管理科学丛书

# 运筹学

## 原理与实践

下册

D. T. 菲利浦斯等 著
运筹学教研室 译

上海交通大学机电分校
管理工程系

管理科学丛书

# 运筹学

## 原理与实践

下册

D. T. 菲利浦斯等　著
运筹学教研室　译

上海交通大学机电分校
管理工程系

· 管理科学丛书 ·

**运 筹 学：原 理 与 实 践**

（下 册）

菲 利 浦 斯 等 著

运 筹 学 教 研 室 译

---

上海交通大学印刷厂 印刷

上海交大机电分校 发售

787×1092毫米 17½印张 41万字

1983年9月第1版 1983年9月第1次印刷

印数 0,001—10,000

定 价： 1.80元

# 译 者 前 言

运筹学自从40年代兴起以后，在国外得到迅速的发展。我国在50年代引进这一学科，并定名为“运筹学”。1956年中国科学院数学研究所成立了运筹学研究室。1958年前后，不少理工科大学相继开设了运筹学课程，目前一些财经院校也增设了这门课程。各校使用的教材很不一致。

我校在1979年准备为企业管理专业学生开这门课时，曾对国内已有的教材和美、日、德、法的一些大学教材进行了选择。我们感到，运筹学没有完全定型，并且与我国的生产实际还存在着很大的距离。不少教材把运筹学作为一门数学分支来处理，也有不少教材为了强调实用只对理论作了一些通俗的介绍。我们觉得，这本由 Phillips，Ravindran 和 Solberg 三人合著的名为《运筹学：原理与实践》的教材，对材料的处理比较合适。书中对线性代数和概率论的必要知识作了有重点的复习，因而在运筹学的理论上处理得比较严谨。书中所举的例子也是尽量结合实际，深入浅出，不同于一般的通俗处理。

我校曾参照该书，针对我国的国情，就其部分材料改写了一本试用教材，使用效果尚好，也受到一些兄弟院校的欢迎和采用。但是运筹学这门学科在我国的发展时间还不长，适当地介绍一些国外较好的教材看来还有必要，为此我们先将这本书的原文译本印出来，供大家参考。译文所根据的本子是第一版第七次印刷。

本书原作兼供研究生和大学本科生之用；在我国，用作工科院校本科生的教本，完全适宜。此外，对具有大学水平的工程技术人员开阔知识领域，也是有益的。本书的全部习题解答同时译出，以供参考。

本书由我校管理工程系运筹学教研室教师协作译出。参加译述工作的有叶庆桐（第一——四、八章）、张优德（五——七章）、徐克绍（第九章、索引）、林同曾（第十章）、梁捷纾（第十一章、附录）。全书译文最后由林同曾、叶庆桐校订；徐克绍、蒋仲刚整理。限于水平，错误之处在所不免，希望读者批评指正。

一九八二年十二月

# 下 册 目 录

# 第七章 排队模式

## §7.1 引　　言

本书大部分章节着眼于方法，使读者对于方法有所理解、有所熟悉，以便根据他的特殊问题，建立自己的模式。这一章却不是这样，考虑的是一类问题；不论什么方法只要适合都可以利用，目的是使读者能理解现实世界中这类问题的特征。

这章主题是排队问题或等待问题。更一般地说，是研究积压问题的原因和对策。似乎不用说这样的研究是值得的。在日常生活中，由于积压所增加的任务，迫使我们去努力弄懂造成延误的现象。然而，这类现象是不简单的；排队不总是需要的人过多或者服务员过少而引起的必然后果。以后，我们将会看到，不考虑随机变化是无法理介排队的。当涉及到平均值时，人们的直觉是很可靠的，但当涉及到随机现象变化时，一般的常识不总是正确的。因此当我们使用随机过程的方法推导排队模式时，会产生一些意想不到的结果。这些模式的重要价值就在于都具有这样的特性，它能显示出我们从通常经验中得不到的真理。只有对它们有了深刻的理解之后，我们才有可能对付现实生活中的积压问题。

## §7.2 一个确定型模式

我们开始用一个简单的确定状态的模式，来描述一种情况，这种情况后面还要用更复杂的方法，即用马尔柯夫过程的方法来描述。这样做有两个原因：一是引进一些没有复杂计算的排队论概念，二是为了表明如果你不了解随机过程，你可能会得到一些能引起误解的结果。

