排队论

——基础与分析技术

唐应辉　唐小我　著

科学出版社
www.sciencep.com

电子科技大学研究生系列教材建设项目

排　队　论

——基础与分析技术

唐应辉　唐小我　著

科　学　出　版　社

北　京

内 容 简 介

全书内容共分为10章，系统地介绍了排队系统的基础理论，重点阐述了几种典型排队系统的瞬态和稳态性质，以及基本的分析方法和技术. 第2章和第3章介绍了无限源和有限源的简单排队系统，第4章在传统分析的基础上，阐述了分析 $M/G/1/\infty$ 型瞬态性质的又一种新思路和技巧，第5章和第6章分别介绍了一般到达的典型排队系统，第7章介绍了一些特殊排队系统，如有优先权服务的排队系统，第8章介绍了排队系统理论的一些应用和实例，而第9章和第10章分别介绍了经典排队系统理论延伸的两个重要方面——休假排队系统与可修排队系统，并在附录中对著名的Little公式，以及 p_j^-，p_j，p_j^+ 三者的关系进行了阐述，使得全书内容更严谨和完善.

本书是作者多年来的科研积累和科研成果的总结，对在应用数学、运筹学、管理科学、计算机科学和通信工程等领域中从事相关研究的科技工作者和工程技术人员有较大的参考价值，同时也可作为应用数学、运筹学和相关专业的本科生和研究生教材及教师的教学参考书.

图书在版编目(CIP)数据

排队论：基础与分析技术/唐应辉，唐小我著. —北京：科学出版社，2006
ISBN 7-03-016588-8

Ⅰ. 排… Ⅱ. ①唐…②唐… Ⅲ. 排队论 Ⅳ. O226

中国版本图书馆CIP数据核字(2005)第145286号

责任编辑：范庆奎 / 责任校对：宋玲玲
责任印制：安春生 / 封面设计：王 浩

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
源海印刷有限责任公司 印刷
科学出版社发行 各地新华书店经销
*
2006年1月第 一 版 开本：B5(720×1000)
2006年1月第一次印刷 印张：17 1/2
印数：1—3 000 字数：331 000

定价：40.00元

(如有印装质量问题，我社负责调换〈环伟〉)

前　言

排队论(Queueing Theory)又名**随机服务系统理论**,是研究拥挤现象的一门数学学科,它通过研究各种服务系统在排队等待中的概率特性,来解决系统的最优设计和最优控制. 排队论是运筹学的重要分支,也是应用概率的重要分支,所研究的问题有很强的实际背景,它起源于 20 世纪初丹麦电信工程师 A. K. Erlang 对电信系统的研究. 之后,经过国内外的数学家和运筹学家的努力,排队论已是一门成熟的理论,其文献数以千计,特别是随着计算机技术的迅猛发展,排队论的科学研究更是日新月异,其应用领域也不断扩大. 目前,排队论的科学研究成果已广泛应用于通信工程、交通运输、生产与库存管理、计算机系统设计、计算机通信网络、军事作战、柔性制造系统和系统可靠性等众多领域,并取得了丰硕成果. 因此,排队论在科学技术及国民经济发展中起到了直接的重要作用,而且已成为从事通信、计算机等领域研究的专家、工程技术人员和管理人员必不可少的重要数学工具之一.

本书内容分为 10 章,系统地介绍了排队系统的经典理论,同时也注意到管理科学、计算机科学、通信等工程科学的专业实际,概括地介绍了排队系统理论的一些应用方面,并较详细地分析了计算机设计中的实时处理和管理科学中的存储问题. 本书尽管是从排队系统的经典理论开始的,但在排队系统经典理论的阐述中,仍然包含着大量的新内容,例如等待时间分布的严谨讨论与表达,$M/G/1$ 排队系统队长的瞬态分布的讨论与稳态分布的递推表达式等,而且这也是理论结构的完整和承接排队论最新成果的必需. 本书在第 9 章和第 10 章分别介绍了排队系统经典理论研究的延伸和拓展——**休假排队系统**与**可修排队系统**. 休假排队系统与可修排队系统是两类更广泛、更复杂的排队系统,提出了许多新的研究课题,是近年来很活跃的研究方面,成果层出不穷,以前的著作(教材)中少有(或没有)介绍,因此这部分内容是排队论最新成果和研究前沿的展现. 书中部分内容参考了国内外有代表性的文献和著作(教材),但也有相当内容是作者多年来的科研成果介绍和为博士生和硕士生讲授这门课程的经验总结,学术思想新颖,如泊松流充分必要条件的简洁证明,$M/G/1$ 排队系统队长的瞬态分布的直接讨论方法与稳态分布的递推表达式的获得,以及系统输出过程的嵌入更新过程分析法和离去顾客平均数的渐进展开等等,特别是关于休假排队系统与可修排队系统的有关内容,更是作者多年辛劳的结晶,反映了排队论研究的最新成果和研究前沿. 另外,本书的又一特点是书末增加了一个附录,介绍书中内容涉及到的、引用到的,但又不适合写进正文中的有关知识,特别是对著名的 Little 公式的阐述,以及对 p_j^-, p_j, p_j^+ 三者关系的

