PAIDUILUN

排队论

陆传赉　编著

北京邮电学院出版社

排 队 论

陆传赉 编著

北京邮电学院出版社

（京）新登字162号

排　队　论

作　　者　陆传赉
责任编辑　郑　捷
*
北京邮电学院出版社出版
新华书店北京发行所发行　各地新华书店经售
北京邮电学院出版社印刷厂印刷
*
850×1168毫米　1/32　印张 8.375　字数 217　千字
1994年5月第一版　1994年5月第一次印刷
印数：1—3000册
ISBN 7-5635-0174-6/O·8　定价：4.60　元

前　言

排队论（又称随机服务系统）是研究系统由于随机因素的干扰而出现排队（或拥塞）现象的规律性的一门学科，它适用于一切服务系统，包括通信系统、交通与运输系统、生产与服务系统、存贮与装卸系统、管理运筹系统以及电子计算机系统等。可以说，凡是出现拥塞现象的系统，都属随机服务系统。

排队论源于电话服务理论的研究，在第二次世界大战以前，其研究多侧重于电话和远距离通信方面，这阶段发展较缓慢，大战以后，由于排队论渗透到军事、经济、生产与服务、管理等多种部门，于是迎来了理论和应用的较大发展，使之成为运筹学和管理科学的一个热门分支，特别是自70年代以来，随着电子计算机的不断更新和发展；通信网的建立和完善；信息科学、生命科学及控制理论的蓬勃发展均涉及到最优设计与最佳服务问题，从而使排队理论与应用获得质和量上飞速的发展。

现在，不仅管理人员和工程设计人员渴求掌握有关排队论的基本知识；而且诸如通信、计算机、管理工程、经济类等专业的本科生、研究生更期待有一本适合其专业和学习程度的排队论教材，鉴于上述，引发我在为信息工程专业本科生讲授该课程五年的讲稿基础上编写了这本小册子。本书是在参阅多种文献基础上按先易后难、循序渐进的原则，以及多讲稳态，少讲瞬态的指导思想下编写整理的，其中也包括作者某些研究成果。本书前两章是排队论讲述中所必需的基本知识与数学工具——马尔可夫链；第三章为单服务窗（此书用服务窗而不用服务员或台完全是为了适应后续课程——通信网理论基础的需要）排队模型；第四章为多服

务窗排队模型；第五章为非马尔可夫排队模型；第六章为离散时间排队模型；第七章为特殊排队模型；第八章为排队系统中的优化模型。书中带 * 号的章节可视学时多少及学生接受能力情况加以增减，或仅作为研究生的必读部分。书中列有大量例题和习题，以便于印证所学的理论和方法。

在本书写作与出版过程中，得到教研室不少同志的帮助；特别是胡正名教授、吴伟陵教授、李道本教授都曾给作者不少指导与鼓励；我的研究生方莉、于凯在多次辅导的基础上提出过不少好的建议；审稿的同志也提出一些宝贵的修改意见，本人在此一并向他们表示衷心的感谢。

由于作者水平所限，书中有可能出现错误或不妥之处，恳请有关专家和广大读者批评指正，以求提高。

作　者

1993年6月

目　录

第一章　预备知识

§1　排队问题的基本概念

1.1　概述

众所周知，由于某些资源、设备或空间（场地）的有限性及社会各部门对它们的需求是存在排队现象的主要因素，而诸如服务机构的管理水平低劣，服务窗（员）的素质差，效率不高，或顾客的无计划性以及其他原因也往往使不该有的排队现象出现。

我们所要讨论的排队论是人们研究大量服务过程的一门数学理论。在社会生活中碰到的排队现象，诸如到商场购物，去图书馆借阅书刊、资料，汽车到加油站加油，船舶需要停靠码头，在公共电话亭打电话，将有毛病的电器送维修部门进行维修，病人去医院挂号看病，将有关数据输入计算机进行存贮等等，均可归结为顾客与服务窗之间的一种服务关系，并可用框图表示这类排队过程，如图1.1-1所示。

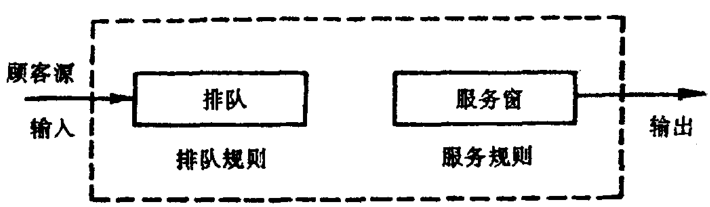

图 1.1-1　排队模型框图

或列表如下：

顾　　　　客	要求服务项目	服　务　窗
待修机（仪）器	修理	维修人员
汽车	加油	加油站
病人	看病	大夫
电话呼唤	通话	交换台
飞机（船舶）	进航空港（港口）	跑道（码头）
球队	比赛	场地
信号	传送	信道
数据	存贮	计算机
……	……	……

