排队论及其应用

董泽清著

西安系统工程学会出版

序

排队论是研究拥挤现象的一门学科。它是运筹学的重要分枝，也是应用概率的一个分枝。它所研究的问题，有着强烈的实际背景，其结果有着广泛的应用性：包括电信、交通运输（陆、海、空）、机器看管、生产线与维修、健康与卫生、水库设计、库存管理、系统可靠性及军事等领域。从六十年代以来另一个极重要的应用领域是计算机系统。如果说排队论起源于电讯领域，那末它的未来发展将与计算机领域结下不解之缘。这包括计算机系统的最优设计与运行管理。涉及离散时间排队系统与网络式排队系统的研究和应用。这是理论与应用相互促进的一个很好例子。

在我国，排队论已应用于电信、纺织、交通、矿山、设算机设计、可靠性、机械加工、电子对抗及军事等领域。但总起说来，还不太普遍。想必在日益强调经济效益、科学管理的方针下，它在我国的社会主义建设中将起愈来愈重要的作用。

排队论的理论研究，特别是瞬时性质的研究，获得了一系列精美的结果，这也是它的研究中最困难的部分。但要使排队论继续保持生命力，重要的是不应割断它与实际问题的联系；而应使它的理论发展与应用的完善化紧密地结合起来。

本书不求理论上的完整，是从应用角度写的，因此瞬时性质很少涉及。重点介绍统计平衡下的结果——在适当的条件下，系统运行足够长的时间之后，呈现稳定（平衡）势（状）态。就应用来说，绝大部分是用统计平衡结果，而瞬时结果，一般说既不便于作数值计算，也不便于对系统进行最优运行控制与设计。因为这些结果随时间是变化的。本书我们试图作到：既注重于应用，又不至于曲解。所涉及的内容力求做到概念准确、论证严格、叙述浅显；对一些较难的、或推导冗长的内容，我们多不加证明地叙述其结果，并讲明前提条件及参考文献，这将有利于初学者与工程技术人员；并尽可能地指出有关问题的新近文献，以利于有志深入学习的读者；书中还包含了丰富的例子，以说明如何使用结果去解决实际问题，其中有些是我们亲自参与实践的一些问题，试图为读者提供一些值得仿效的案例。但好比学游泳一样，一个人要想学会游泳，他就得下水去游。谁要想成为应用排队论解决实际问题的行家、里手，在他了解基本知识之后，关键在于多去实践，熟才能生巧。

第一章对排队论作一简要介绍，第二章讲述书中常用的一些概率论知识，包括负指数分布的特性——无记忆性，泊松过程的特性及生灭过程的极限结果。第三至八章讲述排队论的一些基本模型。我们采取从特殊到一般的叙述方式。第九章是关于排队系统的统计推断，这本质上是属于过程统计，有待进一步发展：利用排队过程的特殊性，发展更有效的统计推断方法。第十章讲述排队论中已用过的主要分析方法、发展概况、研究内容及国内工作概况，作为前面诸章的一个补充。在正文前列有常用符号表。

本书是由讲义逐渐发展而成，第一稿作为运筹学讲义的一部分是由华中工学院打印的，内容较少。第二稿由中国人民大学数学教研室与冶金部马鞍山矿山研究院分别刻印的，已初具规模，第三稿是由中国科技大学打印的，作为数学系77届运筹学班的授课讲义（后由北方交通大学重印）。其后又作了一些扩充和修改，增加了第八章，成为目前的样子。

阅读本书仅需微积分与概率论的基本知识。对于搞应用的同志，在初读时，先略去带星号“*”的节，并不影响对主要内容的了解和应用。本书可作为大学有关专业高年级引论性教材。特别适合于运筹学、管理科学、系统科学、电信、计算机科学以及应用数学等系或专业。可供大学教师、有关科研人员、工程技术人员以及管理人员参考。由于排队论文献浩如烟海、模型变化万千，就是统计平衡结果也不可能全都包含，但在本书的基础上，想必读者会去解决一些实际问题。

在写的过程中，得到了研究室内同志们的指导和帮助；先后在十来个单位讲授过，不少同志提出过许多宝贵意见，在修改过程中大多采用了。在此我对他们表示感谢。恕我不在这里一一列举他们的大名，本书能较快地出版是与杨雯同志的热情支持和操劳分不开的，对她也表示感谢。

