



国家自然科学基金研究专著
NATIONAL NATURAL SCIENCE FOUNDATION OF CHINA



休假随机服务系统

田乃硕 著



北京大学出版社



国家自然科学基金研究专著
NATIONAL NATURAL SCIENCE FOUNDATION OF CHINA

休假随机服务系统

田乃硕 著

北京大学出版社

· 北京 ·

图书在版编目(CIP)数据

休假随机服务系统/田乃硕著. —北京:北京大学出版社,2001

ISBN 7-301-04920-X

I. 休… II. 田… III. 排队论—研究 IV. 0226

中国版本图书馆 CIP 数据核字(2001)第 17870 号

书 名: 休假随机服务系统

著作责任者: 田乃硕 著

责任编辑: 刘 勇

标准书号: ISBN 7-301-04920-X/O · 503

出版者: 北京大学出版社

地址: 北京市海淀区中关村北京大学校内 100871

网 址: <http://cbs.pku.edu.cn/cbs.htm>

电 话: 出版部 62752015 发行部 62754140 理科编辑室 62752021

电子信箱: zpup@pup.pku.edu.cn

排 版 者: 北京因温特有限公司

印 刷 者: 北京大学印刷厂

发 行 者: 北京大学出版社

经 销 者: 新华书店

850×1168 32 开本 11.875 印张 290 千字

2001 年 5 月第 1 版 2001 年 5 月第 1 次印刷

印 数: 0001—2000 册

定 价: 24.00 元

内 容 简 介

休假随机服务系统是经典排队论的最新发展,在一系列高新技术领域有重要应用。本书系统地介绍了位相型分布理论,对国内外近 20 年来在休假排队领域的研究成果给出了统一的、严谨的论述,建立了以随机分解为核心的基本理论体系。书中大量的内容是作者近年来的研究成果,反映了休假随机服务系统的研究概貌。全书共分五章,内容包括:位相型分布、 $M/G/1$ 型休假排队、 $GI/M/1$ 型休假排队、多服务台休假排队系统、休假排队在计算机通讯网络中的应用等。本书叙述深入浅出,说理清楚、论证严谨,注意系统性、先进性和实用性。

本书可供综合大学、高等理工大学随机运筹、应用概率等专业研究生、教师及研究人员阅读;也可供电讯、计算机、交通、管理等领域的工程技术人员参考。

前 言

经典随机服务系统理论(也称排队论),源于 Erlang 关于电话服务的研究,第二次世界大战以后得到迅猛发展,成为随机运筹学与应用概率论中最有活力的研究课题。它不仅建立了较完备的理论体系,而且在军事、生产、经济、管理、交通等领域得到广泛应用。

20 世纪中期以来,随着计算机通讯网络、柔性制造系统(FMS)、异步转换模式(ATM)等高新技术领域的发展,提出了大量的复杂系统设计和控制问题。这些系统的行为通常依赖于随状态而变化的参数,经典排队模型在处理这类问题时表现出极大的局限性。休假(vacation)排队研究正是在这种背景下始于 20 世纪 70 年代。作为经典排队系统的推广,休假排队中允许服务台采取各种在某些时候不接待顾客的策略,这些暂时中断服务的时间(通常是随机变量)统称为休假。导致暂时服务中断的理由可以有多种多样的解释。例如:对服务设施进行维修、保养或能量补充;为提高系统经济效益,在相对清闲时转而从事其他工作。在计算机通讯网络中,把多路复用程序处理某项特定工作的时间,看成对其他类型工作需求的休假;在交通问题研究中,把堵车或机场关闭时间视为服务员休假等等。一方面,休假排队反映了服务可能发生中断这一客观事实;另一方面,各种休假策略为系统的优化设计和过程控制提供了极大的灵活性。因此,休假排队研究受到广泛关注,并迅速成为随机运筹学的一个研究热点。迄今,已经形成了以随机分解(stochastic decomposition)为特色的基本理论和方法,取得丰富研究成果,并在各种高新技术领域得到卓有成效的应用。