说明，使得本书的理论更严谨，结构体系更完整，相信读者阅读后对涉及到的有关问题会有更深刻的理解，也可避免产生误解，因此本书的内容既有广度也有深度.

在本书的写作过程中，力求做到内容层次分明，结构严谨，概念准确，表述清楚.理论分析由浅入深，论证严格，即使对引用到的而又没有给出证明的结果，本书也尽力指明出处，使读者有根可寻，这样内容既兼顾了理论研究的需要，又兼顾了工程应用的需要.

但是，由于作者水平有限，错误在所难免，恳请广大读者指正，以求改进.

本书的出版得到了电子科技大学研究生教材建设基金和学科建设基金的支持，也得到了四川省学术和技术带头人培养基金的支持，作者在此一并表示感谢！

作　者

2005 年 9 月于电子科技大学

常用符号说明

T_n	第 n 个顾客的到达时刻
τ_n	第 n 个到达与第 $n-1$ 个到达之间的间隔时间
$F(t)$	到达间隔时间的分布函数
λ	单位时间内的平均到达率
χ_n	第 n 个顾客所需的服务时间
$G(t)$	服务时间的分布函数
μ	单位时间内的平均服务率(忙的条件下)
ρ	系统的交通强度
$N(t)$	时刻 t 的队长(顾客数)
$N_q(t)$	时刻 t 的等待队长
N	平衡时任意时刻的队长
$\overline{N}$	平衡时的平均队长
N^-	平衡时到达的顾客看到的队长
N^+	平衡时服务完毕离开系统时留在系统中的队长
N_q	平衡时任意时刻的等待队长
$\overline{N}_q$	平衡时任意时刻的平均等待队长
$p_j(t)$	时刻 t 队长为 j 的概率
p_j	平衡时任意时刻队长为 j 的概率
p_j^-	平衡时到达顾客看到队长为 j 的概率
p_j^+	平衡时服务完毕离开系统时留在系统中队长为 j 的概率
$W(t)$	平衡时顾客的逗留时间分布函数
W	平衡时顾客的逗留时间
$\overline{W}$	顾客的平均逗留时间
$W_q(t)$	平衡时顾客的等待时间分布函数
W_q	平衡时顾客的等待时间
$\overline{W}_q$	顾客的平均等待时间
b	忙期长度
$B(t)$	忙期长度的分布函数
$\hat{\tau}$	系统的闲期长度
T_n^+	第 n 个服务完毕的顾客的离去时刻

$M(t)$	$(0,t]$内离去顾客的平均数
V	在休假排队系统中服务员的休假时间长度
$V(t)$	休假时间长度的分布函数
$\bar{V}(t)$	$1-V(t)$
X	在可修排队系统中服务台的寿命长度
$X(t)$	寿命长度的分布函数
Y	服务台故障后的修理时间长度
$Y(t)$	修理时间长度的分布函数
β	平均修理时间长度
L	拉普拉斯(Laplace)变换
LS	拉普拉斯-斯蒂尔切斯(Laplace-Stieltjes)变换
$g(s)$	相应大写$G(t)$的 LS 变换
$g^*(s)$	相应大写$G(t)$的 L 变换
$F^{(k)}(t)$	分布函数$F(t)$自身的k重卷积,$k\geqslant 1$,$F^{(0)}(t)=1$
$E[X]$	某个随机变量X的期望值
$E[X^k]$	某个随机变量X的k阶原点矩
$D[X]$	某个随机变量X的方差
$P^-(z)$	分布律$\{p_j^-\}$的母函数
$P(z)$	分布律$\{p_j\}$的母函数
$P^+(z)$	分布律$\{p_j^+\}$的母函数
$P_v(z)$	在休假排队系统中队长平稳分布的母函数
$\tilde{P}(z)$	在可修排队系统中队长平稳分布的母函数
$\mathscr{R}(s)$	复变量s的实部
$\binom{n}{k}$	n中取k的组合数,$k\leqslant n$,且规定$\binom{n}{0}=1$,$\binom{n}{k}=0\quad(k>n)$

另外,其他有关符号见出处的说明.

