

排队论基础 及应用

孟玉珂

同济大学出版社

排队论基础及应用

孟玉珂 编著

同济大学出版社

内 容 提 要

系统介绍排队论的基本知识,着重叙述在随机服务系统各方面的应用。第1、2章为预备知识,第3章为生死排队系统,第4章为机器维修模型,第5章为爱尔朗排队系统及马尔可夫排队网络,第6、7、8章为一般排队系统,第9章为离散型排队系统。

责任编辑 许纪森

封面设计 王肖生

排队论基础及应用

孟玉珂 编著

同济大学出版社出版

(上海四平路1239号)

新华书店上海发行所发行

同济大学印刷厂印刷

开本: 850×1168 1/32 印张: 11.875 字数: 320 千字

1989年10月第1版 1989年10月第1次印刷

印数: 1—3000 定价: 6.60元

ISBN 7-5608-0161-7/O·25

前 言

排队论是研究排队现象的理论和应用的学科，是专门研究由于随机因素的影响而产生的拥挤现象的科学，有人也称之为随机服务系统，或称之为公用事业的数学方法。公共服务事业中许多重要因素是随机的，往往会出现拥挤现象，电话通话就是这种典型的问题之一。本世纪初，丹麦工程师爱尔朗(Erlang)用概率论方法研究了电话通话问题，从而开创了排队论这门学科。近几十年排队论得到迅速发展，它的理论已相当深入，应用范围已扩大到铁路、公路、水路、航空的运输管理，城市交通管理及控制，车站、码头装卸，仓库贮运，水库蓄水，工厂工人看管机器，制定生产方案，生产任务分配，商店的订货与销售方案的制定，商店和服务行业机构设置及运营，医疗服务，计算机设计、使用和维修，可靠性及军事运筹学等领域。总之，凡是具有公共服务性质的事业和工作，凡是出现拥挤现象的领域，都是排队论的用武之地。从表面现象看，各种拥挤现象杂乱无章、毫无头绪，但利用排队论知识，人们可以找出其中的规律，从而利用或控制这种现象的进程，使我们对这种现象的管理科学化。正因为如此，排队论已越来越受到公用事业机构的设计人员和管理人员的重视。我国四化建设需要大批优秀的管理人才，许多大专院校都十分重视管理技术人才的培养，纷纷成立管理专业或系，或成立管理学院。在管理学院、管理系科的教学，大都开设排队论课程。排队论已成为一个优秀的设计人员和管理人才的必备知识。

尽管从社会和学校两方面看，对排队论知识的需求是迫切的，但到目前为止尚未见到一本系统介绍排队论基本知识的书在我国正式出版。本书的编写目的就是想给学习管理类专业、计算机专业、通讯专业的研究生、本科生及企业的管理人员和设计人员及有关的教师们提供一本入门读物，也可作教材使用。本书是按照读者已具有初等概率论知识（相等于大学工科本科30—40学

时的概率论内容)和初等微积分知识的基础上安排材料的,并充分考虑到与已有知识的衔接。在材料选取上,着眼于那些应用上最感兴趣最重要的内容,限于选稳定态的结果而不选瞬态的结果。为了使读者对排队论有一个较完整的了解,我们系统地介绍了排队论的基本知识和研究排队论的基本方法。我们希望本书能达到这样的目标:学完本书后,读者能用所学的知识和方法去解决实践中遇到的一些简单排队问题,或者能引起某些读者的兴趣,从而立志于排队论的研究并进而去攻读一些更专门的著作。

本书的内容大部分曾在我校运输系、经济系、材料系、计算机系及管理科学研究所、信息研究所的研究生和应用数学专业高年级本科生中讲授过几遍,本书是在原讲义的基础上补充改写而成。本书编写中主要参考了书[1~5]及董泽清编的《排队论及其应用》(内部讲义)。本书的一部分应用题也取自它们,离散时间排队问题在计算机及通讯中有广泛应用,已日益受到人们的重视。鉴于批量运输与装卸是非常普遍和普通的事,因而对批量排队问题作了初步介绍。本书包含大量应用例题和模型,其中也包含作者部分研究成果。为巩固所学内容,第2—7章附有练习题。