由于我的水平所限，难免有不当或错误之处，甚望读者批评指正。

董 泽 清

1983年1月写于科学院应用数学所　　北京

常 用 符 号 及 定 义

瞬时情形

1、A_n：M／G／I／∞系统第n个顾客的服务时间内到达的顾客数。

2、$A_n(t)$：第n个到达间隔时间τ_n的分布函数。

3、B_n：GI／M／I／∞系统第n＋1个到达间隔时间τ_{n+1}内服务结束的顾客数。

4、$B_n(t)$：第n个顾的服务时间分布函数。

5、$N(t)$：时刻t系统所处的状态。

6、$N_q(t)$：时刻t排队等待的顾客数。

7、$N_v(t)$：时刻t正在接受服务的顾客数。

8、$p_n(t)$：时刻t系统处于状态n的概率，即

$p_n(t)=P\{N(t)=n\}$。其中之n有时为(n, i)，(i, n)，(i, n, j)，(i, j, k)等分别表示于时刻t系统处于状态(n, i)，(i, n)，(i, n; j)，(i, j; k)的概率。

9、T_n：第n个顾客的到达时刻，$T_0 \equiv 0$。

10、V_n：第n个顾客的服务时间。

11、$Tq(n)$，Tq_n：第n个顾客的（实）等待时间。

12、$Tq(t)$：设t时刻有个顾客到达，他应等的时间。即虚等待时间。

13、τ_n：第n个到达间隔时间，即$\tau_n = T_n - T_{n-1}$。

统计平衡情形

1、a：M／M／C／m／m系统平均运行的机器数。

2、A（t）：到达间隔时间τ的分布函数。

3、a（t）：到达间隔时间τ的密度幽数。

4、B（t）：服务时间V的分布函数。

5、b（t）：服务时间V的密度函数。

6、c：系统的服务员数。

7、$\overline{c}$：平均忙的服务员数。

8、C_v：服务时间的变异系数。即$C_v = \frac{(D〔V〕)^{1/2}}{E〔V〕}$，

其中E〔V〕、D〔V〕分别见15、13。

9、$C\tau$：到达间隔时间的变异系数。

10、D〔N〕：队长N的方差。

11、D〔N_q〕：排队长N_q的方差。

12、D〔T〕：逗留时间T的方差。

13、D〔V〕：服务时间V的方差。

14、D〔τ〕：到达间隔时间τ的方差。

15、E〔V〕：平均服务时间。

16、E〔τ〕：平均到达间隔时间。

17、E_k：k级（阶）爱尔朗分布符号。

18、G：一般（独立、同分布的）服务时间分布符号。

19、GI：一般（独立、同分布的）到达间隔时间分布的符号。

20、H_k：k阶超指数分布符号。

21、K：系统中最大允许的顾客数。

22、L：平均队长。即L＝E〔N〕。

23、$L^{(A)}$：GI/M/1/∞系统顾客到达时已有顾客的平均数。

24、Lq：平均排队长。即Lq＝E〔Nq〕。

25、$L_q^{(A)}$ GI/M/1/∞系统顾客到达时已有排队顾客数的平均数。

26、M：指数到达间隔时间或指数服务时间的符号。

27、m：M/M/m＋N/m系统（N≥0）中看管的机器数。

28、$\overline{m}$：M/M/m＋N/m系统（N≥0）中平均运行的机器数。

29、N：系统的顾客数.

30、Nq：系统中排队等待的顾客数.

31、N′：顾客到达时系统已有的顾客数.

32、Nv：正在接受服务的顾客数.

33、p_n：任一时刻（在统计平衡下）系统有n个顾客的概率，即$p_n=P\{N=n\}$.

34、q_n：顾客到达时发现系统已有n个顾客的概率.即$q_n=p\{N'=n\}$.

35、q(c)：M/M/m+N/m系统（N≥0）修理工人的损失系数.

36、r(c)：M/M/m+N/m系统（N≥0）由干挠引起的机器损失系数.

37、T：顾客在系统的逗留时间.

38、Tq：顾客在系统的等待时间.

39、V：服务时间、$B(t)=P\{V\leqslant t\}$.

40、W：顾客在系统的平均逗留时间.

41、Wq：顾客在系统的平均等待时间.

42、W_q'：顾客在系统的平均虚等待时间.

43、W(t)：顾客在系统的逗留时间分布函数.

44、Wq(t)：顾客在系统的等待时间分布函数.

45、$W_q'(t)$：顾客的虚等待时间分布函数.

46、λ：单位时间平均到达的顾客数，即$\frac{1}{\lambda}=E[\tau]$；或单位运转时间平均发生的故障数.

47、λe：单位时间实际进入服务系统的平均顾客数.

48、μ：单位时间平均能够服务的顾客数.

49、μe：单位时间实际平均服务的顾客数.

50、ρ：一个服务员的服务率，即$\rho=\frac{\lambda}{\mu}$.

51、ρ_c：c个服务员的服务率，即$\rho_c=\frac{\lambda}{c\mu}$.

52、τ：到达间隔时间.

53、πn：M/G/1/∞系统，服务结束时刻，仍留在系统的顾客数为n的概率.