1.2　排队系统的特征或组成

一、输入过程

1．顾客总体可以有限或无限（如流入水库的水）。

2．顾客到达系统的方式可以逐个或成批。

3．顾客相继到来时间间隔可分为确定型（比如定期航班，定期的课程表等）和随机型（比如看病的病人，候车的旅客，进港口的船舶）。

4．顾客到达系统可以是独立的或相关的，输入过程可以是平稳、马氏、齐次的等。

二、排队规则

排队规则可分为三种制式：

损失制——顾客到达系统时，如果系统中所有服务窗均被占用，则到达的顾客随即离去，比如打电话时碰到占线，用户即搁置重打或离去另找地方或过些时再打，又如旅店客满谢客，挂牌大夫限额挂号，计算机限定的内存等均为此种情形。

等待制——顾客到达系统时，虽然发现服务窗均忙着，但系统设有场地供顾客排队等候之用，于是到达系统之顾客按先后顺序进行排队等候服务。通常的服务规则有先到先服务，后到先服务

（比如仓库中同种物品堆垒后的出库过程），随机服务，优先服务（比如邮政中的快件与特快专递业务，重危病人的急诊，交通中让救火（护）车、警车及迎宾车队优先通过）等。

混合制——它是损失制与等待制混合组成的排队系统，此系统仅允许有限个顾客等候排队，其余顾客只好离去永不再来；或者顾客中有的见到排队队伍长而不愿费时等候，当队伍短时愿排队等候服务；也有排队等候的顾客当等候时间超过某个时间就离队而去均属这种系统。

2. 排队队列可具体或抽象，系统容量可以有限或无限。

3. 排队队列可以单列或多列。

三、服务窗（员）

1. 系统可以无窗口（如自选自付款购物）、一个窗口或多个窗口为顾客进行服务。

2. 在多个服务窗情形，顾客排队可以平行多队排列，串列或并串同时存在的混合排队。

3. 一个服务窗可以为单个顾客或成批顾客进行服务。

4. 各窗口的服务时间可为确定型（如交通路口红绿灯亮的时间，各单位固定的上下班时间）或随机型。服务时间往往假定是平稳的。

四、排队系统的目标参量（或运行指标）

1. 绝对通过能力A，它为单位时间内被服务完顾客的均值。

2. 相对通过能力Q，它为单位时间内被服务完顾客数与请求服务顾客数之比值。

3. 系统排队长度均值L_s，它即是系统内顾客数的均值。

4. 排队等候顾客的平均队列长度L_q，它即是系统内排队等候顾客的均值。

注：如果L_s（或L_q）较大，那么说明该系统的工作效率较低。

5．顾客在系统内逗留时间的均值 W_s；顾客排队等候服务的时间的均值W_q；服务时间的均值 $L_{服} \triangleq \bar{\tau}$；显然有

$$W_s = W_q + \bar{\tau} \qquad (1.1\text{-}1)$$

6．服务窗连续繁忙的时间长度，即忙期T_b。

7．系统的损失概率 $P_{损}$，即系统满员概率。

1.3 排队模型的分类与记号

记X为顾客相继到达系统的间隔时间T的概率分布；Y为服务窗口所耗费的服务时间 τ 的概率分布；Z 为服务系统内服务窗的个数；m 为系统内（最大）排队容量或顾客在系统中排队所允许的（最大）长度（包括正在服务和排队等待的顾客）。

又令M为负指数分布；D为确定型分布；E_k 为 k 阶爱尔兰（Erlang）分布；G为一般分布；GI 为一般独立的分布。

通常用记号 X/Y/Z/m（或∞）来表达排队模型。为方便起见，当系统最大排队容量为∞时，就可略写为X/Y/Z，比如：

M/M/n 排队模型表示顾客相继到达系统的间隔时间服从负指数分布，而服务时间也服从负指数分布，系统内设有 n 个服务窗，系统容量为无限（充分大即可）的等待制排队模型。

