目　录

序 …………………………………………………………………………………… (1)

第一章　线性规划 ……………………………………………………………… (1)
1.1　问题和模型 …………………………………………………………… (1)
1.2　线性规划问题解的性质 ……………………………………………… (5)
1.3　单纯形方法 …………………………………………………………… (10)
1.4　对偶线性规划问题 …………………………………………………… (20)
1.5　参数线性规划和灵敏度分析 ………………………………………… (26)
1.6　微机操作——调用 *YAJ* 中，*LP* 决策支持系统 ………………… (33)

第二章　整数线性规划 ………………………………………………………… (37)
2.1　整数规划的例子 ……………………………………………………… (37)
2.2　整数规划的解法——分支定界法 …………………………………… (39)
2.3　微机操作——调用 *YAJ* 中，*IP* 决策支持系统 ………………… (44)

第三章　运载问题 ……………………………………………………………… (48)
3.1　例题和模型 …………………………………………………………… (48)
3.2　解决平衡问题的表上作业法 ………………………………………… (49)
3.3　微机操作——调用 *YAJ* 中，*TRP* 决策支持系统 ………………… (51)
3.4　不平衡运输问题 ……………………………………………………… (53)

第四章　指派问题 ……………………………………………………………… (55)
4.1　例题和模型 …………………………………………………………… (55)
4.2　匈牙利法 ……………………………………………………………… (56)

4.3 微机操作——调用 YAJ 中,ASMP 决策支持系统 …… (59)
4.4 对象与活动数目不等情形的指派问题 …… (61)

第五章 网络模型 …… (62)
5.1 基本概念 …… (62)
5.2 最短通路问题 …… (65)
5.3 最小生成树问题 …… (69)
5.4 最大流问题 …… (71)
5.5 微机操作——调用 YAJ 中,NET 决策支持系统 …… (73)

第六章 关键路线法 …… (76)
6.1 统筹图 …… (76)
6.2 关键路线 …… (80)
6.3 网络参数计算 …… (81)
6.4 突击施工分析 …… (83)
6.5 微机操作——调用 YAJ 中,CPM 决策支持系统 …… (85)

第七章 计划评审技术 …… (87)
7.1 计划评审技术网络 …… (87)
7.2 工序的时间估计 …… (89)
7.3 网络图的时间参数及计算方法 …… (90)
7.4 关键路线 …… (92)
7.5 事件按时完成的概率 …… (93)
7.6 微机操作——调用 YAJ 中,PERT 决策支持系统 …… (94)

第八章 动态规划 …… (96)
8.1 最短路线问题 …… (96)
8.2 动态规划中的几个名词和符号 …… (98)
8.3 背包问题 …… (100)
8.4 生产和库存问题 …… (103)
8.5 微机操作——调用 YAJ 中,DP 决策支持系统 …… (106)

第九章　库存论 …………………………………………………… (108)
9.1　库存论中的几个要素 ………………………………………… (108)
9.2　确定性库存模型 …………………………………………… (110)
9.3　折扣分析 …………………………………………………… (116)
9.4　随机性库存模型 …………………………………………… (117)
9.5　微机操作——调用 YAJ 中，INVT 决策支持系统 ……………… (122)

第十章　排队论 …………………………………………………… (124)
10.1　排队论的基本概念 ………………………………………… (124)
10.2　顾客到达数的分布和服务时间的分布 ……………………… (128)
10.3　单服务台的 [M/M/1] 模型 ………………………………… (132)
10.4　单服务台的 [M/G/1] 模型 ………………………………… (146)
10.5　多服务台的排队模型 ……………………………………… (150)
10.6　微机操作——调用 YAJ 中，QUEUE 决策支持系统 ………… (153)

第十一章　排队系统模拟——系统仿真 ……………………………… (158)
11.1　模拟概述 …………………………………………………… (158)
11.2　伪随机数发生器 …………………………………………… (164)
11.3　模拟实例 …………………………………………………… (169)
11.4　微机操作——调用 YAJ 中，QSIM 决策支持系统 …………… (171)

第十二章　决策/概率论 …………………………………………… (176)
12.1　决策的概念 ………………………………………………… (176)
12.2　确定型决策问题 …………………………………………… (178)
12.3　风险型决策问题 …………………………………………… (179)
12.4　不确定情况下的决策 ……………………………………… (190)
12.5　微机操作——调用 YAJ 中，DSPB 决策支持系统 …………… (194)

第十三章　马尔科夫 (Markov) 过程 ………………………………… (199)
13.1　正规随机矩阵 ……………………………………………… (199)
13.2　马尔科夫链 ………………………………………………… (200)
13.3　马尔科夫分析 ……………………………………………… (204)
13.4　微机操作——调用 YAJ 中，MKV 决策支持系统 …………… (207)

第十四章　时间序列预测——现象变动的趋势分析 …………………… (209)
14.1　统计预测的一般问题 ……………………………………… (209)
14.2　平均预测法 ……………………………………………… (210)
14.3　趋势预测法 ……………………………………………… (215)
14.4　线性回归预测——最小平方法预测 ………………………… (219)
14.5　预测误差分析 …………………………………………… (221)
14.6　微机操作——调用 *YAJ* 中，*TSFC* 决策支持系统 …………… (225)

参考文献 …………………………………………………………… (228)

序

一、发展生产力离不开现代管理

当前，我们国家正按照邓小平同志关于建设有中国特色的社会主义的理论，一心一意搞经济建设。这就要求加强现代管理，提高效益，努力发展社会生产力。所谓管理，就是管理者为了达到某种目标，运用管理职能作用于管理对象，有意识地不断进行有效且协调的综合活动过程。加强管理，对于我国实现“四化”具有特殊的重要性。不少国外专家对我国当前工业状况进行考察后，一致认为急需解决的问题，第一是管理，第二是管理，第三还是管理。

现代管理内容，就现代企业来讲，包括计划管理、生产管理、技术管理、质量管理、人事劳动管理、物资管理、成本管理、财务管理等等；宏观社会当然还有现代科学管理，现代教育管理。本书侧重讲现代管理的数学方法。

已经形成的系统工程方法是其中最典型的综合管理方法。系统工程是综合性组织管理技术，它也是迄今最大的一门科学管理总体方法。它是立足整体，统筹全局，把分析与综合、定性与定量结合起来，全面组织管理各系统，使之达到整体效果最佳的管理综合技术。我们讲述的正是系统工程实施中的每个具体步骤均要使用的数学模型的最优化技术——运筹学方法。

二、运筹学的产生和发展

凡是要求做出肯定决策的任何问题均属于运筹问题。虽然，自从有了人，运筹问题就已经存在，可是直到第二次世界大战时才出现运筹学（Operations Research）这一名词，并且形成了现代运筹学的科学方法。

第二次世界大战后，随着大企业的出现，许多企业经营状况更加复杂。因此管理者不得不进行分区管理，并且使企业总的效能达到最优。这就要求用一套科学的组织管理方法。在20世纪后半期，又出现了以运筹学为中心内容的新的工程技术——系统工程。

运筹学的迅速发展归功于数字计算机的同时发展。例如,线性规划的单纯形法是在1947年由George B. Dantzig发明的,但它却因为现实条件不具备,直到20世纪50年代中期和末期才得以实现和运用。计算机帮助了许多运筹方法(时至今日还在应用)的发展和(或)实现。运筹学和计算机紧密相关,因为如果没有计算机,运筹学只不过是一种理论科学,不会像现在这样不断的发展。

本书一方面讲清楚运筹学诸分支的数学模型、基本概念、基本理论、计算方法,同时提供了相应的计算机软件的实际应用。

三、运筹学问题的分类

运筹学问题分类,我们在目录中罗列了十四章,它们几乎包含了生产、生活的各个方面。软件YAJ恰是这些内容:

(一)线性规划
(二)整数线性规划
(三)运载问题
(四)指派问题
(五)网络模型
(六)统筹方法①——关键路线法
(七)统筹方法②——计划评审技术
(八)动态规划
(九)库存论
(十)排队论
(十一)排队系统模拟
(十二)决策/概率论
(十三)马尔科夫过程
(十四)时间序列预测

四、运筹学的数学模型

为了得到问题的最优解,通常更有意义和更方便的是用数学形式将问题写出来。这种数学描述或表达称为问题的数学模型。

在过去25年中,已有一批具有适当解法的独特模型为人们所熟悉。例如,线性规划模型、动态规划模型、库存模型、排队模型,这些模型已有现成的解法。任何一个运筹方案的目标是要决定问题的最合适的数学模型,然后用已有方法去解,或者寻求新的解法。

在实际中，有可能无法为给定的问题建立数学模型，或者虽可建立模型，但求解困难（比如无现成方法可用，需要计算机贮量太大，或上机时间太长等等）。对于这种情况，就用直观法或试探法去解（如，Hungarian method 求解资源分配问题）。此外，我们还可以求助仿真研究复杂的大系统。

本书是针对大家熟悉的数学模型的通常解法，给出计算机算法，读者可以得到各种各样问题的解及其分析，以提高求解问题的能力。

五、本书的适用范围

最先考虑的自然是从事实际工作的工程技术人员，他们可以用这些理论和方法去解决与之有关的最优化问题。

本书对于一切运用运筹学知识的人都是适用的，这包括一切管理人员、企业家等等。他们不必去深究运筹学的理论，只要按照书中介绍的方法操作，就可以得到满意的结果。

本书也适合大专院校的师生，那些正在研究和学习应用数学知识——特别是运筹学知识的人们。

本书提供了一个很好的计算机应用软件，对于微机应用人员也是一本难得的参考书。

汪遐昌

2005 年 8 月

第一章　线性规划

线性规划是目前管理中经常使用而又有成效的优化技术。

线性规划和其他学科一样，是由于生产力发展的需要而产生和发展的。随着现代生产的规模越来越大，各部门之间的相互联系越来越紧密和复杂，在生产的组织与计划、交通运输、财贸等方面都要求有新的数学方法来为它们服务。因此，早在20世纪30年代末40年代初，康托洛维奇（Конторвич）和希奇柯克（Hitchcock）等在生产组织和运输问题等方面就开始研究应用线性规划这一数学方法；后来，在20世纪40年代末又由旦茨基（Dantzig）等人提出了单纯形方法并进一步从理论上给线性规划奠定了基础。

计算机的不断发展，不仅为线性规划方法在经济活动中的广泛应用提供了可能性，而且加快了线性规划方法的发展。在1951年，国际水平只能解约束条件为10个方程的线性规划问题，到了1963年就能解1000～10 000个方程的线性规划问题了。另一方面，从时间上看，解一个67个方程的线性规划问题，1956年时需要1小时，到1963年只要28秒钟。当然，现在的情形更是大不一样了。

1.1　问题和模型

1.1.1　运输问题

设某种物资有 m 个产地：$A_1, A_2, \cdots, A_m$，产量分别为 $a_1, a_2, \cdots, a_m$。它们联合供应 n 个销售地：$B_1, B_2, \cdots, B_n$，销量分别为 $b_1, b_2, \cdots, b_n$。又，从 A_i 到 B_j 单位货物运价为 c_{ij}，如表1.1所示。应如何调运，才使总运费最少？

表 1.1

运价(元/吨) 销售地 / 产地	B_1 B_2 … B_n	产量(吨)
A_1	c_{11} c_{12} … c_{1n}	a_1
A_2	c_{21} c_{22} … c_{2n}	a_2
…	… … … …	…
A_m	c_{m1} c_{m2} … c_{mn}	a_m
销量(吨)	b_1 b_2 … b_n	

解 假定从 A_i 运至 B_j 的物资为 x_{ij},得到问题的数学模型:

$$\min z = \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}x_{ij}$$

$$s.t.\begin{cases}\sum_{j=1}^{n} x_{ij} \leqslant a_i, i = 1,2,\cdots,m\\ \sum_{i=1}^{m} x_{ij} \geqslant b_j, j = 1,2,\cdots,n\\ x_{ij} \geqslant 0, i = 1,\cdots,m; j = 1,\cdots,n\end{cases}$$

这里,$\min z$ 表示求 z 的最小值;$s.t.$ 即 subject to(受限于…),后跟约束条件。

1.1.2 布局问题

某生产队要在 n 块土地 $B_1,\cdots,B_n$ 上,种植 m 种作物 $A_1,\cdots,A_m$。各块土地亩数、各种作物计划播种面积以及各种作物在各块地上的单产如表 1.2 所示。问应如何安排种植计划,才使得总产量最多?

表 1.2

产量(公斤/亩) 土地 / 作物	B_1	B_2	…	B_n	播种面积(亩)
A_1	c_{11}	c_{12}	…	c_{1n}	a_1
A_2	c_{21}	c_{22}	…	c_{2n}	a_2
…	…	…	…	…	…
A_m	c_{m1}	c_{m2}	…	c_{mn}	a_m
土地亩数	b_1	b_2	…	b_n	

解 假定作物 A_i 在土地 B_j 上播种面积为 x_{ij},则得数学模型:

$$\max z = \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}x_{ij}$$

$$s.t.\begin{cases}\sum_{j=1}^{n} x_{ij} \geqslant a_i, i=1,2,\cdots,m \\ \sum_{i=1}^{m} x_{ij} \leqslant b_j, j=1,2,\cdots,n \\ x_{ij} \geqslant 0, i=1,\cdots,m; j=1,\cdots,n\end{cases}$$

1.1.3 指派问题

设有 n 件工作$B_1,\cdots,B_n$ 分派给n 个人$A_1,\cdots,A_n$ 去做。设 A_i 完成B_j 的工时为 c_{ij}（$i,j=1,\cdots,n$）。问应如何指派才能使完成全部工作的总工时最少？

解　令

$$x_{ij}=\begin{cases}1 & A_i \text{ 承担} B_j \\ 0 & A_i \text{ 不承担} B_j\end{cases}$$

则得：

$$\min z=\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}x_{ij}$$

$$s.t.\begin{cases}\sum_{i=1}^{n} x_{ij}=1, j=1,2,\cdots,n \text{（每件工作只分派一人去做）} \\ \sum_{j=1}^{n} x_{ij}=1, i=1,2,\cdots,n \text{（每人只做一件工作）} \\ x_{ij}=0 \text{ 或 } 1 \quad i,j=1,2,\cdots,n\end{cases}$$

1.1.4 生产组织与计划问题

某工厂用机床$A_1,\cdots,A_m$ 加工$B_1,\cdots,B_n$ 种零件。在一个生产周期内，A_i 机床只能工作 a_i 机时；工厂必须完成 B_j 零件加工数为 b_j；A_i 机床加工 B_j 零件的时间为 c_{ij}（单位：机时／个），加工 B_j 零件的成本为d_{ij}（单位：元／个）。问在这个生产周期内，怎样安排各机床的生产任务，才能既完成加工任务，又使总的加工成本最低？

解　设在一个周期内，A_i 机床加工B_j 零件 x_{ij} 件。于是：

$$\min z=\sum_{i=1}^{m}\sum_{j=1}^{n} d_{ij}x_{ij}$$

$$s.t.\begin{cases}\sum_{j=1}^{n} c_{ij}x_{ij} \leqslant a_i, i=1,2,\cdots,m \\ \sum_{i=1}^{m} x_{ij} \geqslant b_j, j=1,2,\cdots,n \\ x_{ij} \geqslant 0, i=1,\cdots,m; j=1,\cdots,n\end{cases}$$

1.1.5 合理下料问题

设用某种原料下零件 $A_1,\cdots,A_m$ 的毛坯。根据过去经验,在一件原材料上有 $B_1,\cdots,B_n$ 种不同的下料方式,按 B_j 方式下零件 A_i 的件数为 c_{ij}。此外,需 A_i 零件 a_i 件。问应如何安排下料方式,使得既能满足需要,且用料最少?

解 设用 B_j 种方式下料的原材料数为 x_j,则有数学模型:

$$\min z = \sum_{j=1}^{n} x_j$$

$$s.t.\begin{cases} \sum_{j=1}^{n} c_{ij}x_j \geqslant a_i, i = 1,\cdots,m \\ x_j \geqslant 0, j = 1,\cdots,n \end{cases}$$

1.1.6 配料问题

设用 n 种原料 $B_1,\cdots,B_n$ 制成具有 m 种成分 $A_1,\cdots,A_m$ 的产品。对 A_i 成分的需求量不低于 a_i。B_j 原料的单价为 b_j,其每单位含成分 A_i 量为 c_{ij}。问应怎样配料,才能使产品成本最低?

解 设取原料 B_j 为 x_j 单位($j = 1,\cdots,n$),得到数学模型:

$$\min z = \sum_{j=1}^{n} b_jx_j$$

$$s.t.\begin{cases} \sum_{j=1}^{n} c_{ij}x_j \geqslant a_i, i = 1,\cdots,m \\ x_j \geqslant 0, j = 1,\cdots,n \end{cases}$$

上面我们建立了几个在经济领域中常见的实际问题的数学模型。这些实际问题尽管各式各样,但它们的数学模型却有相似的形式。表示约束条件的数学式子都是线性等式或线性不等式,表示问题最优化指标的目标函数都是线性函数。因为约束条件和目标函数都是线性的,所以我们把具有这种模型形式的问题称为线性规划问题。

线性规划问题的数学模型的一般形式为:

求一组变量 x_j($j = 1,\cdots,n$)的值,使其满足:

$$\min(\text{或 man}) z = \sum_{j=1}^{n} c_jx_j$$

$$s.t.\begin{cases} \sum_{j=1}^{n} a_{ij}x_j \leqslant b_i (\text{或} \geqslant b_i,\text{或} = b_i), i = 1,\cdots,m \\ x_j \geqslant 0, j = 1,\cdots,n \end{cases}$$

其中,a_{ij}、b_i、c_j ($i=1,\cdots,m$;$j=1,\cdots,n$) 为已知量。

若用矩阵符号 $A=(a_{ij})_{m\times n}$,$x=(x_1,\cdots,x_n)^T$,$b=(b_1,\cdots,b_m)^T$,$c=(c_1,\cdots,c_n)$,则可将模型简记为:

$$\min(\text{或}\max)\ z=cx$$
$$s.t.\begin{cases}Ax\leqslant b\ (\text{或}\geqslant b,\text{或}=b)\\ x\geqslant 0\end{cases}$$

我们称满足约束条件的一组变量的值 x_j^0,($j=1,\cdots,n$) 为线性规划问题的一个可行解,使目标函数取得最小 (或最大) 值的可行解称为最优解。

线性规划问题的数学模型是描述实际问题的抽象的数学形式。它反映了客观事物数量间的本质规律。建立数学模型时,要从实际出发,抓住最本质的因素,而将不太重要的因素去掉,建立一个既简单而又比较真实反映客观事物本质规律的模型。

1.2 线性规划问题解的性质

1.2.1 两个变量的线性规划问题的图解法

为了给后面的线性规划问题的基本理论提供较直观的几何说明,我们先介绍线性规划问题的图解法。

例 1 求解:

$$\max z=2x_1+5x_2$$
$$s.t.\begin{cases}x_1\leqslant 4\\ x_2\leqslant 3\\ x_1+2x_2\leqslant 8\\ x_1,x_2\geqslant 0\end{cases}$$

我们视 (x_1,x_2) 为坐标平面上的点。那么,满足条件中每一个不等式的点集就是一个半平面。因为约束条件是由 5 个不等式组成的,所以满足约束条件的点集是 5 个半平面的相交部分:图 1.1 中的凸多边形 $OABCD$。该多边形上的任一点,都是问题的一个可行解,该多边形点的全体,构成线性规划问题全体可行解——又叫问题的可行域。

我们可在可行域中找一个最优解。这只要将等值线 $2x_1+5x_2=z$ 平行移动:对应常数 z 取到一系列不同的值,离原点愈远,z 值愈大。于是问题成为:在上述平行线中,找出一条,使之与凸多边形相交,而又尽可能远离原点的直线。从图可见,经过 B

点的直线符合要求。易解出 $B = (2,3)$，故 $\max z = 2\times 2 + 5\times 3 = 19$。

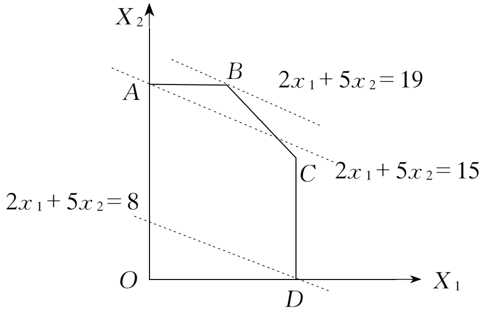


图 1.1

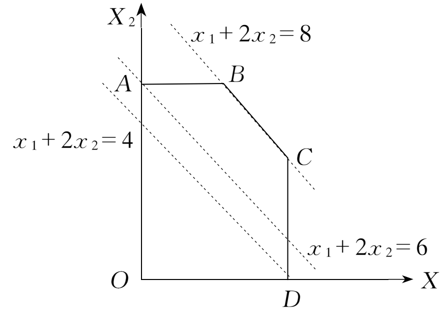


图 1.2

例 2　若把例 1 的目标函数改为 $z = x_1 + 2x_2$，那么，BC 边上每一点都是最优解（离原点最远的一条直线 $x_1 + 2x_2 = 8$ 与 BC 边重合）。这时，最优解有无穷多个，它们对应的目标函数值都是 8，如图 1.2 所示。

例 3　求解

$$\min z = 2x_1 + 2x_2$$

$$s.t.\begin{cases} x_1 - \quad x_2 \geqslant 1 \\ -x_1 + 2x_2 \leqslant 0 \\ x_1, x_2 \geqslant 0 \end{cases}$$

解答如图 1.3。

例 4　把例 3 改为求目标函数的最大值，再求解。

如图 1.3，例 3 最小值为 2；例 4 目标函数无上界，无解。

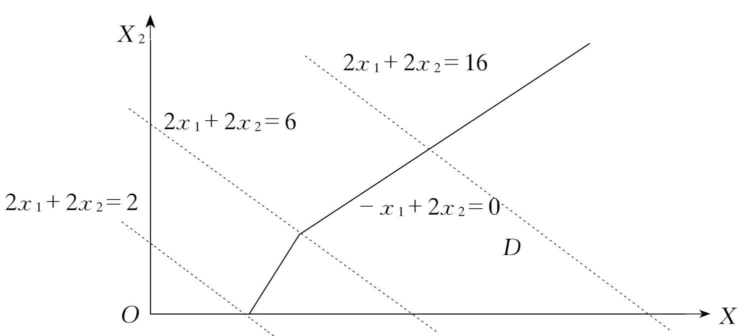


图 1.3

1.2.2　线性规划问题的标准形式

为了讨论方便，我们把线性规划问题化为标准形式，按照如下形式进行：

⑴ 如果第 k 个式子为：

$$a_{k1}x_1 + \cdots + a_{kn}x_n \leqslant b_k$$

则加入变量 $x_{n+k} \geqslant 0$，为：

$$a_{k1}x_1 + \cdots + a_{kn}x_n + x_{n+k} = b_k$$

如果第 e 个式子为：

$$a_{e1}x_1 + \cdots + a_{en}x_n \geqslant b_e$$

则减去变量 $x_{n+e} \geqslant 0$，为：

$$a_{e1}x_1 + \cdots + a_{en}x_n - x_{n+e} = b_e$$

x_{n+k}，x_{n+e} 称为松弛变量。松弛变量在目标函数中的系数为零。

⑵ 如果问题是求目标函数 $z = \sum_{i=1}^{n} c_i x_i$ 的最大值，则化为求目标函数 $z' = -z$ 的最小值。

⑶ 如果对某个变量 x_j 没有非负限制，则引入两个非负变量 $x_j' \geqslant 0$，$x_j'' \geqslant 0$。令 $x_j = x_j' - x_j''$，代入约束条件和目标函数中，使全部变量都有非负限制。

通过上述步骤，可将任何一个线性规划问题，整理成如下标准形式：

$$\min z = \sum_{j=1}^{n} c_j x_j$$

$$s.t.\begin{cases} \sum_{j=1}^{n} a_{ij}x_j = b_i, i = 1,2,\cdots,m \\ x_j \geqslant 0, j = 1,2,\cdots,n \end{cases}$$

按照前面引入的矩阵符号，可简记为：

$$\min z = cx$$

$$s.t.\begin{cases} Ax = b \\ x \geqslant 0 \end{cases}$$

1.2.3　线性规划问题解的性质

1. 线性规划问题解的一些概念

⑴ 可行解、基础可行解、最优解和基础最优解

设有线性规划问题

$$\min z = cx$$

$$s.t.\begin{cases}Ax=b\\x\geqslant 0\end{cases}$$

我们把满足约束条件的

$$x^{(0)}=(x_1^{(0)},x_2^{(0)},\cdots,x_n^{(0)})^T$$

称为线性规划问题的可行解。

若 $x^{(0)}=0$,或 $x^{(0)}$ 的非零分量所对应的系数列向量线性无关时,称可行解 $x^{(0)}$ 为基础可行解。

使目标函数取最小值的可行解,称为最优解。

使目标函数取最小值的基础可行解,称为基础最优解。

(2) 凸集

若连接 n 维点集 S 中任意两点 $x^{(1)},x^{(2)}$ 的线段仍在 S 内,则称 S 为凸集。即,若 $\{x\mid x=\alpha x^{(1)}+(1-\alpha)x^{(2)},0\leqslant\alpha\leqslant 1,x^{(1)}\in S,x^{(2)}\in S\}\subset S$ 则称 S 为凸集。

(3) 极点

若凸集 S 中的点 x,不能成为 S 中任何线段的内点,则称 x 为 S 的极点。即,若对于任意两点 $x^{(1)}\in S,x^{(2)}\in S,x^{(1)}\neq x^{(2)}$ 不存在 $\alpha\in(0,1)$,使得

$$x=\alpha x^{(1)}+(1-\alpha)x^{(2)}$$

则称 x 为 S 的极点。

2. 线性规划问题解的性质

此后,将线性规划问题简记为 (LP) 问题。

定理 1 (LP) 问题的可行解集为凸集。

证明 设其可行解集为

$$S=\{x\mid Ax=b,x\geqslant 0\}$$

并设 $x=\alpha x^{(1)}+(1-\alpha)x^{(2)}$;$\alpha\in[0,1]$。于是,对 $x=\alpha x^{(1)}+(1-\alpha)x^{(2)}$,有 $Ax=A[\alpha x^{(1)}+(1-\alpha)x^{(2)}]=\alpha Ax^{(1)}+(1-\alpha)Ax^{(2)}=\alpha b+(1-\alpha)b=b$ 且显然,$x\geqslant 0$,即 $x\in S$。

定理 2 可行解集 S 中的点 x 是极点,当且仅当 x 是基础可行解。

证明 设 x 是 S 的极点。只需证明当 $x\neq 0$ 时,其非零分量 $x_{j_1},x_{j_2},\cdots,x_{j_k}$ $(k\leqslant m)$ 对应的列向量 $P_{j_1},P_{j_2},\cdots,P_{j_k}$ $(k\leqslant m)$ 线性无关。

如果 $P_{j_1},P_{j_2},\cdots,P_{j_k}$ $(k\leqslant m)$ 线性相关,则必存在不全为零的数 $\delta_1,\delta_2,\cdots,\delta_k$,使得

$$\delta_1P_{j_1}+\delta_2P_{j_2}+\delta_kP_{j_k}=0$$

对任意 λ,也有

$$\lambda(\delta_1 P_{j_1} + \delta_2 P_{j_2} + \delta_k P_{j_k}) = 0 \qquad (\text{i})$$

又因为 $x_{j_1}, x_{j_2}, \cdots, x_{j_k}$ 为可行解 x 的非零分量，所以

$$x_{j_1} P_{j_1} + x_{j_2} P_{j_2} + \cdots + x_{j_k} p_{j_k} = b \qquad (\text{ii})$$

（ii）式 + （i）式，得到

$$(x_{j_1} + \lambda\delta_1) P_{j_1} + \cdots + (x_{j_k} + \lambda\delta_k) P_{j_k} = b$$

（ii）式 − （i）式，得到

$$(x_{j_1} - \lambda\delta_1) P_{j_1} + \cdots + (x_{j_k} - \lambda\delta_k) P_{j_k} = b$$

选取 $\lambda = \min\limits_{1 \leqslant t \leqslant k} \left\{ \dfrac{x_{j_t}}{\delta_t} \mid \delta_t \neq 0 \right\}$，构造

$$\begin{aligned} x^{(1)}: x_{j_1}^{(1)} &= x_{j_1} + \lambda\delta_1 \\ x_{j_2}^{(1)} &= x_{j_2} + \lambda\delta_2 \\ &\cdots \\ x_{j_k}^{(1)} &= x_{j_k} + \lambda\delta_k \\ x_j^{(1)} &= 0, j \neq j_1, j_2, \cdots, j_k \\ x^{(2)}: x_{j_1}^{(2)} &= x_{j_1} - \lambda\delta_1 \\ x_{j_2}^{(2)} &= x_{j_2} - \lambda\delta_2 \\ &\cdots \\ x_{j_k}^{(2)} &= x_{j_k} - \lambda\delta_k \\ x_j^{(2)} &= 0, j \neq j_1, j_2, \cdots, j_k \end{aligned}$$

因此 $x^{(1)} \geqslant 0, x^{(2)} \geqslant 0, x^{(1)} \neq x^{(2)}$。而且，$\sum\limits_{j=1}^{n} x_j^{(1)} P_j = b$，$\sum\limits_{j=1}^{n} x_j^{(2)} P_j = b$。故 $x^{(1)}$，$x^{(2)}$ 为可行解，并且，$x = \dfrac{x^{(1)}}{2} + \dfrac{x^{(2)}}{2}$。这样，$x$ 不是可行解集的极点，与假设矛盾。

下面证明充分性。设 x 是基础可行解。$x = 0$ 时，显然 x 必为极点。故设 $x \neq 0$，其非零分量对应的列向量 $P_{j_1}, P_{j_2}, \cdots, P_{j_k}$ 线性无关。

如果 x 不是极点。则必存在凸集 S 中两个不同的点 $x^{(1)}, x^{(2)}$ 使得

$$x = \alpha x^{(1)} + (1 - \alpha) x^{(2)} \quad \alpha \in (0,1)$$

即有

$$x_j = \alpha x_j^{(1)} + (1 - \alpha) x_j^{(2)}, j = 1, 2, \cdots, n$$

当 $j \neq j_1, j_2, \cdots, j_k$ 时上式变为：

$$0 = \alpha x_j^{(1)} + (1 - \alpha) x_j^{(2)}$$

只有

$$x_j^{(1)}=x_j^{(2)}=0,j\neq j_1,j_2,\cdots,j_k$$

又因为 $Ax^{(1)}=b$，$Ax^{(2)}=b$。两式相减，得

$$(x_{j_1}^{(1)}-x_{j_1}^{(2)})P_{j_1}+\cdots+(x_{j_k}^{(1)}-x_{j_k}^{(2)})P_{j_k}=0$$

而且，其系数不全为零。所以，$P_{j_1},P_{j_2},\cdots,P_{j_k}$ 线性相关，与假设矛盾。

定理 3 （LP）问题的最优值可以在极点上达到。

证明 设 $x^{(0)}$ 为最优解，最优值为 $z(x^{(0)})=cx^{(0)}$。若 $x^{(0)}=0$，根据定理 2，$x^{(0)}$ 已是极点，定理成立。若 $x^{(0)}\neq 0$，且不是极点，其非零向量为 $x_{j_1},x_{j_2},\cdots,x_{j_k}$（$k\leqslant m$），则利用证明定理 2 必要性时类似的方法选择 λ，从 $x^{(0)}$ 出发构造：

$$x^{(1)}=x^{(0)}+\lambda\delta,x^{(2)}=x^{(0)}-\lambda\delta,\delta=(\delta_1,\cdots,\delta_k)^T$$

并且

$$x_j^{(i)}=0,j\neq j_1,\cdots,j_k,i=1,2$$

由这个构造方法，显然 $x^{(1)}$、$x^{(2)}$ 中，至少有一个其非零分量的个数要比 $x^{(0)}$ 的非零分量的个数少一个。而且

$$z(x^{(1)})=cx^{(0)}+\lambda c\delta,z(x^{(2)})=cx^{(0)}-\lambda c\delta$$

前面两式中，必有 $c\delta=0$，不然，$z(x^{(0)})$ 就不是最小值。

故，$z(x^{(0)})=z(x^{(1)})=z(x^{(2)})$。

因此，已知最优解是 $x^{(0)}$，如果 $x^{(0)}$ 不是基础可行解，总可以构造另一个最优解 $x^{(1)}$，其非零分量个数至少比 $x^{(0)}$ 的非零分量个数要少一个。重复这个步骤，可得到最优解：$x^{(2)},x^{(3)},\cdots$，它们的非零分量个数越来越少。到某个时候，一定可以找到最优解 $x^{(e)}$，或 $x^{(e)}=0$，或 $x^{(e)}\neq 0$，但其非零分量所对应的列向量线性无关（因为，到只有一个非零分量时，对应的只有一个列向量，必然线性无关）。这时，$x^{(e)}$ 即为极点，并且 $z(x^{(e)})=z(x^{(0)})$。

因为极点对应矩阵 A 中一组线性无关列向量，而矩阵 A 中共有 n 列，n 个向量中线性无关组的个数是有限的，因而极点个数有限。从这一结论和定理 3，我们可以得出，如果（LP）问题有最优解，就只需从有限的几个极点中去找。后面所讲到的解（LP）问题的单纯形方法，就是根据这个原理：由一个极点，迭代到另一个极点，经有限次迭代，求得（LP）问题最优解，或判断问题无解。

1.3 单纯形方法

设有（LP）问题：

$$\min f = cx$$

$$s.t.\begin{cases} Ax = b \text{ 或 } \sum_{j=1}^{n} x_j P_j = b \\ x \geqslant 0 \end{cases}$$

这里，$c = (c_1, \cdots, c_n)$，$x = (x_1, \cdots, x_n)^T$，$b = (b_1, \cdots, b_m)^T$

$A = (a_{ij})_{m \times n} = (P_1, \cdots, P_n)$，

$P_j = (a_{1j}, \cdots, a_{mj})^T \quad \text{rank} A = m, m < n$。

1.3.1 有初始基 $B_{m \times m} = (P_1, \cdots, P_m)$

改写 $f = cx$ 为 $cx - f = 0$，及 $c_B = (c_1, \cdots, c_m)$，$c_N = (c_{m+1}, \cdots, c_n)$，$x_B = (x_1, \cdots, x_m)^T$，$x_N = (x_{m+1}, \cdots, x_n)^T$，（称 x_B 为基变量，x_N 为非基变量）后，有：

$$c = (c_B, c_N), x = (x_B, x_N)^T, A = (B, N)$$

原（LP）问题与矩阵

	x_1 x_2 $\cdots$ x_m $\cdots$ x_n	右端项
x_1 x_2 $\cdots$ x_m	A	b
$-f$	c	0

即

$$\begin{bmatrix} B & N & b \\ c_B & c_N & 0 \end{bmatrix}$$

相对应。对第一行以 B^{-1} 左乘，得：

$$\begin{bmatrix} I & B^{-1}N & B^{-1}b \\ c_B & c_N & 0 \end{bmatrix}$$

将 c_B 变为零，得到：

$$\begin{bmatrix} I & B^{-1}N & B^{-1}b \\ 0 & c_N - c_B B^{-1} N & -c_B B^{-1} b \end{bmatrix} = \begin{bmatrix} B^{-1}A & B^{-1}b \\ c - c_B B^{-1} A & -c_B B^{-1} b \end{bmatrix} = T(B)$$

——（LP）问题对应于基 B 的单纯形表。

它等价于

$$\begin{cases} x_B + B^{-1}Nx_N = B^{-1}b \\ -f + (c_N - c_BB^{-1}N)x_N = -c_BB^{-1}b \end{cases}$$

得到消去系统：

$$x_B = B^{-1}b - B^{-1}Nx_N$$

$$-f = -c_BB^{-1}b - (c_N - c_BB^{-1}N)x_N$$

易见有：

(1) 令 $x_N = 0$,得 $x_B = B^{-1}b$,即 $x^* = (B^{-1}b, 0)^T$ 满足

$$Ax = b$$

则称此 x^* 为 (LP) 问题关于基 B 的基础解。

若 $B^{-1}b \geqslant 0$,则 $x^* \in S$ (可行域)。称此 x^* 为 (LP) 问题关于基 B 的基础可行解 (显然,x^* 非零分量所对应 A 之列向量线性无关)。而且,$f(x^*) = c_BB^{-1}b$。

(2) 当 $c - c_BB^{-1}A = (c_B, c_N) - c_BB^{-1}(B, N) = (c_B, c_N) - (c_BB^{-1}B, c_BB^{-1}N)$

$= (0, c_N - c_BB^{-1}N) \geqslant 0$ 时,有

$$f(x) = c_BB^{-1}b + (c_N - c_BB^{-1}N)x_N \geqslant c_BB^{-1}b = f(x^*) \quad (\forall x \in S)$$

——x^* 为 (LP) 问题的最优解。称它为基础最优解,B 为最优基。

$c - c_BB^{-1}A$ 给出的一行数叫作检验数。

综合上述结果,得：

判定定理　(LP) 问题对于基 B,若 $B^{-1}b \geqslant 0$,且 $c - c_BB^{-1}A \geqslant 0$,则对应于基 B 的基础解便是最优解。

例 5

$$\min s = x_1 + 2x_2 + 3x_3$$

$$s.t. \begin{cases} 2x_1 - x_2 \quad = 1 \\ x_1 \quad + x_3 = 1 \\ x_i \geqslant 0, i = 1,2,3 \end{cases}$$

给出关于基 $B_1 = (P_1, P_2)$,$B_2 = (P_1, P_3)$,$B_3 = (P_2, P_3)$ 的单纯形表,并分别叙述其基础解的性质 (可行?最优?)

解　原问题有初始表：

$$\begin{bmatrix} 2 & -1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 2 & 3 & 0 \end{bmatrix}$$

基 $B_1 = (P_1, P_2)$ 所对应的单纯形表为：

$$\begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & -2 & -3 \end{bmatrix}$$

看出,此时有可行解 $x^0=(1,1,0)^T$,$z(x^0)=3$,但它不是最优解。

基 $B_2=(P_1,P_3)$ 所对应的单纯形表为:

$$\begin{bmatrix} 1 & -\frac{1}{2} & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & 1 & \frac{1}{2} \\ 0 & 2 & 0 & -2 \end{bmatrix}$$

看出,此时有可行解 $x^0=(\frac{1}{2},0,\frac{1}{2})^T$,$z(x^0)=2$,它也是最优解。

基 $B_3=(P_2,P_3)$ 所对应的单纯形表为:

$$\begin{bmatrix} -2 & 1 & 0 & -1 \\ 1 & 0 & 1 & 1 \\ 2 & 0 & 0 & -1 \end{bmatrix}$$

看出,方程组有解 $x^0=(0,-1,1)^T$,$z(x^0)=1$,但它不是可行解。

1.3.2 基的迭代

若关于初始可行基 B 的某些检验数是负数,不能判定 B 是否为最优基,需要换基迭代。

一般,设

$$B=(P_{j_1},\cdots,P_{j_r},\cdots,P_{j_m})$$
$$B^{-1}A=(b_{ij})_{m\times n}$$
$$B^{-1}b=(b_{1,n+1},\cdots,b_{m,n+1})^T$$
$$c-c_BB^{-1}A=(b_{m+1,1},b_{m+1,2},\cdots,b_{m+1,n})$$
$$-c_BB^{-1}b=b_{m+1,n+1}$$

此处,$c-c_BB^{-1}A\leqslant 0$,比如

$$|b_{m+1,s}|=\max_j\{|b_{m+1,j}|\,|\,b_{m+1,j}<0\}$$

便选定 $P'_s=(b_{1s},b_{2s},\cdots,b_{rs},\cdots,b_{ms})^T$ 为进基向量。

(1) 若 $b_{is}\leqslant 0$,$i=1,2,\cdots,m$,则 (LP) 问题无最优解。

证明　取基础解

$$x:\begin{cases} x_s=\lambda \\ x_{j_1}=b_{1,n+1}-b_{1s}\lambda \\ x_{j_2}=b_{2,n+1}-b_{2s}\lambda \\ \quad\cdots \\ x_{j_m}=b_{m,n+1}-b_{ms}\lambda \\ x_j=0,j\neq j_1,j_2,\cdots,j_m,s \end{cases}$$

其中 λ 为任意大的正数，则 x 都是可行解，而对应的目标函数值为

$$f(x)=-b_{m+1,n+1}+b_{m+1,s}\lambda$$

因 $b_{m+1,s}<0$，所以当 $\lambda\to\infty$ 时，$f(x)\to-\infty$，即问题无最优解（目标函数下方无界，当然无最小值）。

(2) 若 b_{is} 中有正分量，求出

$$\min_{b_{is}>0,1\leqslant i\leqslant m}\left\{\frac{b_{i,n+1}}{b_{is}}\right\}\overset{(设)}{=}\frac{b_{r,n+1}}{b_{rs}}\qquad (P_{j_r}\text{ 退基})$$

（最小者不止一个，选 i 最小者）称 b_{rs} 为轴心项。于是形成新基

$$\hat{B}=(P_{j_1},\cdots,P_{j_{r-1}},P_s,P_{j_{r+1}},\cdots,P_{j_m})$$

矩阵呈

$$B^{-1}A=\begin{array}{c}\begin{matrix} & & (s\text{ 列}) & & (j\text{ 列}) & \end{matrix}\\ \left[\begin{matrix} & & b_{1s} & & & \\ & & \cdots & & & \\ & & b_{ij} & & b_{is} & \\ & & \cdots & & & \\ b_{r1} & \cdots & b_{rs} & \cdots & \cdots & b_{r,n+1} \\ & & \cdots & & & \\ & & b_{ms} & & & \\ & & b_{m+1,s} & & & \end{matrix}\right]\end{array}$$

状。对矩阵进行初等变换：

① 用 b_{rs} 去除第 r 行各数，得新元

$$\hat{b}_{rj}=\begin{cases}1 & j=s\\ \dfrac{b_{rj}}{b_{rs}}, & 1\leqslant j\neq s\leqslant n+1\end{cases}$$

② 将 P'_s 列其余分量变为零，$B^{-1}A$ 中相应新元

$$\hat{b}_{ij}=b_{ij}-b_{is}\left(\frac{b_{rj}}{b_{rs}}\right)\qquad (1\leqslant i\neq r\leqslant m+1,1\leqslant j\leqslant n+1)$$

有结论：Ⅰ.$\hat{B}$ 仍然是一个可行基。

Ⅱ.B 换为 $\hat{B}$ 后，目标函数值不增加。

证明

Ⅰ.B 为可行基，$B^{-1}b\geqslant 0$. 由前述 ②，应有：

$$\hat{b}_{i,n+1} = b_{i,n+1} - b_{is}\left(\frac{b_{r,n+1}}{b_{rs}}\right)$$

当 $b_{is} \leqslant 0$ 时,显然有 $\hat{b}_{i,n+1} \geqslant 0$;

当 $b_{is} > 0$ 时,$\hat{b}_{i,n-1} = b_{is}\left(\frac{b_{i,n+1}}{b_{is}} - \frac{b_{r,n+1}}{b_{rs}}\right)$,

据 r 行之产生,$\hat{b}_{i,n+1} > 0\ (i = 1,\cdots,m)$。

Ⅱ $\hat{b}_{m+1,n+1} - b_{m+1,n+1} = \left[b_{m+1,n+1} - b_{r,n+1}\left(\frac{b_{m+1,s}}{b_{rs}}\right)\right] - b_{m+1,n+1}$

$$= - b_{r,n+1}\left(\frac{b_{m+1,s}}{b_{rs}}\right)$$

因为 $b_{m+1,s} < 0$,$b_{rs} > 0$,故当 $b_{r,n+1} > 0$ 时,

$\hat{b}_{m+1,n+1} - b_{m+1,n+1} < 0$,即 $\hat{b}_{m+1,n+1} > b_{m+1,n+1}$,$-f(\hat{x}) > -f(x)$,$f(\hat{x}) < f(x)$;

当 $b_{r,n+1} = 0$ 时,$\hat{b}_{m+1,n+1} - b_{m+1,n+1} = 0$,即 $\hat{b}_{m+1,n+1} = b_{m+1,n+1}$,$f(\hat{x}) = f(x)$。

这样,我们就完成了基的迭代工作,经有限次迭代,必能求得最优解(若存在)。

1.3.3 单纯形法的求解程序

设有(LP)问题: $\min f = cx$

$$s.t.\begin{cases} Ax = b \text{ 或} \sum_{j=1}^{n} x_j P_j = b \\ x \geqslant 0 \end{cases} \tag{1}$$

又,B 为其可行基,只要算出 B^{-1},便可构造出

$$T(B) = \begin{bmatrix} B^{-1}A & B^{-1}b \\ c - c_B B^{-1}A & -c_B B^{-1}b \end{bmatrix}$$

$$= \begin{bmatrix} I & B^{-1}N & B^{-1}b \\ 0 & c_N - c_B B^{-1}N & -c_B B^{-1}b \end{bmatrix}$$

$$= \begin{bmatrix} & & & b_{1,n+1} \\ I & (b_{ij})_{m\times(n-m)} & & \cdots \\ & & & b_{m,n+1} \\ 0 & b_{m+1,m+1}\ b_{m+1,m+2} & \cdots & b_{m+1,n+1} \end{bmatrix}$$

这里

$b_{m+1,j} = c_j - c_B B^{-1} P_j$ —— 检验数; $(j = 1,2,\cdots,n)$

$B^{-1}P_j = (b_{1j},\cdots,b_{mj})^T$; $(j = m+1,\cdots,n)$

$$B^{-1}b = (b_{1,n+1},\cdots,b_{m,n+1})^T$$

现若 B **非最优基，应当换基** $B \longrightarrow \hat{B}$**。又当用** $\hat{B}^{-1}$ **去计算新的** $T(\hat{B})$**，再判定** $\hat{B}$ **是否最优。下面推导出用** B^{-1} **表示** $\hat{B}^{-1}$ **的公式，以形成固定的计算程序。**

设 $B = (P_{i_l},\cdots,P_{i_r},\cdots,P_{i_m})$**。经** P_{i_r} **退基，**P_s **进基，得：**

$$\hat{B} = (P_{i_1},\cdots,P_{i_{r-1}},P_s,P_{i_{r+1}},\cdots,P_{i_m})$$

有

$$B^{-1}B = (B^{-1}P_{i_1},\cdots,B^{-1}P_{i_r},\cdots,B^{-1}P_{i_m}) = I$$

$$B^{-1}\hat{B} = (B^{-1}P_{i_1},\cdots,B^{-1}P_{i_{r-1}},B^{-1}P_s,B^{-1}P_{i_{r-1}},\cdots,B^{-1}P_{i_m})$$

$$\begin{array}{c} \text{第}(r)\text{列} \\ = \begin{bmatrix} 1 & & b_{1s} & 1 & \\ & 1 & b_{2s} & & \\ & & \cdots & & \\ & & b_{rs} & & \\ & & \cdots & & \\ & & b_{m-1,s} & 1 & \\ & & b_{ms} & & 1 \end{bmatrix} \end{array}$$

$$= B_{rs}$$

由此，$\hat{B}^{-1}B = (B^{-1}\hat{B})^{-1} = B_{rs}{}^{-1}$

$$\begin{array}{c} \text{第}(r)\text{列} \\ = \begin{bmatrix} 1 & & \dfrac{-b_{1s}}{b_{rs}} & & \\ & 1 & -\dfrac{b_{2s}}{b_{rs}} & & \\ & & \cdots & & \\ & & \dfrac{1}{b_{rs}} & & \\ & & \cdots & & \\ & & -\dfrac{b_{m-1,s}}{b_{rs}} & 1 & \\ & & -\dfrac{b_{ms}}{b_{rs}} & & 1 \end{bmatrix} \end{array}$$

$$= E_{rs}$$

故

$$\hat{B}^{-1} = E_{rs}B^{-1}$$

$$x_{\hat{B}} = \hat{B}^{-1}b = (E_{rs}B^{-1})b = E_{rs}(B^{-1}b) = E_{rs}x_B$$

综合得计算框图（见图 1.4）：

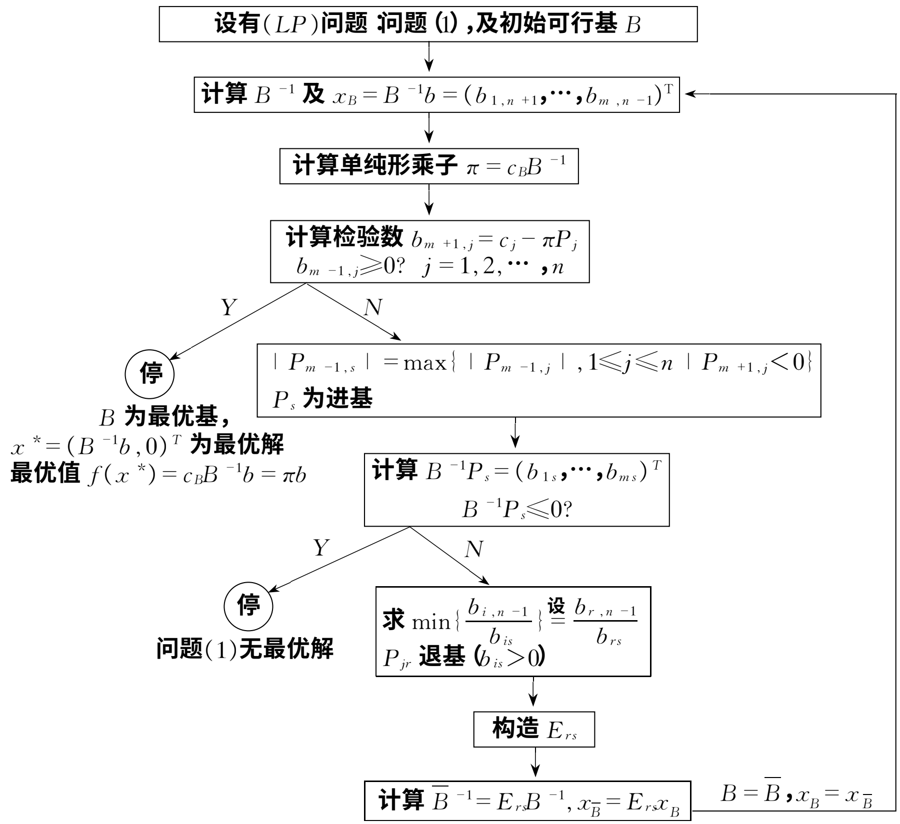


图 1.4

1.3.4 大 M 法 —— 求第一个可行基的一个方法

对 (LP) 问题

$$\min f = \sum_{j=1}^{n} c_j x_j$$

$$s.t.\begin{cases} \sum_{j=1}^{n} a_{ij}x_j = b_i, i = 1,2,\cdots,m\ (m \leqslant n) \\ x_j \geqslant 0, j = 1,2,\cdots,n \end{cases}$$

我们引入人工变量 $A_1, A_2, \cdots, A_m$ 以构造一个辅助规划问题 $(LP)'$：

$$\min z = \sum_{j=1}^{n} c_j x_j + \sum_{i=1}^{m} MA_i$$

$$s.t. \begin{cases} \sum_{j=1}^{n} a_{ij} x_j + A_i = b_i & i = 1,2,\cdots,m \ (m \leqslant n) \\ x_j \geqslant 0, j = 1,2,\cdots,n \end{cases}$$

此问题有一个初始基 $I = (P_{n+1},\cdots,P_{n+m})$ —— 以人工变量对应的系数列向量作为基向量所生成 而在新的目标函数中，我们赋以人工变量一个正系数 M，$M \gg 0$。很明显，只要基变量中还包含人工变量，目标函数就不可能实现最小化。取 M 为一个很大的正数，就是为了在迭代过程中尽快地把人工变量从基变量中替换出来成为非基变量，从而实现目标函数极小化要求。这个方法的核心就在于很大的正数 M，因此称为大 M 法（也叫惩罚法）。下面，我们给出一个具体例子。

例 6　用大 M 法解 (LP) 问题。

$$\min z = -5x_1 - 3x_2 - 2x_3 - 4x_4$$

$$s.t. \begin{cases} 5x_1 + x_2 + x_3 + 8x_4 = 10 \\ 2x_1 + 4x_2 + 3x_3 + 2x_4 = 10 \\ x_j \geqslant 0, j = 1,2,3,4 \end{cases}$$

解　构造 $(LP)'$：

$$\min z = -5x_1 - 3x_2 - 2x_3 - 4x_4 + MA_1 + MA_2$$

$$s.t. \begin{cases} 5x_1 + x_2 + x_3 + 8x_4 + A_1 = 10 \\ 2x_1 + 4x_2 + 3x_3 + 2x_4 + A_2 = 10 \\ x_j \geqslant 0, A_k \geqslant 0, j = 1,2,3,4; k = 1,2 \end{cases}$$

用单纯形法解之：

表 1.3　初始表

		x_1	x_2	x_3	x_4	A_1	A_2		$\frac{B(i)}{A(i,j)}$
基	$c(j)$	-5	-3	-2	-4	M	M	$B(i)$	
A_1	M	5	1	1	8	1		10	0
A_2	M	2	4	3	2		1	10	0
$c(j)-z(j)$		-5	-3	-2	-4	0	0	0	
* 大 M		0	0	0	0	1	1	0	

表 1.4　　　　　　　　　　　　**迭代** 1

A_1	M	5	1	1	$*8$	1		10	1.25
A_2	M	2	4	3	2		1	10	5
$c(j)-z(j)$		-5	-3	-2	-4	0	0	0	
$*$ **大** M		-7	-5	-4	-10	0	0	-20	
上两行即		$-5-7M$	$-3-5M$	$-2-4M$	$-4-10M$			$0-20M$	
				(x_4 **进基**	A_1 **退基**)				

表 1.5　　　　　　　　　　　　**迭代** 2

x_4	-4	$\frac{5}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	1	$\frac{1}{8}$	0	$\frac{5}{4}$	10
A_2	M	$3/4$	$*\frac{15}{4}$	$\frac{11}{4}$	0	$-\frac{1}{4}$	1	$\frac{15}{2}$	2
$c(j)-z(j)$		$-\frac{5}{2}$	$-\frac{5}{2}$	$-\frac{3}{2}$	0	$\frac{1}{2}$	0	5	
$*$ **大** M		$-\frac{3}{4}$	$-\frac{15}{4}$	$-\frac{11}{4}$	0	$5/4$	0	$-\frac{15}{2}$	
				(x_2 **进基**	A_2 **退基**)				

表 1.6　　　　　　　　　　　　**迭代** 3

x_4	-4	$*\frac{3}{5}$	0	$\frac{1}{30}$	1	$\frac{2}{15}$	$-\frac{1}{30}$	1	$\frac{5}{3}$
x_2	-3	$\frac{1}{5}$	1	$\frac{11}{15}$	0	$-\frac{1}{15}$	$\frac{4}{15}$	2	10
$c(j)-z(j)$		-2	0	$\frac{1}{3}$	0	$\frac{1}{3}$	$\frac{2}{3}$	10	
$*$ **大** M		0	0	0	0	1	1	0	
				(x_1 **进基**	x_4 **退基**)				

表 1.7　　　　　　　　　　　　**最后表**

x_1	-5	1	0	$\frac{1}{18}$	$\frac{5}{3}$	$\frac{2}{9}$	$-\frac{1}{8}$	$\frac{5}{3}$	0
x_2	-3	0	1	$\frac{13}{18}$	$-\frac{1}{3}$	$-\frac{1}{9}$	$\frac{5}{18}$	$\frac{5}{3}$	0
$c(j)-z(j)$		0	0	$\frac{4}{9}$	$\frac{10}{3}$	$\frac{7}{9}$	$\frac{5}{9}$	$\frac{40}{3}$	
$*$ **大** M		0	0	0	0	1	1	0	

最优解：$x_1=x_2=\frac{5}{3}$，$x_3=x_4=0$；**目标函数最小值** $=-\frac{40}{3}$。

1.4 对偶线性规划问题

1.4.1 对偶线性规划问题

设有线性规划问题［Ⅰ］

$$\min s = cx$$
$$s.t.\begin{cases} Ax \geqslant b \\ x \geqslant 0 \end{cases}$$

这里，$c = (c_1, \cdots, c_n)$，$x = (x_1, \cdots, x_n)^T$，$A = (a_{ij})_{m \times n}$，$b = (b_1, \cdots, b_m)^T$

与线性规划问题［Ⅱ］

$$\max g = yb$$
$$s.t.\begin{cases} yA \leqslant c \\ y \geqslant 0 \end{cases}$$

这里，$y = (y_1, \cdots, y_m)$。问题［Ⅰ］和问题［Ⅱ］称为“互为对偶的线性规划问题”。

我们可以用下面的对偶表（表1.8），来表达它们的对偶关系：

表1.8

x_1	x_2	$\cdots$	x_n			
c_1	c_2	$\cdots$	c_n		min/max	
	$\geqslant$					
a_{11}	a_{12}	$\cdots$	a_{1n}		b_1	y_1
a_{21}	a_{22}	$\cdots$	a_{2n}		b_2	y_2
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$(\geqslant)$	$\cdots$	$\cdots$
a_{m1}	a_{m2}	$\cdots$	a_{mn}		b_m	y_m

1.4.2 对偶问题的几个基本性质

定理1 如果线性规划问题［Ⅰ］中第 k 个约束条件是等式，那么，它的对偶规划问题［Ⅱ］中第 k 个变量 y_k 无非负限制，反之亦然。

证明 设线性规划问题［Ⅰ］中第 k 个约束条件取等式：

$$a_{k1}x_1 + \cdots + a_{kn}x_n = b_k$$

那么，可将其变为与之等价的两个不等式：

$$a_{k1}x_1 + \cdots + a_{kn}x_n \geqslant b_k$$

$$-\left(a_{k1}x_1+\cdots+a_{kn}x_n\right)\geqslant -b_k$$

于是,线性规划问题[Ⅰ]等价于:

$$\min s=c_1x_1+\cdots+c_nx_n$$

$$s.t.\begin{cases}a_{11}x_1+\cdots+a_{1n}x_n\geqslant b_1\\ \cdots\quad\cdots\quad\cdots\\ a_{k1}x_1+\cdots+a_{kn}x_n\geqslant b_k,\\ -a_{k1}x_1-\cdots-a_{kn}x_n\geqslant -b_k,\\ \cdots\quad\cdots\quad\cdots\\ a_{m1}x_1+\cdots+a_{mn}x_n\geqslant b_m\\ x_j\geqslant 0,j=1,2,\cdots,n\end{cases}$$

其对偶线性规划问题[Ⅱ]应为:

$$\max g=b_1y_1+\cdots+b_k\left(y'_k-y''_k\right)+\cdots+b_my_m$$

$$s.t.\begin{cases}a_{11}y_1+\cdots+a_{k1}\left(y'_k-y''_k\right)+\cdots+a_{m1}y_m\leqslant c_1\\ \quad\cdots\quad\quad\cdots\\ a_{1n}y_1+\cdots+a_{kn}\left(y'_k-y''_k\right)+\cdots+a_{mn}y_m\leqslant c_n\\ y_j\geqslant 0,j=1,\cdots,k-1,k+1,\cdots,n;y'_k\geqslant 0,y''_k\geqslant 0.\end{cases}$$

如果令 $y_k=y'_k-y''_k$,代入上面即得:

$$\max g=\sum_{j=1}^{m}b_jy_j$$

$$s.t.\begin{cases}\sum\limits_{i=1}^{m}a_{ij}y_i\geqslant c_j,j=1,2,\cdots,n;\\ y_j\geqslant 0,j=1,\cdots,k-1,k+1,\cdots,n;y_k\text{ 无非负限制}\end{cases}$$

按此逆推之,第二个结论显然。

例 7 求

$$\min s=x_1-2x_2+3x_3$$

$$s.t.\begin{cases}3x_1+2x_2-x_3\leqslant 1\\ \qquad 2x_2+3x_3\leqslant 3\\ x_1\qquad\quad+x_3\leqslant 5\\ x_j\geqslant 0,j=1,2,3\end{cases}$$

的对偶问题。其对偶问题见表 1.9:

表 1.9

x_1	x_2	x_3			
1	-2	3		min/max	
	$\geqslant$				
-3	-2	1		-1	y_1
	-2	-3	$\geqslant$	-3	y_2
-1		-1		-5	y_3

这样的对偶问题称为对称型的对偶问题。

例 8 **求**

$$\min s = cx$$
$$s.t.\begin{cases}Ax = b \\ x \geqslant 0\end{cases}$$

的对偶问题。

由定理 1**，其对偶问题为：**

$$\max g = yb$$
$$s.t.\begin{cases}yA \leqslant c \\ y\ \text{无非负限制}\end{cases}$$

这样的对偶问题称为非对称型的对偶问题。

例 9 **求**

$$\min s = 3x_1 - 2x_2 + x_3$$
$$s.t.\begin{cases}x_1 + 2x_2 = 1 \\ 2x_2 + x_3 \leqslant -2 \\ 2x_1 + x_3 \geqslant 3 \\ x_1 - 2x_2 + 3x_3 \geqslant 4 \\ x_1, x_2 \geqslant 0, x_3\ \text{无非负限制}\end{cases}$$

解　见表 1.10**：**

表 1.10

x_1	x_2	x_3			
3	-2	1		min/max	
$\geqslant$	$\geqslant$	$\lvert\lvert$			
1	2		$=$	1	y_1 **(无非负限制)**
	-2	1	$\geqslant$	2	y_2
2		1	$\geqslant$	3	y_3
1	-2	3	$\geqslant$	4	y_4