54、$\pi(i)$：分布的i%分位点.

55、$\in$：属于符号，如$X\in A$表示X属于集合A.

目　　录

第　一　章

绪　　言

§1.1 引 言……（1）
§1.2拥挤现象的共性……（1）
§1.3研究的内容与目的……（2）
§1.4排队系统的基本组成部分……（4）
　§1.4.1输入过程……（4）
　§1.4.2排队规则……（6）
　§1.4.3服务机构……（8）
§1.5经典排队模型（系统）的符号表示……（12）
§1.6排队模型的主要数量指标……（13）
　§1.6.1几个数量指标……（13）
　§1.6.2数量指标之间的基本关系……（14）

第　二　章

预　备　知　识

§2.1负指数分布……（16）
　§2.1.1负指数分布的定义……（16）
　§2.1.2负指数分布的特性……（16）
§2.2最简单（Poisson）流……（19）
　§2.2.1最简流的推导……（19）
　* §2.2.2最简单流的特性……（23）
　§2.2.3等待时间与次序统计量……（29）
　* §2.2.4最简单流的转移概率函数及Q矩阵……（31）
§2.3生灭过程……（32）
　§2.3.1生灭过程的定义……（32）
　§2.3.2微分差分方程组……（33）
　§2.3.3统计平衡解……（34）

第　三　章

单服务机构指数服务系统

§3.1M/M/1/∞系统的稳定状态性质……(36)
§3.2几个数字特征……(38)
§3.3等待时间分布……(41)
§3.4一些关系……(47)
§3.5统计平衡下的输出过程……(48)
* §3.6瞬时性质……(49)
* §3.7忙期……(55)
§3.8费用模型的最优化问题……(57)
§3.9服务依赖于状态的系统……(58)
§3.10M/M/1/k系统……(60)
§3.10.1统计平衡解与数字特征……(60)
§3.10.2一个最优化问题及例……(68)

第　四　章

多服务机构指数服务系统

§4.1M/M/∞系统……(68)
§4.1.1统计平衡解与数字特征……(69)
* §4.1.2瞬时解……(71)
§4.2M/M/c/∞系统……(73)
§4.2.1统计平衡性质……(73)
§4.2.2最优化问题及例……(77)
§4.2.3统计平衡下的输出过程……(80)
§4.3M/M/c/k系统……(81)
§4.3.1统计平衡性质……(81)
* §4.3.2M/M/1/1系统的瞬时性质……(86)
§4.4具有不耐烦顾客的M/M/c/∞系统……(87)
§4.5串联排队系统……(89)

第　五　章

有限源服务系统

§5.1指数服务系统（M/M/c/m/m系统）……(93)
§5.1.1状态概率及数字特征……(93)

§5.1.2等待时间分布 …… (100)
§5.1.3最优化问题 …… (102)
§5.1.4M/M/c/c/m系统 …… (103)
§5.2有备用品的M/M/c/m+N/m 系 统 …… (104)
§5.2.1状态概率及数字特征 …… (104)
§5.2.2最优化问题 …… (109)
§5.3循环排队系统 …… (110)
§5.3.1无反馈的循环排队系统 …… (111)
§5.3.2具有反馈的循环排队系统 …… (113)
§5.4一个可靠性问题 …… (114)
§5.4.1有强拆优先权的机器看管问题 …… (115)
§5.4.1双机工作失败的概率p …… (117)

第 六 章

爱尔朗排队系统

§6.1M/E /1/∞排队系统 …… (119)
§6.2E_k/M/1/∞排 队 系统 …… (126)
§6.3E_l/E_k/1/∞排 队 系统 …… (133)

第 七 章

一般服务或（和）一般到达型系统

§7.1M/G/1 / 系 统 …… (137)
§7.1.1嵌入马氏链 …… (137)
* §7.1.2 $\pi_n=p_n$, $n\geqslant 0$ …… (144)
§7.1.3数字特征 …… (146)
§7.1.4等待时间分布 …… (143)
§7.1.5例 …… (153)
§7.2 GI/M/1/∞系 统 …… (158)
§7.3 GI/G/1系统 …… (166)
§7.4逼近解 …… (170)
§7.5 一些结果 …… (171)

第 八 章

优先权排队系统

§8.1M/M/1/∞非强拆两种优先权系统 …… (174)

§8.2M/M/1/∞非强拆多类优先权系统…… (181)
§8.3M/M/1/∞强拆优先权系统…… (184)

第 九 章

排队系统的统计推断

§9.1引言…… (190)
§9.2古典参数估计问题…… (191)
§9.3M/M/1/∞系统的参数估计…… (194)
§9.4生灭排队系统的参数估计…… (198)
§9.5M/G/1/∞系统的参 数估计…… (200)
§9.6排队系统参数的区间估计…… (201)
§9.7假设检验…… (204)
§9.7.1χ^2 检 验 …… (204)
§9.7.2对负指数分布的F检验…… (205)
§9.7.3 K—C检验…… (206)
§9.7.4 A—D检验…… (206)