目　　录

第 1 章　引　　论

1.1　排队系统概述

1.1.1　排队例子及基本概念

排队是日常生活和工作中常见的现象，例如：上下班坐公共汽车，等待公共汽车的排队；顾客到商店购物形成的排队；病人到医院看病形成的排队；往售票处购票形成的排队等；另一种排队是物的排队，例如文件等待打印或发送；路口红灯下面的汽车、自行车通过十字路口. 排队现象是由两个方面构成，一方要求得到服务，另一方设法给予服务. 我们把要求得到服务的人或物(设备)统称为**顾客**，给予服务的**服务人员**或**服务机构**统称为**服务员**或**服务台**(有时服务员专指人，而服务台是指给予服务的设备). 顾客与服务台就构成一个**排队系统**，或称为**随机服务系统**，显然，缺少顾客或服务台任何一方都不会形成排队系统.

排队现象有的是以有形的形式出现，例如上下班坐公共汽车等，这种排队我们称为**有形排队**. 而有的是以无形的形式出现，例如有许多顾客同时打电话到订购处订购车票，当其中一个顾客正在通话时，其他顾客就不得不在各自的电话机旁等待，他们可能分散在各个地方，但却形成一个无形的队列等待通话，这种排队现象称为**无形排队**.

1.1.2　基本的排队系统

1) 单服务员的排队系统，其图解表示如图 1.1 所示.

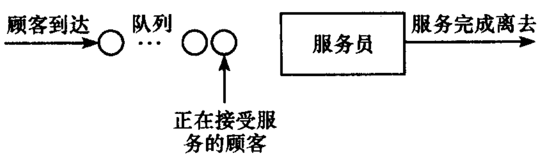

图 1.1

2) 多服务员(台)的排队系统，其图解表示如图 1.2、1.3 和 1.4 所示.

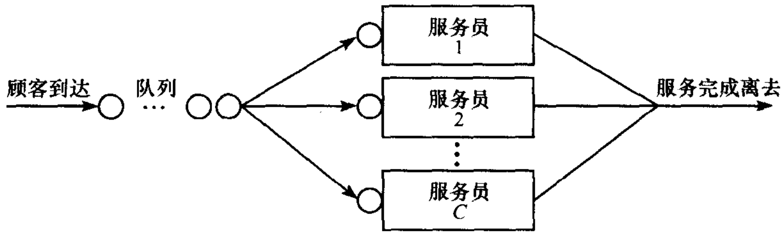

图 1.2　排成一个队列

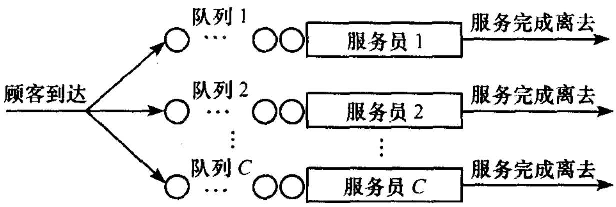

图 1.3　排成多个队列

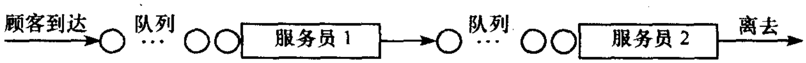

图 1.4　串联

此外，还有串并混合、网络等排队系统.

1.1.3　排队论研究的内容和目的

在各种排队系统中，随机性是它们的一个共同特性，而且起着根本性的作用. 顾客的到达间隔时间与顾客所需的服务时间中，至少有一个具有随机性，否则问题就太简单了. 排队论主要研究描述系统的一些主要指标的概率特性，分为三大部分：

1）排队系统的性态问题

研究排队系统的性态问题就是研究各种排队系统的概率规律，主要包括系统的队长（系统中的顾客数）、顾客的等待时间和逗留时间，以及忙期等的概率分布，包括它们的瞬时性质和统计平衡下的性态. 排队系统的性态问题是排队论研究的核心，是排队系统的统计推断和最优化问题的基础. 从应用方面考虑，统计平衡下的各个指标的概率性质尤其重要.

