方式:

(1) **系统的服务容量**(service volume) —— 如吞吐率或通过率 (throughput), 以单位时间服务 (顾客) 次数来计算. 它直接关联到系统的收益.

(2) **时间的长短**(duration of time) —— 如等待时间 (waiting time), 以从服务需求的提出到服务完成之间逝去的时间作计算. 相当于顾客留在系统的时间, 此直接关系到顾客的满意度.

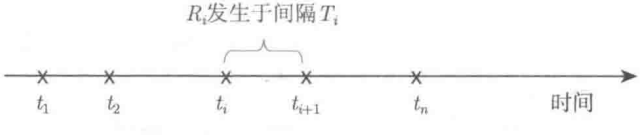
(3) **以"相对频率"**(relative frequency)**为概率** —— 包括系统各类参数或指标的统计分布. 如顾客等待时间的分布, 是指在比例上有多少顾客等待时间小于或等于 t. 又如当新的需求提出时, 先前需求尚未完成的顾客数目 (也可视为队列长度) 的分布, 是指在比例上有多少顾客在提出需求时, 先前还有小于或等于 n 个需求尚未完成.

(4) **以"时间上的占有比率"**(proportion of time)**为概率** —— 如系统使用率 (utilization) 是时间比例上系统正在提供服务 (即被使用). 又如, 即时可用率 (availability) 往往等同闲置率 (未被使用的时间比例). 再者, 队列长度为 n 的概率是指在时间比例上系统的状态为 n (尚有 n 个顾客在等着完成服务).

注意: 上面 (3) 和 (4) 提到队列长度的分布具有不同的定义. 在 (3) 中, 指的是"一个顾客到达系统时, 他看到的队长", (4) 指的是"在时间上, 队长为 n 所占的比例". 又因为在时间轴上任意 (随机) 取一点所观察系统状态为 n 的机会与状态为 n 所占时间比例成正比, (4) 中所指也可解释为"一个随机观察者看到的队长".

(5) **报酬率**(rate of reward) —— 这里介绍一个以后会常用的定理 (证明从略, 读者可参阅 (Ross., 2007)).

定理 1.1(更新报酬定理 (renewal reward theorem))　假设有一系列事件发生的时间为 $(t_1, t_2, t_3, \cdots)$. 连续两事件发生的时间间隔为 $\{T_1 = t_2 - t_1, T_2 = t_3 - t_2, \cdots\}$. 在间隔 T_i 相应报酬为 R_i, 而且 $\{(T_i, R_i)|i = 1, 2, \cdots\}$ 作为成对的随机变数, 具相同而互为独立的分布 (independently and identically distributed, iid). 那么, (长期) 单位时间平均报酬所得 (称为报酬率) 就等于 R_i 与 T_i 期望值之比: $E[R_1]/E[T_1]$. □



在上述的定理中, 因 iid 的假设, 事件发生的间隔, $\{T_i|i=1,2,\cdots\}$ 界定了一个“更新过程”: 在第 i 次事件发生后, $(T_i,T_{i+1},T_{i+2},\cdots)$ 的分布等同于 $(T_1,T_2,T_3,\cdots)$ 的分布. 换言之, 在时序中每当一事件发生, 一切过程就如同回到原点 t_1. 因此事件可称为“更新事件”(renewal), 而整个过程 (process) 称为更新过程 (renewal process). 时间点: t_1, t_2, t_3, $\cdots$ 称为“更新点”或称“再生点”(regeneration point). 与此相对应地, $T_1,T_2,T_3,\cdots$ 就称为“更新间隔”或称“再生周期”(regeneration cycle).

在进行绩效分析时, 如果能找到再生点, 那么绩效量度的计算就只需考察一个再生周期内, 系统状态变化. 举例而言: 倘若在均值为 10 单位时间的再生周期里, 平均有 8.5 个单位时间系统在提供服务, 那么系统使用率 $=8.5/10=0.85$.

下面利用一简单例子来进一步说明绩效量度.

例 1.1(单一服务台的绩效量度) 如图 1.2 所示, 假设一系统仅有一个服务台, 顾客依次到达, 并先后接受服务. “到达间隔”(inter arrival time) 是两个连续到达的时间间隔. 以随机变数 T 来表示. “到达率”(arrival rate) $\lambda=1/E[T]$.

