纯粹数学与应用数学专著　第19号

随机服务系统

（第二版）

徐光辉　著

科学出版社

纯粹数学与应用数学专著 第19号

随机服务系统

(第二版)

徐光辉 著

科学出版社

1988

内 容 简 介

在第二版中，对第一版作了若干修改与补充. 第二版增加了第八章逼近理论.

本书概括地介绍了随机服务系统的基本理论，着重介绍了几种典型系统的瞬时性质；作者以矿山装运过程为例，通俗地介绍了解决随机服务系统的实际问题的有力工具——随机模拟方法，最后，分别阐述了随机服务系统的几个主要的应用方向. 读者只要有微积分与概率论的知识就可以阅读本书.

本书可供高等学校数学系师生以及有关研究人员和工程技术人员参考.

纯粹数学与应用数学专著　第 19 号

随 机 服 务 系 统

（第 二 版）

徐 光 辉 著

责任编辑　杨贤英

科学出版社出版

北京朝阳门内大街 137 号

中国科学院印刷厂印刷

新华书店北京发行所发行　　各地新华书店经售

*

1980 年 2 月第 一 版　　开本：850×1168　1/32

1988 年 9 月第 二 版　　印张：13 7/8

1988 年 9 月第三次印刷　　插页：精 2

印数：平 12,251—13,950　精 1—680　　字数：365,000

ISBN 7-03-000400-4/O · 114（平）

ISBN 7-03-000577-5/O · 147（精）

定价：平　装 6.40 元　布面精装 7.90 元

科技新书目：171-072(平)-073(精)

第 二 版 序

在第二版中对第一版作了若干修改与补充，如最简单流充分必要条件的新的证明，M/G/1 系统与 GI/M/n 系统中的首达时间，M/G/1 系统与 GI/M/n 系统中任意时刻队长平稳分布与嵌入马尔可夫链的平稳队长分布之间的关系等．特别是补写了新的一章——第八章，逼近理论．这是近年来发展起来的比较活跃的一个方向，但这里只能阐述它最基础的一些内容，和前七章合在一起，作为随机服务系统的基本理论介绍给广大读者．

第二版的工作是在中国科学院基金的资助下完成的，作者为此表示由衷的谢意．

徐光辉

1986 年于中国科学院应用数学研究所

第 一 版 序

随机服务系统理论在国民经济和国防建设中有着广泛的应用．它在本世纪初起源于电话话务理论的研究，以后陆续应用于陆空交通、机器管理、水库设计和可靠性理论等领域．六十年代末，随着电子计算机蓬勃发展的需要，又开始了对计算机最优设计的应用．在这将近七十年的历史进程中，随机服务系统不论在理论上还是在应用上都已有了飞速的进展，它的面貌可以说是日新月异．其文献数量，已以千计．

在我国，随机服务系统理论的研究工作是在五十年代的末期才开始发展起来的．在应用方面，主要是配合社会主义建设的需要，与电话、纺织、交通等方面的工作人员合作进行有关问题的研究与计算．在理论方面，则主要着重于几种典型系统的瞬时性质的研究．近年来，随机服务系统理论的应用范围又扩大到矿山、电讯、计算机设计等领域．我们相信，随着我国社会主义建设事业的发展，随机服务系统理论的实际应用将会日益广泛，而应用的深入又必然会进一步促进随机服务系统理论的进展．

本书试图对随机服务系统的基本理论作一概括性的介绍，同时对它几个主要的应用方向分别加以阐述．本书第一章至第七章介绍随机服务系统的基本理论，第八章介绍解决随机服务系统的实际问题的有力工具——随机模拟方法，并以矿山装运过程为例具体地讲述它的应用，第九章介绍随机服务系统的一个重要应用领域——计算机最优设计，第十章介绍随机服务系统的其它应用，包括可靠性问题、水库问题、存储问题、卫星通讯问题等．第一章至第六章的初稿，作者曾于 1964 年在中国科学技术大学应用数学系兼课时作为讲义讲授过，现在进一步作了修改补充，并增添了后四章．本书可供高等学校数学系师生及电讯、计算机、矿山、交通

等领域的工程技术人员参考，阅读本书只需微积分与概率论的基本知识.

本书包括了研究室的同志及作者本人已经发表和尚未发表的部分研究成果，如最简单流与独立负指数分布的等价性的严格证明、GI/M/n 系统的瞬时性质、GI/M/n 系统的 k 阶忙期、到达间隔依赖于队长的系统、矿山装运过程的随机模拟、计算机存储器的性能分析等. 他们有关的工作均已列入书后的“参考文献”.另外，最后的“文献附记”中指出了各章节取材的主要来源，以及某些历史概况和现状，可供读者参考.

