

数学建模方法

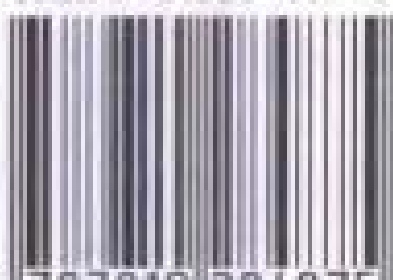
杨学桢 主编

河北大学出版社



责任编辑: 韩建民
马 力
封面设计: 赵 谦
责任印制: 李晓敏

ISBN 7-81028-487-8



9 787810 284875 >

ISBN 7-81028-487-8

O · 23 定价: 18.00元

see more please visit: <https://homeofpdf.com>

数学建模方法

主 编 杨学桢

副主编 王培光

张建辉

吴冬生



A0977623

河北大学出版社

责任编辑 韩建民 马 力

封面设计 赵 谦

责任印制 李晓敏

图书在版编目(CIP)数据

数学建模方法 / 杨学桢等著. — 保定: 河北大学出版社, 2000. 10

ISBN 7-81028-487-8

I. 数… II. 杨… III. 建立模型 IV. 022

中国版本图书馆 CIP 数据核字(2000)第 75479 号

出版: 河北大学出版社(保定合作路 1 号)

经销: 全国新华书店

印制: 河北省〇五印刷厂

规格: 1/32(850mm×1168mm)

印张: 13

字数: 326 千字

印数: 1001~3500 册

版次: 2000 年 10 月第 1 版

版次: 2001 年 6 月第 2 次印刷

定价: 18.00 元

前 言

当前,随着科学技术和计算机的迅速发展,数学向各个领域的广泛渗透已日趋明显,数学不仅在传统的物理学、电子学和工程技术等领域继续发挥着重要作用,而且在经济、人文、体育等社会科学领域也成为必不可少的解决问题的工具。这样,尽快提高学生的数学素质和应用数学知识解决实际问题的能力已是摆在数学教育中的一个重要课题。近年来数学建模(数学模型的建立与分析)课程的设立和全国大学生数学建模竞赛的开展对提高学生的数学素质和能力,培养应用型人才是十分有益的。

目前,数学建模教学及竞赛前的辅导、培训已在全国蓬勃开展,而满足辅导培训需要不同类型、不同特色的教材,尤其是参加竞赛的学生大部分是二、三年级的学生,他们来自不同的系(专业),其所学的数学知识又不能很好地适应应用型人才的需要,因此编写适应建模需要的教材,不仅是建模竞赛的需要,而且也丰富了数学的教学内容和数学应用的领域。鉴于这种情况,我们编写了这本《数学建模方法》。

本书具有以下几个特点:

1. 用较短篇幅介绍了数学建模竞赛中一些必要的数学知识及有关方法。
2. 根据每一章介绍的数学内容,列举了一些不同领域中的建模实例,使学生尽快地将有关数学知识和数学建模联系在一起。部分章节重点强调了如何分析问题和建立模型的方法。
3. 结合我们的体会,对建模中应注意的有关问题做了必要的说明。

全书共分十章,第一、二、三章由王培光执笔,第四、五章由吴冬生执笔,第六、七、八章由杨学桢执笔,第九、十章由张建辉执笔,最后由杨学桢、王培光对全书进行统稿。

由于我们的水平有限,书中难免有不妥、错误之处。敬请各界朋友及同仁们批评指正。

编者

1999.12

目 录

第一章 绪 论	(1)
§ 1.1 模型与数学模型	(1)
§ 1.2 数学模型的分类	(3)
§ 1.3 数学建模举例	(3)
§ 1.4 建立数学模型的方法与步骤	(9)
§ 1.5 建模能力的培养	(11)
习 题	(12)
第二章 微分方程方法建模	(13)
§ 2.1 预备知识	(14)
2.1.1 微分方程的基本概念	(14)
2.1.2 线性微分方程的基本定理	(16)
2.1.3 一阶微分方程	(17)
2.1.4 二阶常系数线性微分方程	(24)
2.1.5 微分方程的稳定性简介	(28)
§ 2.2 建模实例	(30)
2.2.1 小船行走路线问题	(30)
2.2.2 飞向月球问题	(31)
2.2.3 溶液混合问题	(34)
2.2.4 人口问题	(35)
2.2.5 军事战争问题	(41)
2.2.6 宣传运动的效果研究问题	(45)

2.2.7	关于诊断糖尿病的问题	(48)
2.2.8	家用电器销售问题	(52)
习 题	(54)
第三章	差分方程方法建模	(57)
§ 3.1	预备知识	(57)
3.1.1	差分的概念	(57)
3.1.2	差分方程的基本概念	(59)
3.1.3	线性差分方程的基本定理	(60)
3.1.4	一阶常系数线性差分方程	(61)
3.1.5	二阶常系数线性差分方程	(65)
3.1.6	差分方程的稳定性简介	(68)
3.1.7	离散量与连续量的对应关系	(69)
§ 3.2	建模实例	(71)
3.2.1	弦受集中荷载时的位移问题	(71)
3.2.2	生产和库存系统问题	(73)
3.2.3	银行复利问题	(73)
3.2.4	换热器系列的最优设计问题	(74)
3.2.5	市场价格的形成问题	(76)
习 题	(82)
第四章	概率方法建模	(84)
§ 4.1	预备知识	(84)
4.1.1	概率的定义及性质	(84)
4.1.2	随机变量及其分布	(89)
4.1.3	随机变量的数字特征	(94)
4.1.4	大数定律与中心极限定理	(98)
4.1.5	建模中常用的几个分布	(99)
§ 4.2	建模实例	(105)

4.2.1	粒子运动问题	(105)
4.2.2	随机徘徊问题	(107)
4.2.3	彩票问题	(109)
4.2.4	企鹅后代个数问题	(111)
4.2.5	花店老板的进货问题	(112)
4.2.6	病人排队时间问题	(115)
4.2.7	一个更换问题	(118)
4.2.8	建厂问题	(120)
习 题	(124)
第五章 统计方法建模		(127)
§ 5.1	预备知识	(127)
5.1.1	基本概念及统计学中的重要分布	(127)
5.1.2	参数的估计理论	(131)
5.1.3	参数及模型的假设检验	(137)
5.1.4	回归分析方法	(143)
5.1.5	单因子方差分析方法	(145)
§ 5.2	建模实例	(147)
5.2.1	钓鱼问题	(147)
5.2.2	新药药效问题	(148)
5.2.3	婴儿出生时刻问题	(151)
5.2.4	质量控制问题	(152)
5.2.5	运动成绩预测问题	(155)
5.2.6	灯丝材料问题	(158)
5.2.7	蠓虫识别问题	(161)
习 题	(164)
第六章 规划方法建模		(167)
§ 6.1	预备知识	(167)

6.1.1	线性规划	(167)
6.1.2	整数规划	(187)
6.1.3	非线性规划	(198)
§ 6.2	建模实例	(206)
6.2.1	资源最优利用问题	(206)
6.2.2	生产任务的分配问题	(209)
6.2.3	生产力布局问题	(213)
6.2.4	赛事举例	(222)
习 题	(227)
第七章	层次分析法建模	(235)
§ 7.1	预备知识	(235)
7.1.1	AHP 的基本方法和步骤	(235)
7.1.2	不完全判断信息下的排序问题	(241)
7.1.3	群组决策	(245)
§ 7.2	建模实例	(247)
7.2.1	过河的代价与效益分析	(248)
7.2.2	企业留成的分配问题	(255)
7.2.3	农民卖水果问题	(259)
7.2.4	最佳食品搭配问题	(263)
习 题	(267)
第八章	图论方法建模	(269)
§ 8.1	图的基本概念	(269)
§ 8.2	E 图和 H 图	(276)
8.2.1	E 图	(276)
8.2.2	H 图	(277)
8.2.3	建模实例	(278)
§ 8.3	树	(285)

8.3.1	最小树	(286)
8.3.2	有向图	(289)
8.3.3	有根树	(289)
8.3.4	建模实例	(291)
§ 8.4	最短路	(292)
8.4.1	图中某点到其他顶点的最短路	(293)
8.4.2	图中任意两点间的最短路	(297)
8.4.3	建模实例	(299)
§ 8.5	最大流	(302)
8.5.1	最大流与最小割集	(303)
8.5.2	Fold - Fulkerson 标号法	(308)
§ 8.6	匹配	(311)
8.6.1	最大匹配	(311)
8.6.2	建模实例	(314)
§ 8.7	边染色与顶点染色	(315)
8.7.1	基本知识	(315)
8.7.2	建模实例	(317)
§ 8.8	最小复盖、最小控制集、极大独立集	(318)
8.8.1	基本知识	(318)
8.8.2	建模实例	(323)
习 题		(325)
第九章	量纲分析法建模	(330)
§ 9.1	预备知识	(330)
9.1.1	量纲	(330)
9.1.2	单位制与物理定律	(332)
9.1.3	量纲相互独立的充要条件	(333)
9.1.4	量纲分析建模的一般方法	(335)
§ 9.2	建模实例	(339)

9.2.1	考虑空气阻力的单摆周期问题	(339)
9.2.2	航船的阻力问题	(341)
9.2.3	点热源的热扩散问题	(343)
9.2.4	量纲分析法建模的一般步骤	(345)
§ 9.3	方法总结	(346)
习 题	(348)
第十章	其他方法建模	(349)
§ 10.1	变分法建模.....	(349)
10.1.1	变分法的基本知识.....	(349)
10.1.2	建模实例.....	(358)
§ 10.2	对策方法建模.....	(371)
10.2.1	对策论的基本知识.....	(371)
10.2.2	建模实例.....	(379)
§ 10.3	动态规划方法建模.....	(388)
10.3.1	动态规划的基本知识.....	(388)
10.3.2	建模实例.....	(394)
习 题	(400)
参考文献	(404)

第一章 绪 论

随着科学技术的迅速发展,数学与其他学科之间的新的联系不断涌现,数学模型已越来越受到人们的关注,成为人们对各门科学进行量化分析的一个主要手段。例如,在技术领域,从卫星到核电站,从天气预报到家用电器,高技术中的高精度、高自动、高质量、高效率等特点,都是通过数学模型和数学方法并借助计算机的计算控制来实现的。在社会领域,对于当前人类面临的五大问题:(1)人口问题,(2)工业化资金问题,(3)粮食问题,(4)不可再生的资源问题,(5)环境污染问题。如何提出相应的对策予以解决,同样需要借助建立数学模型加以论证,为决策者提供理论依据。

本章主要介绍模型及数学模型的概念及大致分类,通过实例说明建立数学模型的一般步骤和方法。

§ 1.1 模型与数学模型

模型通常是指按比例、形态或其他特征制成的同实物相似的物体,或者是指用一定数学结构作为某种形式语言的解释。它具有以下三方面特点:

1. 它是实际事物的一部分属性的抽象或模仿,而不是全部属性的复制;

2. 它由与被分析的问题有关的要素构成;

3. 它体现了有关要素之间的特殊关系。

模型可以分成型象模型和抽象模型两类。

根据实物,设计或设想,按比例、形态或其他特征制成的看起

来和客观实体基本相似的模型叫做形象模型。例如,教学用建筑模型,人体模型,航海模型等。

借助于符号,图表等来描述客观事物的模型叫做抽象模型。例如,电路图,建工图纸等。

利用变量、等式和不等式以及数学运算等数学符号和语言规则来描述事物的特征及内在联系的模型叫做数学模型。

数学模型的特征是:

1. 它是某事物为实现一种特殊目的而作出的一个抽象化、简单化的数学结构表达式。

2. 它是数学上的抽象,在数值上可以作为公式应用,可以推广到与原物相近的一类问题。

3. 可以作为某事物的数学语言,可以译成算法语言,编写程序进入计算机。

数学模型应用十分广泛的原因是因为在自然界和社会事物中,一种事物或现象,与另一些事物或另一些现象之间,存在着数量(依附某种性质)关系。人们利用数学模型把这种数量(性质)关系刻画出来,用以研究它所反映的特性。例如,爱因斯坦的著名公式 $E = mc^2$, 就是一个反映两种形式的宇宙能之间数量关系的数学模型。它表示包含在物质内部的能量 E , 等于其质量 m 乘以作为另一种形式宇宙能的放射现象所具有的极限均匀速度 c 的平方。

数学模型除了在物理、化学、天文等自然科学领域的广泛应用外,目前也正在越来越多的应用于经济学、管理学、人口学等社会科学领域。例如,马克思的再生产公式就是一个描述社会再生产的经济数学模型:

$I(v + m) = IIc$ 说明在简单生产条件下,两大部类之间所必须保持的一种数量关系。

$I(v + m) > IIc$ 或写成 $I(v + m) = II(c + \Delta c)$ 说明在

扩大再生产的条件下,两大部类间所必须保持的一种数量关系。

§ 1.2 数学模型的分类

数学模型可以按照不同的方式分类,下面简介几种分类方法:

1. 按变量性质分类:根据变量是确定的还是随机的可分为确定性模型和随机性模型。

根据变量是连续的还是离散的可分为连续模型和离散模型。

2. 按时间关系分类:有静态模型和动态模型。

3. 按研究方法分类:有初等模型,微分方程模型,优化模型,运筹模型,概率模型等。

4. 按研究对象所在领域分类:有人口模型,经济模型,社会模型,生态模型,交通模型等。

5. 按研究对象的内部结构和性能的了解程度分类:有白箱模型,灰箱模型和黑箱模型。

模型的分类问题并没有什么重要意义,因为问题本身以及解决问题的方法,依照人们的各种不同分法,它既可属于这个类型又可属于那个类型,重要的是有效地解决问题。本课程主要讨论数学模型的建立及求解问题。