2）排队系统的统计推断

为了了解和掌握一个正在运行的排队系统的规律，就需要通过多次观测、搜集数据，然后用数理统计的方法对得到的数据进行加工处理，推断所观测的排队系统的概率规律，从而应用相应的理论成果来研究和解决该排队系统的有关问题. 排队

系统的统计推断是已有理论成果应用实际系统的基础性工作，结合排队系统的特点，发展这类特殊随机过程的统计推断方法是非常必要的.

3) 排队系统的最优化问题

排队系统的最优化包括系统的最优设计(静态最优)和已有系统的最优运行控制(动态最优)，前者是在服务系统设置之前，对未来运行的情况有所估计，使设计人员有所依据，例如电话局的规模、水库容量的大小、机场跑道数目的设计等；后者是对已有的排队系统寻求最优运行策略，例如库房领取工具，当排队领取工具的工人太多，就增设服务员，这样虽然增加了服务费用，但另一方面却减少了工人领取工具的等待时间，即增加了工人有效的生产时间，这样带来的好处可能远超过服务费用的增加. 因此，对于一个排队系统的设计或运行管理，就需要考虑顾客与服务双方的利益，以便在某种合理的指标上使系统达到最优化. 对大多数实际系统讲，若把输入看作是由客观条件决定的，不受控制(有时也可采取控制输入的手段)，则解决这种问题的关键是确定服务率或服务台数或选取顾客的服务规则或这几种量的组合，使之在某种意义下系统达到最优. 优化的指标函数可以是时间，也可以是费用或收入. 学习和应用排队论知识就是要解决客观系统的最优设计或运行管理，创造更好的经济效益和社会效益.

1.1.4 排队系统的基本组成部分

尽管排队系统是各种各样的，但从决定排队系统进程的主要因素看，它主要由三部分组成：**输入过程**、**排队规则**和**服务机构**. 下面分别加以说明.

1) 输入过程

输入过程是描述顾客来源及顾客是按怎样的规律抵达排队系统. ①**顾客总体数**：顾客的来源可能是有限的，也可能是无限的，例如工厂内发生故障待修的机器是有限的；到达窗口购票的顾客总体可以看成是无限的(因为不存在最大的限制数). ②**到达的类型**：顾客是单个到达，或是成批到达，例如工厂内发生故障待修的机器是单个到达；在库存问题中，进货看成顾客到达，就是成批到达的例子. ③**相继顾客到达的间隔时间服从什么样的概率分布，分布的参数是什么，到达的间隔时间之间是否独立**. 如果设 $T_0=0$，$T_n(n\geqslant 1)$ 表示第 n 个(批)顾客的到达时刻，则

$$T_0 = 0 < T_1 < T_2 < \cdots < T_n < T_{n+1} < \cdots$$

又令 $\tau_n=T_n-T_{n-1}$，$n\geqslant 1$，则 τ_n 表示第 n 个(批)顾客到达时刻与第 $n-1$ 个(批)顾客到达时刻之差，称序列 $\{\tau_n, n\geqslant 1\}$ 为顾客相继到达的**间隔时间序列**. 在排队论研究中，一般假定 $\{\tau_n, n\geqslant 1\}$ 相互独立、同分布. 有定长输入，即顾客是等距时间到达；最简单流输入(Poisson 流输入)，即 $\{\tau_n, n\geqslant 1\}$ 独立、同负指数分布；k 阶埃尔朗输入；超指数分布输入；几何分布输入；一般独立输入；成批输入(每次到达是一批顾客，每批的个数可以是固定的，也可以是随机的)，以及其他形式的输入等.