本书的目的,不仅是全面介绍国内外休假排队研究的主要成果,更重要的是试图以随机分解为核心建立休假排队的完整理论

框架。该领域的许多成果是用各种不同处理方法得到的,并且散见在广泛的文献中,国内外还没有一本以建立完整理论体系为目标的专著。1991年日本学者 Takagi 的《排队分析》(Queueing Analysis)一书,以一章的篇幅介绍了各种 $M/G/1$ 型休假排队系统。该书的目的是向计算机系统工程技术人员提供指标分析的工具。另一部相关的著作是美国的 Dshalalow(1997)主编的《排队论前沿领域》(Frontiers in Queueing),把休假排队列入“参数依赖于状态”的类别加以介绍,主要内容仍是 $M/G/1$ 型休假排队。与这两本书不同,本书致力于理论框架的构建,不仅包括 $M/G/1$ 型休假排队,也包括 $GI/M/1$ 型休假排队和多服务台休假排队的理论和成果。

全书共分五章。第一章介绍位相型(PH)分布。近二十年,为适应多层次、变动因素、随机环境等系统分析的需要,随机模型研究已经从主要依赖指数分布发展到了广泛使用 PH 分布的新阶段。本书中的理论和方法也建立在 PH 分布之上。然而,国内还未出版过关于 PH 分布的著作。本章部分内容取材于 Neuts(1981)《随机模型中的矩阵几何解》一书,补充了 PH 分布理论的最新研究成果,改进了原书中若干定理的证明,使处理方法更具系统性。第二章讲述 $M/G/1$ 型休假排队,包括国内外广泛研究的多重休假、单重休假、启动时间、闸门服务、限量服务、减量服务等各种休假策略的模型。在文献中,这些模型的分析使用了极不相同的方法。本书把他们归结为空竭服务和非空竭服务两个基本模型,建立了统一的随机分解理论。本章中多级适应性休假系统及离散时间休假排队是作者的研究成果,它们包含国内外已研究的部分休假模型为其特例。第三章讲述的 $GI/M/1$ 型休假排队主要是作者的工作。1989年,作者在矩阵几何解方法中发展了一系列新的技巧,并用以处理各种 $GI/M/1$ 型休假系统。平行于 $M/G/1$,建立了 $GI/M/1$ 型休假排队的随机分解理论。第四章研究各种多服务台休假排队系统。在多服务台情况下,休假策略更加灵活多变,系统分析也

更困难。在休假排队研究二十余年的历史进程中,涉及多服务台模型的工作寥寥无几。近几年,作者在各种休假策略的 $M/M/c$, $GI/M/c$ 系统研究中取得了一系列进展。不但给出了稳态指标的详尽刻画,并提出了条件随机分解概念,为确立多服务台休假排队的理论体系奠定了基础。第四章的部分内容及全书各章中关于随机分解的 PH 封闭性等结果,都是首次发表的。最后,第五章简要介绍了休假排队在计算机通讯网络、ATM 等领域的若干应用。

1995 年以来,作者关于休假排队的研究连续得到国家自然科学基金资助,本书正是在这些研究成果的基础上完成的。

阅读本书需要矩阵分析和应用随机过程的基础知识。当然,最好还应对经典随机服务系统理论有一定了解。本书可供随机运筹领域研究人员、高校教师及相关专业研究生阅读,也可供电讯、计算机、交通、管理等领域的工程技术人员参考。

中科院应用数学所徐光辉研究员,是作者学习和从事排队系统研究的老师,他不仅指导我进入这个丰富多彩的领域,还在具体工作中给予许多鼓励和帮助。中科院应用数学所曹晋华、程侃两位研究员,也对作者的工作给予诸多关怀和支持。中科院自动化所博士后李泉林、南开大学数学所博士后岳德权曾参加作者主持的休假排队讨论班,并参与部分研究工作。作者谨向他们表示衷心的感谢。