§ 1.3 数学建模举例

欧几里德几何,牛顿的万有引力定律都是极为成功的数学建模的典型范例,我们还可以举出成百上千成功的数学建模实例,本节给出两个建立数学模型的实际例子,说明从现实对象到数学模型的过程,以便对数学模型全貌的了解有个初步认识。

例 1 贷款买房问题。

房地产已经进入市场,随着住房改革的深入,人人都要考虑买

房。然而,多数人不可能有这么多钱能一次付清,必须贷款买房,从而也就成为我们家庭面临的许多经济决策问题之一。目前市场上不断有各种售房广告出现,人们看到这样的广告之后,除了能知道自己是否有能力去买这样的房子这样的反应之外,也许更多的是提出问题:房子有多大,如果一次付清要多少钱,银行贷款的利率多大等等。为分析这些问题,我们不妨把问题具体化,以便于建立模型分析、解决问题:

小李夫妇为买房要向银行借款6万元,月利率是0.01,贷款期为25年。小李夫妇要知道每个月要偿还多少钱(设为常数),才能决定自己有无能力来买房。这里假设小李夫妇每月能有900元的节余。

解:设 A_0 ($= 6$ 万元) 为向银行的贷款额, $R = (0.01)$ 为月利率, N 表示第 N 个月(时间变量), A_N 表示第 N 个月尚欠银行的款, x 表示 25 年 ($= 300$ 月) 还清本息每月应还的钱。

这里要求的是 x , 因而将 x 看成因变量, A_0, R 看成参数, N 看成自变量。本问题的数学模型可这样建立:

A_0 (一开始借多少钱),

$A_1 = A_0(1 + R) - x$ (一个月后欠银行的钱数, 即本金加利息减去每月要还的钱数)

显然第 N 个月尚欠银行的钱数为:

$$A_N = A_{N-1}(1 + R) - x \quad (1.3.1)$$

其中 A_0 已知。这就是本问题的数学模型。(数学上, (1.3.1) 称为(线性) 差分方程的初值问题)。

将 $A_{N-1}, A_{N-2}, \dots, A_1$ 的表达式依次代入(1.3.1), 即得

$$\begin{aligned} A_N &= A_0(1 + R)^N - x[(1 + R)^{N-1} + (1 + R)^{N-2} + \dots \\ &\quad + (1 + R) + 1] \\ &= A_0(1 + R)^N - \frac{x}{R}[(1 + R)^N - 1] \end{aligned} \quad (1.3.2)$$

当 $N = 300$ 时, $A_{300} = 0$ 就表示还清贷款, 由此得到

$$0 = 60000 \times (1.01)^{300} - \frac{x}{0.01} [(1.01)^{300} - 1]$$

从而解得: $x \approx 632$ 元。这就是说每月要拿出 632 元交付银行。所以, 小李夫妇还是有能力买房的, 但他们的节余减少了, 应急能力降低了。

如果在此时, 某借贷公司针对上述情况出了一个广告: “本公司能帮助您提前三年还清借款, 只要: ① 每半个月向公司还一次钱, 钱数为 $\frac{632}{2}$ 元; ② 由于文书工作增多, 要求您预付三个月的钱, 即预付 1896 元”。提前三年还清。这当然是吸引人的, 条件似乎也不苛刻, 但人们仍有疑惑: 该借贷公司是在从事慈善事业呢, 还是在赚钱? 为此我们对该公司的两个“只要”(孤立地) 进行分析, 就可作出决断。

分析(1) 此时主要是还款周期变了, 从一个月变为半个月, 因而可设 $R = \frac{0.01}{2} = 0.005$, $x = 316$, $A_0 = 60000$, 这时要求的是使 $A_N = 0$ 的 N (表示半个月的数目)。由(1.3.2) 可知

$$N = \frac{\ln(\frac{x}{x - A_0 R})}{\ln(1 + R)} \quad (1.3.3)$$

从而求得 $N \approx 598$ (半个月) $= 299$ 月 ≈ 24.92 年, 即最多只能提前一个月还清。如果只有这一条该借贷公司真的成了慈善机构了。

分析(2) 预付 1896 元表示借贷人只借了 $A_0 = 60000 - 1896 = 58104$ 元, 而 $R = 0.005$, $x = 316$, 要求的是使 $A_N = 0$ 的 N 。由(1.3.3) 求得 $N \approx 505$ (半月) $= 252.5$ 月 $= 21.04$ 年, 即提前四年就还清了, 相当于该借贷公司至少可收入 $632 \times 12 = 7584$ 元。

由上述分析, 小李夫妇明白了, 可以一开始就少借一点钱, 同时也可看到算计(数学) 在家庭经济决策中的重要作用。

由公式(1.3.2),我们也可以在假定了 R 后大致推算出在本例一开始的广告中,如果一次付清应付多少款。

本例只是一个简化的例子,实际的借贷要复杂,因而更证明数学建模的分析的重要性。

例2 放射性废物的处理问题。

环境污染是人类面临的一大公害,放射性污染对人类生命安全和地球上的生物存在严重的威胁,所以特别为人们所关注。和平利用原子能可为人类造福,但是核废料处置不好,又将对人类是一大危害。核废料如何处置为好,必须进行科学论证。

过去一段时间,美国原子能委员会为了处理浓缩的放射性废物,他们把废物装入密封的圆桶,然后扔到水深为 300 英尺的海里。一些生态学家和科学家为此表示担心,圆桶是否会在运输过程中破裂而造成放射性污染?美国原子能委员会向他们保证:“圆桶绝不会破裂”。并作了许多种试验证明他们的说法是正确的。然而又有几位工程师提出了如下的问题:圆桶扔到海洋中时是否会因与海底碰撞而发生破裂?美国原子能委员会仍保证说:“决不会”。

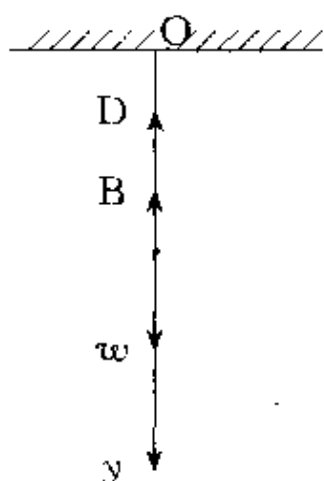


图 1.3.1

这几位工程师进行了大量的实验以后发现:当圆桶的速度超过 40 英尺/秒时,就会因碰撞而破裂。为此我们需计算圆桶同海底碰撞时的速度,是否会超过 40 英尺/秒?

解:如图 1.3.1 选取坐标系, w 表示圆桶重量,使圆桶向下, $w = 527.436$ 磅, $w = mg$, m 表示质量, g 表示重力加速度, $g = 32.2$ 英尺/秒²。

B 表示水作用在圆桶上的浮力,推圆桶向上。原子能委员会使用的是 55 加仑的圆桶,体积为 $7.35(\text{英尺})^3$, $1(\text{英尺})^3$ 海水重量为 63.99 磅。所以 $B = 63.99 \times 7.35 = 470.327$ 磅。

D 表示水作用在圆桶上的阻力,它阻碍圆桶在水中运动,与物体运动方向相反,通常与速度 v 成正比。 $D = cv, c > 0$ 为常数。通过大量实验得出如下结论:圆桶方位对于阻力影响甚小,可以忽略不计。且 $D = 0.08v$ 磅·秒/英尺。则作用在圆桶上的力为

$$F = w - B - cv$$

由牛顿第二定律: $F = ma$, 而 $a = \frac{d^2y}{dt^2}$, 可得

$$\frac{d^2y}{dt^2} = \frac{1}{m}(w - B - cv) = \frac{g}{w}(w - B - cv) \quad (1.3.4)$$

(1.3.4) 是一个二阶常微分方程。

将 $\frac{d^2y}{dt^2} = \frac{dv}{dt}$ 代入(1.3.4), 得

$$\begin{cases} \frac{dv}{dt} + \frac{cg}{w}v = \frac{g}{w}(w - B) \\ v(0) = 0 \end{cases} \quad (1.3.5)$$

解上述一阶线性非齐次方程, 得其解为

$$v(t) = \frac{w - B}{c} [1 - e^{-\frac{cg}{w}t}] \quad (1.3.6)$$

由(1.3.6) 式可知, 圆桶的速度为时间 t 的函数, 要确定圆桶同海底的碰撞速度, 就必须算出桶碰到海底所需的时间 t 。但作为 y 的显函数求出 t 是困难的, 所以不能用(1.3.5) 来求圆桶同海底的碰撞速度。但由(1.3.5) 我们可以得到圆桶的极限速度 v_T , 当 $t \rightarrow \infty$ 时, $v_T = \frac{w - B}{c}$ 。

显然有 $v(t) \leq v_T$, 如果极限速度小于 40 英尺/秒, 那么圆桶就不可能因同海底碰撞而破裂。然而

$$v_T = \frac{w - B}{c} = 713.86 \text{ 英尺/秒}$$

这个数值太大了, 还不能断定 $v(t)$ 究竟是否能超过 40 英尺/秒。

下而将速度 v 作为位置 y 的函数 $v(y)$ 来考虑。

由 $v(t) = v[y(t)]$, 利用复合函数微分法

$$\frac{dv}{dt} = \frac{dv}{dy} \cdot \frac{dy}{dt} = v \frac{dv}{dy}$$

代入(1.3.4)中,得

$$\begin{cases} \frac{w}{g} v \frac{dv}{dy} = w - B - cv \\ v(0) = 0 \end{cases} \quad (1.3.7)$$

对于(1.3.7),利用微分方程数值解法,借助于计算机很容易求出 $v(300)$ 。

下面我们利用一个简便方法得到 $v(300)$ 的一个很好的近似值。

在(1.3.7)中令 $c = 0$ (即不考虑水的阻力),并用 u 代替 v 以示区别,得

$$\begin{aligned} \frac{w}{g} u \frac{du}{dy} &= w - B \\ u(0) &= 0 \end{aligned} \quad (1.3.8)$$

积分(1.3.8),得

$$\frac{w}{g} \frac{u^2}{2} = (w - B)y \quad \text{或} \quad u = \sqrt{\frac{2g}{w}(w - B)y}$$

由此可得

$$u(300) = \sqrt{\frac{2 \times 32.2 \times 57.109 \times 300}{527.436}} \approx 45.7 \text{ (英尺/秒)}$$

$u(300)$ 就是 $v(300)$ 的一个很好的近似值,其理由是:

1. 当不存在与运动方向相反的阻力时,圆桶的速度总会大一些,因此 $v(300) < u(300)$ 。

2. 当 y 增加时,速度 v 增加,所以对于 $y \leq 300$, 有 $v(y) \leq v(300)$ 。由此可以得出水作用在圆桶上的阻力 D 总是小于 $0.08 \times u(300) \approx 3.7$ 磅。

然而使圆桶向下的合力 $w - B \approx 57.1$ 磅,比 D 大得多,因而

可忽略 D 。所以可以认为 $u(y)$ 是 $v(y)$ 的一个很好的近似值。实际上,用数值解法算出的 $v(300) = 45.1$ 英尺 / 秒,与 $u(300) = 45.7$ 英尺 / 秒是比较接近的。

综上所述得出结论,圆桶能够因与海底碰撞而破裂。工程师们的说法是正确的。

这一模型科学地论证了美国原子能委员会过去处理核废料的方法是错误的,从而改变了美国政府过去的错误作法。现在美国原子能委员会条例明确禁止把低浓度的放射性废物抛到海里,改为在一些废弃的煤矿中修建放置核废料的深井。这一模型为全世界其他国家处理核废料提供了经验教训。我国政府决定在甘肃、广西等地修建三个深井放置核废料,防止放射性污染。

§ 1.4 建立数学模型的方法与步骤

由于客观事物的复杂性,建立数学模型的方法也千变万化,它不仅是一种创造性的劳动,甚至可以说是一种艺术,不可设想仅用一些固定不变的条条框框就能迎刃而解。这里所说的步骤仅是一种大体上的规范,真正解决问题还得融会贯通,利用已有的模型,改造已有的模型,创造出更适用的新模型。

建立数学模型一般要经过模型准备、模型假设、模型设计、模型求解、模型分析、模型检验等几个步骤。

1. 模型准备 首先了解问题的实际背景,明确建模的目的和用途,并按要求收集必要的各种信息如现象、数据等,并弄清对象的特征,由此初步确定用哪一类模型。

2. 模型假设 一旦明确了建立数学模型的目的,也就基本上确定了模型所涉及的范围和主要的变量,从而可对问题进行必要的简化,用精确的数学语言作出假设,可以说这是建模的关键一步。一般地说,一个实际问题不经过简化假设就很难翻译成数学问

题,即使可能,也很难求解。不同的简化假设会得到不同的模型。假设作得不合理或过份简单,会导致模型失败或部分失败,于是应该修改和补充假设;假设作得过分详细,考虑的因素过多,会使模型过于复杂而无法进行下一步工作。所以,重要的是,在模型假设时,要首先抓住主要因素,忽略次要因素,尽量将问题均匀化、线性化。

3. 模型建立 根据已有假设,利用适当的数学工具刻画各变量之间的关系,建立相应的数学结构(公式、表格、图形等)。在建模时究竟采用什么数学工具需根据问题的特征、建模的目的要求及建模者的数学特长而定。但应遵循的一个原则是,在有效的基本要求下尽量采用简单的数学工具,以便得到的模型被更多的人了解和使用。

4. 模型求解 根据采用的数学工具,对模型求解,包括解方程、图解、逻辑推理、数值计算等等,要求建模者掌握相应的数学知识,尤其是计算机技术。

5. 模型分析 对模型求解的结果进行数学上的分析,例如根据问题的性质,分析各变量之间的依赖关系或稳定性态,或根据所得结果给出数学上的预测等。不论哪种情况还常常需要进行误差分析,模型对数据的稳定性或灵敏性分析等。

