运筹与管理科学丛书 6

离散时间排队论

田乃硕 徐秀丽 马占友 著

3

科学出版社
www.sciencep.com

运筹与管理科学丛书 6

离散时间排队论

田乃硕 徐秀丽 马占友 著

科学出版社
北京

内 容 简 介

本书系统论述离散时间排队的思想原理和主要结果，建立了一个完整的理论框架. 内容包括 Markov 型、Geom/G/1 型、GI/Geom/c 型、D-BMAP/G/1 型等各种离散时间排队系统的建模和分析，并简要介绍了离散时间排队网络. 除经典模型外，还详细讨论了近些年出现的休假和工作休假离散时间排队系统，并包含计算机通信网络和卫星通信系统性能分析的应用实例. 其中部分内容是作者近年来的研究成果. 本书叙述深入浅出、论证严谨、图文并茂，注意先进性、系统性和实用性.

本书可作为运筹学、管理科学、应用数学、计算机科学、通信科学等专业高年级本科生和研究生的教材或教学参考书，也可供相关专业的科研人员和工程技术人员阅读参考.

图书在版编目(CIP)数据

离散时间排队论/田乃硕，徐秀丽，马占友 著. —北京：科学出版社，2008
(运筹与管理科学丛书；6)
ISBN 978-7-03-021869-8

Ⅰ.离… Ⅱ. ①田… ②徐… ③马… Ⅲ.离散控制-控制系统-排队论
Ⅳ. TP271 O226

中国版本图书馆 CIP 数据核字(2008) 第 063476 号

责任编辑：赵彦超／责任校对：陈玉凤
责任印制：赵德静／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

新蕾印刷厂 印刷

科学出版社发行 各地新华书店经销
*
2008 年 6 月第 一 版 开本：B5(720×1000)
2008 年 6 月第一次印刷 印张：21 3/4
印数：1—3 000 字数：406 000

定价：59.00 元

《运筹与管理科学丛书》序

运筹学是运用数学方法来刻画、分析以及求解决策问题的科学. 运筹学的例子在我国古已有之, 春秋战国时期著名军事家孙膑为田忌赛马所设计的排序就是一个很好的代表. 运筹学的重要性同样在很早就被人们所认识, 汉高祖刘邦在称赞张良时就说道:“运筹帷幄之中, 决胜千里之外.”

运筹学作为一门学科兴起于第二次世界大战期间, 源于对军事行动的研究. 运筹学的英文名字 Operational Research 诞生于 1937 年. 运筹学发展迅速, 目前已有众多的分支, 如线性规划、非线性规划、整数规划、网络规划、图论、组合优化、非光滑优化、锥优化、多目标规划、动态规划、随机规划、决策分析、排队论、对策论、物流、风险管理等.

我国的运筹学研究始于 20 世纪 50 年代, 经过半个世纪的发展, 运筹学队伍已具相当大的规模. 运筹学的理论和方法在国防、经济、金融、工程、管理等许多重要领域有着广泛应用, 运筹学成果的应用也常常能带来巨大的经济和社会效益. 由于在我国经济快速增长的过程中涌现出了大量迫切需要解决的运筹学问题, 因而进一步提高我国运筹学的研究水平、促进运筹学成果的应用和转化、加快运筹学领域优秀青年人才的培养是我们当今面临的十分重要、光荣、同时也是十分艰巨的任务. 我相信,《运筹与管理科学丛书》能在这些方面有所作为.

《运筹与管理科学丛书》可作为运筹学、管理科学、应用数学、系统科学、计算机科学等有关专业的高校师生、科研人员、工程技术人员的参考书, 同时也可作为相关专业的高年级本科生和研究生的教材或教学参考书. 希望该丛书能越办越好, 为我国运筹学和管理科学的发展做出贡献.

袁亚湘

2007 年 9 月

前　言

自 Erlang 的开创性工作以来, 排队论经历了百年历程, 已发展成为运筹学和应用概率论中的重要分支学科. 经典排队论主要研究和处理各种连续时间系统, 已经确立了成熟的理论体系, 并在生产、交通、服务、通信、管理和军事领域得到广泛的应用.