这样的对偶问题称为混合型的对偶问题。

定理2 如果 x,y 分别是互为对偶问题 [Ⅰ] 与 [Ⅱ] 的可行解,那么它们对应的目标函数值间的关系是:$s \geqslant g$。

证明 $s = cx \geqslant (yA)x = y(Ax) \geqslant yb = g$

定理3 若 x', y' 分别为互为对偶线性规划问题 [Ⅰ] 和 [Ⅱ] 的可行解,并且,$cx' = y'b$,那么 x', y' 分别为问题 [Ⅰ] 和问题 [Ⅱ] 的最优解。

证明 由定理2,对于问题 [Ⅰ] 的任何可行解 x,有

$$cx \geqslant y'b = cx'$$

表明 x' 为问题 [Ⅰ] 的最优解。同理可证 y' 为问题 [Ⅱ] 的最优解。

定理4 互为对偶线性规划问题 [Ⅰ] 和 [Ⅱ] 中,任意一个有可行解但无最优解,则另一个就无可行解。

证明 用反证法。设问题 [Ⅰ] 的可行解为 $x^{(0)}$,但 $cx^{(0)}$ 无下界。如果问题 [Ⅱ] 存在可行解 $y^{(0)}$,那么由定理2必有 $cx^{(0)} \geqslant y^{(0)}b$,此与 $cx^{(0)}$ 无下界矛盾。同理可证,若问题 [Ⅱ] 有可行解但无最优解,则问题 [Ⅰ] 无可行解。

定理5 互为对偶线性规划问题 [Ⅰ] 和 [Ⅱ] 都有可行解,则都有最优解。

证明 用反证法。设问题 [Ⅰ] 和问题 [Ⅱ] 都有可行解。如果问题 [Ⅰ] (或问题 [Ⅱ]) 没有最优解,则由定理4,当问题 [Ⅱ] (或问题 [Ⅰ]) 无可行解,矛盾。

定理6 互为对偶线性规划问题 [Ⅰ] 和 [Ⅱ] 中,任意一个有最优解,则另一个也有最优解,并且它们的最优值相等。

证明 引入 $(c,0) = (c_1,\cdots,c_n,0,\cdots,0)$;$x' = (x_1,\cdots,x_n,x_{n+1},\cdots,x_{n+m})^T$,化问题 [Ⅰ] 为标准型:

$$\text{问题 [Ⅰ}'\text{]} \quad \min s = (c,0)x'$$

$$s.t.\begin{cases}(A,-I)x' = b\\ x' \geqslant 0\end{cases}$$

设问题 [Ⅰ] 有最优解 $x^* = (x_1^*,\cdots,x_n^*)^T$,显然,[Ⅰ′] 也有最优解。设其最优解为 $x^{*'} = (x_1^*,\cdots,x_n^*,x_{n-1}^*,\cdots,x_{n-m}^*)^T$,对应的最优基为 B。

根据最优判别定理得:$(c,0) - c_B B^{-1}(A,-I) \geqslant 0$。

令 $c_B B^{-1} = y^*$,则 $(c,0) - y^*(A,-I) \geqslant 0$,即有:

$$(c - y^*A, y^*) \geqslant 0$$

也即

$$y^*A \leqslant c$$

$$y^* \geqslant 0$$

故 y^* 是问题 [Ⅱ] 的可行解。又因为 $(c,0)x^{*'} = cx^* = c_B B^{-1}b = y^*b$,根据定理3,$y^*$ 为问题 [Ⅱ] 的最优解,并且与问题 [Ⅰ] 有相等的最优值

$$cx^* = y^* b = c_B B^{-1} b$$

注 从定理6的证明中可以得到 :问题[Ⅰ]的对偶问题[Ⅱ]的最优解 y^* 就是问题[Ⅰ′]最优基 B 对应的单纯形表中松弛变量所对应的检验数。这是因为：

$$y^* = c_B B^{-1} = [0 - c_B B^{-1}(-I)]$$

1.4.3 对偶问题的经济意义——影子价格

我们通过例子来说明对偶问题的经济意义——影子价格。

例10 假如某作物,全部生产过程中至少需要氮肥32公斤,磷肥24公斤,甲肥42公斤。已知甲、乙、丙、丁四种复合肥料每公斤的价格及含氮、磷、钾的数量,如表1.11：

表1.11

肥料 / 所含成分数量(公斤) / 成分	甲	乙	丙	丁	肥料需要量(公斤)
氮	0.03	0.3	0	0.15	32
磷	0.05	0	0.2	0.1	24
钾	0.14	0	0	0.07	42
每公斤价格(元)	0.04	0.15	0.1	0.13	

问应如何配合使用这些肥料,既能满足作物对氮磷钾的需要,又能使施肥成本最低？

解 用 x_1、x_2、x_3、x_4 分别表示甲、乙、丙、丁四种肥料的用量,那么问题成为：

$$\min s = 0.04x_1 + 0.15x_2 + 0.1x_3 + 0.13x_4$$

$$s.t.\begin{cases} 0.03x_1 + 0.3x_2 + 0.15x_4 \geqslant 32 \\ 0.05x_1 + 0.2x_3 + 0.1x_4 \geqslant 24 \\ 0.14x_1 + 0.07x_4 \geqslant 42 \\ x_j \geqslant 0, j = 1,2,3,4 \end{cases}$$

把这个问题用对偶表表示如表1.12：

表1.12

x_1	x_2	x_3	x_4			
0.04	0.15	0.1	0.13		min⁄max	
		∨/				
0.03	0.3		0.15		32	y_1
0.05		0.2	0.1	⩾	24	y_2
0.14			0.07		42	y_3

其对偶问题可以解释为：某肥料公司计划生产氮、磷、钾三种单成分的化肥，肥料公司要为这三种化肥确定单价（设为 y_1、y_2、y_3），既要使得获利最大，同时又要与生产甲、乙、丙、丁四种复合肥料的公司竞争。因此，必须设想以单成分化肥氮、磷、钾合成的甲肥每公斤价格不超过 0.04 元，即

$$0.03y_1 + 0.05y_2 + 0.14y_3 \leqslant 0.04$$

同样，合成的乙肥、丙肥、丁肥的每公斤价格分别不能超过 0.15 元，0.1 元，0.13 元。所以又有：

$$0.3y_1 \leqslant 0.15$$
$$0.2y_2 \leqslant 0.1$$
$$0.15y_1 + 0.1y_2 + 0.07y_3 \leqslant 0.13$$

由于至少需要氮肥 32 公斤，磷肥 24 公斤，钾肥 42 公斤，所以，要获利最大，就要使

$$g = 32y_1 + 24y_2 + 42y_3$$

的值最大。

利用单纯形方法，可得原问题的最优基 $B = (P_2, P_1, P_7)$。对应单纯形表如表 1.13：

表 1.13

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	
x_2		1	-0.4	0.3	$-\frac{10}{3}$	-2		$\frac{176}{3}$
x_1	1		4	2		-20		480
x_7			0.56	0.21		-2.8	1	25.2
$-s$				0.005	0.5	0.5		-28

最优解为：$x_1 = 480$，$x_2 = \frac{176}{3}$，$x_7 = 25.2$，其他 $x_j = 0$，对应的目标函数最小值 $s = 28$，即应使用钾肥 480 公斤，乙肥 $\frac{176}{3}$ 公斤，不用丙肥和丁肥；最少施肥成本为 28 元。

由上表还可以看出：对偶问题之最优解为 $y_1 = 0.5$，$y_2 = 0.5$，$y_3 = 0$，对应的目标函数最大值 $g = 28$。一般地，我们称对偶问题的最优解为原问题约束条件的影子价格，简称影子价格。

某一约束条件的影子价格表示为当它的约束条件右端的常数增加一个单位时（假设原问题最优基不变），原问题目标函数值增加的数值。

我们证明这个结论。由定理 6，对偶线性规划问题最优值相等：

$$s^* = cx^* = y^* b$$

例如，b 的第一个分量 b_1 增加一个单位时，原问题最优值变为：

$$\begin{aligned}(y_1^*, \cdots, y_m^*)(b_1 + 1, b_2, \cdots, b_m)^T &= y_1^* + \sum_{i=1}^{m} y_i^* b_i \\ &= y_1^* + y^* b = y_1^* + s^*\end{aligned}$$

即，原问题第一个约束条件的影子价格 y_1^* 是它对应的第一个约束条件右端常数项 b_1 增加一个单位时，原目标函数最优值增加的数值，其他以此类推。

还要注意，影子价格（对偶问题的最优解）是原问题最优基对应的单纯形表中松弛变量对应的检验数。但这检验数又与最优基密切联系着：最优基改变，检验数就会改变，因而影子价格也会改变。前面的结论都是在最优基不变的条件下才成立。

1.5 参数线性规划和灵敏度分析

1.5.1 参数线性规划问题

在线性规划的实际问题中，由于某种原因，有时线性规划问题的目标函数的系数 c 和约束条件的常数项 b 的数据不是固定的常数，而会有所波动。这样的一些线性规划问题，便是所谓"参数线性规划"。对于这种线性规划，我们所关心的是在参数的可能范围内，求出问题的最优解，即可以用原来数学模型按实际出现的目标函数的系数或约束条件右端的常数项来决定最优方案。下面我们就讨论这两种情况的单参数线性规划问题的解法。

1. 目标函数的系数含有参数的线性规划问题

例 11 求解 $\min s = (6-\lambda)x_1 + (5-\lambda)x_2 + (-3+\lambda)x_3 + (-4+\lambda)x_4$

$$s.t.\begin{cases} x_1 - x_2 - x_3 + x_5 = 1 \\ -x_1 + x_2 - x_4 + x_6 = 1 \\ -x_2 + x_3 + x_7 = 1 \\ x_j \geqslant 0,\ j = 1,2,\cdots,7 \end{cases}$$

解 $B_1 = (P_5, P_6, P_7)$ 是一个现成可行基，与此对应的单纯形表如表 1.14：

表 1.14

B_1	x_1	x_2	x_3	x_4	x_5	x_6	x_7	
x_5	1	-1	-1		1			1
x_6	-1	1		-1		1		1
x_7		-1	1				1	1
$-s$	$6-\lambda$	$5-\lambda$	$-3+\lambda$	$-4+\lambda$				0

有可行解

$$x_5 = x_6 = x_7 = 1\text{，其他 } x_j = 0$$

为了使这个解是最优解，要求所有检验数为非负：

$$\begin{array}{rcl} 6-\lambda \geqslant 0 & & \lambda \leqslant 6 \\ 5-\lambda \geqslant 0 & \rightarrow & \lambda \leqslant 5 \\ -3+\lambda \geqslant 0 & & \lambda \geqslant 3 \\ -4+\lambda \geqslant 0 & & \lambda \geqslant 4 \end{array}$$

得 $4 \leqslant \lambda \leqslant 5$。此时，前述可行解为最优解，$\min s = 0$。

现在考察 $\lambda < 4$ 时最优解的变化情况：

当 $\lambda < 4$ 时，x_4 对应的检验数 $-4+\lambda < 0$。但 $P'_4 = (0, -4, 0)^T \leqslant 0$，目标函数无下界，问题无最优解。

再考察 $\lambda > 5$ 时最优解的变化情况：

当 $\lambda > 5$ 时，x_2 对应的检验数 $5-\lambda < 0$。需换基 $B_2 = (P_5, P_2, P_7)$ 与此对应的单纯形表如表 1.15：

表 1.15

B_2	x_1	x_2	x_3	x_4	x_5	x_6	x_7	
x_5			-1	-1	1	1		2
x_2	-1	1		-1		1		1
x_7	-1		1	-1		1	1	2
$-s$	$11-2\lambda$		$-3+\lambda$	1		$-5+\lambda$		$-5+\lambda$

仿照上面作法，知道在 $5 \leqslant \lambda \leqslant \dfrac{11}{2}$ 时，得最优解：

$$x_2 = 1, x_5 = x_7 = 2\text{，其他 } x_j = 0\text{；}\min s = -(-5+\lambda) = 5-\lambda\text{。}$$

最后考察 $\lambda > \dfrac{11}{2}$ 时最优解的变化情况：

当 $\lambda > \frac{11}{2}$ 时，x_1 对应的检验数 $11-2\lambda<0$。但 $P''_1 = (0,-1,-1)^T \leqslant 0$，目标函数无下界，问题无最优解。

2. 约束条件右端的常数项中含有参数的线性规划问题

例 12　求解　$\min s = 2x_1 + 6x_2 + 15x_3$

$$s.t.\begin{cases} -2x_1 - 3x_2 - 5x_3 + x_4 = 6 - \lambda \\ x_1 + x_2 + x_3 + x_5 = -2 + \lambda \\ x_2 + 2x_3 + x_6 = -3 + 2\lambda \\ x_j \geqslant 0, j = 1,\cdots,6 \end{cases}$$

解　$B_1 = (P_4, P_5, P_6)$ 是一个现成可行基，相应的单纯形表如表 1.16：

表 1.16

B_1	x_1	x_2	x_3	x_4	x_5	x_6	
x_4	-2	-3	-5	1			$6-\lambda$
x_5	1	1	1		1		$-2+\lambda$
x_6		1	2			1	$-3+2\lambda$
$-s$	2	6	15				0

可行解

$$x_4 = 6-\lambda, x_5 = -2+\lambda, x_6 = -3+2\lambda$$

为最优解，要求同时有：

$$\begin{aligned} 6-\lambda \geqslant 0 & \qquad & \lambda \leqslant 6 \\ -2+\lambda \geqslant 0 & \quad \rightarrow \quad & \lambda \geqslant 2 \\ -3+2\lambda \geqslant 0 & \qquad & \lambda \geqslant \frac{3}{2} \end{aligned}$$

得 $2 \leqslant \lambda \leqslant 6$。这个时候，上述可行解为最优解，且 $\min s = 0$。

现在考察 $\lambda < 2$ 时最优解的变化情况：

当 $\lambda < 2$ 时，基变量 $x_5 < 0$；但 x_5 对应的行没有负数，问题没有最优解。

再考察 $\lambda > 6$ 时最优解的变化情况：

当 $\lambda > 6$ 时，基变量 $x_4 < 0$；用对偶单纯形方法进行换基迭代，得对应于基 $B_2 = (P_1, P_5, P_6)$ 的单纯形表如表 1.17：

表 1.17

B_2	x_1	x_2	x_3	x_4	x_5	x_6	
x_1	1	$\frac{3}{2}$	$\frac{5}{2}$	$-\frac{1}{2}$			$-3+\frac{\lambda}{2}$
x_5		$-\frac{1}{2}$	$-\frac{3}{2}$	$\frac{1}{2}$	1		$1+\frac{\lambda}{2}$
x_6		1	2			1	$-3+2\lambda$
$-s$		3	10	1			$6-\lambda$

欲使可行解

$$x_1=-3+\frac{\lambda}{2},x_5=1+\frac{\lambda}{2},x_6=-3+2\lambda$$

为最优解，当要求它们同时非负。解得 $\lambda\geqslant 6$，对应的目标函数值 $\min s=-6+\lambda$。

注　与前一种参数线性规划不同，这里，对于 B 的最优区间中每个 λ，不但目标函数的最小值是 λ 的函数，而且最优解也是 λ 的函数。

1.5.2　灵敏度分析

假设我们已经根据所给的数据 c、b 及 A，求得了问题的最优解。现在，由于某种原因，c 与 b 有波动，问波动限制在什么范围内，所得的最优基仍是最优基而无须另作规划，这便是“灵敏度分析”的问题。

1. 目标函数的系数的灵敏度分析

例 13　某工厂用甲、乙两种原料生产 A、B、C、D 四种产品。每种产品的利润、现有原料数，及每种产品消耗原料定额如表 1.18：

表 1.18

原料 \ 每万件产品所用原料数(公斤) \ 产品	A	B	C	D	现有原料数(公斤)
甲	3	2	10	4	18
乙			2	$\frac{1}{2}$	3
每万件产品利润(万元)	9	8	50	19	

问应怎样组织生产才能使利润最多？如果产品 A 的价格有波动，问波动应限制在什么范围内，才能使原最优解不变？

解　设用 x_1,x_2,x_3,x_4 分别表示 A,B,C,D 四种产品的生产数(单位：万件)。

按题意,得到:

$$\max s = 9x_1 + 8x_2 + 50x_3 + 19x_4$$

$$s.t.\begin{cases} 3x_1 + 2x_2 + 10x_3 + 4x_4 \leqslant 18 \\ 2x_3 + \dfrac{1}{2}x_4 \leqslant 3 \\ x_1, x_2, x_3, x_4 \geqslant 0 \end{cases}$$

将此化为标准型,再用单纯形法求解,得到最优基 $B = (P_4, P_3)$,对应的单纯形表如表 1.19:

表 1.19

B	x_1	x_2	x_3	x_4	x_5	x_6	
x_4	2	$\frac{4}{3}$		1	$\frac{2}{3}$	$-\frac{10}{3}$	2
x_3	$-\frac{1}{2}$	$-\frac{1}{3}$	1		$-\frac{1}{6}$	$\frac{4}{3}$	1
$-s'$	-4	$-\frac{2}{3}$	0	0	$-\frac{13}{3}$	$-\frac{10}{3}$	-88

最优方案是:生产 1 万件产品 C,生产 2 万件产品 D,不生产 A,B 两种产品。可得最大利润为 88 万元。

现在,假定目标函数中 $c_1 = 9$ 有波动,令 $c_1 = 9 + \lambda$。

由于

$$B = \begin{bmatrix} 4 & 10 \\ \dfrac{1}{2} & 2 \end{bmatrix}, c_B = (19, 50)$$

可求出

$$B^{-1}A = \begin{bmatrix} \dfrac{2}{3} & -\dfrac{10}{3} \\ -\dfrac{1}{6} & \dfrac{4}{3} \end{bmatrix} \cdot A$$

$$= \begin{bmatrix} 2 & \dfrac{4}{3} & 0 & 1 & \dfrac{2}{3} & -\dfrac{10}{3} \\ -\dfrac{1}{2} & -\dfrac{1}{3} & 1 & 0 & -\dfrac{1}{6} & \dfrac{4}{3} \end{bmatrix}$$

以及

$$c - c_B B^{-1}A = (9 + \lambda, 8, 50, 19, 0, 0)$$

$$= \left(-4 + \lambda, -\frac{2}{3}, 0, 0, -\frac{13}{3}, -\frac{10}{3}\right)$$

基 B 对应的单纯形表改为（见表 1.20）：

表 1.20

B	x_1	x_2	x_3	x_4	x_5	x_6	
x_4	2	$\frac{4}{3}$		1	$\frac{2}{3}$	$-\frac{10}{3}$	2
x_3	$-\frac{1}{2}$	$-\frac{1}{3}$	1		$-\frac{1}{6}$	$\frac{4}{3}$	1
$-s'$	$-4+\lambda$	$-\frac{2}{3}$			$-\frac{13}{3}$	$-\frac{10}{3}$	-88

如果要原最优解不变，根据最优判别准则，应有：

$$\left(-4+\lambda,-\frac{2}{3},0,0,-\frac{13}{3},-\frac{10}{3}\right) \leqslant 0 \quad \text{（注意，是最大值问题）}$$

得出

$$\lambda \leqslant 4$$

此即当 $\lambda \leqslant 4$，或 $c_1 = 9+\lambda \leqslant 13$ 时，原最优解：

$$x_1 = x_2 = 0, x_3 = 1, x_4 = 2$$

不变，最大利润仍为 88 万元。

对区间外的 λ 值，最优解的变化情况，可用参数规划方法去求解。

例如，当每万件产品 A 的价格超过 13 万元，即 $c_1 > 13$（即 $\lambda > 4$）时，检验数出现正数（这里是最大值问题），原最优基 B 已经不是最优基。用单纯形方法可得新基 $B' = (P_1, P_3)$，相应的单纯形表如表 1.21：

表 1.21

B'	x_1	x_2	x_3	x_4	x_5	x_6	
x_1	1	$\frac{2}{3}$		$\frac{1}{2}$	$\frac{1}{3}$	$\frac{5}{3}$	1
x_3			1	$\frac{1}{4}$		$\frac{1}{2}$	$\frac{3}{2}$
$-s'$		$\frac{-6+2\lambda}{3}$		$\frac{-4+\lambda}{2}$	$\frac{9+\lambda}{3}$	$\frac{30-5\lambda}{3}$	$84+\lambda$

欲使 B' 为最优基，应有上表中检验数全部非负，由此得到 $4 \leqslant \lambda \leqslant 6$。

即当 $13 \leqslant c_1 = 9+\lambda \leqslant 15$ 时，最优解变为：

$$x_1 = 1, x_3 = \frac{3}{2}, \text{其他 } x_j = 0$$

对应的目标函数最大值为 $s = 84+\lambda$，即 $88 \leqslant s \leqslant 90$。

假如 c_3 有波动，令 $c_3 = 50 + \lambda$。可以依前面的方法（留作练习），求出 $-\frac{5}{2} \leqslant \lambda \leqslant 2$。即当$\frac{95}{2} \leqslant c_3 \leqslant 52$ 时，最优解不变。

当 $c_3 > 52$，或 $c_3 < \frac{95}{2}$ 时，可用上面的方法求出最优解。

2. 约束条件的常数项的灵敏度分析

仍然使用前面的例子。设 b_1 有变动，令 $b_1 = 18 + \lambda$。因为 b 的改变与最优判别准则 $c - c_B B^{-1} A \geqslant 0$ 无关，只影响最优基对应的单纯形表中 $x_B = B^{-1} b$ 是否非负。如果非负，它仍是最优基。因此，当 b 有变动时，如果原来所得的基仍为最优基，应当有 $B^{-1} b \geqslant 0$。

由

$$
\begin{aligned}
B^{-1} b &= \begin{bmatrix} \frac{2}{3} & -\frac{10}{3} \\ -\frac{1}{6} & \frac{4}{3} \end{bmatrix} \cdot \begin{bmatrix} 18 + \lambda \\ 3 \end{bmatrix} \\
&= \left(2 + \frac{2\lambda}{3}, 1 - \frac{\lambda}{6}\right)^T \geqslant 0
\end{aligned}
$$

得到 $-3 \leqslant \lambda \leqslant 6$。即当 $15 \leqslant b_1 \leqslant 24$ 时，原最优基仍为最优基。注意，最优解和目标函数最优值都是 λ 的函数。

$$
\begin{aligned}
\max s &= c_B B^{-1} b = (19, 50)\left(2 + \frac{2\lambda}{3}, 1 - \frac{\lambda}{6}\right)^T \\
&= 88 + \frac{13\lambda}{3} \text{（万元）。}
\end{aligned}
$$

例如，当材料甲限用量增加 3 公斤（即 21 公斤），材料乙限用量不变时，应生产 4 万件 D，$\frac{1}{2}$ 万件 C，可得最大利润 101 万元。

另外，从前面的表可看出，材料甲的影子价格为$\frac{13}{3}$（万元／公斤）。也可以算出当材料甲的限用量增加 3 公斤时，最大利润增加 $3 \times \left(\frac{13}{3}\right) = 13$ 万元，即可得最大利润

$$88 + 13 = 101 \text{（万元）}$$

对于区间外的 λ 值，最优基变化情况可用参数规划求得。请读者自己验证：当 $\lambda > 6$ 时，最优基变为 $B' = (P_4, P_2)$，此时，$\lambda \geqslant 6$。最优解为：

$$x_2 = -3 + \frac{\lambda}{2}, x_4 = 6, \text{其他 } x_j = 0$$

$$\max s = 90 + 4\lambda$$

例如，当材料甲限用量为 50 公斤（即 $\lambda = 32$），乙材料的限用量不变时，就应生

产 13 万件 B,6 万件 D,这时的最大利润为 218 万元。

人们还可以做出,问题中添加新变量时的灵敏度分析,以及添加一个新的约束条件时的灵敏度分析。在这里我们不作进一步讨论了。

1.6 微机操作 —— 调用 YAJ 中,LP 决策支持系统

1.6.1 观察 LP 决策支持系统

这个程序可以解决具有许多变量(不包括松弛/人工变量)和约束条件的线性规划问题。用这个程序所解决问题的大小取决于你的计算机内存,你应该用下面格式输入数据:

Maximize	3.2	$GID1 + 4.0$	$GID2 - 5$	$GID3$
Subject to				
(1)	4	$GID1 + 2.5$	$GID2 + 3$	$GID3 \geqslant 50$
(2)	3.6	$GID1 + 7$	$GID2 - 2.5$	$GID3 \leqslant 86.9$
(3)	15.7	$GID1$	$+ 9$	$GID3 = 20$

(变量假定是非负的)

在这个程序里,你可以用不超过四个字符的办法为变量取名。缺损变量名是 $X1, X2, \cdots, Xn$。问题输入类似于你的公式。你可以选择打印最后解和灵敏度分析,你也可以显示你的问题,如果必要,也可对问题进行修改。借助 LP,你可以将问题存到磁盘上,或从磁盘上读取问题。此外,当问题解出后,你可以显示单纯形法的详细解答步骤。

1.6.2 输入问题

在输入问题时,你应该遵照下列约定:

(1) 100,100.0,+100,+100.0,1E2,和 1.0E+2 作用相同。

(2) −123,−1.23E2,和 −1.23E+2 作用相同。

(3) >=,>,=>,≥ 和 r 作用相同;<=,<,=<,≤ 和 s 作用相同。

(4) 输入一个数据后,请按 ENTER 键。

(5) 在同一荧屏上,你可以移动光标到矫正位置,通过按 BACKSPACE 键以改正错误。

(6) 当你对荧屏上的数据满意后,请按 SPACE BAR 键。

⑺ 在你输入过程中，按 ESC 键回到前一页，按 ／键进到下一页。

并且，按要求输入下列数据：

你要解最大值问题 ⑴，还是最小值问题 ⑵？(输入 1 或 2)

在你的问题中，有多少个变量？(输入数据 $\leqslant 500$)

在你的问题中，有多少个约束条件？(输入数据 $\leqslant 500$)

在你的问题中，有多少个 < 约束条件？(输入数据 $\leqslant 500$)

你要使用缺损变量名 ($X_1, X_2, \cdots, X_n$) 吗？(Y/N)

1.6.3 求解问题

在求解问题时，你可以选 4，显示单纯形法的每个步骤。但这个选择仅当问题是较小时，即当 $N + N1 + N2 + N3 * 2 \leqslant 9$ 时，才被允许。这里 N 是变量的个数，$N1$ 是约束条件“$\leqslant$”的个数，$N2$ 是约束条件“$=$”的个数，$N3$ 是约束条件“$\geqslant$”的个数；否则，仅显示重点项目。当你的问题只有 2 个变量，且约束条件小于 10 时，你也可以选择图形解法。

在图形解法中，表示第一个变量的横轴有 32 个单位，而表示第二个变量的纵轴有 20 个单位。你可以定义轴的标度，或者，让程序自动地定义它们。

下面我们用 1.5 节 1.5.2 中的例子在微机上进行操作。

首先使用 *Lingo*8.0 求解。照下面的式样输入问题 (每行分号结尾)：

$$\max = 9 * x1 + 8 * x2 + 50 * x3 + 19 * x4;$$
$$3 * x1 + 2 * x2 + 10 * x3 + 4 * x4 < 18;$$
$$2 * x3 + 0.5 * x4 < 3;$$

解报告如下：

Global optimal solution found at iteration： 2

Objective value： 88.00000

Variable	*Value*	*Reduced Cost*
X1	0.000000	4.000000
X2	0.000000	0.6666667
X3	1.000000	0.000000
X4	2.000000	0.000000

Row	*Slack or Surplus*	*Dual Price*
1	88.00000	1.000000
2	0.000000	4.333333
3	0.000000	3.333333

下面使用 *YAJ* 中的 *LP* 决策支持系统求解：

初始表

基	C (j)	x1	x2	x3	x4	S1	S2	B (i)	B (i)
		9.000	8.000	50.00	19.00	0	0		A (i , j)
S1	0	3.000	2.000	10.00	4.000	1	0	18.000	0
S2	0	0	0	2.000	0.5000	1	3.0000		
C (j) − Z (j)		9.000	8.000	50.00	19.00	0	0	0	
* Big M		0	0	0	0	0	0	0	

迭代 1

基	C (j)	x1	x2	x3	x4	S1	S2	B (i)	B (i)
		9.000	8.000	50.00	19.00	0	0		A (i , j)
S1	0	3.000	2.000	10.00	4.000	1	0	18.000	0
S2	0	0	0	2.000	0.5000	1	3.0000		
C (j) − Z (j)		9.000	8.000	50.00	19.00	0	0	0	
* Big M		0	0	0	0	0	0	0	

眼前的目标函数值 (Max) = 0,**进基** X3,**退基** S2

迭代 2

基	C (j)	x1	x2	x3	x4	S1	S2	B (i)	B (i)
		9.000	8.000	50.00	19.00	0	0		A (i , j)
S1	0	3.000	2.000	10.00	4.000	1	0	18.00	1.000
X3	50.00	0	0	1.000	0.250	0	0.500	1.500	1E + 20
C (j) − Z (j)		9.000	8.000	0	6.500	0	− 25.0	75.00	
* Big M		0	0	0	0	0	0	0	

眼前的目标函数值 (Max) = 75,**进基** X1,**退基** S1

迭代 3

基	C (j)	x1	x2	x3	x4	S1	S2	B (i)	B (i)
		9.000	8.000	50.00	19.00	0	0		A (i , j)
X1	9.000	1.000	0.667	0	0.500	0.333	− 1.67	1.000	1.500
X3	50.00	0	0	1.000	0.250	0	0.500	1.500	1E + 20
C (j) − Z (j)		0	2.000	0	2.000	− 3.000	− 10.00	84.00	
* Big M		0	0	0	0	0	0	0	

眼前的目标函数值 (Max) = 84,**进基** X2,**退基** X1

迭代 4

基	C(j)	x1	x2	x3	x4	S1	S2	B(i)	B(i)
		9.000	8.000	50.00	19.00	0	0		A(i,j)
X2	8.000	1.500	1.000	0	0.750	0.500	−2.50	1.500	2.000
X3	50.00	0	0	1.000	0.250	0	0.500	1.500	6.000
C(j)−Z(j)		−3.00	0	0	0.500	−4.00	−5.00	87.00	
* Big M		0	0	0	0	0	0	0	

眼前的目标函数值(Max) = 87,**进基** X4,**退基** X2

迭代 5

基	C(j)	x1	x2	x3	x4	S1	S2	B(i)	B(i)
		9.000	8.000	50.00	19.00	0	0		A(i,j)
X4	19.00	2.000	1.333	0	1.000	0.667	−3.333	2.000	0
X3	50.00	−0.500	−0.333	1.000	0	−1.670	1.333	1.000	0
C(j)−Z(j)		−4.000	−0.667	0	0	−4.333	−3.333	88.00	
* Big M		0	0	0	0	0	0	0	

最优目标函数值(Max) = 88

综合结果

变量		解	机会成本	变量		解	机会成本
序号	名字			序号	名字		
1	X1	0	+4.0000000	4	X4	+2.0000000	0
2	X2	0	+0.6666669	5	X5	0	+4.3333335
3	X3	+0.6666669	0	6	X6	0	+3.3333333
目标函数最大值 = 88,**迭代** 4 **次**							

下面再进行灵敏度分析：

目标函数系数的灵敏度分析

C(j)	最小 C(j)	原来值	最大 C(j)	C(j)	最小 C(j)	原来值	最大 C(j)
C(1)	−无穷大	+9.000000	+13.00000	C(3)	+47.500000	+50.000000	+52.000000
C(2)	−无穷大	+8.000000	+8.666667	C(4)	+18.500000	+19.000000	+20.000000

右端常数项的灵敏度分析

B(i)	最小 B(i)	原来值	最大 B(i)	B(i)	最小 B(i)	原来值	最大 B(i)
B(1)	+15.000000	+18.000000	+24.000000	B(2)	+2.2500000	+3.0000000	+3.5999999

第二章　整数线性规划

整数规划是数学规划的一个重要分支。在规划问题中，如果它的某些变量（或全部变量）要求取整数时，这个问题就称为整数规划问题。特别是当这个问题的约束条件是线性的，目标函数为线性函数时，就称此问题是整数线性规划问题。

求解整数规划问题是相当困难的。到目前为止，整数规划问题还没有一个很有效的办法。但是，由于在应用及理论方面提出的许多问题都可以归结为整数规划问题，所以，对整数规划的研究在理论上和实践上都有着重大意义。在这一章，我们只讨论整数线性规划问题的解法。

2.1　整数规划的例子

2.1.1　下料问题

例 1　设用某种型号的圆钢下零件 $A_1, A_2, \cdots, A_m$ 的毛坯，在一根圆钢上，下料的不同方式有 $B_1, B_2, \cdots, B_n$ 种。每种下料方式可以得到各种零件的毛坯数以及每种零件的需要数量如表 2.1：

表 2.1

下料方式 / 各方式下的零件毛坯数 / 零件名称	B_1	B_2	$\cdots$	B_n	各零件的需要量
A_1	c_{11}	c_{12}	$\cdots$	c_{1n}	a_1
A_2	c_{21}	c_{22}	$\cdots$	c_{2n}	a_2
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
A_m	c_{m1}	c_{m2}	$\cdots$	c_{mn}	a_m

问应怎样安排下料方式，使得既满足需要，又使所用原材料最少？

解　设用 B_j 方式下料的圆钢有 x_j 根，则这一问题的数学模型为：

$$\min s = \sum_{j=1}^{n} x_j$$

$$s.t.\begin{cases}\sum_{j=1}^{n} c_{ij}x_j \geqslant a_i, i = 1,\cdots,m \\ x_j \geqslant 0, \text{且全为整数}\end{cases}$$

在这个例子中，所有变量均要求是整数。这类问题称为纯整数规划问题。

2.1.2　工厂设址问题

例 2　要在 m 个不同地点计划修建 m 个规模不全相同的工厂，它们的生产能力分别为 $a_1,\cdots,a_m$（为简单起见，假设生产同一种产品）。第 i 个工厂的建设费用为 f_i，$i = 1,\cdots,m$。又有 n 个零售商店销售这种产品，对这种产品的需求量分别为 $b_1,\cdots,b_n$，从第 i 个工厂运送一个单位产品到第 j 个零售商店的费用为 c_{ij}。试决定应修建哪些工厂，使得既满足零售的需要，又使建设工厂和运输的总费用最小。

解　设 x_{ij} 是第 i 个工厂运往第 j 个零售商店的产品数量（$i = 1,\cdots,m$；$j = 1,\cdots,n$）。同时设

$$y_i = \begin{cases}1 & \text{如果决定修建第 } i \text{ 个工厂 } (i = 1,\cdots,m) \\ 0 & \text{相反}\end{cases}$$

问题则变为：

$$\min s = \sum_{i=1}^{m} \left[f_i y_i + \sum_{j=1}^{n} c_{ij}x_{ij}\right]$$

$$s.t.\begin{cases}\sum_{i=1}^{m} x_{ij} \geqslant b_j, j = 1,\cdots,n \\ \sum_{j=1}^{n} x_{ij} \leqslant a_i y_i, i = 1,\cdots,m \\ y_i = 0 \text{ 或 } 1, i = 1,\cdots,m \\ x_{ij} \geqslant 0, i = 1,\cdots,m; j = 1,\cdots,n\end{cases}$$

这个例子中，y_i 可取 0 或 1，而 x_{ij} 可取任意非负实数。这类整数规划问题称为混合整数规划问题。

例 3　有 m 台同一类型的机床，有 n 种零件在这类机床上进行加工。设各种零件的加工时间分别为 $a_1,\cdots,a_n$。问如何分配，使各机床的总加工任务相等，或者说尽可能均衡。

解　引进变量 x_{ij}：

$$x_{ij}=\begin{cases}1 & \text{若零件 } j \text{ 分配在第 } i \text{ 台上加工,}\\ 0 & \text{若零件 } j \text{ 不分配在第 } i \text{ 台上加工}\end{cases}$$

得到数学模型：

$$\min\left[\max\left(\sum_{j=1}^{n}a_jx_{1j},\sum_{j=1}^{n}a_jx_{2j},\cdots,\sum_{j=1}^{n}a_jx_{mj}\right)\right]$$

$$s.t.\sum_{j=1}^{n}x_{ij}=1,j=1,\cdots,n \text{（一个零件只能在一台机床上加工）}$$

$$x_{ij}=0 \text{ 或 } 1,i=1,\cdots,m;j=1,\cdots,n$$

在这个例子中,所有变量均取 0 或 1。这类整数规划问题称为 $0-1$ 规划问题。

2.1.3　背包问题

例 4　一个旅行者,为了准备旅行的必需物品,要在背包里装一些最有用的东西。但他最多只能携带 b 公斤的物品,而每件物品都只能整件携带。于是他给每件物品规定了一定的“价值”以表示其有用程度。设共有 m 件物品,第 i 件物品重 a_i 公斤,其价值为 c_i。试问在携带物品总重量不超过 b 公斤的条件下,携带哪些物品,可使总价值最大？

解　设

$$x_j=\begin{cases}1 & \text{当携带第 } j \text{ 件物品时}\\ 0 & \text{不携带第 } j \text{ 件物品时}\end{cases}$$

得到数学模型：

$$\max s=\sum_{i=1}^{m}c_ix_i$$

$$s.t.\begin{cases}\sum_{j=1}^{n}a_jx_j\leqslant b\\ x_j=0 \text{ 或 } 1,j=1,\cdots,n\end{cases}$$

这也是一个 $0-1$ 规划问题。我们又将这类规划问题叫做 $0-1$ 背包问题。

2.2　整数规划的解法 —— 分支定界法

对于整数线性规划问题,如果它的可行解集是有限集,我们可以将它的所有的可行解依次代入目标函数中,比较所得目标值的大小,从而求得最优解。

例如,在前面的背包问题中,若 x^0 是它的一个可行解：$x^0=(x_1^0,\cdots,x_n^0)$,这里,

$x_j^0=0$ 或 1。则可能的可行解最多有 $2n$ 个。当 n 较小时,逐个地比较这些解的优劣,则可求得最优解。这样的方法称为完全枚举法。

但是,当 n 较大时,利用完全枚举法几乎是不可能的。这时,我们可以利用部分枚举法来解这一类问题,这种方法也称分支定界法。我们仍以背包问题来说明这一解法的基本思想。

例 5 现有 100 万元资金打算在 5 个不同的地方 $L_1,\cdots,L_5$ 修建某种工厂。由于条件不同,所需投资分别为:$a_1=56$,$a_2=20$,$a_3=54$,$a_4=42$,$a_5=15$(单位:万元)。工厂建成后,每年能得到的利润分别为:$c_1=7$,$c_2=5$,$c_3=9$,$c_4=6$,$c_5=3$(单位:万元)。问应如何确定投资地点,使总投资不超过 100 万元,且使建成后每年所获总利润最多?

解 设

$$x_j=\begin{cases}1 & \text{在 } L_j \text{ 处投资建厂}\\ 0 & \text{不在 } L_j \text{ 处投资建厂}\end{cases}$$

得到数学模型:

$$\max f=7x_1+5x_2+9x_3+6x_4+3x_5$$

$$s.t.\begin{cases}56x_1+20x_2+54x_3+42x_4+15x_5\leqslant 100\\ x_j=0 \text{ 或 } 1, j=1,\cdots,5\end{cases} \qquad [0]$$

问题 [0] 显然有一个可行解 $x^{(0)}=(0,0,0,0,0)^T$,对应的目标函数值 $f(0)=0$。而且,它是目标函数值的一个下界。我们约定:在尚未求出问题的可行解时,目标函数值的下界记为 $-\infty$。

如果我们将问题 [0] 中的 $0-1$ 条件改为 $0\leqslant x_j\leqslant 1$,则问题 [0] 变为普通的线性规划问题:

$$\max f=7x_1+5x_2+9x_3+6x_4+3x_5$$

$$s.t.\begin{cases}56x_1+20x_2+54x_3+42x_4+15x_5\leqslant 100\\ x_j\geqslant 0\\ x_j\leqslant 1, j=1,2,3,4,5\end{cases} \qquad [0]'$$

问题 [0] 的可行解集是问题 [0]′ 的可行解集的子集合,问题 [0]′ 称为问题 [0] 的松弛问题。此问题容易求解,可能出现下列几种情形:

(1) 所得最优解恰好全为整数,当然,这个解也是问题 [0] 的解。

(2) 若松弛问题无可行解,则问题 [0] 也无可行解。

(3) 有最优解,但其分量不全为整数,它当然不是问题 [0] 的最优解。

为了求 [0]′ 的解,我们可以利用单纯形方法。但在这里,我们使用更简单的办法

来求它的解。

首先，计算$\frac{c_i}{a_i}$，$i=1,2,3,4,5$。这个比值的经济意义是投资1万元在第i地所能获得的利润。结果依次是：$\frac{1}{8},\frac{1}{4},\frac{1}{6},\frac{1}{7},\frac{1}{5}$。

将最大的比值所对应的变量取作1（$x_2=1$），再将次大比值所对应的变量取作1（$x_5=1$），… 依次做下去，使满足约束条件（2），就得到一个解：

$$x^{(0)\prime}=\left(0,1,1,\frac{11}{42},1\right)^T$$

对应的目标函数值$f^{(0)\prime}=\frac{130}{7}$。$f^{(0)\prime}$是问题［0］的目标值的下界，但它不是问题［0］的解，因为$x^{(0)\prime}{}_4=\frac{11}{42}$不是整数。

因为$x^{(0)\prime}{}_4$只能取0或1，可将问题［0］化为两个子问题：

$$\max f=7x_1+5x_2+9x_3+3x_5$$

$$s.t.\begin{cases}56x_1+20x_2+54x_3+15x_5\leqslant 100\\ x_4=0\\ x_j=0\text{ 或 }1,j=1,2,3,5\end{cases}\qquad [1]$$

及

$$\max f=7x_1+5x_2+9x_3+6+3x_5$$

$$s.t.\begin{cases}56x_1+20x_2+54x_3+42+15x_5\leqslant 100\\ x_4=1\\ x_j=0\text{ 或 }1,j=1,2,3,5\end{cases}\qquad [2]$$

这一过程我们可以用树形图表示，见图2.1。它形象地表明原问题［0］“分支”为子问题［1］和子问题［2］。

对于任一子问题，如果已求得它的最优解，或者知道它无解，或者断定它的最优解不是所求问题的最优解时，我们就称这一子问题已经探明，否则，称未探明。我们将一个子问题分为若干个子问题的过程称为分支。

用同样的方法，我们可以得到问题［1］的松弛问题，并可以求得这个松弛问题的解：

$$x^{(1)}=\left(\frac{11}{56},1,1,0,1\right)^T$$

对应目标函数值$f^{(1)}=\frac{147}{8}$。$x^{(1)}$仍不是整数解，可分别令$x_1=0$，或$x_1=1$，又将问题［1］分支为（见图2.2）：

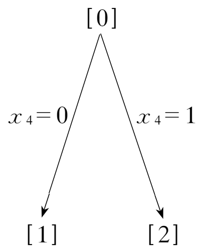


图 2.1

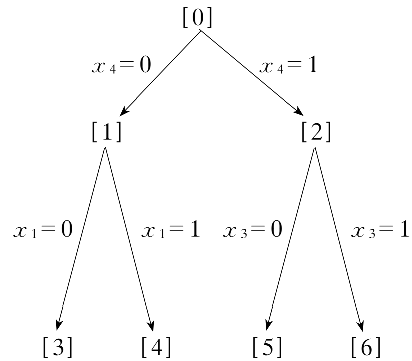


图 2.2

$$\max f = 5x_2 + 9x_3 + 3x_5$$

$$s.t.\begin{cases} 20x_2 + 54x_3 + 15x_5 \leqslant 100 \\ x_4 = 0 \\ x_1 = 0 \\ x_j = 0 \text{ 或 } 1, j = 2,3,5 \end{cases} \qquad [3]$$

及

$$\max f = 7 + 5x_2 + 9x_3 + 3x_5$$

$$s.t.\begin{cases} 56 + 20x_2 + 54x_3 + 15x_5 \leqslant 100 \\ x_4 = 0 \\ x_1 = 1 \\ x_j = 0 \text{ 或 } 1, j = 2,3,5 \end{cases} \qquad [4]$$

解问题[3]的松弛问题得：

$$x^{(3)} = (0,1,1,0,1)^T$$

$$f^{(3)} = 17$$

显然，$x^{(3)}$ 也是问题[3]的最优解。这时，问题[3]已探明。同时，$f^{(3)} = 17$ 也是原问题[0]目标函数值的一个下界。这样，问题[0]的可行解的下界从 0 修改为 17。

再解问题[4]所对应的松弛问题，其最优解为

$$x^{(4)} = \left(1,1,\frac{1}{6},0,1\right)^T$$

$$f^{(4)} = \frac{33}{2}\ (< 17)$$

虽然 $x^{(4)}$ 仍不是问题[4]的解，但可不必再分支了，于是子问题[4]也已探明。

最后，我们总结一下解某一子问题的松弛问题时，可能碰到的几种情况和处理方法。

(1) 对应的松弛问题无解，则该子问题也无解。情况探明，不再分支，凡已探明的子问题均用一横线打住。

(2) 对应的松弛问题的解，恰是该子问题的解。这也就求得了该子问题的最优解。情况探明，并得到新的下界(这一过程叫做定界)。

(3) 对应的松弛问题的解不是这个子问题的解。则当其目标值大于现有最好下界时，该子问题需继续分支；而当其目标函数值不超过现有最好下界时，则该子问题也已探明。到此，可以“剪支”。

现在，再看子问题[2]，可求得其松弛问题的解为

$$x^{(2)} = \left(0,1,\frac{23}{54},1,1\right)^T$$

$$f^{(2)} = \frac{107}{6} \ (> 17)$$

应将子问题[2]继续分支。分别令x_3为0、1，得到(见图2.1)

$$\max f = 7x_1 + 5x_2 + 6 + 3x_5$$

$$s.t.\begin{cases} 56x_1 + 20x_2 + 42 + 15x_5 \leqslant 100 \\ x_4 = 1 \\ x_3 = 0 \\ x_j = 0 \text{ 或 } 1, j = 1,2,5 \end{cases} \qquad [5]$$

及

$$\max f = 7x_1 + 5x_2 + 9 + 6 + 3x_5$$

$$s.t.\begin{cases} 56x_1 + 20x_2 + 54 + 42 + 15x_5 \leqslant 100 \\ x_4 = 1 \\ x_3 = 1 \\ x_j = 0 \text{ 或 } 1, j = 1,2,5 \end{cases} \qquad [6]$$

由于

$$x^{(5)} = \left(\frac{23}{56},1,0,1,1\right)^T$$

$$f^{(5)} = \frac{135}{8} \ (< 17)$$

可剪支，问题[5]已探明。

又

$$x^{(6)} = \left(0, \frac{1}{5}, 1, 1, 0\right)^T$$

$$f^{(6)} = 16 \ (< 17)$$

也可剪支,问题 [6] 也探明。

至此,全部子问题均已探明。原问题的最优解是:

$$x^* = x^{(3)} = (0,1,1,0,1)^T$$

$$\max f = f^{(3)} = 17$$

整个过程可用树形图表示 (见图 2.2)。

如果我们用完全列举法,就要试 $2^5 = 32$ 次才能得出结果。由此例可以看出,对于这类问题,关键首先在于如何确定原问题的松弛问题,而松弛问题又存在比较有效的解决办法;其次,就是如何确定分支的规则,将原问题分支为若干个子问题;最后,在求解过程中,要不断考察能否修改下界 (定界) 以便于剪支。如果所有子问题都已探明,就可以获得原问题的解。

下面,我们再看如何应用分支定界法使用软件 YAJ 去解一般的整数线性规划问题。

例 6　求解问题

$$\max f = x_1 + x_2$$

$$s.t.\begin{cases} x_1 + \dfrac{9}{14}x_2 \leqslant \dfrac{51}{14} \\ -2x_1 + x_2 \leqslant \dfrac{1}{3} \\ x_1, x_2 \geqslant 0 \\ x_1, x_2 \text{ 均为整数} \end{cases}$$

2.3　微机操作 —— 调用 YAJ 中,IP 决策支持系统

2.3.1　观察 IP 决策支持系统

这个程序使用分枝定界方法,解具有许多变量 (不包括松弛/人工变量) 和约束条件 (不包括界) 的混合整数线性规划问题。用这个程序所解问题的大小取决于你的计算机内存。你应该遵循下列格式输入数据:

Maximize　3.2　$GID1 + 4.0$　$GID2 - 5$　$GID3$

Subject to

(1)　4　$GID1 + 2.5$　$GID2 + 3$　$GID3 \geqslant 50$

(2)　3.6　$GID1 + 7$　$GID2 - 2.5$　$GID3 \leqslant 86.9$

界：$GID1 \geqslant 2$，而且是整数（变量假定是非负的）

2.3.2 输入问题

这一部分同前 1.6.2 输入问题。

2.3.3 求解问题

我们就用前面的例 2 在微机上进行求解。首先，使用 *Lingo* 8.0 进行求解。按下面格式输入问题

$$\max = x1 + x2;$$
$$x1 + 0.642875 * x2 < 3.642875;$$
$$-2 * x1 + x2 < 0.333333;$$
$$@Gin(x1);$$
$$@Gin(x2);$$

其解报告为：

Global optimal solution found at iteration：　　0

Objective value：　　4.000000

Variable	*Value*	*Reduced Cost*
X1	2.000000	−1.000000
X2	2.000000	−1.000000

Row	*Slack or Surplus*	*Dual Price*
1	4.000000	1.000000
2	0.3571250	0.000000
3	2.333333	0.000000

下面再用 *YAJ* 中的 *LP* 决策支持系统求解。

当前的分枝－定界解——迭代 1					
下界	变量	上界	变量	解	目标函数系数
0　<	$x1$	< 无穷大	$x1$	+1.5000002	+1.0000000
0　<	$x2$	< 无穷大	$x2$	+3.3333335	+1.0000000
目标函数的非整数解 = 4.833334			*ZL*（下界）= −1*E* + 20		

当前的分枝 – 定界解 —— 迭代 2

下界	变量	上界	变量	解	目标函数系数
2 <	$x1$	< 无穷大	$x1$	+ 2.0000000	+ 1.0000000
0 <	$x2$	< 无穷大	$x2$	+ 2.5555561	+ 1.0000000

目标函数的非整数解 = 4.555556　　ZL **(下界)** = – 1E + 20

当前的分枝 – 定界解 —— 迭代 3

下界	变量	上界	变量	解	目标函数系数
2 <	$x1$	< 无穷大			
3 <	$x2$	< 无穷大			

这个分枝没有可行解

当前的分枝 – 定界解 —— 迭代 4

下界	变量	上界	变量	解	目标函数系数
2 <	$x1$	< 无穷大	$x1$	+ 2.3571429	+ 1.0000000
0 <	$x2$	< 2	$x2$	+ 2.0000000	+ 1.0000000

目标函数的非整数解 = 4.357143　　ZL **(下界)** = – 1E + 20

当前的分枝 – 定界解 —— 迭代 5

下界	变量	上界	变量	解	目标函数系数
3 <	$x1$	< 无穷大	$x1$	+ 3.0000000	+ 1.0000000
0 <	$x2$	< 2	$x2$	+ 1.0000000	+ 1.0000000

目标函数的非整数解 = 4 >　　ZL **(下界)** = – 1E + 20

当前的分枝 – 定界解 —— 迭代 6

下界	变量	上界	变量	解	目标函数系数
2 <	$x1$	< 2	$x1$	+ 2.0000000	+ 1.0000000
0 <	$x2$	< 2	$x2$	+ 2.0000000	+ 1.0000000

当前的目标函数值 = 4 <　　ZL = 4

当前的分枝－定界解 —— 迭代 7					
下界	变量	上界	变量	解	目标函数系数
0 <	$x1$	< 1	$x1$	+ 1.0000000	+ 1.0000000
0 <	$x2$	< 无穷大	$x2$	+ 2.3333330	+ 1.0000000
当前的目标函数值 = 3.333333 <			ZL = 4		

例 2 的综合结果					
变量 序号,名字	解	目标函数系数	变量 序号,名字	解	目标函数系数
1 $x1$	3.0000000	1.0000000	2 $x1$	+ 1.0000000	+ 1.0000000
最大目标函数值 = 4，			总的迭代 7 次		

注意 上面两个不同软件，恰好给出本问题的两个不同最优解。

第三章　运载问题

3.1　例题和模型

例 1　现有一批货物,从 m 个仓库(发点:$S_1,S_2,\cdots,S_m$)运往 n 个销地(收点:$D_1,D_2,\cdots,D_n$),S_i 处有货物 a_i 吨,D_j 处需货物 b_j 吨。又,从 S_i 到 D_j 的运价为 c_{ij} 元/吨。问如何安排,既满足各销地需要,又能使总运费最小。

解　设从 S_i 运往 D_j 的货物量为 x_{ij} 吨。则

$$\min z = \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}x_{ij} \tag{i}$$

$$s.t.\begin{cases}\sum\limits_{j=1}^{n} x_{ij} = a_i, i = 1,\cdots,m & \text{(ii)}\\ \sum\limits_{i=1}^{m} x_{ij} = b_j, j = 1,\cdots,n & \text{(iii)}\end{cases}$$

—— 平衡运输问题。

或者,(ii)式变为

$$\sum_{j=1}^{n} x_{ij} \leqslant a_i, i = 1,\cdots,m \tag{ii$'$}$$

—— 供过于求的运输问题。

或者,(iii)式变为

$$\sum_{i=1}^{m} x_{ij} \leqslant b_j, j = 1,\cdots,n \tag{iii$'$}$$

—— 供不应求的运输问题。

这是一个典型的线性规划问题,可以用单纯形法求解。不过,基于运输问题的特

点，我们有更简单的办法。

3.2　解决平衡问题的表上作业法

3.2.1　最小元素法（MM）

我们通过下面的例子（表 3.1）介绍最小元素法。

表 3.1

运价 D_j / S_i	D_1	D_2	D_3	D_4	供应量
$S1$	1000 \ 3	4000 \ 2	\ 7	\ 6	5000
$S2$	2500 \ 7	\ 5	2000 \ 2	1500 \ 3	6000
$S3$	2500 \ 2	\ 5	\ 4	\ 5	2500
需求量	6000	4000	2000	1500	13 500

求如何安排，运费最小。

解　(1) 求出

$$\min\{c_{ij}: i = 1,2,3; j = 1,2,3,4\} = c_{12} \text{（不止一个，适当选取一个）}$$

及

$$\min\{a_1, b_2\} = b_2 = 4000$$

优先满足 D_2 需求。取 $x_{12} = 4000$，划去第二列。S_1 处余 1000（请记住）。

(2) 求出

$$\min\{c_{ij}: i = 1,2,3; j = 1,3,4\} = c_{23}$$

及

$$\min\{a_2, b_3\} = b_3 = 2000$$

满足 D_3 需求。取 $x_{23} = 2000$，划去第三列。S_2 处余 4000（请记住）。

(3) 求出

$$\min\{c_{ij}: i = 1,2,3; j = 1,4\} = c_{31}$$

及

$$\min\{a_3, b_1\} = a_3 = 2500$$

由 S_3 供应。取 $x_{31} = 2500$，划去第三行 D_1 处尚需 3500（请记住）。

(4) 求出

$$\min\{c_{ij}: i=1,2; j=1,4\} = c_{11}$$

及

$$\min\{S_1 \text{ 余量 } 1000, D_1 \text{ 需要量 } 3500\} = 1000$$

当从 S_1 运 1000 至 D_1。取 $x_{11}=1000$，划去第一行。D_1 处尚需 2500（请记住）。

(5) 求出

$$\min\{c_{ij}: i=2; j=1,4\} = c_{24}$$

及

$$\min\{S_2 \text{ 余量 } 4000, D_1 \text{ 需要量 } 2500\} = 2500$$

当从 S_2 运 2500 至 D_1。取 $x_{21}=2500$，划去第一列。S_2 处余 1500（请记住）。

(6) 运 S_2 处余量 1500 至 D_4。取 $x_{24}=1500$，同时划去第二行和第四列。

至此，给出了一个可行方案：其结果已填入表 3.1 中。相应运费为 42 000 元。

注　在表格内填有数字的格 —— 称满格，应为 $6=3+4-1$ 格。一般，m 行 n 列表，按上述办法满格应为 $m+n-1$ 个（故在中途若同时划去某行某列，则在某空格上填补一个 0）。

3.2.2　运输方案的改进 —— 位势法

与单纯法类似，对上面的可行解，需要判断其是否为最优；若不是，应反复改进，直至求得最优解，我们使用的方法叫位势法。

(1) 求位势。

我们对每行每列分别定义一个位势：$u_i, v_j, i=1,2,3; j=1,2,3,4$。要求它们对满格有：

$$u_i + v_j = c_{ij} \qquad \text{（计 6 个等式）}$$

为求得所有位势，一般常选 $u_1=0$，解出

$$v_1=3, v_2=2, u_2=4, v_3=-2, v_4=-1, u_3=-1$$

注　所求位势 u_i、v_j 实际是原问题关于基本变量组的对偶解。

(2) 求判别数。

据位势，由下面关系求出空格的判别数：$c_{ij}'=c_{ij}-u_i-v_j$

例如，$c_{13}'=7-0-(-2)=9$，$c_{22}'=5-4-2=-1$，…

注　这里的判别数与单纯形法中的判别数一致。

我们可以用闭回路办法（每个空格总可以找到一个回路，除一个顶点在空格内，其他顶点都在满格内）证实，c_{ij}' 的实际意义是：在空格 $(i-j)$ 内增运一个单位所增加的运费。这样，当判别数为“+”的空格，不能增调物资，否则会使运费增大；相反，若为负，该路线可以增调物资，会降低运输费用。于是，当所有 $c_{ij}' \geqslant 0$ 时，该方案已

为最优，计算可以停止。

(3) 方案的调整。

在我们的问题中，仅 $c_{22}' = -1 < 0$。因此，需进行调整：从 S_2 增运物资到 D_2，以降低费用。找到回路 (2 − 2) − (2 − 1) − (1 − 1) − (1 − 2) − (2 − 2)。看出当从 (2 − 1) 减去 2500 成为空格，(1 − 2) 减去 2500，而在 (1 − 1) 和 (2 − 2) 内各增加 2500。得到一个新的可行解：

$x_{11} = 3500$，$x_{12} = 1500$，$x_{22} = 2500$，$x_{23} = 2000$，$x_{24} = 1500$，$x_{31} = 2500$。

此时运费为 39 500 元，较第一次减少了 2500 元。

注　改进过程中需要注意的是，每一新方案都要使满格数为 $m + n - 1$。有时在调整的闭回路中，需要减少运输量的格中有两个以上相等的最小数，调整时在相应的空格中填上这个最小数而这些相等最小数的满格都成了空格。这就需要在这些格中任选一个留为空格，其余补填 0。补填 0 的格视为满格。

(4) 重复 (1) ～ (3) 步骤，直到所有判别数非负为止。

由于对新的可行解所求得的判别数已经非负，这个解已是最优解。

3.3　微机操作 —— 调用 YAJ 中，TRP 决策支持系统

3.3.1　观察 TRP 决策支持系统：

这个程序解运输／转载问题。一个发点或供应点规定只有发出流；一个终点或接收点规定只有进入流；而一个转运点规定两者都有。容量或需求量假定是整数；费用或利润系数假定是实数。本程序提供输入、存储和修改问题的能力；也可以求解问题，显示问题，打印问题的解。

在输入问题后，程序自动将问题转化为运输问题，并用运输单纯形法解之。为了求初始可行解，程序提供了许多选择。对于小的问题，你可以显示解答步骤。程序允许你定义发点、收点、转运点的名字。要求输入的数据是：每点的容量或需求量，点与点之间的运输费用或利润。

3.3.2　输入问题

在输入问题时，请遵照下列约定：

1. 熟悉下列要求，它们规定了关于问题的一般格式。
2. 除非使用缺损名，否则输入每个节点的名字。

3. 然后输入每一点的容量，或需求量。对于转运点，视其纯供应／需求量，输入一个正／负数。

4. 然后输入节点之间的运输费用或利润。当使用固定格式时，为了表示两点没有直接联系（无流），可以输入一个很大的数，或 M／$-M$。

5. 键 *BACKSPACE BAR* 可用于移动光标到你需要修正数据的地方以改正错误；当使用固定格式时，按 *ESC* 键，回到前一页，按／键，进到下一页。

3.3.3 求解问题

在解一个问题时，如果你的问题规模是 $M < 5$，$N < 6$，你可以显示 *MODI* 方法的每一个迭代，这里 M 是发点和转运点的总数，N 是收点和转运点的总数。

你可以选择八个方法之一找初始解。缺损的办法是行最小（*RM*）法。我们讲的是（*MM*）法。

3.3.4 解答及结果显示

前面表 3.1 的结果是：

SN \ DN	D_1	D_2	D_3	D_3	供应量	u_i
S1	1000 \ 3	4000 \ 2	\ 7	\ 6	5000	0
S2	2500 \ 7	\ 5	2000 \ 2	1500 \ 3	6000	0
S3	2500 \ 2	\ 5	\ 4	\ 5	2500	0
需求量	6000	4000	2000	1500	13 500	
v_j	0	0	0	0		

最小目标函数值 = 42 000

SN \ DN	D_1	D_2	D_3	D_3	供应量	u_i
S1	1000 \ 3	4000 \ 2	\ 7	\ 6	5000	0
S2	2500 \ 7	* * \ 5	2000 \ 2	1500 \ 3	6000	4.000
S3	2500 \ 2	\ 5	\ 4	\ 5	2500	−1.000
需求量	6000	4000	2000	1500	13 500	
v_j	3.000	2.000	−2.000	−1.000		