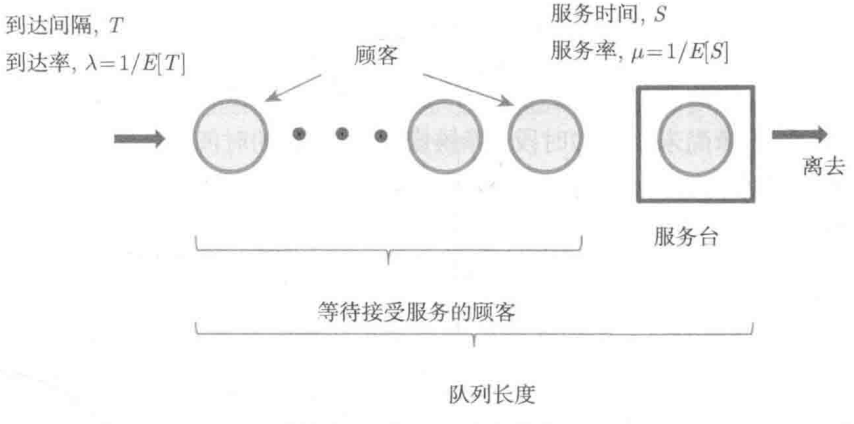


图 1.2 单一服务台系统

提供一位顾客服务所需的时间称为“服务时间”(service time). 以随机变数 S 来表示. “服务率”(service rate) $\mu=1/E[S]$. 对此系统的基本绩效量度包括下列诸项:

- 队列长度 (queue length) —— 系统的顾客数目
- 等待时间 (waiting time) —— 顾客花费在系统的时间. 该名词使用并非统一, 有人称其为“逗留时间”(sojourn time), 而计算机系统研究者却把它叫做“回应时间”(response time)
- 使用率 (utilization) —— 服务台被使用的时间比例
- 繁忙期 (busy period) —— 服务台从闲置状态转为繁忙开始, 直到再度闲置

为止的时间长度. 也被简称为“忙期”

- 吞吐率, 通过率 (throughput) —— 以单位时间实际服务 (顾客) 的次数来计算
- 服务率 (service rate) —— 单位时间可以提供最大的服务次数, 也等于平均服务时间的倒数

注意: 通过率小于或等于服务率, 二者之所以不等, 是因为服务台可能会闲置或无法使用.

从上面各项, 还可引申更多量度:

- 等候接受服务的顾客数 (number of customers waiting for services) —— 相当于队列长度去除正在接受服务的人数
- 延误时间 (delay) —— 等候接受服务的时间, 通常等同于等待时间减去服务时间
- 闲置期 (idle period) —— 每次系统停留在闲置状态的时间长度
- 繁忙周期 (busy cycle) —— 繁忙期 + 闲置期 (此二者交替出现, 故称繁忙周期)
- 可靠性 (reliability) —— 系统功能完好 (无故障) 所占的时间比例

令 X = 连续两次故障发生时间的间隔

Y = 故障尚未修复的时段 (等候修理 + 修理的时间)

根据定理 1.1 (更新报酬定理), 令 $T=X+Y$, $R=X$, 则可靠性 $=E[X]/E[X+Y]$.

同理, 使用率 $=E$[繁忙期]$/E$[繁忙期 + 闲置期]

- 可用率 (availability, accessibility) —— 系统可被使用时所占的时间比例

在现实系统运作中, 无法使用的原因包括待修、维修、定期保养、设备安置 (set-up)、测试、待料、设施故障 (如停电)、人员的疲乏、延误、旷职、交班等.

此外, 服务率与服务的熟练度、技能有关, 有时也和投入的资源有关. 服务率 (服务时间) 可随着服务次数的增加而递增 (减). 图 1.3 所示是一典型的学习过程.

图中每一点代表一日完成的工件数. 总体趋向如图 1.3 中的曲线所示, 可用一指数函数来表示. 较常用的模型 (model) 是以 X_n 为第 n 日完成的 (平均) 工件数:

$$X_n = a \cdot n^b \tag{1.1}$$

该式有两个参数, 即 a 与 b, 可以实际观察为资料, 利用“最小平方法” (least square method) 来确定. 在应用上, n 可以解释成时间 (如 X_n 第 n 月的绩效), 也可代表次数 (X_n 是制造的第 n 件的时间). 那么, 第一次的估计平均值 $X_1 = a$.