21 世纪, 人类进入了信息时代. 计算机和计算机网络的应用已经遍布社会活动和生产生活的各个领域, 人类文明和社会进步从来没有经历过如此深刻的震撼和推动. 信息革命正在改变着传统科学技术领域的思想观念、理论基础和研究方法. 在当代计算机产业和宽带综合业务数字网络 (BISDN) 中, 语音、数据、图像等各种业务都将通过数字化分割成固定长度信元或包, 进行统一的处理和传输. 一个业务量的大小, 用它所包含信元的个数来度量; 一个需求占用的服务时间, 是单位信元处理时间的整倍. 计算机系统和 BISDN 中大量操作环节和运行过程, 更自然地表现出离散时间现象的本质特征. 计算机科学和信息技术的发展, 激发了人们对离散时间系统的关注和探索, 离散时间排队研究正是在这种背景下产生和发展起来的.

Meisling(1958) 的论文是离散时间排队分析的起点. 与经典的连续时间排队不同, 一个离散时间排队模型的基本要素 (到达间隔、服务时间等) 都是非负整值随机变量; 系统运行过程中的各种事件 (顾客到达、服务的开始或结束等) 都只发生在整点时刻. 由于离散时间排队更适合于计算机系统建模和性能分析, 引起了大批排队论和通信工程专家的广泛关注, 并迅速产生了大量理论分析及实际应用成果. Hunter 在 *Mathematical Techniques of Applied Probability* (1983) 一书中, 以一章的篇幅对离散时间排队的早期成果给出了一个系统的处理. Takagi 的著作 *Queueing Analysis* (1993, 卷 3, 离散时间系统) 从为计算机系统性能分析提供数学工具的角度出发, 对各种 Gemo/G/1 系统给出了完整的分析. 使用离散时间排队进行计算机系统和通信网络分析和设计的专著主要有: Brunnel 和 Kim 的 *Discrete-time Models for Communication Systems Including ATM* (1993), Woodward 的 *Communication and Computer Network-Modelling with Discrete Time Queues* (1994). 离散时间排队不仅丰富和扩展了经典排队论研究, 而且为计算机系统和通信网络的设计优化与性能分析提供了理论基础和数学工具. 然而, 或因写作时间较早, 或因局限于具体应用目的, 上述著作均不能反映离散时间排队研究的全貌, 更未建立完整的理论体系.

本书的目的, 是对离散时间排队构建一个较完整的理论框架, 同时较全面地反映该领域的研究成果和方法. 除上述几本著作外, 离散时间排队的大量研究成果散

见在各种杂志上, 并且通常是使用不同方法给出的. 我们对国内外大量成果进行梳理和加工, 结合自己的研究工作, 给出了一个由浅入深、层次分明、有机联系的理论体系. 在内容上, 不仅包括国内外最常见的 Geom/G/1 型系统, 还包括生灭链型和 GI/Geom/1 型模型; 既包括经典模型的分析, 也关注带有各种休假策略、成批 Markov 到达、离散时间排队网络等新近研究成果的系统处理. 对比较简单的模型, 尽量使用嵌入 Markov 链和补充变量等常规方法; 对更复杂的模型, 使用当代流行的拟生灭过程、矩阵几何解和 M/G/1 型结构矩阵分析方法处理. 力争在研究内容和分析方法上反映国内外当代水平. 书中还引入大量计算机系统和通信网络性能分析的例子, 努力做到理论分析难度适中, 应用例子丰富生动、表述流畅、图文并茂.

作者从事排队论研究已有三十多个寒暑, 特别是在休假排队和离散时间排队分析中取得了诸多研究成果. 书中大量内容是作者们自己的研究工作.