当前目标函数最小值 = 42 000，$e(2,2) = -1$

DN \ SN	D_1	D_2	D_3	D_3	供应量	u_i
S1	3500 \ 3	1500 \ 2	\ 7	\ 6	5000	0
S2	\ 7	2500 \ 5	2000 \ 2	1500 \ 3	6000	3.000
S3	2500 \ 2	\ 5	\ 4	\ 5	2500	−1.000
需求量	6000	4000	2000	1500	13 500	
v_j	3.000	2.000	−1.000	0		

当前目标函数最小值 = 39 500,具有多重最优解

综合结果

发地	收地	运量	运价	判别数	发地	收地	运量	运价	判别数
S1	D1	+3500	+3.000	0	S2	D3	+2000	+2.000	0
S1	D2	+1500	+2.000	0	S2	D4	+1500	+3.000	0
S1	D3	0	+7.000	+8.000	S3	D1	+2500	+2.0000	0
S1	D4	0	+6.000	+6.000	S3	D2	0	+5.0000	+4.000
S2	D1	0	+5.000	+1.000	S3	D3	0	+4.0000	+6.000
S2	D2	+2500	+5.000	0	S3	D4	0	+5.000	+6.000

目标函数最小值 = 39 500 (多重解),迭代 1 次

与我们给出的结果一致。

3.4 不平衡运输问题

设问题由(ⅰ)、(ⅱ)′、(ⅲ)组成,且

$$\sum_{i=1}^{m} a_i > \sum_{j=1}^{n} b_j \text{ (}\sum_{i=1}^{m} a_i < \sum_{j=1}^{n} b_j \text{ 时无可行解)}$$

我们可虚设一个收点 D_{n+1}。它的需求量为:

$$b_{n+1} = \sum_{i=1}^{m} a_i - \sum_{j=1}^{n} b_j$$

取 $c_{i,n+1} = 0$,$x_{i,n+1}$ 为不发运的货物数量。这样,问题就化为一个有 m 个发点,$n+1$ 个收 点的收发平衡的运输问题了。

再考虑由(ⅰ)、(ⅱ)及

$$\sum_{i=1}^{m} x_{ij} \geqslant b_j, j = 1,\cdots,n \qquad (ⅲ)''$$

所组成的运输问题。其中,$\sum_{i=1}^{m} a_i > \sum_{j=1}^{n} b_j$ (当 $\sum_{i=1}^{m} a_i < \sum_{j=1}^{n} b_j$ 时无可行解)。

类似，虚设一个 D_{n+1} 收点。令：

$$b_{n+1} = \sum_{i=1}^{m} a_i - \sum_{j=1}^{n} b_j$$

$c_{i,n+1} = \min\{c_{ij};1 \leqslant j \leqslant n\}$，$i = 1,2,\cdots,m$。视 D_{n+1} 为一个转运点。$S_{i\#}$ 处货物先运 $x_{i\#,n+1}$ 到 D_{n+1} 处，再转运到 D_{j*} 处。当解出此问题最优解 $\hat{x}_{ij}$ 后，原问题最优解为：

$$x_{ij} = \begin{cases} x_{ij} & i = 1,\cdots,m;j \neq j^* \\ \hat{x}_{ij} + x_{i,n+1} & i = 1,\cdots,m;j = j^* \end{cases}$$

第四章　指派问题

4.1　例题和模型

例 1　某公司有 m 件任务必须完成，而该公司至少有 m 个雇员能以不同的时间完成 m 项任务中的任何一项。如果指派每个雇员完成一项且仅仅一项任务；雇员 i 完成任务 j 费时 c_{ij}。应如何分配，使完成 m 项任务的总时间最小？

解　设

$$x_{ij} = \begin{cases} 1 & \text{雇员 } i \text{ 分配到任务 } j \\ 0 & \text{雇员 } i \text{ 未分配到任务 } j \end{cases}$$

于是当有：

$$\min z = \sum_{i=1}^{m} \sum_{j=1}^{m} c_{ij} x_{ij}$$

$$s.t.\begin{cases} \sum_{j=1}^{m} x_{ij} = 1, i = 1, \cdots, m \\ \sum_{i=1}^{m} x_{ij} = 1, j = 1, \cdots, m \\ x_{ij} = 0 \text{ 或 } 1, i, j = 1, \cdots, m \end{cases}$$

注　问题中的任务可以是任何类型的工作 $(T_1, T_2, \cdots, T_m)$，雇员可以是任何类型的对象 $(O_1, O_2, \cdots, O_m)$，某雇员完成一项任务的时间可以看做：某对象从事给定工作时，与之相关的效能。

这是典型的整数规划问题；也是更一般运输问题的特殊情况：$a_i = b_j = 1$，对所有 i, j 成立。这自然可以用已经学过的办法求解，但匈牙利法则更为简单且有效。

4.2 匈牙利法

人们把匈牙利法(Hungarian method)归于匈牙利数学家狄·考尼格(D.Konig),他证明了这个方法的主要定理。

我们称矩阵$(c_{ij})_{m\times m}$为问题的效能矩阵。匈牙利法成功地修改了效能矩阵的行和列,直到在每一行和每一列里,至少有一个等于0的元素,从而得到与这些0元素相应的完全分配。当把它应用于原效能矩阵时,这个完全分配是一个最优分配,总的效能将是最小。这种方法总是在有限的步骤内收敛于一个最优解。方法的依据是:任何行或列,加、减一个常数并不改变最优分配。例如,由第i行减去3个单位和第j列加上2个单位,那么,目标函数值将是:

$$\min z = \sum_{i=1}^{m}\sum_{j=1}^{m} c_{ij}x_{ij} - 3\sum_{j=1}^{m} x_{ij} + 2\sum_{i=1}^{m} x_{ij} = \sum_{i=1}^{n}\sum_{j=1}^{m} c_{ij}x_{ij} - 3 + 2$$

从目标函数中加上或减去一个常数不会改变最优解,这是由于每一个基础可行解将在目标函数中加上或减去相同的数量。

上述的概念可以推广为:从效能矩阵中,第i行每个元素减去a_i,$i = 1,\cdots,m$;第j列每个元素减去b_j,$j = 1,\cdots,m$。新的目标函数值为:

$$\min z = \sum_{i=1}^{m}\sum_{j=1}^{m} c_{ij}x_{ij} - \sum_{i=1}^{m} a_i - \sum_{j=1}^{m} b_j$$

从目标函数中再一次减去常数$\sum_{i=1}^{m} a_i$和$\sum_{j=1}^{m} b_j$,不改变基础可行解。

下面通过表4.1给出的实例,说明方法。

表4.1

任课 备课时间 教员	线性规划	排队论	动态规划	回归分析
A	2	10	9	7
B	15	4	14	8
C	13	14	16	11
D	4	15	13	9

问怎么分配,使总的备课时间最少。

(1)修改效能矩阵,使之成为每一行和每一列至少有一个零元素的退缩阵。

从原效能矩阵的第1,2,3,4行分别减去2,4,11,4;再从第3列减去5。得到:

$$\begin{vmatrix} 0 & 8 & 2 & 5 \\ 11 & 0 & 5 & 4 \\ 2 & 3 & 0 & 0 \\ 0 & 11 & 4 & 5 \end{vmatrix}$$

注意，由此得到的完全分配方案的总效能必须加上 $2+4+11+4+5=26$。

(2) 试图制作一个完全分配方案，该方案与上面矩阵的零元素相对应。

为此，从第一行开始，依次检查各行，直到找出只有一个未标记的零元素的一行为止。如果在零元素上有一个符号 △ 或 ×，则称零元素已标记。符号 △ 表示分配 △ 所在行的教员担任 △ 所在那一列的那一门课程。对未做标记的零元素标 △ 后，应对同一列其他的零元素划上 ×。重复这一过程，直到每一行中没有尚未标记的零元素或至少有两个零元素。据此，得：

$$\begin{vmatrix} 0^{\triangle} & 8 & 2 & 5 \\ 11 & 0^{\triangle} & 5 & 4 \\ 2 & 3 & 0^{\triangle} & 0^{\times} \\ 0^{\times} & 11 & 4 & 5 \end{vmatrix}$$

现在，依次检查每列中只有一个未标记的零元素，并给未标记的零元素标 △，对同一行其他零元素 (如果有的话) 划 ×。重复这一过程，直到每列中没有尚未标记的零元素或至少有两个零元素。照此，第三列的 0 划 △，第三行另一 0 划 ×。

显然，它不能给出一个完全分配方案。这就需要划出最少数目的水平线和垂直线，要求它们穿过每行每列的零元素至少一次。其方法如下：

(a) 检查所有尚未标记 △ 的行，并记上 √ (为第四行)。

(b) 检查那些尚未检查过的，而在已检查过的行中有零元素的列，并标上 √ (为第一列)。

(c) 检查那些尚未检查过的，而在已检查过的列中有标记 △ 的行，并标上 √ (为第一行)。

(d) 重复 (b)，(c)，直到不能进一步检查为止 (本例即停止)。

(e) 在所有未检查的行和已检查过的列划直线，得：

$$\begin{vmatrix} 0^{\triangle} & 8 & 2 & 5 \\ 11 & 0^{\triangle} & 5 & 4 \\ 2 & 3 & 0^{\triangle} & 0 \\ 0 & 11 & 4 & 5 \end{vmatrix}$$

如果效能矩阵有 m 行和 m 列，而且可制定出一个总效能为零的完全分配，那么，在现有缩减的效能矩阵内最小线数必须为 m (最小线数是穿过每个零元素至少

一次所必需的线数)。本例中,只需 3 条线穿过每个零元素至少一次,所以,尚不能得到完全的可行分配。换句话说,总的准备时间为 26 单位或更少的完全分配是不可能的。

从第 (e) 步缩减矩阵开始,最后得到总效能为最小的完全解的过程是:

(f) 在上述最后的缩减矩阵中,检查那些没有一根线通过的那些元素。设 K 为其最小元素,找出含有未画线元素的各行,将这些行的每个元素减去 K。本例中,$K=2$,因而由第 1 行、第 4 行减去 2,可得:

$$\begin{vmatrix} -2 & 6 & 0 & 3 \\ 11 & 0 & 5 & 4 \\ 2 & 3 & 0 & 0 \\ -2 & 9 & 2 & 3 \end{vmatrix}$$

注　此时,在完全分配的总效能上需加上 4。

(g) 第 (f) 步的减法使第 (e) 步中划垂直线的各列中某些元素变为负,因此,对第 (e) 步划垂直线的每一列中的所有元素加 K。据此,本例第一列的每个元素加 2 可得:

$$\begin{vmatrix} 0 & 6 & 0 & 3 \\ 13 & 0 & 5 & 4 \\ 4 & 3 & 0 & 0 \\ 0 & 9 & 2 & 3 \end{vmatrix}$$

注　此时,在完全分配的总效能上需减去 2。

可以证明:如果用第 (g) 步中的缩减减阵可以制定只含零元素的完全分配方案,那么,应用同一分配方案对原矩阵总效能也必须是最小。

为了确定完全分配方案,对上述第 (g) 步得到的新的效能矩阵,重复进行制定分配方案的一般过程。具体说:

① 第二行确定一种分配。

② 第四行确定一种分配,且在第一行第一列的零元素上标记 ×。

③ 第一行确定一种分配,且在第三行第三列的零元素上标记 ×。

④ 第三行确定一种分配。

结果为:

$$\begin{vmatrix} 0^{\times} & 6 & 0^{\triangle} & 3 \\ 13 & 0^{\triangle} & 5 & 4 \\ 4 & 3 & 0^{\times} & 0^{\triangle} \\ 0^{\triangle} & 9 & 2 & 3 \end{vmatrix}$$

于是一个完全的最优方案制定出来了：

教员 A 担任动态规划，教员 B 担任排队论，教员 C 担任回归分析，教员 D 担任线性规划。

总的最小备课时间为：$9+4+11+4=28$ 单位（或者，$21+5+2+2-2=28$）。

4.3 微机操作 —— 调用 YAJ 中，ASMP 决策支持系统

4.3.1 观察 ASMP 决策支持系统

这个程序解具有许多对象和任务的指派问题。任务可以表示工作，而对象可以表示工作者。问题标准可以是最小值问题或最大值问题，它取决于对于的每个对象和指派任务的花费/利润系数。程序为数据的输入和修改提供了一个方便的格式。

对于不超过 9 个对象和 9 项任务的小问题，你可以显示匈牙法（Hungarian method）的每一次迭代。本程序可解很大的问题但不显示每一个迭代。

这个程序允许你用不超过 6 个字符给出对象和任务的名字。缺损的名字是 $O1$，$O2$，…，On 和 $T1$，$T2$，…，Tn。

4.3.2 输入问题

在你输入一个问题时，请遵照下列约定：

(1) 熟悉下列规定，从它们可找到处理问题的一般情形。

(2) 如果你不使用缺损名，当输入对象和任务的名字。

(3) 然后对每一个可能的指派，输入花费/利润系数。对于一个不可能的指派，当输入一个很大的 +/- 值，或 $+M/-M$。

其他约定同前。

4.3.3 求解问题

当解问题时，如果你的问题有比 10 小的对象和比 10 小的任务，你可以显示匈牙利法（Hungarian method）的每一步迭代。

前面表 4.1 给出的实例的结果显示：

输入费用／利润系数

对象	任务							
O1	T1	2	T2	10	T3	9	T4	7
O2	T1	15	T2	4	T3	14	T4	8
O3	T1	13	T2	14	T3	16	T4	11
O4	T1	4	T2	15	T3	13	T4	9

初始表

Tk / Ob	T1	T2	T3	T4	
O1	2.000	10.00	9.000	7.000	
O2	15.00	4.000	14.00	8.000	
O3	13.00	14.00	16.00	11.00	
O4	4.000	15.00	13.00	9.000	

迭代 1

Tk / Ob	T1	T2	T3	T4	
O1	0	8.000	2.000	5.000	
O2	11.00	0	5.000	4.000	←
O3	2.000	3.000	0	0	←
O4	0	11.00	4.000	5.000	
	/\				

最后表（总的迭代 2 次）

Tk / Ob	T1	T2	T3	T4	
O1	0	6.000	0	3.000	
O2	13.00	0	5.000	4.000	←
O3	4.000	3.000	0	0	
O4	0	9.000	2.000	3.000	
	/\		/\	/\	

指派问题的综合结果

对象	任务	费用/利润	对象	任务	费用/利润
$O1$	$T3$	9.000	$O3$	$T4$	11.00
$O2$	$T2$	4.000	$O4$	$T1$	4.000
目标函数最小值 = 28，总的迭代 2 次					

4.4 对象与活动数目不等情形的指派问题

假设某经理必须从8辆卡车中安排5辆卡车到5个指定地点去装货。8辆卡车放在不同地点，问题是安排5辆卡车到5个地点的运费最少。运费在表4.2中给出。

因为卡车数较地点数多，将三个虚设的地点F、G、H加到表中，而对每辆卡车到这些地方的运费取为零。于是，对象数与活动数一样。用匈牙利法可求得最优分配法方案如下（见表4.3）。所求最小运费为870。大家可以在微机上验证这一结果。

表4.2

卡 车	装 货 地 点				
	A	B	C	D	E
1	300	290	280	290	210
2	250	310	290	300	200
3	180	190	300	190	180
4	320	180	190	200	170
5	270	210	190	250	160
6	190	200	220	190	140
7	220	300	230	180	160
8	260	190	260	210	180

表4.3

卡 车	装运地点	卡 车	装运地点
1	未指定	5	C
2	未指定	6	E
3	A	7	D
4	B	8	未指定

第五章　网络模型

图论是近二十年来发展很快、应用较广的一个数学分支，它可以用来处理某些离散的数学模型。因此，它在物理、化学、生物、计算机科学、社会科学及经济管理等各方面都有广泛的应用。并且与数学的其他分支也有着密切的联系。在这一章，我们将介绍图论的一些基本概念和在实际问题中的常用算法。

5.1　基本概念

图论中所说的图与一般所说的几何图形和代数函数的图形是完全不同的。它是指一些点的集合 V 和联结其中某些点对的线构成的集合 E 所组成的图形。例如，一个大企业的各管理部门、生产车间，我们可以用点来表示；如果两部门之间有直属的领导关系，则表示这两个部门的点就用线联结起来，从而构成一个图。又如，若干个篮球队要进行循环比赛，我们可以用点表示球队，连线表示两队要进行比赛，也可以构成一个图。显然，对于这样的图，点的位置和线的曲直是无关紧要的，我们只关心点的多少，以及这些线是联结哪些点的。一般地，有下面的定义。

定义 5.1.1　图 G 是由 p 个元素的非空有限集 $V(G)$ 和 $V(G)$ 中给定的 q 个无序对构成的集合 $E(G)$ 所组成的二元组，记作 $G=(V(G),E(G))$，或 $G=(V,E)$。

V 的元素称为点（或顶点），E 的元素称为边（或弧）。通常，我们可以用图形来表示一个图。

例 1　图 $G=(V,E)$，其中 $V=(v_1,v_2,v_3,v_4,v_5)$，$E=(e_1,e_2,e_3,e_4,e_5,e_6,e_7,e_8,e_9)$，如图 5.1 所示。

由定义 5.1.1 可以看出，图只是反映了某些对象间的特定关系。我们用图的形式给以刻画，只不过使得这些对象及其关系显得更加直观和清晰，也便于我们深入

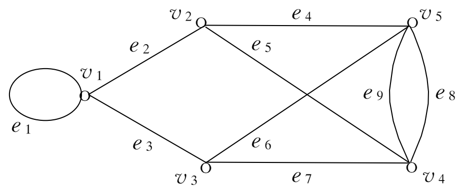


图 5.1

地研究图的某些性质。

以后，我们记：$p=|V(G)|$，表示 G 的顶点个数；$q=|E(G)|$，表示 G 的边数。其中，$p>0$，$q\geqslant 0$，且均为整数。

在图 $G=(V,E)$ 中，如果边 $e=(u,v)\in E(G)$，则称 u 和 v 是 e 的端点，也称 e 是 u（或 v）的关联边。

如果 u、$v\in V(G)$，并且 $(u,v)\in E(G)$，则称点 u 和 v 是相邻的。与同一顶点相关联的边 e_1、e_2 称为邻接的边。如图 5.1 中，v_1，v_2 是相邻的点，e_2、e_3 是邻接的边。

对于边 $e\in E(G)$，如果它的两个端点相同，则称 e 为环。如果联结图中两点的边不止一条，就把这些边称为多重边。如图 5.1 中，边 e_1 是环，e_8，e_9 是多重边，无环、无多重边的图称为简单图。

应当注意的是，一个图的图形表示不是惟一的。如图 5.2 中，几个图所表示的实际上是同一个图 G。

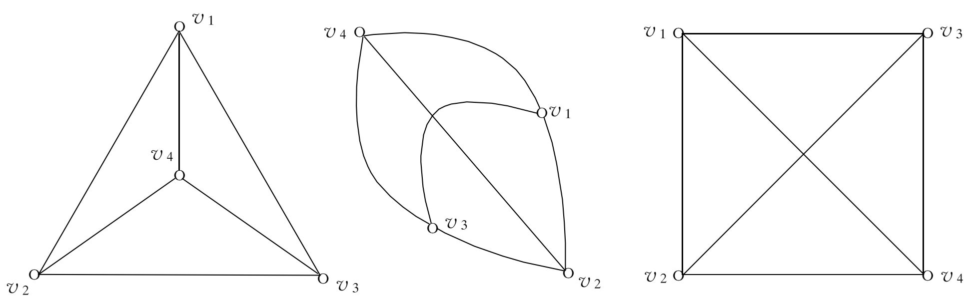
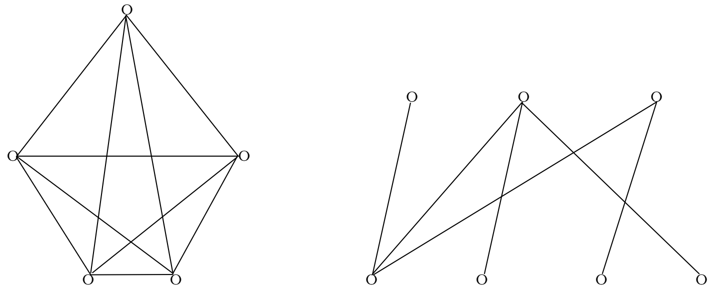


图 5.2

如果图中没有任何边，就称这个图为空图。如果一个简单图 G 中的各对顶点间都有一条边相连，就称图 G 是完全图。n 个顶点的完全图记作 K_n（它有 C_n^2 条边）。

如果在图 $G=(V,E)$ 中，V 可以剖分成两个集合 X、Y，满足 $V=X\cup Y$，$X\cap Y=\varnothing$，使得 G 的每一条边的一个端点在 X 中，另一端点在 Y 中，就称图 G 为二分图（见图 5.3）。



完全图 K_5　　　　二分图

图 5.3

定义 5.1.2　**设** $G_1=(V_1,E_1)$，$G_2=(V_2,E_2)$。**如果** $V_1\subset V_2$，$E_1\subset E_2$，**则称** G_1 **是** G_2 **的子图，记作** $G_1\subset G_2$。

特别，当 $V_1=V_2$，$E_1\subset E_2$ **时，称** G_1 **是** G_2 **的支撑子图，见图** 5.4。

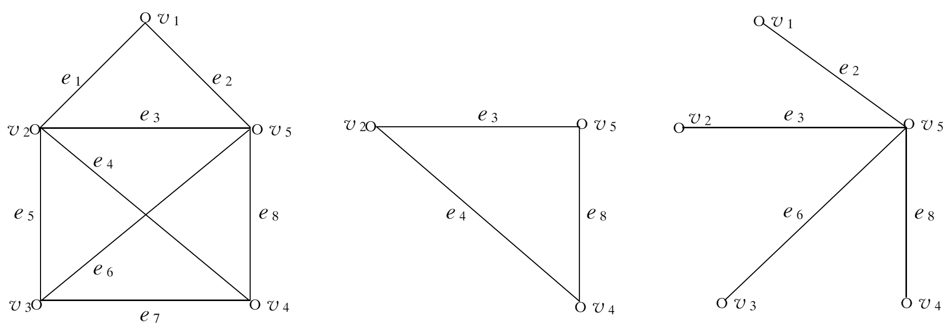


图 G　　　　子图 G'　　　　支撑子图

图 5.4

在图 $G=(V,E)$ **中**，$v\in V(G)$，**我们把与** v **关联的边的数目称为点** v **的次（每个环算作两条边），记作** $d(v)$。$\delta=\min\{d(v)\mid v\in V\}$ **称为** G **的最小次**，$\triangle=\max\{d(v)\mid v\in V\}$ **称为** G **的最大次。我们又将次为奇数的顶点称为奇顶点；反之，称为偶顶点。**

定理 5.1.1　$\sum\limits_{v\in V}d(v)=2q$。

证明　**因图中每条边均与两个顶点相关联，故各点的次之和等于边数的两倍。**

推论　**在任何图中奇次点的个数是偶数。**

我们除了用图形来表示一个图外，还可以用矩阵来表示一个图。在图 $G=(V,E)$

中，$V = \{v_1, \cdots, v_p\}$，$E = \{e_1, \cdots, e_q\}$。如果用 m_{ij} 表示点 v_i 与边 e_j 相关联的次数(0、1 或 2)，则矩阵 $M = (m_{ij})_{p \times q}$ 称为图 G 的关联矩阵。

例 2　若图 G 的图形如图 5.5 所示，则 G 的关联矩阵如图 5.6 所示。

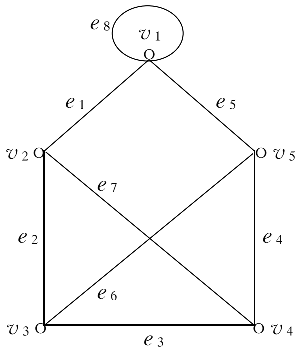


图 5.5

	e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8
v_1	1	0	0	0	1	0	0	2
v_2	1	1	0	0	0	0	1	0
v_3	0	1	1	0	0	1	0	0
v_4	0	0	1	1	0	0	1	0
v_5	0	0	0	1	1	1	0	0

图 5.6

如果用 a_{ij} 来表示联结点 v_i 和点 v_j 的边的数目，则矩阵 $A = (a_{ij})_{p \times p}$ 称为图 G 的邻接矩阵。例如，图 5.5 的图 $G = (V, E)$，其邻接矩阵见图 5.7。

一般说来，图 G 的邻接矩阵比它的关联矩阵的阶数要小得多。所以，通常用邻接矩阵来表示。

	v_1	v_2	v_3	v_4	v_5
v_1	1	1	0	0	1
v_2	1	0	1	1	0
v_3	0	1	0	1	1
v_4	0	1	1	0	1
v_5	1	0	1	1	0

图 5.7

5.2　最短通路问题

最短通路问题是图论中的一个基本问题，它在通讯、石油管道铺设、公路网等实际问题中有着广泛的应用。

例 2　某公司使用一种设备，这种设备在一定年限内随着时间推移逐渐损坏。所以，保留这种设备的时间越长，每年的维修费用就越大。现在，我们假设该公司在

一年开始时必须购置一台这种设备。为简单起见，假设计划使用这台设备的时间为 5 年，估计这台设备的购置费用和维修费用（单位：万元）如表 5.1 所示：

表 5.1

第一年到第五年的购买价格						不同使用年龄的设备的维修费					
年号	1	2	3	4	5	使用年限	0～1	1～2	2～3	3～4	4～5
价格	20	20	22	22	23	维修费	5	7	12	18	25

这家公司希望确定应在哪一年购买一台新设备，使得维修费和新设备的购置费的总和最小。

解　这个问题可以化为图的问题：考虑六个点 $v_1, \cdots, v_6$，其中 v_i（$i = 1, \cdots, 5$）表示在第 i 年初要购买新设备。v_6 是虚设点，表示在第五年的年底才购置新设备。再从点 v_i（$i = 1, \cdots, 5$）引出向点 $v_{i+1}, \cdots, v_6$ 的弧，弧（v_i, v_j）表示第 i 年年初购进的新设备，一直要使用到第 j 年（$j = 2, \cdots, 6$）的年初。弧（v_i, v_j）上所赋的权为第 i 年的购置费加上从第 i 年初使用到第 j 年初这段时机器的维修总费用。例如，弧（v_1，v_4）上所赋的权 $W(v_1, v_4) = 20 + (5 + 7 + 12) = 44$（万元），等等。这样，我们就可以得到一个赋权有向图（图 5.8）。于是，问题就变成：要在这个赋权有向图中，求一条总权最小的 v_1 至 v_6 的单向通路的问题，这样的问题就是最短通路问题。

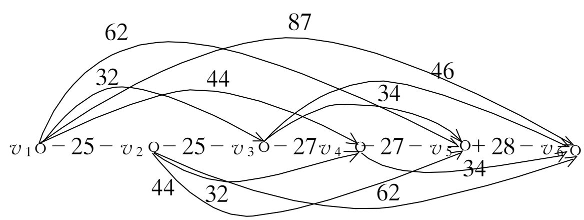


图 5.8

下面，我们介绍求解这一问题的算法。这个算法是 Dijkstra（1959）提出的。它不但求出了图中指定的出发点到终点的最短通路，而且也求出了从出发点到图中其他各点的最短通路。

设在赋权图 $D = (V, A)$ 中，所有弧上的权均非负。如果 $(v_i, v_j) \in A$，则令权 $W(v_i, v_j) = +\infty$。我们可以把权 $W(a_{ij})$ 看作弧 a_{ij} 的长度。现在，要求一条单向 $v_1 - v_p$ 通路 P，使得 $\sum\limits_{a \in P} W(a)$ 最小。

为了简便，我们将起点 v_1 到点 v_k 的最短通路的长记为 $d(v_1, v_k)$ 或 d_k。

Dijkstra 算法的基本思路是很简单的：设 S 是 V 的真子集，并且 $v_1 \in S$。记 $\widetilde{S} =$

$V \setminus S$. 如果 $P = v_1 v_2 \cdots v_k v$ 是从起点 v_1 到 $\widetilde{S}$ 的最短通路，我们用 $d(v_1, \widetilde{S})$ 来表示此通路的长，则显然有 $v_k \in S$，$v \in \widetilde{S}$。并且 P 上的通路 $v_1, \cdots, v_k$ 必然是最短的 $v_1 - v_k$ 通路（图 5.9）。

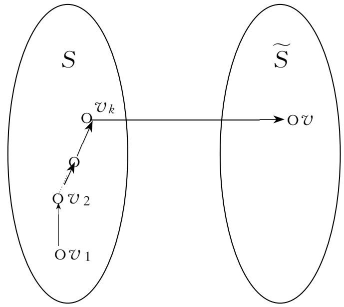


图 5.9

所以

$$d(v_1, v) = d(v_1, v_k) + W(v_k, v)$$

$$d(v_1, \widetilde{S}) = \min\{d(v_1, u) + W(u, v) \mid u \in S, v \in \widetilde{S}\}$$

根据上式，我们提出以下的标号算法：

(1) 开始给顶点 v_1 标上（永久）标号 $d_1 = 0$，其他各点标上临时标号 $d_k(1) = W(1k)$，$(k = 2, \cdots, p)$，令 $S_1 = \{v_1\}$，$\widetilde{S}_1 = \{v_2, \cdots, v_p\}$。如果设 $\min\{W(a_{1j})\} = W(a_{12})$，则 v_2 是与 v_1 距离最近的点，即已求出了一条最短 $v_1 - v_2$ 通路，其长记为 d_2。这时，v_2 获得永久标号 d_2。

令
$$S_2 = \{v_1, v_2\}$$

$$\widetilde{S}_2 = \{v_3, \cdots, v_p\} = \widetilde{S}_1 \setminus \{v_2\}$$

下面我们总是用 S_m 记所有得到永久标号的顶点集，用 $\widetilde{S}_m$ 记所有得到临时标号的顶点集。

(2) 一般地，已有 $S_m \subset V$，并且，$v_1 \in S_m$，$\widetilde{S}_m = V \setminus S_m$。在 $\widetilde{S}_m$ 中求一点 v_k，使得 $d_k = d_k^{(m)} = \min\{d_j^{(m)} \mid v_j \in \widetilde{S}_m\}$。若 $d_k = +\infty$，则说明不存在 v_1 到 v_k 的最短通路。否则，令 $S_{m+1} = S_m \cup \{v_k\}$，$\widetilde{S}_{m+1} = \widetilde{S}_m \setminus \{v_k\}$。

(3) 再修改临时标号：对 $\widetilde{S}_{m+1}$ 中的每一个点 v_j，令 $d_j^{(m+1)} = \min\{d_j^{(m)}, d_k + W_{kj}\}$，其中，$d_k$ 是在 (2) 中已经求出的。再返回步骤 (2)，直到 $m = p - 1$ 为止。

例 3　**就图** 5.10 **给出从** v_1 **到图中各结点的最短通路。**

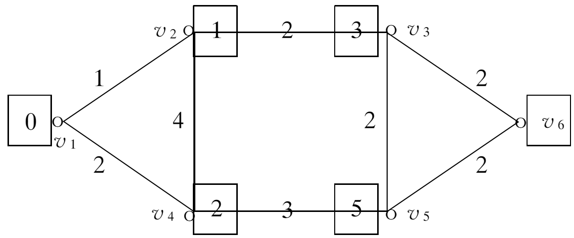


图 5.10

解　(1) **给结点** v_1 **以永久性标号** 0,**并给结点** 2、4 **以临时性标号**(**距离**;**从哪个结点**):**结点** 2(1;1),**结点** 4(2;1);

(2) **给结点** v_2 **以永久性标号** 1,**并给结点** v_3 **以临时性标号**:**结点** 3(3;2)(**因** $2 < 1 + 4$,**结点** 4 **的临时性标号不变**);

(3) **给结点** v_4 **以永久性标号** 2,**并给结点** v_5 **以临时性标号**:**结点** 5(5;4);

(4) **给结点** v_3 **以永久性标号** 3,**并给结点** v_6 **以临时性标号**:**结点** 6(5;3)(**因** $1 + 2 + 2 = 2 + 3 = 5$,**结点** 5 **的临时性标号不变**);

(5) **给结点** v_5 **以永久性标号** 5,**不必更新结点** v_6 **的临时性标号**;

(6) **给结点** v_6 **以永久性标号** 5,**过程结束。**

这样得到从结点 v_1 **到图中各结点的最短通路**:

结点	**距离**	**最短通路**
2	1	$v_1 - v_2$
3	3	$v_1 - v_2 - v_3$
4	2	$v_1 - v_4$
5	5	$v_1 - v_4 - v_5$
6	5	$v_1 - v_2 - v_3 - v_6$

下面,我们就前面讲到的实例,用 Dijkstra **算法,求出最短通路,给出问题的答案。为此,我们将图** 5.6 **重新改画如图** 5.11,**以便于观察。**

给结点 1 **记永久性标记** 0。**结点** 2、3、4、5、6 **分别得到临时性标记** 25、32、44、62、87。

给结点 2 **以永久性标记** 25。**从结点** 1 **出发,通过结点** 2,**分别到达结点** 3、4、5、6、**的"路程"都较已经有的临时性标记大。故不需要更改现有的临时性标记。**

给结点 3 **以永久性标记** 32。**此时由于**:

$d(1,3) + d(3,6) = 32 + 46 = 78 < 87$ (**结点** 6 **已有的临时性标记**),

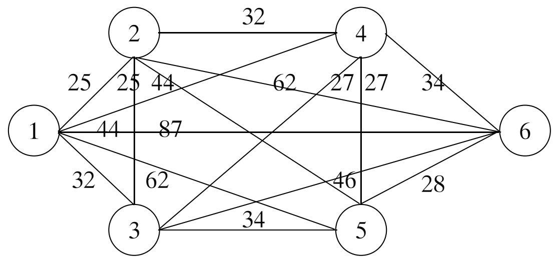


图 5.11

应当更改结点 6 的临时性标记为 78。其余结点的临时性标记不变。

给结点 4 以永久性标记 44。通过比较无需更改其余结点的现有临时性标记（注意，此时也有 $d(1,4)+d(4,6)=44+34=78$）。

给结点 5 以永久性标记 62。通过比较无需更改其余结点的现有临时性标记。

给结点 6 以永久性标记 78。问题解完。

答案如下：

结点	距离	从结点 1 出发的最短通路
2	25	1—2
3	32	1—3
4	44	1—4
5	62	1—5
6	78	1—3—6（或 1—4—6）

所以，应在第三年或第四年购买一台新设备为宜。

5.3 最小生成树问题

定义 5.3.1 不含圈的连通图称为树。一个无圈的图，如果它的每一支都是树，则称为森林。

定理 5.3.1 如果 G 是一棵树，则

(1) G 的任意两个顶点间有惟一的一条通路；

(2) $q=p-1$；

(3) 当 $p\geqslant 2$ 时，G 至少有两个次数为 1 的顶点；

(4) 去掉 G 的任意一条边，则 G 不连通；

(5) 不相邻的两顶点间添加一条边，可以得到一个且仅得一个圈。

证明 设点 $u,v\in V$。如果有两条不同的 $u-v$ 通路 P_1、P_2，且 $P_1\neq P_2$。显然，G 中已有一个圈，矛盾。其余请读者自证。

如果 G 的支撑子图 T 是一棵树，则 T 称为图 G 的支撑树（生成树）（图 5.4）。显然，任何连通图都有支撑树，并且可能有许多不同的支撑树。

下面我们来讨论连接问题 如果要在若干城镇之间架设电话线，使任何两个城镇之间都可以直接通话。设城镇 i、j 间架设电话费用是 W_{ij}。这时，就需要确定一个总费用最小的架设方案。

我们把城镇作为点，连接两点 v_i、v_j 的边赋以权 $W_{ij}>0$。这样，问题就转化为要在赋权图 G 中，求一棵支撑树，使其总权最小。这棵树称为最小支撑树，或最小生成树，简称最小树。这就是所谓的连接问题。Kruskal（1959）提出了解决这一问题的算法，步骤如下：

(1) 取 $e_1\in E(G)$，使得 $W(e_1)=\min\{W(e)\mid e\in E(G)\}$；

(2) 一般，如果已经选好边 $e_1,e_2,\cdots,e_k$，则从边集合 $E_k=E\setminus\{e_1,\cdots,e_k\}$ 中，选取 e_{k+1}，使得满足：

①$e_1,e_2,\cdots,e_k,e_{k+1}$ 所组成的图不含圈；

②$W(e_{k+1})=\min\{W(e)\mid e\in E_{k+1}\}$。直至选取的边数为 $p-1$ 条为止。

例 4 图 5.12 给出了一个实例。

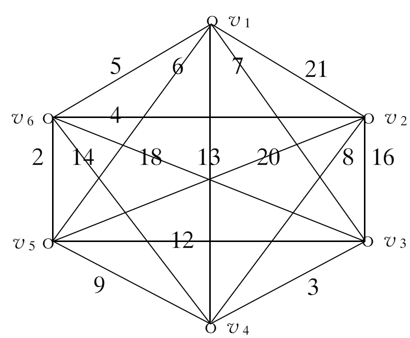


图 5.12

解 软件操作结果如下：

(1) 选 $v_1\in T$，并取与 v_1 联结的最短边 $5=(v_1,v_6)\in T$，$\{v_1,v_6\}\subset T$。

(2) 选取到 $\{v_1,v_6\}$ 的最短边 $2=(v_6,v_5)$，$(v_6,v_5)\in T$，$\{v_1,v_6,v_5\}\subset T$。

(3) 选取到 $\{v_1,v_6,v_5\}$ 的最短边，$4=(v_6,v_2)$，$(v_6,v_2)\in T$，$\{v_1,v_6,v_5,v_2\}\subset T$。

(4) 选取到 $\{v_1,v_6,v_5,v_2\}$ 的最短边 $7=(v_1,v_3)$，$(v_1,v_3)\in T$，$\{v_1,v_6,v_5,

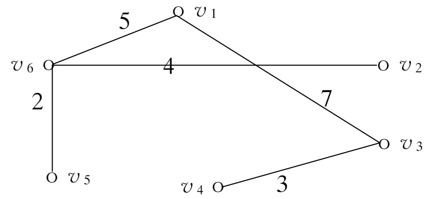


图 5.13

$v_2, v_3\} \subset T$。

⑸ 选取到 $\{v_1, v_6, v_5, v_2, v_3\}$ 的最短边 $3 = (v_3, v_4)$，$(v_3, v_4) \in T$，$\{v_1, v_6, v_5, v_2, v_3, v_4\} \subset T$，整个工作完成。树长 21。

5.4 最大流问题

下面考虑一个有向赋权图 $G = (V, E)$。对 $(v_i, v_j) \in E$，有方向从 v_i 到 v_j，并赋予两个参数：w_{ij} 为该弧的弧容量，f_{ij} 为该弧的物流量。我们按 (w_{ij}, f_{ij}) 的形式写在图中每条边的旁边，得到一个网络图。

我们称满足以下条件的流 f 为可行流：

1. 弧流量限制条件：$0 \leqslant f_{ij} \leqslant w_{ij} \qquad (v_i, v_j) \in E$

2. 平衡条件 $$\sum f_{ij} - \sum f_{ji} = \begin{cases} v(f) & i = s \\ 0 & i \neq s, t \\ -v(f) & i = t \end{cases}$$

式中 $v(f)$ 称为该可行流的流量，即发点的净输入量或收点的净输出量。v_s 为发点，v_t 为收点。

网络中，可行流总是存在的（如，所有 $f_{ij} = 0$ 的流）。我们是要研究如何寻求最大流的问题。网络最大流问题是在网络 G 中，求一个可行流 $f = \{f_{ij}\}$，使其流量 $v(f)$ 达到最大，其数学模型是：

$$\max v(f)$$
$$\text{s.t.} \quad 0 \leqslant f_{ij} \leqslant w_{ij} \qquad (v_i, v_j) \in E$$
$$\sum f_{ij} - \sum f_{ji} = \begin{cases} v(f) & i = s \\ 0 & i \neq s, t \\ -v(f) & i = t \end{cases}$$

这是一个特殊的线性规划问题，自然可用单纯形法求解。但利用图的特点，我们

介绍 Ford − Fulkerson 标号法。

在网络 G 中,设 P 是条从 v_s 到 v_t 的通路,在 P 上并与 P 的方向一致的弧称为前向弧,在 P 上并与 P 的方向相反的弧称为后向弧。

设 $\{f_{ij}\}$ 是 G 的一个可行流,如果存在一条从始点 v_s 到终点 v_t 的通路 P,满足:

① 在 P 的所有前向弧上,$f_{ij} < w_{ij}$;② 在 P 的所有后向弧上,$f_{ij} > 0$,则称 P 是条关于流 $\{f_{ij}\}$ 的可扩充路。

对于网络 G 的一个可行流 $\{f_{ij}\}$,如果能找到可扩充路,我们就可以把 $\{f_{ij}\}$ 调整为流值更大的可行流 $\{f_{ij}'\}$。方法如下:

令 $\varepsilon_1 = \min\{w_{ij} - f_{ij} \mid (v_i, v_j)$ 是 P 上的前向弧$\}$

(如果 P 上无前向弧,就取 $\varepsilon_1 = +\infty$)

$\varepsilon_2 = \min\{f_{ij} \mid (v_i, v_j)$ 是 P 上的后向弧$\}$

(如果 P 上无前向弧,就取 $\varepsilon_2 = +\infty$)

取 $\varepsilon = \min\{\varepsilon_1, \varepsilon_2\}$ 为调整量,定义 $\{f_{ij}'\}$ 如下:

$$f_{ij}' = \begin{cases} f_{ij} & \text{若 } (v_i, v_j) \text{ 不在 } P \text{ 上} \\ f_{ij} + \varepsilon & \text{若 } (v_i, v_j) \text{ 是 } P \text{ 的前向弧} \\ f_{ij} - \varepsilon & \text{若 } (v_i, v_j) \text{ 是 } P \text{ 的后向弧} \end{cases}$$

容易看出,$\{f_{ij}'\}$ 仍然是可行流,并且,其流值比 $\{f_{ij}\}$ 的流值大 ε。反复这样做下去,直到网络中不再存在可扩充路,就求得了网络的最大流。

定理5.4.1 设 $\{f_{ij}\}$ 是网络的一个可行流,并且,对于 $\{f_{ij}\}$ 来说,不存在可扩充路,则 $\{f_{ij}\}$ 就是最大流。

Ford − Fulkerson 于 1956 年提出了找可扩充路的标号算法。我们用下面的一个例子来说明。

我们从一个零可行流开始,寻找一条可扩充路 P。

(1) 给起始点 v_1 以标号 $(0, +)$,成为已标号但尚未检查的点。

(2) 按下面的方法对网络中的各点进行检查和标号。设 v_i 是已标号而未检查的点,则:

① 所有以 v_i 为起点的弧 (v_i, v_j),如果 $f_{ij} < w_{ij}$,并且 v_j 未标号,则给 v_j 以标号 $(i, +)$。v_j 成为已标号未检查的点。

② 以 v_i 为终点的所有弧 (v_k, v_i),若 $f_{ki} > 0$,而 v_k 未标号,则给 v_k 以标号 $(i, -)$。v_k 成为已标号未检查的点。

进行完以上两步后,v_i 改为已检查的点,在标号下面画一横线。

(3) 检查终点是否已得到标号。如果未标号,则重复进行 (2);如果已得到标号,就转入步骤 (4)。

(4) 用“倒向追踪”的方法找出从起点 v_s 到终点 v_t 的可扩充路 P 设 v_t 的标号是 $(i,+)$，则可扩充路 P 上 v_t 的前面的点就是 v_i，弧是 (v_i,v_t)。一般地，考虑 v_i 如果 v_i 的标号是 $(j,+)$，则 v_i 前面的点是 v_j，弧是前向弧 (v_j,v_i) 如果 v_i 的标号是 $(j,-)$，则 v_i 前面的点是 v_j，弧是后向弧 (v_i,v_j)。如此进行，直到追踪到出现标号 $(0,+)$ 的起点 v_s 为止。这样就得到一条从起点 v_s 到终点 v_t 的可扩充路 P，转入步骤 (5)。

(5) 按前面所讲方法进行调整，得到新的可行流 $\{f_{ij}'\}$。抹去所有标号，重新返回步骤 (1)。如果找不到 $\{f_{ij}'\}$ 的可扩充路，则计算终止，已得到最大流。

我们可以在软件 YAJ 上演示这个过程。下面讨论最大流的另外计算方法。

对于网络 $G=(V,E)$，如果将 V 剖分为两个非空集合 S、$\bar{S}$，使得

$$v_s \in S, v_t \in \bar{S}, \text{并且 } V = S \cup \bar{S}, S \cap \bar{S} = \varnothing$$

则所有起点属于 S，而终点属于 $\bar{S}$ 的弧的集合，称为由 S 决定的截集，记作 $(S,\bar{S})$，该截集中所有弧的容量之和，称为这个截集的容量，记作 $C(S,\bar{S})$。其中容量最小的截集称为最小截集。

定理 5.4.2 在任何网络中，最大流的流值等于最小截集的容量。

我们不去证明这个定理，但可以用这个定理对刚才的例子，找出最小截集，验证定理的结论。

调整量 $\varepsilon=2$. 调整后的可行流如图 5.14：

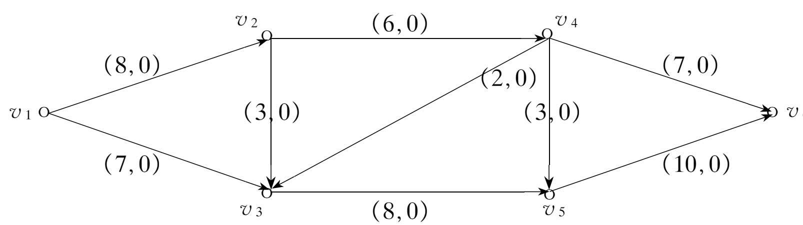


图 5.14

5.5 微机操作 —— 调用 YAJ 中，NET 决策支持系统

5.5.1 观察 NET 决策支持系统

这个程序包含三个所谓纯网络算法：最短通路，最大流，和最小生成树算法。这个程序要求一个包含有由弧所联结的结点的网络，用以表示问题。最短通路算法，找从始点到网络上任一其他点的最短通路；最大流算法，求从源到汇的最大流；最小

生成树算法,求出联结所有结点的弧的总长度。

本程序解具有数十个分枝和结点的网络问题。网络结点从 1 开始,顺序编号。网络数据输入是逐枝进行。对于每一枝的数据,包含始点,终点和这两个点之间的距离,或流入/流出流。在数据输入以后,你就可以解问题,并显示其结果。

5.5.2 问题输入

这个程序可解具有数十个枝/弧和结点的网络问题。选取下列算法中的一个算法:① 最短通路算法,② 最大流算法,③ 最小生成树算法。

(1) 你可以输入枝的名字(不超过 6 个字符),缺损名是 $B_1,\cdots,B_n$。

(2) 从 1 开始,对每个结点顺序编号。

(3) 你可以按任意顺序输入每一枝/弧。

5.5.3 解答实例 —— 分别以前面三类问题的例子做出解答

1. 图 5.10 给出的例子

步骤(1) 永久性标号结点 1,更新下列结点,并标上临时性标号(即距离,结点):
结点 2(1;1),结点 4(2;1)

步骤(2) 永久性标号结点 2,更新下列结点,并标上临时性标号(即距离,结点):
结点 3(3;2)

步骤(3) 永久性标号结点 4,更新下列结点,并标上临时性标号(即距离,结点):
结点 5(5;4)

步骤(4) 永久性标号结点 3,更新下列结点,并标上临时性标号(即距离,结点):
结点 6(5;3)

步骤(5) 永久性标号结点 5,更新下列结点,并标上临时性标号(即距离,结点):
没有

步骤(6) 永久性标号结点 6,更新下列结点,并标上临时性标号(即距离,结点):
没有

最后的最短通路

结点	距离	从结点 1 出发的最短通路
2	1	1—2
3	3	1—2—3
4	2	1—4
5	5	1—4—5
6	5	1—2—3—6

2. 图 5.12 给出的例子

步骤⑴ 连接结点 1,到已连接结点最近的未连接结点是 :结点 6 ;最短距离(从结点 1 到结点 6 具有距离)是 5

步骤⑵ 连接结点 6,到已连接结点最近的未连接结点是 :结点 5 ;最短距离(从结点 6 到结点 5 具有距离)是 2

步骤⑶ 连接结点 5,到已连接结点最近的未连接结点是 :结点 2 ;最短距离(从结点 6 到结点 2 具有距离)是 4

步骤⑷ 连接结点 2,到已连接结点最近的未连接结点是 :结点 3 ;最短距离(从结点 1 到结点 3 具有距离)是 7

步骤⑸ 连接结点 3,到已连接结点最近的未连接结点是 :结点 4 ;最短距离(从结点 3 到结点 4 具有距离)是 3

最后的最短生成树

树上的分枝	距离
1—3	7
1—6	5
3—4	3
6—2	4
6—5	2

整个距离 = 21

3. 可扩充路的标号算法例子(题见第 72 页)

在软件 *YAJ* 上按要求输入,解答结果如下:

最大流算法的详细步骤

步骤 1 流 = 3,由分支 4—5 确定 ;所选择的通路是 :1—2—4—5—6

步骤 2 流 = 3,由分支 2—4 确定 ;所选择的通路是 :1—2—4—6

步骤 3 流 = 7,由分支 1—3 确定 ;所选择的通路是 :1—3—5—6

最后的流

分支	网络流
1—2	6
1—3	7
2—4	6
3—5	7
4—5	3
4—6	3
5—6	10

整个最大流 = 13

第六章　关键路线法

在现代化的工业生产、农业生产、科学研究等各项活动中，如何合理地进行组织计划，统筹安排，使预定的任务能顺利完成，是一件十分重要的事情。一些比较简单的任务，只包含几项活动（工序），只要凭经验或简单的分析，就可以得到比较合理的安排。但是，一项大的工程项目，往往包括成千上万个活动，这些活动之间又有着相互依赖的关系，要对这样的工程进行合理的组织、调度，就不是单凭经验所能解决的了。

我们将本章的关键路线法及下一章的计划评审技术，按照华罗庚教授的称呼，统一叫做统筹方法，又称网络分析技术。统筹方法是利用数学来研究、分析一项工作的合理组织与安排的科学方法，它在众多领域有着广泛的应用。

统筹方法的基础是画出统筹图。统筹图是全面反映一项任务的各个活动（工序）之间的相互关系及先后次序的一个网络图。在统筹图上，对任务的各项活动进行分析和计算，有助于管理人员更合理地、科学地进行组织和管理。

6.1　统筹图

将统筹方法应用于某项工程的组织安排时，首先要将这项任务分解成若干个活动（或称为工序），并弄清这些工序之间的先后关系和完成每道工序所需要的时间。这里，工序是指一项有具体内容的、经过一定时间后才能完成的生产过程。然后，用图论的方法，按这些工序的先后关系，及各项工序的完成时间做出一个网络图（赋权有向图）。最后，给图中的所有顶点进行编号，这时得到的图，称统筹图。

6.1.1　任务的分解和分析

把一项任务分解为若干道工序，确定各工序间的相互关系，是一项复杂的调查研究工作。

例 1　某盖房子的工程有 17 道工序，每道工序所需时间及先后关系如表 6.1：

表 6.1

工序	工序内容	紧前工序	时间（天）
A	基础	–	6
B	准备大桥、柱子	–	3
C	木制房顶骨架	A、B	2
D	砌砖	A、B	2
E	下水道工程	A、B	1
F	运输地面混凝土	E	2
G	卫生工程	C、F	3
H	布线工程	C、F	2
I	煤气、自来水工程	C、F	3
J	抹墙壁（灰泥干燥）	G、H、I	10
K	地面加工	J	3
L	洗脸间、厨房瓷砖工程	J	4
M	屋顶与雨棚	D、C、F	2
N	雨水落水管工程	M	1
O	墙壁涂漆	K、L	1
P	电气工程	O	2
Q	庭园修饰工程	N	5

一般地说，一道工序所需要的时间就是完成这一工序的工时定额。若没有工时定额，或工时定额不能确切定出时，可按下面公式给出一个估计（三点估计）：

$$\text{该工序所需时间} = \frac{a + 4m + b}{6}$$

其中，a—— 最快可能完成该工序的工时，

b—— 最慢可能完成该工序的工时，

m—— 最大可能完成该工序的工时。

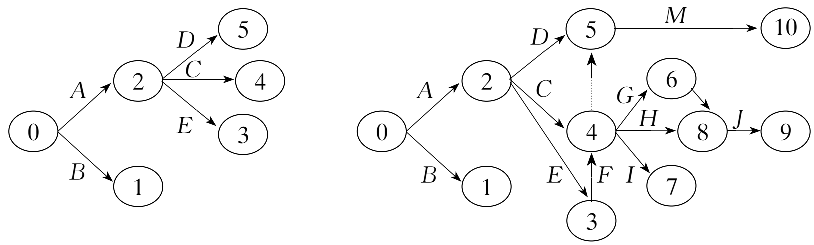
6.1.2　画图

在将一项任务分解成若干工序之后，我们用网络图来表示这一任务。用一条弧表示一道工序，这道工序的开工和完工分别用这条弧的起点和终点来表示。弧的起点和终点称为事件。每个顶点（事件）都给一个编号，并且每条弧起点的编号一定要小于终点的编号，这样的编号称为正规编号。可以证明：在没有有向圈的有向图中，至少存在一个顶点，它不是任何弧的终点。并且，这样的有向图一定有正规编号。

画图时，要根据任务分解表，由第一道工序开始，依先后顺序由左到右画出，直

到最后一道工序为止。

下面我们继续讨论前面的例子。由表 6.1 知，A、B 是平行作业，故要虚设一结点为工程的起点 0。A、B 同是 C、D、E 的紧前工序，如图 6.1 (a)。继续按表往下画，F 是 E 的紧后工序，且 C、F 同是 G、H、I 的紧前工序，J 又是这三道工序的紧后工序，所以这里也要设置虚工序，同时又知，D、C、F 同是 M 的紧前工序，这里也要用虚工序。此时，有图 6.1 (b)。继续按表往下作图，并将每一工序时间写出，按时间先后、结点序号进行调整，得到图 6.2。



(a)　　　　(b)

图 6.1

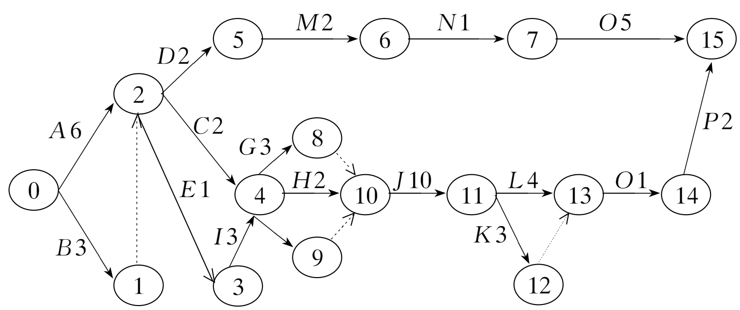


图 6.2

注意 要正确地画出统筹图，必须注意以下几点：

(1) 在统筹图上不能出现有向圈。

(2) 一条弧以及与它关联的顶点 (事件) 只能代表一道工序。

(3) 用好虚工序 为了准确地表达工序之间的先后顺序而增加的工序，其完成时间为零。

(4) 平行作业 为了加快一个工序的完成，往往可以将一个工序分解成两个或更多的工序同时进行。在画图时也可以用虚工序表示这一过程，如图 6.3。

(5) 交叉作业 为加快完成一项任务的进度，我们经常在一道工序未完成时，就

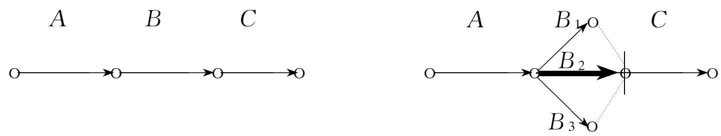


图 6.3

开始紧后工序的工作,而不必等待前一工序全部完成。在画图时要利用虚工序和增加顶点的方法。例如 $A=A_1+A_2+A_3$,$B=B_1+B_2+B_3$,$C=C_1+C_2+C_3$,进行交叉作业,则可画出如下网络图 (图 6.4):

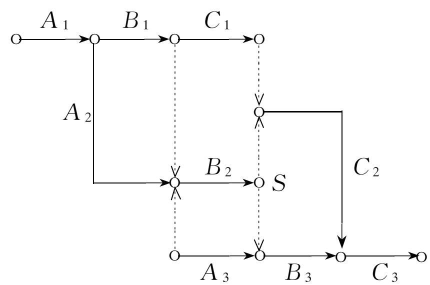


图 6.4

而不能画成:

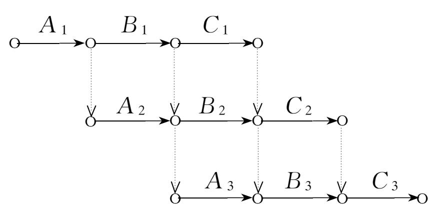


图 6.5

否则,A_3 的紧前工序不但有 A_2,而且还有 B_1;B_3 的紧前工序不但有 A_2、B_2,而且还有 C_1。这显然不符合实际情况。

(6) 外单位协作的工序要明确标在所需工序上。

(7) 先画草图,然后加以整理。比如,在画好的统筹图中,如果不是任何弧的终点的顶点不止一个,则可以把它们合成一个顶点,它是整个图的起点。同时,把不是任何弧的起点 (如不止一个) 合成一个顶点,它是整个图的终点。

(8) 对活动数量大的网络图可以分块绘制,通常可以按工程部位、日历时间或箭

线位置分块。

6.2 关键路线

网络图的作出，这只是计划工作的第一步。往往由于计划的原因造成生产上的不协调，经常发生窝工或浪费劳力、资金等现象，这就要求计划工作必须准确地估算出工程每道工序所用的时间、费用 —— 简称工序时间、工序费用等，以及确定工程从开始到结束总的可利用时间和费用，从而确定在有限资源情况下，如何安排时间、劳力、资金，使得此项工程按时完工或提前完工。另外，有时因外界因素或者一些偶然事件，经常需要考虑某些工序是否可以推延完成时间，但又不至于影响总工程的进度。这一系列问题，就是计划工作中需要进一步研究的重要环节。

我们先从下面一个简单的网络图的分析入手。

图 6.6 中，结点 1 是起点，结点 6 是终点，计 9 道工序。箭杆上所写数字是每道工序需要的时间。从 1 到 6 共有 7 条通路 (相应的时间记为 T)：

$L1$：$1 \to 2 \to 4 \to 6$　　$T1 = 1 + 2 + 5 = 8$

$L2$：$1 \to 2 \to 4 \to 5 \to 6$　　$T2 = 1 + 2 + 2 + 2 = 7$

$L3$：$1 \to 2 \to 3 \to 4 \to 6$　　$T3 = 1 + 2 + 5 + 5 = 13$

$L4$：$1 \to 2 \to 3 \to 4 \to 5 \to 6$　　$T4 = 1 + 2 + 5 + 2 + 2 = 12$

$L5$：$1 \to 3 \to 4 \to 6$　　$T5 = 5 + 5 + 5 = 15$

$L6$：$1 \to 3 \to 4 \to 5 \to 6$　　$T6 = 5 + 5 + 2 + 2 = 14$

$L7$：$1 \to 3 \to 5 \to 6$　　$T7 = 5 + 3 + 2 = 10$

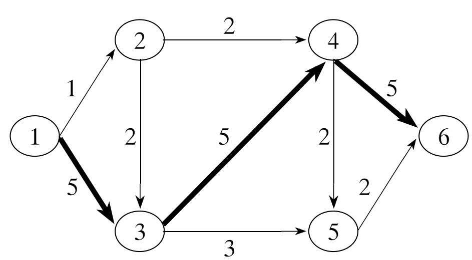


图 6.6

由此可看出，第五条路所用时间最长。如果这条路上所有工序都能按时结束，也就是说，这条路上三道工序都能在五天内完工，则整个工程肯定在 15 天内按时完成。这样，第五条路就成了整个工程的关键路线了。

我们把网络图中时间最长的路线称为关键路线，其他路线称为非关键路线。关键路线上的工序叫做关键工序。

我们可以看出这样的事实，如果关键路线上的所有工序都能按时完成，或提前完成，则总工程一定可以按时完成或提前完成；但若关键路线上哪一道工序延误了时间，则整个工程肯定要推迟完工时间。然而，非关键路线上的非关键工序没有这样的影响（在一定的变化幅度内）。

由此可见，关键路线是总工程的重点。除了在时间上它起到决定性作用外，在降低费用，合理利用人力、物力上也起到重要作用。非关键工序可适当推迟完工时间，也就是可以适当减少人力、物力，而将这些人力、物力加强到关键工序上去，既可以避免生产中的窝工现象，又可节省劳力、减少开支。因此，关键路线的确定在工程管理上具有非常重要的意义。

非关键路线和关键路线在计划工作中也不是一成不变的。因为在调整过程中，非关键路线上的工序若时间长了，就有可能转化为关键工序。所以，两者之间的关系是相对的、可变的。

例 2　找出图 6.2 的关键路线。

它的关键路线有两条：

$0 \to 2 \to 3 \to 4 \to 8 \to 10 \to 11 \to 13 \to 14 \to 15$；

$0 \to 2 \to 3 \to 4 \to 9 \to 10 \to 11 \to 13 \to 14 \to 15$。

长均为 29，所以全工程需要 29 天完工。

6.3　网络参数计算

网络图上仅仅反映出各工序的时间还不够，还必须使管理人员一眼就能看清楚哪些工序什么时间开工、什么时间完工、开工和完工可活动的时间有多少，以便管理人员根据生产进度和资源情况更合理安排和调度生产计划。这就对有关时间的一系列参数进行计算。