由于水平所限,不当之处在所难免,欢迎批评指正。

田 乃 硕

1999 年春节于

秦皇岛燕山大学

· 数 · 理 · 科 · 学 · 系 · 列 ·



国家自然科学基金研究成果专著出版基金资助

目 录

前言	(1)
第一章 位相型分布	(1)
§ 1.1 引言	(1)
§ 1.2 矩阵分析准备知识	(2)
1.2.1 矩阵指数函数	(2)
1.2.2 Kronecker 乘积	(5)
§ 1.3 PH 分布与不可约表示	(7)
1.3.1 连续型 PH 分布	(7)
1.3.2 离散型 PH 分布	(10)
1.3.3 不可约表示	(12)
§ 1.4 封闭性	(14)
1.4.1 卷积与剩余寿命	(14)
1.4.2 PH 分布的混合	(17)
1.4.3 极值分布	(22)
1.4.4 随机化	(24)
§ 1.5 稠密性和渐近指数性	(27)
1.5.1 PH 分布的稠密性	(27)
1.5.2 渐近指数性	(29)
§ 1.6 PH 更新过程	(30)
1.6.1 PH 计数过程	(30)
1.6.2 更新函数	(34)
第二章 M/G/1 型休假排队	(36)
§ 2.1 休假排队系统	(36)

2.1.1	历史背景	(36)
2.1.2	随机分解	(39)
§ 2.2	经典 M/G/1 排队	(41)
2.2.1	模型及嵌入 Markov 链	(41)
2.2.2	队长分布	(43)
2.2.3	等待时间与忙期分析	(47)
§ 2.3	空竭服务 M/G/1 休假模型	(49)
2.3.1	M/G/1 的边界状态变体	(49)
2.3.2	随机分解定理	(53)
§ 2.4	多重休假 M/G/1 排队	(56)
2.4.1	模型的描述	(56)
2.4.2	随机分解与 PH 封闭性	(58)
2.4.3	忙期和忙循环	(62)
§ 2.5	单重休假 M/G/1 排队	(63)
2.5.1	单重休假模型	(63)
2.5.2	随机分解与 PH 封闭性	(64)
2.5.3	忙期分析	(70)
§ 2.6	多级适应性休假的 M/G/1 排队	(72)
2.6.1	多级适应性休假模型	(72)
2.6.2	随机分解与 PH 封闭性	(74)
2.6.3	连续时间过程	(78)
2.6.4	忙期分析	(81)
§ 2.7	启动时间的 M/G/1 排队	(83)
2.7.1	模型和随机分解	(83)
2.7.2	PH 封闭性和忙期分析	(86)
2.7.3	相关模型	(88)
§ 2.8	N -策略 M/G/1 排队	(91)
2.8.1	模型和队长	(91)
2.8.2	忙期和等待时间	(93)
2.8.3	各种相关模型	(97)
§ 2.9	闸门服务 M/G/1 型休假排队	(99)
2.9.1	非空竭服务模型与再生循环方法	(99)

2.9.2	闸门服务与多重休假系统	(103)
2.9.3	闸门服务与单重休假系统	(107)
2.9.4	Bernoulli 闸门服务系统	(110)
§ 2.10	限量服务 M/G/1 型休假排队	(114)
2.10.1	纯限量服务休假系统	(114)
2.10.2	G-限量服务休假系统	(116)
2.10.3	E-限量服务休假系统	(123)
2.10.4	B-限量服务休假系统	(128)
2.10.5	T-限量服务休假系统	(132)
§ 2.11	减量服务 M/G/1 型休假排队	(137)
2.11.1	纯减量服务休假系统	(137)
2.11.2	一般减量服务休假系统	(140)
2.11.3	Bernoulli 减量服务休假系统	(145)
§ 2.12	Geom/G/1 型离散时间休假排队	(148)
2.12.1	Geom/G/1 型离散时间排队	(148)
2.12.2	多级适应性休假 Geom/G/1 系统	(152)
2.12.3	队长随机分解与 PH 封闭性	(154)
2.12.4	等待时间	(158)
2.12.5	忙期分析、特例	(162)