全书共分八章. 第 1 章论述离散时间排队中出现的一些新现象和独有的特性. 第 2 章简要介绍本书中需要的预备知识, 包括 Markov 链、离散 PH 分布、可逆 Markov 链等内容. 第 3 章处理各种离散时间生灭链型排队系统, 其中离散时间消失系统、非齐次成批到达的无穷服务台系统是作者的研究工作. 第 4 章致力于各种 Geom/G/1 型系统的分析, 包括成批到达、有限等待场所及各种休假机制的模型. 本章中多级适应性休假模型的引入和分析是作者的研究工作, 它包含国际上研究的众多 Geom/G/1 休假模型作为特例. 第 5 章处理各种 GI/Geom/c 型离散时间排队, 包括多服务台和带有休假机制的模型. 本章中 GI/Geom/1 型休假排队系统的分析是作者首先给出的. 第 6 章关于各种工作休假排队模型的分析全部是作者近期的研究成果. 工作休假机制由 Servi 与 Fann(2002) 引入, 是经典休假排队的一种推广, 并在光纤通信系统中有重要的应用背景. 众所周知, 在 BISDN 中, 需求的到达过程表现出明显的相依性和突发性, 传统的更新到达过程已不能描述这些新的特征. 近些年引入的 Markov 到达过程克服了传统到达模式的局限性, 能很好地适用于 BISDN 建模和分析. 第 7 章系统地处理各种有离散成批 Markov 到达过程的排队模型. 第 8 章简要介绍离散时间排队网络的理论和成果. 排队网络是尚处于发展过程中的研究领域, 但是在通信网络建模与分析中极为重要. 本章结合卫星通信和局域网领域的应用例子, 给出离散时间排队网络的一个系统的处理. 每一章的最后都设置文献评述一节, 给出本章内容和相关工作的出处, 便于读者阅读和研究进一步的文献.

阅读本书只需要 Markov 过程和矩阵分析的基本知识. 本书可供随机运筹学及管理科学领域的研究人员、高校教师和相关专业研究生使用, 也可供计算机系统和通信网络的工程技术人员阅读参考.

1994年以来, 我们关于排队论的研究得到国家自然科学基金 (19871072;

10271102; 10671170) 的连续资助, 对此作者表示衷心的感谢.

美国 Western Washington 大学 Zhe George Zhang 教授长期与作者合作研究, 燕山大学信息科学与工程学院申利民教授、金顺福教授、刘洺辛博士参与并承担部分研究工作, 博士研究生李继红、陈海宴、孙威、霍占强和田志斌也在课题研究和书稿撰写过程中做了大量的工作. 在此, 作者向他们表示感谢.

作　者

2007 年春节

于秦皇岛燕山大学

目　录

第 1 章　引　　论

§ 1.1　离散时间排队模型

自 Erlang 关于排队论的开创性工作以来, 排队论经历了近百年历史. 排队系统, 也称为随机服务系统, 是研究服务过程和拥挤现象的随机模型. 顾客随机地到达一个服务场所, 要求进行某种服务. 服务可能立即开始, 也可能需要排队等待一段时间后才开始, 完成服务的顾客离开系统. 在这类模型的描述中, 有两个随机变量是最基本的. T_n 表示第 $n+1$ 个顾客与第 n 个顾客到达时刻之间的间隔, 在简单的情况下, 假定 $\{T_n, n \geqslant 1\}$ 是独立同分布的, 简记为 T, 称为到达间隔, 有分布函数 $A(x) = P\{T \leqslant x\}, x > 0$. 以 S_n 表示第 n 个顾客要求的服务时间, 通常假定 $\{S_n, n \geqslant 1\}$ 是独立同分布的, 简记为 S, 称 $B(x) = P\{S \leqslant x\}, x > 0$ 为服务时间分布. 当然, 完整地刻画一个排队系统, 除了到达时间间隔和服务时间分布以外, 还需要指定一个排队规则 —— 以怎样的顺序安排顾客进行服务的法则.

到达间隔和服务时间是非负连续随机变量的排队系统, 称为连续时间排队. 在这类模型中, 顾客到达或离去这些事件的发生时刻可以取任何正实数. 排队论的大量文献和著作, 主要集中于连续时间排队系统. 到达间隔和服务时间都是正整值随机变量的排队系统, 称为离散时间排队. 这相当于把时间轴分割成等长的部分, 称为时隙 (slot). 顾客的到达和离去只能发生在时隙的分点处. 与连续时间排队相比, 离散时间排队的研究是较晚才开始的. Meisling(1958) 的论文是关于离散时间排队的开创性工作.