在本书的写作过程中，研究室的同志给予了不少的指导和帮助，作者谨向他们表示衷心的感谢. 此外，作者在中国科学技术大学兼课期间，杨德庄同志担任辅导，他对本书的初稿提出过很多宝贵意见，作者也谨向他表示深切的谢意.

随机服务系统理论又名排队论，有人也称之为公用事业理论中的数学方法，我们认为用随机服务系统理论这个名称更为恰当，因为这既指出了它所研究的各种问题可以用服务系统这个概念来加以统一的共性，又强调了它从数学研究的范畴来说具有随机性的特性. 而各种服务系统不仅包括有排队等待的(等待制)，也包括无法排队的(损失制)，因此统称为排队论不尽适宜.

由于作者水平所限，错误在所难免，欢迎广大读者批评指正，以求改进.

徐光辉

1977 年于中国科学院数学研究所

常 用 符 号 表

	输入间隔	服务时间	队 长	等待时间	忙 期
随机变量	t	v	q	w	d
分布函数	$A(x)$	$B(x)$	$\{p_j\}$ (或$\{\pi_j\}$)	$W(x)$	$D(x)$
数学期望	$\frac{1}{\lambda}$ (或 α)	$\frac{1}{\mu}$ (或 β)	$\bar{Q}$	$\bar{W}$	$\bar{D}$
二 阶 矩	$\alpha^{(2)}$	$\beta^{(2)}$			
L-S 变换 (或母函数)	$A^*(s)$	$B^*(s)$	$Q(z)$ (母函数)	$W^*(s)$	$D^*(s)$
分布函数的 j 重卷积	$A^{(j)}(x)$	$B^{(j)}(x)$			

目　　录

第一章 引 论

§1. 概 述

1. 例子 日常生活中人们经常遇到各种各样的服务系统．如上下班坐公共汽车，汽车与乘客就构成一个服务系统．到商店买东西，售货员与顾客就构成一个服务系统．

还有许多场合，服务系统的构成没有那么明显．例如有很多旅客想打电话到火车站定购车票，当其中一个旅客正在通话时，其它旅客就不得不在各自的电话机前等待．虽然车站定票处与这些旅客可能分散在城市的各个地区，但是他们却构成一个服务系统．

在服务系统中，要求服务的"顾客"可以是人，也可以是某种物品．如在有自动机床的工厂里，因故障而停止运转的机器等待工人去修理，在此服务系统中，服务机构是修理工人，而要求服务的"顾客"就是待修的机器．这种服务系统的例子，还可以举出很多．例如码头的船只等待装卸；执行空战任务归来而急需降落的飞机因跑道不空在空中盘旋等待；通过水库调度来控制水的泄放等等．

在上述各种服务系统中，顾客到来的时刻与进行服务的时间都随不同的时机与条件而变化，因此服务系统的状况也是随机的，即随各种时机与条件而波动．所以，我们在考察这些系统时，为了强调其随机性，就称之为随机服务系统．

从上面的例子可以看出，各种随机服务系统具有下列共同组成部分：

(i) 输入过程：就是指各种类型的"顾客"按怎样的规律到来．这些顾客可以是公共汽车的乘客、商店的顾客、打电话的用户、损坏待修的机器、等待装卸的船只、急需降落的飞机、水库上游的来水等等．他们陆续到来，要求服务．

(ii) 排队规则：就是指到来的顾客按怎样的规定次序接受服务。例如，公共汽车的乘客按到达先后次序上车，电话用户按随机次序接通电话，而急需降落的飞机却应按它们的迫切程度降落。

(iii) 服务机构：就是指同一时刻有多少服务设备可接纳顾客，每一设备可接纳多少顾客，以及每一顾客服务多少时间。例如商店有两个售货员可以接待顾客，机场有三条跑道可供飞机降落，有六辆公共汽车装载乘客，每辆最多坐 30 人等等。购货时间、降落时间、坐车时间就是这些顾客的服务时间。