第三章 GI/M/1 型休假排队 (166)

§ 3.1	GI/M/1 型结构矩阵	(166)
3.1.1	经典 GI/M/1 排队	(166)
3.1.2	矩阵几何解	(170)
3.1.3	GI/PH/1 排队	(175)
§ 3.2	多重指数休假的 GI/M/1 排队	(180)
3.2.1	模型和率阵	(180)
3.2.2	队长和等待时间的随机分解	(185)
3.2.3	连续时间队长过程	(191)
§ 3.3	单重指数休假的 GI/M/1 排队	(194)
3.3.1	模型和嵌入 Markov 链	(194)

3.3.2	状态分类	(197)
3.3.3	随机分解和 PH 结构	(201)
§ 3.4	位相型休假的 GI/M/1 排队	(208)
3.4.1	模型和嵌入 Markov 链	(208)
3.4.2	率阵和稳定条件	(212)
3.4.3	随机分解和 PH 结构	(216)
§ 3.5	位相型启动时间的 GI/M/1 排队	(223)
3.5.1	嵌入 Markov 链	(223)
3.5.2	稳态分布与随机分解	(226)
§ 3.6	N -策略 GI/M/1 排队	(231)
3.6.1	稳态队长及随机分解	(231)
3.6.2	等待时间、特例	(238)
§ 3.7	GI/Geom/1 离散时间休假排队	(240)
3.7.1	GI/Geom/1 排队系统	(240)
3.7.2	多重休假模型	(243)
3.7.3	随机分解结果	(246)

第四章 多服务台休假排队系统 (252)

§ 4.1	引言	(252)
4.1.1	历史和现状	(252)
4.1.2	休假策略分类	(253)
4.1.3	条件随机分解	(255)
§ 4.2	理论和方法的准备	(256)
4.2.1	拟生灭过程	(256)
4.2.2	矩阵几何解的条件随机分解	(260)
§ 4.3	同步休假的 M/M/c 排队	(263)
4.3.1	同步休假的统一模型	(263)
4.3.2	率阵和稳态队长分布	(266)
4.3.3	等待时间及其 PH 结构	(273)
4.3.4	条件随机分解	(277)
4.3.5	若干特例	(281)

§ 4.4	同步 N -策略休假的 $M/M/c$ 排队	(283)
4.4.1	N -策略和多重休假模型	(283)
4.4.2	队长及随机分解	(285)
4.4.3	条件等待时间的随机分解	(290)
§ 4.5	异步休假的 $M/M/c$ 排队	(292)
4.5.1	异步休假的统一模型	(292)
4.5.2	率阵 R 及其算法	(297)
4.5.3	稳态指标及条件随机分解	(300)
4.5.4	若干例子	(305)
§ 4.6	同步休假的 $GI/M/c$ 排队	(308)
4.6.1	模型及结构矩阵	(308)
4.6.2	稳态队长及其随机分解	(312)
4.6.3	等待时间及随机分解	(316)
第五章 休假排队在计算机通讯网络中的应用		(321)
§ 5.1	数据存取调度与分时系统	(321)
5.1.1	磁鼓模型	(321)
5.1.2	分时多址连接	(324)
§ 5.2	信道与磁盘子系统	(327)
5.2.1	模型的描述	(327)
5.2.2	广义服务时间分布	(329)
5.2.3	指标分析	(333)
§ 5.3	环形局域网性能分析	(336)
5.3.1	环形网络和规约	(336)
5.3.2	广义服务时间	(338)
5.3.3	休假参数的逼近	(341)
5.3.4	网络性能指标	(344)
§ 5.4	ATM 网络的虚通道分析	(347)
5.4.1	ATM 网络与虚通道连接	(347)
5.4.2	模型及指标分析	(350)
参考文献		(355)

第一章 位相型分布

§ 1.1 引言

20 世纪 70 年代以来,位相型(phase type,简记 PH)分布成为排队、存储、可靠性及各种相关随机模型分析的重要工具. 取代指数分布的特殊地位,随机模型分析发展到了 PH 分布的新阶段. 本章系统地介绍 PH 分布的基本理论和技巧.