1. 工序最早可能开工时间

紧前工序最早完工的时间就是紧后工序最早可能开工时间，记为 $t_{ES}(i,j)$。

2. 工序最早可能完工时间

按工序最早可能开工时间开始，且完成该工序任务所需要的时间称为工序最早可能完工时间，记为 $t_{EF}(i,j)$。有

$$t_{EF}(i,j) = t_{ES}(i,j) + t(i,j)$$

3. 工序最迟开工时间

网络图中非关键工序在不影响紧后工序生产的情况下，可推迟开工时间的上界为工序最迟开工时间。这就是说，若再推迟开工时间的话，就会影响紧后工序按时开工了。将此记为 $t_{LS}(i,j)$。

4. 工序最迟完工时间

按工序最迟开工时间开始，且完成该工序任务所需时间为工序最迟完工时间，记为 $t_{LF}(i,j)$。有

$$t_{LF}(i,j) = t_{LS}(i,j) + t(i,j)$$

5. 工程最早完工时间

网络图中，关键路线上所有工序都是按最早开工时间开工的情况下，总工程所用的时间称为工程最早完工时间，记为 T_E。

6. 最早结点时间

某一结点 (i) 最早可能开工时间的时间称为最早结点时间，记为 $t_E(i)$。这个时间是从工程始点到该结点最长路线各工序的时间和。显然有

$$t_E(0) = 0, t_E(i) = \max_{k<i}\{t_E(k) + t(k,i)\}, i = 1,\cdots,n$$

其中，$t_E(k)$ 是结点 (i) 相邻的前面一个结点 (k) 的最早结点时间。n 是工程终点的标号。在网络图中，为了表示出每个最早结点时间，就将 $t_E(i)$ 写在方框内，填在结点左下方。

7. 最迟结点时间

某一结点 (i) 最迟必须完成的时间称为最迟结点时间，记作 $t_L(i)$。这个时间是从计划完成时间中减去某一结点 (i) 到终点 (n) 的最长时间间隔。所以从终点开始由右向左递推，即可得到所有结点的最迟时间，显然有

$$t_L(n) = t_E(n) = TE$$

$$t_L(i) = \min_{i<k}\{t_L(k) - t(i,k)\}, i = n-1, n-2,\cdots,1,0$$

其中，$t_L(k)$ 是结点 (i) 后面的 (按工序顺序) 结点 k 的最迟时间。

在网络图上，将结点 (i) 的 $t_L(i)$ 写在三角形框内，填在该点右下方。应用前面的公式，对例 1 上各结点算出最早、最迟结点时间 (请读者自己验证) 如图 6.7。

8. 总机动时间

某工序可以推迟开工的时间称为该工序总机动时间，记为 $R(i,j)$。有

$$R(i,j) = t_{LS}(i,j) - t_{ES}(i,j) = t_{LF}(i,j) - t_{EF}(i,j)$$

对于关键工序来说，因 $t_{LS} = t_{ES}, t_{LF} = t_{EF}$，故关键工序总机动时间为零。我们常常用此判定一工序是否为关键工序。

9. 自由机动时间

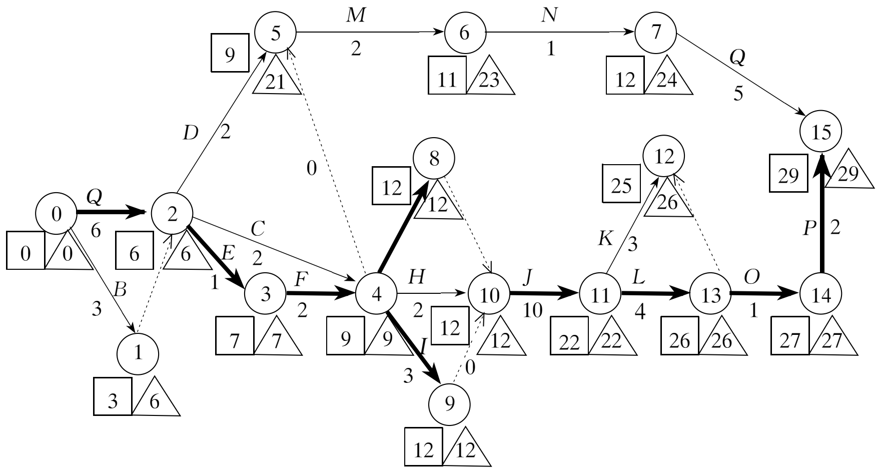


图 6.7

在不影响紧后工序最早开工的前提下，工序最早结束时间可以推迟的时间称为该工序的自由机动时间，记为 $r(i,j)$：

$$r(i,j) = t_{ES}(j,k) - t_{EF}(i,j),\quad i < j < k$$

在网络图上，我们清楚地看到，关键路线上的每个结点有 $t_E(i) = t_L(i)$。而且 $R(i,j) = 0$，$r(i,j) = 0$。

但在非关键工序上，$R(i,j) \neq 0$。这样，我们可以在非关键工序上减少人力、物力，推迟工序时间，减少费用支出，增加到关键工序上去，当然，推迟时间不能超过 $R(i,j)$。

在实际工作中，人们有多种办法对网络计划进行优化。

6.4 突击施工分析

我们的软件 YAJ 中，提供的 CPM 决策支持系统可以帮助我们进行突击施工分析。根据需要，在可能条件下，可以适当压缩工程的完工时间，出现突击施工期限。按这个新的期限完工的线路可能不止一条。我们要求列举出所有这样的关键线路，以及算出由于突击施工所增加的突击施工费用。

要想做到这一点，需要事先给出每一个工序的有关参数：正常施工期限 a、突击施工期限 b、正常施工费用 c、突击施工费用 d。显然，条件 $b < a$，$d > c$ 是必要的。我们可以用表格或每边赋有四个权数的赋权图来记录这一事实。就图 6.2，我们再增添几个参数如表 6.2：

表 6.2

工序名	始点	终点	正常施工期限	突击施工期限	正常施工费用	突击施工费用
$A1$	1	2	1	0.5	1	1.5
$A2$	1	3	5	3	5	10
$A3$	2	3	2	1	2	3
$A4$	2	4	2	1.5	2	2.5
$A5$	3	4	5	4	5	8
$A6$	3	5	3	2	3	4
$A7$	4	5	2	1	2	3
$A8$	4	6	5	2	5	12
$A9$	5	6	2	1.5	2	2.5

前面已经求出了没有突击施工条件下的关键线路：$1 \to 3 \to 4 \to 6$；关键工序 $A2$、$A5$、$A8$；完工时间 15。容易求出整个工程正常施工费用：

$$1 + 5 + 2 + 2 + 5 + 3 + 2 + 5 + 2 = 27$$

现在，我们研究如何进行突击施工分析。比如，根据需要，欲压缩工期四天，问当如何安排？

先观察上表，发现对关键路线上的关键工序可作这样的压缩：$A2$ 减少 2 天；$A5$ 减少 1 天；$A8$ 减少 1 天（不是通过 $A2$ 减少 2 天，$A8$ 减少 2 天来完成，虽然那样将可减少些突击施工费用，但又要涉及到工序 $A7$ 或 $A9$）。新的施工期限分别为 3 天，4 天，4 天。增加突击施工费用分别为 5，3，及 2.3333。关键路线完工时间为 11 天，总的费用为 37.3333。而且，在 11 天完工的路线应该有四条：

$1 \to 2 \to 3 \to 4 \to 5 \to 6$，即工序 $A1$，$A3$，$A5$，$A7$，$A9$；

$1 \to 3 \to 4 \to 5 \to 6$，即工序 $A2$，$A5$，$A7$，$A9$；

$1 \to 2 \to 3 \to 4 \to 6$，即工序 $A1$，$A3$，$A5$，$A8$；

$1 \to 3 \to 4 \to 6$，即工序 $A2$，$A5$，$A8$。

它们都是关键路线。

再比如，要压缩工期 5 天。除在原来关键工序 $A2$，$A5$ 上分别减少 2 天、1 天外，只有对工序 $A8$ 减少 2 天了（即使要较多地增加施工费用）。另外，为了保证整个工程 10 天完成，对前面已经求出的第一、第二道关键工序上的关键工序 $A7$ 必须压缩 1 天。这样，就需要增加突击施工费用：

$$\left(\frac{10-5}{2}\right) \times 2 + \left(\frac{8-5}{1}\right) \times 1 + \left(\frac{12-5}{3}\right) \times 2 + \left(\frac{3-2}{1}\right) \times 1$$
$$= 5 + 3 + 4.6666 + 1 = 13.6666$$

总的施工费用就是 $27 + 13.6666 = 40.6666$，整个工程 10 天完工。

关键工序仍然是前面所列出的四条不变。

6.5 微机操作 —— 调用 *YAJ* 中,*CPM* 决策支持系统

6.5.1 观察 *CPM* 决策支持系统

这个程序使用关键路线方法(*CPM*)给出数百个活动的设计。每个活动假定有持续的活动时间和费用。输入和修改数据都很容易。而且,问题的数据可以存储到磁盘上,并从上面读出。这个程序假定,每个网络使用“活动 – 弧”格式。

在一个设计方案做出后,你不仅可以显示最早开始时间(*ES*),最迟开始时间(*LS*),最早完成时间(*EF*)和最迟完成时间(*LF*),而且,也可以显示每个活动的松弛时间和临界状态。此外,如果工程有多重评审通路,这个程序也将显示网络的近 10 个通路。

这个程序允许你定义活动的名字(不超过 6 个字符)。缺损的名字是 *A*1,*A*2,…,*An*。注意,像结点一样,对活动也必须从 1 开始,按照领先关系顺序编号。

6.5.2 问题输入

这个 *CPM* 程序,可以解决数百个活动的安排设计。

在输入问题时,请按照下列约定:

⑴ 活动名字可以包含到 6 个字符。

⑵ 持续时间,连同小数点,可以包含 6 个阿拉伯数字。

⑶ 费用,连同小数点,可以包含 6 个阿拉伯数字。

⑷ 如果你没有 *crash* 时间和费用,就输入空格。

⑸ 开始结点编号,必须小于结束结点编号。

6.5.3 问题求解

当解一个问题时,你可以选择显示 *CPM* 方法的所有中间结果。如果你已经输入了 *crash* 时间和 *crash* 费用,你也可以实施 *crash* 分析。

如果你的工程网络有多重评审通路,这个程序可以显示到 10 条通路。下面我们对图 6.6 给出的例子进行微机操作,并按表 6.2 的数据进行突击施工分析。在 *YAJ* 上按要求输入数据,给出解答:

1. **不进行突击施工分析的结果：**

工序号	工序名	最早开始	最晚开始	最早完成	最晚完成	松弛时间 $LE-ES$
1	$A1$	0	2.0000	1.0000	3.0000	2.0000
2	$A2$	0	0	5.0000	5.0000	**关键工序**
3	$A3$	1.0000	3.0000	3.0000	5.0000	2.0000
4	$A4$	1.0000	8.0000	3.0000	10.000	7.0000
5	$A5$	5.0000	5.0000	10.000	10.000	**关键工序**
6	$A6$	5.0000	10.000	8.0000	13.000	5.0000
7	$A7$	10.000	11.000	12.000	13.000	1.0000
8	$A8$	10.000	10.000	15.000	15.000	**关键工序**
9	$A9$	12.000	13.000	14.000	15.000	1.0000
完成时间 = 15,**整个费用** = 27						

$$CP\#1:\ 1 \xRightarrow{(A2)} 3 \xRightarrow{(A5)} 4 \xRightarrow{(A8)} 6$$

2. **进行突击施工分析**

(1) **压缩工期**：4 **天。**

(2) **突击施工执行情况：**

突击施工工序 $A2$,**压缩时间单位** 2,**新的期限** 3,**增加费用** 5,

突击施工工序 $A5$,**压缩时间单位** 1,**新的期限** 4,**增加费用** 3,

突击施工工序 $A8$,**压缩时间单位** 4,**新的期限** 4,**增加费用** 2.333 333。

(3) **完成时间** = 11,**整个费用** = 37.333 333。

(4) **关键线路：**

$$CP\#1:\ 1 \xRightarrow{(A1)} 2 \xRightarrow{(A3)} 3 \xRightarrow{(A5)} 4 \xRightarrow{(A7)} 5 \xRightarrow{(A9)} 6$$

$$CP\#2:\ 1 \xRightarrow{(A2)} 3 \xRightarrow{(A5)} 4 \xRightarrow{(A7)} 5 \xRightarrow{(A9)} 6$$

$$CP\#3:\ 1 \xRightarrow{(A1)} 2 \xRightarrow{(A3)} 3 \xRightarrow{(A5)} 4 \xRightarrow{(A8)} 6$$

$$CP\#4:\ 1 \xRightarrow{(A2)} 3 \xRightarrow{(A5)} 4 \xRightarrow{(A8)} 6$$

第七章　计划评审技术

任何一项大的任务是否能完成，与该任务各阶段的计划、调度和控制的质量密切相关。一般来说，除非运用某种计划和协调方法，阶段数不可能非常多，否则就会在管理上失去控制。有一种运筹方法是计划评审技术（Program Evaluation and Review Technique，PERT），或称计划评审法。它有助于在大型任务中管理和控制人力、材料、设备和时间的利用。PERT 可以用来找出任务中的关键部分，做出必要的调整，为的是能按期完成任务的进度计划。它最适用于包含很多不确定因素的大规模的研究和开发任务。

计划评审技术是在 1958—1959 年发展起来的，当时美国海军北极星导弹的计划采用了这种方法，使之比原计划期限提前了 18 ~ 24 个月。从 1959 年起，计划评审技术几乎已成功地应用于各种大企业中。

关键路线法（Critical Path Method，CPM）是另一种计划和协调办法。它在建筑工业中得到发展。因为这两种方法是建立在网络模型的基础上，所以又统称为网络分析技术。

我国在 20 世纪 60 年代初期便开始推广和应用网络分析技术，取得了较好的效果。1965 年，著名数学家华罗庚教授在推广这项技术时，称其为“统筹法”，即统筹安排的意思。实践证明，统筹法是一个十分有效的科学管理方法。

7.1　计划评审技术网络

PERT 分析的第一阶段是要确定在整个任务完成以前必须先做完的工作（或工序），每一工序需要的时间和资源。资源可以是人、钱、材料、设备或空间。很明显，并不是所有工序都能同时完成的，某些工序未完成以前，有一些工序是不能开始的。

一项任务的各工序之间的关系可以用网络来表示，网络由点和箭头组成。箭头

表示工序：箭头尾部表示某工序开始，而箭头前端表示该工序已结束；点表示工序的开始和完成；小圆圈表示的点称为事件（event）。事件表示所有进入该点的工序为已完成，而所有离开该点的工序为刚开始。事件并不需要时间和资源，它们只表示某些工序完成而其余工序开始的时刻。当所有到达该事件的工序都已完成，则称事件已完成。

在某些给定事件未完成以前，从该事件出发的工序是不能开始的。某工序在事件 i 开始而在事件 j 结束，则用工序 (i,j) 表示。

图 7.1 中，事件 1 表示任务开始，而事件 6 表示任务完成。工序 (1,3) 和工序 (2,3) 都完成以后，事件 3 才完成。同样，事件 3 完成以前，工序 (3,6) 不能开始。当工序 (5,6)、工序 (3,6)、工序 (4,6) 全部完成时，最终事件 6 才算完成。

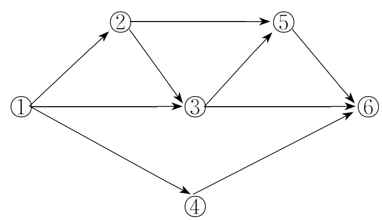


图 7.1

网络表示各工序间的相互关系，构成网络的规则如下：

① 事件 1 表示事件开始，从事件 1 出发的工序前面没有其他先行工序；

② 事件 M 表示事件的完成，M 是事件的最大数；

③ 工序 (i,j) 在事件 i 开始，在事件 j 结束；

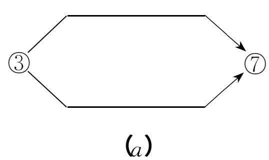
④ 对每一工序 (i,j) 来说，$i < j$；

⑤ 对于固定一个 j 值来说，在事件 j 完成以前，所有工序 (i,j) 必须完成；

⑥ 每个工序 (i,j) 必须是惟一的。

规则 ⑥ 的含意如下：如果两个或两个以上的工序从事件 i 开始到事件 j 结束，则我们需要虚设一些事件和时间为零的工序。例如，在图 7.2 中，有两个不同的工序 a 和 b，要在事件 3 开始，在事件 7 结束。(a) 图所示是错误的。而要引入虚设事件 6 和虚设工序 (6,7)，如 (b) 图所示。虚箭头表示虚设工序。虚设事件 6 表示工序 a 完成，而事件 7 表示工序 a 和 b 都完成。

根据规则 ⑥，假设 a,b,c,d 是必须完成的 4 个工序，工序 a 和 b 必须在工序 c 之前完成，但只有工序 b 要在工序 d 之前完成。这可以用图 7.3 表示，加一条虚线表示虚设工作。



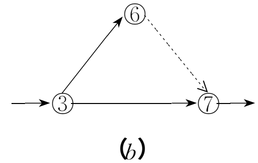


图 7.2

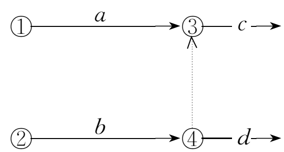


图 7.3

7.2 工序的时间估计

必须先有 $PRET$ 网络中各工序的时间估计，才能使网络成为有用的计划和协调方法。通常有如下三种估计：

a—— 最乐观时间（Optimistic Time），每件事进行得比预期的顺利，而没有延迟的期望时间。

m—— 最可能时间（Most Likely Time），实际最可能的完成时间。

b—— 最悲观时间（Pessimistic Time），差不多每件事都出差错的期望时间。

根据这三个估计时间 a，m，b，每个工序的平均完成时间，即期望完成时间估计为

$$ET = \frac{a + 4m + b}{6}$$

$$\left(= \frac{1}{2}\left(\frac{a + 2m}{3} + \frac{2m + b}{3}\right) \text{— 华罗庚解释}\right)$$

与此类似，每个工序完成的方差可估计为

$$\sigma_{ET}{}^2 = \left(\frac{b - a}{6}\right)^2$$

我们可以在工序网络图上，将这些数据写在该工序箭头旁边。

7.3 网络图的时间参数及计算方法

7.3.1 事件的最早时间(Earliest Time,TE)

算出各工序的期望完成时间以后,可用来计算各事件的期望完成时间。对每个事件来说,从起始事件到该事件可能经过的每条路线上各工序所需时间之和,其最大总和即为该事件最早期望完成时间。实际上,我们可以采用如下算法:假设

$ET(i,j)$—— 工序 (i,j) 的期望完成时间

$TE(j)$—— 事件 j 的最早期望完成时间

则对某固定 j 值,有:

$$TE(j) = \max_{i}[TE(i) + ET(i,j)]$$

$$TE(1) = 0$$

这里,(i,j) 为到达 j 所有工序。每一事件的 TE 时间,表示事件能完成的最早时间。其条件为 通向该事件的每条路线上的各工序 (i,j) 能准确地在 $ET(i,j)$ 时间内完成。计算最终事件的 TE 值的路线称为网络的关键路线,关键路线上的工序时间总和大于通过网络的任何其他路线的工序时间总和。

7.3.2 事件的最迟时间(Lastest Time,TL)

标题内的参数,是在假设所有后续事件都能按期完成的条件下,不延误总任务完工期限的最迟允许的某事件发生时间。为了计算事件的 TL 值,可沿着给定事件和最终事件之间的各条路线,从任务的计划完成时间中减去累计的各工序时间,所得结果中的最小一个就是该事件的 LF 值。即,

$TL(i)$—— 事件 i 的最迟时间,对某个固定 i 值,可得

$$TL(i) = \min_{j}[TL(j) - ET(i,j)]$$

这里,(i,j) 是从事件 i 出发的所有工序。这个公式是个递推公式,可以计算 $TL(i)$,$i = M,M-1,M-2,\cdots,2,1$。而取 $TL(M) = TE(M)$,或 $TL(M) = SD$—— 任务的计划完成时间。

7.3.3 工序最早开始时间(Earliest Start,ES)

一个工序必须等它之前工序结束后才能开始,在这之前是不具备开始条件的,这个时间就叫做工序的最早开始时间或称最早开工期。工序的最早开始时间等于箭

尾事件的最早时间：

$$ES(i,j) = TE(i)$$

7.3.4 工序最早完成时间（Earliest Finish，EF）

工序的最早完成时间，或称最早完工期，就是该工序的最早开始时间加上本工序所需要的时间：

$$EF(i,j) = ES(i,j) + ET(i,j)$$

7.3.5 工序最迟完成时间（Latest Finish，LF）

工序的最迟完成时间，或称最迟完工期，等于该工序箭头事件的最迟时间：

$$LF(i,j) = TL(j)$$

7.3.6 工序最迟开始时间（Latest Start，LS）

一项工序，有一个或几个紧后工序。为了不影响紧后工序如期开工，该工序必须有一最迟必须开始的时间，这个时间就叫做工序的最迟开始时间或称最迟开工期。工序最迟开始时间等于该工序的最迟结束时间减去该工序时间的差：

$$LS(i,j) = LF(i,j) - ET(i,j)$$

7.3.7 工序的总时差——$S(i,j)$

工序的总时差，是指在不影响整个任务的完工期的前提下，一项工序的完工期可以推迟多长时间。其计算是：

$$S(i,j) = LS(i,j) - ES(i,j)$$

或

$$S(i,j) = LF(i,j) - EF(i,j)$$

我们称总时差为零的工序（事件）为关键工序（事件）；总时差为零的路线为关键路线，在这条路线上总周期最长。

7.3.8 工序的单时差——$s(i,j)$

工序的单时差，等于紧后工序的最早开工时间减去本工序的最早完成时间，或等于箭头事件的最早时间减去本工序的最早完成时间：

$$s(i,j) = ES(j,k) - EF(i,j)$$

或

$$s(i,j) = TE(j) - EF(i,j)$$

7.3.9 事件的时差 —— $S(i)$

$$S(i) = TL(i) - TE(i)$$

计算和利用时差，是网络分析的一个重要问题。时差又常常叫做松弛时间（Slack Time）。下面，我们给出一个例子（见表 7.1）：

表 7.1

	工序									
6→7	时间	1→2	1→3	1→4	2→5	3→4	3→6	4→5	4→7	5→7
a	2.20	1.51	5.00	6.00	2.00	5.51	1.76	4.00	5.00	2.51
m	2.45	6.12	9.90	7.50	5.00	10.21	4.06	8.16	10.38	5.61
b	6.00	10.00	15.39	12.00	8.00	14.00	6.00	11.35	13.49	11.00

其相应的网络图（见图 7.4）尽可能地给出了参数：

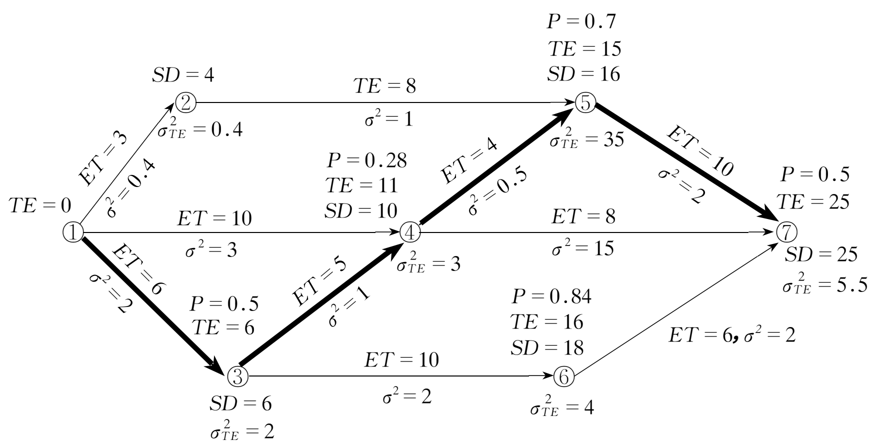


图 7.4

7.4 关键路线

通过 $PERT$ 网络的最长路线称为关键路线（Critical Path），沿着这条路线可得最终事件的 TE 值。

取 $TL(M) = SD(M)$ —— 任务的计划完成时间，作为一个参考点。相应的松弛时间为：

$$S(M) = TL(M) - TE(M) = SD(M) - TE(M)$$

如果希望最终事件严格如期完成，则 $S(M) = 0$。关键路线上各事件的松弛时间等于最终事件的松弛时间，将为 0。

$PERT$ 分析的一个主要目的是突出关键路线，可以将系统外的资源或非关键路线上的各工序的资源提供给关键路线上各工序。与此相关的是，第一项任务的实际工作开始后，$PERT$ 网络可以根据所有最新的信息每周、每月或根据需要进行更新。这些信息可以包括计划日期，现有的资源，工序时间估计，实际完成时间，等等。

需要注意，工序 (i,j) 是关键路线上的工序，它必须满足三个条件：

①$TL(i) - TE(i) = S(M)$，

②$TL(j) - TE(j) = S(M)$，

③$TE(j) - TE(i) = TL(j) - TL(i) = ET(i,j)$。

这里，$S(M)$ 是最终事件的松弛时间。

图 7.4 中的关键路线用宽箭头表示。其上每一个工序满足上面三个条件，结果是：

关键路线为 ①—③—④—⑤—⑦ 完工周期为 25。

7.5 事件按时完成的概率

$PERT$ 网络中规定了一个（或一个以上）事件的计划完成时间，特别是最终事件的计划完成时间几乎总是给定的。应用计划完成时间，以及每个工序完成时间的平均值和方差可以决定事件按期完成的概率。惟一必要的假设是：每事件完成时间服从正态分布其平均值为 TE，方差为 σ_{TE}^2，σ_{TE}^2 为各工序时间的方差之和，这些工序都处在用以计算 TE 的那条路线上。

例如，在图 7.4 中，

$$TE(4) = ET(1,3) + ET(3,4) = 6 + 5 = 11$$

于是

$$\sigma_{TE}^2(4) = \sigma_{(1,3)}^2 + \sigma_{(3,4)}^2 = 2 + 1 = 3$$

值得注意的是，$\sigma_{(i,j)}^2$ 是工序 (i,j) 完成时间的方差。

对 $PERT$ 网络中较早的事件作正态分布的假设可能不太恰当，但是对较迟的事件上述假设是十分恰当的。较迟事件的完成时间是几个工序完成时间之和，其中每个都有确定的平均值和方差，根据中心极限定理，事件的完成时间趋于正态分布。

假设：T—— 某事件完成时间，

TE——T 的期望值，

$\sigma_{TE}{}^2$—— T 的方差，

SD—— 该事件预期完成时间。

则
$$T \sim N(TE,\sigma_{TE}^2)$$
可以证明
$$Z = \frac{T - TE}{\sigma_{TE}} \sim N(0,1)$$
如果我们要计算某事件按期或提前完成的概率,此概率为：
$$P(T \leqslant SD) = \int_{-\infty}^{SD} N(TE,\sigma_{TE}^2)\,\mathrm{d}t = \int_{-\infty}^{\frac{SD-TE}{\sigma_{TE}}} N(0,1)\,\mathrm{d}t$$
例如,在图 7.4 中,我们决定任务按期或提前完成的概率,与此类似,事件 5 [$SD(5) = 16$] 按期或提前完成的概率为：
$$P[T \leqslant SD(5)] = P(T \leqslant 16) = \int_{-\infty}^{16} N(15,3.5)\,\mathrm{d}t = \int_{-\infty}^{\frac{16-15}{\sqrt{3.5}}} N(0,1)\,\mathrm{d}t = 0.702$$

7.6 微机操作 —— 调用 YAJ 中,PERT 决策支持系统

7.6.1 观察 PERT 决策支持系统

这个程序解决上百个活动的工程网络，使用的方法是“计划评审技术”(PERT)。活动持续时间假定是 β 分布,具有三个时间估计：乐观的、可能的和悲观的。对于输入和修改数据,程序提供了一个方便的格式。PERT 假定每个网络都用“弧上记录活动”的格式。

在你的网络被解出以后,你可以选择显示最早开始时间 (ES),最晚开始时间 (LS),最早完成时间 (EF) 和最晚完成时间 (LF)。如果工程网络有多重关键线路,PERT 将显示网络的多到 10 条关键线路。此外,你还可以实施概率分析。

程序的概率计算,假定诸活动是无关的,而且,所有线路也是无关的。还假定你的网络拥有足够大的活动数目,以致可以使用正态分布。因此,当活动不是无关时,或者活动数目不是很大时,下面分析可能会高度偏倚。

7.6.2 输入问题

PERT 程序可以分析数百个活动的网络。

(1) 活动名可以包含到 6 个字符。

(2) 持续时间可以包含 6 个阿拉伯数字(连同小数位)。

(3) 始点编号必须小于终点编号。

其他约定同前 1.6.2。

7.6.3 求解问题

当解一个问题时,你可以显示 *PERT* 方法的中间结果。如果你的工程网络有预先的特殊计划,你也可实施概率分析。当你的工程网络有多重关键线路时,*PERT* 程序可显示到 10 条关键线路。

我们以表 7.1 所给数据进行微机操作。

在 *YAJ* 上,按要求输入后,显示结果如下:

工序		工序		最早开始	最晚开始	最早完成	最晚完成	松弛
号	名	期望时间	方差					*LS* − *ES*
1	*A*1	3.0000	0.4011	0	3.9984	3.0000	6.9983	3.9983
2	*A*2	5.9983	2.0022	0	0	5.9983	5.9983	关键工序
3	*A*3	9.9983	2.9987	0	1.0000	9.9983	10.998	1.0000
4	*A*4	8.0000	1.0000	3.0000	6.9983	11.000	14.998	3.9983
5	*A*5	5.0000	1.0000	5.9983	5.9983	10.998	10.998	关键工序
6	*A*6	10.058	2.0022	5.9983	8.9433	16.057	19.002	2.9450
7	*A*7	4.0000	0.4994	10.998	10.998	14.998	14.998	关键工序
8	*A*8	7.9983	1.5006	10.998	17.002	18.997	25.000	6.0033
9	*A*9	10.002	2.0022	14.998	14.998	25.000	25.000	关键工序
10	*A*10	5.9983	2.0022	16.057	19.002	22.055	25.000	2.9450
整个工程期望完成时间 = 25								

关键线路:*CP* #1 (具有方差 5.503 828):

1⇒3⇒4⇒5⇒7

下面再进行概率分析。

1. 什么是你的计划时间?25

在 *CP* #1 上:方差 = 5.503 828,标准差 = 2.346 024,

在 25 天完成概率是 0.5,全部工程完工概率,在 25 天是 0.5。

2. 什么是你的计划时间?28

在 *CP* #1 上:方差 = 5.503 828,标准差 = 2.346 024,

在 28 天完成概率是 0.899 519 1,全部工程完工概率,在 28 天是 0.899 519 1。

3. 什么是你的计划时间?22

在 *CP* #1 上:方差 = 5.503 828,标准差 = 2.346 024,

在 22 天完成概率是 0.100 480 9,全部工程完工概率,在 22 天是 0.100 480 9。

第八章　动态规划

动态规划是用于许多类型的数学方法。在一些领域中动态规划已用来解决各种问题，例如，资源分配、货物装运、设备更新、排序、生产计划和存贮等。然而，动态规划只是解决问题的一种“方法”，而不是用来解决这些问题的一种算法。因此，对每种类型的问题需要单独的（动态规划）算法。

动态规划法涉及多级决策的最优化。它把已知问题分为许多级（阶段）或许多子问题，然后，按顺序解决子问题，直到最后解决初始的问题。这种方法将问题按时间分为阶段，决策随时间阶段而变动，“动态”的含义就寓于其中了。动态规划法的核心是 Bellman 所提出的最优化原理，即：最优策略具有这样的性质，即不论其初始状态和初始决策如何，其余的决策必须按照第一个决策所造成的状态构成最优策略，这是一个十分重要的概念。虽然，可以应用动态规划确定大量各种各样问题的最优解，但并不是用这一方法求解所有问题时都是有效的方法。要确定什么情况下应用动态规划为宜，经验和创造性是指导性的因素。

在这一章，我们重点讲解三个问题：最短路线问题、背包问题、生产和存储问题。

8.1　最短路线问题

最短路线问题（Stagecoach Problem）即为公共马车问题。

图 8.1 为一线路网络，其中有 10 个结点，①、②、…、⑩。要从 ① 点铺设一条管道到 ⑩ 点，弧上数字表示两点间的距离（公里）。要求选择一条从 ① 到 ⑩ 的铺管路线，使总距离最短。

如何求解这个问题？最容易联想到的方法是穷举法：把整个线路前后综合一起考虑，列出所有可能的路线。这里一共有 16 条不同的路线，比较得结果是

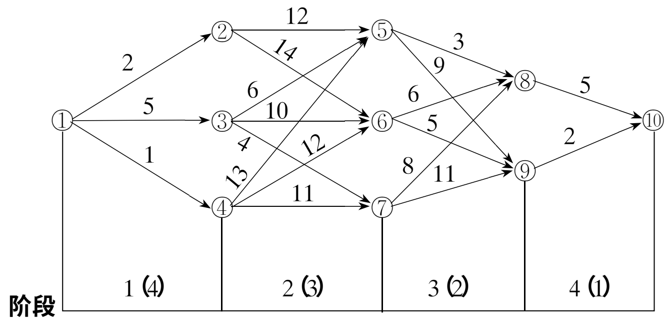


图 8.1

①——③——⑤——⑧——⑩，这条路线的距离最短，其距离为 19 公里。

这个方法尽管求出了结果，但却不可取。当网络很复杂时，几乎办不到。动态规划就是针对这类问题，为减少计算工作量而提出的一种新方法。应用于此例，求解方法如下：

首先将整个网络分成四个阶段，仍如图 8.1 所示：顺序编号，逆序思考。我们从最右边阶段做起。称 ⑧、⑨ 为本阶段的输入结点，⑩ 为输出结点。这个阶段结果为

表 8.1

输入结点	决策线路	到终点的最短距离
8	(8,10)	5
9	(9,10)	2

考虑阶段 3。从 ⑤ 出发，有两条路线到阶段 4 的输入结点。由

$$\min\{(5,8)+(8,10),(5,9)+(9,10)\}=\min\{3+5,9+2\}=8=(5,8)+(8,10)$$

这里，(i,j) 也表示弧长，应取决策线路 ⑤——⑧。对 ⑥、⑦ 进行同样的计算，得到表 8.2。

表 8.2

输入结点	决策线路	输出结点	到终点的最短距离
5	(5,8)	8	8
6	(6,9)	9	7
7	(7,9)	9	12

考虑阶段 2。有三个输入结点，三个输出结点。比如，考虑 ③，由

$$\min\{(3,5)+(5,8),(3,6)+(6,9),(3,7)+(7,9)\}=(3,5)+(5,8)=6+8=14,$$

当取决策线路 ③——⑤。对 ②、④ 进行同样的计算，得到表 8.3。

表 8.3

输入结点	决策线路	输出结点	到终点的最短距离
2	(2,5)	5	20
3	(3,5)	5	14
4	(4,6)	6	19

大家或许已经注重到，当计算一个阶段中任一输入结点到终点的最短距离时，都用到了前阶段中算出的最优结果，这样做大大节省了工作量。

现在，再考虑阶段 1。容易求出 ① 到 ⑩ 的最短距离为：

$$\min\{2+20, 5+14, 1+19\} = 19$$

这个阶段的决策线路为 ①——③。

最后，纵观整个网络，就得到从始点到终点的最短路径（见表 8.4）：

表 8.4

阶段	决策线路
1（4）	(1,3)
2（3）	(3,5)
3（2）	(5,8)
4（1）	(8,10)

最优线路为 ①——③——⑤——⑧——⑩，距离为 $5+6+3+5=19$ 公里。

最短路线问题优化原理：如果某一结点是整个网络最优化路线中的一点，则该点到终点的最短路径也在网络的最优化路径中。此原理容易用反证法证明，前面指出的计算特点，就是基于这个优化原理。

8.2 动态规划中的几个名词和符号

上一节已经讲过，求最短线路，不是从整个网络上看，而是从将问题分成几个阶段，每一阶段形成一个子问题，所有子问题解决了，整个问题也就解决了。至于每个子问题要注意哪些问题，我们以阶段 3（2）为例，作进一步分析。

最短路径在阶 3（2）为 (5,8)，但在求得最优解之前，起始点 ⑤ 是不知道的，也可能是 ⑥ 点或 ⑦ 点。这是在求解过程中的一个变量，称为这个阶段的输入变量，用 x_3 表示，它的取值范围是 ⑤、⑥、⑦。然而，最优路径经过 ⑤ 点不是在阶段 3（2）中，而是在阶段 2（3）中决定的，也就是 x_3 的解不是阶段 3 中所求的内容。因此阶段 3 中的子

问题只好分别假定 ⑤、⑥、⑦ 是输入结点，找出它们的输出结点 —— 要利用阶段 4 的结果。

当假定 ⑤ 是阶段 3 的输入结点时，由它走向阶段 4 的路径，可能是 (5,8)，也可能是 (5,9)，这又是阶段 3 要解决的一个变量，称为阶段 3 的决策变量，用 d_3 表示。求解 d_3 的决策标准是 ⑤ 点到终点的路径距离最短，要利用阶段 4 的结果。求得 d_3 后，自然就得到了这个阶段的输出结点 ⑧。阶段 3 的输出也相应有一个变量，称为阶段 3 的输出变量，用 x_4 表示。类似的，可以求出 ⑥ 和 ⑦ 的决策变量和输出变量。

用以上符号，阶段 3 的子问题可以表示为图 8.2。

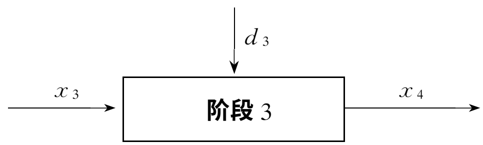


图 8.2

阶段 3 的输出变量 x_4 也正是阶段 4 的输入变量。

将上述求解过程推广到一般，对有 n 个阶段的动态规划问题，令，

x_k = 阶段 k 的输入（阶段 $k-1$ 的输出）变量，

d_k = 阶段 k 的决策变量。

则求解过程为：

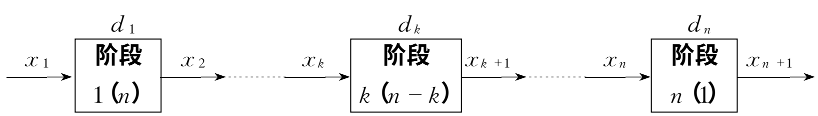


图 8.3

在整个路线中，输入输出变量 x_k 是重要的因素，它们将 n 个阶段偶合一起。在任何阶段 k，我们要知道输入变量 x_k，从而选出最好的决策变量 d_k。输出变量 x_{k+1} 取决于输入变量和决策变量：

$$x_{k+1} = f_k(x_k, d_k)$$

我们叫它是阶段变换函数，又叫这个阶段某输入结点到某输出结点的距离为该阶段的指标函数值，它取决于输入变量和决策变量。在不同的问题中，指标函数有不同的具体含义，如它可以表示为一个阶段的支付或者价值。它一般表示为 $r_k = (x_k, d_k)$。

考虑了阶段的变换函数和指标函数，n 个阶段的动态规划问题可以用下面的形式表达其变换关系：

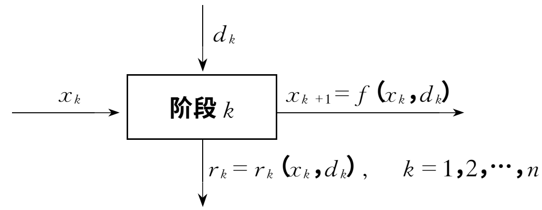


图 8.4

每个阶段有两个输入：状态变量和决策变量。也有两个输出：一个新的状态变量 $f(x_k,d_k)$ 和一个指标函数 $r_k(x_k,d_k)$。

若对阶段 k 输入一个 x_k，则可以求得该阶段和已讨论阶段的最优总指标函数，以符号 $g_k(x_k)$ 表示。例如，前面例题中，有：

$$g_3(x_3=\text{结点⑤})=8,g_3(x_3=\text{结点⑥})=7,\cdots$$

8.3　背包问题

背包问题(Knapsack Problem)是动态规划中经常遇到的一类问题，它的原始意义是：旅行者携带的背包中所能装的物品重量有一定的限度，它是已知的。旅行者要从 n 种物品中挑选若干数量装入背包，一方面总重量不超过背包所能携带的限度，另一方面又要使所携带物品的总价值最大。问题是应当挑选哪几种物品，各物品的份量又是多少？假定每件物品的单位重量和单位价格都是已知的。

例 1　某工厂要做来年的生产计划，准备在四种产品中挑选一定的数量进行生产，情况如表 8.5 所示。

表 8.5

产品	待加工的数量	单位产品加工时间(月)	单位产品价格(元)
第一种	5	1	1500
第二种	3	3	5000
第三种	2	4	7000
第四种	3	5	8000

这里，所挑选的产品，一方面要能在一年内加工完毕，另一方面要使总生产价格最大。可供生产的总时间是一年的 12 个月，相当于“背包”的“总容量”，单位产品的加工时间相当于这种物品的“单位重量”，所要生产的产品总的加工时间不能超过

“背包的总容量”12 个月。所以该问题可归结为背包问题来解。

首先将问题划分为四个阶段。在第 k 阶段决定生产第 k 种产品的件数，$k=1,2,3,4$。

阶段 4　考虑所有可能的输入状态，具体见表 8.6：

表 8.6

x_4	最优 d_4	x_5	$g_4(x_4)=r_4(x_4,d_4)$
0	0	0	0
1	0	1	0
2	0	2	0
3	0	3	0
4	0	4	0
5	1	0	8000
6	1	1	8000
7	1	2	8000
8	1	3	8000
9	1	4	8000
10	2	0	16 000
11	2	1	16 000
12	2	2	16 000

阶段 3　考虑所有可能的输入状态，列表如下：

表 8.7

x_3	最优 d_3	x_4	$g_3(x_3)=r_3(x_3,d_3)+g_4(x_4)$
0	0	0	0
1	0	1	0
2	0	2	0
3	0	3	0
4	1	0	7000
5	0	5	8000
6	0	6	8000
7	0	7	8000
8	2	0	14 000
9	1	5	15 000
10	0	10	16 000
11	0	11	16 000
12	0	12	16 000

阶段 2　考虑所有可能的输入状态，见表 8.8：

表 8.8

x_2	最优 d_2	x_3	$g_2(x_2) = r_2(x_2, d_2) + g_3(x_3)$
0	0	0	0
1	0	1	0
2	0	2	0
3	1	0	5000
4	0	4	7000
5	0	5	8000
6	2	0	10 000
7	1	4	12 000
8	0	8	14 000
9	0	9	15 000
10	2	4	17 000
11	1	8	19 000
12	1	9	20 000

阶段 1　因为计划内总计有 12 个月，$x_1 = 12$，d_1 的取值范围为 0，1，2，3，4，5。最优决策为 $d_1 = 1$。即：

表 8.9

x_1	最优 d_1	x_2	$g_1(x_1) = r_1(x_1, d_1) + g_2(x_2)$
12	1	11	20 500

最后结果为：

表 8.10

产品	生产件数	收益	剩余生产时间
1	1	1500	11
2	1	5000	8
3	2	14 000	0
4	0	0	0

另　若当年只能安排 10 个月生产，又当如何安排？—— 只在第四阶段考虑即可，这实际是最优解的灵敏度计算（请大家给出答案）。

8.4 生产和库存问题

某工厂生产一种产品，经过对市场情况的调查了解后，预测了今后三个月内可能出售的产品数量，同时也估算了厂内三个月对该产品的生产能力、存储容量、单位产品的生产成本和库存费用，如表 8.11：

表 8.11

月份	可销售量	生产能力	存储容量	单位产品生产成本	单位产品存储费
1	2	3	2	800	150
2	3	2	3	700	150
3	3	3	2	1000	200

问如何安排生产计划，在厂内生产和库存能力许可的条件下，既满足市场需求，又能使总的生产和库存总费用最少（在一月份开始生产时已有库存一件）？

用动态规划解决此问题，首先还是将整个问题分为若干阶段。这里，按时间分为三个阶段，每月为一个阶段，仍然逆序考虑。

在第 i 个阶段，我们用 x_i 表示状态变量 —— 库存件数；d_i 表示决策变量 —— 生产件数；$r_i(x_i, d_i)$ 表示指标函数 —— 生产与库存费用之和。

阶段 3　这个阶段是这样一个子问题：

$$\begin{aligned} \min g_3(x_3) &= r_3(x_3, d_3) \\ &= 1000d_3 + 200(x_3 + d_3 - 3) \\ &= 1200d_3 + 200x_3 - 600 \end{aligned}$$

$$s.t.\begin{cases} x_3 + d_3 - 3 \leqslant 2\text{，或 } x_3 + d_3 \leqslant 5 \\ d_3 \leqslant 3 \\ x_3 + d_3 \geqslant 3 \\ d_3 \geqslant 0 \end{cases}$$

我们将其结果列如下表（d^* 表示最优决策）：

表 8.12

x_3	d_3^*	x_4	$g_3(x_3)$
0	3	0	3000
1	2	0	2000
2	1	0	1000
3	0	0	0

阶段 2　解子问题：

$$\begin{aligned}\min z &= g_2(x_2)\\ &= 700d_2 + 150(x_2 + d_2 - 3) + g_3(x_3)\\ &= 850d_2 + 150x_2 - 450 + g_3(x_3)\end{aligned}$$

$$s.t.\begin{cases} x_2 + d_2 - 3 \leqslant 3, \text{或 } x_2 + d_2 \leqslant 6 \\ d_2 \leqslant 2 \\ x_2 + d_2 \geqslant 3 \\ d_2 \geqslant 0 \end{cases}$$

现在，考虑 x_2 的所有可能值得到最优决策 d_2^*，如表 8.13：

表 8.13

	$r_2(x_2,d_2)+g_3(x_3)$					
x_2 \ d_2	0	1	2	d_2^*	$g_2(x_2)$	$x_3 = x_2 + d_2 - 3$
0	–	–	–	–	M	–
1	–	–	4400	2	4400	0
2	–	3700	3550	2	3550	1

表中横线为不可行解；M 是一个很大的价格，因为当输入 $x_2 = 0$ 时无可行解，用一个很大的数 M 是避免 $x_2 = 0$ 发生在最优解中。

阶段 1　目的是根据开始时库存量 $x_1 = 1$，决定阶段的生产量 d_1，并使本阶段的指标函数值 $r_1(x_1,d_1) + g_2(x_2)$ 最小。应解子问题：

$$\min z = 950d_1 + 150x_1 - 300 + g_2(x_2)$$

$$s.t.\begin{cases} x_1 + d_1 \leqslant 4 \\ d_3 \leqslant 3 \\ x_1 + d_1 \geqslant 2 \\ d_1 \geqslant 0 \end{cases}$$

结果如下：

表 8.14

	$r_1(x_1,d_1)+g_2(x_2)$						
d_1 \ x_1	0	1	2	3	d_1^*	$g_1(x_1)$	$x_2=x_1+d_1-2$
1	–	M	6150	6250	2	6150	1

至此，三个子问题已经求解完毕。将三个子问题综合来看，从最后的第一个问题往回反推至第三个子问题，就得到整个问题的解。

已知

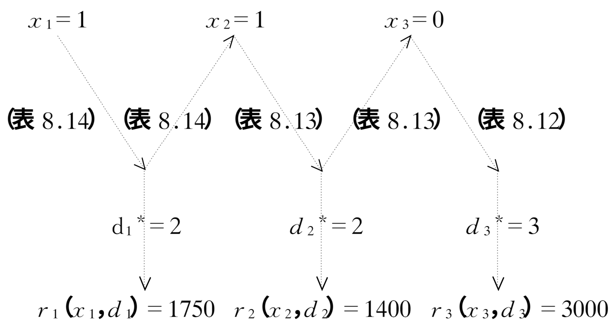


图 8.5

最优总指标函数即与最优解 $d_1^*=2, d_2^*=2, d_3^*=3$ 相对应的总生产和库存费用为：

$$g_1(x_1)=1750+1400+3000=6150\ (\text{元})$$

与表 8.14 的结果相同。

由以上各节的内容可看出，动态规划适用于这样的问题：一个大问题可以划分为几个阶段，每个阶段形成一个子问题。各个阶段是相互联系的，每个阶段都要作出决策，并且一个阶段的决策，常影响下一个阶段的决策，从而影响整个过程的活动路线。各个阶段所确定的决策构成一个决策序列，称为一个策略。由于各个阶段可供选择的决策往往可能不止一个，因而就形成有许多策略可供我们选取，对应于一个就有确定的活动效果，效果是可以用数量来衡量的。这样，策略不同，效果也不同。多阶段决策问题，就是要在允许选择的那些策略中间，选择一个最优策略，使在给定的标准下达到最好的效果。

在动态规划求解过程中，作为整个过程的最优策略具有这样的性质：无论过去

状态和决策如何，对其决策所形成的状态而言，余下的决策必须构成最优策略，这是动态规划的最优化原理。利用这个原理可以把多阶段决策的求解过程看成是一个连续的递推过程，由后向前逐步推算。在求解时，各状态前面的状态和决策，对其后面的子问题而言，只不过相当于其初始条件而已，并不影响后面过程的最优策略。

在动态规划递推过程中，每个阶段的优化方法可按不同的问题而不同。除表格方法外，还有线性规划方法、整数规划中的分支定界法、网络方法（最小通路）、匈牙利法、微积分方法等等，去求解各个子问题。但任何方法，对解一系列子问题比解整个原问题总是要容易。这是动态规划方法的一个特点，也是一个优点。

动态规划求解问题的一个突出弱点是所谓维数问题，即当问题中的变量个数（维数）太大时，由于计算机存储器容量和计算速度的限制而会无法求解。

8.5 微机操作 —— 调用 YAJ 中，DP 决策支持系统

8.5.1 观察 DP 决策支持系统

这个程序包含三个所谓动态规划算法：最短路线问题、背包问题和生产存储控制问题。DP 程序可解的问题，允许拥有数十个阶段，每个阶段有数十个状态。解答过程使用逆序解法：从最后一个阶段（到达终点的阶段）到第一个阶段。

对于最短路线问题，要求给出的数据是：阶段数，每个阶段的状态数，以及逐次阶段上两个状态的距离。对于背包问题，要求给出每个阶段的空间需求量和返回值。对于生产／存贮控制问题，要求给出需求量，单位产品存储费用，单位产品生产费用和每个阶段的组织费用。在这个程序中所使用的阶段变换函数和返回函数，假定是线性的。

8.5.2 输入问题

(1) 回答下列问题，以便确定你的问题的大小。
(2) 联结发点的阶段定义成第一阶段，而联结终点的阶段定义成最后一个阶段。
(3) 下列问题是为配合你的特殊 DP 问题所设计的。
(4) ESC 键，允许你从开始重新进入。
(5) ／键，允许你重新进入前一个问题。
(6) 在你输入一个问题时，函数菜单允许你修改数据。

8.5.3 求解问题

⑴ 最短路线问题

仍以 8.1 的例题为例，显示结果如前面叙述，最后结果如表 8.4。

⑵ 背包问题

仍以 8.3 的例题为例，显示结果几乎与前面讲述内容完全一样。

⑶ 生产和存储问题

仍以 8.4 的例题为例，显示结果如下：

阶段 (时期) 3：

输入存储量	生产	输出存储量	最小总费用
0	3	0	3000
1	2	0	2000
2	1	0	1000
3	0	0	0

阶段 (时期) 2：

输入存储量	生产	输出存储量	最小总费用
0	0	−3	不能用
1	2	0	4400
2	2	1	3550

阶段 (时期) 1：

输入存储量	生产	输出存储量	最小总费用
1	2	1	6150

最后解：

时期(阶段)	开始存储量	生产	开办费	生产费	结束存储量	保管费	总的费用
1	1	2	0	1600	1	150	1750
2	1	2	0	1400	0	0	1400
3	0	3	0	3000	0	0	3000
总费用 = 6150							

第九章　库存论

商店为了满足顾客的需要,必须有适当数量的库存货物来支持经营。一旦某种物品缺货,就要损失营业额;工厂中零备件缺货,会造成停工损失;医院里血库贮血不够就不能供应急救手术。总之,为了保证各类系统的正常运行,需要后备资源,库存物品要不断得到补充。

那么,是否库存量越大越好呢?过高的库存,对于经营是不利的:它占有大量流动资金;需要大量的行政管理和库存费用;长期保管的物品会变质、过期失效等,也要造成经济损失。所以需要有一个科学的方法来处理何时补充库存以及补充多少的问题,这就是本章研究的对象。

库存理论(Inventory Theory)是运筹学最早成功的应用领域之一。早在1915年,哈里斯(Harris)对商业中的库存问题建立了一个简单模型,并求得了最优解,但未被人们重视。1918年威尔逊(Wilson)重新得出了哈里斯公式,这就是下面介绍的确定性模型和威尔逊公式。模型称作确定性的是因为模型中不考虑随机因素。第二次世界大战后,带有随机因素的库存模型也得到了研究,在模型中考虑到了需求量的随机性。目前,库存理论的兴趣已转到了多种商品、多个库存点的理论。

9.1　库存论中的几个要素

商品不断地通过供——存——销三个环节来满足需要。在这样一个系统中,管理人员可以通过控制购入的时间及购入量来调节系统的运行,使供求关系更合理。典型的库存问题如图9.1。

1. 需求

对库存系统来讲,需求就是它的输出。需求量可以通过市场销售的数据进行处理得到,它可以是一个随机变量。例如,某种型号的自行车每天的销量就可以看作一个随

购(补充供应) ⟶ 库存 ⟶ 销(需求)

图 9.1

机变量。需求量也可以是一个常量,如自动生产线上每个班对某种零件的需求量。

2. 补充供应

它是使库存系统运行的输入。库存物品的补充可以是订货,也可以是生产,要解决的是什么时间补充以及补充多少的问题。有些情况下,从订货到收货之间有一段滞后时间,滞后时间也分常数与随机变量两种类型。

3. 库存策略

库存策略给出何时补充以及补充多少的一个方案。一个库存系统,为了了解库存量的多少需要进行检查。一类是连续型的,即对任一时刻 t,我们可以知道库存水平 $I(t)$;另一类是周期性的检查,即时刻 kt $(k=1,2,\cdots)$ 的库存水平是已知的,t 是一个常数,称检查周期。在库存策略中常用的是 (s,S) 型策略,s 叫订货点,S 叫库存水平。在连续检查的系统中,一旦库存量达到 s,则立即订货(或生产)。其订货量 $Q=S-s$。因此,由 s 确定订货时间,由 $Q=S-s$ 确定订货量。在周期检查的系统中,若在某一个检查 kt,库存水平 $I(kt)<s$,则立即订货,订货量为 $Q=S-I(kt)$。若 $I(kt)\geqslant s$,则该时刻不订货。

4. 费用

它包括进货、保管、缺货损失引起的费用等。进货费用通常有如下形式:

$$C(z)=\begin{cases}K+cz & z>0\\ 0 & z=0\end{cases}$$

这里 z 为进货量,K 为每进一次进货所需的固定费用(如手续费,最小的起运费或生产准备费等),c 为货物的单位价格(或单位成本)。

保管费用,它包括库存费用、保险费、税金以及保管过程中的损耗等。

缺货费用,用来衡量因为缺货而带来的损失。缺货造成的直接损失是营业额的损失,使收入减少。缺货还使企业的信誉受到损失。如果是医院血库缺血,还会危及人的生命安全。

5. 目标函数

要在一类策略中选择一个最优策略,就需要有一个赖以衡量优劣的准绳,这就是目标函数。通常把目标函数取为平均费用函数或平均利润函数。选择的策略应使平均费用达到最小值,或使平均利润达到最大值。

9.2 确定性库存模型

9.2.1 模型 I —— 无限供给率、不许缺货

假定：连续检查，系统运行的时间无限，且：

(1) 需求率 D (Demand，单位时间的需求量) 是常数；

(2) 从订货到交货间的时间为 0 (当交货能力或生产能力很大，从订货到交货间的时间很短时，我们就可以这样处理 Shortage cost per unit，independent of time = 0，Replenishment or production rate per year = ∞)；

(3) 不允许缺货 (这意味着缺货的损失是非常大的，Shortage cost per unit per year = ∞)；

(4) 采用 (s,S) 型策略；

(5) $t=0$ 时，库存水平为 S；

(6) 费用为进货费用与保管费用之和；

(7) 目标函数为单位时间的平均费用。

我们来寻求最优策略，即最优 s，S 值。记 $I(t)$ 为时刻 t 的库存水平，设 k 为订货周期，即两次相邻订货之间的时间，因在 $0<t<k$ 中无订货 (图 9.2)，故，

$$I(t+dt)=I(t)-Ddt$$

因此，$I(t)$ 满足微分方程

$$I'(t)=-D,\quad 0<t<k \tag{9.1}$$

$$I(0)=S \tag{9.2}$$

解之，得：

$$I(t)=S-Dt,\quad 0\leqslant t<k \tag{9.3}$$

在一个周期结束时库存水平为 s，即 $I(k)=s$，于是由 (9.3) 式得：

$$Dk=S-s=Q \tag{9.4}$$

这里 Q 就是订货量。

下面考虑费用情况：

一个周期中只有一次订货，订货量为 Q，故每个周期进货费用为 $K+cQ$，因而单位时间平均进货费用为：

$$\frac{1}{k}\cdot(K+cQ)$$

每个周期的平均库存水平为：

$$\frac{1}{k}\cdot\int_0^k I(t)\,\mathrm{d}t$$

设单件的保管费为 h，则单位时间平均保管费用为故总平均费用为。故总平均费用为：

$$\bar{F}(Q,S,k)=\frac{K+cQ}{k}+\frac{h}{k}\cdot\int_0^k I(t)\,dt \tag{9.5}$$

由（9.3）、（9.4）式知

$$\int_0^k I(t)\,\mathrm{d}t=\int_0^k (S-Dt)\,\mathrm{d}t=Sk-\frac{Qk}{2}=k\left(S+\frac{Q}{2}\right)$$

故

$$\bar{F}(Q,S,k)=\frac{K+cQ}{k}+\left(S+\frac{Q}{2}\right)h$$

利用（9.4）式消去上式中的 k，可得目标函数：

$$F(S,Q)=cD+\frac{KD}{Q}+\left(S+\frac{Q}{2}\right)h \tag{9.6}$$

因此，在（s，S）策略下，最佳的 S 及 Q 由

$$\min\{F(S,S);S\geqslant 0,Q>0\}$$

给出。由于目标函数 $F(S,Q)$ 中未知量 S 和 Q 是可以分开的，即：

$$F(S,Q)=(cD+Sh)+\left(\frac{KD}{Q}+\frac{hQ}{2}\right)$$

故可以分别求 $cD+Sh$ 及 $\frac{KD}{Q}+\frac{hQ}{2}$ 的最小值点。易见在约束条件 $S\geqslant 0$ 下，当 $S=0$ 时 $cD+sh$ 达到最小值，故最优解为 $s^*=0$。记：

$$\bar{F}(Q)=\frac{KD}{Q}+\frac{hQ}{2}$$

解 $\bar{F}'(Q)=0$，得到最优解为：

$$Q^*=\sqrt{\frac{2KD}{h}}$$

Q^* 的确定可如图 9.2 所示。我们得到了 S 和 Q 的最优解分别为

$$s^*=0,\quad Q^*=\sqrt{\frac{2KD}{h}} \tag{9.7}$$

Q^* 的公式（9.7）通常称作经济订货量公式，或威尔逊公式，或最佳批量公式。当 s，S 取最优值时，平均费用达到最小，其值为：

$$\begin{aligned}F(s^*,Q^*)&=cD+\frac{KD}{Q^*}+\left(s^*+\frac{Q^*}{2}\right)h\\&=cD+\sqrt{\frac{2KD}{h}}\end{aligned} \tag{9.8}$$

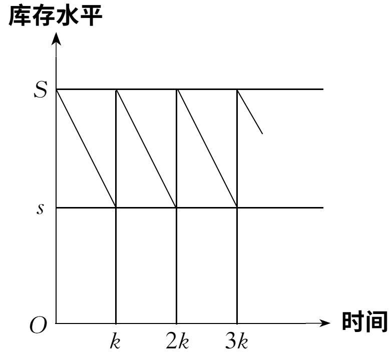


图 9.2

经济订货量公式广泛地用于工业中，部分原因是由于其概念及计算简单，另外也由于对费用及需求估计不太准时对结果的影响不太大。

例 1　某工厂的自动装配线每年要用 480 000 个某种型号的电子管。生产该电子管的成本是每个 5 元，而每开工一次，生产的准备费用为 1000 元。估计每年该电子管的保管费为成本的 25%。若不允许缺货，每次生产的批量应多大？每年开工几次生产该电子管？

解　$D = 480\ 000$（个／年），$K = 1000$（元），$c = 5$（元／个），

$h = 0.25 \times 5 = 1.25$（元／个·年）

由（9.7）式算出：

$$Q^* = \sqrt{\frac{2KD}{h}} = \sqrt{2 \times 1000 \times \frac{480\ 000}{1.25}} = 27\ 713\text{（个）。}$$

因此每年开工次数为 $\dfrac{D}{Q^*} = \dfrac{480\ 000}{27\ 713} = 17$。

9.2.2　模型 Ⅱ —— 无限供给率、允许缺货

在这个模型中缺货是允许的，在单位时间内缺货一件的损失费为 c_S。缺货时未能得到满足的需求，在收到下批订货时要予以满足，到货时直接输出而无须经过库存。其余的假定与模型 Ⅰ 相同。问题仍然是确定最佳 s、S。使单位时间的平均总费用（包括进货费、保管费及缺货损失）达到极小值。