正由于任何随机服务系统都有上述共同组成部分，因此，才有可能建立处理这些问题的统一理论——随机服务系统理论。

2. 研究的目的与方法 从上述例子可以看出，服务机构过小，便不能满足顾客的需要，并使服务质量降低。因此对顾客来说，服务机构愈大，他们就愈方便。

但是，服务机构大了，人力物力的开支也就相应增多，有时就会造成不必要的浪费。因此就产生顾客需要与机构经济之间的协调问题。

在有些情况下，可以在服务机构设置以后，根据顾客到来的情况加以调整。但在另外一些情形，必须在服务机构设置之前就根据顾客输入与服务过程对系统未来的进程作出正确的估计，以便使设计工作有所依据。例如电话局的设计、机场跑道的设计、计算机的设计等就是如此。

如何正确地了解系统的性态，以便最终能合理地设计与控制随机服务系统，使得它既能满足顾客的需要，又使机构的花费最为经济，这就是随机服务系统理论的研究目的。

随机因素在随机服务系统中起着根本性的影响。顾客到来的时刻一般是无法事先规定的，到来顾客要求服务的时间也是随不同顾客而异的。例如运转的机器不知道在何时会发生故障，发生故障后修复的时间也随故障的性质和工人的技术水平而各异。因此，在研究随机服务系统时，自然就要采用研究随机现象规律性的一门数学分支——概率论的方法。

3. 随机服务系统的三个组成部分 前面已经提及随机服务系统的三个组成部分：输入过程、排队规则和服务机构. 现在分别对它们作比较详细的描述.

(i) 输入过程：可以有各式各样的输入过程，例如

1）定长输入：顾客有规则地等距到达，如每隔时间 α 到达一个顾客，此时相继顾客到达间隔 $\boldsymbol{t}$ 的分布函数 $A(t)$ 为

$$A(t) = P\{\boldsymbol{t} \leqslant t\} = \begin{cases} 1, & t \geqslant \alpha; \\ 0, & t < \alpha. \end{cases} \tag{1}$$

产品通过传送带进入包装箱就是这种输入的例子.

2）最简单流：或者称为普阿松（Poisson）输入.

满足下列四个条件的输入称为最简单流：

a）平稳性：在区间 $[a, a+t)$ 内有 k 个顾客到来的概率与 a 无关，而只与 t, k 有关. 记此概率为 $v_k(t)$.

b）无后效性：不相交区间内到达的顾客数是相互独立的.

c）普通性：令 $\psi(t)$ 表示长为 t 的区间内至少到达两个顾客的概率，则

$$\psi(t) = o(t), \quad t \to 0.$$

d）有限性：任意有限区间内到达有限个顾客的概率为 1. 因而

$$\sum_{k=0}^{\infty} v_k(t) = 1.$$

这种输入的应用最为广泛，并且最容易处理，因此我们在下一节将予以详细讨论. 在该处将证明：对最简单流，长为 t 的时间内到达 k 个顾客的概率 $v_k(t)$ 遵从普阿松分布，即

$$v_k(t) = e^{-\lambda t} \cdot \frac{(\lambda t)^k}{k!}, \quad k = 0, 1, 2, \cdots, \tag{2}$$

其中 $\lambda > 0$ 为一常数. 令第 i 个顾客到达的时刻为 $\boldsymbol{\tau}_i (i = 1, 2, \cdots)$, $\boldsymbol{\tau}_0 = 0$, 并令 $\boldsymbol{t}_i = \boldsymbol{\tau}_i - \boldsymbol{\tau}_{i-1}$, $i = 1, 2, \cdots$, 则一个输入过程是最简单流的充分必要条件为：相继顾客到达间隔 $\{\boldsymbol{t}_i\}$ 是相互独立相同分布的，且其分布函数为负指数分布：

$$A(t)=\begin{cases}1-e^{-\lambda t}, & t\geqslant 0;\\ 0, & t<0.\end{cases} \tag{3}$$

因而平均到达间隔与方差分别为

$$Et=\int_0^{\infty} t\,dA(t)=\frac{1}{\lambda} \tag{4}$$

与

$$\sigma_t^2=\frac{1}{\lambda^2}. \tag{5}$$

3）爱尔朗（Erlang）输入 E_k：它的到达间隔相互独立，具有相同的 k 阶（k 为正整数）爱尔朗分布函数：

$$A(t)=\begin{cases}1-\sum\limits_{i=0}^{k-1} e^{-\lambda t}\dfrac{(\lambda t)^i}{i!}, & t\geqslant 0;\\ 0, & t<0,\end{cases} \tag{6}$$

其中 $\lambda>0$ 为一常数．易知其密度函数为

$$a(t)=\frac{\lambda(\lambda t)^{k-1}}{(k-1)!}e^{-\lambda t},\ t\geqslant 0, \tag{7}$$