假定有一人（称服务员）在执行某项零件或部件的制作任务(称事项或顾客)，这些任务以周期性的间隔到达他的工作台。我们假定每隔3分钟到达一项任务，这个人执行任务需要一段时间（称服务时间），并假定这时间为2分钟。至于这些任务在到达之前或服务完之后发生什么，我们不感兴趣。

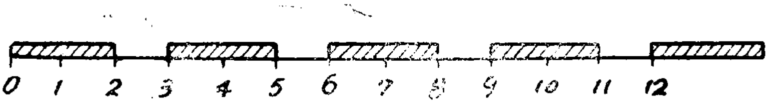


图 7.1 服务员的工作状况

假定我们用图表示一个服务员在不同时间的工作状况，当一项任务刚到达就开始服务，情况就象图7.1所表明的那样。在图中，阴影线表示服务正在进行的时间，而空白表示服务员空闲。显然从长远来看，服务员有1/3的时间空着。略为思考一下，就可以得出

结论说，服务员的空闲时间为总时间的 $(a-s)/a$ 或 $1-s/a$，其中 $s$ 为服务时间，$a$ 为二项任务到达之间的时间，如果要以时间利用率来表示，那末服务员的忙期为总时间的 $s/a$，当然假定 $s\leqslant a$。

无疑，按一般的常识，更不用说按传统的管理原则，都会认为服务员的时间没有充分利用。假如这个过程是能控制的话，例如能控制两个到达之间的时间间隔的话，就将很直观地要求提高到达的速度（即减少到达之间的时间）直到服务员没有空为止。这就等于要求服务的时间利用率增加到 1 。这将在 $s=a$ 时发生，即到达时间间隔正好是服务时间。任何企图进一步减少两个到达之间的时间，将导致"瓶颈"的阻塞现象。即任务的到达速度大于服务完成的速度，由此形成一个不断增长的队伍。

概括一下，这个模式告诉我们：

如果 $s<a$，则服务员的空闲时间为总时间的 $[1-(s/a)]$；

如果 $s=a$，则服务员使用了全部服务时间，但没有形成队伍。

如果 $s>a$，则将形成一个不断增长的等待服务队伍。

用这个模式的人（也许他没有认识到）可能认为第一和第三种情况同样都是不合适的。于是他努力使服务率和到达率保持"平衡"。这或许是工业工程中的概念。对于同样的情况，在完成一个马尔柯夫模式后，将重新检验这个看上去似乎最佳的原则。

在建立一个正式的随机模式之前，让我们非正式地先来考虑引进随机变量在到达和服务过程中的作用。我们假定到达的发生是随机的时间间隔以代替固定的时间间隔。这个随机的时间间隔的平均值为 3 分钟（或更一般地，为 $a$ 分钟）。类似地，我们假定服务时间平均为 2 分钟（或 $s$ 分钟）。这样，一个新的现象发生了，即当两个到达之间的时间间隔小于第一项任务完成的时间时，则第二位到达者必须等待一些时间之后才能得到服务。换句话说，形成了队伍。这两个时间都是平均值，所以，这种排队情况不一定经常发生，但它必将会偶然发生。甚至同样情况可能接连地发生两次或多次，以致可以想象，这队伍也会偶然排得相当长。

注意，平均而言，到达过程和服务过程恰好是同我们第一个模式的情况一样。我们现在说的排队是暂时的，波动的队，是因为到达和服务时间的随机性所产生的，它不同于服务员超负荷工作所形成的排队，因为那种排队会造成瓶颈阻塞现象恶化。当到达时间间隔和（或）服务时间为随机时，即使平均服务时间 $s$ 小于平均到达时间间隔 $a$，也可能发生排队。这种现象是由于服务和（或）到达的时间间隔的随机性所造成的，如果上述服务和到达的间隔是固定的，那末这样的情况就绝对不会发生。因为这个随机性是很复杂的，这样自然就会在考察时忽略它而把服务和到达时间间隔作为定值来处理，认为把两个相互影响的因素的"平均"状态来简化这些可变量就足以对系统的平均状态作出予计。然而这样做的话，就会取消我们正在设法理解的真实现象。