6. 模型检验 将模型分析的结果翻译回到实际对象中,用实际的现象、数据等检验模型的合理性和适用性,即验证模型的正确性。通常,一个较成功的模型不仅能解释已知现象,还应当能预言一些未知的现象,并能被实践所证明。应该说,模型检验对于建模的成败是非常重要的。当然,有些模型如核战争模型就不可以要求接受实际的检验了。

如果对模型检验的结果表明不符合或者部分不符合,此时就要进一步分析原因,对模型进行修改。一般讲,问题通常出在模型假设上,应该修改、补充假设,重新建模。有些模型要经过几次反复,不断完善,直到检验结果获得某种程度上的满意。

建模一般步骤如图 1.4.1 所示:

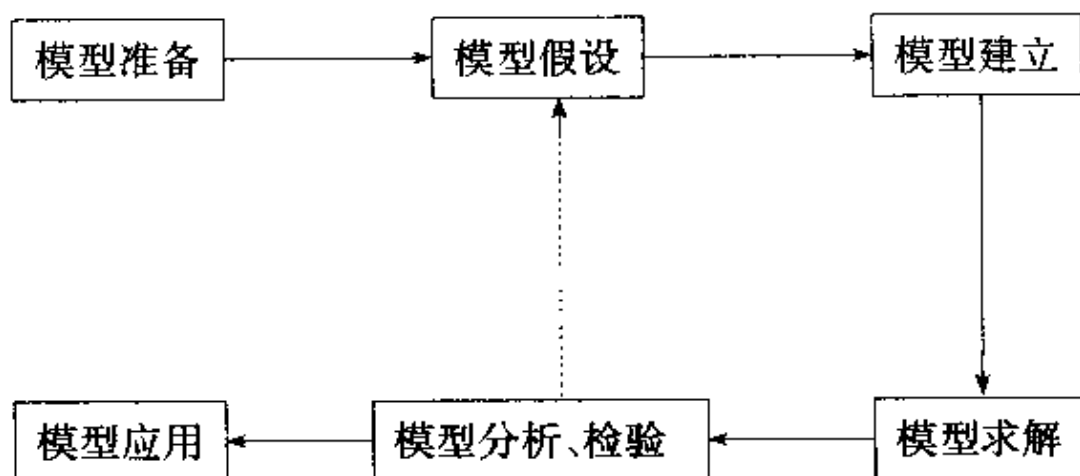


图 1.4.1 建模步骤示意图

§ 1.5 建模能力的培养

前面说过,数学建模可以看成一门艺术。艺术在某种意义下是无法归纳出几条准则或方法的。从建模步骤和原则来看,一个人的建模能力的培养包括以下几个方面。

1. 培养“翻译”的能力,即把经过一定抽象、简化的实际问题用数学的语言表达出来并给出它们的关系特征,形成数学模型,对应用数学的方法进行或计算得到的结果,能用“常人”能懂的语言“翻译”(表达)出来。

2. 应用已学到的数学方法和思想进行综合应用和分析,并能理解合理的抽象和简化,特别是对问题进行数学分析的重要性。

3. 发展联想能力。因为对于不少完全不同的实际问题,抽象出主要数学特征,在一定的简化层次下,它们的数学模型是相同或相似的,这也体现出数学的应用广泛性。所以我们应培养广泛的兴趣,多思考,勤奋踏实工作,通过熟能生巧而逐步达到触类旁通的境界。

4. 逐渐发展形成一种洞察能力。即在充分占有资料的基础上,

经过初步分析能迅速抓住主要矛盾,舍弃次要因素,对可以用哪些方法解决面临的问题,以及不同方法的优劣作出判断。

5. 熟练使用技术手段(目前主要是计算机及相应的各种软件包)。

由此可见,为了培养建模能力,一要大量阅读、思考别人做过的模型,二要接触实际问题,亲自动手,认真做几个实际题目。

应当指出的是,数学建模过程是一种创造性思维过程,除了想象、洞察、判断这些属于形象思维、逻辑思维范畴的能力之外,直觉和灵感往往也起着不可忽视的作用。当由于各种限制利用已有知识难以对研究的问题作出有效的推理和判断时,凭借相似、类比、猜测、外推等思维方式及不完整、不连续、不严密的,带启发性的直觉和灵感,去“战略性”地认识对象,是人类创造性思维的特点之一。历史上不乏在科学家的直觉和灵感的火花中诞生的假说、论证和定律。当然,直觉和灵感不是凭空产生的,它要求人们具有丰富的背景知识,对问题进行反复思考和艰苦探索,对各种思维方法运用娴熟。特别是由于现代科技的分工越来越细,各门学科互相渗透,不同专业成员之间的相互讨论和思想交锋,是激发直觉和灵感的重要因素。目前跨行业专门人才间的协作和联合越来越受到重视。

习 题

1. 什么是模型?试叙述模型的分类。
2. 试叙述数学模型的定义、特征。
3. 试叙述数学模型的分类。
4. 举实例说明建立数学模型的必要性。包括实际问题的背景,建模目的,大体上需要什么样的模型以及怎样应用这种模型等。
5. 简要叙述建立数学模型的一般步骤和原则。

第二章 微分方程方法建模

微分方程建模是数学建模的重要方法之一,因为在自然科学以及在工程、经济、军事、社会等学科中,许多实际问题的数学描述将导致求解微分方程的定解问题。把各种实际问题化成微分方程的定解问题,建立微分方程模型,大体上可以按以下几步:

1. 根据实际要求确定要研究的量(自变量、未知函数、必要的参数等)并确定坐标系。
2. 找出这些量所满足的动态特征和基本规律。
3. 运用这些规律列出方程和定解条件,从而建立起微分方程模型。

建立微分方程模型,其方法可归纳为:

1. 根据规律列方程。利用数学、力学、物理、化学等学科中的定理或许多经过实践或实验检验的规律和定律,如牛顿第二定律、放射性物质的放射性规律等等建立问题的微分方程模型。

2. 微元分析法。自然界中有许多现象所满足的规律是通过变量的微元之间的关系式来表达的。对于这类问题,我们不能直接列出自变量和未知函数及其变化率之间的关系式,而是通过微元分析法,利用已知规律建立变量的微元之间的关系式,再通过取极限的方法得到微分方程模型。

3. 模拟近似法。在生物、经济等学科中,许多现象的规律并不很清楚,而且相当复杂。常常需要根据实际资料或大量的实验数据,提出各种假设,根据假设,模拟实际现象所满足的规律,然后利用适当的数学方法给出微分方程模型。

本章将结合一些典型的例子讨论几个在不同领域中微分方程

模型的建模方法。

§ 2.1 预备知识

2.1.1 微分方程的基本概念

定义 2.1.1 含有未知函数的导数(或微分)的函数方程称为微分方程。

如果未知函数是一个变量的函数,由这样的函数及其导数构成的函数方程称为常微分方程。

例如 (1) $\frac{dy}{dx} = -\frac{x}{y}$

(2) $\frac{d^2x}{dt^2} + b\frac{dx}{dt} + cx = 0$

(3) $a_0(x)y^{(n)} + a_1(x)y^{(n-1)} + \cdots + a_n(x)y = f(x)$

本章只讨论常微分方程问题。简称为微分方程,甚至简称为方程。

定义 2.1.2 在微分方程中,未知函数最高阶导数的阶数,称为微分方程的阶。

例如(1)为一阶常微分方程;(2)为二阶常微分方程,(3)为 n 阶常微分方程。

定义 2.1.3 一个函数代入微分方程中,使得它成为关于自变量的恒等式,称此函数为微分方程的解。

由于微分方程的解是函数,将这个函数代入方程,是经过微分运算使等式成立的,因此微分方程的解有无穷多个。

定义 2.1.4 对于 n 阶微分方程,含有 n 个(相互独立的)任意常数的解,称为微分方程的通解。

例 1 $\frac{d^2y}{dx^2} = a$ (a 为已知常数)。

解:因为函数 $y = \frac{1}{2}ax^2 + c_1x + c_2$ (c_1 和 c_2 为任意常数) 代入方程后,使两端恒等,且此解中含有两个任意常数,所以它是方程的通解。

为确定微分方程一个特定的解,通常要给出这个解所必须满足的条件,此条件在数学上称为定解条件。

定解条件一般分初始条件和边界条件两类。在此我们只讨论初始条件。

n 阶微分方程(3) 的初始条件是,

当 $x = x_0$ 时, $y = y_0, \frac{dy}{dx} = y_0^{(1)}, \dots, \frac{d^{n-1}y}{dx^{n-1}} = y_0^{(n-1)}$

定义 2.1.5 求微分方程满足定解条件的问题称为定解问题。

特别在初始条件下求解方程的问题称为初值问题。

定义 2.1.6 满足初始条件的解称为该微分方程的一个特解。

一般来讲,初始条件不同,对应的特解也不同。特解可通过初始条件的限制,从通解中确定任意常数而得到。

例 2 求解初值问题。

$$\begin{cases} y'' = a & (a \text{ 为已知常数}) \\ x = 0; y = 0, y' = 1 \end{cases}$$

解:由例 1 知,方程的通解是

$$y = \frac{1}{2}ax^2 + c_1x + c_2$$

对上述关于 x 求导,得

$$y' = ax + c_1$$

将初始条件代入 y 及 y' 的表示式中,得到 $c_1 = 1, c_2 = 0$,初值问题的解为

$$y = \frac{1}{2}ax^2 + x$$

2.1.2 线性微分方程的基本定理

在力学,经济学等领域中用得最多的,最简单的一类方程是线性微分方程。从数学观点上看,这也是最容易处理的一类微分方程。下面给出线性微分方程的几个基本定理。

n 阶线性微分方程的一般形式为

$$a_0(t)x^{(n)} + a_1(t)x^{(n-1)} + \cdots + a_{n-1}(t)x' + a_n(t)x = f(t) \quad (2.1.1)$$

其中 $x^{(n)}, x^{(n-1)}, \cdots, x'$ 分别表示 x 对 t 的 n 阶, $n-1$ 阶, \cdots , 1 阶导数, $a_0(t), a_1(t), \cdots, a_n(t), f(t)$ 为 t 的已知函数, 且 $a_0(t) \neq 0$ 。

若 $f(t) \neq 0$, 则称方程(2.1.1) 是非齐次方程; $f(t) \equiv 0$, 则称方程是齐次方程; (2.1.1) 对应的齐次方程是

$$a_0(t)x^{(n)} + a_1(t)x^{(n-1)} + \cdots + a_{n-1}(t)x' + a_n(t)x = 0 \quad (2.1.2)$$

对于方程(2.1.1), (2.1.2), 我们有如下定理。

定理 2.1.1 若 $x_1(t), x_2(t)$ 是方程(2.1.2) 的两个解, 则 $c_1x_1(t) + c_2x_2(t)$ 也是方程(2.1.2) 的解。(c_1, c_2 为任意常数)

定理 2.1.2 若 $x_1(t)$ 是方程(2.1.2) 的解, $\bar{x}(t)$ 是方程(2.1.1) 的解, 则 $cx_1(t) + \bar{x}(t)$ 是方程(2.1.1) 的解。(c 为任意常数)

定义 2.1.7 对于函数 $x_1(t), x_2(t), \cdots, x_n(t)$, 若存在 n 个不全为零的常数 c_1, c_2, \cdots, c_n , 使恒等式

$$c_1x_1(t) + c_2x_2(t) + \cdots + c_nx_n(t) = 0$$

对一切 $t \in [a, b]$ 成立, 则称函数 $x_1(t), x_2(t), x_n(t)$ 在区间 $[a, b]$ 上线性相关; 否则称为线性无关。

定义 2.1.8 n 阶齐次方程(2.1.2) 的 n 个线性无关解, 称为方程(2.1.2) 的一个基本解组。

定理 2.1.3 (齐次方程解的结构定理) 若 $x_1(t), x_2(t), \cdots,$

$x_n(t)$ 是方程(2.1.2) 的一个基本解组, 则方程(2.1.2) 的通解可表示为

$$x(t) = c_1 x_1(t) + c_2 x_2(t) + \cdots + c_n x_n(t) \quad (2.1.3)$$

其中 c_1, c_2, \cdots, c_n 为 n 个(独立的) 任意常数。

定理 2.1.4 (非齐次方程解的结构定理) 若 $\bar{x}(t)$ 是方程(2.1.1) 的一个特解, $x_1(t), x_2(t), \cdots, x_n(t)$ 是对应的齐次方程的基本解组, 则方程(2.1.1) 的通解可表示为

$$x(t) = \bar{x}(t) + c_1 x_1(t) + c_2 x_2(t) + \cdots + c_n x_n(t) \quad (2.1.4)$$

定理 2.1.5 (非齐次方程的叠加原理) 若 $x_1(t), x_2(t)$ 分别是下列两个方程的解

$$\begin{aligned} a_0(t)x^{(n)} + a_1(t)x^{(n-1)} + \cdots \\ + a_{n-1}(t)x' + a_n(t)x = f_1(t) \end{aligned} \quad (2.1.5)$$

$$\begin{aligned} a_0(t)x^{(n)} + a_1(t)x^{(n-1)} + \cdots \\ + a_{n-1}(t)x' + a_n(t)x = f_2(t) \end{aligned} \quad (2.1.6)$$

则 $x(t) = x_1(t) + x_2(t)$ 是方程

$$\begin{aligned} a_0(t)x^{(n)} + a_1(t)x^{(n-1)} + \cdots \\ + a_{n-1}(t)x' + a_n(t)x = f_1(t) + f_2(t) \end{aligned} \quad (2.1.7)$$