第 十 章

小 结

§10.1排队论中的理论分析方法…… (210)
§10.2排队论的发展概况…… (214)
§10.3排队论的研究内容…… (215)
§10.4国内工作概况…… (216)
附表1.M/M/1/N系统ρ的最优值…… (218)
附表2.F (x) 的参数已知时，K—C检验的临界 值K α…… (219)
附表3.F (x) 的参数未知时，K—C检验的临界值K α …… (220)
附表4.A—D检验的临界 值 Aα …… (221)
附表5.GI/M/1/∞系统r_0作为ρ的 函数表…… (221)

参 考 文 献

排队论及其应用

第一章 绪 言

§1.1 引 言

排队论（随机服务理论、随机服务系统）是研究各种排队系统的概率规律性（属于认识世界）；从而解决有关排队系统的最优化问题（属于改造世界）；为此，还有相应过程统计的推断工作。简而言之，排队论是研究拥挤现象的一门学科。在日常的工作与生活的各个方面，人们都会遇到各种各样的拥挤问题——为了获得某种服务而排队等待。可以说是司空见惯的。如进饭馆就餐、到图书馆借书、至车站乘公共汽车、去医院看病、往售票处购票、上工具房领物品，都免不了会要排队等待。饭馆的服务员与顾客、图书馆的出纳员与借阅者、公共汽车与乘客、医生与病人、售票员与买票者、工人与管理员均分别构成一个排队系统，或称为服务系统。比如，当买票者到达售票处时，若所有售票员均忙着，他就只好排队等待。随着生产与服务事业的社会化，这种拥挤现象会变得愈来愈普遍。

除了有形的排队之外，还可以是无形的队列。如有几个旅客同时都打电话到火车站（飞机场、码头）电话售票处订购车（飞机、船）票，则当一个旅客正在通话时，其他旅客就只得在各自的电话机前等待。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。

排队的不一定是人，也可以是物。如生产线的原料、半成品等待加工；因出故障而停止运转的机器等待工人修理；码头的船只等待装或卸；要降落的飞机因跑道不空而在空中盘旋，进入我方阵地上空的敌飞行器，待我方火力射击。

进行服务的也不一定是人，也可以是物。如机场的跑道、自动售货站的电子设备、公共汽车。

顾客也不一定是一个一个地可数的，而是一个取连续值的变量。如在水库问题里，上游的水滚滚而来，就是不可数的。若调节得好，能充分发挥水库的效益（通常是多目标的）。

服务员也不一定固定在一个地方对顾客进行服务。如出租汽车站，车子回站与乘客到来都是随机的。有的乘客是用电话要车。

§1.2 拥挤现象的共性

为了用语一致起见，今后凡是要求服务的对象统称为“顾客”。进行服务者统称为

“服务机构”，或“服务员”。因此，顾客与服务机构（员）完全是广义的。在不同的问题中可以有不同的含义。

实际的排队系统虽然千差万别，然而它们却有一些共同的特征，使我们能对它们进行统一的处理。一个排队系统能抽象地描述为如下：为了获得某种服务而到达的顾客；若不能立即获得服务，而又允许排队等待，则加入等待队伍；获得服务之后离开系统。用图形表示如下。