## § 7.3 排队参数

排队论是研究随机波动的等待问题，如我们在上述例子中所描写的。在推导马尔柯夫模式应用于同样的例子之前，让我们先来考虑一下那些与排队有关的基本因素和那些可能发生的变动。

仍用本章开始的例子，很明显一个服务员不一定指一个人，到达也不一定指一个另件或部件。实际上，过程的状态最主要的是在时间方面。所谓“服务”几乎指的是任何事情在某一时间间隔内占用一个人，一个设备或一个地点，并在服务完成之前阻止后续到达者利用它。类似地，所谓“到达”就是不论什么事，只要在特定时刻出现并要求服务的就是“到达”。除了诸如在市场上结账、售票窗口等具有服务和到达的明显排队状态之外，排队问题还包括：在打电话时等待交换台接通，飞机等待跑道，文件等待打字，计算机程序等待输入等等。

到达过程可以用几种方法来描述，一般最方便的办法是采用相当于到达时间间隔的随机变量来描述（见图7.2）

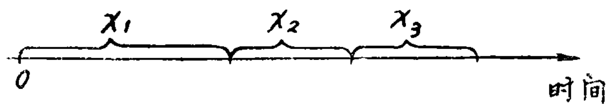


图 7.2 到达过程

这些随机变量几乎总是些假定相互独立并具有相同的分布，因此在这种情况下，指出随机变量的分布形式（如负指数分布，Erlang 分布，一般分布）和规定它的参数就足够了。

类似地服务过程通常也用随机变量 $S_i$ 表示，$S_i$ 代表 i 个到达顾客的服务时间。$S_i$ 也几乎总是假定为相互独立并且有相同的分布。这样也只要指出分布的形式和具有的参数就足够了。把服务过程假定独立于到达过程，意思是服务时间的长短不依赖于到达在什么时间发生（同样，到达的发生也不依赖于服务时间的长短），但要注意，服务时间只有在到达发生以后才能开始，在这个意义上，服务过程与到达过程就是有关的了。

图7.3表示通常用来代表具有一个服务员的排队图。有时，系统往往有几个服务员，这样就如图7.4那样表示。

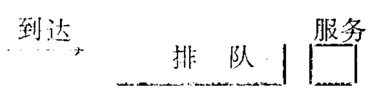


图 7.3 一个服务员的排队

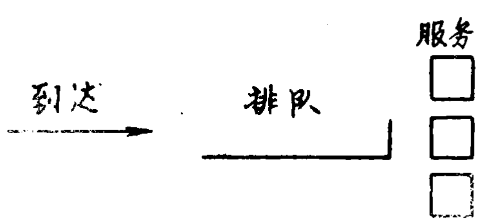


图 7.4 多个服务员的排队

讨论各种形式的排队时有一种特殊的符号，称作肯道尔(Kendall)符号。即约定用符号$(i/j/c)$叙述排队形式，其中 $i$ 和 $j$ 代表字母，每个字母表示几个可能分布形式中的一个。第一个字母表示到达过程，第二个字母表示服务过程。最后一项 $c$ 代表一个数，即表

示服务员数。规定分布用的字母为：

M——负指数分布(negative exponential)

$E_K$——爱尔朗—K分布(Erlang—K)，

D——定长分布(Deterministic or fixed)

GI——到达时间间隔的一般分布(general distribution of interarrival times)，

G——服务时间的一般分布(general distribution of service times)。

于是，一个排队具有到达时间间隔为负指数分布，具有3个服务员，每个服务员的服务时间为Erlang-2分布，则这个排队系统可表示为$M/E_2/3$。

在某些情况下，限定排队的长度为预定的容量是很重要的。另一些情况下，可以允许排队的容量是无限的。Kendall符号不区别这两种情况。此外，从总体中到来的顾客可以是无限的，在这种情况下，它的到达速度不受正在等待服务的顾客数的影响。或者从总体中到来的顾客是有限的，在这种情况下，当等待服务数增加时，到达的速度必然会减少（在极端情况下，当总体中的所有顾客都进入系统时，就不再有新的到达发生）。我们称这两种情况分别为“无限源”和“有限源”。除另作说明外，一般都假定为“无限源”。