的解。

以上定理的证明, 在任何一本常微分方程教科书中都能找到, 在此略去这些定理的证明。

如何求方程(2.1.2) 的线性无关解以及方程(2.1.1) 的特解的具体方法, 我们将在后面讨论中给出。

2.1.3 一阶微分方程

一阶微分方程的一般形式为

$$F(x, y, y') = 0 \quad (2.1.8)$$

下面介绍几种可用初等方法求解的类型

1. 可分离变量的一阶微分方程

$$\text{形如 } f(x)dx = g(y)dy \quad (2.1.9)$$

的微分方程,称为变量分离的微分方程。

将(2.1.9)式两边同时积分,得

$$\int f(x)dx = \int g(y)dy + c \quad (2.1.10)$$

其中 c 是任意常数,(2.1.10) 就是方程(2.1.9) 的通解。

例3 求微分方程 $dp = kp(N-p)dt$ ($N, k > 0$ 为常数) 的解。此处假设 $0 < p < N$ 。

解:将该方程改写为

$$\frac{dp}{p(N-p)} = kdt$$

等式两边积分得到

$$\frac{1}{N} \ln \frac{p}{N-p} = kt + C$$

$$\text{或 } \frac{p}{N-p} = e^{N(kt+C)} = Ae^{at}$$

其中 $A = e^{NC}, a = Nk$

解上式,得

$$p = \frac{NAe^{at}}{Ae^{at} + 1} = \frac{N}{1 + Be^{-at}}$$

其中 $B = \frac{1}{A}$, 这个方程称为 Logistic 曲线方程。

2. 齐次微分方程

$$\text{形如 } \frac{dy}{dx} = f\left(\frac{y}{x}\right) \quad (2.1.11)$$

的微分方程,称为齐次微分方程。

对于齐次方程(2.1.11),作变量替换

$$v = \frac{y}{x} \quad (y = xv) \quad (2.1.12)$$

得 $\frac{dy}{dx} = x \frac{dv}{dx} + v$

将 $y, \frac{dy}{dx}$ 代入(2.1.11), 得可分离变量的微分方程

$$x \frac{dv}{dx} = f(v) - v$$

即 $\frac{dv}{f(v) - v} = \frac{dx}{x}$

它的通解为

$$\int \frac{dv}{f(v) - v} = \ln x - \ln c$$

或 $x = ce^{\int \frac{dv}{f(v) - v}}$ (2.1.13)

求出积分 $\int \frac{dv}{f(v) - v}$ 后, 将 $v = \frac{y}{x}$ 代入(2.1.13), 即可得到方程(2.1.11)的通解。

例 4 求解微分方程

$$\frac{dy}{dx} = \frac{y^2}{xy - x^2}$$

解: 将原方程改写为

$$\frac{dy}{dx} = (\frac{y}{x})^2 / (\frac{y}{x}) - 1$$

令 $v = \frac{y}{x}$, 得

$$x \frac{dv}{dx} = \frac{v}{v-1}$$

$$\frac{v-1}{v} dv = \frac{1}{x} dx$$

两边积分, 得

$$v = \ln(xv) + c_1$$

即 $xv = e^{v-c_1} = ce^v \quad (c = e^{-c_1})$

将 $v = \frac{y}{x}$ 代入上式, 得所给微分方程的通解为

$$y = ce^{\frac{y}{x}}$$

3. 一阶线性微分方程

$$\text{形如 } \frac{dy}{dx} + p(x)y = q(x) \quad (2.1.14)$$

的微分方程, 称为一阶线性微分方程。

方程(2.1.14) 相应的齐次方程为

$$\frac{dy}{dx} + p(x)y = 0 \quad (2.1.15)$$

关于(2.1.14) 的通解可由以下两步实现。

1) 一阶线性齐次微分方程的通解

将(2.1.15) 分离变量后得

$$\frac{dy}{y} = -p(x)dx$$

两边积分得

$$\ln y = -\int p(x)dx + \ln c_1 \quad (c_1 \neq 0)$$

$$\text{即 } y = ce^{-\int p(x)dx} \quad (c \text{ 为任意常数}) \quad (2.1.16)$$

(2.1.16) 即为齐次方程(2.1.15) 的通解。

2) 一阶线性非齐次微分方程的通解

方程(2.1.14) 的解可用“常数变易法”求得。即将其对应的齐次方程通解(2.1.16) 中的任意常数 c , 换成待定函数 $c(x)$, 设(2.1.14) 具有如下形式的解

$$y = c(x)e^{-\int p(x)dx} \quad (2.1.17)$$

对上式关于 x 求导, 得

$$y' = c'(x)e^{-\int p(x)dx} - c(x)p(x)e^{-\int p(x)dx}$$

将 y, y' 代入(2.1.14), 得

$$\begin{aligned} c'(x)e^{-\int p(x)dx} - c(x)p(x)e^{-\int p(x)dx} \\ + p(x)c(x)e^{-\int p(x)dx} = q(x) \end{aligned}$$

即 $c'(x) = q(x)e^{\int p(x)dx}$

积分上式得

$$c(x) = \int q(x)e^{\int p(x)dx} + c$$

其中 c 是任意常数,代入(2.1.17) 式得

$$y = e^{-\int p(x)dx} \left(\int q(x)e^{\int p(x)dx} + c \right) \quad (2.1.18)$$

不难验证,(2.1.18) 就是方程(2.1.14) 的通解。

例 5 解一阶线性微分方程。

$$y' - \frac{2}{x+1}y = (x+1)^3$$

解:(1) 由 $y' - \frac{2}{x+1}y = 0$ 有

$$\frac{dy}{y} = \frac{2dx}{x+1}$$

解上式得

$$y = c(x+1)^2$$

(2) 令原方程的通解为 $y = c(x)(x+1)^2$

则 $y' = c'(x)(x+1)^2 + 2c(x)(x+1)$

将 y, y' 代入原方程后得

$$c'(x) = x+1$$

积分上式得

$$c(x) = \frac{1}{2}(x+1)^2 + c \quad (c \text{ 为任意常数})$$

将 $c(x)$ 代入(2) 中,可求得原方程的通解为

$$y = \frac{1}{2}(x+1)^4 + c(x+1)^2$$

4. 一阶常系数线性微分方程

在一阶方程的求解中,实际问题经常遇到常系数方程的求解问题,其一般形式为

$$y' + ay = f(x) \quad (2.1.19)$$

其中 a 为已知常数; $f(x)$ 是 x 的已知函数。

(2.1.19) 对应的齐次微分方程为

$$y' + ay = 0 \quad (2.1.20)$$

由(三)的讨论知, (2.1.20) 的通解为

$$y = Ae^{-ax} \quad (A \text{ 为任意常数})$$

函数 $y(x) = Ae^{-ax}$ 的性质依赖于常数 a 和 A 。

关于非齐次微分方程的通解,除了常数变易法外,根据 $f(x)$ 的特殊情形,可用待定系数法确定。下面介绍几种情况。

情形 1° $f(x)$ 为常数

若 $f(x) = b$, 此时方程(2.1.19) 变为

$$y' + ay = b \quad (2.1.21)$$

设(2.1.21) 有特解 $\bar{y}(x) = \mu$, μ 为待定常数。代入(2.1.21) 得 $a\mu = b$

若 $a \neq 0$, 则特解 $\bar{y} = \mu = \frac{b}{a}$

若 $a = 0$, 改设特解 $\bar{y}(x) = \mu x$, 代入(2.1.21) 得 $\mu = b$, 特解 $\bar{y}(x) = bx$

情形 2° $f(x)$ 是指数函数

若 $f(x) = B\alpha^{\beta x}$, B, α, β 为已知常数, 且 $\alpha > 0, B \neq 0, \beta \neq 0$, 则方程(2.1.19) 变为

$$y' + ay = B\alpha^{\beta x} \quad (2.1.22)$$

注意到 $B\alpha^{\beta x} = Be^{(\beta \ln \alpha)x}$, 上式可改写为

$$y' + ay = Be^{dx} \quad (d = \beta \ln \alpha) \quad (2.1.23)$$

设(2.1.23) 有特解 $\bar{y}(x) = \mu e^{dx}$, μ 为待定常数。代入(2.1.23), 得

$$[(d + a)\mu - B]e^{dx} = 0$$

若 $d + a \neq 0$, 则 $\mu = \frac{B}{d + a}$, (2.1.23) 有特解

$$\bar{y}(x) = \frac{B}{d+a}e^{dx}$$

若 $d+a=0$, 改设特解 $\bar{y}(x) = \mu x e^{dx}$ 代入(2.1.23), 得

$$(d+a)\mu x e^{dx} + (\mu - B)e^{dx} = 0$$

由 $e^{dx} \neq 0$, 得 $\mu - B = 0$, 即 $\mu = B$, 于是(2.1.23) 有特解

$$\bar{y}(x) = B x e^{dx}$$

情形 3° $f(x)$ 是 x 的 m 次多项式

为简便, 考虑 $f(x) = c_0 + c_1 x$ 的情况, 其中, c_0, c_1 为已知常数, 且 $c_1 \neq 0$ 。此时方程变为

$$y' + ay = c_0 + c_1 x \quad (2.1.24)$$

设特解为 $\bar{y}(x) = \alpha + \beta x$, α, β 为待定常数。代入(2.1.24) 得

$$\beta + a\alpha + a\beta x = c_0 + c_1 x$$

上式对任何 x 成立, 则有

$$\begin{cases} \beta + a\alpha = c_0 \\ a\beta = c_1 \end{cases}$$

若 $a \neq 0$, 则解得 $\alpha = \frac{ac_0 - c_1}{a^2}$, $\beta = \frac{c_1}{a}$, (2.1.24) 有特解

$$\bar{y}(x) = \frac{ac_0 - c_1}{a^2} + \frac{c_1}{a}x$$

若 $a = 0$, 改设特解 $\bar{y}(x) = x(\alpha + \beta x)$, 代入(2.1.24) 后同上理由可求得特解为

$$\bar{y}(x) = c_0 x + \frac{c_1}{2}x^2$$

情形 4° $f(x)$ 是正弦、余弦函数

若 $f(x) = B_1 \cos wx + B_2 \sin wx$, 其中 w, B_1, B_2 为已知常数, 且 $B_1^2 + B_2^2 \neq 0, w \neq 0$, 此时方程为

$$y' + ay = B_1 \cos wx + B_2 \sin wx \quad (2.1.25)$$

设特解 $\bar{y}(x) = \alpha \cos wx + \beta \sin wx$, α, β 为待定常数, 代入(2.1.25) 后得

$$(w\beta + a\alpha - B_1)\cos wx + (a\beta - w\alpha - B_2)\sin wx = 0$$

上式对任何 x 成立,则

$$\begin{cases} a\alpha + w\beta = B_1 \\ -w\alpha + a\beta = B_2 \end{cases}$$

由于 $D = \begin{vmatrix} a & w \\ -w & a \end{vmatrix} = a^2 + w^2 \neq 0$, 因此由上式可惟一确定出 α, β 。

$$\alpha = \frac{aB_1 - wB_2}{D}, \beta = \frac{aB_2 + wB_1}{D}$$

情形 5° $f(x)$ 是上述四类函数的线性组合

例如: $f(x) = A + \cos wx$, 此时, 方程变为

$$y' + ay = A + \cos wx \quad (2.1.26)$$

可用两种方法解方程(2.1.26)

解法 1: 设(2.1.26) 的特解为

$$\bar{y}(x) = \mu + \alpha \cos wx + \beta \sin wx$$

μ, α, β 为待定常数, 然后按前四种情形类似进行求解。

解法 2: 利用非齐次方程的叠加原理, 将(2.1.26) 拆成以下两个方程

$$\begin{cases} y' + ay = A \end{cases} \quad (2.1.27)$$

$$\begin{cases} y' + ay = \cos wx \end{cases} \quad (2.1.28)$$

然后由情形 1° 和情形 4° 的方法分别求出(2.1.27), (2.1.28) 的特解 $\bar{y}_1(x)$ 和 $\bar{y}_2(x)$, 则

$$\bar{y}(x) = \bar{y}_1(x) + \bar{y}_2(x)$$

就是方程(2.1.26) 的特解。

找到方程(2.1.19) 的特解 $\bar{y}(x)$ 之后, 由定理 2.1.4

$$y(x) = Ae^{-ax} + \bar{y}(x)$$

就是(2.1.19) 的通解。其中 A 为任意常数。

2.1.4 二阶常系数线性微分方程

二阶常系数线性微分方程的一般形式是

$$y'' + ay' + by = g(x) \quad (2.1.29)$$

其中 a, b 为已知常数, $g(x)$ 为 x 的已知函数。

对应于方程(2.1.29)的齐次方程是

$$y'' + ay' + by = 0 \quad (2.1.30)$$

下面对方程(2.1.29), (2.1.30)的解法分别进行讨论。

1. 二阶常系数齐次方程

仿照一阶方程, 设(2.1.30)有指数型非零特解

$$\bar{y}(x) = e^{\lambda x} \quad (\lambda \text{ 为待定常数}) \quad (2.1.31)$$

将 $y(x)$ 代入方程(2.1.30), 得

$$\lambda^2 e^{\lambda x} + a\lambda e^{\lambda x} + be^{\lambda x} = e^{\lambda x}(\lambda^2 + a\lambda + b) = 0$$

消去非零因子 $e^{\lambda x}$, 得

$$\lambda^2 + a\lambda + b = 0 \quad (2.1.32)$$

这就是 $\bar{y}(x)$ 为(2.1.30)的特解的充分必要条件。称(2.1.32)为方程(2.1.30)的特征方程, 其解称为特征根。

特征方程(2.1.32)的两个特征根是

$$\lambda_{1,2} = \frac{1}{2}[-a \pm (a^2 - 4b)^{\frac{1}{2}}] \quad (2.1.33)$$

方程(2.1.30)的解的性质, 依赖于特征根 $\lambda_{1,2}$ 。我们按(2.1.32)的判别式 $\Delta \equiv a^2 - 4b$ 的符号进行讨论。