2）排队规则

排队规则是指服务允许不允许排队，顾客是否愿意排队. 在排队等待的情形下服务的顺序是什么，分为：①**损失制**：顾客到达时，若所有服务台均被占，服务机构又不允许顾客等待，此时该顾客就自动离去，例如通常使用的损失制电话系统. ②**等待制**：顾客到达时，若所有服务台均被占，他们就排队等待服务. 在等待制系统中，服务顺序又分为：**先到先服务**，即顾客按到达的先后顺序接受服务；**后到先服务**，例如情报系统、天气预报资料总是后到的信息越重要，要先处理；**随机服务**，即在等待的顾客中随机地挑选一个顾客进行服务，例如电话员接线就是用这种方式工作；**有优先权的服务**，即在排队等待的顾客中，某些类型的顾客具有特殊性，在服务顺序上要给予特别待遇，让他们先得到服务，例如病危人先治疗；带小孩的顾客先进站等. 优先权又分**强拆型优先权**和**非强拆型优先权**. 强拆型优先权是指这类顾客到达时，无论正在接受服务的顾客是否服务完毕，都必须立即中止服务而转为接受这类顾客并给予服务，例如医院对病危人的服务. 非强拆型优先权是指这类顾客到达时，必须等待正在接受服务的顾客服务完毕后才会得到服务. 在多个服务台的情形，顾客到达是排成一个队列，或是排成多个队列，例如在并联服务台的排队系统中. 而对于循环排队，如纺纱工管理 m 台机器，总是在各机器之间按固定路线巡回，遇到哪台机器故障就处理哪台. ③**混合制**：损失制与等待制的混合，分为队长（容量）有限的混合制系统，等待时间有限的混合制系统（等待时间≤固定的时间 t_0，否则就离去），以及逗留时间有限制的混合制系统.

3）服务机构

刻划服务机构的主要方面为：①**服务台的数目**. 在多个服务台的情形下，是串联或是并联；②**顾客所需的服务时间服从什么样的概率分布，每个顾客所需的服务时间是否相互独立，是成批服务或是单个服务等**. 常见顾客的服务时间分布有：定长分布、负指数分布、超指数分布、k 阶埃尔朗分布、几何分布、一般分布等.

由于输入过程、排队规则和服务机构的复杂多样性，形成了各种各样的排队模型，因此在研究一个排队系统之前，首先要弄清这三部分的具体内容和结构.

1.1.5　经典排队系统的符号表示

一个排队系统是由许多条件决定的，为了简明起见，在经典排队系统中，常采用 3～5 个英文字母表示一个排队系统，字母之间用斜线隔开：第一个字母表示输入分布类型，第二个字母表示服务时间的分布类型，第三个字母表示服务台的数目，第四个字母表示系统的容量，有时用第五个字母表示顾客源中的顾客数目. 例如：

$M/M/c/\infty$表示输入过程是 Poisson 流，服务时间服从负指数分布，系统有 c 个服务台平行服务（$0<c\leqslant\infty$），系统容量为无穷，于是 $M/M/c/\infty$系统是等待制

系统；

$M/G/1/\infty$表示输入过程是 Poisson 流，顾客所需的服务时间为独立、服从一般概率分布，系统中只有一个服务台，容量为无穷的等待制系统；

$GI/M/1/\infty$表示输入过程为顾客独立到达且相继到达的间隔时间服从一般概率分布，服务时间是相互独立、服从负指数分布，系统中只有一个服务台，容量为无穷的等待制系统；

$E_k/G/1/K$ 表示相继到达的间隔时间独立、服从 k 阶埃尔朗分布，服务时间为独立、服从一般概率分布，系统中只有一个服务台，容量为 $K(1\leqslant K<\infty)$的混合制系统；

$D/M/c/K$ 表示相继到达的间隔时间独立、服从定长分布，服务时间相互独立、服从负指数分布，系统中有 c 个服务台平行服务，容量为 $K(c\leqslant K<\infty)$的混合制系统；

$M^r/M/1/\infty$表示顾客是成批到达，每批到达为固定的 r 个顾客$(1\leqslant r<\infty)$，批与批的到达间隔时间独立、服从负指数分布，顾客的服务时间独立、服从负指数分布，有一个服务台，系统容量为无穷的等待制系统；

$M^X/M^r/1/\infty$表示顾客是成批到达，每批到达的数量 X 是具有某个离散型概率分布律的随机变量，批与批的到达间隔时间独立、服从负指数分布，而系统中有一个服务台，顾客是成批服务，每批为 r 个顾客，且服务时间独立、服从负指数分布，容量为无穷的等待制系统.

1.1.6 描述排队系统的主要数量指标

1）队长与等待队长

队长是指在系统中的顾客数(包括正在接受服务的顾客)，而等待队长是指系统中排队等待的顾客数，它们都是随机变量，是顾客和服务机构双方都十分关心的数量指标，应确定它们的分布及有关矩(至少是期望平均值). 显然，队长等于等待队长加上正在被服务的顾客数.