如图所示，到达的服务对象似乎总是按到达的顺序依次服务的，但不一定总是这样规定，任务或顾客服务次序的各种规定称“服务规则”(queue discipline)。其中常用的有：

FCFS——先到先服务(First come, First served)

LCFS——后到先服务(Last come, First served)

RSS——随机选择服务(Random selection for service)

PR——优先服务(Priority)

最后一个表达方式是不很明显的，因为优先服务可能有许多种不同的优先条件。

为了能在排队问题的符号中包括这些附加条件，在1971年国际排队符号标准化会议上，同意增加三项说明，扩大肯道尔符号，现在的符号读作：

（到达过程/服务过程/服务员数/在系统中顾客极限数/在源中顾客数/排队规则。）

如果它是 $(\cdot/\cdot/\cdot/\infty/\infty/FCFS)$

形式，则后两项或三项可以省略，只要写$(\cdot/\cdot/\cdot/\infty)$或$(\cdot/\cdot/\cdot)$就可以了。换句话说，除非另有说明这些项都取标准形式。即使有了这样冗长的符号，要表示你所想要的不同模式，还是不够用的。例如，有几个服务员，其服务速度各不相同，符号中就没有提供所要表示的内容。此外，这些符号仅表示排一个队，但在实际的服务系统中常常包含几个队。但尽管有这些缺点，这些符号已包含了很多有用的情况。

从排队模式中，可以期望求出一些不同的结果，其中较常用到的有：

$L$＝在稳态下，系统中的期望单元（人）数（包括正在服务的单元（人数)），

$L_q$＝在稳态下，等待服务的期望单元（人）数（除去正在服务的单元人数)。

$W$＝在稳态下，一个单元(人)在系统中期望花费时间（包括服务时间)。

$W_q$＝在稳态下，一个单元(人)在系统中期望用于排队的时间(除去服务时间)

上述四者是相关的，如果知道其中一个，则其余三个就很容易求得。

## § 7.4 M/N/1 排 队

现在我们再回到本章开始时作为确定型模式处理的那个单个服务员的随机过程模式。因为这个模式在许多可行的排队模式中是最简单的，我们将作详细的推导。通过对这一模式的理解，可为以后许多有用的其它模式打下基础。

在前面我们假定需要的服务零件（人）的到达时间间隔固定为 $a$ 和服务时间间隔固定为 $s$，现在我们假定到达和服务时间间隔都为随机变量，分别具有平均值 $a$ 和 $s$。为了满足马尔柯夫性，迫使我们假定这些随机变量都是负指数分布。其他分布可能合适，但将不允许采用简单的连续时间马尔柯夫技术。如果 $\lambda$ 为到达时间间隔的分布参数(即到达速度)，$\mu$ 为服务时间参数（即服务速度），则 $a=1/\lambda$，$s=1/\mu$。在现实世界状态中，$a$ 和 $s$ 可以通过到达和服务时间间隔的抽样来估计平均值。$\lambda$ 和 $\mu$ 的值将由它们相应的倒数给出，这些值足以描述分布的全部特性。

我们定义马尔柯夫过程的状态为系统中任一时刻具有的部件（人）数，包括正在服务的那一个在内。如果我们不限定系统的容量，则可能的状态为 0,1,2,3……，趋向无限。因为到达是一次来一个这样发生的，系统中的数每来一次就增加一个，因为服务是一次完成一个，系统中的数每服务完一次就减少一个，这个过程是“生灭过程”。此外，因为目前系统中的数不影响到达速度和服务速度，状态转移速度保持不变。总之，其转移图如图7.5所示：

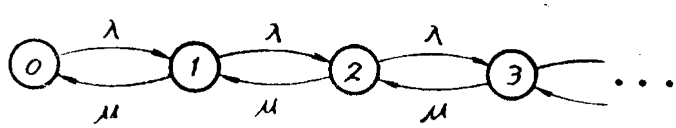


图 7.5 $M/M/1$ 转移图

从这一观点开始，马尔柯夫过程已很好地确定了，此后可用正常的一套来分析了。然而，有必要停下来考虑一下建立模式时的一些假设。首先看到，如果我们忽略服务完成的情况，只考虑到达的总数，我们将得到一个普通的泊松过程。因此，我们可以单独地把到达过程描述为泊松过程。然而我们不能用同样的方法来描述服务过程，因为单独地考虑服务过程是没有意义的。只有在系统中有服务对象时，才有服务过程。注意到转移图中很自然的表示了两者不同的指向；从每一点有一个箭离开，对应于到达转移，另一个箭对应于服务完成，但仅当这些点的“顾客”数大于零时才有。