情形 1° $\Delta > 0$

这时, λ_1 与 λ_2 为两个互异的实根。(2.1.30)的两个特解为

$$\bar{y}_1(x) = e^{\lambda_1 x}, \quad \bar{y}_2(x) = e^{\lambda_2 x}$$

且 $\bar{y}_1(x), \bar{y}_2(x)$ 线性无关, 由定理 2.1.3, (2.1.30)的通解

$$y(x) = c_1 e^{\lambda_1 x} + c_2 e^{\lambda_2 x} \quad (2.1.34)$$

其中 c_1, c_2 为任意常数。

情形 2° $\Delta \equiv 0$

这时有实重根 $\lambda_1 = \lambda_2 = \lambda^* = -\frac{a}{2}$, $\bar{y}_1(x) = e^{\frac{a}{2}x}$ 是(2.1.30)

的一个特解。不难验证, $xe^{-\frac{a}{2}x}$ 是(2.1.30) 的另一个特解。且特解 $e^{\lambda^* x}$ 与 $xe^{\lambda^* x}$ 是线性无关的。

由定理 2.1.3, (2.1.30) 的通解是

$$y(x) = (c_1 + c_2 x)e^{\lambda^* x} \quad (2.1.35)$$

其中, c_1, c_2 为任意常数。

情形 3° $\Delta < 0$

这里 λ_1 与 λ_2 为一对共轭复根。

$$\lambda_{1,2} = \alpha \pm i\beta$$

其中 $i = \sqrt{-1}, \alpha = -\frac{a}{2}, \beta = \frac{1}{2}(4b - a^2)^{\frac{1}{2}}$

(2.1.30) 的通解是

$$y(x) = c'_1 e^{(\alpha+i\beta)x} + c'_2 e^{(\alpha-i\beta)x}$$

其实值解为

$$y(x) = e^{\alpha x}(c_1 \cos \beta x + c_2 \sin \beta x) \quad (2.1.36)$$

$$\text{或 } y(x) = Ae^{\alpha x} \cos(\beta x - \theta) \quad (2.1.36)'$$

其中 c_1, c_2, A, θ 为任意常数。

2. 二阶常系数线性非齐次方程

由通解结构定理 2.1.4, 方程(2.1.29) 的通解等于对应的齐次方程通解与(2.1.29) 的一个特解之和。前面我们已得到了齐次方程的通解, 因而下面的问题是讨论如何求(2.1.29) 的一个特解。

情形 1° $g(x)$ 为多项式

若 $g(x) = b_0 x^m + b_1 x^{m-1} + \cdots + b_{m-1} x + b_m$

当 $b \neq 0$ 时, 可设(2.1.29) 的某一个特解具有如下形式

$$\bar{y} = B_0 x^m + B_1 x^{m-1} + \cdots + B_{m-1} x + B_m$$

其中 B_0, B_1, \cdots, B_m 为待定系数。

将 \bar{y} 代入方程(2.1.29), 得到常数 B_0, B_1, \cdots, B_m 所必须满足

的条件。

$$\begin{cases} bB_0 = b_0 \\ bB_1 + maB_0 = b_1 \\ bB_2 + (m-1)aB_1 + m(m-a)B_0 = b_2 \\ \dots\dots \end{cases} \quad (2.1.37)$$

逐一求出 B_0, B_1, \dots, B_m 即可求得特解 \bar{y} 。

情形 2° $g(x)$ 为指数函数

若 $g(x) = Ae^{ax}$, 且 a 不是特征根时,

设特解 $\bar{y} = Be^{ax}$, B 为待定常数。

代入(2.1.29)后得

$$Ba^2 + aBa + bB = A$$

解上式,得

$$B = \frac{A}{a^2 + aa + b}$$

$$\text{特解 } \bar{y}(x) = \frac{A}{a^2 + aa + b} e^{ax} \quad (2.1.38)$$

情形 3° $g(x)$ 是正弦、余弦函数

若 $g(x) = A_1 \cos wx + B_1 \sin wx$, 其中 w, A_1, B_1 为已知常数, 且 $A_1^2 + B_1^2 \neq 0, w \neq 0, wi$ 不是特征根。

设特解 $\bar{y} = \alpha \cos wx + \beta \sin wx$, α, β 是待定常数。代入方程(2.1.29)后得 α, β 所必须满足的条件

$$\begin{cases} b\alpha + aw\beta - w^2\alpha = A_1 \\ b\beta - aw\alpha - w^2\beta = B_1 \end{cases} \quad (2.1.39)$$

当 wi 不是特征根时, 此方程组一定有解。由上式解出 α, β , 即可求得特解 \bar{y} 。

说明:

(1) 对于二阶方程所提供的解法, 可完全类似地推广到 n 阶方程中。

(2) 在求解问题过程中, 可视具体情况将高阶方程转化成一阶微分方程组进行讨论。

例如, 对于二阶方程

$$\frac{d^2x}{dt^2} + a \frac{dx}{dt} + bx = 0 \quad (2.1.40)$$

令 $\frac{dx}{dt} = y$, 则与方程(2.1.40) 等价的一阶线性微分方程组为

$$\begin{cases} \frac{dx}{dt} = y \\ \frac{dy}{dt} = -bx - ay \end{cases} \quad (2.1.41)$$

2.1.5 微分方程的稳定性简介

在现实世界中, 任何系统总会受各种各样的干扰作用, 这种作用常常使系统偏离原来的给定的运动状态, 因而有必要研究这种作用对原来给定运动的影响。这就是微分方程的稳定性问题。

下面简单介绍方程的平衡点及稳定性概念, 并给出判断方程解稳定的初等方法。

设有微分方程

$$\frac{dx}{dt} = f(x) \quad (2.1.42)$$

右端不显含自变量 t , 代数方程

$$f(x) = 0 \quad (2.1.43)$$

的实根 $x = x_0$ 称为方程(2.1.42) 的平衡点。它也是方程(2.1.42) 的解。

定义 2.1.9 若从所有可能的初始条件出发, 方程(2.1.42) 的解 $x(t)$ 均满足

$$\lim_{t \rightarrow +\infty} x(t) = x_0$$

则称平衡点 x_0 是稳定的(稳定性理论中称为渐近稳定); 否则, 称

x_0 是不稳定的(不渐近稳定)。

完全类似的可定义二阶方程和 n 阶方程的平衡点及稳定性概念。

稳定性的判别方法是实际问题中经常要讨论的核心问题。下面给出稳定性判别的两个基本定理。

定理 2.1.6 对于常系数齐次微分方程

$$a_0 y^{(n)} + a_1 y^{(n-1)} + \cdots + a_{n-1} y' + a_n y = 0 \quad (a_0 \neq 0) \quad (2.1.44)$$

其特征方程(由(2.1.44)的系数构成的 n 次代数方程)为

$$a_0 \lambda^n + a_1 \lambda^{n-1} + \cdots + a_{n-1} \lambda + a_n = 0 \quad (a_0 \neq 0) \quad (2.1.45)$$

设 $\lambda_1, \lambda_2, \cdots, \lambda_n$ 是(2.1.45)的 n 个特征值(实或复的,重根按重数计算),且

$$\lambda_j = \alpha_j + i\beta_j \quad (j = 1, 2, \cdots, n) \quad (2.1.46)$$

则方程(2.1.44)零解渐近稳定的充分必要条件是

$$\alpha_j < 0 \quad (j = 1, 2, \cdots, n) \quad (2.1.47)$$

若特征根不易求出,而我们只关心各根实部符号,则可由下述定理判别各根的实部符号,从而进一步判断方程(2.1.44)的零解的稳定性质。

定理 2.1.7 (Hurwitz 判据) 给定 n 次代数方程(2.1.45)(不失一般性,可设 $a_0 > 0$)

作行列式

$$\Delta_n = \begin{vmatrix} a_1 & a_0 & 0 & 0 & \cdots & \cdots & \cdots & 0 \\ a_3 & a_2 & a_1 & a_0 & \cdots & \cdots & \cdots & 0 \\ a_5 & a_4 & a_3 & a_2 & a_1 & a_0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \vdots & & & & & & & \\ a_{2n-1} & a_{2n-2} & \cdots & \cdots & \cdots & \cdots & \cdots & a_n \end{vmatrix}$$

当 $i > n$ 时定义 $a_i = 0$, 则方程(2.1.45)所有的根有负实部的充分

必要条件是 Δ_n 的一切主子行列式都大于零,即

$$\Delta_1 = a_1 > 0, \Delta_2 = \begin{vmatrix} a_1 & a_0 \\ a_3 & a_2 \end{vmatrix} > 0, \Delta_3 = \begin{vmatrix} a_1 & a_0 & 0 \\ a_3 & a_2 & a_1 \\ a_5 & a_4 & a_3 \end{vmatrix} > 0, \\ \dots, \Delta_n = a_n \Delta_{n-1} > 0$$

§ 2.2 建模实例

2.2.1 小船行走路线问题

一小船从河边某点驶向对岸码头,若考虑水的流速影响,船行走的路线如何?

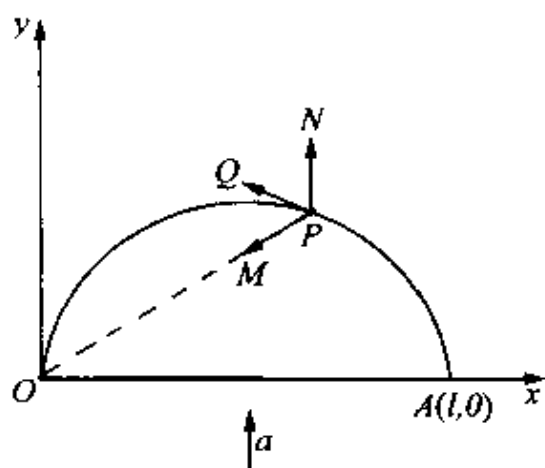


图 2.2.1

如图建立坐标系,设
(1) 水流方向为 y 轴正向,速度为 a ; (2) 船在 A 处,轮船匀速行驶,速度为 b ,为了到达码头,总是朝向码头 O 前进; (3) 船所行走的路线为 $y = y(x)$; (4) 河宽为 l 米。

在曲线 $y = y(x)$ 上任取一点 $P(x, y)$, 我们分析三个向量; 船速度向量, 水流速度向量和船行的速度向量。如图 2.2.1, 它们分别是 \overrightarrow{PM} , \overrightarrow{PN} 和 \overrightarrow{PQ} , 且

$$\overrightarrow{PQ} = \overrightarrow{PN} + \overrightarrow{PM}$$

因为水流方向为 y 轴正向, 大小为 a , 所以水流速度向量

$$\overrightarrow{PN} = \{0, a\}$$

划船方向指向原点 $O(0,0)$, 大小为 b , 所以划船速度向量

$$\overrightarrow{PM} = \left\{ \frac{-bx}{\sqrt{x^2 + y^2}}, \frac{-by}{\sqrt{x^2 + y^2}} \right\}$$

因此船行速度向量

$$\overrightarrow{PQ} = \left\{ \frac{-bx}{\sqrt{x^2 + y^2}}, \frac{-by}{\sqrt{x^2 + y^2}} + a \right\}$$

由船行速度向量的方向为船行路线的切线方向, 有

$$y' = \frac{\frac{-by}{\sqrt{x^2 + y^2}} + a}{\frac{-bx}{\sqrt{x^2 + y^2}}}$$

$$\text{或 } y' = \frac{y}{x} - \frac{a}{b} \sqrt{1 + \left(\frac{y}{x}\right)^2} \quad (2.2.1)$$

对于我们的问题是求上述方程满足条件

$$y(l) = 0 \quad (2.2.2)$$

的解。

方程(2.2.1) 是齐次方程, 由齐次方程解法得

$$y + \sqrt{x^2 + y^2} = cx^{1-k}, (k = \frac{a}{b} < 1)$$

由(2.2.2), 可求出任意常数 $c = l^k$ 。此外, 因

$$\frac{1}{y + \sqrt{x^2 + y^2}} = \frac{\sqrt{x^2 + y^2} - y}{x^2} = \frac{1}{c} x^{k-1}$$

结合前式, 可解出

$$y = \frac{l}{2} \left[\left(\frac{x}{l}\right)^{1-k} - \left(\frac{x}{l}\right)^{1+k} \right]$$

当 $x = 0$ 时确有 $y = 0$, 故小船一定能到达码头。

2.2.2 飞向月球问题

从太古时代以来, 人类一直关注着月球, 幻想着许多事情, 许

多人梦想着有那么一天,人们可以到月球上去旅行,我国民间就流传有“嫦娥奔月”的故事。随着近代科学技术的发展,特别是电子技术、火箭技术等高科技的发展,现在,幻想已经变成现实,人们到月球上去旅行的理想终于实现了。下面我们通过数学模型来论证“飞向月球”的理论依据。由此可转化为如下问题进行讨论。

设有一质量为 m 的发射体(飞船),由地面垂直地射向月球(如图 2.2.2),讨论其运动规律。

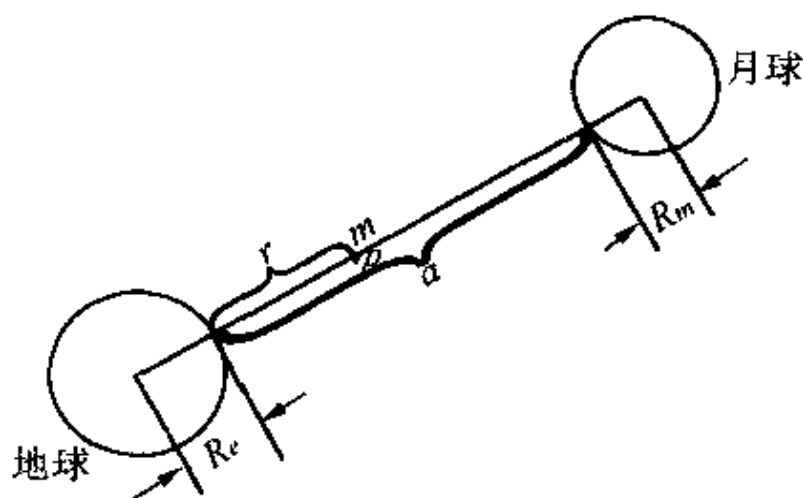


图 2.2.2

为使问题简化,我们做如下基本假设:

(1) 地球和月球是空间中的均匀球体,半径分别为 R_e 和 R_m , 质量分别为 M_e 和 M_m ;

(2) 质量为 m 的发射体(飞船)以初速 v_0 垂直地射向月球球心;

(3) 不考虑地球和月球的旋转,以及太阳和其他星球对它们的影响;

(4) 不计空气阻力。

为建立数学模型,我们将利用牛顿万有引力定律:任何两个物体都是相互吸引的,引力的大小跟两物体的质量的乘积成正比,而

与它们的距离平方成反比,即

$$F = G \frac{M_1 M_2}{r^2} \quad (2.2.3)$$

其中 M_1, M_2 是物体的质量, r 是它们之间的距离, G 是引力常数。

如图 2.2.2, 取地球向月球的方向为正, 根据牛顿第二定律和万有引力定律, 得

$$m \frac{d^2 r}{dt^2} = - \frac{GM_e m}{(r + R_e)^2} + \frac{GM_m m}{(a + R_m - r)^2}$$

或 $\frac{d^2 r}{dt^2} = - \frac{GM_e}{(r + R_e)^2} + \frac{GM_m}{(a + R_m - r)^2} \quad (2.2.4)$

注意到质量为 m 的物体对地球的引力是它的重力 mg , 由(2.2.3) 得

$$\frac{GM_e m}{R_e^2} = mg \quad \text{或} \quad GM_e = gR_e^2 \quad (2.2.5)$$

类似地, 以 g_m 表示在月球上的重力加速度, 有

$$GM_m = g_m R_m^2 \quad (2.2.6)$$

将(2.2.5), (2.2.6) 代入(2.2.4), 得

$$\frac{d^2 r}{dt^2} = - \frac{gR_e^2}{(r + R_e)^2} + \frac{g_m R_m^2}{(a + R_m - r)^2} \quad (2.2.7)$$

初始条件是

$$t = 0, r = 0, r' = v_0$$

注意到 $\frac{dr}{dt} = v, \frac{d^2 r}{dt^2} = v \frac{dv}{dr}$, (2.2.7) 可变为

$$v \frac{dv}{dr} = - \frac{gR_e^2}{(r + R_e)^2} + \frac{g_m R_m^2}{(a + R_m - r)^2} \quad (2.2.8)$$

初始条件化为

$$v|_{r=0} = v_0$$

利用可分离变量方法, 求得

$$v^2 = v_0^2 + \frac{2gR_e^2}{r + R_e} + \frac{2g_m R_m^2}{a + R_m - r} - 2gR_e - \frac{2g_m R_m^2}{a + R_m} \quad (2.2.9)$$

这就是速度 v 和距离 r 间的函数关系式;再将 v 换成 $\frac{dr}{dt}$, 我们就能确定 r 为时间 t 的函数。由此可以计算到达月球的时间(理论上的时间)。

事实上,由(2.2.8) 求出 v , 得

$$v = \frac{dr}{dt} = \sqrt{\varphi(r)}$$

其中 $\varphi(r)$ 由(2.2.9) 给出,积分上式得

$$t = \int_0^r \frac{dr}{\sqrt{\varphi(r)}}$$

上式右端的积分可化为椭圆积分来求值。

2.2.3 溶液混合问题

设有一容器装有某种浓度的溶液,以流量 v_1 注入浓度为 c_1 的同样溶液,假定溶液立即被搅匀,并以 v_2 的流量流出这种混合后的溶液,试建立容器中浓度与时间关系的数学模型。

设容器中溶液溶质的质量为 $x(t)$, 原来的初始质量为 x_0 , $t = 0$ 时溶液的体积为 v_0 。

由物质守恒定律,在 Δt 的时间间隔内,容器内溶质的改变量等于流入溶质的数量减去流出溶质的数量,即

$$\Delta x = c_1 v_1 \Delta t - c_2 v_2 \Delta t \quad (2.2.10)$$

其中 c_1 是输入溶液的浓度, c_2 为 t 时刻容器中溶液的浓度

$$c_2 = \frac{x}{v_0 + (v_1 - v_2)t}$$

用 Δt 除(2.2.10) 式两端,并令 $\Delta t \rightarrow 0$, 得

$$\begin{cases} \frac{dx}{dt} = c_1 v_1 - c_2 v_2 \\ x(0) = x_0 \end{cases} \quad (2.2.11)$$

这就是混合溶液的数学模型。

以上是我们利用数学、物理规律和微元分析法建立的数学模型。下面我们给出生物、社会、军事、医学等学科中的数学模型实例。

2.2.4 人口问题

人口问题是当今世界上人们普遍关注的问题之一。从我国建国以来的历史发展情况和当前的现实说明,建设社会主义现代化必须认真对待人口问题,有计划地控制人口增长的速度是我国的一项基本国策。建立人口增长模型,可用来描述人口增长过程,通过分析对人口增长进行预测,制定相应的人口政策以控制人口增长。

影响人口增长的因素很多:人口的基数,人口的自然增长率和各种诸如自然灾害、传统习惯、医疗条件、战争等扰动因素等等,如果一开始把诸多因素都考虑进去,则无从下手,因此我们可先把问题简化,抓住主要因素——净增长率(出生率减去死亡率),舍去众多的次要因素,建立比较简单的模型,再通过对解的讨论逐步修改,得到较完善的模型。

模型1 Malthus 模型

英国人口统计学家 Malthus(1766 ~ 1834) 在担任牧师期间,查看了教堂 100 多年人口出生统计资料,于 1798 年提出了著名的人口指数增长模型,这个模型的基本假设是:在人口自然增长过程中,人口的净相对增长率是常数,或者说,单位时间内人口的增长量与当时的人口成正比。

如何建立 Malthus 的数学模型呢?

设时刻 t 的人口数为 $N(t)$,当考察一个国家或很大地区的人口时, $N(t)$ 是很大的整数。为了便于利用分析的工具和方法,视 $N(t)$ 为连续,可微函数,设 $t = t_0$ 时的人口为 N_0 ,人口增长率为 r ,根据 Malthus 的假设,在 t 到 $t + \Delta t$ 时间内人口的增长量为

$$N(t + \Delta t) - N(t) = rN(t)\Delta t$$

令 $\Delta t \rightarrow 0$, $N(t)$ 满足的方程为

$$\frac{dN}{dt} = rN \quad (2.2.12)$$

解上式, 并利用 $N(t_0) = N_0$, 得

$$N(t) = N_0 e^{r(t-t_0)} \quad (2.2.13)$$

(2.2.12) 是一个线性方程, 称为 *Malthus* 人口模型。

由(2.2.13)式给出的模型, 与19世纪以前欧洲一些地区的人口统计数据可以很好地吻合。一些人口增长率长期稳定不变的国家 and 地区用这个模型进行预报, 结果也令人满意。

但是当人们用19世纪以后许多国家的人口统计资料与 *Malthus* 模型计算结果比较时, 却发现了相当大的差异。表2.2.1列出了美国19世纪、20世纪的人口统计数据与这个模型的比较结果。表2.2.1的第2列是美国从1790年以来的实际人口, 第3列是本世纪初人们按 *Malthus* 模型计算结果, 其中取1790年为 $t = t_0 = 0$ (t 以10年为单位), 用这一年的实际人口数 $N_0 = 3.9 \times 10^6$ 与1800年的实际人口 $N(1) = 5.3 \times 10^6$, 确定出 $r = 0.307$, 然后按(2.2.13)式计算。

再用这个模型预测未来人口, 将会发现世界人口总数在2510年将是 2×10^{14} 人(2万亿), 2635年将是 1.8×10^{15} 人(18万亿), 2670年将是 3.6×10^{15} 人(36万亿)。显然, 这些数学表明 *Malthus* 人口模型对长期的预测是不正确的。

由以上分析可以看出, *Malthus* 人口模型对未来的总人口数预测不正确, 应该予以修正。

模型2 *Logistic* 模型

为什么 *Malthus* 模型只符合人口的过去结果而不能用于预测未来人口呢? 主要原因是, 随着人口的增加, 自然资源、环境条件等因素对人口继续增长的阻滞作用越来越显著。如果当人口较少时, 人口的自然增长率可看作常数, 那么当人口增加到一定数量后, 增

长率就会随着人口的继续增加而逐渐减少。许多国家人口增长的

表 2.2.1 美国的实际人口
与按两种模型计算的人口的比较

年	实际人口 ($\times 10^6$)	指 数 模 型		Logistic 模 型	
		($\times 10^6$)	误差(%)	($\times 10^6$)	误差(%)
1790	3.9				
1800	5.3				
1810	7.2	7.3	1.4		
1820	9.6	10.0	4.2	9.7	1.0
1830	12.9	13.7	6.2	13.0	0.8
1840	17.1	18.7	9.4	17.4	1.8
1850	23.2	25.6	10.3	23.0	-0.9
1860	31.4	35.0	10.8	30.2	-3.8
1870	38.6	47.8	23.8	38.1	-1.3
1880	50.2	65.5	30.5	49.9	-0.6
1890	62.9	89.6	42.4	62.4	-0.8
1900	76.0	122.5	61.2	76.5	0.7
1910	92.0	167.6	82.1	91.6	-0.4
1920	106.5	229.3	115.3	107.0	0.5
1930	123.2			122.0	-1.0
1940	131.7			135.9	3.2
1950	150.7			148.2	-1.7
1960	179.3			158.8	-11.4
1970	204.0			167.6	-17.8
1980	226.5				

实际情况完全证实了这一点。如果我们利用表 2.2.1 第 2 列给出

的数据计算一下美国人口每 10 年的增长率,也可以知道大致是逐渐下降的。显然,应该对 *Malthus* 模型关于净相对增长率为常数的假设进行修改。

将增长率 r 表示为人口 $N(t)$ 的函数 $r(N)$, 按照前面的分析, $r(N)$ 应是 N 的减函数。一个最简单的假定是设 $r(N)$ 为 N 的线性函数。

$$r(N) = r - sN \quad (r, s > 0)$$

这里 r 相当于 $N = 0$ 时的增长率, 称固有增长率。显然对于任意的 $N > 0$, 增长率 $r(N) < r$ 。为确定系数 s 的意义, 引入自然资源和环境条件所能容纳的最大人口数量 N_m , 称最大人口容量。当 $N = N_m$ 时增长率应为零, 即 $r(N_m) = 0$, 由此可确定出 s 。人口增长率 $r(N)$ 可表示为

$$r(N) = r(1 - \frac{N}{N_m})$$

其中 r, N_m 是根据人口统计数据或经验确定的常数。

按照上述假定, (2.2.12) 应修改为

$$\begin{cases} \frac{dN}{dt} = r(1 - \frac{N}{N_m})N \\ N(t_0) = N_0 \end{cases} \quad (2.2.14)$$

上式称为 *Logistic* 模型。其解为

$$N(t) = \frac{N_m}{1 + (\frac{N_m}{N_0} - 1)e^{-r(t-t_0)}} \quad (2.2.15)$$

由解(2.2.15), 可得出人口总数具有如下规律:

(1) 当 $t \rightarrow \infty$ 时, $N(t) \rightarrow N_m$, 即无论人口初值如何, 人口总数趋向于极限值 N_m 。

(2) 当 $0 < N_0 < N_m$ 时, $\frac{dN}{dt} = r(1 - \frac{N}{N_m})N > 0$, 所以 $N(t)$ 是时间 t 的单调递增函数, 又

$$\frac{d^2N}{dt^2} = r^2(1 - \frac{N}{N_m})(1 - \frac{2N}{N_m})N$$

显然, 当 $N < \frac{N_m}{2}$ 时, $\frac{d^2N}{dt^2} > 0$, 曲线向上凹;

当 $N > \frac{N_m}{2}$ 时, $\frac{d^2N}{dt^2} < 0$, 曲线向下凹。

从 $N(t) \sim t$ (S形曲线) 和 $\frac{dN}{dt} \sim N$ 的曲线图, 见图 2.2.3, 可得如下结论:

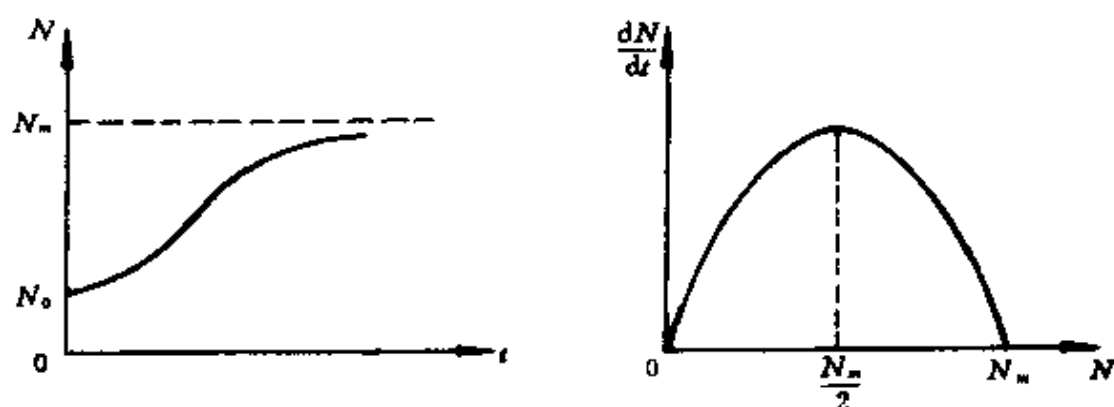


图 2.2.3

人口增长率 $\frac{dN}{dt}$ 由增变减, 在 $\frac{N_m}{2}$ 处最大, 即在人口总数达到极限值一半以前是加速生长时期, 过这一点以后, 生长的速度逐渐变小, 并且迟早会达到零, 这是减速生长时期。