仍设进货周期为 k，在一个周期中库存量由 S 变到 0 所需时间长度记为 k_1（如图 9.3）。与模型 Ⅰ 中的（1）–（3）完全一样的推导可得：

$$I(t) = S - Dt, \quad (0 \leqslant t < k) \tag{9.9}$$

在 $0 \leqslant t < k$ 这段时间内库存水平有时为正，有时为负，负值表示缺货量。在一

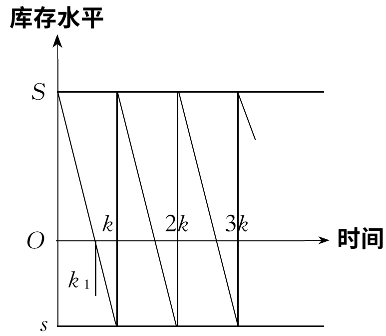


图 9.3

个周期内的平均库存水平为：

$$\frac{1}{k}\cdot\int_0^{k_1} I(t)\,\mathrm{d}t = \frac{1}{k}\cdot\int_0^{k_1}(S-Dt)\,\mathrm{d}t = \frac{1}{k}\cdot\left(Sk_1-\frac{Dk_1^2}{2}\right) \tag{9.10}$$

而在一个周期中的平均缺货量为：

$$-\frac{1}{k}\cdot\int_{k_1}^{k}(S-Dt)\,\mathrm{d}t = -\frac{1}{k}\cdot\left[S(k-k_1)-\frac{D(k^2-k_1^2)}{2}\right] \tag{9.11}$$

单位时间中总平均费用为下面三部分之和：

(1) 单位时间平均进货费$\frac{K+cQ}{k}$，这里到 Q 是进货量，$Q=S-s$；

(2) 单位时间平均保管费$\frac{h}{k}\cdot\left[Sk_1-\frac{D}{2}\cdot k_1{}^2\right]$；

(3) 单位时间平均缺货损失费$\frac{cs}{k}\cdot\left[S(k-k_1)-\frac{D}{2}\cdot(k^2-k_1{}^2)\right]$。

利用

$$S=Q+s=Dk_1$$

$$Q=Dk$$

消去总费用中的 k_1，k，整理以后得到目标函数为：

$$F(Q,S)=\frac{KD}{Q}+cD+\frac{hS^2}{2Q}+\frac{c_s(Q-S)^2}{2Q} \tag{9.12}$$

问题化为，在约束条件 $Q>0$，$S>0$ 下，求使 (9.12) 式达到极小值的 Q^*，S^*。

由极值必要条件，

$$\frac{\partial F}{\partial Q}=-\frac{KD}{Q^2}-\frac{hS^2}{2Q^2}+\frac{c_s}{2}\left(1-\frac{S^2}{Q^2}\right)=0$$

$$\frac{\partial F}{\partial S}=\frac{hS}{Q}+\frac{c_s}{Q}\cdot(S-Q)=0$$

即

$$c_s Q^2 - (c_s + h) S^2 = 2KD$$
$$(c_s + h) S = c_s Q$$

解得(舍去负解)

$$Q^* = \sqrt{\frac{2KD}{h} \cdot \left(1 + \frac{h}{c_s}\right)}$$
$$S^* = \frac{c_s}{c_s + h} \cdot Q^* = \sqrt{\frac{2KD}{h} \cdot \frac{c_s}{c_s + h}} \qquad (9.13)$$

容易验证,它们使(9.12)式达到极小值。

由 Q 及 S 的最优值可进一步求出 s 的最优值为:

$$s^* = S^* - Q^* = \sqrt{\frac{2KD}{h} \cdot \frac{c_s}{c_s + h}} \left(1 - \frac{c_s + h}{c_s}\right) = -\sqrt{\frac{2KDh}{c_s (c_s + h)}} \qquad (9.14)$$

关于模型Ⅱ,还有两点说明如下:

(1)若记 α 为在最优方案下系统缺货实践的比例,则有

$$\alpha = 1 - \frac{k_1^*}{k} = 1 - \frac{S^*}{Q^*} = 1 - \frac{c_s}{c_s + h} = \frac{1}{1 + c_s / h} \qquad (9.15)$$

例如,若缺货损失与保管损失之比为 $\frac{c_s}{h} = 10$,则系统缺货的时间比例为 $\alpha = \frac{1}{11} \approx 9\%$。有时候 c_s 是不易估计的。而决策者倒比较容易定出一个适当的缺货时间比例 α,这时反过来由(9.15)式可求出 c_s 的值。

(2)若 $c_s \to \infty$,则经济订货量 Q^* 的公式(9.13)与模型Ⅰ中的(9.7)式是一致的。

9.2.3 模型Ⅲ——有限供给率,允许缺货

设 R($R > D$)为单位时间交货率,其余假设与模型Ⅱ相同。

图9.4是这个模型的库存变化图。在一个周期(0,k)中,从开始到时刻 k_1 库存水平以单位时间需求量减少;从 k_1 到 k 进货(生产)和需求同时存在,因此库存水平以单位时间 $R - D$ 的速度增加。

下面利用简单的几何论证把这个模型化为模型Ⅱ。

在一个周期中,保管费用与 $\triangle AOS$、$\triangle BkF$ 两个三角形的面积之和成正比,而缺货的损失与 $\triangle ABC$ 的面积成正比。连接 SE,由相似三角形之间对应边成比例,有:

$$AG/CE = SA/SC = SO/Ss = Bk/CE$$

由此得到 $AG = Bk$。因而 $\triangle ABC$ 的面积等于 $\triangle GkE$ 的面积;$\triangle SOG$ 的面积等

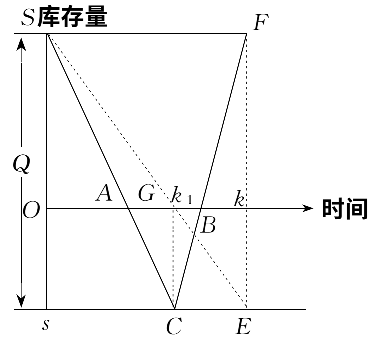


图 9.4

于 $\triangle AOS$ 与 $\triangle BkF$ 面积之和。这样，图 9.4 中的每一个周期对应图 9.4 的一个周期。相应的需求率是 SE 的斜率的绝对值，记作 D_1，则：

$$D_1 = \frac{Q}{k} \tag{9.16}$$

而 SC、CF 的斜率的绝对值分别是：

$$D = \frac{Q}{k_1}, R - D = \frac{Q}{k - k_1}$$

因此得：

$$k = \frac{Q}{R - D} + \frac{Q}{D}$$

代入 (9.16) 式得到：

$$D_1 = \frac{D(R - D)}{R} \tag{9.17}$$

于是，在 (9.13)、(9.14) 式中把 D 取作 D_1 便得到最佳方案，即：

$$S^* = \sqrt{\frac{2KD(R - D)}{hR}} \cdot \sqrt{\frac{c_s}{c_s + h}} \tag{9.18}$$

$$s^* = \sqrt{\frac{2KD(R - D)}{c_s R}} \cdot \sqrt{\frac{h}{c_s + h}} \tag{9.19}$$

并且，进一步可以算出最佳订货周期长度为：

$$k^* = \frac{Q^*}{D_1} = \sqrt{2KP[hD(R - D)]} \cdot \sqrt{1 + \frac{h}{c_s}} \tag{9.20}$$

在这一周期中开始生产的最佳时间为：

$$k_1^* = \frac{Q^*}{D} = \sqrt{\frac{2K(R - D)}{DhR}} \cdot \sqrt{1 + \frac{h}{c_s}} \tag{9.21}$$

生产进行的时间长度为：

$$k^{*}-k_{1}^{*}=\sqrt{2KD[hR(R-D)]}\cdot\sqrt{1+\frac{h}{c_{s}}}$$

特例1　若不允许缺货，交货能力有限，此时有 $c_s=\infty$，相应的公式成为

$$S^{*}=\sqrt{\frac{2KD(R-D)}{hR}},s^{*}=0,Q^{*}=S^{*}$$

$$k^{*}=\sqrt{2KR[hD(R-D)]}$$

$$k_{1}^{*}=\sqrt{\frac{2K(R-D)}{DhR}}$$

特例2　当 $R=\infty$ 时，即意味着交货能力是无限的，模型Ⅲ中的公式与模型Ⅱ中的相应公式是一致的。

9.3　折扣分析

现在考虑具有数量折扣的经济订货量模型。所谓数量折扣，就是提供存贮货物的企业为鼓励用户多用货物，对于一次购买较多数量的用户在价格上给予一定的优惠。换句话说，单位货物购置费 c 应看作是 Q 的函数 $c(Q)$。通常，$c(Q)$ 是下降的阶梯函数。

为讨论方便，我们仅对最简单的情况进行分析，其方法对一般情况同样适用。设

$$c(Q)=\begin{cases}c & 0<Q<Q_0\\ c(1-\beta) & Q\geqslant Q_0\end{cases}$$

其中 $0<\beta<1$，β 被称为价格折扣率。这时，(9.6) 式就变成下列条件极值问题：

$$\min F(s,Q)=c(Q)^{*}D+\frac{KD}{Q}+[S+\frac{Q}{2}]h$$

$$s.t.\ S\geqslant 0,Q>0 \tag{9.22}$$

在对其求最优解时，当注意 $c(Q)$ 并不总是与 Q 无关的常量。当然，Q 在 $(0,Q_0)$ 内或在 $[Q_0,\infty)$ 内时 $c(Q)$ 都是常量，仍有结果 (9.7) 式。因为 $c(Q)$ 是分段函数，所以，(9.22) 中的 F 也是分段函数，如图 9.5 所示。

当 $Q^{*}>Q_0$ 时 (如图 9.5 (c))，Q^{*} 就是 (22) 的最优解。当 $Q^{*}<Q_0$ 时 (如图 9.5 (a)，(b))，还要比较 $f(Q^{*})$ 与 $f(Q_0)$ 的大小。若 $f(Q^{*})<f(Q_0)$，则 Q^{*} 为问题的最优解，否则，Q_0 为最优解。

当 $Q^{*}<Q_0$ 时，若把订货量从 Q^{*} 提高到 Q_0，则易见：购置费从 cD 下降为

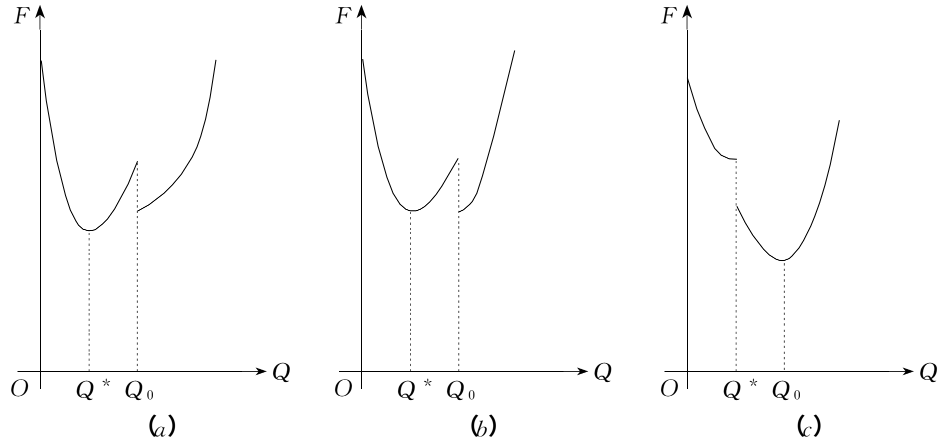


图 9.5

$c(1-\beta)D$，订货费从$\frac{KD}{Q^*}$下降为$\frac{KD}{Q_0}$，而保管费从$\frac{Q^*h}{2}$上升为$\frac{Q_0h}{2}$(库存量增加)。

例2　某工厂生产某种产品。原料充分供应，一年生产1500套($D=1500$)。每套原料费48元($c=48$)。估计每套原料年库存保管费(h)为原料费的22%。此外，每次订货要订货费25元($c_0=25$)，求经济订货量EOQ若该厂在订货购买6个月原料时，原料供应者可给予5%的折扣，问该厂会接受这个建议吗？

解　代入(9.7)式——注意，$K=25$

$$EOQ=Q^*=\sqrt{\frac{2\times25\times1500}{0.22\times48}}=84.27$$

代入(9.6)式，得到年总花费为27 889.9(元)。

当享受折扣时，每套原料费为$48(1-0.05)=45.6$，$Q=750$

于是，这时，总费用 = 套材料费 + 年订货费 + 年库存费

$$F=1500\times45.6+\frac{25\times1500}{750}+\frac{10.03\times750}{2}$$
$$=68\,400+50+3761.25=72\,211.25\text{(元)}$$

较原来减少了678.70元，所以该建议可以接受。

9.4　随机性库存模型

在很多情况下库存模型中的参数不是固定不变的常量，比如某种商品在一天中

的销量就是这样的。然而在变化之中往往显示出某种规律性，通过对历史资料的统计可以看出它服从一定的分布规律。在这种情况下我们就应该用随机变量来描述它。本节介绍一个需求量是随机变量的模型——单阶段随机需求模型。在学习这个模型时要注意，它和前面三个模型的区别不仅表现在需求量由常量变成了随机变量，还表现在时间变量的作用是不同的。在前面，时间变量是连续变量；而在这个模型中一个阶段（或称一个库存周期）是时间的最小单位，这就是单阶段的特点。在这样的模型中仅在一个阶段的开始做一次决策，比如决定进货量。由于一个阶段的需求是个随机变量，不管所确定的进货量是多少，总有三种情况可能发生：即供求正好相等，供过于求或供不应求。在供过于求的情况下，说明进货量及原有库存量之和超过了需求量，货物不能全部销售出去，需为所剩的货物支付保管费（也包括货物变质等损失）；如果发生了供不应求的情况，即进货量与原有库存之和小于需求量，则需承担缺货损失费。

假设：① 需求量 X 是随机变量，分布为已知：

$$P(X=i)=p_i, \quad i=0,1,2,\cdots$$

② 采用 (s,S) 策略。设 $t=0$ 时库存水平为 i，若 $i \geqslant s$，则不订货；若 $i<s$，则订货量为 $S-i$，且假定立即可以进货。

③ 费用有进货费、保管费和缺货损失费三项。

④ 目标函数是平均费用。

下面寻求使目标函数达到极小值的最佳 (s,S) 策略。先考虑费用函数：

1. 进货费。设进货量为 z，则进货费用为：

$$C(z)=\begin{cases}0 & z=0\\ K+cz & z>0\end{cases} \tag{9.23}$$

其中 K 为固定费用，c 为单价。

2. 保管费。记进货量与原有库存之和为 y，单价保管费为 $h \geqslant 0$，则保管费是个随机变量，它等于

$$H(y;x)=\begin{cases}0 & x \geqslant y\\ h(y-x) & x<y\end{cases}$$

保管费的期望值为：

$$EH(h;x)=\sum_{i=0}^{y} h(y-i)p_i$$

3. 缺货损失费也是个随机变量，它等于：

$$P(y;x)=\begin{cases}0 & x \leqslant y\\ c_s(x-y) & x>y\end{cases}$$

它的期望值为：

$$EP(y;x)=\sum_{i=y+1}^{\infty}c_s(i-y)p_i$$

其中 c_s 是单件缺货损失，记：

$$\begin{aligned}L(y)&=EH(y;x)+EP(y;x)\\&=\sum_{i=0}^{y}h(y-i)p_i+\sum_{i=y+1}^{\infty}c_s(i-y)p_i\end{aligned}\tag{9.24}$$

于是，当初始库存水平为 i 时，总费用函数为：

$$g(y\mid i)=\begin{cases}L(i) & y=i\\K+c_s(y-i)+L(y) & y>i\end{cases}\tag{9.25}$$

而使得 $g(y\mid i)$ 取到最小值的 y 便是最优的库存水平，有：

$$\begin{aligned}f(i)&=\min_{y\geqslant i}g(y\mid i)\\&=\min\begin{cases}\min\limits_{y>i}g(y\mid i)\\g(i\mid i)\end{cases}\\&=\min\begin{cases}K-c_s+\min\limits_{y>i}[cy+L(y)]\\L(i)\end{cases}\end{aligned}$$

先来求最优策略中的 S。这时我们不妨设上式中的最小值不等于 $L(i)$，否则将意味着这时不需要订货。于是问题化为求 y，使 $[cy+L(y)]$ 达到极小值。

注意，cy 与 $L(y)$ 均为凸函数，即对于任意的整数 y 均有：

$$L(y)\leqslant\frac{L(y-1)+L(y+1)}{2}$$

cy 的凸性是显然的；下面验证 $L(y)$ 的凸性。我们去验证另一个等价的不等式

$$L(y+1)-L(y)\geqslant L(y)-L(y-1)$$

事实上，

$$\begin{aligned}&L(y+1)-L(y)\\&=\sum_{i=0}^{y+1}h(y+1-i)p_i+\sum_{i=y+2}^{\infty}c_s(i-y-1)p_i-\sum_{i=0}^{y}h(y-i)p_i-\sum_{i=y+1}^{\infty}c_s(i-y)p_i\\&=h\sum_{i=0}^{y}p_i-c_s\sum_{i=y+1}^{\infty}p_i=(h+c_s)\sum_{i=0}^{y}p_i-c_s\end{aligned}\tag{9.26}$$

类似地推导可得：

$$L(y)-L(y-1)=(h+c_s)\sum_{i=0}^{y-1}p_i-c_s$$

易见所要验证的不等式显然成立。两个凸函数之和仍为凸函数，故 $cz+L(z)$ 为凸函数。而凸函数必定存在一个极小值点，并且是全局的极小，我们可以用下面的方法

来找到它。设 $y=S$ 是函数取极小值的点，即：

$$\min_y [cy+L(y)]=cS+L(S) \tag{9.27}$$

则必有：

$$c(S+1)+L(S+1)\geqslant cS+L(S)$$

即

$$L(S+1)-L(S)\geqslant -c$$

代入（9.25）式得到：

$$L(S+1)-L(S)=(h+c_s)\sum_{i=0}^{s}p_i-c_s$$

于是得：

$$(h+c_s)\sum_{i=0}^{s}p_i\geqslant c_s-c$$

因而 S 是使：

$$\sum_{i=0}^{s}p_i\geqslant\frac{c_s-c}{c_s+h} \tag{9.28}$$

成立的最小正整数。

下面再来讨论 s 的求法。在 (s,S) 策略之下，当初始库存量等于 s 时不订货。因此，$y=s$ 时的总费用（此时订货量 0）为 $L(s)$，它应小于 $y=S$（此时订货量为 $S-s$）的总费用 $K+c(S-s)+L(S)$。因而有：

$$L(s)\leqslant K+c(S-s)+L(S) \tag{9.29}$$

另一方面，当初始库存小于 s 时，应该订货。故当（9.29）式中的 s 成为 $s-1$ 时不等号要反过来。因此，最佳的订货点 s 是使（9.29）式成立的最小整数。

把 s,S 的求法归纳如下：

（1）由（9.27）式求出 S。对于通常的离散分布随机变量 X，$\sum_{i=0}^{s}p_i$ 都可以通过查表得到。

（2）由（9.28）式求出 s。或者换成（9.29）式的等价形式：

$$L(s)+cs\leqslant K+cS+L(S) \tag{9.30}$$

（9.29）式的右端不含 s，算起来更方便些。

例 3　设需求量 X 满足

$$p_i=P(X=i)=\frac{1}{5},\quad i=0,1,2,3,4$$

$$h=5,c_s=10,c=0.1,K=4$$

求 s,S。

解
$$\sum_{i=0}^{s} p_i = \sum_{i=0}^{s} \frac{1}{5} = \frac{S+1}{5}$$
$$\frac{c_s - c}{c_s + h} = 0.66$$

故满足$\frac{S+1}{5} \geqslant 0.66$的最小整数 $S = 3$。S 也可以由 X 的分布函数图形（图 9.6）解出。

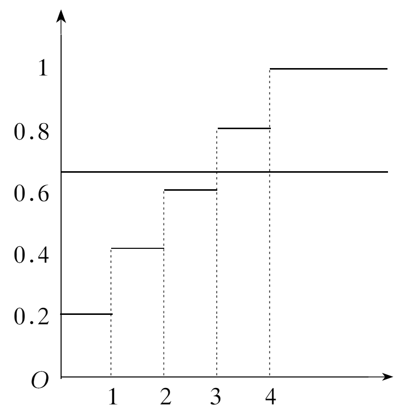


图 9.6

因
$$L(y) = \begin{cases} h(y-2) & y \geqslant 4 \\ \frac{h}{5} \cdot \sum_{i=0}^{y}(y-i) + c_s \cdot \sum_{i=y+1}^{4}(i-y) & 0 \leqslant y \leqslant 3 \end{cases}$$

对于 $S = 3, K = 4, c = 0.1, h = 5, cs = 10$，(9.29) 式右边为：
$$K + cS + L(S) = 4 + 0.1 \times 3 + \sum_{i=0}^{3}(3-i) + 2 = 12.3$$
算出 $L(2) = 9, L(1) = 13$，有：
$$L(2) + 0.1 \times 2 = 9.2, \quad L(1) + 0.1 \times 1 = 13.1$$

因此 $s = 2$ 是使 (9.29) 式成立的最小整数，得最优库存策略为 (2,3)。

在随机性库存模型中还有需求量和交货时间都是随机变量的情形，以及多种物资的库存问题。另外，从解决库存问题的方法上讲，许多决定性的模型可以化为线性规划和动态规划来解。

9.5 微机操作 —— 调用 YAJ 中，INVT 决策支持系统

9.5.1 观察 INVT 决策支持系统

这个程序使你能计算三个经典的库存论问题。它们是经济订货量问题（EOQ），带有折扣分析的经济订货量问题，以及单周期随机系统。EOQ 要求的数据是存储保管费用，订货费用，缺货费用和存储单位的购进费用（生产成本）。附带的数据要求是需求量，补充速度，采购间隔期。对于折扣分析，除了 EOQ 数据外，还要求折扣和价格下降数。单周期随机问题要求售价，残值，单位和缺货费用以及需求量分布。

9.5.2 问题选择

为了输入你的模型，对下列选项做出选择：

① 确定经济订货量（EOQ），② 确定折扣分析，③ 单周期随机需求量问题，④ 返回到功能菜单。

9.5.3 输入和求解

1. EOQ 问题 —— 仍以例 1 为例

EOQ 输入：

需求量，每年（D）$= 480\ 000$

每次订货的订货或组织费用（Co）$= 1000$

存贮费用，每一单位，每年（Ch）$= 1.25$

缺货费用，每一单位，每年（Cs）$= \infty$

每单位的缺货费用，与时间无关（∞）$= 0$

补充或生产速度，每年（P）$= \infty$

每年内对于一次新的订货的采购间隔期（LT）$= 0$

单位成本（C）$= 5$

EOQ 输出：

$EOQ = 27\ 713.473$

最大存贮量 $= 27\ 712.143$

最大延期交货期 $= 0.000$

订货区间 $= 0.058$

再订货点 = 0.000

订货费用 = 17 320.096

保管费用 = 17 320.088

缺货费用 = 0.000

存贮费用小计，每年 = 34 640.184

原料成本，每年 = 2 400 000.000

总计成本，每年 = 2 434 640.200

2. 折扣分析 —— 仍以例 2 为例

数据输入：

你要再次进入 EOQ 数据吗（y/n）？y

需求量，每年？1500

订货或生产费用，每次订货？25

存贮费用，每一单位，每年？10.56

缺货费用，每一单位，每年（缺损值 = ∗）？

缺货费用，每一单位，与时间无关（缺损值 = 0）？

补充或生产速度，每年（缺损值 = ∗）？

在一年内对于一次新的订货的采购间隔期（缺损值 = 0）？

不具有折扣的单位花费？48

你要分析多少个折扣下降（< 50）？1

进入折扣分析，用第一个价格（最低的）开始：

折扣下降 ♯1

下降量？750

折扣（%）？5

折扣分析：

没有折扣：EOQ = 84.275，总计成本 = 72 889.945

折扣：5%，量：750

EOQ = 86，总计成本 = 69 267.406，不可能的

下降 = 750，总计成本 = 72 212.000，处理这个价格降

最优决策：折扣 5%，订货 750，总计成本 = 72 212

第十章 排队论

排队是人们日常生活中经常遇到的现象。当要求服务的对象到达服务台不能立刻得到服务时，就必须等候，因而出现了排队现象。

增加服务台，自然可以减少排队现象；但如果增加太多，服务设施又会出现空闲浪费；增加太少，顾客排队等待的时间就会太长，对顾客和社会都会带来不良影响。排队论就是要研究在要求服务的顾客和提供服务的服务台之间，取得平衡，提高服务质量，降低服务费用。

排队论（Queuing Theory）是20世纪初由丹麦工程师爱尔朗（Erlang A.K.）在研究电话系统时初创的。由于它所研究的实际问题有着强烈的实际背景，所以，几十年来，它的应用领域越来越广泛，在理论上也得到很快的发展。在我国它已应用于电信、交通、纺织、矿山及军事等领域。20世纪60年代以来，由于计算机技术的飞速发展，又对排队论的应用开拓了更加广阔的前景。

排队论有时被称为等待线问题，又因为它主要研制随机型的服务问题，所以有时也被称为随机服务系统理论。

10.1 排队论的基本概念

10.1.1 排队现象的共同特征

实际的排队现象虽然千差万别，但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。这些共同的特征是：为了获得某种服务而到达的顾客，如果不能立刻得到服务，而又允许排队等待，则加入等待队伍；获得服务之后，则离开。我们把包含有这些特征的系统称为排队系统。表10.1罗列了现实世界中形形色色的排队系统。

表 10.1

到达的顾客	要求服务的内容	服务台
1. 待修的机器	修理	修理技工
2. 修理技工	领取新零件	零件库保管员
3. 病人	诊断或做手术	医生
4. 电话呼唤	通话	交换台
5. 文件稿	打字	打字员
6. 提货单	提取货物	仓库保管员
7. 到达机场的飞机	降落	跑道
8. 驶入港口的货船	装(卸)货	装(卸)货的泊位
9. 河水进入水库	放水,调整水位	水库的水闸
10. 进入阵地的敌机	高射炮进行射击	我方高射炮

我们还可以用图形来表示一些抽象的排队系统:

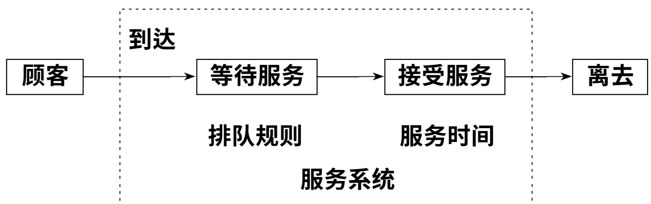


图 10.1　排队服务模型

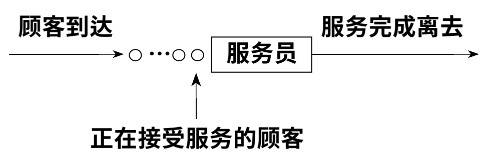


图 10.2　单服务员排队系统的图解表示

10.1.2　排队系统的组成

一般的排队系统有下面三个组成部分:

1. 输入过程

输入过程讨论的是顾客按怎样的规律到达。要完全描述一个输入过程,需要三方面的内容。

(1) 顾客的总体数或顾客源数,它是指可能到达服务机构的顾客总数。这可能是

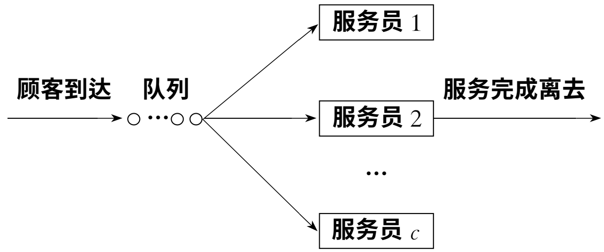


图 10.3　c 个服务员,一个公共队的图解表示

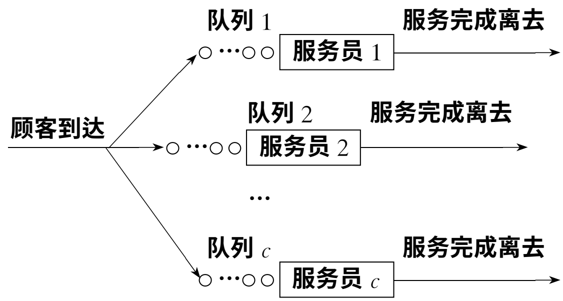


图 10.4　c 个服务员,c 个队列的图解表示

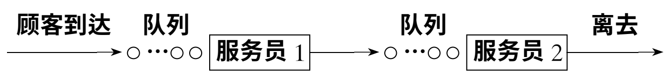


图 10.5　单服务员串联排队系统的图解表示

有限的,也可以是无限的。当顾客相当多时,可近似看作是无限的。

⑵ 顾客到达的类型。顾客是单个到达,还是成批到达。货物进库,可视为顾客成批到达。

⑶ 顾客相继到达的时间间隔分布。从表面上看,顾客到达服务台,似乎是杂乱无章的,其实,常常服从某种统计规律,如定长分布,负指数分布等。

2. 排队规则

排队规则所研究的是顾客接受服务的先后次序问题,有下面几种情形:

⑴ 损失制。顾客到达时,如果所有的服务台都在被占用,则顾客将随即离去,不再接受服务。这将失去许多顾客,称为损失制。

⑵ 等待制。顾客到达时,如所有服务台都被占用,则顾客将排队等待。这又分为:① 先到先服务,这是最通常的情形;② 后到先服务,如堆积存放的钢板,在需要

时，总是从最上面取出使用，这就是后到先服务的例子；③ 随机服务，如电话交换台接通市内电话的呼唤就是一个例子；④ 优先权服务，如危重病人到医院，将先于一般病人得到治疗；加急电报将先于普通电报译送，都是优先权服务的例子。

此外，从占有的空间来看，有的排队系统要规定队伍容量的最大限度，有的则没有这种限制。

(3) 混合制。顾客起初进入排队，但后来觉得等待时间太长，又离开队伍。

3. 服务机构，又称服务台

完全描述服务机构，应有以下内容：

(1) 服务台的数目。有单、多服务台的不同情形。

(2) 任一时刻接受服务的顾客数。有单个接受服务，也有成批接受服务。

(3) 服务时间的分布。一般说来，对每个顾客的服务时间都是随机变量，其分布类型有定长分布、负指数分布等。

10.1.3 排队系统的描述符号和分类

为了以后讨论方便，我们对排队系统定义如下符号：

n：系统中顾客的数目；

λ：顾客到达的平均速率，即单位时间内到达的顾客数；

μ：系统的平均服务速率，即单位时间内可以服务完的顾客数；

$P_n(t)$：系统中有 n 个顾客的概率；

c：服务台的个数；

$FCFS$：先到先服务的排队规则；

$LCFS$：后到先服务的排队规则；

PR：优先权服务的排队规则；

M：到达过程为泊松过程，或服务时间服从负指数分布，因为泊松过程属于马尔柯夫过程，负指数分布具有马尔柯夫性，M 来自 $Markov$ 的第 1 个字母；

D：确定型分布；

E_k：参数为 k 的爱尔朗分布。

1953 年，康道尔（Kendall D.G.）归纳了一种符号来描述排队系统，[$A/B/C$]，其中 A 表示输入过程的特点，B 表示服务过程的特点，而 C 表示服务台的数目。如 [$M/M/1$] 系统表示的是这样一个系统：泊松到达过程，服务时间服从负指数分布，只有一个服务台。

但是，系统中其他的重要因素在康道尔的描述方式中却没有得到反映。1966 年，艾姆·里氏（Lee A.M.）提出再增加三个因素。它的一般表示方式为：

$$[A/B/C]:[d/e/f]$$

其中，A 顾客到达过程的概率分布；

B 服务过程的概率分布；

C 服务台的数目；

d 排队系统的最大容量；

e 顾客总体的数量；

f 排队规则。

例如，$[M/M/4]:[N/\infty/FCFS]$ 表示这样一个系统：泊松到达过程，服务时间为负指数分布，有 4 个服务台，系统最大容量为 N 个顾客，顾客源无限，先到先服务制。

10.2　顾客到达数的分布和服务时间的分布

10.2.1　泊松流

在概率论中，我们曾学习过泊松分布。设随机变量 X 所有可能取的值为 0，1，2，… 而取每个值的概率为

$$P\{X=n\}=\frac{\lambda^n e^{-\lambda}}{n!},\quad n=0,1,2,\cdots \tag{10.1}$$

式中 $\lambda>0$ 是常数，则称随机变量 X 服从参数为 λ 的泊松分布。概率论的知识还告诉我们，泊松分布是一类二项分布的逼近，即每次抽样只能有两个结果，其中一种结果在一次抽样中发生的概率很小，当抽样的次数足够多时，则该事件发生 n 次的概率就近似服从泊松分布。

现在，把上式中的参数 λ 引入时间参数 t，即得：

$$P_n(t)=\frac{(\lambda t)^n e^{-\lambda t}}{n!}\quad t>0,n=0,1,2,\cdots \tag{10.2}$$

此式表示长为 t 的时间内到达 n 个顾客的概率 $P_n(t)$，服从泊松分布。称它为泊松流，或称泊松过程，或称最简单流。它是排队论中最常见的一种理论分布。

应当指出，引入时间参数的随机问题已经超出简单的概率论知识了，这类问题称为随机过程。就随机过程而言，我们既要讨论描述到达顾客的随机变量 X，在下文还要讨论与时间参数有关的前后顾客到达的时间间隔。这是泊松流或泊松过程的两个侧面的问题。

设 t 时间内到达的顾客数为随机变量 $N(t)$，则其数学期望值

$$E[N(t)] = \sum_{n=1}^{\infty} n \cdot P_n(t) = \sum_{n=1}^{\infty} n \left[\frac{(\lambda t)^n}{n!}\right] e^{-\lambda t} = e^{-\lambda t} \cdot \frac{\sum_{n=1}^{\infty} (\lambda t)^n}{(n-1)!}$$

$$= e^{-\lambda t} \cdot (\lambda t) \cdot \frac{\sum_{n=1}^{\infty} (\lambda t)^{n-1}}{(n-1)!} = e^{-\lambda t} \cdot (\lambda t) \cdot e^{\lambda t} = \lambda t \qquad (10.3)$$

同理,可证明泊松流的方差:

$$D[N(t)] = \lambda t \qquad (10.4)$$

10.2.2　负指数分布

当输入过程是泊松流时,研究两个顾客先后到达的时间间隔 T 的概率分布。由于在单位时间里到达的顾客数是随机变量,那么,对应地,前后两个顾客到达的时间间隔也就是随机变量了,即有时间隔长一些,有时间隔又短一些。

设 T 的分布函数为 $F_T(t)$,则

$$F_T(t) = P\{T \leqslant t\}$$

这个概率也就是在 $[0,t)$ 区间内至少有一个顾客到达的概率。由前面结果,在 $[0,t)$ 内没有顾客到达的概率为

$$P_0(t) = e^{-\lambda t} \quad (\text{当 } n = 0 \text{ 时})$$

所以,至少有一个顾客到达的概率:

$$F_T(t) = 1 - P_0(t) = 1 - e^{-\lambda t}, \quad t > 0 \qquad (10.5)$$

而概率密度函数为:

$$f_T(t) = \frac{dF_T(t)}{dt} = \lambda e^{-\lambda t}, \quad t > 0 \qquad (10.6)$$

就是说,到达间隔时间 T 服从负指数分布。由上可知,如果顾客到达是泊松流,那么,先后顾客到达的间隔时间必定服从负指数分布。反过来,如果相继到达的间隔时间服从负指数分布,那么,可以证明,t 时间内到达的顾客数一定服从泊松分布,即是泊松流。

负指数分布是连续型概率分布,其密度函数如图 10.6,它是一个单调递减函数。

负指数分布的期望:

$$E[T] = \int_0^{\infty} t \cdot \lambda e^{-\lambda t} dt = \frac{1}{\lambda} \qquad (10.7)$$

方差:

$$D[T] = \frac{1}{\lambda^2} \qquad (10.8)$$

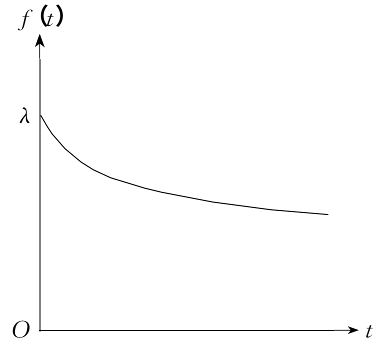


图 10.6

式(10.1)中的 λ 表示单位时间内到达的顾客数的平均值,这时的间隔时间的平均值显然是 $\frac{1}{\lambda}$,这同负指数的期望值是一致的。

服务时间 v 的分布。

为了讨论问题方便,一般地,我们假设顾客接受服务的时间 v 也服从负指数分布。设其概率分布函数和密度函数分别是:

$$F_v(t) = 1 - \mathrm{e} - \mu t \tag{10.9}$$

$$f_v(t) = \mu \mathrm{e} - \mu t \tag{10.10}$$

式中,μ 表示单位时间内能够服务完的顾客数,称为服务速率,而 $\frac{1}{\mu} = E(v)$ 就表示一个顾客的平均服务时间。排队论中的“平均”就是指概率论中的期望值。

(10.1)~(10.7)式中的 λ 和 μ 的比值:

$$\rho = \frac{\lambda}{\mu} \tag{10.11}$$

称为服务强度,它是刻划服务效率和服务机构利用强度的重要标志。当 $\rho < 1$,ρ 值越小,单位时间内到达顾客的平均数比能服务完的顾客数要小得多,这种情况下顾客就越能及时得到服务,等待时间就越少,相应地服务员空闲的时间就越多,服务设施的利用率就越低。反过来,当 $\rho < 1$,ρ 值越大,单位时间内到达的顾客的平均数与能够服务完的平均数相差不多,在这种情况下,顾客等待的时间就越多,相应地,服务员空闲的时间就越少,服务设施的利用率就高。

10.2.3 爱尔朗分布

设 $v_1, v_2, \cdots, v_k$ 是相互独立的随机变量,且都服从相同参数 $k\mu$ 的负指数分布,那么,令随机变量 T

$$T = v_1 + \cdots + v_k \tag{10.12}$$

则 T 的概率密度

$$P_k(t) = \left[\frac{\mu k(\mu kt)^{k-1}}{(k-1)!}\right] \cdot e^{-\mu kt}, \quad t > 0 \tag{10.13}$$

则称随机变量 T 服从 k 阶爱尔朗分布。且它的期望为：

$$E[T] = \frac{1}{\mu} \tag{10.14}$$

方差为：

$$D[T] = \frac{1}{k\mu^2} \tag{10.15}$$

这是因为随机变量和的期望等于各随机变量期望的和，所以

$$\begin{aligned} E[T] &= E(v_1 + \cdots + v_k) \\ &= E(v_1) + \cdots + E(v_k) \\ &= \frac{1}{\mu k} + \cdots + \frac{1}{\mu k} \qquad (\text{计 } k \text{ 项}) \\ &= \frac{1}{\mu} \end{aligned}$$

又因为相互独立的随机变量和的方差等于各随机变量方差的和，即

$$\begin{aligned} D[T] &= D(v_1 + \cdots + v_k) \\ &= D(v_1) + \cdots + D(v_k) \\ &= \frac{1}{(\mu k)^2} + \cdots + \frac{1}{(\mu k)^2} \qquad (\text{计 } k \text{ 项}) \\ &= \frac{1}{k\mu^2} \end{aligned}$$

爱尔朗分布存在于这样一类排队系统中：有 k 个串联的服务台，每个服务台服务时间相互独立，服从相同的负指数分布（参数为 k_μ），那么，一个顾客走完这 k 个服务台总共需要的服务时间就服从 k 阶爱尔朗分布。显然，$k = 1$ 时，k 阶爱尔朗分布就是负指数分布，而当 $k \to \infty$，即若有无穷多个服务台相串联时，爱尔朗分布将退化为确定型分布。

注　因为理论分布有一定的规律，人们研究得比较彻底，分析起来也比较方便，所以，在研究某一实际排队系统时，常常要把经验分布拟合成某种理论分布（如果能够拟合的话），这时，就应先从调查和统计数据入手，把这些数据加以整理，然后分析数据的特点，看看它们可能适合何种理论分布。

对于排队系统，常见的理论分布是泊松流，泊松流的主要假设条件是：

① 无后效性，即前面到达的顾客数并不影响后面到达的顾客数；

② 平稳性，即顾客到达的多少只同时间间隔有关，而同统计时的时刻无关；

③ 普通性，即在很小的时间间隔内，到达两个或两个以上的顾客的概率极小，可以忽略。

尽管在实际系统中，上述三个假设条件并不一定都成立，但是，理论证明和实际统计资料都表明了，排队系统中的实际现象往往比较符合泊松流。例如，商店到达的顾客数，电话的呼唤次数，一般都比较符合泊松分布。而负指数分布的服务时间，由于其简单易于处理，也往往成为研究实际系统的起点。当然，由于实际现象的千差万别，也还有其他形式的输入流。

10.3 单服务台的 [M/M/1] 模型

对于随机型排队系统，一般是给定顾客到达的分布形式和服务时间的分布形式，即在给定输入条件和服务条件下，研究系统的运行指标，这些指标有：

1. 在排队系统中顾客数的期望值 L，它是指正在被服务的和排队等待的两部分顾客的期望值；

2. 排队等待的顾客的期望值 L_q；

3. 顾客在系统中全部时间的期望值 W，它是指顾客排队等待的时间和被服务时间之和的期望值；

4. 顾客排队等待的时间的期望值 W_q。

为了求解上述这些指标，要先求解系统状态为 n（即有 n 个顾客）的概率 P_n（$n=1,\cdots$），然后根据概率 P_n 推算这些指标。对于常见的几种排队系统，将推算出若干计算公式。先考虑单服务台情形。

10.3.1 [M/M/1]：[∞/∞/FCFS] 系统

此即泊松流输入，负指数分布的服务时间，只有一个服务台，排队系统内的顾客数没有限制，顾客源无限，先到先服务制的系统。

1. 系统稳态概率 $P(n)$ 的计算

在分析这个模型时，首先要求出系统在任意时刻 t 的状态 n 的概率 $P_n(t)$，它决定了系统运行的特征。我们以细菌分裂和死亡的例子来推导这个模型中有关 $P_n(t)$ 的微分方程组。

假定有一堆细菌，在 t 时刻有 n 个细菌的概率为 $P_n(t)$，考虑从 t 到 $t+\triangle t$ 时，系统中有 n 个细菌的概率变化情况。这里，$\triangle t$ 是很小的时间间隔。我们把 $\triangle t$ 限制到足够小，得到下列事件中只有一种可以发生：

(1) 只有一个细菌分裂而没有细菌死亡的概率为 $\lambda\triangle t+o(\triangle t)$ ($o(\triangle t)$ 是 $\triangle t$ 的高阶无穷小)；

(2) 只有一个细菌死亡而没有细菌分裂的概率为 $\mu\triangle t+o(\triangle t)$；

(3) 两个或两个以上细菌分裂或死亡的概率为 $o(\triangle t)$；

(4) 由全概率之和为 1，则既没有细菌分裂，又没有细菌死亡的概率为：

$$1-\lambda\triangle t-\mu\triangle t-o(\triangle t)$$

当 $n=0$ 时，即系统中一个细菌也没有时，在 $\triangle t$ 内不会有细菌死亡，故此时 $\mu\triangle t=0$。于是，

$$P_0(t+\triangle t)=P_0(t)[1-\lambda\triangle t-o(\triangle t)]+P_1(t)[\mu\triangle t+o(\triangle t)]$$

当 $n>0$ 时，即系统中至少有一个细菌时，在 $\triangle t$ 内既可能有细菌增生，其概率为 $\lambda\triangle t+o(\triangle t)$，也可能有细菌死亡，其概率为 $\mu\triangle t+o(\triangle t)$，还可能既不增生也不死亡，其概率为 $1-\lambda\triangle t-\mu\triangle t-o(\triangle t)$。于是：

$$\begin{aligned}P_n(t+\triangle t)=&P_{n-1}(t)[\lambda\triangle t+o(\triangle t)]+P_{n+1}(t)[\mu\triangle t+o(\triangle t)]\\&+P_n(t)[1-\lambda\triangle t-\mu\triangle t-o(\triangle t)]\end{aligned}$$

从而，将前面两个式子变形，并取极限，得：

$$\begin{aligned}\frac{\mathrm{d}P_0(t)}{\mathrm{d}t}&=\frac{\lim\limits_{\triangle t\to 0}[P_0(t+\triangle t)-P_0(t)]}{\triangle t}\\&=\mu P_1(t)-\lambda P_0(t)\ ;\end{aligned}\tag{10.16}$$

$$\begin{aligned}\frac{\mathrm{d}P_n(t)}{\mathrm{d}t}&=\frac{\lim\limits_{\triangle t\to 0}[P_n(t+\triangle t)-P_n(t)]}{\triangle t}\\&=\lambda P_{n-1}(t)+\mu P_{n+1}(t)-(\lambda+\mu)P_n(t)\ ,\quad n>0\end{aligned}\tag{10.17}$$

(10.16) ~ (10.17) 式所表示的是一组微分方程，又称差分微分方程。因为它是从细菌的增生和死亡的假设条件推导出来的，又称它是生灭过程微分方程组。可以验证，这组微分方程所描述的，就是泊松流输入，负指数分布的服务时间，只有一个服务台的排队系统，更准确地说，就是 $[M/M/1]:[\infty/\infty/FCFS]$ 排队系统。

这组微分方程的解称为瞬态解。因求瞬态解比较麻烦，而且在工程实际中应用较少，我们省略了。

现在，只研究稳态情形。这时，$Pn(t)$ 与时间无关，故写成 $P(n)$，对 t 的导数为零。一般情况下，系统运行相当时间后，就可进入稳态运行，即 $P(n)$ 趋于一个常数。稳态解是我们最关心的问题，服务设施的设置就是根据稳态的顾客情况来考虑的。稳态解又比较容易求出，所以，我们重点研究排队系统的稳态情形。

在稳态时，从 (10.16) 和 (10.17) 式得到：

$$0=-\lambda P(0)+\mu P(1)$$

$$0=\lambda P(0)+\mu P(2)-(\lambda+\mu)P(1)$$

$$0=\lambda P(1)+\mu P(3)-(\lambda+\mu)P(2)$$

$$\cdots$$

$$0=\lambda P(n-1)+\mu P(n+1)-(\lambda+\mu)P(n),n\geqslant 1$$

从这些式子,可顺次求得:

$$P(1)=\frac{\lambda}{\mu}P(0)$$

$$P(2)=\frac{\lambda}{\mu}2P(0)$$

事实上,

$$P(2)=\frac{\lambda+\mu}{\mu}P(1)-\frac{\lambda}{\mu}P(0)=\left(\frac{\lambda}{\mu}+1\right)\cdot\frac{\lambda}{\mu}P(0)-\frac{\lambda}{\mu}P(0)=\left(\frac{\lambda}{\mu}\right)^2P(0)$$

$$\cdots$$

$$P(n)=\frac{\lambda}{\mu}nP(0),n\geqslant 1$$

事实上,

$$\begin{aligned}P(n+1)&=\frac{\lambda+\mu}{\mu}P(n)-\frac{\lambda}{\mu}P(n-1)\\&=\left(\frac{\lambda}{\mu}+1\right)\cdot\left(\frac{\lambda}{\mu}\right)^nP(0)-\frac{\lambda}{\mu}\cdot\left(\frac{\lambda}{\mu}\right)^{n-1}P(0)=\left(\frac{\lambda}{\mu}\right)^{n+1}P(0)\end{aligned}$$

由 $\sum_{n=0}^{\infty}P(n)=1$,得到:

$$P(0)\sum_{n=0}^{\infty}\left(\frac{\lambda}{\mu}\right)^n=1$$

或 $\left(\rho=\frac{\lambda}{\mu}\right)$

$$P(0)\sum_{n=0}^{\infty}\rho^n=1$$

所以,

$$\begin{aligned}P(0)&=\frac{1}{\sum_{n=0}^{\infty}\rho^n}\\&=\frac{1}{1/(1-\rho)}\quad(\text{当 }0\leqslant\rho<1\text{ 时})\\&=1-\rho。\end{aligned}$$

由此,又有:

$$P(n) = \rho^n(1-\rho), n \geqslant 1, 0 \leqslant \rho < 1 \tag{10.18}$$

2. **状态转移图**

稳态方程可以通过所谓状态转移图列出，见图 10.7，然后求解。

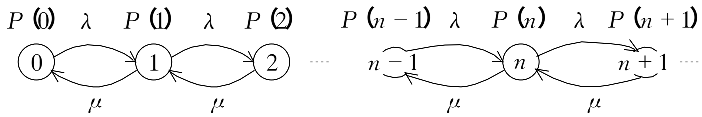


图 10.7

图中的圆圈表示状态，圆圈中的标号是状态符号，它表示系统中稳态顾客数。图中的箭头表示从一个状态到另一个状态的转移，箭头上面的符号表示状态转移速率。如由状态到状态的转移速率为 λ**，表示系统由** 0 **个顾客到** 1 **个顾客的到达速率为** λ**，而由** 1 **个顾客到** 0 **个顾客的服务速率为** μ**。系统处于稳态时，对每个状态来说，转入率应等于转出率。例如，对于状态** n **来说，转入率是** $\lambda P(n-1) + \mu P(n+1)$**，而转出率是** $\lambda P(n) + \mu P(n) = (\lambda + \mu) P(n)$**，因此，**

$$\lambda P(n-1) + \mu P(n+1) = (\lambda + \mu) P(n), n \geqslant 1$$

这个方程称为平衡方程。对状态 0 **来说，**

$$\lambda P(0) = \mu P(1)$$

它们和前面的结果是等价的。这样，利用状态转移图得到稳态方程。对于一些比较复杂的排队问题，用状态转移图列稳态方程有时是比较方便的。由稳态方程求解稳态概率的方法和前面一样。

3. **系统运行指标** L, L_q, W, W_q **的求解**

根据稳态时系统状态为 n **的概率，可以计算出稳态时系统的运行指标，它们是**

(1) **在系统中的平均顾客数或顾客数的期望值** L**。**

由离散型随机变量的期望值定义

$$L = \sum_{n=0}^{\infty} nP(n)$$

将 (10.18) **式结果代入，得到**

$$\begin{aligned} L &= \sum_{n=0}^{\infty} n\rho^n(1-\rho) \\ &= \sum_{n=0}^{\infty} n(\rho^n - \rho^{n+1}) \\ &= \sum_{n=1}^{\infty} \rho^n \\ &= \frac{\rho}{1-\rho} \quad 0 \leqslant \rho < 1 \end{aligned}$$

将 $\rho=\dfrac{\lambda}{\mu}$ 代入，得

$$L=\frac{\lambda}{\mu-\lambda} \tag{10.19}$$

(2) 在队列中等待服务的顾客的期望值 L_q。

对于单服务台的排队系统，当没顾客时，当然不存在排队现象。当系统中有顾客时，排队等待的顾客数，必定比系统中顾客总数少一个。如果系统中有 n 个顾客，那么，排队等待的顾客数为 $n-1$。所以，排队等待的顾客数的期望值：

$$\begin{aligned}L_q&=\sum_{n=1}^{\infty}(n-1)P(n)\\&=\sum_{n=1}^{\infty}nP(n)-\sum_{n=1}^{\infty}P(n)\end{aligned}$$

而

$$\sum_{n=1}^{\infty}nP(n)=\sum_{n=0}^{\infty}nP(n)=L$$

$$\sum_{n=1}^{\infty}P(n)=1-P(0)=1-(1-\rho)=\rho$$

从而

$$L_q=L-\rho=\frac{\lambda}{\mu-\lambda}-\frac{\lambda}{\mu}=\frac{\lambda^2}{\mu(\mu-\lambda)}=\frac{\rho^2}{1-\rho} \tag{10.20}$$

看出，排队等待的顾客数的期望值 L_q 比系统中的顾客数的期望值 L 少 ρ。

(3) 顾客在系统中全部时间 (逗留时间) 的期望值 W。

顾客在系统中的时间 T_s 是一个随机变量。在本模型中，它服从参数为 $\mu-\lambda$ 的负指数分布 (证明从略)。

分布函数

$$F(t)=1-e^{-(\mu-\lambda)t},t\geqslant 0$$

密度函数

$$f(t)=(\mu-\lambda)e^{-(\mu-\lambda)t},t\geqslant 0$$

所以，在系统中全部时间 T_s 的期望值。

$$W=E[T_s]=\frac{1}{\mu-\lambda} \tag{10.21}$$

(4) 在队列中排队等待时间的期望值 W_q

在队列中排队等待时间的期望值，应等于在系统中全部时间的期望值，减去服务时间的期望值，而服务时间的期望值是 $\dfrac{1}{\mu}$，故

$$W_q = W - \frac{1}{\mu} = \frac{1}{\mu - \lambda} - \frac{1}{\mu} = \frac{\rho}{\mu - \lambda} \tag{10.22}$$

公式(10.18)－(10.22),是我们所讨论模型的基本公式,读者应当理解,并用此进行计算。

例1　小汽车作过境检查,到达平均速率为100辆/小时,是泊松流;检查一辆汽车平均需要15秒钟,为负指数分布。试求稳态概率$P(0)$、$P(1)$、$P(2)$和系统中汽车数的期望值L,排队等待的汽车数的期望值L_q,过境检查全部时间的期望值W,等待检查时间的期望值W_q。

解　由给定条件,$\lambda = 100$(辆/小时),$\mu = 4$(辆/分)$= 240$(辆/小时)。所以,

$$\rho = \frac{100}{240} = \frac{5}{12}$$

$$P(0) = 1 - \rho = 1 - \frac{5}{12} = 0.5833$$

$$P(1) = \rho P(0) = \left(\frac{5}{12}\right)\left(1 - \frac{5}{12}\right) = 0.2431$$

$$P(2) = \rho^2 P(0) = \left(\frac{5}{12}\right)^2\left(1 - \frac{5}{12}\right) = 0.1013$$

$$L = \frac{\lambda}{\mu - \lambda} = \frac{100}{240 - 100} = 0.714\text{(辆)}$$

$$L_q = \frac{\lambda^2}{\mu(\mu - \lambda)} = \frac{100^2}{240 \times (240 - 100)} = 0.298\text{(辆)}$$

$$W = \frac{1}{\mu - \lambda} = \frac{1}{240 - 100} = 0.007\ 14\text{(小时)} = 25.7\text{(秒)}$$

$$W_q = \frac{\lambda}{\mu(\mu - \lambda)} = \frac{100}{240 \times (240 - 100)} = 0.003\text{(小时)} = 10.7\text{(秒)}$$

4.L、L_q与ρ的关系讨论

由

$$L = \frac{\rho}{1 - \rho}$$

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

这两个指标都只同ρ有关,或者说都只与比值$\frac{\lambda}{\mu}$有关,而与λ,μ的绝对值大小没有关系。这同人们的生活经验是一致的:顾客来的虽多,如果服务得快,那么排队仍然不会长。反之,即使来的顾客不多,但如果服务得慢,那么排队仍然会长。

又

$$L - L_q = \rho, 0 \leqslant \rho < 1$$

即 L 与 L_q 之差恰等于服务强度 ρ。因此，对于 $[M/M/1]:[\infty/\infty/FCFS]$ 系统，可以取不同的 ρ 值，绘出 $\rho - L, L_q$ 曲线，如图 10.8。

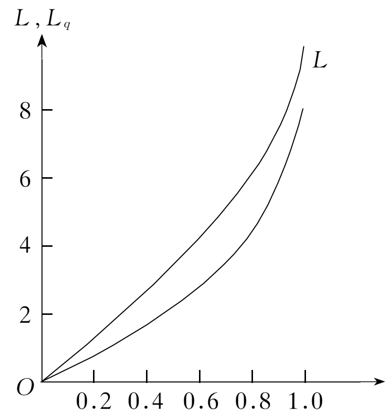


图 10.8

根据曲线，若已知某一系统的 ρ 值，立刻可求出 L 和 L_q 值。如已知 $\rho = 0.5$，从图上可求出 $L = 1, L_q = 0.5$。

反之，如果为了把排队等待的顾客数的期望值 L_q 限制在某一数值以内，那么也可以立刻求出 ρ 值必须限制在何值以内。如已知 $L \leqslant 5$，则由曲线立刻可求出 $\rho \leqslant 0.85$，即为了避免排队长度超过 5 人，还必须给服务员留下 15% 的空闲时间。当然，排队长度期望值为 5，表明有时将少于 5 人，有时会多于 5 人。从曲线还可看出，当 ρ 值在 0.7 以上变化时，引起 L 和 L_q 的变化是很敏感的。

根据曲线，还可深入思考一下随机型模型和确定型模型之间的差异。对确定型模型，即等间隔到达的顾客流，等服务时间的服务规律，如果顾客到达的速率同服务速率相等，系统将达到平衡，即顾客不用等待就能得到服务。而对于随机型模型，情况就不同了。如果到达速率趋近于服务速率，即 $\rho \to 1$，则系统中顾客数的期望值 L 和排队顾客数的期望值 L_q 将趋于无穷。显然，如果希望服务员达到更高的服务强度，那么等待的队列就会比较长；而如果希望等待的队列比较短，那么服务员的服务强度就会低一些，即将有较多的空闲时间。比较理想的服务强度的选择应当兼顾这两个方面，即服务强度不太小，而排队的人数又不太多。

5. L, L_q，和 W, W_q 之间的关系 —— 里特公式

人们发现，通常计算 L 和 L_q 比计算 W 和 W_q 要容易些。那么，是不是可以根据已经计算出来的 L 和 L_q 直接推算 W 和 W_q 呢？就是说，L, L_q 和 W, W_q 之间有没有一种简单关系呢？

里特(Little,J.D.C.)证明了,如果 λ_e 为顾客到达系统中的速率的平均值,那么,在很宽的条件下,都有以下关系式:

$$W = \frac{L}{\lambda_e} \tag{10.23}$$

和

$$W_q = \frac{L_q}{\lambda_e} \tag{10.24}$$

由于每个顾客的平均服务时间为$\frac{1}{\mu}$,那么,W 和 W_q 的差别在于平均服务时间

$$W = W_q + \frac{1}{\mu} \tag{10.25}$$

在 [M/M/1]:[∞/∞/FCFS] 系统中,由于顾客到达的速率是一常数 λ,而系统可容纳的顾客数又没有限制,即在各种状态下到达系统的顾客的速率均为 λ,这时

$$\lambda_e = \lambda$$

如果对于不同的状态 n,顾客到达的速率 λ_n,不是常数,则可用公式:

$$\lambda_e = \sum_{n=0}^{\infty} \lambda_n P(n) \tag{10.26}$$

计算。

公式(10.23)–(10.26),称为里特公式。

10.3.2 [M/M/1]:[N/∞/FCFS] 系统

这种模型是泊松到达过程,服务时间服从负指数分布,一个服务台,系统内只允许有 N 个顾客(即正在接受服务的和排队等待的总人数不得超过 N 个),顾客源是无限的,先到先服务制。模型示意如图 10.9。

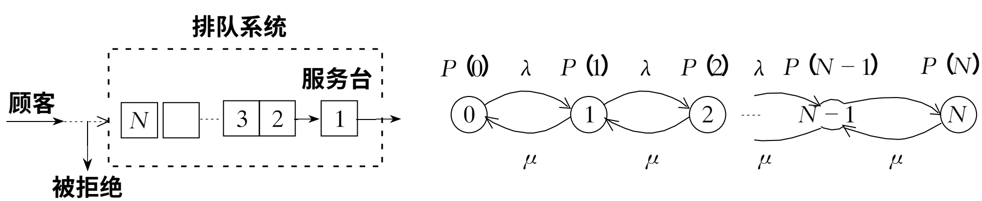


图 10.9 [M/M/1]:[N/∞/FCFS]系统及其状态转移图

1. 系统稳态概率的求解

对于任一状态,转入率应等于转出率。据此,得到各状态的平衡方程:

对状态 0,

$$\lambda P(0)=\mu P(1)$$

对状态 i，

$$\lambda P(i-1)+\mu P(i+1)=(\lambda+\mu)P(i),i=1,2,\cdots,N-1$$

对状态 N，

$$\lambda P(N-1)=\mu P(N)$$

由这些平衡方程，可得到：

$$P(n)=\rho^n P(0),n=1,2,\cdots,N$$

代入 $\sum_{n=0}^{N}P(n)=1$，而有

$$P(0)=\frac{1}{\sum_{n=0}^{N}\rho^n}=\begin{cases}\dfrac{1-\rho}{1-\rho^{N+1}} & \rho\neq 1\\ \dfrac{1}{N+1} & \rho=1\end{cases}$$

从而

$$P_n=\begin{cases}\dfrac{\rho^n(1-\rho)}{1-\rho^{N+1}} & \rho\neq 1\\ \dfrac{1}{N+1} & \rho=1\end{cases}$$

2. 系统稳态运行指标的计算

(1) 系统中顾客数的期望值 L

$$\begin{aligned}L&=\sum_{n=0}^{N}nP(n)=\sum_{n=0}^{N}n\rho^n P(0)\\&=\rho P(0)\sum_{n=1}^{N}n\rho^{n-1}\\&=\rho P(0)\frac{\mathrm{d}\sum_{n=1}^{N}\rho^n}{\mathrm{d}\rho}\\&=\rho P(0)\frac{\mathrm{d}[(\rho-\rho^{N+1})/(1-\rho)]}{\mathrm{d}\rho}\\&=\rho\left(\frac{1-\rho}{1-\rho^{N+1}}\right)\cdot\frac{\mathrm{d}[(\rho-\rho^{N+1})/(1-\rho)]}{\mathrm{d}\rho}\\&\cdots\\&=\frac{\rho}{1-\rho}-\frac{(N+1)\rho^{N+1}}{1-\rho^{N+1}},\rho\neq 1\end{aligned}\qquad(10.27)$$

$$L=\sum_{n=0}^{N}nP(n)$$
$$=\frac{1}{N+1}\sum_{n=1}^{N}\cdot n$$
$$=\frac{1}{N+1}\cdot\frac{N(N+1)}{2}=\frac{N}{2}\quad \rho=1 \tag{10.28}$$