当过程处于非零的任一状态时，下二个状态是不肯定的。可以认为，到达和服务完成两者竞赛的结果决定了过程向哪个方向发展。假定系统开始是空的，并且第一个正好到达，那末下一个状态是什么，取决于服务员在下一个顾客到达之前是否完成对第一个顾客的服务。假定服务员没有完成，第二个顾客在第一个顾客服务完成之前就到达了。那末下一个状态是什么，将取决于服务员在第三个顾客到达之前是否能完成对第一个顾客的服务。现在我们或许会有这样的假设：因为第一个顾客在第二个顾客到达之前已经开始受到服务，

余下的服务时间就短了一些，因此，很可能服务员在再一次到达发生之前，已完成了对第一个顾客的服务，但这样的论述与马氏性假设有矛盾。第二个一到达（进入达态 2 ），服务员就“忘记”已经为第一个服务所花的时间，这余下的服务时间的分布与整个服务时间的分布是相同的。如果认为，这样的说法太荒谬，则我们就不能使用马氏模式。另一方面，如果我们坚持认为过去所花的时间会影响余下服务时间的分布（或者直到下次到达为止的余下服务时间），则我们必须求助于更加复杂的数学描述。这样甚至转移图也会失去它的意义，因为这样就必须在确定怎样离开某一状态之前，知道它的进入状态。

回忆一下转移速率矩阵的形式：

$$\begin{pmatrix} -\lambda & \lambda & & 0 \\ \mu & -(\lambda+\mu) & \lambda & \\ & \mu & -(\lambda+\mu) & \lambda \\ & \cdot & \cdot & \cdot \\ & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot \\ & \cdot & \cdot & \cdot \end{pmatrix}$$

这个矩阵可无限地扩展右下角。但它的无限性并不防碍问题的解决，因为它的结构是正规的。

状态转移函数 $P_{ij}(t)$ 的微分方程表达式是容易立出的。可用两个方程式给出如下：

$$\frac{dP_{i0}(t)}{dt}=-\lambda P_{i0}(t)+\mu P_{i1}(t) \qquad \text{当任一} i\geqslant 0$$

$$\frac{dP_{ij}(t)}{dt}=-\lambda P_{ij-1}(t)-(\lambda+\mu)P_{ij}(t)+\mu P_{ij+1}(t)$$

$$\text{当任一} i\geqslant 0,\ \text{任一} j>0$$

从一般形式的连续时间马氏过程的方程得

$$\frac{dP(t)}{dt}=P(t)\lambda$$

在这个意义上，$P_{ij}(t)$ 表示系统在时刻为零时有 $i$ 个顾客，而在时刻 $t$ 时有 $j$ 个顾客的概率，从这个 $P_{ij}(t)$ 我们能有效地确定几乎所有有关的事情。

遗憾的是，这微分方程组看上去非常简单，可是很不容易求解。这个解已经求得，但它包含有变形的贝塞尔函数，它是那样复杂以致不能直观地提供排队的状态。一般说来，去求得排队模式的 $P_{ij}(t)$ 函数既不可行，也无必要。因此，几乎所有的已知解都是针对稳态的情况。当然要知道在短时间内将发生什么，那就很遗憾了。但是，另一方面，关于为排队系统的设计或运行作的大多数决策是针对相当长的时间内的情况，这就使稳态的解足以满足要求。

所考虑的模式的稳态方程为：

$$0=-\lambda\Pi_0+\mu\Pi_1$$
$$0=\lambda\Pi_0-(\lambda+\mu)\Pi_1+\mu\Pi_2$$

$$0=\lambda\Pi_1-(\lambda+\mu)\Pi_2+\mu\Pi_3$$
$$\vdots \quad \vdots \quad \vdots \quad \vdots$$