为了与 *Malthus* 模型比较, 表 2.2.1 第 5 列是本世纪人们用 *Logistic* 模型计算的美国人口, 其中仍取 1790 年为 $t = t_0 = 0$, $N_0 = 3.9 \times 10^6$, 而取 $N_m = 197 \times 10^6$, $r = 0.31$ 。可以看出, 直到 1930 年计算结果都能与实际数据较好地吻合。后来的误差越来越大, 一个明显的原因是在 60 年代美国的实际人口数已经突破了本世纪

初所假设的极限人口 N_m 。由此可见这个模型缺点之一就是 N_m 不易确定。事实上,随着生产力的发展和人们认识能力的改变, N_m 也是可以改变的。

下面我们用 *Logistic* 模型来预测世界未来人口总数,某些生物学家估计, r 的自然值为 0.029。又据估计,全世界 1961 年人口总数为 3.06×10^9 ,人口每年以 2% 的速率增长。由(2.2.14):

$$\frac{1}{N} \frac{dN}{dt} = r(1 - \frac{N}{N_m})$$

$$\text{即 } 0.02 = 0.029(1 - \frac{3.06 \times 10^9}{N_m})$$

从而得到

$$N_m = 9.86 \times 10^9$$

即世界人口总数极限值近 100 亿。

由上预测知,1961 年世界人口总数为 30 亿左右,尚未达到极限值 100 亿的一半。因此世界人口总数将处于加速生长时期。这与 1961 年以后的一段时期世界人口增长很快的实际情况是吻合的。

几点说明:

(1) 人也是一种生物,本节讨论的人口模型,原则上可适用于其他生物种群,如森林中的树木,池塘中的鱼等, *Logistic* 模型在生物数量的分析和预测中有着广泛的应用。

(2) 在 *Malthus* 模型和 *Logistic* 模型中,把总数看作是处于同等地位的成员组成的。严格讲是不对的。要建立更精确的数学模型,应根据成员的年龄分组,还应把总数的成员按性别分开,比如可应用 *Leslie* 模型等。

我国学者为了解决我国人口迅猛增长问题,作了大量的调查研究,建立了人口模型,利用数学模型与实际人口普查结果比较,误差很小,在五年预测中误差仅在千分之一数量级上。有兴趣的读者可参考有关人口问题的文献。

2.2.5 军事战争问题

战争是一个非常复杂的问题,涉及因素很多,如兵员的数量和质量,武器的先进与落后,地理位置的有利与不利,指挥员的艺术、后勤供应、气候条件等。因此,如果把战争所涉及的因素都考虑进去,这样的模型是既难建立又难解决。但是对于一个通常情况下的局部战争,在合理的假设下选择主要因素建立一个作战数学模型,我们将会看到得出的结论是具有普遍意义的。

在第一次世界大战期间, $F \cdot W \cdot Lanchester$ 就投身于作战模型的研究,并得到了一些可以从中得到交战结果的数学模型。

对于一次局部战斗,有些因素可以不考虑,如气候、后勤供应、士气等,而有些因素可认为双方是相同的,如武器装备、指挥艺术等。

模型1 正规战争模型

令 $x(t)$ 表示 t 时刻甲军人数, $y(t)$ 表示 t 时刻乙军人数,假设

- (1) 每一方人数减员率与另一方人数成正比;
- (2) 两军士兵都处于双方火力范围内;
- (3) 不考虑双方支援部队;
- (4) 双方的初始兵力分别是 x_0 和 y_0 。

由以上假设可得正规部队的作战模型为

$$\begin{cases} \frac{dx}{dt} = -ay \\ \frac{dy}{dt} = -bx \\ x(0) = x_0, y(0) = y_0 \end{cases} \quad (2.2.16)$$

其中 $a > 0, b > 0$ 均为常数,表示战斗系数。

积分(2.2.16),得

$$ay^2 - bx^2 = ay_0^2 - bx_0^2 = c \quad (2.2.17)$$

这就是著名的“Lanchester 平方定律”。(2.2.17) 式在 $x-y$ 平面上是一簇双曲线,如图 2.2.4 所示。双曲线上箭头表示战斗力随着时间而变化的方向。

可以看出,如果 $c > 0$, 轨线将与 y 轴相交,这就是说存在 t_1 , 使 $x(t_1) = 0, y(t_1) = \sqrt{\frac{c}{a}} > 0$, 即当甲方兵力为零时乙方兵力为正值,表明乙军获胜。同理可知 $c < 0$ 时甲军获胜。当 $c = 0$ 时双方战平。

进一步分析可知,乙军要想获胜,即要使 $ay_0^2 > bx_0^2$ 成立。可采用两种方式:(1) 增加 a , 即配备更先进的武器。(2) 增加最初投入战斗的人数 y_0 。但是,值得注意的是:在(2.2.17) 中, a 增大两倍,结果 ay_0^2 也增大两倍,但 y_0 增大两倍则会使 ay_0^2 增大四倍。这正是两军摆开战场作正规战时 Lanchester 平方定律的意义,说明兵员增加战斗力将大大增加。

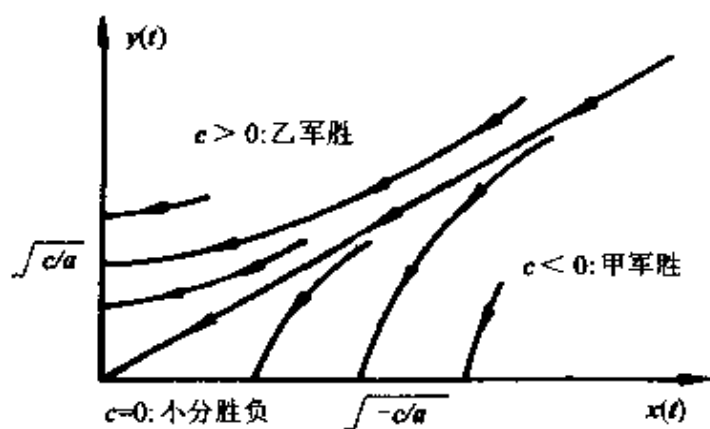


图 2.2.4

为此我们考虑有增援情况的战争模型。即除了模型 1 的基本假定之外,进一步假设 $f(t)$ 和 $g(t)$ 分别表示甲军和乙军 t 时刻的增援率。(所谓增援率,就是增援战士投入战斗或战士撤离战斗

的速度)。此时正规作战模型为

$$\begin{cases} \frac{dx}{dt} = -ay + f(t) \\ \frac{dy}{dt} = -bx + g(t) \end{cases} \quad (2.2.18)$$

模型2 混合战争模型

如果甲军是游击队,乙军是正规部队,由于游击队对当地地形熟,常常位于不易发现的有利地形。设游击队占据区域 R ,由于乙军看不清楚甲军,只好向区域 R 射击,但并不知道杀伤情况。为此,我们假设:

(1) 游击队 $x(t)$ 的战斗减员率应当与 $x(t)$ 成正比,(因为 $x(t)$ 越大,目标越大;被敌方子弹命中的可能性越大)。

(2) 游击队 $x(t)$ 的战斗减员率还与 $y(t)$ 成正比,(因为 $y(t)$ 越大,火力越强, $x(t)$ 的伤亡人数也就越大)。

(3) 游击队和正规部队的增援率分别为 $f(t)$ 和 $g(t)$ 。

由以上假设可知,游击队的战斗减员率等于 $cx(t)y(t)$,作战模型为

$$\begin{cases} \frac{dx}{dt} = -cxy + f(t) \\ \frac{dy}{dt} = -dx + g(t) \end{cases} \quad (2.2.19)$$

其中 c 称为敌方 y 的战斗有效系数。

若无增援 $f(t)$ 和 $g(t)$,则

$$\begin{cases} \frac{dx}{dt} = -cxy \\ \frac{dy}{dt} = -dx \end{cases} \quad (2.2.20)$$

积分(2.2.20) 式得

$$cy^2 - 2dx = cy_0^2 - 2dx_0 = M \quad (2.2.21)$$

(2.2.21) 式在 $x - y$ 平面上定义了一簇抛物线,如图 2.2.5。

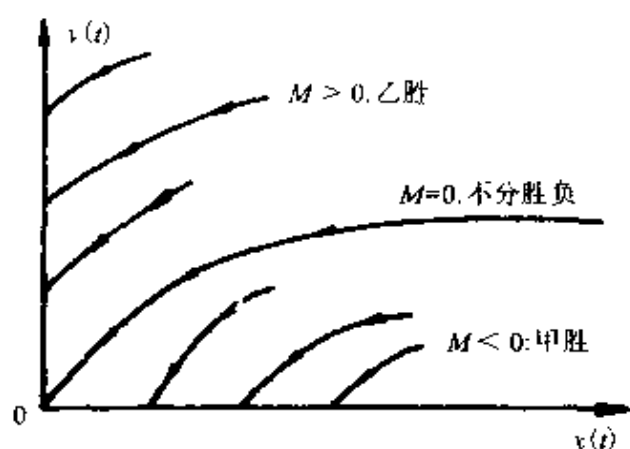


图 2.2.5

如果 $M > 0$, 则正规部队胜。因为当 $y(t)$ 减小到 $\sqrt{\frac{M}{c}}$, 游击队 $x(t)$ 已被消灭。同样, 如果 $M < 0$, 则游击队胜。

几点说明:

(1) 在模型(2.2.18)中, 如果 $a, b, f(t)$ 和 $g(t)$ 已知, 则可用显式求解。在模型(2.2.19)中, 因方程组是非线性的, 求解困难, 可利用计算机求解。

(2) 事前确定战斗系数 a, b, c 和 d 的数值通常是不可能的。但是如果对已有的战役资料确定 a 和 b (或者 c 和 d) 的适当系数值, 那末对于其他类似于同样条件下进行的战斗, a 和 b (或 c 和 d) 这些系数可认为是已知的。

(3) 由(2.2.21)式有

$$\left(\frac{y_0}{x_0}\right)^2 > \frac{2b}{cx_0}$$

利用上式可以估计出正规部队要取得胜利需投入多少初始兵力。美国人曾用这个模型分析越南战争, 以及二战时美国和日本硫黄岛战役, 发现模型结果与实际数据吻合的很好, 这就说明了

Lanchester 作战模型是能够用来描述实际战争的。有兴趣的读者可参考有关文献。

(4) 上述模型没有考虑交战双方的政治、经济、社会以及自然环境等因素,因而仅靠战场上兵力的优劣是很难估计战争胜负的,所以我们认为用这些模型判断整个战争的结局是不可能的,但是对于局部战役来说或许还有参考价值。更重要的是,建模的思路和方法为我们借助数学模型讨论社会科学问题提供了可借鉴的示例。

2.2.6 宣传运动的效果研究问题

对于大众关心问题,政府总要通过各种新闻媒介进行宣传和教育活动,如开展预防传染病流行的宣传运动,那么这种宣传效果对防止传染病的蔓延会起到多大作用,为什么宣传运动要常抓不懈,并且要有一定强度,下面我们从数学模型的角度来分析一下宣传运动的必要性。

模型1 不开展宣传运动模型

设总人数 N 不变, $x(t)$ 为 t 时刻得病人数,它传染给正常人的传染率为 r (r 为常数)。

显然,从 t 到 $t + \Delta t$ 时间内平均传染率是

$$\frac{x(t + \Delta t) - x(t)}{\Delta t(N - x(t))}$$

令 $\Delta t \rightarrow 0$, 得 t 时刻的传染率

$$\frac{1}{N - x} \frac{dx}{dt} = r$$

因此,我们得到 $x(t)$ 所满足的数学模型

$$\begin{cases} \frac{dx}{dt} = r(N - x) \\ x(0) = x_0 \end{cases} \quad (2.2.22)$$

解(2.2.22), 得

$$x(t) = N[1 - (1 - \frac{x_0}{N})e^{-rt}]$$

对上式分析,我们发现,当 $t \rightarrow +\infty$ 时, $\lim_{t \rightarrow +\infty} x(t) = N$ 。这说明,最终每个人都要染上疾病。

模型2 持续宣传运动模型

为了预防传染病流行,“宣传”有如在人们耳边敲起警钟“要当心!”因而是很重要的,下面讨论开展持续宣传运动的效果。

假设宣传的开展将使得传染上疾病的人数 $x(t)$ 减少,且

(1) 减少的速度与总人数 N 成正比,比例常数取决于宣传强度。

(2) 从 $t = t_0 > 0$ 开始,开展一场持续的宣传运动,宣传强度为 a ($0 < a < r$)。

由假设可得数学模型

$$\begin{cases} \frac{dx}{dt} = r(N - x) - aNH(t - t_0) & (t \geq 0) \\ x(0) = x_0 \end{cases} \quad (2.2.23)$$

其中

$$H(t - t_0) = \begin{cases} 1 & t \geq t_0 \\ 0 & t < t_0 \end{cases}$$

即 Heaviside 函数。

利用分段求解方法,可求得(2.2.23),即求解

当 $0 \leq t < t_0$ 时解方程

$$\begin{cases} \frac{dx}{dt} = r(N - x) \\ x(0) = x_0 \end{cases}$$

由模型1,得

$$x(t) = N[1 - (1 - \frac{x_0}{N})e^{-rt}]$$

当 $t \geq t_0$ 时解一阶非齐次方程的初值问题:

$$\begin{cases} \frac{dx}{dt} = r(N - x) - aN \\ x(t_0) = N[1 - (1 - \frac{x_0}{N})e^{-rt_0}] \end{cases}$$