而平均到达间隔与方差分别为

$$Et=\int_0^{\infty} t a(t)dt=\frac{k}{\lambda} \tag{8}$$

与

$$\sigma_t^2=\frac{k}{\lambda^2}. \tag{9}$$

当 $k=1$ 时的一阶爱尔朗输入 E_1 就是最简单流．

4）超指数输入 H_k：它的到达间隔相互独立，具有相同的 k 阶（k 为正整数）超指数分布函数：

$$A(t)=\begin{cases}1-\sum\limits_{i=1}^{k}\alpha_i e^{-\lambda_i t}, & t\geqslant 0;\\ 0, & t<0,\end{cases} \tag{10}$$

其中 $\lambda_i>0,\alpha_i>0\ (i=1,2,\cdots,k)$ 均为常数，且 $\sum\limits_{i=1}^{k}\alpha_i=1$．易知其密度函数为

$$a(t)=\sum_{i=1}^{k}\alpha_i\lambda_i e^{-\lambda_i t},\ t\geqslant 0, \tag{11}$$

而平均到达间隔与方差分别为

$$Et=\int_0^\infty ta(t)dt=\sum_{i=1}^{k}\frac{\alpha_i}{\lambda_i}=\frac{1}{\lambda} \tag{12}$$

与

$$\sigma_t^2=2\sum_{i=1}^{k}\frac{\alpha_i}{\lambda_i^2}-\frac{1}{\lambda^2}. \tag{13}$$

当 $k=1$ 时的一阶超指数输入 H_1 就是最简单流.

5）一般独立输入：它的到达间隔相互独立，相同分布．分布函数记为 $A(t)$．上面所有的输入都是一般独立输入的特例．

6）成批到达的输入：假定有一系列到达点，它们的间隔分布可以是上述的各种分布，但在每一到达点上到来的不是单独一个顾客，而是一批顾客，每批顾客的数目 $\boldsymbol{n}$ 为一随机变量，其分布为

$$P\{\boldsymbol{n}=k\}=a_k,\ k=0,1,2,\cdots. \tag{14}$$

(ii) 排队规则：

1）损失制：顾客到达时，若所有服务台均被占，该顾客就自动消失．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，他们就排成队伍，等待服务，服务次序可以采用下列各种规则：

a）先到先服务：即按到达次序接受服务．这是最通常的情形．

b）后到先服务：例如将钢板堆入仓库看成是顾客的到达，需要使用时将它们陆续取走看作是服务，则一般都是取用放在最上面的，也就是最后放上的钢板．又如在通讯系统中，最后到达的信息一般说是最有价值的，因而有时会采取“后到先服务”的方式．

c）随机服务：当服务机构得空时，在等待的顾客中随机地选取一名进行服务，也即每一等待的顾客被选到的概率相同．

d）优先权服务：如码头上载有重要物资的船只先行装卸；电报分普通电报、加急电报；长途电话比市内电话优先，甚至可中断

市内电话的通话.

e) 多个（n 个）服务台的情形：当顾客到达时可以按如下规则在每个服务台前排成一个队：第 $1, n+1, 2n+1, \cdots$ 个顾客排入第一个队，第 $2, n+2, 2n+2, \cdots$ 个顾客排入第二个队等等. 或者排成一个公共的队，当有一服务台得空时，队首顾客进入服务. 也可以这样来排成 n 个队：第 m 个顾客到达时，以概率 $C_i^{(m)}$ 排入第 i 个队 $\left(\sum_{i=1}^{n} C_i^{(m)}=1,\ m=1,2,\cdots\right)$. 显然，第一种情形是这种情形的特例.事实上，令 $C_i^{(kn+i)}=1, i=1,2,\cdots, n,\ k=0,1,2,\cdots$，就得到第一种情形.

3) 混合制：

a) 队长有限制的情形：顾客到达时，若队长 $<N$，就排入队伍；若队长 $=N$，顾客就离去.

b) 等待时间有限制的情形：顾客在队伍中的等待时间不能超过 T，超过 T 后顾客就离去.

c) 逗留时间（等待时间与服务时间之和）有限制的情形：顾客在系统中的逗留时间不能超过 T，超过 T 后顾客就离去. 例如高射炮阵地射击来空袭的敌机，敌机飞过高射炮射击区域所需的时间为 T，若敌机飞过该区域后还未被击落，就算消失.