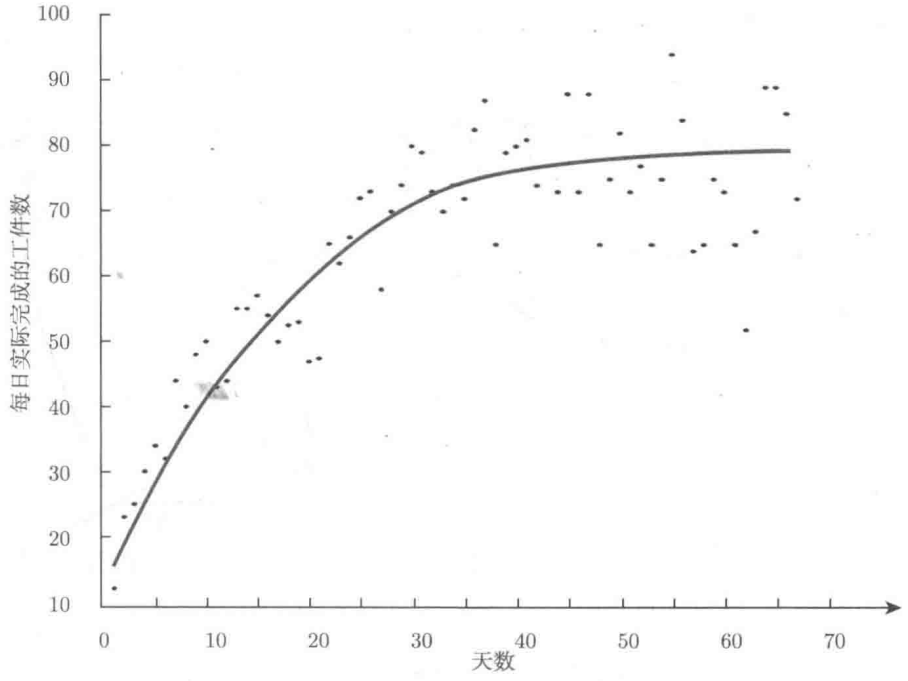


图 1.3 学习过程

资料来源: Glover J H. Manufacturing progress functions. *International J. of Production Research*, v.5, 1966

整个改进的速率可由下面比值来解释:

$$h = X_{2n}/X_n = 2^b \quad \text{或者} \quad b = \log_2 h \tag{1.2}$$

由经验得知, 如$\{X_j\}$为手工操作时间 (此时 X_j 为减函数), 则 $h \in (0.7, 0.95)$. 换言之, 当工作次数加倍时, 操作时间减少 5%~30%. 在设计新系统时, 通常尚无观察样本, 此时可以假设 $h = 85\%$, 也即 $b = \log_2 0.85 = -0.2345$; 反之, 图 1.3 中的 X_j 为增函数, h 的值应介于 1.05 和 1.43 之间.

另一较复杂的模型由 J. H. Glover 在他的论文 (*Manufacturing Progress Functions. International Journal of Production Research*, v.5, 1966) 中提出, 以逐日累积工件数 $Y_n = \sum_{i=1}^{i=n} X_n$ 为变数,

$$Y_n = c \cdot n^d \tag{1.3}$$

其中 c 和 d 的求法与 a 和 b 同. 因为每一个 Y_n 都由 X_1 起算, 此模式特别注重前期的行为.

察看图 1.3 中的前后期可知: (i) 前期改进速度较快, 因此强调前期数据的 (1.3) 模型往往较之 (1.1) 更为精准, 而 (ii) 后期虽有较为稳定的平均工件数, 但每日差异变大.

1.2　服务系统的排队问题及其模型

从广义面来说, 一个服务系统的运作不外乎接纳到来的顾客 (可为一种任何实体), 并经过逐个的处理、必要的留置, 最后再释出系统. 其应用范围涵跨生产线、物料搬运系统、存货系统、水坝、计算机系统、通信网络、维修服务、交通系统、旅馆租赁、银行、饭店、零售店、医院、紧急救护等. 甚至在森林资源的利用维护中的树木, 社区发展的建筑物都可视为系统上的顾客.

由于顾客等候或接受服务时常呈现“排队”(queues) 的状态, 分析一个随机服务系统就自然会以“排队论”(queuing theory) 的理论 (亦称等候线原理, waiting line theory) 为基础来建立数学模型 (mathematical model).