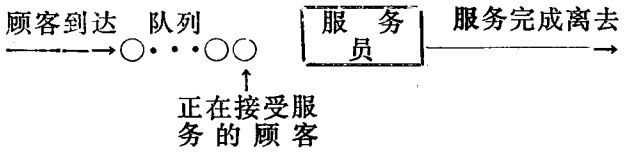

图1.1　单服务员排队系统的图解表示。

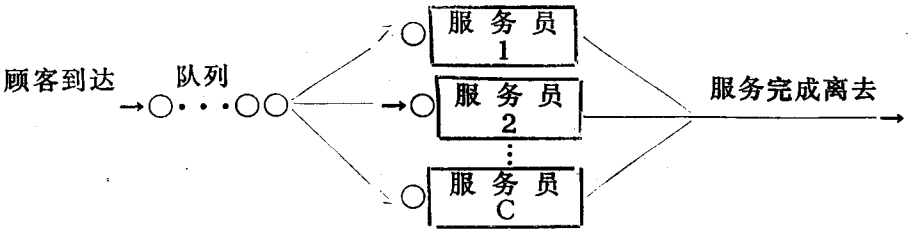

图1.2　C个服务员，一个公共队的图解表示。

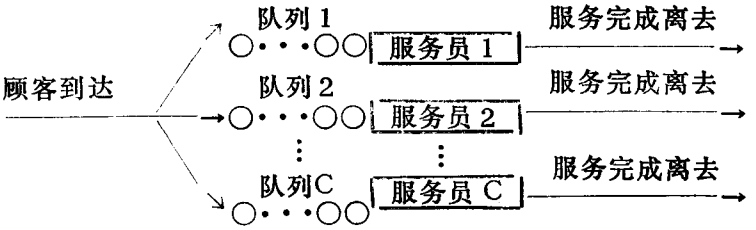

图1.3　C个服务，C个队列的图解表示。

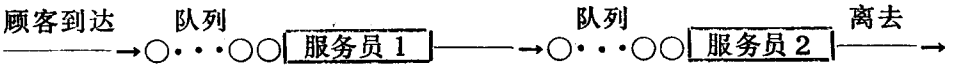

图1.4　单服务员串连排队系统的图解表示。

类似地可画出串、并联排队系统，网络排队系统的图解表示。

在各种排队系统中，相继（邻）顾客到来的间隔时间与每个顾客的服务时间，往往是无法事先确切预知的。因此，随机性是排队系统的一个共性。事实上，随机性在排队论中起作根本性的作用。顾客的相继到达间隔时间与服务时间两者，至少有一个具有随机性；否则问题就简单了，只需作简单的安排即可。后者我们将不讨论。正因为如此，排队论有时称为随机服务理论，或称为随机服务系统理论。

§1.3　研究的内容与目的

由于排队系统具有随机性，因此我们的首要任务就是研究具有随机性的拥挤现象的规律性，研究大量偶然性中所蕴含的必然性。在这种意义上讲，排队论是一种特殊的随

机过程论——排队过程论。因为现代随机过程论研究的一个主要内容，是研究样本函数的性质（如见王梓坤〔65〕*）。而排队论研究的首要任务是研究排队过程（模型）的几个数量指标的概率规律（见§1.6），即研究排队过程的一些整体性质。然后，进一步探讨排队系统的最优化问题。为此，还有相应的随机过程的统计推断工作。

为什么会产生排队？回答是相当简单的：在系统中要求服务的顾客数多于可用的服务机构数。为何如此呢？原因是多种多样的。例如，可能是服务员不够；提供防止等待所必须的服务员数（或服务率）可能是经济上不可行的，或不合算的；可能是提供的服务场地有限制等等。当然，若服务员的数量（或服务速率）得到充分保证，就可以防止排队现象；至少不会排得太长。如每台机器都有一个工人专职看管，一出故障，就能立即修理，因此不会产生机器等待修理的现象。故，对顾客来说，服务机构的数量（或服务速率），当然愈多（高）愈方便。但另一方面，服务机构愈多（或服务速率愈高），人力、物力的支出也就愈大。显然也是不经济的。因此，就产生了顾客的等待与服务机构的数量（或服务速率）之间的合理平衡问题。对大多数实际排队模型，问题的关键在于确定服务率、或服务机构数，或确定选取顾客进行服务的规则，或确定这几个量的某种组合，使得在某种意义下达到最优；当然也可能是确定到达率，或与别的量的某种组合。总之，要寻求排队模型中某（些）参变量的最优值。当然“最优”的含义随问题的不同而异。简单的费用模型如下图所示。

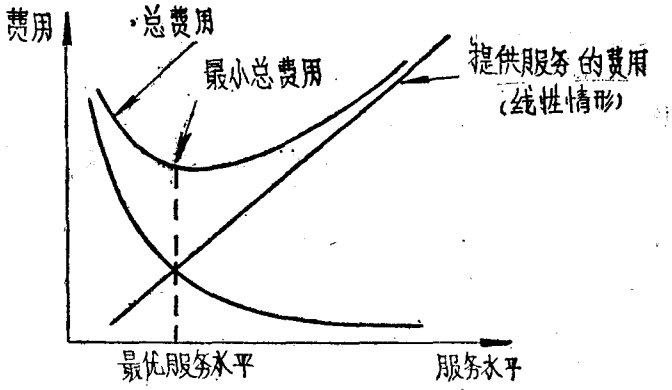

图1.5　费用模型

这里应区别两种情形：一种是对已有服务系统的最优运营（控制），称为动态最优；一种是服务系统的最优设计，称为静态最优。前者是对现有排队系统寻求最优运营策略（规则），也叫系统的实时控制。如在工具房的例子中，当排队领物的工人太多，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熟练工人领物的等待时间，即增加了机器的有效生产时间，这样带来的好处，可能远超过服务费用的增加。这