在(10.27)式中,令 $N\to\infty$,有 $L=\frac{\rho}{1-\rho}$——[M/M/1]:[∞/∞/FCFS]系统。

其余几个运行指标须借助于里特公式求解。不过要注意,里特公式中的 λ_e 是有效到达速率,即实际到达系统的顾客的到达速率。在有顾客数限制的系统中,系统"客满"时,新到达的顾客便不能进入系统排队,只能离去,因此,在计算有效到达率时,认为离去的这一部分顾客到达率为 0。

对本模型,"客满"即顾客为 N 时,新到达的顾客离去,故有效到达率为:

$$\lambda_e=\lambda\sum_{n=0}^{N-1}P(n)+0\cdot P(N)=\lambda[1-P(N)]$$

利用里特公式得:

(2) $$W=\frac{L}{\lambda_e}=\frac{L}{\lambda[1-P(N)]} \tag{10.29}$$

(3) $$L_q=L-\frac{\lambda_e}{\mu}=L-\frac{\lambda[1-P(N)]}{\mu} \tag{10.30}$$

(4) $$W_q=\frac{L_q}{\lambda_e}=\frac{L}{\lambda[1-P(N)]}-\frac{1}{\mu} \tag{10.31}$$

例 2　某单人理发店内有 4 把椅子接待人们排队等待理发。当 4 把椅子都坐满时,后来的顾客就不进店而离开。顾客平均到达率为 4 人/小时,理发时间平均 10 分钟。设到达过程为泊松流,服务时间服从负指数分布,求:① 顾客一到达就能理发的概率;② 系统中顾客数的期望值 L 和排队等待的顾客数的期望值 Lq;③ 顾客在理发店内逗留的全部时间的期望值 W;④ 在可能到达的顾客中因客满离开的概率。

解　由题设,这是一个[M/M/1]:[N/∞/FCFS]系统。

总容量(加上供理发的一个椅子)为

$$N=4+1=5$$

又

$$\lambda=4\text{(人/小时)},\mu=6\text{(人/小时)},\rho=\frac{4}{6}=\frac{2}{3}$$

① 顾客一到就理发,即系统中没有顾客,这个概率

$$P(0)=\frac{1-\rho}{1-\rho^{N+1}}=\frac{1-2/3}{1-(2/3)^{5+1}}=0.365$$

②$L=\dfrac{\rho}{1-\rho}-\dfrac{(N+1)\rho^{N+1}}{1-\rho^{N+1}}$

$$=\frac{2/3}{1-(2/3)}-(5+1)\left(\frac{2}{3}\right)^5+\frac{1}{1-(2/3)^{5+1}}=1.423\text{（人）}$$

$$\lambda_e=\lambda[1-P(5)]=\lambda[1-\rho^5P(0)]=4\times\left[1-\left(\frac{2}{3}\right)^5\times0.365\right]=3.808\text{（人／小时）}$$

$$L_q=L-\frac{\lambda_e}{\mu}=1.423-\frac{3.808}{6}=0.788\text{（人）}$$

③$W=\dfrac{L}{\lambda_e}=\dfrac{1.423}{3.808}=0.374\text{（小时）}=22.42\text{（分）}$

④ 由题意，因客满而离开的概率，即

$$P(5)=\rho^5P(0)=\left(\frac{2}{3}\right)^5\times0.365=0.048$$

这也是理发店的损失率。

附　以本例比较系统容量有限和无限两种情况的几个指标见表 10.2：

表 10.2

$\lambda=4$（人／小时） $\mu=6$（人／小时）	L （人）	L_q （人）	W （分）	W_q （分）	$P(0)$	可能到来的顾客中 离开的概率
$N=5$	1.423	0.788	22.42	12.42	0.365	4.8%
无限	2.000	1.333	30.00	20.00	0.333	0

由于有 4.8% 的顾客离开了，故容量有限的排队系统中，等待时间要短一些。

10.3.3　[M/M/1]：[N/N/FCFS] 系统

即泊松流到达过程，负指数分布的服务时间，一个服务台，系统容纳的顾客数有限，顾客的总体数有限，先到先服务的系统。顾客总体虽只有 N 个，但每个顾客到来并经过服务后，仍回到原来的总体，所以仍然可以到来，模型如图 10.10。

1. 稳态概率的计算

首先要说明，本模型到达率 λ，其含义同无限顾客源模型中的到达率是不同的。在无限源模型中，是按全体顾客来考虑到达率的，其 λ 表示单位时间内到达的顾客数。在本模型中，λ 表示单位时间内到达某一顾客的速率。比如，机修模型中，一台机器修理完毕开始工作，到再次出现故障，平均$\dfrac{1}{\lambda}$ 天，那么，λ 是一台机器到达排队系统要求修理的平均速率。假设一台机器由开始运行到再次出现故障，平均为 20 天，那么

$$\frac{1}{\lambda}=20\text{（天）}$$

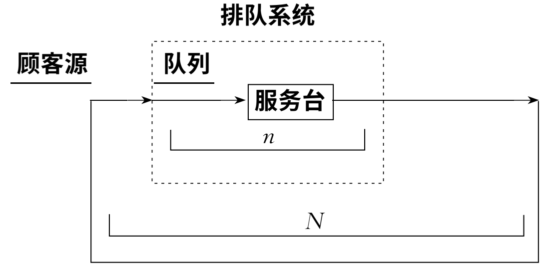


图 10.10

或到达速率

$$\lambda = \frac{1}{20}\text{(台／天)}$$

设某工厂总共有 N 台机器，如果只有一台机器在排队系统之外，即在正常工作，那么到达排队系统要求修理的速率为 λ。如果有 2 台机器在排队系统之外，且两台到达的速率都相等，则机器到达排队系统的速率为 2λ。显然，在排队系统外正常运行的机器的数量，就是将要到达排队系统要求修理的机器的数量。设排队系统内的机器数为 n，那么，在排队系统外正常运行的机器的数量就是 $N-n$，于是，进入排队系统的机器的速率为

$$\lambda_n = (N-n)\lambda,\quad n \leqslant N \tag{10.32}$$

$$\lambda_n = 0, n \geqslant M$$

上面两式中，λn 表示排队系统中有 n 个顾客时的到达速率，这个速率是随着系统中的顾客数的变化而变化的。系统中的顾客数越多，顾客到达的速率越小，而系统中的顾客数越少，顾客到达的速率就越大。这种情形同顾客源无限的模型是不一样的。在那类模型中，认为顾客到达的速率是一个常数 λ，而同排队系统中的顾客数无关。服务速率，仍然假定对所有的顾客都是 μ。

由顾客到达速率的计算和服务速率的假定，可以画出本模型状态转移图：

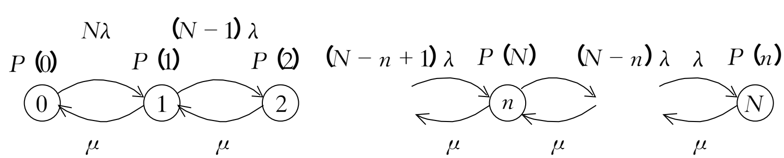


图 10.11

稳态下对每个状态都达到平衡，转入率等于转出率：

对状态 0，

$$N\lambda P(0) = \mu P(1)$$

对状态 n，

$$[(N-n)\lambda + \mu]P(n) = (N-n+1)\lambda P(n-1) + \mu P(n+1),$$
$$n = 1,\cdots,N-1$$

对状态 N，

$$\mu P(N) = \lambda P(N-1)$$

利用 $\sum_{n=0}^{N} P(n) = 1$,得到：

$$P(0) = \frac{1}{\sum_{n=0}^{N} (\lambda/\mu)^n [N!/(N-n)!]}$$

$$P(n) = \left(\frac{\lambda}{\mu}\right)^n \left[\frac{N!}{(N-n)!}\right] P(0), \quad 0 < n \leqslant N \tag{10.33}$$

2. 运行指标的计算

在计算运行指标之前,必须先算出系统中顾客有效到达速率 λ_e,它是一个期望值。

$$\lambda_e = E[(N-n)\lambda] = E[N\lambda - n\lambda] = E[N\lambda] - E[n\lambda]$$

上式中第一项是常数,又 $E(n) = L$—— 排队中顾客数的期望。于是：

$$\lambda_e = N\lambda - \lambda E(n) = N\lambda - \lambda L = \lambda(N-L)$$

由里特公式，

$$L_q = L - \frac{\lambda_e}{\mu} = L - \frac{\lambda(N-L)}{\mu}$$

注意到

$$L = L_q + 1 - P(0)$$

将上面两式联立解之,得：

$$L = N - \frac{\mu}{\lambda} + \left(\frac{\mu}{\lambda}\right) P(0) \tag{10.34}$$

$$L_q = N - \frac{\mu+\lambda}{\lambda} \cdot [1 - P(0)] \tag{10.35}$$

再由里特公式，

$$W = \frac{L}{\lambda_e} = \frac{L}{\lambda(N-L)} \tag{10.36}$$

$$W_q = \frac{L_q}{\lambda_e} = \frac{L_q}{\lambda(N-L)} \tag{10.37}$$

例 3　一台机修工负责 3 台机器的维修工作。设每台机器在维修之后平均可运行 5 天,而平均维修一台机器的时间为 2 天。试求稳态下的各状态概率和各运行指标。

解 由题设，$\lambda = \frac{1}{5}$（台/天），$\mu = \frac{1}{2}$（台/天），$N = 3$，$\frac{\lambda}{\mu} = \frac{2}{5}$

由（10.33）式，

$$P(0) = \{\sum_{n=0}^{3}(\frac{\lambda}{\mu})^n[\frac{3!}{(3-n)!}]\}^{-1}$$
$$= [(\frac{2}{5})^0(\frac{3!}{3!}) + (\frac{2}{5})^1(\frac{3!}{2!}) + (\frac{2}{5})^2(\frac{3!}{1!}) + (\frac{2}{5})^3(\frac{3!}{0!})]^{-1}$$
$$= 0.282$$

$$P(1) = (\frac{\lambda}{\mu})^1(3-0)P(0) = \frac{2}{5}\times 3\times 0.282 = 0.339$$

$$P(2) = (\frac{\lambda}{\mu})^2(3-0)(3-1)P(0) = (\frac{2}{5})^2\times 3\times 2\times 0.282 = 0.271$$

$$P(3) = (\frac{\lambda}{\mu})^3(3-0)(3-1)(3-2)P(0)$$
$$= (\frac{2}{5})^3\times 3\times 2\times 1\times 0.282 = 0.108$$

由（10.34）-（10.37）式，

$$L = N - \frac{\mu}{\lambda} + \frac{\mu}{\lambda}P(0) = 3 - \frac{5}{2} + \frac{5}{2}\times 0.282 = 1.205\text{（台）}$$

$$L_q = N - (\frac{\mu-\lambda}{\lambda})[1-P(0)] = 3 - (\frac{7}{2})\times(1-0.282)$$
$$= 0.487\text{（台）}$$

$$W = \frac{L}{\lambda(N-L)} = \frac{1.205}{0.2\times(3-1.205)} = 3.36\text{（天）}$$

$$W_q = \frac{L_q}{\lambda(N-L)} = \frac{0.487}{0.2\times(3-1.205)} = 1.36\text{（天）}$$

附 单服务台排队模型的计算公式见表 10.3。

表 10.3

参数	[M/M/1]：[∞/∞/FCFS]	[M/M/1]：[N/∞/FCFS]	[M/M/1]：[N/N/FCFS]
$P(0)$	$1-\rho$	$\frac{1-\rho}{1-\rho^{N+1}}$	$[\sum_{n=0}^{N}(\frac{\lambda}{\mu})^n\frac{N!}{(N-n)!}]^{-1}$
$P(n)$	$\rho^n(1-\rho)$	$\frac{\rho^n(1-\rho)}{1-\rho^{N+1}h}$	$(\frac{\lambda}{\mu})^n[\frac{N!}{(N-n)!}]P(0)$
L	$\frac{\rho}{1-\rho}$	$\frac{\rho}{1-\rho} - \frac{(N+1)\rho^{N+1}}{(1-\rho^{N+1})}$	$L_q + 1 - P(0)$
L_q	$\frac{\rho^2}{1-\rho}$	$L-\rho[1-P(N)]$	$N-(\frac{\mu+\lambda}{\lambda})[1-P(0)]$

表 10.3(续)

参数	$[M/M/1]:[\infty/\infty/FCFS]$	$[M/M/1]:[N/\infty/FCFS]$	$[M/M/1]:[N/N/FCFS]$
W	$\dfrac{1}{\mu-\lambda}$	$\dfrac{L}{\lambda[1-P(N)]}$	$\dfrac{L}{\lambda(N-L)}$
W_q	$\dfrac{\lambda}{\mu(\mu-\lambda)}$	$\dfrac{L_q}{\lambda[1-P(N)]}$	$\dfrac{L_q}{\lambda(N-L)}$
备注	$\rho=\dfrac{\lambda}{\mu}<1$	$\rho=\dfrac{\lambda}{\mu}\neq 1$,$\lambda_e=\lambda[1-P(N)]$	$\lambda_e=\lambda(N-L)$

10.4 单服务台的 $[M/G/1]$ 模型

上一节讨论的单服务台排队系统都是 $[M/M/1]$ 模型。根据统计资料和理论推断,通常,顾客到达还是比较符合某一参数的泊松分布的,但服务时间的分布则要取决于服务机构状况和顾客状况等多种因素,它并不一定符合负指数分布。因此,对于排队系统的研究,如果仅仅局限于负指数分布的服务时间,就不够了。这就是一般服务时间分布,单服务台的 $[M/G/1]$ 模型研究的实际背景。

10.4.1 $[M/G/1]:[\infty/\infty/G]$ 系统

对于 $[M/G/1]$ 模型,服务时间 v 的分布函数是任意函数,其他条件和 $[M/M/1]$ 系统相同。这里,要求服务时间的期望值 $E(v)$ 和方差 $D(v)$ 存在,并为已知,且为了达到稳态,服务强度 $\rho<1$,其中 $\rho=\lambda E(v)$。

在上述条件下,我们有

$$L=\rho+\frac{\rho^2+\lambda^2 D(v)}{2(1-\rho)} \tag{10.38}$$

或

$$L=\lambda E(v)+\frac{\lambda^2\{[E(v)]^2+D(v)\}}{2[1-\lambda E(v)]} \tag{10.39}$$

这就是布拉切克—欣钦(Pollaczek-Khintchine)公式,又叫($P-K$)公式。我们引用它。这样,只要知道 λ,$E(v)$ 和 $D(v)$,不管 v 是什么分布,就可以求出系统中顾客数的期望值 L。由公式可以看出,L 不仅同顾客到达的速率 λ 和服务时间的期望值 $E(v)$ 有关,而且同服务时间的方差 $D(v)$ 有关。方差越大,L 值就越大。所以,要想改进系统的运行指标,除考虑服务时间的期望值 $E(v)$ 之外,还要考虑改变方差 $D(v)$。

$[M/G/1]$ 模型中,里特公式仍然适用,它的形式为:

$$W = \frac{L}{\lambda} \tag{10.40}$$

$$W_q = \frac{L_q}{\lambda} \tag{10.41}$$

$$W = W_q + E(v) \tag{10.42}$$

利用 $[P - K]$ 公式和里特公式,我们可以求出系统的运行指标。

最后要提到的,本模型的服务规则不写成 $FCFS$,而写成 G,即任意规则。原因在于如果仅仅考虑系统运行指标的期望值,而不考虑其他因素,那么,可以证明,这些期望值同服务规则无关。因此就没有必要限制服务规则,以后的模型都不作这一限制。

例 4　某修理店只有一个工人,每小时平均有 4 个顾客带来器具要求修理。这个工人为一个顾客服务平均需 0.1 小时,方差 $D(v) = \frac{1}{8}$ (小时/人)^2。若顾客到达为泊松流,求:① 系统中的顾客数 L;② 排队等待的顾客数 L_q;③ 顾客在系统中的全部时间 W;④ 顾客排队等待的时间 W_q。

解　这是一个 $[M/G/1]$ 模型。

$\lambda = 4$ (人/小时),$E(v) = 0.1$,$D(v) = \frac{1}{8}$ (小时/人)^2

$$\text{标准差} \sqrt{D(v)} = 0.353\,553\,4$$

由 $(P - K)$ 公式,求得:

$$① L = \lambda E(v) + \frac{\lambda^2 \{[E(v)]^2 + D(v)\}}{2[1 - \lambda E(v)]}$$

$$= 4 \times 0.1 + \frac{4^2 (0.1^2 + 1/8)}{2(1 - 4 \times 0.1)} = 2.2 \text{ (人)}$$

$$② L_q = L - \lambda E(v) = 2.2 - 4 \times 0.1 = 1.8 \text{ (人)}$$

$$③ W = \frac{L}{\lambda} = \frac{2.2}{4} = 0.55 \text{ (小时)}$$

$$④ W_q = \frac{L_q}{\lambda} = \frac{1.8}{4} = 0.45 \text{ (小时)}$$

10.4.2　$[M/D/1]$:$[\infty/\infty/G]$ 系统

这是定长服务时间的模型。通常,在自动装配线上完成一件工作的时间可以认为是定长时间。自然,可以认为 $[M/D/1]$ 模型是 $[M/G/1]$ 模型的特例。这时,服务时间方差 $D(v) = 0$。

用 $(P - K)$ 公式求解:

$$L = \lambda E(v) + \frac{\lambda^2\{[E(v)]^2 + D(v)\}}{2[1-\lambda E(v)]}$$

$$= E(v) + \frac{\lambda^2[E(v)]^2}{2[1-\lambda E(v)]}$$

$$L_q = L - \lambda E(v) = \frac{\lambda^2[E(v)]^2}{2[1-\lambda E(v)]}$$

$$W = \frac{L}{\lambda} = E(v) + \frac{\lambda[E(v)]^2}{2[1-\lambda E(v)]}$$

$$W_q = \frac{L_q}{\lambda} = \frac{\lambda[E(v)]^2}{2[1-\lambda E(v)]}$$

在 $[M/M/1]:[\infty/\infty/G]$ 系统中，$E[v] = \frac{1}{\mu}$，

$$W_q = \frac{\lambda}{\mu(\mu-\lambda)} = \frac{\lambda[E(v)]^2}{1-\lambda E(v)}$$

比较这两个结果，可知在其他条件都相同的情况下，$[M/M/1]:[\infty/\infty/G]$ 模型中的等待时间的期望值 W_q 是 $[M/D/1]:[\infty/\infty/G]$ 模型中的等待时间的期望值 W_q 的两倍。

10.4.3 $[M/E_k/1]:[\infty/\infty/G]$ 系统

这是服务时间服从 k 阶爱尔朗分布的单服务台系统。这类系统的示意图如图 10.12。它表明每个顾客必须经过 k 个服务站，在每个服务站的服务时间 v_i 相互独立，并服从相同的参数为 $k\mu$ 的负指数分布。那么，总的服务时间 $v = \sum_{i=1}^{k} v_i$ 服从k 阶爱尔朗分布。

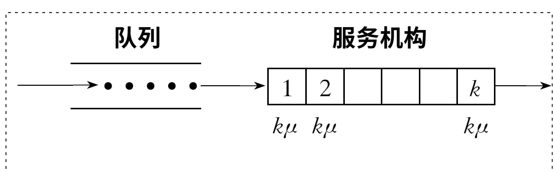


图 10.12

由前面的结果知道，

$$E(v) = \frac{1}{\mu}$$

$$D(v) = \frac{1}{k\mu^2}$$

在计算时，可以认为 $[M/E_k/1]$ 模型是 $[M/G/1]$ 模型的特例，用布拉切克—欣钦公式：

$$\begin{aligned} L &= \lambda E(v) + \frac{\lambda^2\{[E(v)]^2 + D(v)\}}{2[1 - \lambda E(v)]} \\ &= \frac{\lambda}{\mu} + \frac{\lambda^2(1/\mu^2 + 1/k\mu^2)}{2(1 - \lambda/\mu)} \\ &= \rho + \frac{(k+1)\rho^2}{2k(1-\rho)} \end{aligned}$$

由里特公式：

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

$$W = \frac{L}{\lambda}$$

$$W_q = \frac{L_q}{\lambda}$$

读者可以自己证明，当 $k = 1$ 时，$[M/E_k/1]$ 模型即是 $[M/M/1]$ 模型，而当 $k \to \infty$ 时，$[M/E_k/1]$ 模型即是 $[M/D/1]$ 模型。

例 5　小汽车以 10 辆/小时的速率，按泊松流到达加油站，设服务一辆可划分为 5 个子工作：① 询问加油数量；② 检查油箱；③ 开泵加油；④ 加足给定数量，关泵；⑤ 向司机收款。设每个子工作为负指数分布，且相互独立，均值为 $\frac{1}{60}$ 时/辆，试求各运行指标。

解　由题设，本系统可看作 $[M/E_k/1]:[\infty/\infty/G]$ 模型。$k = 5$，$\frac{1}{k\mu} = \frac{1}{60}$。故 $\mu = 12$，$\lambda = 10$

$$L = \frac{10}{12} + \frac{10^2[(1/12)^2 + 1/(5\times 12^2)]}{2(1 - 10/12)} = 3.33\ (\text{辆})$$

$$L_q = L - \frac{\lambda}{\mu} = 3.33 - \frac{10}{12} = 2.5\ (\text{辆})$$

$$W = \frac{L}{\lambda} = \frac{3.33}{10} = 0.333\ (\text{小时})$$

$$W_q = W - \frac{1}{\mu} = \frac{1}{3} - \frac{1}{12} = \frac{1}{4}\ (\text{小时})$$

注意　$D(v) = \frac{1}{k\mu^2} = \frac{1}{720}$，$\sqrt{D(v)} = 0.037\,267\,8$

附　本节三种模型运行指标计算公式表（表 10.4）：

表 10.4

参数	$[M/G/1]:[\infty/\infty/G]$	$[M/D/1]:[\infty/\infty/G]$	$[M/E_k/1]:[\infty/\infty/G]$
L	$\lambda E(v)+\dfrac{\lambda^2\{[E(v)]^2+D(v)\}}{2[1-\lambda E(v)]}$	$\dfrac{\lambda}{\mu}+\dfrac{\lambda^2}{2\mu(\mu-\lambda)}$	$\dfrac{\lambda}{\mu}+\dfrac{(k+1)\lambda^2}{2k\mu(\mu-\lambda)}$
L_q	$L-\lambda E(v)$	$\dfrac{\lambda^2}{2\mu(\mu-\lambda)}$	$L-\dfrac{\lambda}{\mu}$
W	$\dfrac{L}{\lambda}$	$\dfrac{L}{\lambda}$	$\dfrac{L}{\lambda}$
W_q	$\dfrac{L_q}{\lambda}$	$\dfrac{L_q}{\lambda}$	$\dfrac{L_q}{\lambda}$
备注			$E(v)=\dfrac{1}{\mu}$,$D(v)=\dfrac{1}{k\mu^2}$

10.5 多服务台的排队模型

10.5.1 $[M/M/C]:[\infty/\infty/G]$ 系统

系统的示意图如图 10.13。

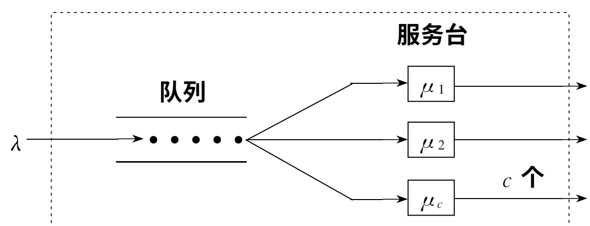


图 10.13 $[M/M/C]:[\infty/\infty/G]$系统

设顾客到达速率为 λ,每个服务台的服务速率均为 μ。则整个系统的最大服务速率为 $C\mu$。令服务强度

$$\rho=\frac{\lambda}{C\mu}$$

显然,欲使系统能稳态运行,必须有

$$\rho<1$$

此模型的特点在于整个系统的服务速率与系统中的顾客数有关。如果系统中只

有一个顾客，则系统的服务速率等于 μ，因为其他服务台处于闲置状态。如果系统中有两个顾客，则系统的服务速率为 2μ。依次类推，如果系统中的顾客达到 C 个，则系统达到最大服务速率 $C\mu$，所以服务台均投入服务。而当系统中的顾客数超过 C 个时，多余的顾客只能进入排队等待服务，系统的服务速率仍为 $C\mu$。得到如下的状态转移图（图 10.14）：

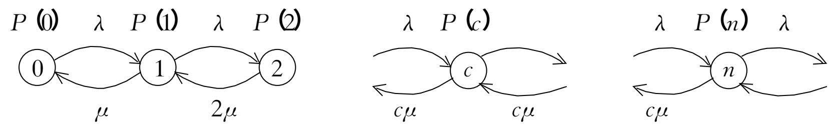


图 10.14

根据状态转移图可以列出平衡方程，由平衡方程和所谓正常化方程（即各状态稳态概率之和为 1），可以求出稳态概率。我们直接给出全部运行指标的计算结果：

$$P(0) = \left[\frac{\sum_{n=0}^{C-1}\lambda^n}{\mu^n n!} + \left(\frac{1}{C!}\right)\left(\frac{\lambda}{\mu}\right)^C\left(\frac{1}{1-\rho}\right)\right]^{-1}$$

$$P(n) = \begin{cases} \left(\dfrac{\lambda^n}{\mu^n n!}\right)P(0), & 1 \leqslant n \leqslant C \\ \dfrac{\lambda^n P(0)}{\mu^C C!\,C^{n-C}}, & n \geqslant C \end{cases}$$

$$L = L_q + \frac{\lambda}{\mu}$$

$$L_q = \frac{\lambda^C \rho P(0)}{\mu^C C!(1-\rho)^2}$$

$$W = \frac{L}{\lambda}$$

$$W_q = \frac{L_q}{\lambda}$$

例 6　有一交通检查站，汽车到达速率为 504 辆/小时，检查站有 3 个检查员，每个检查员以速率 240 辆/小时检查。设到达为泊松流，服务时间服从负指数分布，试求稳态概率 $P(0)$ 和 $P(n)$，以及运行指标。

解　由题设，本题可看作 $[M/M/C]:[\infty/\infty/G]$ 模型：

$$\rho = \frac{\lambda}{C\mu} = \frac{504}{3 \times 240} = 0.7$$

$$\frac{\lambda}{\mu} = \frac{504}{240} = 2.1$$

由前述公式，得：

$$P(0)=\left[\left(1+2.1+\frac{4.41}{2}\right)+\left(\frac{1}{6}\right)(2.1)^3\left(\frac{1}{1-0.7}\right)\right]^{-1}=0.0957$$

$$P(1)=\frac{\lambda}{\mu}P(0)=2.1\times0.0957=0.201$$

$$P(2)=\left(\frac{\lambda^2}{\mu^2 2!}\right)P(0)=\left(\frac{504}{240}\right)^2\times\frac{1}{2}\times0.0957=0.211$$

$$P(3)=\left(\frac{\lambda^3}{\mu^3 3!}\right)P(0)=2.1^3\times\frac{1}{6}\times0.0957=0.148$$

$$P(4)=\left(\frac{\lambda^4}{\mu^4 3!3^{4-3}}\right)P(0)=2.1^4\times\left(\frac{1}{3!3}\right)\times0.0957=0.103$$

先计算运行指标 L_q

$$L_q=\frac{\lambda^C\rho P(0)}{\mu^C C!(1-\rho)^2}=\frac{2.1^3\times0.7\times0.0957}{3!(1-0.7)^2}=1.149$$

$$L=L_q+\frac{\lambda}{\mu}=1.149+2.1=3.249$$

$$W=\frac{L}{\lambda}=\frac{3.249}{504}=0.006\,45\text{（小时）}=23.21\text{（秒）}$$

$$W_q=\frac{L_q}{\lambda}=\frac{1.149}{504}=0.002\,28\text{（小时）}=8.21\text{（秒）}$$

10.5.2 单队多服务台和多个单队单服务台系统的比较

例 7 设有两个业务科室，如生产科和销售科，都要求打印文件和信函。若每个科室平均每天需打印 4 份文件，而一个打字员平均每天可打印 5 份文件。那么，是每个科室都设一名打字员工作效率高呢，还是两个科室合在一起设两名打字员工作效率高呢？也就是说，建立一个单队多服务台系统，还是建立两个单队单服务台系统？

$\lambda=4\rightarrow$ - - - - - - - - □ $\mu=5$

$\lambda=4\rightarrow$ - - - - - - - - □ $\mu=5$

两个单服务台系统

□ $\mu=5$

$\lambda=8\rightarrow$ - - - - - - - - - -

□ $\mu=5$

单队多服务台系统

解 对于两个单队系统，

$$\rho=\frac{\lambda}{\mu}=\frac{4}{5}=0.8$$

等待时间

$$W_q=\frac{\lambda}{\mu(\mu-\lambda)}=\frac{4}{5(5-4)}=0.8\text{（天）}$$

即平均每份文件要等 0.8 天才能开始打印。

对于单队两个服务台系统，

$$\rho = \frac{\lambda}{C\mu} = 2 \times \frac{4}{2 \times 5} = 0.8$$

$$P(0) = \left[1 + \frac{8}{5} + \left(\frac{1}{2}\right)\left(\frac{8}{5}\right)^2\left(\frac{1}{1-0.8}\right)\right]^{-1} = 0.11$$

$$W_q = \frac{\lambda^{C-1}\rho P(0)}{\mu^C C!(1-\rho)^2} = \frac{8^{2-1} \times 0.8 \times 0.11}{5^2 \times 2!(1-0.8)} = 0.35\text{（天）}$$

即平均每份文件只需等待 0.35 天即可打印。很显然，单队多服务台系统比相当的多个单队单服务台系统的工作效率大为提高。也就是说，单队多服务台系统的等待时间必小于相当的多个单队单服务台系统。本例中，如果四个科室集中使用四名打字员，则每份文件的平均等待时间将缩短到 0.16 天。

可惜，这个原则在实际生活和生产过程中，往往由于种种原因，被人们忽视。如同类材料，或同类零件的仓库，常常由于本位主义而分别设立在各个车间，这将造成材料或零件的积压，增加流动资金。如能建立比较集中的零件或材料库，将能显著地减少积压。

读者可将多服务台的三种模型计算公式汇总成表。

10.6 微机操作——调用 YAJ 中，QUEUE 决策支持系统

10.6.1 观察 QUEUE 决策支持系统

这个程序，使你能分析几个排队论问题。有单服务台模型——包括：具有无限或有限排队，具有无限或有限人数，具有不同服务时间分布等等。多服务台模型也可分析。该程序能使用惯用的排队论术语显示分析，并完成度量：到达率、服务率、利用率、系统中的平均顾客数、队伍中的平均顾客数、系统中一个顾客所花费的平均时间、在队列中一个顾客所等待的平均时间以及系统状况的概率。

10.6.2 问题求解

1. 以例 1 为例。

输入数据：

M/M/1

顾客到达率（λ）= 100.000

分布 :*Poisson*

服务台个数 = 1

服务率,每个服务台 = 240.000

分布 :*Poisson*

平均服务时间 = 0.004 (小时)

标准差 = 0.004 (小时)

队列极限 = 无限

顾客人数 = 无限

输出数据:

M/M/1

每小时,具有 (λ) = 100 顾客,和服务率 = 240 顾客,

利用率因子 (ρ) = 0.4166667

系统中顾客的平均人数 (L) = 0.7142857

队列中顾客的平均人数 (L_q) = 0.2976191

系统中一个顾客逗留的平均时间 (W) = 7.142857E − 03

队列中一个顾客等待的平均时间 (W_q) = 2.97619E − 03

所有服务台是闲置的概率 (P_0) = 0.5833334

顾客来到时需要等待的概率 (P_W) = 0.4166667

$P(1) = 0.24306$ $P(2) = 0.10127$ $P(3) = 0.04220$ $P(4) = 0.01758$

$P(5) = 0.00733$ $P(6) = 0.00305$ $P(7) = 0.00127$ $P(8) = 0.00053$

$P(9) = 0.00022$ $P(10) = 0.00009$

$$\sum_{i=1}^{10} P(i) = 0.416601$$

2. 仍以例 2 为例。

输入数据:

M/M/1

顾客到达率 (λ) = 4.000

分布 :*Poisson*

服务台个数 = 1

服务率,每个服务台 = 6.000

分布 :*Poisson*

平均服务时间 = 0.167 (小时)

标准差 = 0.167 (小时)

队列极限 = 5

顾客人数 = 无限

输出数据：

$M/M/1$

每小时，具有 (λ) = 4 顾客，和服务率 = 6 顾客，

利用率因子 (ρ) = 0.6666667

系统中顾客的平均人数 (L) = 1.422557

队列中顾客的平均人数 (L_q) = 0.7879701

系统中一个顾客逗留的平均时间 (W) = 0.3736177

队列中一个顾客等待的平均时间 (W_q) = 0.2069511

所有服务台是闲置的概率 (P_0) = 0.3654135

顾客来到时需要等待的概率 (P_W) = 0.6345865

$P(1) = 0.24361$ $P(2) = 0.16241$ $P(3) = 0.10827$ $P(4) = 0.07218$

$P(5) = 0.004812$ $P(6) = 0.00000$ $P(7) = 0.00000$ $P(8) = 0.00000$

$P(9) = 0.00000$ $P(10) = 0.00000$

3. 以例 4 为例。

输入数据：

$M/G/1$

顾客到达率 (λ) = 4.000

分布：$Poisson$

服务台个数 = 1

服务率，每个服务台 = 10.000

分布：一般的

平均服务时间 = 0.100 (小时)

标准差 = 0.354 (小时)

队列极限 = 无限

顾客人数 = 无限

输出数据：

$M/G/1$

每小时，具有 (λ) = 4 顾客，和服务率 = 10 顾客，

利用率因子 (ρ) = 0.4

系统中顾客的平均人数 (L) = 2.204213

队列中顾客的平均人数 (L_q) = 1.804213

系统中一个顾客逗留的平均时间 (W) = 0.5510534

队列中一个顾客等待的平均时间（W_q）= 0.4510534

所有服务台是闲置的概率（P_0）= 0.6

顾客来到时需要等待的概率（P_W）= 0.4

$P(1) = 0.24000$ $P(2) = 0.09600$ $P(3) = 0.03840$ $P(4) = 0.01536$

$P(5) = 0.00614$ $P(6) = 0.00246$ $P(7) = 0.00098$ $P(8) = 0.00039$

$P(9) = 0.00016$ $P(10) = 0.00006$

$$\sum_{i=1}^{10} P(i) = 0.399958$$

4. **以例** 5 **为例。**

输入数据：

$M/G/1$

顾客到达率（λ）= 10.000

分布：Poisson

服务台个数 = 1

服务率，每个服务台 = 12.000

分布：一般的

平均服务时间 = 0.083（**小时**）

标准差 = 0.037（**小时**）

队列极限 = **无限**

顾客人数 = **无限**

输出数据：

$M/G/1$

每小时，具有（λ）= 10 **顾客，和服务率** = 12 **顾客**

利用率因子（ρ）= 0.8333333

系统中顾客的平均人数（L）= 3.327366

队列中顾客的平均人数（L_q）= 2.494033

系统中一个顾客逗留的平均时间（W）= 0.3327366

队列中一个顾客等待的平均时间（W_q）= 0.2494033

所有服务台是闲置的概率（P_0）= 0.1666667

顾客来到时需要等待的概率（P_W）= 0.8333333

$P(1) = 0.13889$ $P(2) = 0.11574$ $P(3) = 0.09645$ $P(4) = 0.08038$

$P(5) = 0.06698$ $P(6) = 0.05582$ $P(7) = 0.04651$ $P(8) = 0.03876$

$P(9) = 0.03230$ $P(10) = 0.02692$

$$\sum_{i=1}^{10} P(i) = 0.698745$$

5. **以例 6 为例。**

输入数据：

$M/M/1$

顾客到达率（λ）$= 504.000$

分布：$Poisson$

服务台个数 $= 3$

服务率，每个服务台 $= 240.000$

分布：$Poisson$

平均服务时间 $= 0.004$（**小时**）

标准差 $= 0.004$（**小时**）

队列极限 $=$ **无限**

顾客人数 $=$ **无限**

输出数据：

$M/M/1$

每小时，具有（λ）$= 504$ **顾客，和服务率** $= 240$ **顾客**

利用率因子（ρ）$= 0.7$

系统中顾客的平均人数（L）$= 3.248804$

队列中顾客的平均人数（L_q）$= 1.148804$

系统中一个顾客逗留的平均时间（W）$= 6.446039E-03$

队列中一个顾客等待的平均时间（W_q）$= 2.279372E-03$

所有服务台是闲置的概率（P_0）$= 9.569379E-02$

顾客来到时需要等待的概率（P_W）$= 0.4923445$

$P(1) = 0.20096$　$P(2) = 0.21100$　$P(3) = 0.14770$　$P(4) = 0.10339$

$P(5) = 0.07237$　$P(6) = 0.05066$　$P(7) = 0.03546$　$P(8) = 0.02482$

$P(9) = 0.01738$　$P(10) = 0.01216$

$$\sum_{i=1}^{10} P(i) = 0.875924$$

这些结果与前面解答均一致。

第十一章　排队系统模拟——系统仿真

11.1　模拟概述

11.1.1　什么是模拟

模拟，又称为仿真，是一种在计算机上进行实验的数学方法，它是数学分析方法和数学模拟的一种延伸。

本书以前几章所述的运筹学的各分支，都是建立在数学模型的基础上，运用数学解析方法求解各有关变量，从而使实际问题得到优化和改进。但是，在很多情况下，往往由于问题本身的随机性质，或数学模型过于复杂，或要求在求解过程中进行人机对话，这时，单纯采用数学分析方法不容易或根本无法求出解析解。在这种情况下，构造面向问题的仿真模型，通过计算机进行模拟仿真，将可以得到实际问题的静态或动态特性，以及系统的主要参数或某些性能指标，从而对所研究的实际问题有一个概括的了解。

由于对实际过程进行仿真实验，并不需要实际过程介入，因此，它特别适用于费用十分昂贵的实验，如两军对垒的军事和武器系统的演习，还适用于短期内无法完成的实验，如未来几十年内的经济和社会发展的预测。

近年来，由于系统科学和计算机科学的发展，把仿真方法推向一个崭新阶段，即系统仿真阶段。系统仿真方法已在各个方面，包括管理方面，得到了广泛的应用，成为系统工程的有力工具。

因为仿真结果在很大程度上取决于构模人员的技术熟练程度，因此，它常被称为是一种“艺术”。这就要求在构造仿真模型时必须对客观的实际系统有深入的了解，并对仿真程序的编制有熟练的技巧。

11.1.2　蒙特卡洛（Monte Carlo）法

先介绍用随机方法解决肯定型问题或随机型问题的一种方法——蒙特卡洛法，这是一种特殊的数值计算方法。

客观实际系统有多种属性，有的是肯定型，有的是随机型，而有的系统在某些方面是肯定型的，在另一方面又是随机型的。对这些实际系统构造仿真模型，可以构造成肯定型的，也可以构造成随机型的。也就是说，可以用随机的方法来解决肯定型的问题。例如，现有一个单变量函数 $f(x)$，欲求该函数在 $[a,b]$ 区间的定积分，有三种方法供选择：① 如果 $f(x)$ 比较简单，则可用积分公式求解；② 如果 $f(x)$ 比较复杂，无法用积分公式求解，可以用一些数值方法，如辛卜生公式或龙贝方法求解；③ 用随机数采样的方法，即蒙特卡洛法。

见图 11.1，假定在 $[a,b]$ 区间 $f(x)\geqslant 0$，且 $f(x)$ 的最大值为 c。那么，所求定积分的值就等于其下方图形的面积，这个图形被包含在由边 $b-a$ 及 c 所构成的矩形之内。

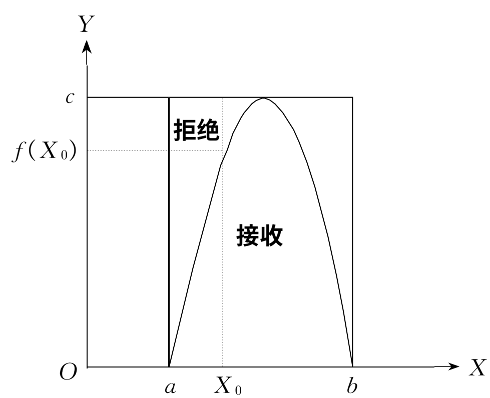


图 11.1

如果我们在矩形里边选择随机点，而且确定随机点是落在曲线之下还是落在曲线之上，那么，显而易见，当所提供的随机点的分布是均匀地出现在矩形之内时，则降落在曲线下方的随机点数占全部随机点数的比例，应该近似地等于曲线下方的面积与矩形面积的比值。设随机点总数为 N，而降落在曲线之下的随机点数为 n，则近似地有：

$$\frac{\int_a^b f(x)\,\mathrm{d}x}{c(b-a)}=\frac{n}{N} \tag{11.1}$$

或

$$\int_a^b f(x) = [c(b-a)] \cdot \frac{n}{N} \tag{11.2}$$

而且，当 N 的数值增加时，计算的精确度将提高。当点数足够多时，积分值$\int_a^b f(x)\,\mathrm{d}x$ 就可以用矩形面积 $c(b-a)$ 和比值$\frac{n}{N}$ 的乘积来估计。

随机点的选择方法可以分两步 第一步，在 a 和b 之间随机选择一个x，记为 X_0；第二步，在 0 和 c 之间随机选择一个y 值，设为 Y_0。如果 $Y_0 \leqslant f(X_0)$，这点就意味着被接收，并计入 n 内，否则它就被拒绝，然后再挑选下一个随机数。

当然，在实际上并没有广泛地使用蒙特卡洛法去估计单变量的积分值，因为有一些数值计算方法更为有效。但是，蒙特卡洛法却常常用于多变量的积分。尤其重要的是，蒙特卡洛法给了我们一个有益的启示，即一个肯定型的问题可以采用随机数的方法求解。自然，采用随机数的方法也可以解决随机型的问题。采用随机数的方法求解问题，是蒙特卡洛法的基本点，也是仿真技术的基本点。一般说来，蒙特卡洛法是用于静态模型的一种计算技术，仿真技术则用于动态模型的计算。也有人认为仿真是蒙特卡洛法的一种应用。

11.1.3 仿真的一般步骤

我们先用人工方法来仿真一个商店服务系统，然后再进一步研究用计算机仿真的系统。

假设顾客到达的时间间隔均匀分布在 1 到 10 分钟之间，而每一顾客所需要的服务时间均匀分布在 1 到 6 分钟之间。仿真的任务是要研究顾客在服务系统中所花费的平均时间和售货员的空闲时间占全部工作班时间的百分比。

为了描述商店服务系统活动，需要设计一种方法来产生反映顾客流情况的数据。最简单的方法是使用有从 1 到 10 的号码的十张扑克牌和一个从 1 到 6 点数的六面体骰子。任意抽一张扑克牌，读出上面的号码，表示当前顾客到达和前一个顾客之间的时间间隔，然后把扑克牌放回；再掷一下骰子，读出上面的点数，表示当前顾客所需的服务时间。重复这个过程，就产生出一系列表示顾客到达的时间间隔和相应顾客需要的服务时间数据。表 11.1 列出了 20 个顾客的样本。

表 11.1

顾客	顾客到达间的时间间隔(分)	服务时间(分)	到达的时间	服务时间		顾客在系统中全部时间	售货员空闲时间(分)
				开始	结束		
1	–	1	8:00	8:00	8:01	1	0
2	3	4	8:03	8:03	8:07	4	2
3	7	4	8:10	8:10	8:14	4	3
4	3	2	8:13	8:14	8:16	3	0
5	9	1	8:22	8:22	8:23	1	6
6	10	5	8:32	8:32	8:37	5	9
7	6	4	8:38	8:38	8:42	4	1
8	8	6	8:46	8:46	8:52	6	4
9	8	1	8:54	8:54	8:55	1	2
10	8	3	9:02	9:02	9:05	3	7
11	7	5	9:09	9:09	9:14	5	4
12	3	5	9:12	9:14	9:19	7	0
13	8	3	9:20	9:20	9:23	3	1
14	4	6	9:24	9:24	9:30	6	1
15	4	1	9:28	9:30	9:31	3	0
16	7	1	9:35	9:35	9:36	1	4
17	1	6	9:36	9:36	9:42	6	0
18	6	1	9:42	9:42	9:43	1	0
19	7	2	9:49	9:49	9:51	2	6
20	6	2	9:55	9:55	9:57	2	5
总计						68	55

由表数据可知,20 名顾客在服务系统中全部时间是 68 分钟,售货员空闲时间是 55 分钟,售货员从 8 点开始到 9 点 57 分,在班上时间共 117 分钟。于是可以算出顾客在服务系统中平均时间

$$\omega = \frac{68}{20} = 3.4 \text{(分)}$$

售货员空闲时间的百分比

$$\eta = \frac{55}{117} \times 100\% = 0.47 \times 100\% = 47\%$$

根据排队论的知识,本例的服务员空闲所占比例过大,应加以调整。

如果不仅要考虑系统在一段时间内的稳态性能,如前面所计算的;而且要考虑系统的动态特性,即顾客的到达是随时间而变化的实际背景,如高峰时顾客到的多,低峰时顾客到的少。那么可以在不同的时间段内设计不同的随机数的均值,就可以

得到在不同的时间内商店服务系统的性能。可见，仿真技术不仅能评价复杂的静态模型，而且能评价复杂的动态模型。

当然，用人工方法仿真系统既麻烦又不可靠。高速电子计算机的出现，才为复杂系统的仿真开辟了道路。计算机仿真主要特点是在极短的时间内产生随机性很好的随机数列，用程序模拟系统中实体的运动情况，最后输出所希望的数据。

计算机仿真的一般步骤可用图 11.2 表示。

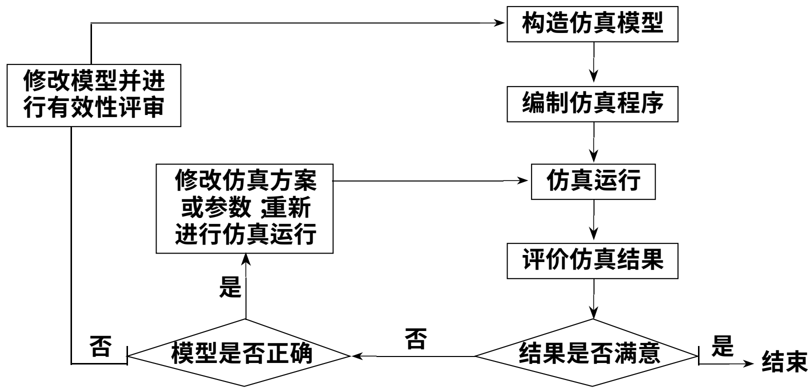


图 11.2

1. 构造仿真模型

这里的构造模型是指：

⑴ 对所研究的系统进行详细地了解，包括该系统的范围、系统的外部环境和系统的目标。系统的范围要确定哪些部分必须包括在内，哪些部分可以忽略不计；系统的外部环境，着重分析那些对系统影响较大的因素；仿真的目标将决定仿真结果的具体要求，即仿真结果将输出哪些参数，如对一般排队问题，应输出平均等待时间、平均队列长度、服务设施的利用率以及这些参数的概率分布等等。

⑵ 确定模型中所包含的变量。系统的变量可以分为外生变量和内生变量。所谓外生变量，就是外部环境影响系统的因素，构模时表现为输入模型的变量，如排队模型中顾客到达的间隔时间和服务时间属于这一类变量。内生变量是仿真模型内部生成的变量，如顾客的等待时间，服务员空闲时间等。

⑶ 确定模型的结构，即模型由哪几部分组成，它们之间的逻辑关系和数学关系是什么。如在排队问题中，服务设施（理发员、生产机床等），一般称为模型中的固定实体；顾客（或被加工零件等），称为模型中的流动实体；还有随着流动实体而变动的各种参数。这些实体之间由一定的逻辑关系和数学关系联结起来，就构成了一个完整的仿真模型。

⑷ 收集和分析必要的数据。比如，各个变量相对应的概率密度函数或累积分布函数，各种定额或标准数据等。在收集数据的同时必须进行必要的数据处理，这时将用到一些统计方法，如回归分析、方差分析、置信度测定。难以想像，输入不准确的数据，经过仿真运行，会得到准确的数据输出。

2. 根据模型编程序

一旦构模完成之后，主要的工作就是编程序。一般应先绘制出相应的框图或模块图，然后再写源程序。任何通用语言都可用于仿真模型编程。用以下方法可鉴定所编程序的有效性：

⑴ 先将模型简化，如将随机变量改为非随机变量，对模型先行定性运行，证明模型基本正确以后，再加入随机因素继续进行仿真运行。

⑵ 验证已知情况。如将系统过去运行数据或参数输入仿真模型，如果仿真结果与已知实际结果基本一致，则说明构模是成功的。

⑶ 对输出变量进行统计分析。用同样的数据对仿真模型进行多次运行，每次运行只改变其随机数的顺序，那么，其输出结果的变动不应过大，即输出结果的方差不应过大。如果是这样，则模型是可取的。否则，模型中可能存在较大的问题。

⑷ 对模型进行分部测试。如果各部分子模型都与相应的子系统一致，则模型总体很可能是可信的。

3. 仿真运行

当我们确信所构造的模型是正确的，并且所编的程序已能成功地进行工作，这时仿真运行工作才能真正开始。任何仿真运行不能期望一次得到全部结果，因为每次运行只相当于一次统计试验。因此，常常要对输入参数，甚至模型作必要的变动，才能看出实际系统在多方面的特性。

另一个必须注意的问题是每次仿真运行时间必须足以使系统进入稳态。仿真运行一般由初始条件开始，经过一定时间的瞬态过程以后，才能进入稳态，如图 11.3。

进入稳态以后，还须有足够的运行时间，才能取得稳态参数。只是系统由瞬态转入稳态的时间的长短取决于系统本身的特性。因此，对于一个具体的系统，究竟何时进入稳态，要对输出参数进行分析观察以后，才能大致确定。

4. 分析与评价仿真结果

仿真过程中，对实际系统进行了抽象化的构思成为模型，输入的数据有些是推测出来的，这些都包含了主观因素。因此对仿真结果必须进行全面的分析，确认有价值，才将结果送交有关部门。对仿真结果的分析主要达到两个目的。其一是所得结果对于分析解决实际问题是否满足了，如果还不满足就要找出新的方案，或改变模型的结构增加新的子模型，或改变原有的参数以取得更多的数据；其二是将仿真结果

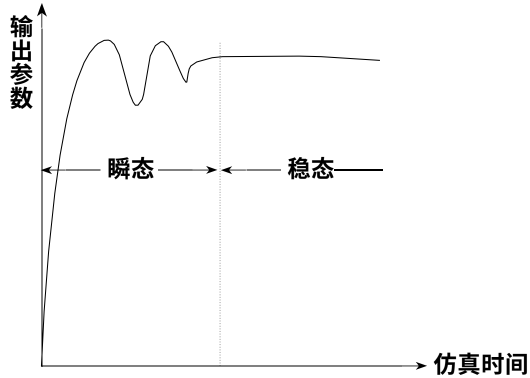


图 11.3

进行归总、整理，以提供决策的有效数据。

◆ 11.2 伪随机数发生器

在前一节，我们曾用人工方法仿真一个商店服务系统，所谓人工方法，是指商店顾客到达和服务时间的随机变量，是用人工抽扑克牌和投骰子的方法来产生的，我们并且说明了用人工方法产生的随机数其随机性是不可靠的。用计算机产生随机数，则要优越得多，速度快，效率高，而且随机性能好。那么，计算机是怎样产生随机数的呢？

11.2.1 均匀分布的随机数产生办法

最简单的随机数是均匀分布的随机数。设在 $[a, b]$ 区间内均匀分布的随机变量是由下式定义：

$$f(x) = \begin{cases} \dfrac{1}{b-a} & a \leqslant x \leqslant b \\ 0 & \text{其他} \end{cases} \tag{11.3}$$

式中，$f(x)$ 为概率密度函数。相应的，其累积分布函数 $F(x)$ 可表示为：

$$F(x) = \begin{cases} 0 & x < a \\ \dfrac{x-a}{b-a} & a \leqslant x \leqslant b \\ 1 & x > b \end{cases} \tag{11.4}$$

特别，当 $a = 0, b = 1$ 时，有：

$$f(x) = \begin{cases} 1 & 0 \leqslant x \leqslant 1 \\ 0 & \text{其他} \end{cases} \tag{11.5}$$

$$F(x) = \begin{cases} 1 & 0 \leqslant x \leqslant 1 \\ 0 & \text{其他} \end{cases} \tag{11.6}$$

所谓均匀分布的随机数，就是在 $[0,1]$ 区间内以同样的概率所产生出来的一系列随机数。这种概率均等的随机数，可以考虑以下几种途径产生。

第一种方法，将事先准备好的随机数存入计算机的磁带或磁盘，在使用时直接调用。

第二种方法，在计算机外部建立专门的物理装置，如热噪声源，按其噪声电压的大小表示不同的随机数。

第三种方法，用一定的算法，在计算机内产生所需要的随机数，这是目前使用最广泛的方法。通常，第 $i+1$ 个随机数是用第 i 个随机数按一定公式计算出来的。这样，使得产生出来的随机数并不是真正的随机数，因此称为伪随机数。实践证明，这种伪随机数发生器能够充分满足仿真及其他随机过程的计算的需要，并且，其产生过程也是方便的，可行的。

既然伪随机数是通过一定的算法过程实现的，那么对于这些算法，也称为伪随机数发生器，就应具有一定的要求，主要有：

1. 所产生的数字必须尽可能接近均匀分布；
2. 此发生器必须具有较长的循环周期，即应产生大量的随机数之后才出现重复现象；
3. 算法过程不能发生退化现象，即不能出现反复产生同一常数的情况；
4. 算法过程应有再现性，即在同样条件下，可复现相同的随机数序列。

下面介绍几种比较常见的伪随机数发生器。有些是历史性的过渡方法，供参考；有些是目前常使用的方法。

(1) 平方取中法

首先任选一个大于 32 位的 $2n$ 位，即偶数位的整数 X_0 作为初始值，然后把它平方，得到一个 $4n$ 位的整数，取出中间的 $2n$ 位数作为第一个随机数 X_1；再把 X_1 平方，取出中间的 $2n$ 位数作为第二个随机数 X_2 如此不断重复下去，就可得到 $X_0, X_1, \cdots$ 的均匀分布的伪随机数列。如：

$$X_0 = 2152; X_0^2 = 04631104;$$

$$X_1 = 6311; X_1^2 = 39828721;$$

$$X_2 = 8287; X_2^2 = 68674369;$$

$X_3 = 6743$； …

这样一直下去，即产生一个四位数的伪随机数列，这种伪随机数列只要通过一个简单的程序在计算机上就能实现。

⑵ 乘积取中法

首先选 $2n$ 位的整数 X_0 和 X_1 两个数作为初始值，取它们乘积中间的 $2n$ 位作为 X_2 然后从 X_1 和 X_2 的乘积中以同样的方法得到 X_3 如此不断地重复下去，就可以得到一个伪随机数序列。如：

$X_0 = 1234$；　　$X_1 = 5768$；

$X_0X_1 = 07006652$；　　$X_2 = 0066$；

$X_1X_2 = 00374748$；　　$X_3 = 3747$；

$X_2X_3 = \cdots$

⑶ 乘同余数法

这是应用最广泛的一种算法，其数学公式为：

$$X_{n+1} = AX_n \pmod{m} \tag{11.7}$$

式中，X_{n+1}—— 第 $n+1$ 个随机数；X_n—— 第 n 个随机数；A 和 m 是已知的正整数，$A < m$；$AX_n \pmod{m}$ 是表示 AX_n 除以 m 所得的余数。这种方法有时又叫余数法。

假设 $A = 11$，$X_0 = 7$，$m = 2^4 = 16$，依公式 (11.7) 可得伪随机数列：

$X_1 = 13$，(因 $A\dfrac{X_0}{m} = 11 \times \dfrac{7}{16} = 4\cdots13$，余数是 13)

$X2 = 15$，(因 $A\dfrac{X_1}{m} = 11 \times \dfrac{13}{16} = 8\cdots15$，余数是 15)

$X3 = 5$，(因 $A\dfrac{X_2}{m} = 11 \times \dfrac{15}{16} = 10\cdots5$，余数是 5)

$X4 = 7$，(因 $A\dfrac{X_3}{m} = 11 \times \dfrac{5}{16} = 3\cdots7$，余数是 7)

…

由上面推算可知，这个随机数列取 4 个数后就开始循环重复，其循环长度为 2^{4-2}。很明显，要用这种方法来产生一个永不重复的数列是不可能的。然而，如果能细心地选择有关常数，就可以在数列重复之前得到一个很长的数列，可以将此视为随机数。

一个理想的伪随机数发生器，应能产生均匀分布的随机数列，且各数字之间的相关性差，数列的循环周期长。此外，在实际中还要考虑这些随机数在数字计算机上能有效地实现。

11.2.2 产生给定分布的随机数的方法

产生了均匀分布的随机数之后,便为进一步产生各种所需要的分布的随机数创造了条件。将均匀分布的随机数列,转化为其他分布的随机数列,应用最广泛的一种方法,是概率论中的概率积分变换运算法,通常称为反变换法,它同求一个函数的反函数的方法类似。

反变换定理可以叙述为:如果 U_i ($i=1,2,\cdots$) 是在从0到1区间内均匀分布的独立随机变量,而且 $F^{-1}(x)$ 是随机变量 X 的反累积分布函数,则由 $X_i=F^{-1}(U_i)$ 定义的随机变量就是变量 X 的随机抽样。也就是说,为了产生具有给定分布的随机数 x,必须先求出它的累积分布函数 $F(x)$,并且使在 $[0,1]$ 区间内均匀分布的随机数 U 等于 $F(x)$,然后求出反函数 $F^{-1}(x)$,就是给定分布的随机数。按取反变换的方法,又分为解析法和数值法。

1. 解析法

对于一些比较简单的函数表达式,可用解析法直接求反函数,由此得到给定分布的随机数。

例1 试由 $[0,1]$ 区间均匀分布的随机数 U,产生一组符合以下概率密度的随机数:

$$f(x)=\begin{cases}3x^2 & 0\leqslant x\leqslant 1\\ 0 & \text{其他}\end{cases}$$

先算出该变量的累积分布函数 $F(x)$,

$$F(x)=\int_0^x 3t^2\mathrm{d}t=x^3$$

令均匀分布的随机数 U 与 $F(x)$ 相等,得:

$$U=F(x)=x^3$$

解出由 U 为自变量的 x 的表达式:

$$x=F^{-1}(U)=U^{\frac{1}{3}}$$

例2 试由 $[0,1]$ 区间均匀分布的随机数 U,产生一组随机数,使这组随机数符合以下概率密度函数:

$$f(x)=\begin{cases}2(1-x) & 0\leqslant x\leqslant 1\\ 0 & \text{其他}\end{cases}$$

第一步,算出该随机变量的累积分布函数,

$$F(x)=\int_0^x 2(1-t)\mathrm{d}t=2x-x^2$$

第二步,令

$$U = F(x) = 2x - x^2$$

第三步,解出 x,得:

$$x = 1 \pm (1 - U)^{\frac{1}{2}}$$

又因为 $0 \leqslant x \leqslant 1$,故舍去一个根,得:

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

即为所求。

负指数分布的概率分布函数在仿真中是经常用到的,怎样通过 $[0,1]$ 区间均匀分布的随机数产生负指数分布的随机数呢?根据反变换定理,先列出负指数分布的概率密度函数:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geqslant 0 \\ 0 & \text{其他} \end{cases}$$

其累积分布函数

$$F(x) = 1 - e^{-\lambda x}$$

令

$$U = F(x) = 1 - e^{-\lambda x}$$

故

$$1 - U = e^{-\lambda x}$$

由于 U 是在 $[0,1]$ 区间均匀分布的随机数,故 $1 - U$ 也是在 $[0,1]$ 区间均匀分布的随机数,因此可以用 U 代替 $1 - U$,得:

$$U = e^{-\lambda x}$$

取反函数,得:

$$x = -\ln\frac{U}{\lambda} \tag{11.8}$$

这就是负指数分布的随机数产生的方法。

2. 数值法

对于表达形式比较复杂的概率累积分布函数,用解析法求反函数常常是十分困难的,甚至是完全不可能的,在这种情况下必须采用数值计算方法。

例 3　有一电话系统,按电话呼叫时间长短统计各自的概率,得累积分布函数如图 11.4 所示。试用数值方法计算其输出的几个随机数。

我们不妨将从 0 到 1 之间均匀刻度的纵坐标看作是均匀分布的随机数,是输入的随机数 U,将横坐标看作是输出的随机数 X。那么,对任一给定随机数 U_j,可在曲线上找出对应于 U_j 值的横坐标 x 值,即是输出的随机数 x_j。需要注意的是,如果给

定的 Uj 值不是恰恰在有精确值的曲线的数值点上，那么可以近似地认为两个相邻的数值点之间是直线，这样就可以用线性插值的方法估计输出的随机数值。

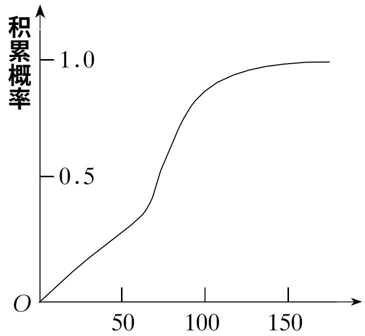


图 11.4　呼叫时间长度

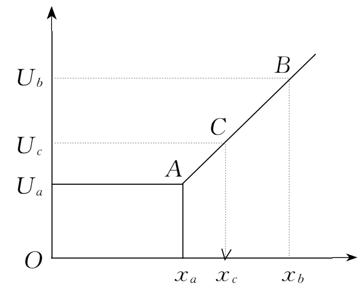


图 11.5

线性插值的原理见图 11.5，近似认为曲线上的点 A 和 B 之间是一线段，如果输入值 Uc 对应曲线上的点 c，并且 Ua，Ub，xa，xb 是已知的随机数，由线段比例性质，得：

$$\frac{U_c - U_a}{x_c - x_a} = \frac{U_b - U_a}{x_b - x_a}$$

由上式可解出输出的随机数

$$x_c = x_a + \frac{U_c - U_a}{(U_b - U_a) \cdot (x_b - x_a)} \tag{11.9}$$

设本例输入随机数 U_j 为 0.1009，0.3754，0.0842，0.9901，0.1280，则可用数值方法（需要插值），得到输出的随机数 x_i 为 69.23，88.48，66.65，142.33，71.82。

3. 舍弃法

它同蒙特卡洛法类似。

11.3　模拟实例

我们研究一个多服务员的排队系统。

例 4　某工厂工具间有两名管理员，负责向生产工人出借工具。根据实地观测工人来到工具间的平均间隔为 5 分钟，服务时间的频率分布如表 11.2：

表 11.2

服务时间	6	7	8	9	10
频　　率	0.1	0.2	0.3	0.3	0.1

每个管理员工作一小时的工资为 0.5 元，每个生产工人平均每小时可为工厂创造利润 3 元，试用模拟方法研究这一系统的效益。

假设进行模拟的总时间确定为 3 小时，工人到达的间隔服从均匀分布，也即有：

表 11.3

间 隔 (分)	1	2	3	4	5	6	7	8	9	10
概　　率	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1

为了生成抽样数据，可先分别求出到达间隔的分布函数 $F(x_i) = P(X \leqslant x_i)$ 和服务时间的积累分布 (表 11.4 和表 11.5)。

表 11.4

间隔 x_i	1	2	3	4	5	6	7	8	9	10
$F(x_i)$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0

表 11.5

服务时间	6	7	8	9	10
积累频率	0.1	0.3	0.6	0.9	1.0

这是一个具有 2 个服务员、排队规则为等待制的先到先服务的排队系统。

我们可以这样来模拟这一随机服务系统：

1. 产生一个随机数，利用这一随机数由分布 F 得出到达间隔 T_1，于是 1♯ 顾客到达系统的时刻 $t = T_1$。此时，服务员有空，马上对 1♯ 顾客服务。接着再产生一个随机数，根据分布 G 得出 1♯ 顾客的服务时间 τ_1。于是，1♯ 顾客在时刻 $t = T_1 + \tau_1$ 接受服务后离去。

2. 产生一个随机数，根据分布 F 得到到达间隔 T_2 ;再产生一个随机数，根据分布 G 得到 2♯ 顾客的服务时间 τ_2。因此，2♯ 顾客在时刻 $t = T_1 + T_2$ 到达系统。若他到达时，1♯ 顾客已离去，即若 $T_1 + T_2 \geqslant T_1 + \tau_1$，则他马上可得到服务，然后在时刻 $t = T_1 + T_2 + \tau_2$ 离开系统。若他到达时，1♯ 顾客尚未离去，即若 $T_1 + T_2 < T_1 + \tau_1$，那么他需等到时刻 $t = T_1 + \tau_1$ 才能受到服务，然后于时刻 $t = T_1 + \tau_1 +$

τ_2 离开系统。

3. 类似地可以模拟3＃,4＃,…,n＃顾客的到达、服务和离去。直到预定的模拟时间 T 达到后再结束。

4. 在若干预定的时刻算出这一排队系统的数量指标,如平均队长(L),平均等待队长(L_q),平均逗留时间(W),平均等待时间(W_q),每个服务台的利用率($Util.$),等等。

由于在我们给出的软件 *YAJ* 中,*QSIM* 程序已自动地给出了随机数,只要给出相应的数据,上述数量指标均可从微机输出结果中读到。

11.4 微机操作 —— 调用 *YAJ* 中,*QSIM* 决策支持系统

11.4.1 观察 *QSIM* 决策支持系统

这个程序用蒙特卡洛模拟以分析具有直到20个服务台和20个队列的排队系统。在一行中允许顾客数达到100人。服务技巧取决于服务时间和分布形式。服务机构由排队容量和优先类型所规定。当你在定义服务和到达模型时,有5种分布形式允许使用:指数的,爱尔兰的,一致的,正态的,常数的。指定的优先规则可以是 *FIFO*(先到先服务),*LIFO*(后到先服务),或随机的。交替到达时间分布也由5种分布形式中之一种规定。

11.4.2 问题输入

当输入一个问题时,请熟悉下列约定:

(1) *ESC* 键,允许你从开始重新输入数据。

(2)/键,允许你对前一个问题重新输入数据。

(3) 在输完数据以后,函数菜单允许你修改数据。

然后回答下列问题:

有多少个服务台(不超过20)?