熟知,对各类随机模型进行解析处理的主要障碍,是条件概率引起的高度复杂性. 经典随机模型理论中广泛使用指数随机变量,通过指数分布的无后效性绕过这一困难. 以指数分布为核心的经典随机模型理论见[1]~[6]. 但是,这样一来不仅大大缩小了所研究模型的范围,也极大地降低了模型的实际应用价值. 假定服务时间、部件寿命等服从指数分布,其不合理性是显而易见的. 一方面,离开指数分布,随机模型的定量分析面临实质困难;另一方面,在大多数情况下,假定所研究的随机变量服从指数分布又明显违背客观事实. 这一直是随机模型分析中的突出矛盾.

寻求保持指数分布易于进行解析处理的优点,又能包罗更广的新分布类的努力,可以追溯到 Erlang 的早期工作. 他从指数分布出发构造了位相型 Erlang 分布. 在其后漫长的岁月里,许多作者尝试过各种混合型分布,例如超指数分布、广义 Erlang 分布等. 所有这些分布都具有有理的 Laplace-Stieltjes 变换(LST),这就推动了 Cox 在一系列工作中发展了有理变换分布类的理论. 但是,这一分布类又过于宽泛,在很大程度上失去了指数分布易于处理的优点. PH 分布正是有理变换分布类中保持了指数分布特色的一个子类.

PH 分布是一个有限状态 Markov 过程吸收时间的分布. Jensen(1954)首先把这一分布用于一类经济模型,但他只作为符合模型结构的一个随机分布来使用,未能提供有力的处理方法.使 PH 分布成为有力工具的关键性工作是由 Neuts(1975)发展的矩阵分析处理技巧. PH 分布函数被表成矩阵指数形式,相当于指数分布从数值参数到矩阵参数的一种推广.其 LST 和各阶矩也都有简洁的矩阵表示,显示了与指数分布间的深刻类比. PH 分布类具有良好的封闭性,含有 PH 变量的各类随机模型,其指标通常仍服从 PH 分布,从而有简明的矩阵表示.使用这种矩阵分析方法,以简洁优美的形式统一并推广了分散在文献中表面上极不相同的大量结果,这些结果通常是在有微小差别的各种混合型分布假设下得到的.这就显示了 PH 分布在保持指数分布特性上的独特魅力.

使 PH 分布得到普遍重视的另一个原因,是它能更好地适应各种复杂的应用背景.当代高新技术的发展,要求在随机模型的分析中考虑多层次和变动因素等复杂状况.例如波动到达率和服务率的排队系统,带有休假或可修机制的服务台,有折旧行为或在变动环境中运行的部件,随季节变化而改变输入和库容的存储系统等等. PH 分布较好地适应了这类复杂随机模型的描述和分析.例如:假定部件寿命服从 PH 分布,可将它的各个位相解释为折旧的不同阶段或随机环境的不同状态.

最后,PH 分布的算法意义也是不容忽视的.含有 PH 分布的模型分析中出现的各种矩阵公式,通常都有明确的概率意义. PH 分布甚至使随机模型分析中的一些复杂数值积分转化成简单的矩阵运算,从而在算法上更容易实现.

§ 1.2 矩阵分析准备知识

1.2.1 矩阵指数函数

作为 PH 分布的准备知识,本节简要汇集若干矩阵分析结果.

它们是涉及 Markov 链问题时经常使用的. 对文献中容易找到的结果, 只叙述结论而不再证明.