*）〔65〕意即〔1965〕。我们将始终采用类似的简写法。

是对服务机构数量的控制。也可以是控制顾客的到达。如一个理发店，快下班时，若等待的顾客已较多，就决定不再接受新来的顾客。也可以是控制服务规则。最优设计是在服务系统设置之前，就对未来的情形有所估计，使设计人员有所依据。如电话局的规模、水库库容、港口码头数、机场跑道数等等的设计。

总之，当今管理者面对的一个极其重要的问题，是如何正确的设计与有效地运营一个服务系统，既能适当地满足顾客的需要，又使服务构花费最小（或收益最大）。这就是排队论研究的最终目的。这种问题往往为管理者所忽略，造成投资不能最大地发挥效益，甚至亏损。排队论为解决这类问题提供了一种有效的方法。即排队论为现代管理决策提供了一种科学分析的方法。

§1.4 排队系统的基本组成部分

在实践中可能发生的排队系统是千变万化的。但有其共同的特征，概括起来总有三个基本组成部分（对一些特殊问题，可能还有别的组成部分）：输入过程、排队规则和服务机构。下面对它们分别作较详细的叙述。

§1.4.1 输入过程

输入过程就是刻划顾客按怎样的规律到达。要完全刻划一个输入过程需要如下三方面。

Ⅰ）顾客总体（顾客源）数。可能是有限的，也可能是无限的，甚至非可数无限的。工厂内出故障的机器数显然是有限的，到达售票窗口前的顾客总体可看作是无限的，因不存在最大限制数。上游流入水库的河水，就是非可数无限顾客总体的例子。

Ⅱ）到达的类型。是单个到达，还是成批到达。在库存问题中，进货看作顾客，就是成批到达的例子。

Ⅲ）相继顾客（批）到达的间隔时间分布。

令$T_0=0$，T_n表示第n（$\geqslant 1$）个顾客的到达时刻。在顾客单个到达情形，我们有

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

令 $\tau_n=T_n-T_{n-1},\quad n=1,2,3,\cdots$

τ_n是第n个顾客到达时刻与第n－1个顾客到达时刻之差，称为第n个到达间隔时间（interarrival time）。

一般均假定$\{\tau_n\}$独立同分布。其分布函数记作A（t）。关于$\{\tau_n\}$的分布，在排队论中将它分类为如下几种情形。

1.定长分布（D）（deterministic）

顾客规则地等距时间到达。如每隔时间c_0到达一个顾客。我们有

$$A(t)=P\{\tau_n\leqslant t\}=\begin{cases}0, & \text{若}t<c_0,\\ 1, & \text{若}t\geqslant c_0.\end{cases}\qquad(1.1)$$

医院予约的病人，如规定每半时到达一个；产品通过传送带进入包装箱都是定长输入的例子。

2.最简单流（或称Poisson流）（M）（Markov）

若{ τ_n }独立、同负指数分布，密度函数为

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

则称到达过程服从参数为λ的最简流（Poisson流）。平均（期望）间隔时间为$\frac{1}{\lambda}$，方差$D[\tau_n]=\frac{1}{\lambda^2}$，$E[\tau_n^k]=k!E^k[\tau_n]$。

令N（t）表示o到t时刻到达的顾客总数。具有N（o）＝0（这个限制可去掉）。随机过程{ N（t），t≥o }如具有如下性质而称为“流”：对任给的t≥o，N（t）仅取非负整数值；概率为1的样本函数是非降右连续的（如见方开泰等〔65〕）。

对任给的t_1，t_2满足$o\leqslant t_1<t_2$，令

$p_n(t_1,t_2)=P\{N(t_2)-N(t_1)=n\}$，$n\geqslant o$，

它表示在（t_1，t_2〕里到达n个顾客的概率。现在我们来给出最简单流的一个等价定义。

一个流{ N（t），t≥o }是最简单流的充要条件是它满足如下三个条件：

①对任给的t（≥o），一个增量△t，有

$$p_1(t,t+\triangle t)=\lambda t+o(\triangle t),$$

其中λ（>o）是个常数（它不依赖于t，反应了流{ N（t），t≥o }的平稳性；从$p_1(t,t+\triangle t)$的表达式，也反应了{ N（t），t≥o }的有限性）；o（△t）表示

$$\lim_{\triangle t\to o}\frac{o(\triangle t)}{\triangle t}=o,$$

这个条件意即在（t，t＋△t〕恰好到达一个顾客的概率为λt＋o（△t）；

②独立性。在不相交的时间区间内到达的顾客数是相互独立的。即{ N(t),t≥o }是独立增量过程；

③普通性。$\sum_{n=2}^{\infty}p_n(t,t+\triangle t)=o(\triangle t)$，这意即在（t，t＋△t〕内到达两个及其以上的顾客的概率是可忽略不计的，当△t足够小时。