所有服务台都是相同的吗(Y/N)?

平均服务时间(分钟)是多少?

可用的'服务/到达'时间分布是:

① 指数的,② 爱尔兰的,③ 一致的,④ 正态的,⑤ 常数的。

服务时间分布是什么(1-5)?

平均服务时间是多少？

有多少队列（不超过 20，缺损值 = 1）？

队长限制是多少（≤100，缺损值 = 100）？

排队优先规则是：

①*FIFO*（先到先服务），②*LIFO*（后到先服务），③ 随机的。

优先规则是什么（1 – 3）？

平均交替到达时间是多少？

什么是交替到达时间分布（1 – 5）？

　　最少时间？

　　最多时间？

你要做多长的模拟（分钟）？

你要从什么时间收集数据？

输入每个队列中顾客的最初人数：

在队列中有多少顾客？

我们就前面 11.3 中的模拟实例进行输入：

服务台个数：2 个　两个服务台的平均服务时间：8.1（分）　分布：指数

类型：*FIFO*（先到先服务）　顾客交替到达时间：5（分）

11.4.3　模拟显示

1）时间：0.00，当前事件：初始状态，时间极限：100.00，从：0.0

服务台状况	队列	Qmax	Lq	L	Wq	W	$Dtil$
1，闲	0	0	0.00	0.00	0.00	0.00	0.00000
2，闲				0.00	0.00	0.00	0.00000

综合：Lq = 0.00　L = 0.00　Wq = 0.00　W = 0.00　$Util.$ = 0.0000　$Balking$ = 0

2）时间：6.04，当前事件：新的到达，时间极限：100.00，从：0.00

服务台状况	队列	Qmax	Lq	L	Wq	W	$Dtil$
1，闲	0	0	0.00	0.00	0.00	0.00	0.00000
2，闲				0.00	0.00	0.00	0.00000

综合：Lq = 0.00　L = 0.00　Wq = 0.00　W = 0.00　$Util.$ = 0.0000　$Balking$ = 0

3）时间：11.12，当前事件：服务结束，时间极限：100.00，从：0.00

服务台状况	队列	Qmax	Lq	L	Wq	W	$Dtil$
1，闲	0	0	0.00	0.00	0.00	0.00	0.00000
2，忙				0.46	0.00	5.07	0.45631

综合：Lq = 0.00　L = 0.46　Wq = 0.00　W = 5.07　$Util.$ = 0.4563　$Balking$ = 0

4) 时间:11.19,当前事件:新的到达,时间极限:100.00,从:0.00

服务台状况	队列	Qmax	Lq	L	Wq	W	Dtil
1,闲	0	0	0.00	0.00	0.00	0.00	0.00000
2,闲				0.45	0.00	5.07	0.45343

综合:$Lq = 0.00$ $L = 0.45$ $Wq = 0.00$ $W = 5.07$ $Util. = 0.4534$ $Balking = 0$

5) 时间:11.60,当前事件:新的到达,时间极限:100.00,从:0.00

服务台状况	队列	Qmax	Lq	L	Wq	W	Dtil
1,闲	0	0	0.00	0.00	0.00	0.00	0.00000
2,忙				0.47	0.00	5.83	0.47273

综合:$Lq = 0.00$ $L = 0.47$ $Wq = 0.00$ $W = 5.83$ $Util. = 0.4727$ $Balking = 0$

6) 时间:17.77,当前事件:服务结束,时间极限:100.00,从:0.00

服务台状况	队列	Qmax	Lq	L	Wq	W	Dtil
1,忙	0	0	0.00	0.35	0.00	9.00	0.34722
2,忙				0.66	0.00	5.83	0.65581

综合:$Lq = 0.00$ $L = 1.00$ $Wq = 0.00$ $W = 6.88$ $Util. = 1.0030$ $Balking = 0$

7) 时间:20.60,当前事件:服务结束,时间极限:100.00,从:0.00

服务台状况	队列	Qmax	Lq	L	Wq	W	Dtil
1,忙	0	0	0.00	0.44	0.00	9.00	0.43692
2,闲				0.57	0.00	5.83	0.56568

综合:$Lq = 0.00$ $L = 1.00$ $Wq = 0.00$ $W = 6.88$ $Util. = 1.0026$ $Balking = 0$

8) 时间:28.23,当前事件:新的到达,时间极限:100.00,从:0.00

服务台状况	队列	Qmax	Lq	L	Wq	W	Dtil
1,闲	0	0	0.00	0.32	0.00	9.00	0.31881
2,闲				0.41	0.00	5.83	0.41276

综合:$Lq = 0.00$ $L = 0.73$ $Wq = 0.00$ $W = 6.88$ $Util. = 0.7316$ $Balking = 0$

9) 时间:28.84,当前事件:服务结束,时间极限:100.00,从:0.00

服务台状况	队列	Qmax	Lq	L	Wq	W	Dtil
1,闲	0	0	0.00	0.32	0.00	9.00	0.31881
2,忙				0.41	0.00	5.83	0.41276

综合:$Lq = 0.00$ $L = 0.73$ $Wq = 0.00$ $W = 6.88$ $Util. = 0.7316$ $Balking = 0$

10) 时间:31.06,当前事件:新的到达,时间极限:100.00,从:0.00

服务台状况	队列	Qmax	Lq	L	Wq	W	Dtil
1,闲	0	0	0.00	0.29	0.00	9.00	0.28974
2,闲				0.39	0.00	4.09	0.39482

综合:$Lq = 0.00$ $L = 0.68$ $Wq = 0.00$ $W = 5.32$ $Util. = 0.6846$ $Balking = 0$

11) 时间:32.89,当前事件:服务结束,时间极限:100.00,从:0.00

服务台状况	队列	Qmax	Lq	L	Wq	W	Dtil
1,闲	0	0	0.00	0.27	0.00	9.00	0.27359
2,忙				0.43	0.00	3.52	0.42854

综合：$Lq = 0.00$　$L = 0.70$　$Wq = 0.00$　$W = 4.62$　$Util. = 0.7021$　$Balking = 0$

12）时间：35.68,当前事件：新的到达,时间极限：100.00,从：0.00

服务台状况	队列	Qmax	Lq	L	Wq	W	Dtil
1,闲	0	0	0.00	0.25	0.00	9.00	0.25220
2,闲				0.40	0.00	3.52	0.39503

综合：$Lq = 0.00$　$L = 0.65$　$Wq = 0.00$　$W = 4.62$　$Util. = 0.6472$　$Balking = 0$

13）时间：39.46,当前事件：服务结束,时间极限：100.00,从：0.00

服务台状况	队列	Qmax	Lq	L	Wq	W	Dtil
1,忙	0	0	0.00	0.32	0.00	6.39	0.32372
2,闲				0.36	0.00	3.52	0.35372

综合：$Lq = 0.00$　$L = 0.68$　$Wq = 0.00$　$W = 4.48$　$Util. = 0.6810$　$Balking = 0$

14）时间：40.35,当前事件：新的到达,时间极限：100.00,从：0.00

服务台状况	队列	Qmax	Lq	L	Wq	W	Dtil
1,闲	0	0	0.00	0.32	0.00	6.39	0.31653
2,闲				0.35	0.00	3.52	0.34931

综合：$Lq = 0.00$　$L = 0.67$　$Wq = 0.00$　$W = 4.48$　$Util. = 0.6658$　$Balking = 0$

15）时间：42.13,当前事件：新的到达,时间极限：100.00,从：0.00

服务台状况	队列	Qmax	Lq	L	Wq	W	Dtil
1,闲	0	0	0.00	0.30	0.00	6.39	0.30320
2,忙				0.38	0.00	4.06	0.37671

综合：$Lq = 0.00$　$L = 0.68$　$Wq = 0.00$　$W = 4.72$　$Util. = 0.6799$　$Balking = 0$

16）时间：45.03,当前事件：新的到达,时间极限：100.00,从：0.00

服务台状况	队列	Qmax	Lq	L	Wq	W	Dtil
1,忙	0	0	0.00	0.35	0.00	7.00	0.34810
2,忙				0.42	0.00	4.06	0.41687

综合：$Lq = 0.00$　$L = 0.76$　$Wq = 0.00$　$W = 5.16$　$Util. = 0.7650$　$Balking = 0$

17）时间：46.54,当前事件：服务结束,时间极限：100.00,从：0.00

服务台状况	队列	Qmax	Lq	L	Wq	W	Dtil
1,忙	1	1	0.00	0.37	0.00	7.00	0.36921
2,忙				0.44	0.00	4.06	0.43575

综合：$Lq = 0.00$　$L = 0.80$　$Wq = 0.00$　$W = 5.16$　$Util. = 0.8050$　$Balking = 0$

18）时间：48.93,当前事件：新的到达,时间极限：100.00,从：0.00

服务台状况	队列	Qmax	Lq	L	Wq	W	Dtil
1,忙	0	1	0.03	0.43	0.00	7.00	0.40005
2,忙				0.46	0.25	4.06	0.46334

综合:$Lq = 0.03$ $L = 0.89$ $Wq = 0.17$ $W = 5.38$ $Util. = 0.8634$ $Balking = 0$

19) 时间:50.35,当前事件:服务结束,时间极限:100.00,从:0.00

服务台状况	队列	Qmax	Lq	L	Wq	W	Dtil
1,闲	1	1	0.03	0.45	0.00	7.00	0.41700
2,忙				0.48	0.25	4.56	0.47851

综合:$Lq = 0.03$ $L = 0.93$ $Wq = 0.17$ $W = 5.38$ $Util. = 0.8955$ $Balking = 0$

20) 时间:50.73,当前事件:服务结束,时间极限:100.00,从:0.00

服务台状况	队列	Qmax	Lq	L	Wq	W	Dtil
1,闲	0	1	0.06	0.48	0.36	5.70	0.42131
2,忙				0.48	0.25	4.56	0.48236

综合:$Lq = 0.06$ $L = 0.96$ $Wq = 0.29$ $W = 5.02$ $Util. = 0.9037$ $Balking = 0$

……

11.4.4 综合结果

服务台	*Util.*	*Wq.*	*Var.(Wq)*	*W.*	*Var.(W)*	*Obsvtn.*
1	0.5996	1.5973	6.4485	6.7112	14.92	12
2	0.6209	1.2059	4.4995	7.5608	31.68	10

数据采集周期:0 到 102.3548(分)

队列	Qmax	Qmin	*Current Q*	*Lq.*	*Var.(Lq)*	*L.*
1	3	0	0	0.3051	0.4521	0.9046

数据采集周期:0 到 102.3548(分)

$Util. = 1.22042$ $Lq. = 0.3051$ $Var.(Lq) = 0.4521$ $L. = 1.5255$

$Wq. = 1.4194$ $Var.(Wq) = 5.6005$ $W. = 7.0974$ $Var.(W) = 22.72$

数据采集周期:0 到 102.3548(分)

第十二章　决策/概率论

所谓决策，就是为了达到某种预定的目的，决策者在若干可供选择的行动方案中，选择最佳方案的过程。决策是人们在工作和生活中普遍存在的一种活动。

在一定的条件下，从某些可能做出的决定中，选择最优策略的有关理论，就是决策分析所研究的内容。

本章将介绍决策/概率理论中的基本概念，着重讨论确定型决策问题、风险型决策问题和不确定型决策问题。

使用的方法有均值和方差分析、贝叶斯分析、支付表分析和决策树分析。

12.1　决策的概念

12.1.1　决策的基本概念

在工业生产、农业生产、交通运输、国防建设以及科学研究等活动中，在工程中的设计与施工，工业新产品的开发与生产批量的确定，企业的发展，商品的经营等，经常需要面对几种不同的自然状态（或称客观条件），对可能采取的几种不同方案进行选择，这就提出了决策问题。

例 1　某工厂考虑生产甲、乙两种产品。根据以往统计资料推测，如果市场上销售情况好，则生产甲种产品可获利 5 万元，生产乙种产品获利 9 万元；销售一般时，生产甲种产品可获利 3 万元，生产乙种产品可获利 1 万元；销路差时，生产甲种产品可获利 1 万元，生产乙种产品获利 -1 万元（即赔本 1 万元）。根据调查和推测，得知产品销路好、一般、差的概率分别是 0.3，0.5，0.2。试决定工厂生产哪一种产品可能获利最大？

解　这个决策问题是面对三种自然状态（销路好，一般，差），有两种可供选择

的行动方案的决策问题。

自然状态是决策者在决策时可能碰到的自然情况,或称客观条件。很显然,这些状态的出现是不以决策者的意志为转移的,是决策者不可控制的因素。我们用 S_j 表示第 j 个自然状态。

策略,就是可供选择的行动方案。策略的选取要由决策者决定。因此,策略是决策者可控制的因素。我们用 A_i 表示第 i 个策略。策略常称为决策变量。

收益值,是指对应每一个自然状态,采取一种策略,决策者就会得到一个确定的收益。我们用 C_{ij} 表示在状态 S_j 下决策 A_i 所得到的收益值。所有收益值可用收益矩阵 $D = (C_{ij})$ 表示。

12.1.2 决策的三要素

状态,策略,收益值称为决策的三要素,我们下面来做出进一步的分析。

1. 状态

因为状态是决策者无法控制的,所以在决策前必须要了解状态的信息,这才有可能估计在采取相应的策略后可能发生的效果。状态一般有确定型的,即必然发生的;随机型的,即有统计规律可循;不确定型的。根据状态的上述三种情况,而有确定型决策,风险型决策,不确定型决策。

2. 策略

一般在两个或两个以上的策略中,才可以进行评比择优,做出相应的决策。策略可以用离散变量表示,也可以用连续变量表示,可以是静态的,也可以是动态的。

3. 收益

通常是用货币或其他计量标准衡量。对于多目标决策,往往折算成同一种标准衡量。

12.1.3 决策过程

(1) 确定决策目标。当确定决策所要达到的目的时,要兼顾各方面的要求,否则,会造成决策失误。当有多个目标时,应主次恰当,统筹兼顾。

(2) 对状态进行科学分析。一般是根据过去和现在的信息对未来进行预测。预测能为决策提供可靠的科学远景,是决策的前提。要用各种预测技术,为决策提供未来的信息,以便使决策更加准确和正确。

(3) 提出拟采用的策略。拟定策略时,应以科学技术手段作为基础,所选择的策略应是切实可行的策略。策略的可行性是决策的先决条件。

(4) 评价策略效果。尽可能对各种策略的效果进行科学计算和分析。要注意定量

与定性分析相结合,局部效果要服从整体效果,当前效果和长远效果相结合。

(5) 做出决策,选择最优策略或满意策略。

12.1.4 决策的分类

(1) 按自然状态分类:有确定型决策,风险型决策或叫随机型决策,不确定型决策。

(2) 按决策目标分类:有单目标决策问题,多目标决策问题。

(3) 按决策的内容分类:有战略决策,管理决策,日常业务决策。

(4) 按决策的方法分类:有程序化决策,非程序化决策。

12.2 确定型决策问题

确定型决策问题的主要特征是:

(1) 只有一个确定的自然状态;

(2) 存在着决策希望达到的一个明确目标;

(3) 存在着可供决策人选择的两个或两个以上的行动方案;

(4) 不同的行动方案在确定情况下的收益值或损失值可以计算出来。

这是一种逻辑比较简单的决策,只需要在多个备选方案中,选择一个最有利的方案就行了。

例 2 某企业生产两种产品:A 和 B,有关资料如下表。试求企业盈利最大时 A 和 B 的产量。

表 12.1

	生产 1 公斤 A 需要	生产 1 公斤 B 需要	资源限制
原料 (公斤)	8	5	400
使用设备 (小时)	2	5	200
劳动力 (人)	9	10	500
盈利 (元)	20	35	

解 这是为了获得企业最大盈利,如何组织两种产品的产量的决策问题。用线性规划来建立决策模型。设 A 产品生产 x_1 公斤,B 产品生产 x_2 公斤。有数学模型:

$$\max z = 20x_1 + 35x_2$$

$$s.t.\begin{cases}8x_1 + \ \ 5x_2 \leqslant 400 \\ 2x_1 + \ \ 5x_2 \leqslant 200 \\ 9x_1 + 10x_2 \leqslant 500 \\ x_1, x_2 \geqslant 0\end{cases}$$

用单纯形求解得到：$x_1 = 20$（公斤），$x_2 = 32$（公斤），最大盈利为 1520（元）。

12.3 风险型决策问题

12.3.1 风险型决策的概念

这是决策分析研究得最多的一个领域，也是我们讨论的重点。它的主要特征是：

⑴ 存在着两个或两个以上的自然状态，未来究竟出现哪种状态，决策人不能事先肯定。但是，各种状态出现的概率可以事先知道。

⑵ 存在着决策希望达到的明确的目标，如收益最大或损失最小。

⑶ 存在着两个或两个以上的策略可供决策人选择，最后只选定一个策略。

⑷ 不同的策略在不同状态下的收益值或损失值可以计算出来。

根据状态概率的估计方式，可分为仅仅根据历史统计资料的经验分布的决策问题和依据理论分布推断的理论分布的决策问题。

例 3　在石油开发时，会遇到须决定是否在某处钻井的问题。尽管是在初步探明的石油层构造带上钻井，但由于地下情况极为复杂，在一个特定的位置上钻井钻后能否出油，油产量的高低等等，很难预先准确知道，只能根据过去类似地质情况的钻井记录或进一步的地面勘探结果进行估计。现在考虑一种简化了的情况：假定根据过去类似地质情况的钻井记录，某钻井位置地下无油的概率为 0.3，有油但产量一般的概率为 0.5，有油且产量很高的概率为 0.2。又知道开井后若地下无油则亏损 25 万元，有油但产量一般则净收益 20 万元，有油且产量很高则净收益 45 万元，如在该处不钻井则收益为零。试问，在上述情况下应作何种决定，钻井还是不钻井？

本例属于经验分布的风险决策问题。对此，仅仅考虑各个可能状态下的收益情况是不够的，还必须考虑各种状态出现的概率。经常采用的综合评价方案优劣的方法是期望值法。

10.3.2　期望值法

这里主要是指离散随机变量的数学期望，因为自然状态常常是离散的。然后，根

据算出的期望值的大小决定方案的优劣。

现在我们来计算例 3 的期望值。

设钻井方案为 A_1，不钻井方案为 A_2，无油状态为 S_1，有油但产量一般状态为 S_2，有油且产量很高状态为 S_3。由题设，收益值（单位：万元）为：

$$c_{11}=-25 \quad c_{12}=20 \quad c_{13}=45$$
$$c_{21}=0 \quad c_{22}=0 \quad c_{23}=0$$

各种状态的概率为：

$$P(S_1)=0.3 \quad P(S_2)=0.5 \quad P(S_3)=0.2$$

于是，钻井期望收益：

$$\begin{aligned}E(A_1)&=\sum_{j=1}^{3}c_{1j}P(S_j)\\&=-25\times0.3+20\times0.5+45\times0.2=11.5\text{（万元）}\end{aligned}$$

不钻井期望收益：

$$E(A_2)=\sum_{j=1}^{3}c_{2j}P(S_j)=0$$

由此，$E(A_1)>E(A_2)$，选择钻井方案比不钻井方案要好。

应当指出，如果只钻一口井，那么仍然存在钻井结果是无油而亏损的情况，这时钻井就不如不钻了。但是，如果在同等情况下，钻十口井，钻二十口井，或更多，那么可以肯定，钻井方案的收益要比不钻井方案高。

如果地下无油的概率增大，而地下有油的概率减少，那么不钻井方案将优于钻井方案。所以，决策问题中自然状态出现的概率对判断一个方案的优劣具有很大的影响。因此，对各种可能情况出现的概率要作认真调查和科学分析。

期望值法的计算步骤：

(1) 根据统计资料计算各个自然状态的概率；

(2) 计算每个策略在各个自然状态下的收益值；

(3) 计算每个策略的期望收益值；

(4) 根据期望收益值评价策略的优劣。

12.3.3 期望机会损失法

机会损失是指在任一自然状态下，没有选择最优策略而造成的利润损失或资源浪费，它在数值上等于在某种自然状态下可能采用的最优策略的收益减去在这种自然状态下实际采用的策略的收益之差。即机会损失

$$COL=\text{最优策略收益}-\text{实际策略收益}$$

下面先讨论例 3 的情形。

假定在这个策略中，我们做出了决定：钻井。可是钻井的结果地下无油，于是损失了 25 万元。对此，我们会感到某种遗憾：假若当初决定不钻井的话，可以避免这笔巨额损失。相反，如果钻井结果，发现地下有油，那么我们就不会为做出不钻井的决定而感到任何遗憾，因为没有承受任何损失，损失值为零。也就是说，对于地下无油的自然状态，钻井策略将损失 25 万元，不钻井损失为 0。对于地下有油但产量一般的自然状态，钻井策略损失为 0，而不钻井策略将有 20 万元的利润损失，也会使人感到某种遗憾。对于地下有油且产量很高的自然状态，钻井策略损失为 0，而不钻井将有 45 万元的利润损失。上述这些损失，在决策分析中称为机会损失，我们把这些结果列成机会损失表（表 12.2）。

表 12.2

自然状态	$S1$	$S2$	$S3$	期望机会损失 $E[COL_i]$
机会损失 COL_{ij} 概率 策略	0.3	0.5	0.2	
A_1	25	0	0	7.5（万）
A_2	0	20	45	19（万）

与计算期望收益值方法类似，计算期望机会损失采用公式

$$E[COL_i] = \sum_{j=1}^{n} COL_{ij}P(S_j), i = 1,2,\cdots,m$$

式中，$E[COL_i]$ 表示第 i 种策略的机会损失，COL_{ij} 表示第 i 种策略在第 j 种状态下的机会损失。

对于所讨论的钻井问题，应用此公式所得结果如上表。请注意，机会损失值都是正值，而收益值在盈利时是正值，亏损时是负值。因为 $E[COL_1] < E[COL_2]$，故应选择 A_1 为最优策略。用期望收益值法与机会损失值法所得结果完全相同，事实上两者是等价的。

12.3.4　决策树法

对经验分布的决策问题，基本方法是分析每一策略的收益的期望值，或分析其机会损失的期望值，除去前面介绍的方法外，还可以采用决策树方法进行。用树枝状的图形来表示这种方法的思路。我们通过一个多级决策的例子，来说明这个方法。

例 4　有一个化工原料工厂，由于某项工艺不够好，产品成本较高。现在计划将

该项工艺改进。取得新工艺有两种途径，一是自行研究，估计成功的可能性是 0.6；二是从国外引进，成功的可能性是 0.8。不论研究成功，还是谈判成功，生产规模都考虑两种方案，一是产量不变，一是增加产量。如果自行研究和谈判都失败，则采用原工艺进行生产，并保持原产量不变。

根据市场预测，估计今后 5 年内跌价的可能性是 0.1，保持中等价格的可能性是 0.5，涨价的可能性是 0.4。各状态下的收益值见表 12.3。试用决策树进行决策。

表 12.3

益损值 方案 / 价格状态（概率）	按原工艺生产	引进技术成功（0.8）		自行研究成功（0.6）	
		产量不变	增加产量	产量不变	增加产量
价格低落（0.1）	−100	−200	−300	−200	−300
价格中等（0.5）	0	50	50	0	−250
价格高涨（0.4）	100	150	250	200	600

解　第一步　画决策树，如图 12.1。

从左往右画。决策点 1，有两个方案分支：引进技术和自行研究。两方案分支的端点是方案节点 2 和 3。从节点 2 引出两个状态分支：失败（概率 0.2）和成功（概率 0.8），其中失败分支的端点是方案节点 4，而成功分支的端点是决策节点 5。由于对失败分支节点 4，意味着维持原工艺不变，不需作决策了，因此，只需要从节点 4 引出价格低、中、高三个状态分支即可，并标上相应的概率和收益值。对于成功分支的决策 5，又需引出两个方案分支：产量不变和增加产量。产量不变的方案分支，其节点 8 引出三个状态分支，填上各自的状态概率和相应的收益值。产量增加的方案分支，其节点 9 引出三个状态分支，也填上各自的状态概率和相应的收益值。

对方案节点 3，也引出两个状态分支，即自行研究成功和自行研究失败，各自标上状态概率 0.6 和 0.4。其中，成功分支的端点是决策点 6，它有两个可选方案，产量不变和增加产量，每个方案又各有三种状态。失败分支的端点是方案节点 7，它同方案节点 4 一样，标上三种状态的概率和相应的收益。

从图可看出，例中的决策有两级，第一级要作产量不变和产量增加两个方案的选择和决策，第二级才作引进技术和自行研究两个方案的选择和决策。

第二步　计算各方案的期望收益值。

从上往下计算。

对节点 4，其期望值为

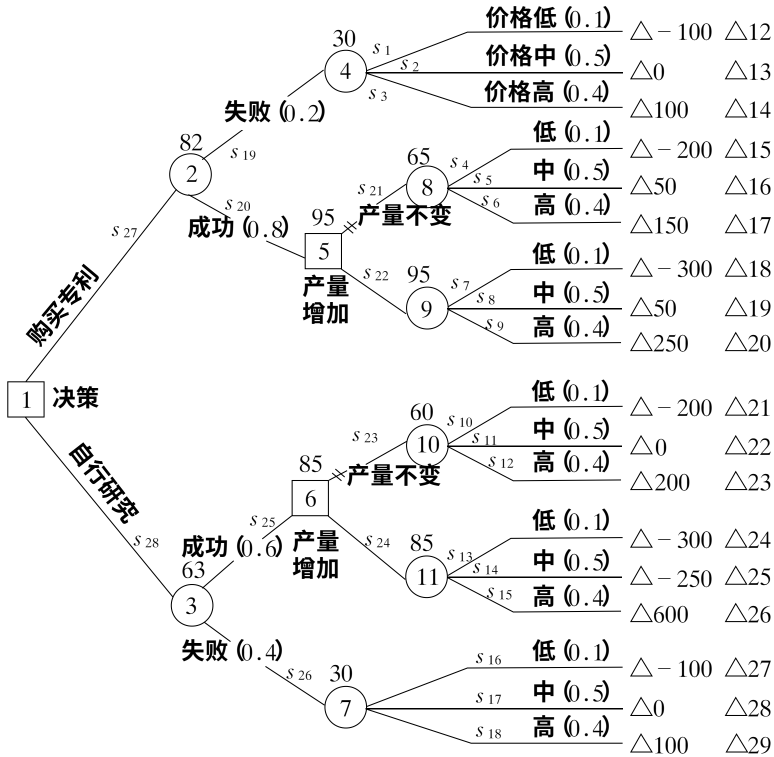


图 12.1 多级决策树

$$-100\times0.1+0\times0.5+100\times0.4=30$$

将此记在节点 4 处。

对节点 8,其期望值为:

$$-200\times0.1+50\times0.5+150\times0.4=65$$

将此记在节点 8 处。

节点 9,10,11,7 类似,所有结果都记在相应的节点处。

第三步 进行第一级决策。

从上往下进行。对节点 5,增加产量的方案的期望值大于产量不变方案的期望值,选择增加产量的方案,并将增加产量的期望值 95 记在决策点 5 处。

同理,对决策点 6,舍去产量不变方案,选择增加产量方案。并将该方案的期望值 85 记在决策点 6 处。

第四步 计算第二级决策方案的期望收益。

引进技术方案期望收益值为

$$30 \times 0.2 + 95 \times 0.8 = 82$$

记在节点 2 处。

自行研究方案期望收益值为：

$$85 \times 0.6 + 30 \times 0.4 = 63$$

将此记在节点 3 处。

第五步　进行第二级决策。

从图上清楚看出，应选取引进技术方案。至此，这个两级决策问题全部完成。

利用决策树方法进行决策的主要特点是：它的决策依据仍然是各方案的期望值，因此，在本质上它仍然是一种期望值决策方法。但是，由于这种方法可以用图形把决策过程形象地表示出来，因而使决策人可以有顺序有步骤地去考虑，以科学的推理去考虑各有关因素。

对于较复杂的决策问题，用决策树方法比较有效，对多级决策问题，尤其方便简捷。

12.3.5　矩阵法

记状态 S_j 的概率 $P(S_j)$ 为 P_j，$P=(P_1,\cdots,P_n)^T$，$A=(A_1,\cdots,A_m)$。收益矩阵 $D=(c_{ij})$。那么，方案 A_i 的期望

$$E(A_i)=\sum_{j=1}^{n} c_j P_j$$

若再记 $E(c)=[E(c_1),\cdots,E(c_m)]^T$。便有：

$$E(c)=DP$$

当给出收益矩阵 D 和状态概率向量 P 时，我们便可计算期望收益向量 $E(c)$，然后根据各期望收益值的大小选择最优策略。这就是决策问题中的矩阵法。当可选方案较多，自然状态也比较多时，使用矩阵法特别方便。用计算机进行两个矩阵的乘法也是办得到的。

12.3.6　贝叶斯（Bayesi）决策

在随机性决策中，对自然状态的概率分布 $P(S)$ 所作估计的精确性直接影响到收益期望值。因此，如果时间上允许，就往往要考虑出钱购买（或调查）有关的新信息。然后利用这些新信息修正原先对 $P(S)$ 做出的估计。并利用经过修正的概率分布作决策，这就是所谓的贝叶斯决策。其常用作法是：

1. 验前分析

决策者根据自己的经验和判断，估计 $P(S)$。然后，凭借这种验前概率分布并利

用收益值，由期望值准则做出决策。假定得到相应收益期望值为 E_1。

2．预验分析

在实际调查前，可先对购买（或调查）新信息是否合算做出分析，并做出相应的选择。

3．验后分析

实际调查后，根据所得结果对验前概率分布作修正，得出验后概率分布，再重新做出决策。

4．阶段分析

将调查收集信息的过程划分为若干阶段，在每一阶段上都作预验分析和验后分析。

在预验分析中如何进行决策呢？设调查可能得到的结果共有 r 种：$Z_1,\cdots,Z_r$。请注意，调查结果未必一定与客观的自然状态相符合。先根据以往的经验，估算在客观自然状态是 S_j，而调查结果为 A_k 的条件概率 $P(A_k \mid S_j)$（$k=1,\cdots,r$；$j=1,\cdots,n$）。再利用概率论中的贝叶斯公式，我们可以算出调查结果为 A_k 的条件下自然状态为 S_j 的条件概率：

$$P(S_j \mid A_k) = \frac{P(A_k \mid S_j)P(S_j)}{\sum_{i=1}^{n} P(A_k \mid S_i)P(S_i)}$$

$$k=1,\cdots,r;j=1,\cdots,n$$

假定调查结果为 A_k，根据上式算出 $P(S_j \mid A_k)$ 后，用它取代 $P(S_j)$，可算出最优策略下的收益期望值 $\hat{E}_k$

$$\hat{E}_k = \max\{\sum_{j=1}^{n} c_{ij}P(S_j \mid A_k);i=1,\cdots,m\}$$

此外，利用概率论中的全概率公式，可知调查结果为 A_k 的概率是：

$$P(A_k) = \sum_{j=1}^{n} P(A_k \mid S_j)P(S_j)$$

因此，估算出调查后决策所得的收益期望值为：

$$E_2 = \sum_{k=1}^{r} \hat{E}_k P(A_k)$$

而 $E_2 - E_1$ 就是进行调查使收益期望值增大的数值。显然，如果 $E_2 - E_1$ 大于调查的费用，那就可以认为调查是合算的。当然，要是实际进行了调查，并出现结果 A_k，那么 $P(S_j \mid A_k)$ 就是概率，$\hat{E}_k - E_1$ 就是这次调查真正增加出来的收益期望值。此时，如果要考虑再作一轮新的调查，那么，上次调查所得的验后概率可以作为下一次调查的验前概率来使用。

例 5　**为了开发某种新产品,某工厂需要对生产设备的投资规模作决策。设现有三种可供选择的方案,**A_1**:购买大型设备,**A_2**:购买中型设备,**A_3**:购买小型设备,未来市场对这些产品的需求情况也有三种,**S_1**:需求量较大,**S_2**:需求量中等,**S_3**:需求量较小,**

收益矩阵(单位:万元)
$$D=\begin{bmatrix}50 & 20 & -20\\ 30 & 25 & -10\\ 10 & 10 & 10\end{bmatrix}$$

并且,先期概率为 $P(S_1)=0.3$**,**$P(S_2)=0.4$**,**$P(S_3)=0.3$**。**

现在,调查结果的条件概率为(表 12.4**):**

表 12.4

$P(A_i \mid S_j)$	A_1	A_2	A_3
S_1	0.6	0.3	0.1
S_2	0.2	0.5	0.3
S_3	0.2	0.2	0.6

试做出贝叶斯分析。

解　**先求出** $P(A_1)$、$P(A_2)$、$P(A_3)$**,如表** 12.5

表 12.5

$P(A_i \mid S_j)P(S_j)$	A_1	A_2	A_3
S_1	0.18	0.09	0.03
S_2	0.08	0.20	0.12
S_3	0.06	0.06	0.18
$P(A_i)$	0.32	0.35	0.33

再算出 $P(S_j \mid A_i)$**,见表** 12.6**:**

表 12.6

$P(S_j \mid A_i)$	A_1	A_2	A_3
S_1	0.56	0.26	0.09
S_2	0.25	0.57	0.36
S_3	0.19	0.17	0.55

最后得到图 12.2 **所示的一个决策树。**

由图上看出，$E_1 = 17$（万元），$E_2 = 19.99$（万元）。因此，当调查费用不超过 $19.99 - 17 = 2.99$（万元）时，调查是合算的。

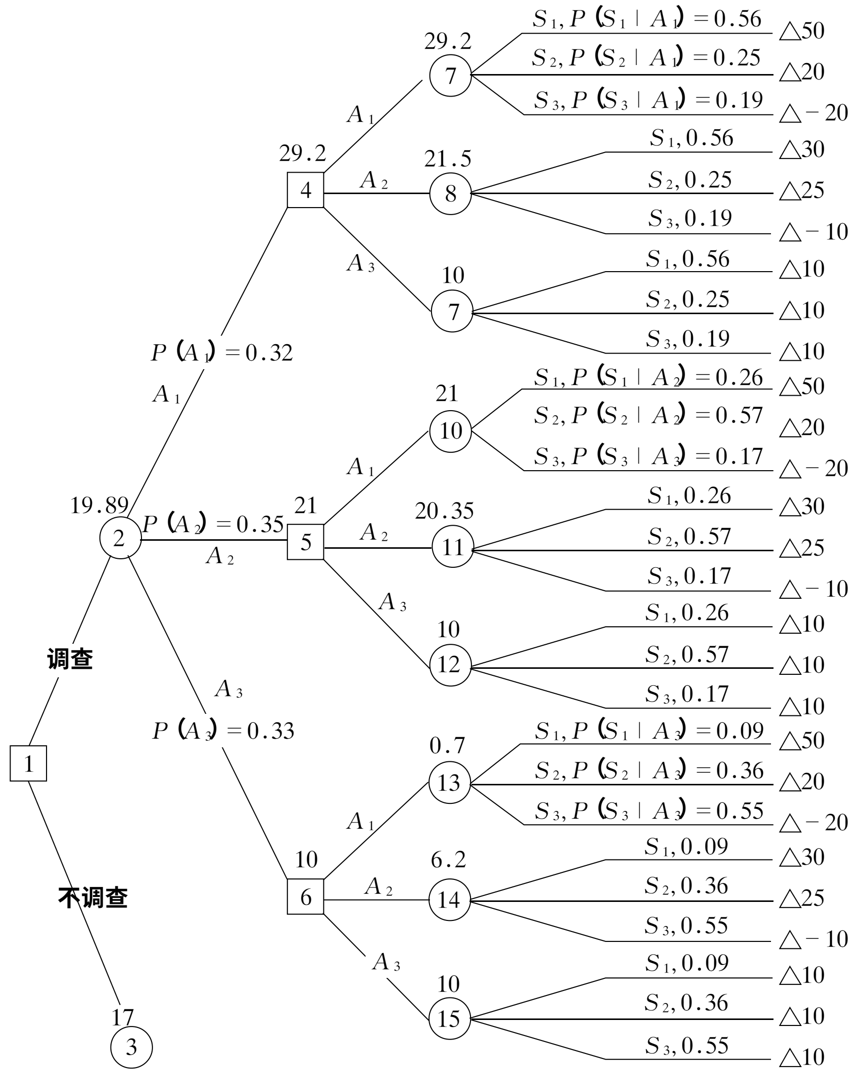


图 12.2

12.3.7 多行为决策的报童模型

街头的卖报童，每天要从报馆批发一批报纸，然后零售。报纸只能在当天出售，剩余的只能作废品处理，而每天对报纸的需求量又是不确定的。报童便面临这样一

个决策问题:应批发多少份报纸,使得每天卖报的收入或叫期望收益最多,换句话说,使得期望机会损失最少。

1. 连续变量的报童模型

本模型的主要假设条件是,产品的需求量 S 是连续变量,需求量的概率密度为 $f(S)$。相应地,决策变量也是连续变量,这里的决策变量是指从批发部进货的数量。设:

S 状态变量,指需求量;

A 决策变量,指批发进货量;

k_0 单位进货过量损失,指因销售不出造成的积压损失,它在数值上等于进货批发价格减去滞销处理价格;

k_u 单位进货不足损失,它在数值上等于零售价格减去批发价格之差。

则当决策变量 $A \geqslant S$ 时所产生进货过量损失为:

$$k_0(A - S)$$

期望过量损失为:

$$k_0\int_0^A (A - S) f(S)\,\mathrm{d}S, \quad 0 \leqslant S \leqslant A$$

同理,当决策变量 $A \leqslant S$ 时,将产生进货不足损失,期望不足损失为:

$$k_u\int_A^\infty (S - A) f(S)\,\mathrm{d}S, A \leqslant S \leqslant \infty$$

总期望机会损失应为期望过量损失和期望不足损失之和,故为:

$$K = k_0\int_0^A (A - S) f(S)\,\mathrm{d}S, 0 \leqslant S \leqslant A + k_u\int_A^\infty (S - A) f(S)\,\mathrm{d}S, A \leqslant S < \infty$$

当选择了最优策略 $\hat{A}$ 时,必有期望机会损失最小,即满足:

$$\left.\frac{\mathrm{d}K}{\mathrm{d}A}\right|_{\hat{A}} = 0$$

解方程

$$\frac{\mathrm{d}K}{\mathrm{d}A} = 0$$

得到

$$F(\hat{A}) = \frac{k_u}{k_0 + k_u}$$

这里,$F(A) = \int_0^A f(S)\,\mathrm{d}S$—— 积累分布函数,图 12.3 给出了其模型解释。

例 6　设有某种商品,其需求量 S 符合均匀分布。

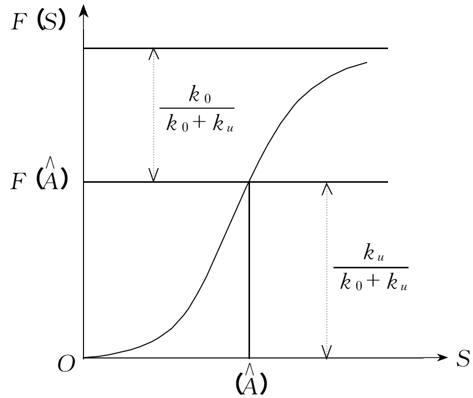


图 12.3

其概率密度为：

$$f(S)=\begin{cases}\dfrac{1}{5} & 10\leqslant S\leqslant 15\\ 0 & \text{其他}\end{cases}$$

已知该商品的购进单价为 12.5 元，出售单价为 15 元；如当天未能售出，第二天的处理价格为 11.25 元。试求合理的进货数量。

解　本例中的单位不足损失为：

$$k_u = 15 - 12.5 = 2.5\ (\text{元})$$

单位过量损失为：

$$k_0 = 12.5 - 11.25 = 1.25\ (\text{元})$$

则最优进货量 $\hat{A}$ 必定适合

$$F(\hat{A}) = \frac{k_u}{k_0+k_u} = \frac{2.5}{1.25+2.5} = \frac{2}{3}$$

即

$$\int_{10}^{\hat{A}} \frac{1}{5}\mathrm{d}S = \frac{2}{3} \Rightarrow \hat{A} = \frac{2}{3}\times 5 + 10 = 13.3333$$

注　最优决策的确定，不仅同需求量 S 的分布有关，而且同单位过量损失 k_0 和单位不足损失 k_u 有关。一般情况下，最优进货量 $\hat{A}$，同商品需求量的平均值是不同的。$\hat{A}$ 所满足的条件，只需是连续分布。

2. 离散变量的报童模型

现设商品的需求量 S 是离散变量，其概率为 $P(S)$ 决策变量 A 也是离散变量。

则最优决策变量 $\hat{A}$ 适合

$$\sum_{S=0}^{\hat{A}-1} P(S) \leqslant k_u (k_0 + k_u) \leqslant \sum_{S=0}^{\hat{A}} P(S)$$

即最优策略应当这样选择,它对应的累积概率将是大于比值$\frac{k_u}{k_0 + k_u}$的最小的数。

例 7　设有某商店的需求量 S,是离散型的,即只能整件出售。有过去的统计资料推知今后的需求量分布如下表:

表 12.7

需求量	10	11	12	13	14	15	合计
概率 $P(S)$	0.15	0.20	0.19	0.18	0.17	0.11	1.00

其余参数同例 5。试求最优订货量。

解　同例 5:$k_u = 2.5$,$k_0 = 1.25$,$\frac{k_u}{k_0 + k_u} = \frac{2}{3} = 0.6666$

再计算需求量的累积分布概率

$$P(10) = 0.15 < 0.6666$$

$$\cdots$$

$$P(10) + P(11) + P(12) = 0.54 < 0.6666$$

$$P(10) + P(11) + P(12) + P(13) = 0.72 > 0.6666$$

所以,最优订货量 $\hat{A} = 13$。

12.3.8　理论分布的决策问题

一般步骤是:

(1) 根据统计资料或事件特点,推断自然状态属于何种概率分布。常用的理论分布有正态分布,均匀分布,泊松分布等;

(2) 估计理论分布的参数;

(3) 应用报童模型进行决策。

12.4　不确定情况下的决策

前面讨论的确定型决策问题,是指已知的某种自然状态必然发生;风险型决策问题,虽然预先不知道哪种状态必然发生,但是每种自然状态发生的概率是可以通

过历史资料或主观估计的方法得到。而不确定情况下的决策，则只知道有几种自然状态可能发生，这些状态发生的概率并不知道。

下面介绍几种不确定型决策方法 —— 用到了支付表分析。

12.4.1 乐观法

也叫最大最大（maximax）原则。方法的思想基础是抱乐观态度：即使出现不利的情况，也未必会有多大损失，但一旦出现最有利的情况，却能得到很大的利益。

例8 今有5个行动方案 $A1,\cdots,A5$，4个自然状态 $S1,\cdots,S4$。自然状态发生的概率不知道。相应的收益值如表12.8：

表12.8

	S_1	S_2	S_3	S_4	各方案最大收益
A_1	4	5	6	7	7
A_2	2	4	6	9	9
A_3	5	7	3	5	7
A_4	3	5	6	8	8
A_5	3	5	5	5	5
决策	选择方案 A_2，收益9				

决策步骤是：

(1) 把各方案在各种自然状态下的最大收益求出来。

$$\max\{4,5,6,7\}=7$$
$$\max\{2,4,6,9\}=9$$
$$\max\{5,7,3,5\}=7$$
$$\max\{3,5,6,8\}=8$$
$$\max\{3,5,5,5\}=5$$

(2) 求各最大收益中的最大值。

$$\max\{7,9,7,8,5\}=9$$

得出最优策略 A_2，最大收益值为9。

12.4.2 悲观法

也叫最大最小（maximin）决策方法，或保守法。方法的思想基础是抱悲观态度：冒失行动会造成重大失误，最好还是从最不利的情况出发，向最好的方向努力。

例9 试将例8用悲观法决策。

决策步骤是：

⑴ 把每个方案在自然状态下的最小收益求出来。

$$\min\{4,5,6,7\} = 4$$
$$\min\{2,4,6,9\} = 2$$
$$\min\{5,7,3,5\} = 3$$
$$\min\{3,5,6,8\} = 3$$
$$\min\{3,5,5,5\} = 3$$

⑵ 求各最小收益值中的最大值。

$$\max\{4,2,3,3,3\} = 4$$

得出最优策略是 A_1,相应收益值是 4。

12.4.3 乐观系数法

又叫赫威兹(Hurwicz)决策准则。这个方法的特点是,对客观条件的估计既不那么乐观,也不那么悲观,主张折衷考虑(故又叫折衷值准则),用一个数字表示乐观程度,该数字称为乐观系数,记为 α,并规定 $0 \leqslant \alpha \leqslant 1$。用下式计算结果：

$$CV_i = \alpha \max\{c_{ij} \mid j\} + (1-\alpha)\min\{c_{ij} \mid j\}$$

若

$$CV_k = \max\{CV_i \mid i\}$$

则选取 $\hat{A} = A_k$。

显然,$\alpha = 1$ 时,即为乐观法;$\alpha = 0$ 时,即为悲观法。

例 10　设 $\alpha = 0.8$,用乐观系数法对例 8 进行决策。

解　易求

$$CV_1 = 0.8 \times 7 + 0.2 \times 4 = 6.4$$
$$CV_2 = 0.8 \times 9 + 0.2 \times 2 = 7.6$$
$$CV_3 = 0.8 \times 7 + 0.2 \times 3 = 6.2$$
$$CV_4 = 0.8 \times 8 + 0.2 \times 3 = 7.0$$
$$CV_5 = 0.8 \times 5 + 0.2 \times 3 = 4.6$$

从而

$$CV_2 = \max\{6.4,7.6,6.2,7.0,4.6\} = 7.6$$

对应策略为 A_2,故 $\hat{A} = A_2$。

注　α 值不同,可能得到不同的决策结果。如果该问题的形势比较乐观,α 值可取大一点;反之,α 取小一点。

12.4.4 后悔值准则

令 $w_j = \max\{c_{1j}, c_{2j}, \cdots, c_{mj}\}, j = 1,2,\cdots,n$. 它表示决策者在状态 S_j 下可获得的最大利益。又定义

$$r_{ij} = w_j - c_{ij}$$

它是选择 A_i 而产生的后悔值。

后悔值准则是：

(1) 计算后悔值 r_{ij} $(i = 1,2,\cdots,m; j = 1,2,\cdots,n)$

(2) 对每个 A_i 确定可能产生后悔最大的数值

$$R_i = \max\{r_{i1}, r_{i2}, \cdots, r_{in}\}, i = 1,2,\cdots,m$$

(3) 选取 A_k，使得

$$R_k = \min\{R_1, R_2, \cdots, R_m\}$$

例 11　用后悔值准则解例 8。

解　决策过程如下表，可知应采用策略 A_1 或 A_4。

表 12.9

c_{ij}	S_1	S_2	S_3	S_4	r_{i1}	r_{i2}	r_{i3}	r_{i4}	R_i
A_1	4	5	6	7	1	2	0	2	(2)
A_2	2	4	6	9	3	3	0	0	3
A_3	5	7	3	5	0	0	3	4	4
A_4	3	5	6	8	2	2	0	1	(2)
A_5	3	5	5	5	2	2	1	4	4
w_j	5	7	6	9					

12.4.5 不足理由原理 —— 等可能法

当决策人在决策过程中，不能肯定各种状态出现的概率时，就认为它们出现的概率是相等的(故又叫等可能法)：$P(S_j) = \dfrac{1}{n}, j = 1,\cdots,n$。然后用风险型决策方法，求出期望收益值，视其大小进行决策。这个方法是法国数学家拉普拉斯首先提出来的，所以又称拉普拉斯方法。

例 12　用等可能法解例 8。

解　因 $P(S_j) = 0.25,\ j = 1,2,3,4$。故

$$E[c] = D \cdot P(S)$$

$$
= \begin{bmatrix} 4 & 5 & 6 & 7 \\ 2 & 4 & 6 & 9 \\ 5 & 7 & 3 & 5 \\ 3 & 5 & 6 & 8 \\ 3 & 5 & 5 & 5 \end{bmatrix} \cdot \begin{bmatrix} 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{bmatrix} = \begin{bmatrix} 5.00 \\ 5.25 \\ 5.00 \\ 5.50 \\ 4.50 \end{bmatrix}
$$

$E[c_1] = E[c_4] = 5.50$，所以对 A_1 和 A_4 还得加以比较。通过计算

$$D[c_1] = E[c_1] - \min\{c_{1j} \mid j\} = 5.50 - 4 = 1.50$$

$$D[c_4] = E[c_4] - \min\{c_{4j} \mid j\} = 5.50 - 3 = 2.50$$

由 $D[c_1] < D[c_4]$，当选取方案 A_1（仍然是考虑机会损失最小）。

本节介绍了不确定型决策问题的几种方法，还有其他一些方法。通过各例可以看出，方法不同，结果可能各异。这里，除了决策者的主观态度和经验之外，关键是要对自然状态即客观情况作认真地调查研究，这包括对历史统计资料的整理和分析，对未来情况的估计和预测。这样做出的决策才会比较符合客观实际，才能达到预期的目的。

12.5 微机操作 —— 调用 YAJ 中，DSPB 决策支持系统

12.5.1 观察 DSPB 决策支持系统

这个程序提供了四个功能：平均值和方差分析，贝叶斯分析，支付表分析和决策树分析。借助均值和方差分析，在你输入不超过 100 个结果和概率后，程序将显示其简单平均和方差。借助贝叶斯分析，在你提供自然状态的先验概率和结果的条件概率以后，程序可以计算联合概率，边缘概率和后验概率。支付表分析，DSPB 要求你提供具有直到 40 个自然状态和 40 个策略的支付表，程序将显示在不同标准下的决策。

决策树分析，要求你提供不超过 80 个分支的树。分支和结点将按照支的先后关系从 1 开始，顺序编号。每一支应举出其策略或概率。对结点要指出它是机会点，或决策点，或终点。

12.5.2 问题输入

当输入一个问题时，请观察下列约定：

⑴ 为了输入你的问题，在下列方案中选出一个方案。

⑵ 然后输入你问题的数据。

⑶ 当输完数据后,按 *ENTER* 键。

⑷ 在同一荧屏上,你可以移动光标到需要改错的地方,按 *BACKSPACE* 键修改错误。

⑸ 当你对一屏的数据满意后,按 *SPACE BAR* 键。

⑹ 为了输入缺损值,按 *ENTER* 键即可。

为了输入你的问题,做出一个选择:

1. 均值和方差分析
2. 贝叶斯分析
3. 支付表分析
4. 决策树分析
0. 返回到函数菜单

12.5.3 问题解答

1. 下面以例 3 为例。

我们按照下面形式输入(使用支付表分析):

先验概率:S_1:0.3 S_2:0.5 S_3:0.2

状态	选择	
S_1	A_1:−25	A_2:0
S_2	A_1:20	A_2:0
S_3	A_1:45	A_2:0

结果:

决策标准 期望值法

决策将是:A_1

期望值:11.5

2. 仍以例 3 为例。

输入同上。

结果:

决策标准 期望后悔(期望机会损失)

决策将是:A_1

期望后悔值:7.5

3. 以例 4 为例。

严格按照图 12.1 进行输入。该图有 28 条分支。

支编号	支名字	开始结点	末结点	开始结点类型	概率	盈利
1	$s1$	4	12	2	0.1	−100
2	$s2$	4	13	2	0.5	0
3	$s3$	4	14	2	0.4	100
4	$s4$	8	15	2	0.1	−200
5	$s5$	8	16	2	0.5	50
6	$s6$	8	17	2	0.4	150
7	$s7$	9	18	2	0.1	−300
8	$s8$	9	19	2	0.5	50
9	$s9$	9	20	2	0.4	250
10	$s10$	10	21	2	0.1	−200
11	$s11$	10	22	2	0.5	0
12	$s12$	10	23	2	0.4	200
13	$s13$	11	24	2	0.1	−300
14	$s14$	11	25	2	0.5	−250
15	$s15$	11	26	2	0.4	600
16	$s16$	7	27	2	0.1	−100
17	$s17$	7	28	2	0.5	0
18	$s18$	7	29	2	0.4	100
19	$s19$	2	4	2	0.2	0
20	$s20$	2	5	2	0.8	0
21	$s21$	5	8	1	0	0
22	$s22$	5	9	1	0	0
23	$s23$	6	10	1	0	0
24	$s24$	6	11	1	0	0
25	$s25$	3	6	2	0.6	0
26	$s26$	3	7	2	0.4	0
27	$s27$	1	2	1	0	0
28	$s28$	1	3	1	0	0

结果：

结点	结点类型	期望值	决策
1	决策	82	$s27$
2	机会	82	
3	机会	63	

4	机会	30	
5	决策	95	s22
6	决策	85	s24
7	机会	30	
8	机会	65	
9	机会	95	
10	机会	60	
11	机会	85	

4. **以例 5 为例。**

结果显示：

1）**贝叶斯分析 —— 自然状态的先验概率**

S_1 :0.3　　S_2 :0.4　　S_3 :0.3

2）**贝叶斯分析 —— 条件概率**

状态	选择		
S_1	A_1 :0.6	A_2 :0.3	A_3 :0.1
S_2	A_1 :0.2	A_2 :0.5	A_3 :0.3
S_3	A_1 :0.2	A_2 :0.2	A_3 :0.6

3）**贝叶斯分析 —— 联合概率**

状态	选择		
S_1	A_1 :0.18	A_2 :0.09	A_3 :0.03
S_2	A_1 :0.08	A_2 :0.20	A_3 :0.12
S_3	A_1 :0.06	A_2 :0.06	A_3 :0.18

4）**贝叶斯分析 —— 选择的边缘概率**

A_1 :0.32　　A_2 :0.35　　A_3 :0.33

5）**贝叶斯分析 —— 后验或修正概率**

状态	选择		
S_1	A_1 :0.562500	A_2 :0.257143	A_3 :0.090909
S_2	A_1 :0.250000	A_2 :0.571429	A_3 :0.363636
S_3	A_1 :0.187500	A_2 :0.171429	A_3 :0.545455

5. **分别用四种方法对例 8 进行求解。**

使用该软件的支付表分析部分。

结果显示：

1）**决策标准：乐观法**（Maximax）

决策将是：A_2

乐观法值：9

2）**决策标准：悲观法**（Maximin）

决策将是：A_1

悲观法值：4

3）**决策标准：后悔值准则**（Minimax regret）

决策将是：A_1

后悔值准则值：2

4）**决策标准：不足理由原理**（Principle of insufficient reason）

决策将是：A1

不足理由原理值：5.5

第十三章　马尔科夫(Markov)过程

根据当前的状态和发展趋向去预测未来的状态,这是管理决策中重要的环节之一,马尔科夫分析就是这方面常用的方法。在系统论述这一方法之前,我们先介绍随机矩阵的概念。

13.1　正规随机矩阵

设 $U=(u_1,u_2,\cdots,u_n)^T$ 是一个 n 维随机向量,它的每个分量 u_i 都是非负数,且 $\sum_{i=1}^{n}u_i=1$,则称 U 为一个概率向量。如果 n 阶方阵 A 的每一行组成的向量都是 n 维概率向量,则称 A 是一个随机矩阵。

性质 1　若 $U\in R_n$ 是一个概率向量,$A=[a_{ij}]_{n\times n}$ 是一个随机矩阵,则 $Y=A^TU$ 也是一个概率向量。

证　$Y^T=U^TA=(\sum_{i=1}^{n}u_ia_{i2},\cdots,\sum_{i=1}^{n}u_ia_{in})$。注意到:

$$\sum_{j=1}^{n}a_{ij}=1,i=1,2,\cdots,n$$

于是便有

$$\sum_{j=1}^{n}y_j=\sum_{j=1}^{n}\sum_{i=1}^{n}u_ia_{ij}=\sum_{i=1}^{n}\sum_{j=1}^{n}u_ja_{ij}=\sum_{i=1}^{n}u_i(\sum_{j=1}^{n}a_{ij})=\sum_{i=1}^{n}u_i=1$$

性质 2　$A=[a_{ij}]_{n\times n}$,$B=[b_{ij}]_{n\times n}$ 都是随机矩阵,则 AB 也是一个随机矩阵。

证　用 $A_{i.}$ 表示由矩阵 A 的第 i 行组成的向量,由性质 1 可知 B^TA_i,都是概率向量。而

$$AB = \begin{bmatrix} A_1^T . B \\ M \\ A_n^T . B \end{bmatrix}$$

故 AB 的每一行组成的向量均为概率向量,即 AB 为随机矩阵。

设 $A = [a_{ij}]_{n\times n}$ 是一个随机矩阵,且存在一个正整数 k 使矩阵A^k (A^k 必是一个随机矩阵) 中每个元素都是正数,则称 A 是一个正规随机矩阵。

在马尔科夫分析中要用到下列重要结论:

定理 13.1 设 A 是一个正规随机矩阵,则

(1) 一定存在一个概率向量 X,使得 $A^TX = X$,且有

$$x_j > 0, \qquad j = 1,2,\cdots,n$$

(2) $A^k \to B$ ($k \to +\infty$),且 B 的每一行相同,全为 X^T;

(3) 对于任一 n 维概率向量 U,总有 $(A^k)^TU \to X$ ($k \to +\infty$)。

注 对矩阵

$$A = \begin{bmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

由

$$A^2 = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{3}{4} \end{bmatrix}$$

知 A 是一个正规随机矩阵,作为练习,大家可以用它验证这个定理。

13.2 马尔科夫链

设有一位顾客每天向一家商店买一包香烟,他购买香烟并不固定于某一种牌号。商店中 A,B,C,D,E 五种牌号的香烟,他都有可能购买。现以 ξm 表示此顾客在第 m 天购买的香烟牌号,则 ξm 是一个随机变量,它可取到的“值”为 A,B,C,D,E,这些值也称为 ξm 的状态,全体状态组成的状态集记为

$$N = \{A,B,C,D,E\}$$

一般地,设 $\{\xi_1,\cdots,\xi_m,\cdots\}$ 为一个随机变量的序列,

$$N = \{N_1,\cdots,N_n\}$$

(N 也可以是可列集) 为这些随机变量的状态集。若对于任意 k 和任意的正整数 $i_1,\cdots,i_{k+1}$,下面两个条件概率相等:

$$P(\xi_{k+1}=N_{i_{k+1}}\mid \xi_1=N_{i_1},\cdots,\xi_k=N_{i_k})=P(\xi_{k+1}=N_{i_{k+1}}\mid \xi_k=N_{i_k}) \tag{13.1}$$

则称随机变量序列 $\{\xi_1,\cdots,\xi_m,\cdots\}$ 是一个马尔科夫链——其中随机变量 ξ_{k+1} 所处某一状态的概率仅受紧接在它前面的随机变量 ξ_k 所处状态的影响,而与再前面的随机变量 $\xi_{k-1},\cdots,\xi_1$ 所处状态无关。这就是所谓的无后效性。

设 $\{\xi_m\mid m=1,2,\cdots\}$ 是一个马尔科夫链,i,j,k 均为正整数。如果下面的等式对于任一正整数 s 都成立,则称此链为一均匀的马尔科夫链:

$$P(\xi_{s+k}=N_j\mid \xi_s=N_i)=P(\xi_{k+1}=N_j\mid \xi_1=N_i) \tag{13.2}$$

式(13.2)说明,条件概率 $P(\xi_{s+k}=N_j\mid \xi_s=N_i)$ 与 s 无关。

对于均匀的马尔科夫链 $\{\xi_m\mid m=1,2,\cdots\}$,我们称条件概率

$$p_{ij}^{(k)}=P(\xi_{k+1}=N_j\mid \xi_1=N_i),i,j=1,2,\cdots,n \tag{13.3}$$

为从状态 N_i 到状态 N_j 的 k 步转移概率(在我们讨论的问题里假设 N 为有限集),并称矩阵

$$P^{(k)}=[p_{ij}^{(k)}]_{n\times n}$$

为 k 步转移矩阵。显然,

$$p_{ij}^{(k)}\geqslant 0,(i,j=1,2,\cdots,n);\sum_{j=1}^{n}p_{ij}^{(k)}=1,(i=1,2,\cdots,n)$$

例 1　某商店对前一天来店分别购买 A,B,C 牌号的顾客各 100 名的购买情况进行了统计(每天都买一包),统计结果如表 13.1。

表 13.1

顾客数目 / 本次购买的牌号 / 前次购买的牌号	A	B	C
A	20	50	30
B	20	70	10
C	30	30	40

不妨认为顾客在每次购买香烟时,只对他前次所吸香烟牌号有印象,因此商店可以采用一个均匀马尔科夫链 $\{\xi_m\mid m=1,2,\cdots\}$ 来描述顾客对香烟的需求状况。随机变量 ξ_m 表示顾客在第 m 天购买的香烟牌号,令 $N_1=A,N_2=B,N_3=C$。故此马尔科夫的转移概率矩阵(即一步转移矩阵 $P^{(1)}$)为:

$$P=\begin{bmatrix}0.2 & 0.5 & 0.3\\0.2 & 0.7 & 0.1\\0.3 & 0.3 & 0.4\end{bmatrix} \tag{13.4}$$

假定一位顾客在第一天购买牌号 A 的香烟,试问他在第三天购买牌号 B 的概率是多少?