离散时间排队有广泛的应用背景, 特别是计算机通信技术的发展, 极大地推动了离散时间排队的研究和应用. 例如, 在宽带综合业务数字网络 (B-ISDN) 中, 异步转移模式 (ATM) 被国际电联电信标准部确立为未来的通用技术之一. ATM 网络以单一的网络结构和综合的方式处理语音、数据、图形和电视各种信息传输. 所有的消息数字化以后都被分割成有固定长度的 ATM 信元 (cell), 每个信元由 53 个字节 (8 位数) 构成. 前 5 个字节为首标, 其余 48 个字节为消息域. 在这个系统中, 顾客是随机到达并要求传输的消息, 排队指消息在缓冲器中等待传输, 服务台是计算机通信系统或传输路线. 任何消息的服务 (传输) 时间是单位 ATM 信元持续时间的整数倍, 因此是正整值随机变量. 这一系统中最自然和最基本的时间单位是一个 ATM 信元在线路中的传送时间, 服务开始和结束都只能在有确定间隔的分点处发生. 类似地, 在各种计算机系统中, 比特或二进制码的持续时间是最基本的时间单

位, 一个程序所需要的服务时间就是它包含的二进制码的数量. 此外, 在生产自动线上, 待加工部件以等距方式传递; 在交通领域, 火车、飞机在确定时刻到站. 这些系统均可用离散时间排队模型刻画.

离散时间排队模型分析的整个过程, 几乎每一步骤都可给出直观而具体的概率解释. 关于系统性能指标的结果, 不仅有简单的形式, 而且有明确的概率意义. 另外, 正如 Meisling 指出的, 若干连续时间排队问题可以通过首先研究一个离散时间排队模型, 然后令时隙无限缩短取极限的方法解决.

同连续时间排队一样, 一个离散时间排队系统主要由下列三部分组成:

(1) 到达模式. 顾客到达只能发生在时间轴上的整点处. 若在每个时隙分点上最多到达一个顾客, 称为单个到达. τ_n 表示第 n 个顾客的到达时刻, $\tau_0=0, T_n=\tau_n-\tau_{n-1}, n\geqslant 1$ 称为到达间隔, 它是正整值随机变量. 在简单的情况下, 假定 $\{T_n, n\geqslant 1\}$ 是独立同分布的, 这时, 到达过程是有离散更新间隔的更新过程. 更新间隔 $T_n\equiv T$ 有分布

$$p_n=P\{T=n\},\quad n\geqslant 1,$$

数学期望 $E(T)$ 称为平均到达间隔. 一个重要的特例是 T 服从几何分布

$$p_n=P\{T=n\}=p(1-p)^{n-1},\quad n\geqslant 1,\quad 0<p<1.$$

这一到达过程可解释为: 在每个时隙分点处以概率 p 到达一个顾客, 以概率 $1-p$ 无到达, 并且不同整点处到达行为相互独立. 长为 n 个时隙的一段时间到达顾客数 C_n 服从二项分布

$$P\{C_n=k\}=\binom{n}{k}p^k(1-p)^{n-k},\quad k=0,1,\cdots,n.$$

因此, 有几何分布到达间隔的离散时间到达过程, 称为 Bernoulli 到达过程. 类似于连续时间排队中的指数分布, 在离散随机变量中只有几何分布是无后效的. 离散时间的 Bernoulli 到达过程相当于连续时间的 Poisson 到达过程, 有许多简单的性质, 并且容易进行解析的数学处理.