建立数学模型是一个抽象过程, 参见图 1.4, 其目的是利用一个代表实际系统的“逻辑结构”来作分析与优化系统模型以服务绩效的量度为其输出 (output), 输入 (input) 部分为系统的参数以及运作规则 (rules of operations), 通常是经由观察与资料分析得到, 具体项目包括:

- 系统的结构布局 (topological structure)
- 服务需求发生的模式 (demand pattern)
- 对服务要求 (service requirements)
- 服务的顺序 (order of service), 也即“排队规则”(queuing discipline)

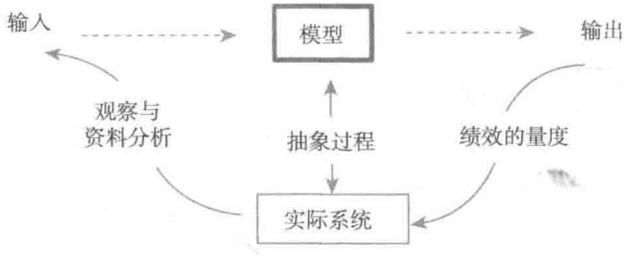


图 1.4　建立数学模型

建立模型过程通常经由下列工作才算完成:

- 完备界定的逻辑结构及其运作规则
- 输入的资料收集与整理以作为数学分析之用
- 求取模型解 (model solution), 以作为绩效分析与评估

除了绩效量度之外, 通过建立模型还可:

- 进一步了解系统的结构与逻辑
- 决定系统中具关键性的参数
- 增进模型求解的能力

- 利用模型实验改善系统的设计与运作
- 减少成本与时间的浪费

以后诸章节将由简入繁的顺序逐次讨论不同的模型. 在此特别需要强调的是: 任何数学模型仅仅是用来分析与了解实际系统行为的一个方便措施. 不论它的精确性多高, 也只能是近似于实际. 因此从应用面来看, 合理而方便的解往往远比精确而繁复的解更有价值.

1.3 图示系统上队列的特性

系统的状态与顾客接受服务的质量可简单地表现在图 1.5 中. 图中的横坐标代表时间, 纵坐标为累积次数, 令:

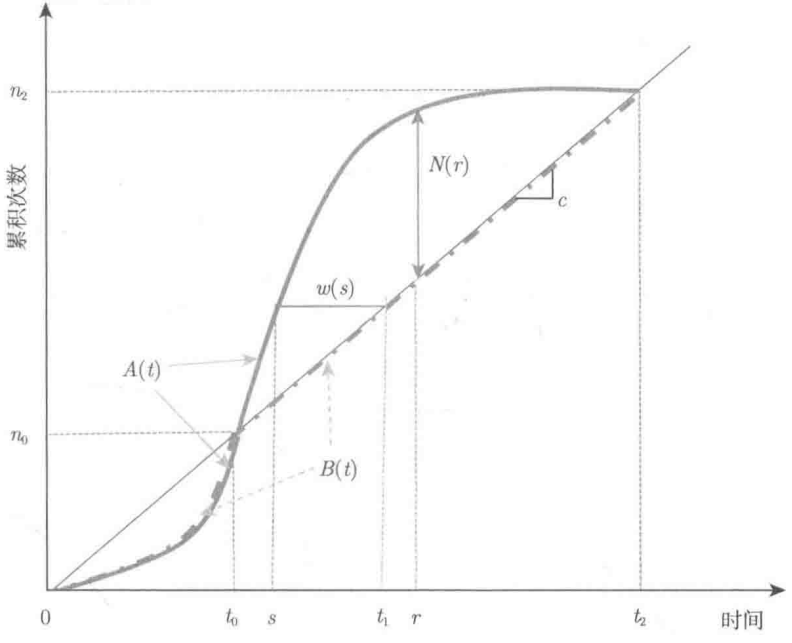


图 1.5 图示系统队列特性

斜线的斜率 c 为服务率;

S 形曲线 $A(t)$ 为在时间 t 之前, 累计到达的顾客数;

虚线 $B(t)$ 为在时间 t 之前, 累计离去的顾客数;

$N(r)$ 为在时间点 r 的队列长度;

$w(s)$ 为在时间点 s 到达者花费在系统上的等待时间.

则: (1) 在 $t \in (0, t_0)$ 时段, 由于 $A(t) < ct$, $B(t) = A(t)$.

(2) 当$t\in(t_0,t_2)$, $A(t)>ct$, 到达系统的速率超过了服务的速率c, 故$B(t)=ct$.

(3) 任意时刻$r\in(t_0,t_2)$的队列长度, $N(r)=A(r)-B(r)$.

(4) 在s到达者必须等到前面顾客完成服务离去后, 才能轮到自己接受服务, 所以他离去的时间点会在累计到达数等于累计离去数的时刻, $t_1\in A(t_1)=ct_1$等待时间$w(s)=t_1-s$.