设 $f(z)$ 是解析函数, 则在圆 $|z| < r$ 上可展成幂级数

$$f(z) = \sum_{n=0}^{\infty} a_n z^n.$$

设 A 是 m 阶方阵, 有特征值 $\eta_j (j=1, \dots, m)$, A 的谱半径定义为:

$$\text{SP}(A) = \max\{|\eta_1|, \dots, |\eta_m|\} < r.$$

进一步我们可定义矩阵函数 $f(A)$. 若 A 可对角化, 有可逆矩阵 H 使 $A = H^{-1} \Lambda H$, $\Lambda = \text{diag}(\eta_1, \dots, \eta_m)$, 定义

$$f(A) = H^{-1} f(\Lambda) H,$$

其中 $f(\Lambda) = \text{diag}(f(\eta_1), \dots, f(\eta_m))$. 若 A 不能对角化, 存在可对角化的矩阵列 $\{A_n\}$, A_n 收敛于 A , 这时定义

$$f(A) = \lim_{n \rightarrow \infty} f(A_n).$$

可以证明上述定义是相容的, 极限 $f(A)$ 不依赖于逼近 A 的可对角化矩阵列的选取.

本书中大量应用下列两个矩阵函数: 一个是

$$(I - A)^{-n} = \sum_{k=0}^{\infty} \binom{n+k-1}{k} A^k, \quad \text{SP}(A) < 1;$$

另一个是矩阵指数函数

$$\exp(A) = \sum_{k=0}^{\infty} \frac{1}{k!} A^k,$$

后者对任何方阵 A 都收敛. 一般地, 我们熟悉的解析函数均可扩展到矩阵变元, 除非幂级数表达式本质上依赖于乘法的可换性或变量的可逆性.

今后, 设 I_m 表 m 阶单位阵, e_m 表 $m \times 1$ 列向量, 它的所有分量都是 1. 不易发生混淆时将省去下标. 矩阵指数函数的下列性质是常用的.

若 A 与 B 可换, 即 $AB = BA$, 则

$$\exp(A + B) = \exp(A) \exp(B).$$

特别地,若 $B = \lambda I$, 则

$$e^{\lambda t} \exp(A) = \exp(\lambda I + A).$$

对任何方阵 A , 都有

$$\frac{d}{dt} \exp(At) = A \exp(At) = \exp(At) A,$$

并且 $\exp(At)$ 是微分方程

$$\frac{d}{dt} V(t) = V(t) A, \quad V(0) = I$$

的惟一解. 若 A 是可逆方阵, 成立下列分部积分公式

$$\begin{aligned} \int_0^x f(t) \exp(At) dt &= f(x) \exp(Ax) A^{-1} - f(0) A^{-1} \\ &\quad - A^{-1} \int_0^x f'(t) \exp(At) dt. \end{aligned}$$

当 $\operatorname{Re}(s) \geq 0$ 时, $\exp(At)$ 的 Laplace 变换是

$$\int_0^\infty e^{-st} \exp(At) dt = \int_0^\infty \exp[-(sI - A)t] dt = (sI - A)^{-1},$$

其中假定矩阵 A 的逆阵存在. 类似的积分公式在 A 不可逆时不再成立. 作为例子, 若 Q 是一个有限生成元, 我们来计算

$$M(x) = \int_0^x \exp(Qt) dt.$$

$M(x)$ 的元素 $M_{ij}(x)$ 是以 Q 为生成元的 Markov 过程从状态 i 出发, 在 $(0, x)$ 内逗留在状态 j 上的条件平均时间. 设 π 是 Q 的平稳概率向量, 满足 $\pi Q = 0, \pi e = 1$, 则 $e\pi - Q$ 是非奇异的. 因 $Qe = 0$, 我们有

$$\begin{aligned} M(x)(e\pi - Q) &= \int_0^x \exp(Qu)(e\pi - Q) du \\ &= xe\pi - \int_0^x \exp(Qu) Q du \\ &= xe\pi - [\exp(Qx) - I]. \end{aligned}$$

注意到 $e\pi = e\pi(e\pi - Q)$, 给出

$$M(x) = \int_0^x \exp(Qu) du = xe\pi + [I - \exp(Qx)](e\pi - Q)^{-1}.$$