(iii) 服务机构：服务台的个数可以是一个或几个，可以是单个服务，也可以是成批服务，例如公共汽车一次就装载大批乘客. 下面来描述一下各种服务分布：

1) 定长分布：每一顾客的服务时间都是常数 β，此时服务时间 v 的分布函数为

$$B(x)=P\{v\leqslant x\}=\begin{cases}1, & x\geqslant\beta;\\ 0, & x<\beta.\end{cases} \tag{15}$$

2) 负指数分布：即各个顾客的服务时间相互独立，具有相同的负指数分布：

$$B(x)=\begin{cases}1-e^{-\mu x}, & x\geqslant 0;\\ 0, & x<0,\end{cases} \tag{16}$$

其中 $\mu > 0$ 为一常数. 平均服务时间与方差分别为

$$Ev = \int_0^\infty x dB(x) = \mu\int_0^\infty xe^{-\mu x}dx = \frac{1}{\mu} \tag{17}$$

与

$$\sigma_v^2 = \frac{1}{\mu^2}. \tag{18}$$

3）爱尔朗分布 E_k：即各个顾客的服务时间相互独立，具有相同的 k 阶（k 为正整数）爱尔朗分布：

$$B(x) = \begin{cases} 1 - \sum_{i=0}^{k-1} e^{-k\mu x}\frac{(k\mu x)^i}{i!}, & x \geqslant 0; \\ 0, & x < 0, \end{cases} \tag{19}$$

其中 $\mu > 0$ 为一常数. 易知其密度函数为

$$b(x) = \frac{k\mu(k\mu x)^{k-1}}{(k-1)!}e^{-k\mu x}, \quad x \geqslant 0, \tag{20}$$

而平均服务时间与方差分别为

$$Ev = \int_0^\infty xb(x)dx = \frac{1}{\mu} \tag{21}$$

与

$$\sigma_v^2 = \frac{1}{k\mu^2}. \tag{22}$$

$k=1$ 时爱尔朗分布化归为负指数分布；$k\to\infty$ 时得到长度为 $\frac{1}{\mu}$ 的定长服务分布.

4）超指数分布 H_k：即各个顾客的服务时间相互独立，具有相同的 k 阶（k 为正整数）超指数分布：

$$B(x) = \begin{cases} 1 - \sum_{i=1}^{k}\alpha_i e^{-\mu_i x}, & x \geqslant 0; \\ 0, & x < 0, \end{cases} \tag{23}$$

其中 $\mu_i > 0$，$\alpha_i > 0 (i = 1, 2, \cdots, k)$ 均为常数，且 $\sum_{i=1}^{k}\alpha_i = 1$. 易知其密度函数为

$$b(x)=\sum_{i=1}^{k}\alpha_i\mu_i e^{-\mu_i x},\quad x\geqslant 0, \tag{24}$$

而平均服务时间与方差分别为

$$Ev=\int_0^\infty xb(x)dx=\sum_{i=1}^{k}\frac{\alpha_i}{\mu_i}=\frac{1}{\mu} \tag{25}$$

与

$$\sigma_v^2=2\sum_{i=1}^{k}\frac{\alpha_i}{\mu_i^2}-\frac{1}{\mu^2}. \tag{26}$$

超指数分布服务可以看成顾客以概率 α_i 接受参数为 μ_i 的负指数分布服务. 当 $k=1$ 时的一阶超指数分布就是负指数分布.

5）一般服务分布： 所有顾客的服务时间是相互独立相同分布的随机变量,其分布函数记为 $B(x)$. 前面所有的服务分布都是一般服务分布的特例.

6）多个服务台的情形： 可以假定各个服务台的服务分布参数不同或分布类型不同.

7）服务时间依赖于队长：这反映了一般服务者的心理,排队的人愈多,服务的速度也就愈高.

下面,我们引入一个常用的关于随机服务系统分类的记号,令

M：代表普阿松输入或负指数服务分布；

D：代表定长输入或定长服务分布；

E_k：代表 k 阶爱尔朗输入或 k 阶爱尔朗服务分布；

H_k：代表 k 阶超指数输入或 k 阶超指数服务分布；

GI：代表一般独立输入；

G：代表一般服务分布.

并令 X/Y/n 代表输入为 X，服务分布为 Y 的 n 个服务台的随机服务系统. 于是例如

M/M/n：即普阿松输入、负指数服务分布的 n 个服务台的系统；

D/M/1：即定长输入、负指数服务分布的单个服务台的系统；

M/G/1：即普阿松输入、一般服务分布的单个服务台的系统；