图 1.5 中斜线左上方与曲线之间的面积

$$a=\int_{t_0}^{t_2}N(r)dr=\int_{n_0}^{n_2}w(s)ds \tag{1.4}$$

可视为累积的队列长度, 或者累积的等待时间. 那么

平均队长:

$$L=a/(t_2-t_0)$$

平均等待时间:

$$W=a/(n_2-n_0)=a/[A(t_2)-A(t_0)]$$

$(0,\ t_2)$的到达率

$$\lambda=[A(t_2)-A(t_0)]/(t_2-t_0)$$

由上列三式可得

$$L=\lambda W \tag{1.5}$$

1.4　高速公路交通问题

1.3 节的作图法可以很好地解释高速公路交通拥塞的状况. 假设某段公路经历拥挤时刻, 令

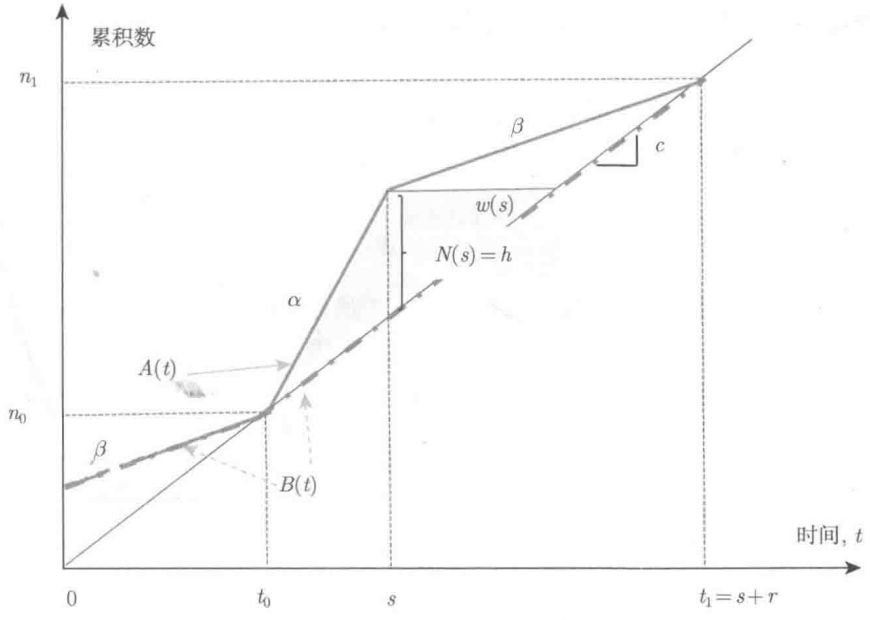
c为公路可承受的最大流量 (车辆数/时);

α为拥挤时刻每小时需求使用公路的车辆数, $\alpha>c$;

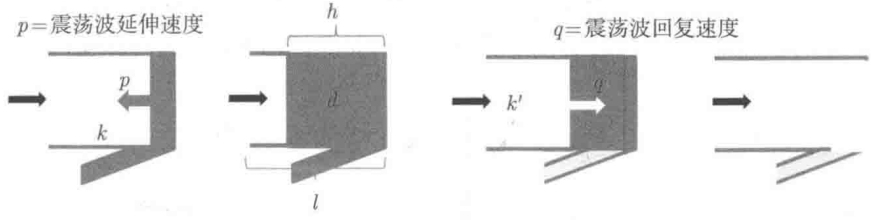
β为平常时刻每小时使用公路的车辆数, $\beta<c$.

在图 1.6(a) 中, 拥挤时段为$(t_0,\ s)$, 累计需求曲线$A(t)$的斜率为α, 而$(0,t_0)$与$(s,\ t_1)$为平常时段, 需求率为β.

图 1.6 (b) 是一个比较接近真实情况描述. (i) 左一图中, 公路进口有过多车辆出现而造成拥塞, (ii) 拥塞区域逐渐向上游延伸 (如左二图示), (iii) 然后进入车辆减少, 拥塞区相应变小 (左三图), (iv) 最终回复平常状态.



(a) 公路交通变化模型



(b) 公路交通的拥塞与恢复

图 1.6

拥塞区的扩张 (收减) 看起来就如一向上 (下) 游传送的“震荡波”(shock wave) 一样, 其速度 $p(q)$ 可由密度 (公路上每单位长度的平均车辆数) 与交通流量 (每单位时间通过的车辆数) 求得. 令:

d 为拥塞区车辆密度;

k, k' 为非拥塞区车辆密度;

u 为 (上游) 非拥塞区的车流量;

v 为非拥塞区平均行车速度;

l 为路段的长度;

h 为拥塞区最大的长度 (排队模型中队列最长时的长度).

当公路段无法让增加车辆全数“顺利”通行时, 车辆密度就会快速增加, 而车速放缓, 流通量随之降低, 由高流量进入低流量而多出的车辆就如同被震荡波所“吸