M/M/n/m 排队模型表示顾客到达的间隔时间和服务时间均为负指数分布，有 n 个服务窗且系统容量为m的损失制排队模型。

G/M/1 排队模型表示顾客到达的间隔时间为一般分布，服务时间为负指数分布，只设有一个服务窗的等待制排队模型。

G/GI/1 排队模型表示间隔时间为一般分布，服务时间为一般独立分布，只设有一个服务窗且系统容量为无限的等待制排队模型。

M^k/M/1 排队模型表示每批有k 个顾客到达系统，且批与批

到达间隔时间是负指数分布，服务时间为负指数分布，只有一个服务窗，且系统容量为无限的等待制排队模型。

M/M/n/m/m 排队模型表示顾客到达的间隔时间与服务时间均为负指数分布，系统内有 n 个服务窗，顾客源为 m 且系统容量也为 m 的闭合式（循环）排队模型。

§2 排队问题中常见的事件流

我们将同类事件一个（批）个（批）随机地来到服务窗口要求服务的序列称作事件流，如电话局接到的呼唤流、计算机出现的故障流、进站的汽车流、看病的人流、去某公司应聘考试的考生流等均是事件流。显然，这些事件流均为随机变量，由于顾客（用户）到达系统的间隔时间与服务时间均为非负，故它们还是非负的随机变量。常用的有下述几个分布：

1．二项分布：在 n 次独立试验中，某事件A出现 m 次的概率为 $C_n^m p^m(1-p)^{n-m}$，其中 p 为事件A出现的概率。

2．泊松（Poisson）流，又称最简单事件流。它具有如下特点：

(1) 平稳性。在任何一段长度为 t 的时间区间内，出现任意数量事件的概率只与 t 有关，而与 t 所处的位置（或与起始时刻）无关。记 λ 为平稳流的强度。

(2) 无后效性(又称无记忆性或马氏性)。在互不相交的两时间区间 T_1、T_2 内所出现的事件数是相互独立的。比如到商店购物的顾客，待修的机器，进站的列车等均具有无后效性。

(3) 普通性。在同一瞬间，多于一个事件出现的概率（或同时到达系统有两个或两个以上顾客的概率）可忽略不计。

如果用数学语言的话，记 $N(t)$ 为时间区间 $[0, t]$ 内到达的事件（顾客）数；令

$$P_n(t_1,t_2)=P\{N(t_2)-N(t_1)=n\} \qquad (t_2>t_1) \quad n\geqslant 0 \tag{1.2-1}$$

倘若它满足下列条件：

(a) 不重迭区间内到达的（顾客）事件是彼此独立的；

(b) $P_1(t,t+\Delta t)=\lambda\Delta t+0(\Delta t)$； (1.2-2)

其中0（Δt）是Δt的高阶无穷小量（当$\Delta t\to 0$时），λ是正常数。

(c) $\sum\limits_{n=2}^{\infty}P_n(t,t+\Delta t)=0(\Delta t)$； (1.2-3)

那么，就称顾客到来的事件流是一泊松流。简记$P_n(t)=P_n(0,t)$，容易验证

$$P_0(t,t+\Delta t)=1-\lambda\Delta t+0(\Delta t) \tag{1.2-4}$$

且

$$\begin{aligned}P_n(t+\Delta t)&=P(N(t+\Delta t)-N(0)=n)\\&=P_n(t)P_0(t,t+\Delta t)+P_{n-1}(t)P_1(t,t+\Delta t)\\&\quad+P_{n-2}(t)P_2(t,t+\Delta t)+\cdots\cdots\\&=P_n(t)(1-\lambda\Delta t)+P_{n-1}(t)\lambda\Delta t+0(\Delta t)\end{aligned} \tag{1.2-5}$$

由此可得

$$\frac{P_n(t+\Delta t)-P_n(t)}{\Delta t}=-\lambda P_n(t)+\lambda P_{n-1}(t)+0(1) \tag{1.2-6}$$

令$\Delta t\to 0$，有

$$\begin{cases}P_n'(t)=-\lambda P_n(t)+\lambda P_{n-1}(t)\\ P_n(0)=0 \qquad n\geqslant 1\end{cases} \tag{1.2-7}$$

及

$$\begin{cases}P_0'(t)=-\lambda P_0(t)\\ P_0(0)=1\end{cases} \tag{1.2-8}$$

由上两式解得

$$P_n(t)=\frac{(\lambda t)^n}{n!}e^{-\lambda t} \qquad n=0,1,2,\cdots(t>0) \qquad (1.2\text{-}9)$$