本书是按这样的次序安排的:第1章为排队系统概论,第2章为预备知识,第3—9章是本书的主要内容,这几章基本上是按方法分类:包括生死过程排队方法(3、4章)、相位方法(5章§1—§3)、马尔可夫网络方法(5章§5、§6)、嵌入马尔可夫链方法(6、7章)、因式分解法(8章),第9章是介绍离散时间的排队问题,方法上与前几章平行。3—5章是初等部分,6、7章是中等部分,8章更难一些,9章相应也分成三部分。

在我国排队论的研究和应用是从50年代末开始发展起来的。我国既重视理论研究,也紧密结合社会主义建设实际开展应用研究,并取得了不少成绩,同时也培养了一批人才。然而四化建设事业的迅猛发展又对我们提出了更高的要求,当前我国正掀起学习管理科学的高潮,我们应珍惜人们的求知欲望,向广大管理人员和技术人员普及一些排队论知识。如果我们的管理人员和技术

人员能够懂得并会应用一些排队论知识，我们的管理水平必将有很大提高，从而大大促进我们的公用事业使之发挥出更大的经济效益和社会效益。

最后，作者对侯振挺教授表示衷心的感谢，侯教授审阅了本书全部手稿并提出了重要的修改意见。感谢陈安岳副教授，他审阅了全部手稿并和作者一起对部分章节的修改进行了细致的研究和确定。由于他们的工作，使本书增色不少。

由于作者水平所限，书中存在的问题肯定还是不少的，恳切希望专家和读者提出宝贵意见并指正。

孟玉珂

1988年4月，

目 录

第一章 随机服务系统概论.....	(1)
§ 1 基本概念.....	(1)
§ 2 排队系统几个重要特征.....	(7)
§ 3 排队系统的符号表示.....	(9)
§ 4 排队系统最优化问题.....	(10)
第二章 几个重要的概率分布、几个重要的随机过程.....	(12)
§ 1 输入分布、输入流、服务分布、输出流.....	(12)
§ 2 几个重要的连续型分布.....	(14)
§ 3 几个重要的离散型分布.....	(22)
§ 4 普阿松过程.....	(26)
§ 5 离散时间的马尔可夫链.....	(36)
§ 6 连续时间的马尔可夫链.....	(47)
§ 7 生死过程.....	(54)
第二章 练习题.....	(59)
第三章 无限源的生死排队系统.....	(63)
§ 1 $M/M/1/\infty$ 排队系统.....	(63)
§ 2 有可变输入率的 $M/M/1/\infty$ 排队系统.....	(78)
§ 3 Little 公式——队长与等待时间关系.....	(83)
§ 4 可变服务率的 $M/M/1/\infty$ 排队系统.....	(86)
§ 5 服务台的忙期、闲期.....	(89)
§ 6 $M/M/\infty$ 排队系统——无穷多服务台排队系统	(91)
§ 7 $M/M/m/\infty$ 排队系统——有 m 个服务台情形	(93)
§ 8 具有不耐烦顾客的 $M/M/m/\infty$ 排队系统	(102)
§ 9 $M/M/1/K$ ——混合制排队系统(一).....	(105)

§ 10	$M/M/m/K$ ——混合制排队系统(二).....	(112)
§ 11	$M/M/m/m$ ——损失制排队系统.....	(117)
§ 12	M/M 排队系统输出过程.....	(120)
§ 13	队长分布与到达时刻看到的队长分布的关系	(123)
	第三章 练习题.....	(125)
第四章	有限源的生死排队系统及其变种——机器维修模型	(130)
§ 1	$M/M/m/n/n$ 维修模型—— n 部机器 m 个维修工情形	(130)
§ 2	损失制 $M/M/m/n$ 模型——恩格塞特公式	(141)
§ 3	有备用品的 $M/M/m/n + N/n$ 模型.....	(143)
§ 4	二阶段循环排队系统.....	(147)
§ 5	排队论方法在可靠性分析中的应用举例——依系统故障程度、派维修工与抢修工的系统可靠性.....	(151)
	第四章 练习题.....	(162)
第五章	爱尔朗排队系统与马尔可夫排队网络.....	(165)
§ 1	相位法、 r 阶爱尔朗分布.....	(165)
§ 2	$M/E_r/1$ 排队系统.....	(167)
§ 3	$E_r/M/1$ 排队系统.....	(175)
§ 4	爱尔朗排队模型与批量排队模型的关系.....	(182)
§ 5	一般马尔可夫排队网络.....	(187)
§ 6	有反馈的循环排队系统.....	(200)
	第五章 练习题.....	(204)
第六章	普阿松到达、一般服务的 $M/G/1$ 排队系统.....	(207)
§ 1	嵌入马尔可夫链.....	(207)
§ 2	转移概率矩阵.....	(208)
§ 3	平稳分布.....	(211)
§ 4	平均队长.....	(213)