2）顾客在系统中的等待时间与逗留时间

顾客的**等待时间**是指从顾客进入系统的时刻起直到开始接受服务止这段时间，而**逗留时间**是顾客在系统中的等待时间与服务时间之和. 在假定到达与服务是彼此独立的条件下，等待时间与服务时间是相互独立的. 等待时间与逗留时间是顾客最关心的数量指标，应用中关心的是统计平衡下它们的分布及期望平均值.

3）系统的忙期与闲期

从顾客到达空闲的系统，服务立即开始，直到系统再次变为空闲，这段时间是系统连续繁忙的时间，我们称为系统的**忙期**，它反映了系统中服务员的工作强度.

与忙期对应的是系统的**闲期**，即系统连续保持空闲的时间长度. 在排队系统

中，统计平衡下忙期与闲期是交替出现的.

而**忙期循环**是指相邻的两次忙期开始的间隔时间，显然它等于当前的忙期长度与闲期长度之和.

4）输出过程

输出过程也称**离去过程**，是指接受服务完毕的顾客相继离开系统的过程. 刻画一个输出过程的主要指标是相继离去的间隔时间和在一段已知时间内离去顾客的数目，这些指标从一个侧面也反映了系统的工作效率.

此外，在不同的排队系统中，还会涉及其他数量指标，例如在损失制与混合制排队系统中，顾客的损失率及单位时间内损失的平均顾客数，在多服务台并行服务的系统中，某个时刻正在忙的服务台数目，以及系统的利用率等.

1.2 几个重要的概率分布

1.2.1 定长分布(单点分布)

定义 1.2.1 设随机变量 X 以概率 1 取常值 a，即 $P\{X=a\}=1$，则称 X 服从定长分布或单点分布. 它的概率分布函数为

$$F(t)=P\{X\leqslant t\}=\begin{cases}0, & t<a,\\ 1, & t\geqslant a.\end{cases} \tag{1.2.1}$$

1.2.2 负指数分布

定义 1.2.2 一个连续型随机变量 X，若它的分布密度函数为

$$f(t)=\begin{cases}\lambda e^{-\lambda t}, & t\geqslant 0,\\ 0, & t<0,\end{cases} \tag{1.2.2}$$

其中 $\lambda(>0)$ 为常数，则称随机变量 X 服从参数 λ 的负指数分布，其概率分布函数为

$$F(t)=\begin{cases}1-e^{-\lambda t}, & t\geqslant 0,\\ 0, & t<0.\end{cases} \tag{1.2.3}$$

k 阶原点矩 $E[X^k]=\frac{k!}{\lambda^k}(k=1,2,\cdots)$，方差 $D[X]=\frac{1}{\lambda^2}$.

引入一个无量纲的量

$$\varphi=\frac{\sqrt{D[X]}}{E[X]}, \tag{1.2.4}$$

其中，假定 $E[X]\neq 0$，称 φ 为随机变量 X 的变异系数.

显然，服从负指数分布的随机变量的变异系数 $\varphi=1$，这是一个随机变量服从负指数分布的必要条件. 如果一个随机变量的变异系数远离于 1，则可以认为该随

机变量不会服从负指数分布.

定理 1.2.1 设连续型随机变量 X 服从参数 $\lambda(>0)$ 的负指数分布,则

1) 对任意 $t\geqslant 0, s\geqslant 0$,有

$$P\{X>t+s\,|_{X>s}\}=P\{X>t\}=e^{-\lambda t};\tag{1.2.5}$$

2) 对任意一个与 X 相互独立的非负随机变量 Y, 和任意 $t\geqslant 0$, 在 $P\{X>Y\}>0$ 的条件下,有

$$P\{X>Y+t\,|_{X>Y}\}=P\{X>t\}=e^{-\lambda t}.\tag{1.2.6}$$

证明 1) 由条件概率公式,有

$$\begin{aligned}P\{X>t+s\,|_{X>s}\}&=\frac{P\{X>t+s,X>s\}}{P\{X>s\}}\\&=\frac{P\{X>t+s\}}{P\{X>s\}}=\frac{e^{-\lambda(t+s)}}{e^{-\lambda s}}=e^{-\lambda t};\end{aligned}$$

2) 同理,由条件概率公式和全概率分解,有

$$\begin{aligned}P\{X>Y+t\,|_{X>Y}\}&=\int_0^\infty P\{X>y+t\,|_{X>y}\}dP\{Y\leqslant y\}\\&=\int_0^\infty e^{-\lambda t}dP\{Y\leqslant y\}=e^{-\lambda t}.\end{aligned}$$