通式为

$$0=\lambda\Pi_{j-1}-(\lambda+\mu)\Pi_j+\mu\Pi_{j+1} \qquad 当 j>0$$

但第一个方程是特殊情况。因为第一个方程仅含 $\Pi_0$ 和 $\Pi_1$，它可用来求 $\Pi_1$：

$$\Pi_1=\left(\frac{\lambda}{\mu}\right)\Pi_0$$

代入第二方程，使它仅含 $\Pi_0$ 和 $\Pi_2$ 得

$$\Pi_2=\left(\frac{\lambda}{\mu}\right)^2\Pi_0$$

继续使用这方法，我们得一般形式为

$$\Pi_j=\left(\frac{\lambda}{\mu}\right)^j\Pi_0$$

为了保证这个形式确实是无限个方程的全部解，需要把它代入通式中来检验。余下的未知数可代入下列正则方程(normalizing-equation)求得

$$1=\Pi_0+\Pi_1+\Pi_2+\cdots$$

$$1=\Pi_0+\left(\frac{\lambda}{\mu}\right)\Pi_0+\left(\frac{\lambda}{\mu}\right)^2\Pi_0+\cdots$$

$$1=\Pi_0+\left[1+\left(\frac{\lambda}{\mu}\right)+\left(\frac{\lambda}{\mu}\right)^2+\cdots\right]$$

$$1=\Pi_0\left(\frac{1}{1-\frac{\lambda}{\mu}}\right)$$

$$\left(只要\frac{\lambda}{\mu}<1\right)$$

$$\Pi_0=1-\frac{\lambda}{\mu}$$

注意到无限级数收敛的要求 $\frac{\lambda}{\mu}<1$，对这个必要条件的含义下文即将加以考虑。

总之，稳态概率可由下式给出：

$$\Pi_0=\left(1-\frac{\lambda}{\mu}\right)$$

$$\Pi_1=\left(\frac{\lambda}{\mu}\right)\left(1-\frac{\lambda}{\mu}\right)$$

$$\Pi_2=\left(\frac{\lambda}{\mu}\right)^2\left(1-\frac{\lambda}{\mu}\right)$$

$$\vdots \quad \vdots \quad \vdots$$

或用

$$\Pi_j=\left(\frac{\lambda}{\mu}\right)^j\left(1-\frac{\lambda}{\mu}\right) \qquad j\geqslant 0$$

这是几何分布的形式。

因为这两个参数$\lambda$和$\mu$总是作为一个比值同时出现的，习惯上就用一个参数$\rho$来代替，即$\rho=\frac{\lambda}{\mu}$。这样稳态概率就简化为：

$$\Pi_j=\rho^j(1-\rho)$$

这个新参数$\rho$称为"服务强度"(traffic intensity)。显然它是到达速度与服务速度之比，但也可以从另一个角度看这个比值。它是到达速度($\lambda$)乘以平均服务时间($1/\mu$)，这样就可以把$\rho$看作为在一个平均服务时间内到达顾客的平均值。如以我们开始时所用的$a$和$s$（平均到达时间间隔和平均服务时间）来表示，则

$$\rho=\frac{\lambda}{\mu}=\frac{\left(\frac{1}{\mu}\right)}{\left(\frac{1}{\lambda}\right)}=\frac{s}{a}。$$

回忆一下，本章开始所讲到的确定型模式，为了避免排队不断增长，服务强度$\rho=\frac{s}{a}$必须小于或等于1。在马氏模式中要使$\frac{\lambda}{\mu}$（或$\rho$）小于1，它是获得正则方程的解所必须的，这两个要求基本上是相同的。所不同的是确定型模式所要求的$\frac{s}{a}=1$，不仅是可行，而且是希望得到的。它表示到达速度和服务速度处于完全"平衡"状态。但是我们在马氏模式中得到的解并不包括$\rho=1$这个情况。我们以后就要分析这种情况。