如果允许一个整点处有许多顾客同时到达, 称为成批到达. 这时, 除了到达间隔分布以外, 还需给出批量分布. 假设同时到达的顾客数为 A, 有批量分布

$$a_k=P\{A=k\},\quad k\geqslant 1,\quad \sum_{k=1}^{\infty}a_k=1.$$

当间隔 T 服从几何分布且与批量 A 独立时, 任何时隙分点处到达顾客数 C 有分布

$$b_0=P\{C=0\}=1-p,\quad b_k=P\{C=k\}=pa_k,\quad k\geqslant 1.$$

在更一般情况下, 还需要引入到达间隔有相依性结构的离散时间到达过程. 例如, 离散时间 Markov 到达过程 (DMAP).

(2) 服务机制. 首先应规定服务场所中服务员的个数, 只有一个服务员称为单服务台系统; 多个服务员并列地接待顾客称为多服务台系统. 以 S_n 表示第 n 个顾客需要的服务时间, 在简单的情况下假定 $\{S_n, n \geqslant 1\}$ 是独立同分布的, 并记 $S_n \equiv S$, 有分布

$$P\{S=k\}=g_k, \quad k \geqslant 1,$$

称 $E(S)$ 为平均服务时间. 一个特殊情况是 S 服从几何分布

$$P\{S=k\}=\mu(1-\mu)^{k-1}, \quad k \geqslant 1, \quad 0<\mu<1.$$

这可解释为: 如果服务在某个时隙上正在进行, 则在该时隙结束时以概率 μ 完成服务, 以概率 $1-\mu$ 服务将持续到下一个时隙.

如果在任何时刻, 服务台最多只能接待一个顾客, 称为单个服务的, 服务台可同时接待多个顾客称为成批服务的. 对后一种情况, 服务员可同时接待的顾客数称为服务批量, 它可以是定数, 也可以是随机变量.

(3) 排队规则. 当系统中有多个顾客排队等待时, 一旦服务台空出, 如何安排队列中顾客进入服务台的方式称为排队规则. 在大多数情况下, 按到达顺序安排顾客服务, 称为先到先服务, 简记为 FIFO(first in first out). 依据不同应用背景, 也需要引入其他排队规则. LIFO(late in first out)指服务台空出时, 队列中最后到达者先开始进行服务. 如果当服务台空出时, 在排队中的顾客任意选出一名顾客开始服务, 称为按随机顺序服务, 简记为 SIRO (service in random order). 在许多实际应用中, 顾客可能分成多种类型, 每一类顾客享有不同等级的优先权. 一旦服务台空出, 队列中具有高优先权的顾客先开始服务. 在优先权服务中, 如果高优先权顾客一旦到达, 立即驱逐正在接受服务的低优先权顾客, 称为抢占型优先权服务. 如果容忍低优先权顾客完成正在进行的服务, 然后高优先权顾客再进入服务台, 称为非抢占型优先权服务. 根据各种实际问题的需要, 在排队论研究中引入了形形色色的排队规则.

与连续时间排队一样, 使用 Kendell 引入的记法表示一个离散时间排队系统. 按这种表示法, 除了需指定排队规则外, 使用 3 个符号描述一个排队系统, 写成 A/B/c 形式, 符号 A 和 B 分别用于描述到达间隔和服务时间的分布类型, c 是数字, 用于表示系统中服务台的个数. A 与 B 将用一些特定的符号取代, 通用记号是:

D: 表示定长分布. A=D 意味着到达间隔为一个确定的长度 (正整数), B=D 表明顾客服务时间为定长.

Geom: 表示几何分布, A=Geom 意味着到达是 Bernoulli 过程, B=Geom 表明服务时间服从几何分布.

G: 表示一般分布, 只假定到达间隔或服务时间是一般的正整值随机变量, 对其分布不加具体限制.

按 Kendell 表示法, 例如, Geom/G/1 表示到达为 Bernoulli 过程, 服务时间服从一般分布的单服务台离散时间排队系统. D/Geom/c 表示到达间隔定长, 服务时间服从几何分布的 c 个服务台离散时间排队系统, 如此等等.

与连续时间排队相同, 对离散时间排队模型, 主要关心的仍然是系统中顾客数、等待时间和忙期等性能指标. 区别在于, 不仅系统中顾客数, 现在, 等待时间和忙期也是整值随机变量.