§ 5	等待时间分布	(214)
§ 6	例题	(219)
§ 7	三种分布(到达时队长分布、平稳分布、离开时队长分布)的关系	(226)
§ 8	批量到达 $M^x/G/1$ 排队模型	(232)
§ 9	剩余寿命分布	(245)
	第六章 练习题	(247)
第七章	一般输入、负指数服务分布的 $GI/M/m$ 排队系统	(249)
§ 1	嵌入马尔可夫链的转移概率矩阵	(249)
§ 2	到达时刻队长的平稳分布	(254)
§ 3	等待队长的平稳分布	(256)
§ 4	等待条件下的等待时间分布	(256)
§ 5	平均队长	(257)
§ 6	平衡状态下等待时间分布	(257)
§ 7	$GI/M/1$ 排队系统	(258)
§ 8	$GI/M/1$ 排队系统队长平稳分布	(264)
§ 9	$GI/M/2$ 排队系统	(268)
§ 10	批量到达的 $GI^*/M/1$ 排队系统	(272)
§ 11	$GI/E_r/1$ 排队系统	(281)
	第七章 练习题	(282)
第八章	一般到达、一般服务的 $GI/G/1$ 排队系统	(284)
§ 1	林德勒积分方程法	(284)
§ 2	等待时间的平稳分布	(288)
§ 3	例题	(291)
§ 4	重服务强度下等待时间分布的近似公式	(299)
第九章	离散时间排队系统	(304)
§ 1	引言	(304)
§ 2	$Geom/Geom/1$ 排队系统	(306)
§ 3	$Geom/Geom/m$ 排队系统	(313)

§ 4	$Geom/G/1$ 排队系统	(315)
§ 5	批量到达 $Geom^x/G/1$ 排队系统	(323)
§ 6	$GI/Geom/1$ 排队系统	(332)
§ 7	批量到达 $GI^x/Geom/1$ 排队系统	(340)
§ 8	一般到达、一般服务的 $GI/G/1$ 排队系统	(342)
附 录		
1.	母函数	(347)
2.	拉普拉斯-斯蒂尔阶斯变换 ($L-S$ 变换)	(348)
3.	排队论的符号	(352)
4.	排队论重要结果一览表	(354)
主要参考书目		(370)

第一章 随机服务系统概论

§ 1 基本概念

排队是日常生活和工作中常见的现象。例如等公共汽车排队，到商店购物排队，交款排队，到医院看病等待排队，买火车票排队，托运行李排队，取货排队，这是人的排队。还有另一种排队，例如文件等待打印或发送，报告等首长批示，路口红灯下面的汽车、自行车等待通过路口，这是物或设备排队。排队现象由两个方面构成。一方有需求，要求得到满足，通常称为要求得到服务，我们把要求服务的人或设备称为**顾客**。一方要设法满足对方的要求给予服务，我们把服务人员或服务机构称为**服务员或服务台**。顾客与服务台构成一个**排队系统**，或称之为**随机服务系统**。很显然，缺少顾客或服务台任何一方都不会出现排队现象。

顾客源源而来，受许多因素的影响，顾客到达服务台的时刻是随机的。服务台也受许多因素的影响，服务完一个顾客的时间长短也是随机的。因此，不可避免地会出现排队现象。例如，在某一时刻要求服务的顾客数超过服务台的服务容量时，顾客就必须排队等待服务，直到服务台出现空闲时才得到服务，服务完后离开。由于顾客到达服务台时得不到及时服务，服务台前形成排队，从而出现拥挤现象。因此，哪里有拥挤现象，哪里就有排队论的问题。排队论是研究拥挤现象的一门科学。