□

我们把负指数分布的这个性质称为“**无记忆性**”或“**无后效性**”. 它的直观意义是,如果把 X 解释为机器的使用寿命,则该机器在使用了一段时间(这段时间可以是随机时间)还没有坏的条件下,它的剩余寿命仍然服从原来参数的负指数分布,与已使用过的时间无关. 正因为负指数分布具有这个性质,在排队论和可靠性理论等领域内起着非常重要的作用. 下面证明“无记忆性”也是负指数分布的充分条件,为此,不加证明地引进数学分析中的一个结果,其证明也可见参考文献[12].

引理 1.2.1 设 $g(x)$ 是一元函数,则 $g(x)=a^x(a>0)$ 的充分必要条件是

1) $g(x)$ 是 x 的连续(或单调)函数, $g(1)\neq 0$;

2) 对任意 x、y, $g(x+y)=g(x)\cdot g(y)$.

定理 1.2.2 设 X 是取非负值的连续型随机变量,

1) 若 X 具有“无记忆性”,则 X 服从负指数分布;

2) X 服从参数 $\lambda(>0)$ 的负指数分布的充分必要条件是对任给的 $t\geqslant 0$, $E[X-t\,|_{X>t}]=\dfrac{1}{\lambda}$.

证明 1) 任取 $x>0$,令 $g(x)=P\{X>x\}$,则 $g(x)$ 是右连续且单调. 由于 X 的“无记忆性”,则对任取的 $y>0$,有

$$\begin{aligned}g(x+y)=P\{X>x+y\}&=P\{X>x+y\,|_{X>x}\}\cdot P\{X>x\}\\&=P\{X>y\}\cdot P\{X>x\}=g(x)\cdot g(y).\end{aligned}$$

显然对一切 $x\geqslant 0, g(x)\geqslant 0$，所以 $g(1)\geqslant 0$. 下面证 $g(1)>0$.

事实上，若 $g(1)=0$，则对任意正整数 n，有

$$g(1)=\left[g\left(\frac{1}{n}\right)\right]^n=0,$$

于是

$$g\left(\frac{1}{n}\right)=0.$$

令 $n\to\infty$，有 $g(0)=0$，这与 $g(0)=P\{X>0\}$ 相矛盾. 于是 $g(1)>0$ 成立. 再根据引理 1.2.1，有

$$g(x)=a^x(a>0).$$

由于 $g(x)=P\{X>x\}$ 是概率，所以 $0<a<1$（$a=0$ 或 $a=1$ 均使得 $1-g(x)$ 不是概率分布函数）. 不妨令

$$a=\mathrm{e}^{-\lambda},\quad \lambda>0,$$

则

$$g(x)=\mathrm{e}^{-\lambda x},\quad x\geqslant 0.$$

此式说明 X 服从负指数分布.

2）由负指数分布的"无记忆性"，必要性是明显的，现证充分性. 仍令

$$g(x)=P\{X>x\},\ \bar{g}(x)=1-g(x),\quad x\geqslant 0,$$

则

$$\begin{aligned}E[X-t\mid_{X>t}]&=\int_t^{\infty}(x-t)\mathrm{d}P\{X\leqslant x\mid_{X>t}\}\\&=\int_t^{\infty}(x-t)\mathrm{d}\left[\frac{\bar{g}(x)-\bar{g}(t)}{g(t)}\right]\\&=\frac{1}{g(t)}\int_t^{\infty}(x-t)\mathrm{d}\bar{g}(x)=\frac{1}{\lambda},\end{aligned}$$

于是

$$\int_t^{\infty}(x-t)\mathrm{d}\bar{g}(x)=\frac{1}{\lambda}g(t).$$

此式左边关于 t 可导，当然右边关于 t 也可导，因此

$$\frac{1}{\lambda}\frac{\mathrm{d}}{\mathrm{d}t}[g(t)]=-\int_t^{\infty}\mathrm{d}\bar{g}(x)=-g(t),$$

于是

$$g(t)=c\mathrm{e}^{-\lambda t},\quad t\geqslant 0.$$

再由 $g(0)=1$ 定出 $c=1$，即 $g(t)=\mathrm{e}^{-\lambda t}, t\geqslant 0$. 证毕.

□