这个等价性的证明见方开泰等〔65〕，对最简单流，我们能证明（见第二章§2）

$$P\{N(t)=k\}=\frac{(\lambda t)^k}{k!}e^{-\lambda t},\qquad t>o,\quad k=0;\quad 1,\quad 2,\quad \cdots \quad (1.2)$$

在长为t的时间内到达顾客的平均数$E[N(t)]=\lambda t$，方差 $D[N(t)]=Var[N(t)]=\lambda t$。

这种输入，应用最为广泛，而且数学上易于处理，有重要的特性，我们将在第二章详细地予以讨论。

3. k阶爱尔朗（Erlang）输入（E_k）

{ τ_n }独立、同k阶爱尔朗分布，其密度函数为

$$a(t)=\frac{\lambda(\lambda t)^{k-1}}{(k-1)!}e^{-\lambda t}\qquad t\geqslant o, \qquad (1.3)$$

其中k为正整数，其分布函数为

$$A(t)=1-e^{-\lambda t}\left(1+\frac{\lambda t}{1!}+\frac{(\lambda t)^2}{2!}+\cdots+\frac{(\lambda t)^{k-1}}{(k-1)!}\right),\ t\geqslant 0 \qquad (1.4)$$

则称为爱尔朗输入。爱尔朗分布的平均数$E[\tau_n]=\frac{k}{\lambda}$，方差$D[\tau_n]=\frac{k}{\lambda^2}$。k阶爱尔朗分布实为k个相互独立、具有相同负指数分布(其密度函数为 $\lambda e^{-\lambda t}, t\geqslant 0$)的随机变量和的分布（见§2.2*）。我们有 $k=\frac{E^2[\tau_n]}{D[\tau_n]}$。爱尔朗分布是丹麦电话工程师A、K、Erlang在本世纪初叶所研究，它是熟知的Gamma分布（其密度函数为$f(x)=\frac{1}{\Gamma(\alpha)\beta^{\alpha}}X^{\alpha-1}e^{-x/\beta}, x\geqslant 0$)的特殊情形(即 $\alpha=k$, $\beta=\frac{1}{\lambda}$)。当k=1时，即为负指数分布。

4.一般独立输入（GI） (general independent)

$\{\tau_n\}$独立、同分布，称为一般独立输入。上述各种输入均是一般独立输入的特殊情形。

5.成批（batches）输入

设T'_n为第n批顾客的到达时刻（$T_0'=0$），令$\tau'_n=T'_n-T'_{n-1}$, n=1, 2, 3, …。在时刻T'_n到达的顾客数ρ_n。一般假定$\{\tau'_n\}$独立同分布，其分布可以是上述各种情形之一；$\{\rho_n\}$是独立同分布。若还假定τ'与ρ_r也独立。这种输入称为平稳无后效流。

6.具有不耐烦（impatient）顾客的输入

到达的顾客发现排队过长，就不加入队列而离去，叫做有障碍（balke）；也可能顾客在队列中等了一定时间之后，决定离去，这叫做放弃(renege)，如见Gavish与Schweitzer〔76/77〕；在有多个平行队列的情形，顾客可能中途从一队转换到另一个队。这些统称为具有不耐烦顾客的输入。

7.非平稳输入。

相继到达时刻的间隔的分布类型与分布的参数值，均与服务系统运行时间的长短无关的，称为平稳输入。以上1到5均是平稳输入。若相继到达时刻的间隔的分布或(和)参数值随时间变化的，称为非平稳输入。如电话系统，呼叫流就是非平稳的，而具有明显的周期性。

8.其它输入

如有调度员的输入：当有多个平行的排队队列时，顾客到达由调度员按排加入哪个队，因他对各个服务设备的性能瞭如指掌。半马氏输入（如见Cinlar〔67$_2$〕）。

§1、4、2 排队规则

排队规则刻化顾客接受服务的先后次序。也分几种情形。

1.损失制（Lossing system）

---

*) §2.2，表示第二章第二节。以后我们采用类似的符号。

当一个顾客到达时，若所有服务机构均被占用，它就自动离去，永不再来（这仅是为了数学上处理方便所作的假定，乃实际问题的一种近似）。如损失制电话系统。

2.等待制（Waiting system）

当顾客到达时，若所有服务机构均被占用，他就加入队列，等待服务。如长途电话系统。服务员选取顾客进行服务所遵循的规则有如下几种。

①先到先服务。即按顾客到达的先后次序接受服务。这是最普遍的情形。如自由售票窗口。

②后到先服务。在许多库存系统中就是这种情形，只要存放物品是不变质的。如将钢板存入仓库，看成是顾客到达，需要时，总是从最上面取出，即后到先服务。在情报系统中，愈是后来的信息，往往愈重要，更应先译出来。