排队现象不只是以有形的形式出现，还可以无形的形式出现。例如，向某单位打电话正好占线打不通，这就需等一会儿再拨号，等候在各分散的电话机前的人形成无形的排队。发生故障的机器等待维修，靠码头的轮船或进站的车辆等待装卸，进入调车场的列车等待编组，飞机因跑道不空等待着陆，货物等待运往仓库或销售地等都是无形排队的例子。

排队系统是由顾客一方和服务台一方组成的，从整个社会效益考虑，既要照顾到顾客一方的利益，又要照顾到服务台一方的利益，因此在设计排队系统时，应使整个系统达到最优。事实上，为了减小排队长度，可行的方法是增加服务设备，这样就需增加服务机构的投资。另一方面，服务台过多可能造成较多的空闲时间也会造成浪费。如果服务设备太少，排队现象会严重，给顾客造成的损失会增加。如果排队等待时间过长，顾客等待不耐烦大批离开服务系统，服务机构也会遭受损失。排队论的研究目的就在于求出各种排队系统的规律性，使设计人员掌握这种规律，设计出最优化的排队系统，使管理人员掌握这种规律，调整与控制排队系统使它处于最佳运营状态，从而使社会能得到最大效益。

排队论研究的内容有三个部分：

1. 排队系统的性态问题。研究各种类型的排队系统的规律性。一个排队系统排队的长短（队长）是随机的，顾客等待时间的长短（等待时间）是随机的，服务忙的时间长短（忙期）是随机的。研究排队系统的规律主要是研究队长分布，等待时间分布，忙期分布，包含瞬时状态及统计平衡条件下的稳定状态两种。

2. 排队系统的统计推断。为了了解和掌握一个正在运营的排队系统的规律，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符合哪种模型，从而可用相应的理论来研究和解决该排队系统的问题。

3. 排队系统的最优化问题。排队系统在运营中需对它进行控制，使之达到最优运营状态，这就是动态最优化。要建立一个排队系统，首先要给定排队系统的一些指标和特征，设计人员根据给定的指标和特征对排队系统进行设计，使排队系统达到最优化，这是静态最优化。设计最优化，运营最优化，这是我们学习和运用排队论的目的。

排队系统的性态问题，是排队论的核心内容，它是排队系统

的统计推断及最优化问题的基础，本讲义将着重介绍这方面的内容。从实用角度来讲，更关心统计平衡条件下排队系统的状态，我们介绍的内容也限于稳定状态。对瞬时态的问题感兴趣的读者可阅读有关专著。

§ 1.1 排队系统的一般表示

排队系统尽管千差万别，但都可以抽象为顾客到达服务台前，服务台有空闲便立刻得到服务，若服务台不空闲，则需等待服务台出现空闲时再接受服务，服务完后离开服务台。因此排队模型可用图 1.1 表示。图中虚线框图为排队系统。顾客要求得到服

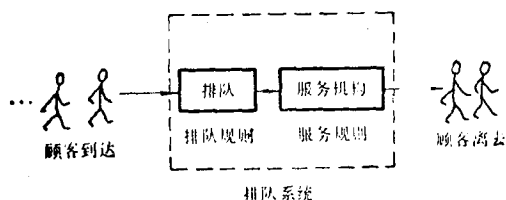


图 1.1 排队系统图

务，不断到达服务台，顾客数量超过服务台容量便形成排队，等待服务。排队，服务机构组成一个排队系统。顾客到达，排队等待，服务机构给予适当的服务以满足顾客的需求，顾客离开服务机构，这四个环节便组成一个排队模型。任何一个排队模型都可用上述抽象的方式进行表示。

由于现实世界变化万千，因而排队现象也是多种多样的，因此对顾客和服务台应作广泛的理解，它们可以是人也可以是某种物质或设备。排队可以有形的，也可以是无形的。为了更好地说明顾客、要求服务的内容及服务机构的多种多样性，我们用表 1.1（见本书第 4 页）给出一些排队现象的例子。