§1.2　入 口 协 议

在连续时间排队中, 顾客到达和离去等事件发生于同一时刻的概率为零. 因此, 总可以区别这些事件发生的先后顺序. 在离散时间排队中, 所有事件发生在时间轴上的整点处, 时刻 $t = n$ 处既可能发生顾客到达, 也可同时发生完成服务顾客离去. 如果以 $Q(n)$ 表示时刻 n 系统中顾客数, 为了使 $Q(n)$ 有明确意义, 就必须进一步规定时刻 $t = n$ 处发生事件的顺序, 从而决定在 $t = n$ 处到达或离去的顾客是否记入 $Q(n)$. 关于在整点时刻到达和离去、服务开始或结束规则的更细致的约定, 统称为入口协议.

事实上, 可以讨论两种不同的情况: 第一类称为早到系统 (early arrival system), 到达发生在时隙首端 (n, n^+) 处, 服务完成只能发生于时隙末端 (n^-, n) 处. 第二类称为晚到系统 (late arrival system), 到达只能发生在时隙末端 (n^-, n) 处, 服务完成发生在时隙首端 (n, n^+) 处. 图 1.2.1 显示了这两类模型的区别.

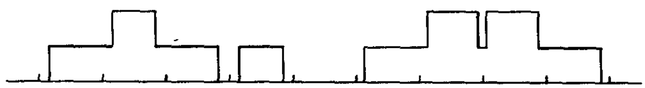
(a) 早到系统　　(b) 晚到系统

图 1.2.1　两类到达模式

在晚到系统中, 还要区别两类不同情况. 一类是 (n^-, n) 到达的顾客遇空闲服务台可立即开始服务, 并且服务可能在 (n, n^+) 处结束, 称为有直接入口 (immediate access) 的晚到系统. 例如, 当服务时间服从参数 μ 的几何分布时, 一个在 (n^-, n) 到达并遇空出的顾客, 在有直接入口的晚到系统中, 将以概率 μ 于 (n, n^+) 处完成服

务离去. 另一类是 (n^-,n) 处到达的顾客遇空闲服务台, 其服务不能在 (n,n^+) 处结束, 称为有延迟入口 (delayed access) 的晚到系统. 这可看成顾客在 (n^-,n) 处不能直接进入空闲服务台, 需延迟到 (n,n^+) 才开始接受服务. 因此, 在 (n^-,n) 到达并遇服务台空闲的顾客, 于 (n,n^+) 处完成服务的概率为零.

如上所述, 按入口协议, 一个离散时间排队系统可以表现为三种不同的模型: 早到系统、有直接入口的晚到系统、有延迟入口的晚到系统. 离散时间排队系统状态的描述不仅取决于入口协议的分类, 还与考察时刻上的细微差别有关. $Q(n^-),Q(n)$, $Q(n^+)$ 分别表示时刻 n^-,n,n^+ 系统中的顾客数, 即使在入口行为完全确定的模型中, 它们也将取不同的值. 应该强调指出的是, 即使在完全相同的到达时刻和服务时间实现下, 上述三种入口行为模型中系统中顾客数过程也可以有不同的样本轨道. 作为例子, 设到达时刻是 0, 1, 3, 5, 6, 7(到达间隔是 1, 2, 2, 1, 1), 服务时间是 2, 1, 1, 2, 1, 1. 图 1.2.2 给出了三种入口行为下顾客数过程的样本轨道, 以及在确定的入口行为下 $Q(n^-),Q(n)$ 和 $Q(n^+)$ 取值的区别.