③随机服务。当一个服务结束时，从等待的顾客中随机地选取一个进行服务。如市内电话的接线员。这是唯一可行的选取规则。

④优先权服务。进入服务系统的顾客有不同的优先权（由重要程度或顾客花钱购买），具有较高优先权的顾客先于低优先权的顾客得到服务，而不管他们到达的先后次序。优先权的等级可以有许多种。而且同种优先权的顾客同时在服务系统中多于一个，他们被选取的先后次序也必须刻划。可以是前面诸规则的任何一种。

优先权规则又分为两类。一类叫做强拆的（Preemptive）。即当新到达的顾客的优先权高于正在接受服务顾客的优先权，则终止正进行的服务，直接服务新到的具有较高优先权的顾客；当较高优先权的顾客服务结束时，系统中也再无较高优先权的顾客，再次对被强拆的顾客进行服务。另一类是无强拆。当较高优先权的顾客到达时，正在接受服务的顾客的优先权虽然较低，仅当它的服务完成之后，从等待的顾客中选取具有最高优先权的顾客进行服务。

如码头载有重要物资的船只先行卸货；电报分普通报、急电、加急电、军政报等；长途电话对市内电话有强拆权。在电子侦察中，常常把雷达威胁按制导、炮瞄、轰炸、瞄准、引导和警戒的次序分类排队，并是强拆型。

⑤几个服务机构的情形

这可以有多种排队规则。如所有顾客排成一个公共队。队首到先结束服务的机构处去接受服务，或别的方式接受服务；到达的顾客按下规则在每个服务台前各排成一个队：第i个、第n＋i个、第2n＋i个，…顾客排入第i个队（i＝1，2，…，n）；按下方式排成几个队：第m（m＝1，2，3，…）个顾客到达时，以概率$p_i^{(m)}$排入第i个队（i＝1，2，…，n）。对每个队仍可按前面四种规则之一进行服务。

⑥循环排队

如纺纱（织布、落筒）工，看管m台机器，她总是在各机器之间按固定线路巡回，遇到哪台出故障了，就处理哪台。

3.混合制（损失制与等待制的混合）

①队长有限制的情形。

等待空间有限制的情形，最多只能容纳k个顾客在服务系统中。当顾客到达时，若

队长小于k，就加入队列；若队长等于k，该顾客就离去，且永不再来。如水库问题，库容是有限的；旅馆的床位也是有限的。

②等待时间有限制的情形。

顾客在系统中的等待时间不能超过给定的t_0，超过t_0，顾客就自动离去，永不再来。如易损电子元器件的库存问题，超过一定时间元器件就失效了。如见Gavish与Schweitser〔77〕

③逗留时间（等待时间与服务时间之和）有限制的情形。

顾客在系统的逗留时间不能超过事先给定的t_0，若超过，顾客就离去。如用高射炮射击敌机，当敌机飞越高炮射击空域的时间为t_0时，在这个时间内还未被击落，也就不可能被该防空阵地击落，就算消失。

应指出：损失制与等待制可看成是混合制的特殊情形。如$k=C$（C是服务系统服务员数）就变为损失制；$k=\infty$就变为等待制。

§1、4、3 服务机构

完全刻划服务机构有以下几个方面：服务员的数目，在多个服务员的情形，是串联或是并联；在任一时刻接受服务的顾客数，即是单个地进行服务还是成批地进行服务；服务时间的分布。分时计算机就是并行地进行处理，旅游车载客游览就是成批服务的例子。在计算机网络里我们必须同时处理串、并联的复杂系统。

下面描述各种服务时间分布。设服务时间V的分布函数记作B（t），密度函数（如果有）记作b（t）。

1、定长分布（D）

每个顾客的服务时间是一个常数c，因此

$$B(t)=p\{V\leqslant t\}=\begin{cases}0, & \text{当}t<c\\ 1, & \text{当}t\geqslant c\end{cases} \tag{1.5}$$

2、指数分布（M）

各个顾客的服务时间是相互独立，具有相同的负指数分布：

$$B(t)=\begin{cases}0, & \text{当}t<0,\\ 1-e^{-\mu t} & \text{当}t\geqslant 0.\end{cases} \tag{1.6}$$

密度函数为

$$b(t)=\begin{cases}0, & t<0,\\ \mu e^{-\mu t}, & t\geqslant 0.\end{cases} \tag{1.7}$$

其中$\mu>0$，为一常数；其平均服务时间

$$E[V]=\frac{1}{\mu}=\int_0^{\infty} t\,\mu e^{-\mu t}\,dt \tag{1.8}$$

方差为

$$D[V]=\int_0^{\infty}\left(t-\frac{1}{\mu}\right)^2\mu e^{-\mu t}\,dt=\frac{1}{\mu^2} \tag{1.9}$$

变异系数