日常见到的拥挤现象种类繁多，机制也是各式各样的，有的顾客来源有限（如工厂等待维修的机器），有的顾客来源是无限的（如电话呼唤），有的排队系统顾客是离散地到达、离散地离

表 1.1

序 号	到 达 的 顾 客	要 求 服 务 内 容	服 务 机 构
1	到货场的车辆	装车（卸车）	装卸工（或机械）
2	进入调车场的车辆	编 组	调 车 员
3	驶入港口的货船	装 卸 货	码 头
4	提 货 单	提取货物	仓库管理员
5	提款或存款入	提款或存款	银行出纳员
6	电话呼唤	通 话	交 换 台
7	病人就医	诊断或手术	医 生
8	文 件 稿	打字或排印	打字员或排字工
9	进入阵地的敌机	被 击 毁	高 射 炮
10	需上计算机的题目	计 算	计 算 机
11	上游河水进入水库	放水调节水位	水闸管理员
12	出故障的机器	修 理	修理工人

开，有的排队系统顾客是连续地到达且连续地离开（如河水进入水库，水库放水）。服务方式也五花八门各式各样。数学上研究这种拥挤现象的方法是抛去其具体内容，找出它们的共同本质。人们要问，是哪几个因素决定了这种排队系统的进程呢？经过研究，人们发现，顾客到达的规律、服务的规律，排队方式及服务规则起着决定性作用。因此，首先应了解它们。

§ 1.2 排队系统的三个组成部分

尽管排队系统是各式各样的，但从决定排队系统进程的因素来看，它是由三个基本部分组成的，这就是输入过程、排队规则及服务机构。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的。首先应说明顾客的数量，即顾客的来源是有限的还是无限的。例如，工厂内停工待修的机器是有限源的，电话呼唤是无限源的。其次说明顾客达到服务机构是成批到达的（每批数量是随机的还是非随机的）还是单个到达的，即说明顾客的到达方式。

三是说明相继到达的顾客（批或单个）之间的**间隔时间**的分布是什么。是定长分布还是随机分布，分布参数是什么，到达的间隔时间之间是否独立，分布是否随时间变化即是否平稳。

2. 排队规则

排队规则是指服务机构允许不允许排队，顾客愿意不愿意排队，在排队等待情形下服务的顺序是什么。分成下列三种情形。

（1）损失制系统。顾客到达服务机构，如果所有服务台都被占用，服务机构又不允许顾客等待，于是顾客只好自动离去到其他地方接受服务或被迫放弃服务要求。这些顾客都叫离去，而且假定他们不返回。例如，旅店客满谢绝客，戏院客满拒绝售票，医院挂号限额等均属于这种情形。

（2）等待制系统。当顾客到达服务机构时，如果所有的服务台都被顾客占用无空闲，这时该顾客就自动地加入队列排队等待服务，服务完才离开。例如货车到站场等待装卸排队，人们到食堂就餐排队，买火车票排队，发电报排队等均属于这种情形。

（3）混合制系统。这是损失制与等待制混合组成的系统。服务机构只允许有限个顾客等待，多余的顾客被迫离开永不再回，这种系统就是混合制系统。另外还有顾客见到队长不愿等待，看到队短则耐心等待，或顾客的等待时间不能超过某时间 T 否则就离开都是属于混合制系统。例如，药品出厂供病人使用，用量是随机的，超过有效期则作废，是顾客等待时间有限的例子。

具有等待性质的排队系统，相应的服务规则主要有以下几种：

（1）先到先服务，即顾客按到达先后顺序接受服务。这是最普遍最常见的服务形式。

（2）后到先服务，即服务顺序与到达的顺序相反。例如仓库同一类东西堆放在一起，一般是后进的货放在上面，取货时是后进的先运出。情报系统、天气预报资料总是后到的消息越重要要先处理。