	0	1	2	3	4	5	6	7	8	9
$Q(n^-)$	0	1	2	1	1	0	1	2	2	1
$Q(n)$	0	1	1	0	0	0	1	1	1	0
$Q(n^+)$	1	2	1	1	0	1	2	2	1	0

(a) 早到系统

	0	1	2	3	4	5	6	7	8	9
$Q^{(i)}(n^-)$	0	1	1	0	0	0	1	1	1	0
$Q^{(i)}(n)$	1	2	1	1	0	1	2	2	1	0
$Q^{(i)}(n^+)$	1	1	0	0	0	1	1	1	0	0

(b) 有直接入口的晚到系统

	0	1	2	3	4	5	6	7	8	9
$Q^{(d)}(n^-)$	0	1	2	1	1	0	1	2	2	1
$Q^{(d)}(n)$	1	2	2	2	1	1	2	3	2	1
$Q^{(d)}(n^+)$	1	2	1	1	0	1	2	2	1	0

(c) 有延迟入口的晚到系统

图 1.2.2 不同样本轨道的比较

虽然在同一模型中, $Q(n^-), Q(n), Q(n^+)$ 各不相同, 但对三种入口行为模型, 它们之间也存在一些明显的联系. 容易证明: $Q^{(i)}(n^-) = Q(n)$, $Q^{(i)}(n) = Q[(n+1)^-]$, $Q^{(i)}(n^+) = Q(n+1)$, $Q^{(d)}(n^-) = Q(n^-)$, $Q^{(d)}(n^+) = Q[(n+1)^-]$ 等等, 这里上标 (i) 和 (d) 分别表示直接入口和延迟入口的晚到系统, 无上标表示早到系统. 基于这些联系, 我们可以只研究一种入口行为的模型, 并集中于三个状态变量中的一个. 本书主要研究早到系统和有延迟入口的晚到系统, 其他模型的结果容易从得到的结论中推出. 无论如何, 在分析一个离散时间排队模型时, 必须明确规定入口协议及考察时刻, 才能确保系统状态有明确的意义.

在文献中, 关于到达和离去模式引入了有微小差别的各种变体. 不同的入口规则, 取决于控制系统运行中对同时发生事件顺序的不同规定和协议. 入口行为的规定, 对所关注的过程是否具有准可逆性及顾客服务质量的性能指标都有实质性的影响. 详见文献 (Chao et al., 1999; Gravey et al., 1992).

§1.3 文献评述

排队论文献主要集中于连续时间模型, 已经出版了大量连续时间排队论的著作. 使用经典方法处理连续时间排队的著作如文献 (Saaty, 1961; Takacs, 1962; Prabhu, 1965; Cohen, 1982; Kleinrock, 1976; Cooper, 1981; Gross et al., 1985; Takagi, 1991) 等. 近些年, 出版了一批使用点过程和鞅论等现代方法处理连续时间排队的专著, 如文献 (Bremaud, 1981; Franken et al., 1982; Brandt et al., 1990; Baccelli et al., 2003) 等. 国内关于连续时间排队的主要著作如文献 (徐光辉, 1988; 孙荣恒等, 2002; 唐应辉等, 2006) 等.

与连续时间排队比较, 离散时间排队研究是较晚才开始的. 1958 年, Meisling 的论文研究了 Geom/G/1 离散时间排队模型, 并指出平行于连续时间排队论建立和发展离散时间排队理论的可能性. 随后, Dafermos 与 Netus (1971), Netus 等 (1973,I-IV), Hsu 与 Burke (1976), Chan 与 Maa (1978), Morrison (1979), Bharath-Kumar (1980) 等研究了各种离散时间排队模型的稳态指标和计算方法. 连续时间排队中大量经典结果, 平行地扩展到离散时间模型, 并开始将离散时间排队应用于计算机通信网络性能分析. 文献 (Kobayashi et al., 1977) 是一篇关于离散时间排队在计算机通信领域应用的长篇综述, 指出离散时间排队更适合于计算机网络的建模和分析, 这篇综述对离散时间排队的研究和应用起到了积极的推动作用. 迄今, 已经建立了较完整的离散时间排队理论体系. 然而, 与连续时间排队情况不同, 离散时间排队的著作只出版了少数几本. 关于离散时间排队早期结果的一个系统论述如文献 (Hunter, 1983, 卷 2, 第 9 章). 本书 §1.2 中关于入口性质的描述, 主要来源于 Hunter (1983). Gravey 与 Hebuterne (1992) 引入并研究了各种各样的入口行为