这实际上要求的是二步转移概率

$$p_{12}^{(2)} = P(\xi_3 = N_2 \mid \xi_1 = N_1)$$

由(13.4)式可知,顾客在第二天购买牌号 A,B,C 的概率分别是:

$$p_{11} = 0.2, p_{12} = 0.5, p_{13} = 0.3$$

而第二天购买的牌号分别是 A,B,C 时,第三天购买牌号为 B 的概率,分别为:

$$p_{12} = 0.5, p_{22} = 0.7, p_{32} = 0.3$$

故可得下面结果(参看图 13.1(a)):

$$\begin{aligned} p_{12}^{(2)} &= p_{11}p_{12} + p_{12}p_{22} + p_{13}p_{32} \\ &= [p_{11} \quad p_{12} \quad p_{13}] \cdot \begin{bmatrix} p_{12} \\ p_{22} \\ p_{32} \end{bmatrix} = [0.2 \quad 0.5 \quad 0.3] \cdot \begin{bmatrix} 0.5 \\ 0.7 \\ 0.3 \end{bmatrix} = 0.54 \end{aligned}$$

一般地,均匀马尔科夫链的二步转移矩阵 $P^{(2)}$ 中任一元素 $p_{ij}^{(2)}$ 可应用下面公式来计算(参看图 13.1(b)):

$$p_{ij}^{(2)} = P_{i\cdot}^{T} P_{\cdot j} \tag{13.5}$$

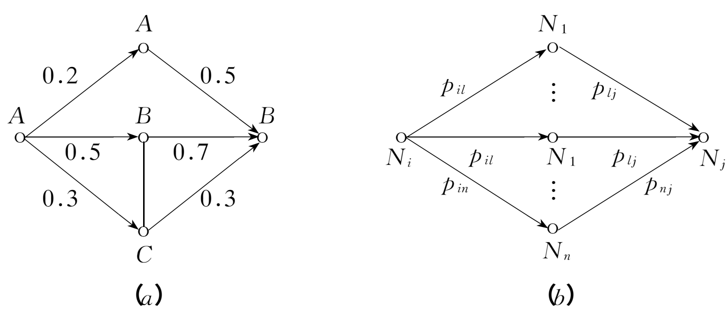


图 13.1

故由(13.5),便知上例中二步转移矩阵为:

$$P^{(2)} = P^{(1)}P^{(1)} = P^2 = \begin{bmatrix} 0.23 & 0.54 & 0.23 \\ 0.21 & 0.62 & 0.23 \\ 0.24 & 0.48 & 0.28 \end{bmatrix}$$

不难知道,式(13.4)的转移概率矩阵 P 是一个随机矩阵,而且还是一个正规随机矩阵。由定理 13.1 知,存在一个概率向量 $\pi = (\pi_1, \pi_2, \pi_3)^T$,使得:

$$P^T\pi = \pi \tag{13.6}$$

此 π_j 称为状态 N_j 的稳态概率。在用方程:

$$\pi_1 + \pi_2 + \pi_3 = 1$$

取代(13.6)式中第三个方程以后，解得 $\pi = (0.22, 0.57, 0.21)^T$。

由例可知，一般的均匀马尔科夫链有如下一些性质：

(1) 转移概率矩阵 $P = P^{(1)}$ 是一个随机矩阵。

(2) 由图 13.2 可知，k 步转移概率为：

$$p_{ij}^{(k)} = \sum_{l=1}^{n} p_{il}^{(k-1)} p_{lj}^{(1)} \tag{13.7}$$

因而 k 步转移矩阵为：

$$P^{(k)} = P^{(k-1)} \cdot P^{(1)} = P^k \tag{13.8}$$

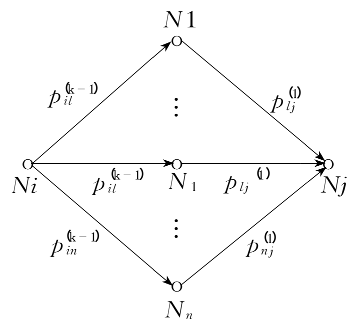


图 13.2

(3) 记 $v_{jk} = P(\xi_k = N_j)$，作概率向量

$$v_k = (v_{1k}, \cdots, v_{nk})^T \tag{13.9}$$

(v_1 就是均匀马尔科夫链 $\{\xi_k \mid k = 1, 2, \cdots\}$ 的初始分布) 则有：

$$v_j^{k+1} = \sum_{i=1}^{n} v_i^l p_{ij}^{(k)} \tag{13.10}$$

从而有：

$$v^{k+1} = (P^k)^T v^1 \tag{13.11}$$

(4) 若转移概率矩阵 P 是一个正规随机矩阵，则存在一个概率 $\pi = (\pi_1, \cdots, \pi_n)^T$，使得

$$P^T \pi = \pi \tag{13.12}$$

π_j 称为状态 N_j 的稳态概率，π 称为稳态概率向量。同时，由定理 13.2 和 13.3 可知：

①$P^{(k)} \to B \ (k \to +\infty)$，$B$ 的每一行向量相同，全为 π^T；

②$v^{k+1} = (P^k)^T v^1 \to \pi \ (k \to +\infty)$，也即

$$\lim_{K \to +\infty} P(\xi_k = N_j) = \pi_j \tag{13.13}$$

换句话说,均匀马氏链在经历一定时间的状态转移后,最终达到完全与初始状态无关的一种平衡状态。

13.3 马尔科夫分析

我们将以几个例子来说明马尔科夫分析法及其应用。

例2 某计算机实验室的一台计算机经常出故障,研究者每隔15分钟观察依次计算机运行的状况,收集了12个小时的数据(共作49次观测)。用1表示正常,用0表示不正常,所得的数据序列如下:

1110010011111110011110111111001111111110001101101

试问:① 若计算机在前一时段(15分钟)的状态为0,从本时段起此计算机能正常工作1小时(4个时段)的概率为多少?② 在运行较长时间后,计算机能正常工作1小时的概率为多少?

解 设 ξ_m 为第 m 个时段的计算机状态,它是一个随机变量,取值为0和1。令 $N_1 = 0$,$N_2 = 1$。可以认为,这是一个均匀的马尔科夫链。48次状态的转移情况是:

$$0 \to 0,6\text{ 次};0 \to 1,8\text{ 次};1 \to 0,8\text{ 次};1 \to 1,26\text{ 次}$$

因此,转移概率矩阵可以近似地假定为:

$$P = \begin{bmatrix} \frac{6}{6+8} & \frac{8}{6+8} \\ \frac{8}{8+26} & \frac{26}{8+26} \end{bmatrix} = \begin{bmatrix} \frac{6}{14} & \frac{8}{14} \\ \frac{8}{34} & \frac{26}{34} \end{bmatrix}$$

设稳态概率向量 $\pi = (\pi_1, \pi_2)^T$,则 π_1, π_2 应满足方程组:

$$\begin{cases} \frac{6}{14} \times \pi_1 + \frac{8}{34} \times \pi_2 = \pi_1 \\ \pi_1 + \pi_2 = 1 \end{cases}$$

由此可解得

$$\pi = (\pi_1, \pi_2)^T = \left(\frac{14}{48}, \frac{34}{48}\right)^T$$

这个结果与直接利用数据中0和1出现的频率统计结果相符。这样,题中①所求概率为:

$$p_{12} \cdot p_{22} \cdot p_{22} \cdot p_{22} = \frac{8}{14} \times \left(\frac{26}{34}\right)^3 = 0.256$$

题中②所求概率为:

$$\pi_2 \cdot \pi_2 \cdot \pi_2 \cdot \pi_2 = \left(\frac{34}{48}\right)^4 = 0.252$$

例 3　A, B, C **三家公司生产同一种产品。根据调查得知，各家公司产品上月占据市场的份额分别为** 50%，30%，20%。**设第** m **个月顾客对这种产品的各种牌号的需求状况可用一个均匀的马尔科夫链** $\{\xi_m \mid m = 1, 2, \cdots\}$ **来表示。令状态** N_1, N_2, N_3 **分别表示购买** A, B, C **公司产品，相应的概率转移矩阵为：**

$$P = \begin{bmatrix} 0.70 & 0.10 & 0.20 \\ 0.10 & 0.80 & 0.10 \\ 0.05 & 0.05 & 0.90 \end{bmatrix}$$

试问：本月和下月三家公司产品占据市场的份额各为多少？一年后三家公司产品占据市场的份额各为多少？

解　**上月购买** A **公司产品的顾客中，本月将有** 70% **仍买** A **公司产品；上月购买** B **公司产品的顾客中，本月将有** 10% **改买** A **公司产品；上月购买** C **公司产品的顾客中，本月将有** 5% **改买** A **公司产品。因而** A **公司产品本月占据市场的份额为：**

$$0.5 \times 0.70 + 0.3 \times 0.10 + 0.2 \times 0.05$$

$$= [0.5 \quad 0.3 \quad 0.2] \cdot \begin{bmatrix} 0.70 \\ 0.10 \\ 0.05 \end{bmatrix} = 0.39$$

其中初始概率向量 $(0.5, 0.3, 0.2)^T$ **表示上月三家公司产品占据市场的份额。于是描述本月各公司产品占据市场份额的概率向量为：**

$$(0.5, 0.3, 0.2)P = (0.39, 0.30, 0.31)$$

描述下月各公司产品占据市场份额的概率向量为：

$$\begin{aligned}(0.39, 0.30, 0.31)P &= (0.5, 0.3, 0.2)P^2 \\ &= (0.32, 0.29, 0.39)\end{aligned}$$

而描述一年后各公司产品占据市场份额的概率向量为：

$$(0.5, 0.3, 0.2)P^{12}$$

因为 P **是一个正规随机矩阵，故根据定理** 13.1(3) **可知，此概率向量近似于稳态概率向量** π**。求解方程组：**

$$\begin{cases} 0.70\pi_1 + 0.10\pi_2 + 0.05\pi_3 = \pi_1 \\ 0.10\pi_1 + 0.80\pi_2 + 0.05\pi_3 = \pi_2 \\ \pi_1 + \pi_2 + \pi_3 = 1 \end{cases}$$

可得 $\pi = (0.18, 0.23, 0.59)^T$

故 A **公司产品占据市场的份额，一年后将由年初的** 50% **下降到** 18%。**情况令人**

忧虑，A 公司必须考虑改进的措施。

例 4* 某商店经营一种易腐食品，出售后一个单位可获利 $a = 5$ 元，若当天售不出去，则每单位损失 $b = 3$ 元。该商店经理统计了连续 40 天的需求情况（不是实际销售量），现将所得数据列出如下：

$$3,3,4,2,2,4,2,3,4,4,4,3,2,4,2,3,3,4,2,2$$
$$4,3,4,3,2,3,4,2,3,2,2,3,4,2,4,4,3,2,3,3$$

经理想应用马尔科夫链来预测需求量，确定明天进货量。① 已知当天需求量为 3 个单位，明日应进货多少个单位？② 若不知当天需求量，明日应进货多少个单位？

解　令状态 $N_1 =$ 需求量为 2 个单位，状态 $N2 =$ 需求量为 3 个单位，状态 $N3 =$ 需求量为 4 个单位。状态转移的情况是：

$$N_1 \to N_1, 3\text{ 次}; N_1 \to N_2, 6\text{ 次}; N_1 \to N_3, 4\text{ 次}$$
$$N_2 \to N_1, 4\text{ 次}; N_2 \to N_2, 3\text{ 次}; N_2 \to N_3, 6\text{ 次}$$
$$N_3 \to N_1, 6\text{ 次}; N_3 \to N_2, 4\text{ 次}; N_3 \to N_3, 3\text{ 次}$$

于是，此马尔科夫链的转移概率矩阵 P 可近似地写为：

$$P = \begin{bmatrix} \frac{3}{13} & \frac{6}{13} & \frac{4}{13} \\ \frac{4}{13} & \frac{3}{13} & \frac{6}{13} \\ \frac{6}{13} & \frac{4}{13} & \frac{3}{13} \end{bmatrix}$$

求解方程组：

$$\begin{cases} \frac{3}{13} \times \pi_1 + \frac{4}{13} \times \pi_2 + \frac{6}{13} \times \pi_3 = \pi_1 \\ \frac{6}{13} \times \pi_1 + \frac{3}{13} \times \pi_2 + \frac{4}{13} \times \pi_3 = \pi_2 \\ \pi_1 + \pi_2 + \pi_3 = 1 \end{cases}$$

得稳态概率向量

$$\pi = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)^T$$

这样，应用边际分析的思想就可以回答经理的问题。由于：

$$\frac{b}{a+b} = \frac{3}{5+3} = \frac{3}{8} = 0.375$$

而

$$P(\text{明日需求量} = 4 \mid \text{当天需求量} = 3) = p_{23} = \frac{6}{13} \approx 0.462 > \frac{b}{a+b}$$

故对①的回答为：明日应进货 4 个单位。又因为：

$$P\text{（明日需求量}=4\text{）}=\pi_3=\frac{1}{3}<\frac{b}{a+b}$$

$$P\text{（明日需求量}\geqslant 3\text{）}=\pi_2+\pi_3=\frac{2}{3}>\frac{b}{a+b}$$

所以对②的回答为：明日进货 3 个单位。

13.4 微机操作——调用 YAJ 中，MKV 决策支持系统

13.4.1 观察 MKV 决策支持系统

这个程序使用马尔科夫过程模型，使你能求出系统在特殊状态和在特殊时间周期的概率。程序输入的数据包括：初始状态概率向量和转移概率矩阵。这个程序允许的状态最大数是 50。如果不知道初始状态概率，程序假定每个状态的概率相等。你可以用不超过 6 个字符为状态取名。缺损的名字是 $S_1, S_2, \cdots, S_n$。

你可以显示每个周期每个状态的概率。当两个连续周期之间相同状态的所有概率的差小于 10^{-6} 时，将会看到逼近稳定状态概率。你也可以指定周期数，而且，程序将停止在那个周期上，周期的最大数是 32 000。

13.4.2 问题输入

当输入一个问题时，请观察下列约定：

(1) 回答下列问题，它们反映了有关问题的一般信息。

(2) 如果不用缺损名，就输入状态名。

(3) 如果已经知道，就输入状态的初始概率向量。

(4) 输入转移概率矩阵。

在你的问题中，有多少状态（输入数 $\leqslant 50$）？

你知道初始状态概率向量吗（Y/N）？

你要用状态的缺损名（$S_1, S_2, \cdots, S_n$）吗（Y/N）？

然后输入初始状态概率向量，和转移概率矩阵，便可进行求解。

13.4.3 解答结果

我们以例 3 为例。

输入初始状态概率：S_1：0.5，S_2：0.3，S_3：0.2

输入状态转移矩阵：

从	到		
S_1	S_1∶0.7	S_2∶0.1	S_3∶0.2
S_2	S_1∶0.1	S_2∶0.8	S_3∶0.1
S_3	S_1∶0.05	S_2∶0.05	S_3∶0.9

结果显示：

初始状态概率 —— 迭代 0

S_1∶0.5000 S_2∶0.3000 S_3∶0.2000

状态概率 —— 迭代 1

S_1∶0.3900 S_2∶0.3000 S_3∶0.3100

状态概率 —— 迭代 2

S_1∶0.3185 S_2∶0.2945 S_3∶0.3870

状态概率 —— 迭代 3

S_1∶0.2717 S_2∶0.2868 S_3∶0.4415

状态概率 —— 迭代 4

S_1∶0.2410 S_2∶0.2787 S_3∶0.4803

状态概率 —— 迭代 5

S_1∶0.2206 S_2∶0.2711 S_3∶0.5084

状态概率 —— 迭代 6

S_1∶0.2069 S_2∶0.2643 S_3∶0.5287

状态概率 —— 迭代 7

S_1∶0.1977 S_2∶0.2586 S_3∶0.5437

状态概率 —— 迭代 8

S_1∶0.1914 S_2∶0.2538 S_3∶0.5547

状态概率 —— 迭代 9

S_1∶0.1817 S_2∶0.2499 S_3∶0.5629

状态概率 —— 迭代 10

S_1∶0.1841 S_2∶0.2468 S_3∶0.5691

……

最后迭代 —— 总迭代数 = 43

S_1∶0.1765 S_2∶0.2353 S_3∶0.5882

对每个状态的循环周期

S_1∶5.67 S_2∶4.25 S_3∶1.70

第十四章　时间序列预测——现象变动的趋势分析

14.1　统计预测的一般问题

14.1.1　预测的意义和种类

在社会经济活动中，无论从宏观的角度还是从微观的角度，都存在着许多未知的因素，影响着各级的管理决策。为了克服未知因素可能带来的消极后果，必须进行有科学根据的预测。所谓预测，是人们在观察和分析客观事物发展过程的历史及现状的基础上，通过对客观事物发展规律的认识，进而推断其未来状况的过程。换句话说，预测是在制定切实可行的计划时，为了避免可能产生的缺点和失误，而对事物的未来发展预先进行的多种方案的设计和研究。

预测可以按不同的标准进行分类。按预测的范围分，可分为宏观预测和微观预测；按预测时间的长短分，可分为长期预测（5 年以上），中期预测（2 ～ 5 年），短期预测（1 ～ 2 年），近期预测（1 年以内）；按预测的方法分，可以分为定性预测法和定量预测法。

14.1.2　预测的一般步骤

(1) 确定预测目的。

(2) 整理、分析历史和现实资料。

(3) 选择预测方法。

(4) 进行实际预测。

(5) 论证预测结果，分析预测误差。

14.1.3 预测中数据的修正与调整

运用统计预测方法对社会经济现象进行预测往往离不开数据。充分利用各种统计数据，从中寻求客观事物的变化规律，从而建立起比较合适的数学模型，才能保证预测的质量。但在实际预测时，常常会遇到有关数据的缺损或异常，因而要对其修正或调整。所用到的方法，通常有线性插值法、算术平均法、几何平均法。

14.1.4 影响动态序列水平的因素

动态序列各项发展水平的变化，是由许多复杂因素共同作用的结果。影响因素归纳起来大体有四类：

1. 长期趋势

它是指现象在一段较长的时间内，由于普遍的、持续的、决定性的基本因素的作用，使发展水平沿着一个方向，逐渐向上或向下变动的趋势。认识和掌握事物的长期趋势，可以把握事物发展变化的基本特点。

2. 季节变动

它是指现象受季节的影响而发生的变动。随着时序的更换，使现象呈周期重复的变化。认识和掌握季节变动，对于近期行动决策有重要的作用。

3. 循环变动

它是指现象发生周期比较长的涨落起伏的变动。通常所指的是经济发展荣衰不绝相替的变动。

4. 不规则变动

它是指现象除了受以上各种变动的影响之外，还受临时的、偶然因素或不明原因而引起的非周期性、非趋势性的随机变动。不规则变动是无法预知的。

现象变动趋势分析就是要把动态序列受各类因素的影响状况分别测定出来，搞清研究对象发展变化的原因及其规律，为预测未来和决策提供依据。

14.2 平均预测法

当现象的数量变化比较平稳，没有明显的递增或递减趋势时，运用平均预测法。平均预测法的基本思想是以时间序列中一定时期的平均水平作为未来的预测值。它的理论根据是时间序列的各项数值不仅受基本因素作用的影响，而且受随机因素变动的影响，通过多项数值的平均可以消除随机因素的影响，显示出现象的基本水平，

用来作为未来的预测值。

14.2.1 简单平均预测

简单平均预测就是以时间序列中近期的一段数值的简单算术平均数作为预测的依据。多用于预计生产计划完成程度。例如，预计月度生产计划完成情况，通常在每月的下旬，这时基本上已经掌握了上中旬的生产数据，只需预计下旬的产量，就可以推算该月生产计划可能完成的程度。我们可用中旬最后 5 天的平均日产量乘以下旬计划生产天数作为下旬预计完成数。这样，

$$\text{本月预计计划完成程度(\%)} = \frac{\text{至中旬止实际完成数} + \text{下旬预计完成数}}{\text{本月计划任务数}} \times 100\%$$

14.2.2 移动平均预测

移动平均预测法就是从时间序列中选择包括本期在内的最近几个时期的数值，计算它们的平均值，作为下一个时期的预测值。随着预测时期向前推移，相邻的 n 个时期数据也向前移动，故称为移动平均预测法。

设时间序列共有 t 个时期的数值，本期为 t 时期。先计算 t 期的移动平均数(移动时距取为 n)，称为 t 期的修匀值，记作 M_t，并以 M_t 作为 $t+1$ 期的预测值，记作 $\hat{X}_{t+1}$。计算方法如下：

$$\hat{X}_{t+1} = M_t = \frac{X_t + X_{t-1} + \cdots + X_{t-n+1}}{n}$$

移动平均数时距 n 的确定应根据时间序列的数值变动情况而定，一般说来，被平均的时间越长则修匀的能力也越大，因而消除随机因素的可能性也越大。所以，如果数值变动比较平缓，时距可取长一些，如果数值变动很剧烈，时距可取短一些。

现在以某百货公司逐月销售额预测为例。某公司各月销售额如表 14.1 所示，试以三个月的简单移动平均法，逐月预测下个月的销售额。

表 14.1

月份	销售额	三月移动总和	逐月预测值 $\hat{X}_{t+1} = M_t$
1	350	—	—
2	345	—	—
3	361	1056	—
4	356	1062	352
5	363	1080	354
6	358	1077	360

表 14.1（续）

月份	销售额	三月移动总和	逐月预测值 $\hat{X}_{t+1} = M_t$
7	347	1068	359
8	354	1059	356
9	370	1071	353
10	368	1092	357
11	348	1086	364
12	373	1089	363

如表所示，4～12 月的预测值就是各自上一月的移动平均数。第二年 1 月的销售额预测值就是本年 12 月的移动平均数，即：

$$\hat{X}_{12+1} = M_{12} = \frac{368 + 348 + 373}{3} = 363\text{（万元）}$$

简单移动平均预测法是假定平均数的各项资料，不论是近期还是远期，对于预测值起同等的作用。这种方法计算简便，适用于在趋势变动不明显的时间序列中使用。

通常认为，在时间序列中，近期的数值对预测期的影响较大，而远期的数值对预测期的影响较小。因此，在预测的时候，根据距离预测期的远近，给 n 个时期的数值以不同的权数，所求得加权平均数称为加权移动平均数。这种以加权移动平均数作为下一个时期的预测数的方法称为加权移动平均预测法。由于此方法的局限性，已被指数修匀法所代替。

14.2.3　指数修匀预测

指数修匀预测法也称指数平滑预测法。按修匀次数的多寡，而有指数一次修匀预测、指数二次修匀预测、指数三次修匀预测等等。这里所讨论的是指数一次修匀预测法，它实质上是移动平均预测法的改进。

按移动平均预测的规定，

$$\hat{X}_{t+1} = M_t = \frac{X_t + X_{t-1} + \cdots + X_{t-n+1}}{n}$$

$$\hat{X}_t = M_{t-1} = \frac{X_{t-1} + X_{t-2} + \cdots + X_{t-n}}{n}$$

因而有：

$$\hat{X}_{t+1} = M_t = \frac{X_t}{n} + M_{t-1} - \frac{X_{t-n}}{n}$$

由于 M_{t-1} 是 X_{t-n} 的最优估计值，用 M_{t-1} 代替 X_{t-n} 则有：

$$\hat{X}_{t+1} = M_t = \left(\frac{1}{n}\right) X_t + \left(1 - \frac{1}{n}\right) M_{t-1}$$
$$= \left(\frac{1}{n}\right) X_t + \left(1 - \frac{1}{n}\right) \hat{X}_t$$

这个式子表明，$t+1$ 期的预测值 X_{t+1} 等于 t 期的实际值 X_t 和 t 期的预测值 $\hat{X}_t$ 的加权平均数，权数分别为 $\frac{1}{n}$ 和 $\left(1-\frac{1}{n}\right)$。

整理上式可得：

$$\hat{X}_{t+1} = \hat{X}_t + \frac{1}{n}\left(X_t - \hat{X}_t\right)$$

这个式子表明，$t+1$ 期的预测值实际上是 t 期预测值加以调整的结果，这个调整数就是 t 期实际值和预测值差额的 $\frac{1}{n}$。

以上都是就简单的移动平均数而言。如果令 $\frac{1}{n} = \alpha$，$(0 < \alpha < 1)$，并以 $S_t^{(1)}$ 代 M_t，则有：

$$\hat{X}_{t+1} = S_t^{(1)} = \alpha X_t + (1-\alpha) S_{t-1}^{(1)}$$
$$= \alpha X_t + (1-\alpha) \hat{X}_t$$

这是指数修匀（平滑）预测的基本公式。式中 $S_t^{(1)}$ 称为 t 期的平滑值，α 称为平滑常数。这个预测公式表明，$t+1$ 期的预测值是 t 期实际值 X_t 与 t 期预测值 $\hat{X}_t$ 加权平均的结果，其权数分别是 α 和 $1-\alpha$。因此，在进行预测时，我们只需要有本期的实际值、预测值和适当的平滑常数 α，就可以由上述预测公式对下一期进行预测。方法简便易行，资料可以保存逐期连续使用。

整理上式可得：

$$\hat{X}_{t+1} = S_t^{(1)} = S_{t-1}^{(1)} + \alpha\left(X_t - S_{t-1}^{(1)}\right)$$
$$= \hat{X}_t + \alpha\left(X_t - \hat{X}_t\right)$$

这是指数修匀预测公式的变形。它 $t+1$ 期预测值 $\hat{X}_{t+1}$ 实际上是 t 期预测值加以调整的结果。这个调整数就是 t 期实际值和预测值差额乘上一定比例数 α。调整的分量主要决定于平滑常数 α 的取值。

下面证明，指数修匀预测也就是加权移动平均预测，只不过把平均范围扩展到整个时间序列。实因：

$$\hat{X}_{t+1} = \alpha X_t + (1-\alpha) \hat{X}_t$$
$$= \alpha X_t + (1-\alpha)\left[\alpha X_{t-1} + (1-\alpha) \hat{X}_{t-1}\right]$$

$$= \alpha X_t + \alpha(1-\alpha)X_{t-1} + (1-\alpha)^2\hat{X}_{t-1}$$

$$\cdots$$

$$= \alpha X_t + \alpha(1-\alpha)X_{t-1} + \cdots + \alpha(1-\alpha)^{t-1}X_1 + (1-\alpha)^t\hat{X}_1$$

当 t 很大时，上式最后一项接近于 0，可以略去不计。故有：

$$\hat{X}_{t+1} = \sum_{j=0}^{t-1}\alpha(1-\alpha)^j X_{t-j}$$

上式表示，预测值 $\hat{X}_{t+1}$ 是时间序列 $X_t, X_{t-1}, \cdots, X_1$ 的加权平均数，其权数分别是 $\alpha, \alpha(1-\alpha), \cdots, \alpha(1-\alpha)^{t-1}$。权数之和的极限为 1：

$$\lim_{t\to\infty} A = \lim_{t\to\infty}\alpha(1-\alpha)^j = \frac{\alpha}{1-(1-\alpha)} = 1$$

于是，我们得到：

$$\hat{X}_{t+1} = S_t^{(1)} = \alpha\sum_{j=0}^{\infty}(1-\alpha)^j X_{t-j}$$

这是指数一次修匀（平滑）预测的一般公式，以后还要多次应用，要认真掌握。当然，实际生活中，t 总是有限的，所以上述公式计算的预测值只是近似的加权平均数。

恰当选择平滑常数 α 在预测中是个关键问题。从上式可以看出，移动平均数的权数分配是按指数方式递减的。而递减的速度则决定于 α 值的大小。α 越大，权数递减速度越快，强调近期变量的作用。一般，如序列的变化有明显的上升或下降趋势，受趋势性因素影响较大，应取较大的 α 值 如序列的变化比较平稳，或突然上升又突然下降，受随机性因素影响明显的，应取较小的 α 值。

关于初始平滑值 $S_0^{(1)}$ 的确定，也即确定时间序列第一期的预测值 $\hat{X}_1$，这是关系到预测的滑动起点水平，缺少这个数值则 t 期的预测也就没有条件。通常认为，由于 t 期离序列的初期相对说很远，$\hat{X}_1$ 值对于 t 期的预测值影响已经很小，为方便起见，可以将 X_1 作为 $\hat{X}_1$。这样，我们就可以根据时间的实际数据进行逐期预测。

现仍以表 14.1 的资料为例。平滑常数分别取 α 为 0.8 和 0.2，进行各月预测，并比较两种预测的平均绝对误差。假定第一个月的预测值和实际值相同。

我们是这样算的：

$\alpha = 0.8$ 时，三月份 $X_3 = 361$，

$\hat{X}_3 = \alpha X_2 + (1-\alpha)\hat{X}_2 = 0.8\times 345 + 0.2\times 350 = 346$，

预测绝对误差 $= |X_3 - \hat{X}_3| = |361 - 346| = 15$；

四月份 $X_4 = 356$，$\hat{X}_4 = \alpha X_3 + (1-\alpha)\hat{X}_3 = 0.8\times 361 + 0.2\times 346 = 358$；

预测绝对误差 $= |X_4 - \hat{X}_4| = |356 - 358| = 2$，其余各项以此类推。

$\alpha = 0.8$ 时，预测平均绝对误差 $= \dfrac{\sum |X_t - \hat{X}_t|}{n} = \dfrac{107.8}{12} = 8.98$；

$\alpha = 0.2$ 时，预测平均绝对误差 $= \dfrac{\sum |X_t - \hat{X}_t|}{n} = \dfrac{99.2}{12} = 8.27$。

$\alpha = 0.2$ 时，预测平均绝对误差较小，所以 α 取 0.2 值进行指数修匀预测更为适宜。

14.3 趋势预测法

14.3.1 趋势线的测定

平均预测法是以现象过去已经达到的平均数量水平作为未来的预测值。这种方法只适宜于现象的数量变化比较平稳的水平型时间序列。如果现象的数量变化具有明显的上升或下降的趋势，用时间序列的某一变量或几个变量的平均水平作为未来的预测值就会发生显著的时间滞后，有比较大的系统预测误差。

当时间序列具有明显的趋势变动时，首先要判断趋势变动的形式。所谓趋势变动是指变量依时间顺序按一定方向、可用数学方程式表达的轨迹运动。社会经济现象的发展趋势决定于现象的内在联系和外部条，只有深入分析，把握现象发展的根据和特征，才能了解事物变化的趋势。就发展的数量关系来探讨，主要可以从以下两方面考虑。

一是从现象变化的增量特征来考虑。例如，变量按时间顺序其单位时间一级增量大体相等即存在线性变动趋势；如果单位时间二级增量大体相等即存在抛物线变动趋势等等。

二是从现象发展的速度特征来考虑。例如，现象各期的递增速度大体相同，即存在指数曲线趋势等等。

在这个基础上就可以引出现象变化趋势的类型，并给出相应的趋势线的方程式，作为趋势预测的根据。按方程式的表现形式来分，可以分为线性趋势和非线性趋势。

14.3.2 指数二次修匀预测

指数二次修匀预测也称指数二次平滑预测。它是在一次修匀的基础上，对第一次修匀值再进行第二次修匀，使序列变动更加平滑，然后观察它是否接近于线性趋势。如果属于线性型趋势，则以二次修匀值为基础确定直线方程参数，进行预测。

设 $S_t^{(1)}$、$S_t^{(2)}$ 分别表示序列 t 期的第一次修匀值和第二次修匀值。$S_t^{(2)}$ 修匀值的计算程序和第一次修匀计算 $S_t^{(1)}$ 时相同，只是现在把 $S_t^{(1)}$ 作为原序列再进行一次修匀。在第一次修匀时，规定 t 期的修匀值是由 t 期的实际值 X_t 与 t 期预测值 $\hat{X}_t$ 加权平均得到的，其权数分别为 α 和 $1-\alpha$，而 t 期的预测值就是 $t-1$ 期的修匀值，$\hat{X}_t = S_{t-1}^{(1)}$，因此可以说，$S_t^{(1)}$ 是由 X_t 与 $S_{t-1}^{(1)}$ 分别以权数 α 和 $1-\alpha$ 的加权平均得到的。以此类推，t 期的二次修匀值 $S_t^{(2)}$ 就是由 t 期的一次修匀值 $S_t^{(1)}$ 与 $t-1$ 期的二次修匀 $S_{t-1}^{(2)}$ 加权平均得到的，其权数分别是 α 和 $1-\alpha$。经过整理，我们得到以下三个基本关系式：

$$S_t^{(2)} = \alpha S_t^{(1)} + (1-\alpha) S_{t-1}^{(2)}$$

$$S_t^{(2)} = S_{t-1}^{(2)} + \alpha (S_t^{(1)} - S_{t-1}^{(2)})$$

$$S_t^{(2)} = \alpha \sum_{j=0}^{\infty} (1-\alpha)^j S_{t-j}^{(1)}$$

但在这里，我们没有理由以 $S_t^{(2)}$ 作为 $t+1$ 期的预测值，因为在存在趋势变动的情况下这样将发生更大的时间滞后误差。而应该求出合适的直线趋势方程，进行预测。设 $\hat{X}_{t+T}$ 为 $t+T$ 期的预测值，则有：

$$\hat{S}_{t+T} = a_t + b_t T$$

式中，a_t 与 b_t 表示 t 期的直线方程参数。为了确定方程的参数值，必须建立 a_t、b_t 与修匀值 $S_t^{(1)}$、$S_t^{(2)}$ 之间的联。

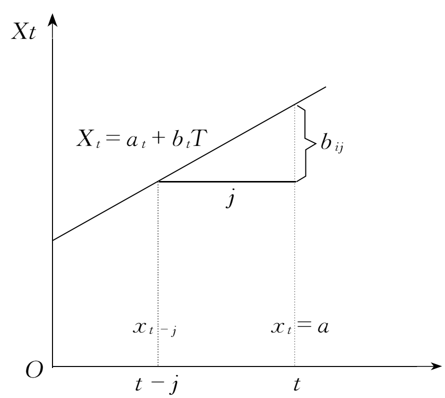


图 14.1

从图 14.1 的关系中可以看出：

$$X_{t-j} = a_t - b_{tj}$$

所以，

$$S_t^{(1)}=\alpha\sum_{j=0}^{\infty}(1-\alpha)^j(a_t-b_tj)=a_t\alpha\sum_{j=0}^{\infty}(1-\alpha)^j-b_t\alpha\sum_{j=0}^{\infty}(1-\alpha)^jj$$

$$=a_t-(1-\alpha)\frac{b_t}{\alpha}$$

式中：

$$\alpha\sum_{j=0}^{\infty}(1-\alpha)^j=1$$

$$\alpha\sum_{j=0}^{\infty}(1-\alpha)^jj=\frac{1-\alpha}{\alpha}$$

同理，

$$S_{t-j}^{(1)}=\alpha\sum_{i=0}^{\infty}(1-\alpha)^iX_{t-j-i}=\alpha\sum(1-\alpha)^i[a_t-b_t(j+i)]$$

$$=a_t\alpha\sum_{i=0}^{\infty}(1-\alpha)^i-b_tj\alpha\sum_{i=0}^{\infty}(1-\alpha)^i-b_t\alpha\sum_{i=0}^{\infty}i(1-\alpha)^i=a_t-b_tj-(1-\alpha)\frac{b_t}{\alpha}$$

又由于：

$$S_t^{(2)}=\alpha\sum_{j=0}^{\infty}(1-\alpha)^jS_{t-j}^{(1)}=\alpha\sum_{j=0}^{\infty}(1-\alpha)^j[a_t-b_tj-(1-\alpha)\frac{b_t}{\alpha}]$$

$$=a_t-(1-\alpha)\frac{b_t}{\alpha}-(1-\alpha)\frac{b_t}{\alpha}=a_t-2(1-\alpha)\frac{b_t}{\alpha}$$

联立方程：

$$\begin{cases}S_t^{(1)}=a_t-1(1-\alpha)\dfrac{b_t}{\alpha}\\[2ex] S_t^{(2)}=a_t-2(1-\alpha)\dfrac{b_t}{\alpha}\end{cases}$$

解得：

$$a_t=2S_t^{(1)}-S_t^{(2)}$$

$$b_t=\frac{\alpha}{1-\alpha}\cdot(S_t^{(1)}-S_t^{(2)})$$

这样确定了平滑常数 α 值，就可以从已知的 $S_t^{(1)}$、$S_t^{(2)}$ 值计算出方程的参数 a_t、b_t。当然，参数随着 $S_t^{(1)}$、$S_t^{(2)}$ 的改变而改变，每一期的修匀值都可以有自己的趋势线。由于预测期的逐期向前推移，趋势线必然不断改变。

二次修匀预测也必须先确定初始的修匀值 $S_0^{(1)}$ 和 $S_0^{(2)}$ 作为滑动的起点。由于二次修匀的现象具有直线的趋势，所以，不宜应用第一期的实际值 X_1 作为预测值 $S_0^{(1)}$，也不宜用 $S_0^{(1)}$ 来代替 $S_0^{(2)}$，应该考虑到时间滞后对它们的影响。一般可以用分

割平均法估计时间序列直线趋势的两个参数 $\hat{a}$、$\hat{b}$，然后代入下面方程求初始的修匀值：

$$\begin{cases} S_0^{(1)} = \hat{a} - (1-\alpha)\dfrac{\hat{b}}{\alpha} \\ S_0^{(2)} = \hat{a} - 2(1-\alpha)\dfrac{\hat{b}}{\alpha} \end{cases}$$

现在，举例说明二次修匀的计算过程。我国 1978－1987 年的汽车产量如表 14.2 第三栏所示。

表 14.2

年序 (t)	年份	汽车产量（万辆）(x_t)	一次修匀值 $S_t^{(1)} = \alpha X_t + (1-\alpha) S_{t-1}^{(1)}$	二次修匀值 $S_t^{(2)} = \alpha S_t^{(1)} + (1-\alpha) S_{t-1}^{(2)}$
1	1978	14.9	12.3	10.0
2	1979	18.6	16.7	14.7
3	1980	22.2	20.3	18.6
4	1981	17.6	18.4	18.5
5	1982	19.6	19.3	19.1
6	1983	24.0	22.6	21.5
7	1984	31.6	28.9	26.7
8	1985	43.7	39.3	35.5
9	1986	37.0	37.7	37.0
10	1987	47.2	44.3	42.1

确定平滑常数 $\alpha = 0.7$，初始修匀值 $S_0^{(1)}$、$S_0^{(2)}$ 估计如下：

设直线趋势方程：$X = \hat{a} + \hat{b}t$。其中，

$$(t_1, X_1) = \left(\frac{1+2+3+4+5}{5}, \frac{14.9+18.6+22.2+17.6+19.6}{5}\right) = (3, 18.6)$$

$$(t_2, X_2) = \left(\frac{6+7+8+9+10}{5}, \frac{24+31.6+43.7+37+47.2}{5}\right) = (8, 36.7)$$

代入直线趋势方程 $X = \hat{a} + \hat{b}t$ 得：

$$\hat{a} + 3\hat{b} = 18.56$$

$$\hat{a} + 8\hat{b} = 36.7$$

解得：$\hat{a} = 7.708$，$\hat{b} = 3.624$

再将 $\hat{a}$、$\hat{b}$ 值代入下式求初始修匀值：

$$S_0^{(1)} = a_t - (1-\alpha)\frac{\hat{b}}{\alpha} = 7.708 - 0.3 \times \frac{3.624}{0.7} = 6.16$$

$$S_0^{(2)} = a_t - 2(1-\alpha)\frac{\hat{b}}{\alpha} = 7.708 - 0.6 \times \frac{3.624}{0.7} = 4.60$$

根据方程：

$$S_t^{(1)} = \alpha X_t + (1-\alpha) S_{t-1}^{(1)}$$
$$S_t^{(2)} = \alpha S_t^{(1)} + (1-\alpha) S_{t-1}^{(2)}$$

逐年计算一次修匀值和二次修匀值。并将这些结果填入表 14.2 的第 4、5 栏内。

进一步，根据最近一年的修匀值，计算 t 年趋势方程的参数 a_t 和 b_t：

$$a_t = 2S_t^{(1)} - S_t^{(2)} = 2 \times 44.3 - 42.1 = 46.5$$

$$b_t = \alpha \frac{S_t^{(1)} - S_t^{(2)}}{1-\alpha} = 0.7 \times \frac{44.3 - 42.1}{0.3} = 5.1$$

预测的直线方程为：

$$\hat{X}_{t+T} = a_t + b_t T = 46.5 + 5.1T$$

如果要预测 1990 年（即第 13 期）的汽车产量，则 $T = 3$，代入方程即可：

$$\hat{X}_{13} = 46.5 + 5.1 \times 3 = 61.8 \text{（万辆）}$$

14.4 线性回归预测 —— 最小平方法预测

在趋势预测中，也可以利用时间序列以最小平方法求现象变动的趋势线，并进行预测。这时以时间 t 为自变量，以时间序列 X_i 为因变量。直线预测方程为：

$$\hat{x} = a + bt$$

我们要求

$$S(a,b) = \sum (x_i - \hat{x})^2 = \sum (x_i - a - bt_i)^2$$

最小。因此，a、b 应该满足条件：

$$\begin{cases} \dfrac{\partial S}{\partial a} = 2\sum (x_i - a - bt_i)(-1) = 0 \\ \dfrac{\partial S}{\partial b} = 2\sum (x_i - a - bt_i)(-t_i) = 0 \end{cases}$$

也即

$$\begin{cases} \sum x_i = na + b\sum t_i \\ \sum t_i x_i = a\sum t_i + b\sum t_i^2 \end{cases}$$

利用这一方程组就可以确定参数 a、b：

$$b = \begin{vmatrix} n & \sum x_i \\ \sum t_i & \sum t_i x_i \end{vmatrix} \Big/ \begin{vmatrix} n & \sum t_i \\ \sum t_i & \sum t_i^2 \end{vmatrix}$$

$$= \left[n \sum t_i x_i - \sum t_i \sum x_i\right] \Big/ \left[n \sum t_i^2 - \left(\sum t_i\right)^2\right]$$

$$a = x - bt$$

由此给出具体的预测方程。

这实际是关于时间和时间序列之间的一条回归直线方程。

我们仍然以表 14.2 为例，列表计算如表 14.3：

表 14.3

	年序 (t)	年份	汽车产量 (x)	t^2	tx	x^2	$\hat{x}$
	1	1978	14.9	1	14.9	222.01	11.98
	2	1979	18.6	4	37.2	345.96	15.46
	3	1980	22.2	9	66.6	492.84	18.94
	4	1981	17.6	16	70.4	309.76	22.42
	5	1982	19.6	25	98.0	384.16	25.90
	6	1983	24.0	36	144.0	576.00	29.38
	7	1984	31.6	49	221.2	998.56	32.86
	8	1985	43.7	64	349.6	1909.69	36.34
	9	1986	37.0	81	333.0	1369.00	39.82
	10	1987	47.2	100	472.0	2227.84	43.30
总和	55		276.4	385	1806.9	8835.82	

代入 a、b 计算公式：

$$b = \frac{10 \times 1806.9 - 55 \times 276.4}{10 \times 385 - 552}$$

$$= \frac{2867}{825}$$

$$= 3.48$$

$$a = \frac{276.4}{10} - b\frac{55}{10} = 27.64 - 3.48 \times 5.5 = 8.5$$

得到线性回归直线方程式：

$$\hat{x} = 8.5 + 3.48t$$

据此，预测 1990 年 ($t = 13$) 汽车产量数为：

$$\hat{x}_{13} = 8.5 + 3.48 \times 13 = 53.74 \text{ (万辆)}$$

附　变量 t 和变量 x 之间的线性相关系数 r 由下式测定：

$$
\begin{aligned}
r &= \frac{n\sum t_i x_i - \sum t_i \sum x_i}{\sqrt{n\sum t_i^2 - (\sum t_i)^2} \times \sqrt{n\sum x_i^2 - (\sum x_i)^2}} \\
&= \frac{10 \times 1806.9 - 55 \times 276.4}{\sqrt{10 \times 385 - 55^2} \times \sqrt{10 \times 8835.82 - 276.4^2}} \\
&= \frac{2867}{\sqrt{825} \times \sqrt{11\,961.24}} = \frac{2867}{28.72 \times 109.36} = \frac{2867}{3140.82} = 0.91
\end{aligned}
$$

看出，属于高度相关。

14.5　预测误差分析

14.5.1　预测误差的意义

统计预测误差是指未来的实际观测值与预测值之间的离差。预测误差综合反映预测工作的质量，是预测工作中最引人注意的问题。如果预测误差超过一定的限度，则统计预测本身就失去意义，甚至会带来计划、决策、行动等一系列损失。但我们也要认识到，由于社会经济条件和影响因素的消长变化是如此的复杂，统计预测充其量也不过是对未来现象预测平均水平的估计，要求统计预测准确无误与未来实际完全相符合也是不现实的。我们的目的在于尽可能利用现有的信息，分析原因，改进工作，努力减少或控制误差。

产生统计预测误差的原因主要有：

⑴ 预测方法选择不当，建立的预测模型不符合客观实际，不反映现象的基本数量关系；

⑵ 历史资料不充分，不完整，存在虚假因素；

⑶ 外部条件发生根本的变化；

⑷ 预测人员的认识水平和知识技能的局限。

对于统计预测误差要有一个正确的态度。虽然说统计预测不可避免地存在着误差，但这并不能抹杀预测的重要性，因为统计资料反映客观事物的准确程度是相对的，对统计预测准确性的要求也是相对的。如果在预测过程中，有真实可靠的实际资料为依据，有一定的经济理论为指导，有科学的分析方法和实事求是的精神，预测的结果是可以做到比较接近或基本符合实际情况的。

14.5.2 预测误差的表现形式

预测误差 e_i 一般表现为实际值与预测值之差：

$$e_i = y_i - \hat{y}_i$$

为了反映误差的一般水平,可以有下面几种指标：

1. 平均误差

$$\bar{e} = \frac{\sum e_i}{n} = \frac{\sum (y_i - \hat{y}_i)}{n}$$

优良估计的预测误差的数学期望等于 0,如果实际计算结果不为 0,就表示预测存在偏误,其数值越大就表示平均偏误水平越高。

2. 平均绝对偏差 MAD

$$MAD = \frac{\sum |y_i - \hat{y}_i|}{n}$$

平均误差存在正误差和负误差相互抵消的缺陷,平均绝对偏差则是对每项误差先取绝对值,因而它反映了绝对误差的总平均水平。

3. 均方误差 MSE

$$MSE = \frac{\sum e_i^2}{n} = \frac{\sum (y_i - \hat{y}_i)^2}{n}$$

这就是估计标准方差,它反映实际观察值和估计理论值方差的平均水平,是最常用的预测误差公式。

4. 均方根误差 $RMSE$

$$RMSE = \sqrt{\frac{\sum e_i^2}{n}} = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n}}$$

它是均方误差的平方根,其作用与方差相同,也是一种常用的预测误差公式。

5. 偏倚 $BIAS$

$$BIAS = \frac{\sum (y_i - \hat{y}_i)^3 / n}{RMSE^3}$$

它反映预测的不对称程度,它是以估计标准方差为标尺单位。

我们可以就前面的例子,进行各类误差的计算 (表 14.4 及表 14.5)：

表 14.4

年份	产量	最小平方法			
	y_i	$\hat{y}_i$	e_i	e_i^2	e_i^3
1978	14.9	11.98	2.92	8.53	22.43
1979	18.6	15.46	3.14	9.86	30.96
1980	22.2	18.94	3.26	10.63	34.65
1981	17.6	22.42	−4.82	23.23	−111.98
1982	19.6	25.90	−6.30	39.69	−250.05
1983	24.0	29.38	−5.38	28.94	−155.72
1984	31.6	32.86	−1.26	1.59	−4.02
1985	43.7	36.34	7.36	54.17	398.69
1986	37.0	39.82	−2.82	7.95	−22.43
1987	47.2	43.30	3.90	15.21	59.32
合计	276.4	276.40	0	199.80	1.85

表 14.5

	e	MAE	MSE	$BIAS$
最小平方法	0	4.12	19.98	

14.5.3 预测误差的分析

确切地说，预测误差是指未来实际值与预测值之差，即 $y_i+n-\hat{y}_i+n$。因而模型中各期观察值与相应的理论值的误差并不等同于预测误差。但是，分析模型中的实际误差 $e_i=y_i-\hat{y}_i$，还是有重要意义的。通过实际的误差分析，可以检验预测模型的合理性；实际误差水平为未来预测误差提供重要的信息；对实际误差的跟踪分析提供改进预测方法的依据。

我们可以从不同方面进行实际误差分析。

1．误差的方差分析

均方误差（MSE）反映实际误差的总水平，这种误差可能来自两个方面。一方面各期的平均误差水平 $e=\dfrac{\sum e_i}{n}$，平均误差水平低，一般表示预测精度高，因而均方误差就小；但另一方面，平均误差水平并不能保证各项误差对平均误差的分散程度，如果各项误差的离差很大，就说明预测模型的稳定性差，用于预测可能产生很大偏差，在这种情况下均方差也就增大。因此有必要计算误差的方差指标 σ_e^2 来反映误差

变异情况：

$$\sigma_e^2 = \frac{\sum (e_i - e)^2}{n}$$

很容易证明均方误差 MSE 是平均误差平方与误差方差之和，即：

$$MSE = \frac{\sum e_i^2}{n} = e^2 + \sigma_e^2$$

事实上，

$$\text{右边} = \left[\frac{\sum (y_i - \hat{y}_i)}{n}\right]^2 + \left(\frac{1}{n}\right) \cdot \sum \left[(y_i - \hat{y}_i) - \frac{\sum (y_i - \hat{y}_i)}{n}\right]$$

$$= \left[\frac{\sum (y_i - \hat{y}_i)}{n}\right]^2 + \frac{1}{n} \cdot \sum \left[(y_i - \hat{y}_i)^2 - 2(y_i - \hat{y}_i)\frac{\sum (y_i - \hat{y}_i)}{n} + \left(\frac{\sum (y_i - \hat{y}_i)}{n}\right)^2\right]$$

$$= \left[\frac{\sum (y_i - \hat{y}_i)}{n}\right]^2 + \frac{1}{n}\Big) \cdot \sum (y_i - \hat{y}_i)^2 - 2\left[\frac{\sum (y_i - \hat{y}_i)}{n}\right]^2 + \left[\frac{\sum (y_i - \hat{y}_i)}{n}\right]^2$$

$$= \frac{1}{n}\sum (y_i - \hat{y}_i)^2$$

$$= MSE = \text{左边。}$$

当平均误差 $e = 0$ 时，MSE 即由 σ_e^2 来决定。在比较不同预测时，优良的方法应该是在保证平均误差小的基础上误差方差也小的方法。如果一种预测方法虽然误差方差小但平均误差却很大，这说明预测模型可能存在系统性的偏误。

现在，我们继续以我国 1978—1987 年汽车产量为例，计算最小平方方法的均方误差，并分解为平均误差平方和误差方差两项指标，以表明误差的来源，见表 14.6。

表 14.6

	均方误差		平均误差平方		误差方差	
	MSE	%	e^2	%	σ_e^2	%
最小平方法	19.975	100	0	0	19.975	100

从表看出，最小平方法的误差最小，它全部来自误差方差。指数平滑法的均方误差有 2.08% 来自平均误差。误差方差所占比重大，这说明由于序列的变动度大引起误差的变动度也大，产量的变动不但受趋势性因素影响，其他非趋势性因素也影响干扰着序列的变动。

2. 误差的模拟分析

以上预测误差分析都是就预测模型中的观察值和相应理论值的离差及其变动情况而言。但是，离差大也未必都属于不利现象。因为预测模型的目的在于提供未来时间的预测值，以使它尽可能接近未来的实际值，为人们提供目前行动决策的依据。为了更有效地达到预测的目的，对预测模型进行种种权宜的处理，例如对近期的资料加重权数，调整平滑常数值等等，这时模型理论值与观察值离差虽然加大了，然而对于预测却是更切实际的。因此，对于预测误差还需要进一步进行模拟分析，即先将近期的资料删去，利用其余的资料设置模型，然后由模型计算出近期的预测值以便与近期观察的实际值进行对比，检查预测的结果是否符合实际情况。这种分析方法的好处在于近期的资料在模型之外，未进入模型，它只作为检查模型之用。不过这也是一种权宜之计，避去自己检查自己之嫌。近期的个别资料可能受非基本的随机因素影响很大，检查结果未必就能说明问题；再者近期也不等同于将来，情况还会不断变化。这种分析方法的结论只能提供参考。

14.6 微机操作 —— 调用 YAJ 中，TSFC 决策支持系统

14.6.1 观察 TSFC 决策支持系统 .

这个程序能让你实施时间序列预测，使用的是下列通俗的方法：简单平均，具有常数过程的移动平均，具有线性趋势过程的移动平均，单、复指数修匀，适应指数修匀，以及线性回归。TSFC 所要求的时间序列数据，可以从键盘上输入，也可从磁盘的文件上录入。为了预测将来的趋势，允许 100 个历史数据。然而，不超过 20 个的观测值，能够显示在荧屏上，或在选择图形表示结果时，被打印出来。TSFC 可预测未来 15 个周期。这些均能被显示或打印。

TSFC 允许同时使用三个预测方法，并且在荧屏上显示图形或结果。

14.6.2 问题输入

以表 14.2 给出的数据输入：

编号	历史数据	编号	历史数据	编号	历史数据	编号	历史数据	编号	历史数据
1	14.9	3	22.2	5	19.6	7	31.6	937	
2	18.6	4	17.6	6	24	8	43.7	10	47.2

14.6.3 问题解答选择及结果显示(可同时选择三项进行解答)

(1)就移动 3 个周期、预测 10 个周期进行移动平均预测

结果显示:

① **简单平均(AV)**:$MAD = 9.569155$

$MSE = 152.0487$

$BIAS = -9.35434$

② **具有常数过程的移动平均(MA)**:$MAD = 6.971428$

$MSE = 85.97588$

$BIAS = -6.695238$

③ **具有线性趋势的移动平均(LT)**:$MAD = 6.828572$

$MSE = 64.51588$

$BIAS = 4.761968E-02$

综合结果:

观察值		预测结果			
			AV	MA	LT
1	14.9000	11	27.6400	42.6333	46.1333
2	18.6000	12	27.6400	42.6333	47.8833
3	22.2000	13	27.6400	42.6333	49.6333
4	17.6000	14	27.6400	42.6333	51.3833
5	19.6000	15	27.6400	42.6333	53.1333
6	24.0000	16	27.6400	42.6333	54.8833
7	31.6000	17	27.6400	42.6333	56.6333
8	43.7000	18	27.6400	42.6333	58.3833
9	37.0000	19	27.6400	42.6333	60.1333
10	47.2000	20	27.0000	42.6333	61.8833

(2)就修匀常数取 0.7、预测 10 个周期,进行指数修匀预测

结果显示:

① **简单指数修匀(SE)**:$MAD = 5$

$MSE = 51.72529$

$BIAS = -4.673538$

② **就取线性趋势 0.3,作指数修匀(ET)**:$MAD = 5.631692$

$MSE = 46.41265$

$BIAS = -4.144278$

③ **适应指数修匀（*AE*）：** $MAD = 5.986079$

$MSE = 52.13937$

$BIAS = -4.546244$

综合结果：

观察值		预测结果			
			SE	*ET*	*AE*
1	14.9000	11	44.3433	50.3273	44.4313
2	18.6000	12	44.3433	54.6934	44.4313
3	22.2000	13	44.3433	59.0595	44.4313
4	17.6000	14	44.3433	63.4256	44.4313
5	19.6000	15	44.3433	67.7919	44.4313
6	24.0000	16	44.3433	72.1579	44.4313
7	31.6000	17	44.3433	76.5240	44.4313
8	43.7000	18	44.3433	80.8901	44.4313
9	37.0000	19	44.3433	85.2562	44.4313
10	47.2000	20	44.3433	89.6223	44.4313

(3) 就 10 个周期进行线性回归（*LR*）预测

结果显示：

$MAD = 4.504646$　　$MSE = 31.10507$　　$BIAS = -3.153134$

斜率 $= 3.475153$　　**截距** $= 8.526657$ **（请对照前面解答结果）**

综合结果：

观察值		预测结果	
			LR
1	14.9000	11	46.7533
2	18.6000	12	50.2285
3	22.2000	13	53.7036
4	17.6000	14	57.1788
5	19.6000	15	60.6540
6	24.0000	16	64.1291
7	31.6000	17	67.6043
8	43.7000	18	71.0794
9	37.0000	19	74.5546
10	47.2000	20	78.0297

参考文献

[1] 胡富昌. 线性规划. 北京:中国人民大学出版社,1990
[2] 中国人民大学数学教研室. 运筹学通论. 北京:中国人民大学出版社,1987
[3] 刘龄德,夏国平. 系统工程基础. 北京:中央广播电视大学出版社,1985
[4] 魏国华,傅家良,周仲良. 实用运筹学. 上海:复旦大学出版社,1987
[5] 黄良文. 社会经济统计学原理. 北京:中国财政经济出版社,1989