（3）随机服务，一个服务结束，服务员从等待的顾客中随机

地选取一个顾客进行服务。电话员接线就是用这种方式工作。

(4) 有优先权的服务。在排队系统中某些顾客有时特别受到重视, 在服务顺序上给以特殊待遇, 让他们优先得到服务。例如病情严重的病人在医院先治疗, 荣誉军人优先购票, 火车站有母子候车室, 带小孩的旅客先进站。

优先权也是有等级的, 如一单位接到长途电话, 还在进行的内部电话暂停, 先通长途, 长途电话结束, 内部电话再接着通话, 这叫强拆优先权。另一种是无强拆优先权, 即接受服务的顾客服务完后, 具有优先权者再接受服务。优先购票属于这种情形。

排队和服务模型是各式各样的, 上述几种情况不可能全部概括, 我们只是介绍了几种常见的或比较重要的情形。除去上述情形外还有许多其它情形, 比如循环排队, 排队网络, 循环服务; 有几个服务台时, 到达的顾客可以排成一个队也可以按一定要求排成几个队; 服务台的多少也可根据排队等待人数多少进行调节或是增加或是减少等。由于模型不同, 研究和处理的方法也不同, 因此在遇到实际问题时, 必须灵活使用数学方法和模型。

在排队系统的研究中, 排队的长度和服务规则无关, 而顾客在系统中的等待时间及逗留时间的长短和服务规则有关, 不同的服务规则直接影响顾客在系统中耗费时间的长短。

3. 服务机构

服务机构主要指服务台的数目, 服务机构有多个服务台进行服务时服务的方式是并联的还是串联的, 接受服务的顾客是成批的还是单个的, **服务时间**服从什么分布, 顾客们接受服务的时间是否独立。例如学生注册、机车车辆进厂检修是串联服务的例子, 运煤装车是成批服务的例子。列车到编组站编组也是成批服务的例子, 而且每批(一列车)包含的车辆数(顾客)也可近似看作是随机的。⁴

抛开排队系统的实际背景, 抽象出它们的数学结构, 可以看出, 每一个排队系统都由三部分即输入过程、排队规则和服务机构唯一决定。不同类型的排队模型的差别也是由于输入过程、排

队规则或服务机构不同而形成的。因此在研究一个排队系统前，首先要弄清这三部分的具体内容。由于三部分的具体内容和结构不同，在数学处理上往往会采用不同的方法才能得到解决。

§ 2 排队系统几个重要特征

对于一个排队系统设计的优劣及运营的好坏，涉及到顾客和服务机构的利益，也有社会效果好坏的问题或效益大小的问题。顾客和服务机构关心的是队长、顾客在系统中的等待时间与逗留时间及其分布、服务台忙期的分布等。这些指标是排队系统最重要的特征。下面分别加以说明。

§ 2.1 队长

我们把系统中正在接受服务及等待服务的顾客数目称为**队长**。由于顾客到达是随机的，服务台的服务时间是随机的，因此队长也是随机的。顾客一方及服务机构一方都十分关心这个量的平均值及其分布。这是排队系统的一个重要指标。设 t 时刻系统的队长为 $N(t)$ ，而排队等待的顾客数为 $N_q(t)$ 也叫**等待队长**。队长、等待队长都取非负整数，它们的分布相应记作

$$p_n(t) = p(N(t) = n), \quad n = 0, 1, 2, \dots$$

$$q p_n(t) = p(N_q(t) = n), \quad n = 0, 1, 2, \dots$$

t 时刻平均队长为 $E[N(t)]$ ，平均等待队长为 $E[N_q(t)]$ 。在实际中关心的是在统计平衡条件下的平稳值，即 $t \rightarrow +\infty$ 时，上述几个量的极限值。若上述极限存在，分别记作：

$$p_n = \lim_{t \rightarrow +\infty} p_n(t), \quad n = 0, 1, 2, \dots$$

$$p'_n = \lim_{t \rightarrow +\infty} q p_n(t), \quad n = 0, 1, 2, \dots$$

$$\bar{N} = \lim_{t \rightarrow +\infty} E[N(t)] \quad \bar{N}' = \lim_{t \rightarrow +\infty} E[N_q(t)]$$

§ 2.2 顾客在系统中的等待时间（逗留时间）

从顾客抵达排队系统直到接受服务这段时间称为顾客的**等待**