对$\rho$的另一种理解是从稳态解中获得的。$\Pi_0$是系统空着时的稳态概率。它可以被看作是服务空闲的那部分时间(从长远来看)。因此$(1-\Pi_0)$可看作是服务繁忙的那部份时间。因为$1-\Pi_0=1-(1-\rho)=\rho$，因而我们可以把$\rho$解释为服务员的有效利用率。模式又一次提示：如果控制得了的话，为充分使用服务员起见，应使参数$\rho$接近于1。

现在我们有了系统中顾客数的概率分布（在稳态中）就可以计算系统中顾客数的平均值$L$。

$$\begin{aligned}L&=\sum_j^{\infty} j\Pi_j\\&=0\Pi_0+1\Pi_1+2\Pi_2+\cdots\\&=0(1-\rho)+1\rho(1-\rho)+2\rho^2(1-\rho)\cdots\\&=(1-\rho)(\rho+2\rho^2+3\rho^3+\cdots)\\&=\frac{\rho}{1-\rho}\end{aligned}$$

这个无限级数仅当$\rho<1$时趋向于有限和，但这是事先假定了的。

我们也可以用$\Pi_j$计算稳态系统中队长$L_q$的平均数。当系统中只有一个服务员时，在队中有$j$个部件（顾客）的概率等于在系统中有$j+1$个部件（顾客）的概率。唯一的例外是没有人在排队，在这种情况下，系统中的数（顾客）或者为零或者为1。因此，

$$\begin{aligned}L_q&=0(\Pi_0+\Pi_1)+1\Pi_2+2\Pi_2+\cdots\cdots\\&=1\rho^2(1-\rho)+2\rho^3(1-\rho)+\cdots\cdots\end{aligned}$$

$$=(t-\rho)(\rho^2+2\rho^3+3\rho^4+\cdots\cdots$$

$$=(1-\rho)\rho^2\frac{1}{(1-\rho)^2}$$

$$=\frac{\rho^2}{1-\rho}$$

请注意，$L$系统中（顾客的）平均数，和$L_q$队中（顾客的）平均数的差是：

$$L-L_q=\frac{\rho}{1-\rho}-\frac{\rho^2}{1-\rho_0}=\rho\left(\frac{1-\rho}{1-\rho}\right)=\rho$$

这是可以预料的，因为早已说明，$\rho$可被看作为占用服务的部分时间。

图7.6表示$L$和$L_q$为服务强度（或服务员利用率）$\rho$的函数，在许可的范围内，$0\leqslant\rho<1$。这个图能表明，当我们试图充分地利用服务员时，将是什么情况——那就是排队人数或系统中人数的期望值趋向于无限大。如果说队长偶尔趋向无限大，或者队长有可能趋向无限大，这已是很不好了；但结果甚至更坏，它表明平均数的队长要趋向无限大。

假如认为这个结果似乎很不正常，以致要怀疑到模式本身，那你必须了解，为什么在我们日常生活中没有见到过队是排得无限长的。因为每当一个队越来越长时，人们就会自动调整，而改变这个系统。换句话说，即使我们设法达到$\rho=1$，我们决不会见到这个系统趋向稳态。如果允许系统运行下去，那末这个系统的模式是能正确地反映所要发生的情况的。

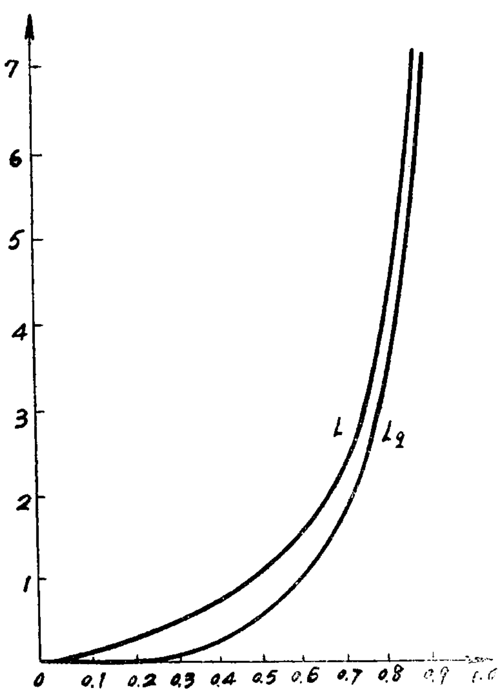


图 7.6　L，$L_q$ 作为$P$的函数图