此即为泊松分布。

3．负指数分布

当顾客流为泊松流时，用T表示两个相继顾客到达系统的时间间隔，记其分布函数为

$$F(t)=P(T\leqslant t)=1-P(T>t)=1-P_0(t)$$

由于$P_0(t)=e^{-\lambda t}$，故有

$$F(t)=1-e^{-\lambda t} \qquad (1.2\text{-}10)$$

相应的分布密度为

$$f(t)=\lambda e^{-\lambda t} \qquad (1.2\text{-}11)$$

它就是负指数分布的密度函数。

易知

$$ET=\frac{1}{\lambda} \qquad DT=\frac{1}{\lambda^2}$$

通常，服务窗为一顾客服务所需的时间τ的分布函数与分布密度为

$$F_\tau(t)=1-e^{-\mu t} \quad f_\tau(t)=\mu e^{-\mu t} \qquad (1.2\text{-}12)$$

其中参数μ为单位时间内服务窗所完成服务的顾客均值数，且有

$$\bar{\tau}=E\tau=\frac{1}{\mu} \qquad (1.2\text{-}13)$$

4．爱尔兰分布

考察最简单的事件流，记各事件到达系统的时间间隔序列为T_1，T_2，…，T_i，…，且它们是同负指数分布的随机变量序列，如图 1.2-1 所示。

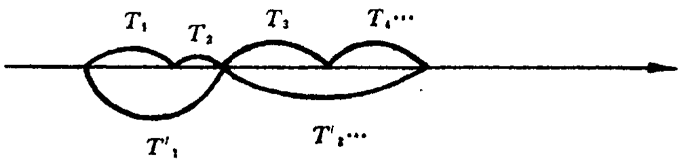

图 1.2-1 爱尔兰随机流

置

$$T'_1=T_1+T_2,\ T'_2=T_3+T_4,\ T'_3=T_5+T_6,\ \cdots$$

$$T'_n=T_{2n-1}+T_{2n},\ \cdots$$

那么，称 T'_1，T'_2，…，T'_n，…是二阶爱尔兰流。

如令

$$T'_1=\sum_{i=1}^{k}T_i\ ,\ T'_2=\sum_{i=1}^{k}T_{k+i},\ \cdots$$

则称T'_1,T'_2，…，T'_n，…是 k 阶爱尔兰流。由于各个T_i是相互独立且服从相同参数 λ 的负指数分布，为求 k 阶爱尔兰流的分布，可记 $T=\sum\limits_{i=1}^{k}T_i$，注意$f_{T_1}(t)=\lambda e^{-\lambda t}(t>0)$，故由$f_{T_1}(t)$的Laplace变换

$$L[f_{T_1}(t)]=\int_0^{\infty}\lambda e^{-\lambda t}\cdot e^{-st}dt=\frac{\lambda}{s+\lambda}$$

及拉氏变换的性质得到

$$L[f_T(t)]=\frac{\lambda^k}{(s+\lambda)^k}$$

再查反拉氏变换表知

$$f_T(t)=\frac{\lambda(\lambda t)^{k-1}}{(k-1)!}e^{-\lambda t}\qquad(t>0,\ \lambda>0)\qquad(1.2\text{-}14)$$

并且

$$ET = \sum_{i=1}^{k} ET_i = \frac{k}{\lambda}$$

$$DT = \sum_{i=1}^{k} DT_i = \frac{k}{\lambda^2}$$

5．广义爱尔兰分布

设$\{T_i, i=1,2,\cdots\}$如上段所述为独立且有分布密度

$$f_{T_i}(t) = \lambda_i e^{-\lambda_i t}$$

的各事件到达的时间间隔序列，那么，$T = \sum_{i=1}^{k} T_i$ 的广义爱尔兰分布密度为

$$f_T(t) = \prod_{i=1}^{k} \lambda_i \sum_{i=1}^{k} \frac{e^{-\lambda_i t}}{\prod_{j \neq i}^{k} (\lambda_j - \lambda_i)} \qquad (1.2\text{-}15)$$

实用上，常取$k=2$，那么有

$$f_T(t) = \lambda_1 \lambda_2 \frac{e^{-\lambda_1 t} - e^{-\lambda_2 t}}{\lambda_2 - \lambda_1}$$

这时的广义爱尔兰流的强度为

$$\lambda = \frac{\lambda_1 \lambda_2}{\lambda_1 + \lambda_2}$$

6．超指数分布（记作H_k）

设随机变量T的分布密度为

$$f(t) = \sum_{i=1}^{k} \theta_i \mu_i e^{-\mu_i t} \qquad (1.2\text{-}16)$$