解上式,得

$$x(t) = N[1 - (1 - \frac{x_0}{N})e^{-rt}] - \frac{aN}{r}[1 - e^{-r(t-t_0)}]$$

两段表达式合起来,得(2.2.23)的解为

$$x(t) = N[1 - (1 - \frac{x_0}{N})e^{-rt}] - \frac{aN}{r}H(t - t_0)[1 - e^{-r(t-t_0)}]$$

令 $t \rightarrow +\infty$, 得

$$\lim_{t \rightarrow +\infty} x(t) = N(1 - \frac{a}{r}) < N$$

这说明持续的宣传是起作用的,最终会使发病率减少($0 < a < r$)。

模型3 瞬时宣传运动模型

如果宣传运动是短暂进行的,这在日常生活中是常见的,例如仅仅是听一个报告,或街头散发传单等,即在 $t = t_1, t_2, \dots, t_m$ 等 m 个时刻进行 m 次宣传,宣传强度分别为 a_1, a_2, \dots, a_m , 则数学模型变成。

$$\begin{cases} \frac{dx}{dt} = r(N - x) - N \sum_{j=1}^m a_j \delta(t - t_j) \\ x(0) = x_0 \end{cases} \quad (2.2.24)$$

其中我们用 $a_j \delta(t - t_j)$ 来数学地近似描述时间极为短暂的宣传。

用拉氏变换可求解方程(2.2.24),得

$$x(t) = x_0 e^{-rt} + N[1 - e^{-rt}] - N \sum_{j=1}^m a_j H(t - t_j) e^{-r(t-t_j)}$$

令 $t \rightarrow +\infty$, 得

$$\lim_{t \rightarrow +\infty} x(t) = N$$

这说明短暂的宣传(即使是多次的短暂宣传)是不起作用的,最终还是所有的人都传染上疾病。

几点说明:

(1) 由以上分析可知,要想对某个问题进行宣传,引起人们的注意,并产生一定的效果,必须开展持续的宣传活动。

(2) 对于瞬时宣传模型,实际情况是,即使短暂的宣传也往往会在人们的头脑中留下一定时间延续的印象,可考虑进一步建立模型问题。

2.2.7 关于诊断糖尿病的问题

糖尿病是一种新陈代谢疾病,其特征是血液和尿中糖份过分。糖尿病常常是通过葡萄糖耐量试验来诊断。病人要在前一天晚上禁食以后来医院,并在口服大量葡萄糖后 3 ~ 5 小时内,测定病人的血液中葡萄糖的浓度,然后利用测量结果进行诊断。这种诊断碰到的最大困难是怎样解释试验结果。六十年代,美国阿克曼等人为了建立了一个相当可靠的准则。

假设在进行葡萄糖耐量试验时,主要考虑两个浓度:血糖浓度 G ; 胰岛素浓度 H , 后者代表有关激素的总效果。象胰岛素这样的降低血糖浓度的激素被认为将增加 H , 而象皮质醇这样的增加血糖浓度的激素被认为将减小 H 。这个模型可用下列方程组来描述

$$\begin{cases} \frac{dG}{dt} = F_1(G, H) + J(t) \\ \frac{dH}{dt} = F_2(G, H) \end{cases} \quad (2.2.25)$$

其中 $J(t)$ 是血糖浓度增加的外界速度。 F_1 和 F_2 依赖于 G 和 H , 表明 G 和 H 的变化由 G 和 H 的值共同决定。

进一步假设,暂时断食的病人到医院时, G 和 H 达到该人的最佳值 G_0 和 H_0 。这意味着

$$F_1(G_0, H_0) = 0, F_2(G_0, H_0) = 0$$

因为 G 和 H 对它们的最佳值的偏差是我们关心的量, 所以作变换

$$g = G - G_0, h = H - H_0$$

于是, (2.2.25) 可写为

$$\begin{cases} \frac{dg}{dt} = F_1(G_0 + g, H_0 + h) + J(t) \\ \frac{dh}{dt} = F_2(G_0 + g, H_0 + h) \end{cases} \quad (2.2.26)$$

注意到

$$\begin{aligned} & F_1(G_0 + g, H_0 + h) \\ &= F_1(G_0, H_0) + \frac{\partial F_1(G_0, H_0)}{\partial G} g + \frac{\partial F_1(G_0, H_0)}{\partial H} h + e_1 \\ & F_2(G_0 + g, H_0 + h) \\ &= F_2(G_0, H_0) + \frac{\partial F_2(G_0, H_0)}{\partial G} g + \frac{\partial F_2(G_0, H_0)}{\partial H} h + e_2 \end{aligned}$$

其中 e_1 和 e_2 与 g 和 h 相比, 都是高阶小量。因此, 假定 G 和 H 只是稍微偏离 G_0 和 H_0 , 则可以略去 e_1 和 e_2 , 得

$$\begin{cases} \frac{dg}{dt} = \frac{\partial F_1(G_0, H_0)}{\partial G} g + \frac{\partial F_1(G_0, H_0)}{\partial H} h + J(t) \\ \frac{dh}{dt} = \frac{\partial F_2(G_0, H_0)}{\partial G} g + \frac{\partial F_2(G_0, H_0)}{\partial H} h \end{cases} \quad (2.2.27)$$

对于 (2.2.27), 我们无法预先确定系数 $\frac{\partial F_i(G_0, H_0)}{\partial G}$, $\frac{\partial F_i(G_0, H_0)}{\partial H}$ 的值, ($i = 1, 2$), 但是我们可确定这四个常数符号。

从医学角度看, 当 $g > 0, h = 0$ 时, $\frac{dg}{dt} < 0$ 。

因此通过组织对葡萄糖的吸收和将过量的葡萄糖以糖原的形式贮藏在肝脏内, 血糖浓度将降低, 所以 $\frac{\partial F_1(G_0, H_0)}{\partial G} < 0$ 。

因为通过促进组织对葡萄糖的吸收和增加葡萄糖向糖原转化的速率,血糖水平将降低,所以 $\frac{\partial F_1(G_0, H_0)}{\partial H} < 0$ 。

因为 g 的一个正值将引起内分泌腺分泌激素使 H 增加,所以 $\frac{\partial F_2(G_0, H_0)}{\partial G} > 0$ 。

因为通过激素的新陈代谢,血液中的浓度将会降低,所以 $\frac{\partial F_2(G_0, H_0)}{\partial H} < 0$ 。

由以上分析,可将方程(2.2.27)简写成:

$$\begin{cases} \frac{dg}{dt} = -m_1g - m_2h + J(t) \\ \frac{dh}{dt} = -m_3h + m_4g \end{cases}$$

其中 m_1, m_2, m_3 和 m_4 均为正常数。

由于我们只测量血液中葡萄糖的浓度,所以对于方程(2.2.27)采用消元法,不难得到关于 $g(t)$ 的二阶线性方程

$$\begin{aligned} \frac{d^2g}{dt^2} + (m_1 + m_2) \frac{dg}{dt} + (m_1m_3 + m_2m_4)g &= m_3J + \frac{dJ}{dt} \\ \text{或} \quad \frac{d^2g}{dt^2} + 2\alpha \frac{dg}{dt} + w_0^2g &= s(t) \end{aligned} \quad (2.2.28)$$

其中

$$\alpha = \frac{1}{2}(m_1 + m_2), w_0^2 = m_1m_3 + m_2m_4, s(t) = m_3J + \frac{dJ}{dt}$$

注意到,只要不是在很短的时间内吞食大量的葡萄糖,方程(2.2.28)的右端就恒等于零。为了解决我们的问题,设 $t = 0$ 时完全吃下大量葡萄糖,此时,当 $t \geq 0$ 时, $g(t)$ 满足

$$\frac{d^2g}{dt^2} + 2\alpha \frac{dg}{dt} + w_0^2g = 0 \quad (2.2.29)$$

其特征方程为

$$\lambda^2 + 2\alpha\lambda + w_0^2 = 0$$

特征根 $\lambda_{1,2} = -\alpha \pm \sqrt{\alpha^2 - w_0^2}$, 因为 $\alpha > 0, w_0 > 0$, 二个特征根皆有负实部。(2.2.29) 的通解为

$$g(t) = Ae^{-\alpha t} \cos(\omega t - \delta) \quad (2.2.30)$$

其中 $\omega^2 = w_0^2 - \alpha^2, \alpha > 0$

显然当 $t \rightarrow +\infty$ 时, $g(t) \rightarrow 0$, 解稳定。可见这个数学模型确实符合于所预料的实际情况, 血糖浓度将返回到它的最佳值 G 。

由 $g = G - G_0$, 我们有

$$G(t) = G_0 + Ae^{-\alpha t} \cos(\omega t - \delta) \quad (2.2.31)$$

其中 G_0, A, α, ω 和 δ 为未知常数。可由下列方法确定。 G_0 用病人刚到医院口服大量葡萄糖前测量的血糖浓度值确定; 而其余四个数 A, α, ω 和 δ 可以用在 t_1, t_2, t_3 和 t_4 时刻测量的血糖浓度 G_1, G_2, G_3 和 G_4 , 由方程

$$G_i = G_0 + Ae^{-\alpha t_i} \cos(\omega t_i - \delta) \quad (i = 1, 2, 3, 4)$$

中求出。

然后以 $T_0 = \frac{2\pi}{\omega_0}$ 值为标准, 当 T_0 小于 4 小时, 是正常的; 而当 T_0 明显大于 4 小时, 则意味着受检人有轻微糖尿病。

几点说明:

(1) 以下确定 G_0, A, α, ω_0 和 δ 的方法更准确。即在 t_1, t_2, \dots, t_n 时刻测量出病人的血糖浓度 G_1, G_2, \dots, G_n 。典型地, n 取 6 或 7。然后求出 G_0, A, α, ω_0 和 δ 的最佳值, 使得误差

$$E = \sum_{j=1}^n [G_j - G_0 - Ae^{-\alpha t_j} \cos(\omega t_j - \delta)]^2$$

最小。 E 的最小化问题可以由计算机解决。这种方法比上一种方法更可取, 因为(2.2.31) 只是关于 $G(t)$ 的一个近似公式。从而, 可能出现这样的情况, 求出的 G_0, A, α, ω_0 和 δ 的值在 t_1, t_2, t_3, t_4 四个点处精确地满足(2.2.31), 而在其他时刻拟合甚差。第二种方

法由于用到了更多的测量值,因而在整个时间区间内提供了与数据更好的拟合。

(2) 上述模型只适用于诊断轻微的糖尿病或发现糖尿病前兆,因为我们始终都假设了 G 与其最佳值 G_0 的偏差 g 很小。 G 与 G_0 的偏差很大通常表示严重的糖尿病或尿毒症,这是脑垂体后叶的一种失调。

2.2.8 家用电器销售问题

随着科学技术的发展,人们生活日益得到改善和提高,家用电器化已是大势所趋。例如电饭锅在发达国家已是很受欢迎的一种。对于电饭锅制造厂的计划部门,应该对产品的需求进行预测,以便科学地安排生产计划,防止盲目生产,避免不必要的浪费和损失。

设 $x(t)$ 表示开始销售以来, t 时刻电饭锅售出总数, x_∞ 表示最大需求量。

模型 1 最简模型

商品的销售与广告密切相关,除了利用电视、广播、报刊作广告外,实物广告的效果更好。据统计,实物广告的作用相当于电视、广播、报刊广告的 10 倍。即电饭锅的销售速度 $\frac{dx}{dt}$ 与 $x(t)$ 成比例,设比例系数为 λ ,则有

$$\frac{dx}{dt} = \lambda x \quad (2.2.32)$$

积分(2.2.32) 式得

$$x = ce^{\lambda t} \quad (2.2.33)$$

即电饭锅的售销是按指数函数增加的。

设 $t = t_0$ 时, $x = a > 0$,

则此时(2.2.33) 式为

$$x = ae^{\lambda(t-t_0)}$$

关于 λ 值的确定,可利用销售以来几个月内所记录的销售量

求出 $\frac{dx}{dt}$, 再求出各 t 值的 $(\frac{dx}{dt})/x$, 加以平均即可。

类似于宣传效果模型问题的讨论, 电饭锅销售模型 1 在销售初期, 可以认为是合适的。但是当产品已经普及到大多数家庭时, 模型 1 中的假设不能成立, 为此需作修改。

模型 2 改进模型

设产品相当普及时, 销售速度 $\frac{dx}{dt}$ 与潜在的需要量 $(x_{\infty} - x)$ 成正比。从该产品销售的全过程看, 应该将这个性质与第一步时的情形结合起来考虑。电饭锅销售速度 $\frac{dx}{dt}$ 应与 $x(x_{\infty} - x)$ 成比例, 由此得

$$\frac{dx}{dt} = \lambda x(x_{\infty} - x) \quad (2.2.34)$$

整理上式, 得

$$\frac{dx}{x(x_{\infty} - x)} = \lambda dt$$

积分得

$$\frac{1}{\lambda_{\infty}} \int \left(\frac{1}{x_{\infty} - x} + \frac{1}{x} \right) dx = \int \lambda dt + c' \quad (2.2.35)$$

解(2.2.35) 式, 得

$$x(t) = \frac{x_{\infty}}{1 + ce^{-\lambda x_{\infty} t}} \quad (2.2.36)$$

(2.2.36) 式中 c 和 λ 的确定

由(2.2.35) 式有

$$\frac{1}{\lambda_{\infty}} (-\ln(x_{\infty} - x) + \ln x) = \lambda t + c'$$

$$\ln \frac{x_{\infty} - x}{x} = -x_{\infty} \lambda t - x_{\infty} c'$$

令 $Y = \ln \frac{x_{\infty} - x}{x}$, $X = -x_{\infty} t$, $b = -x_{\infty} c'$, 则上式变成

$$Y = \lambda X + b。$$

于是问题转化成根据过去的数据 (X_i, Y_i) 按最小二乘法求 $Y = \lambda X + b$ 的回归直线问题。

电饭锅销售模型具有普遍的实用性,一个工厂要开发新产品或者扩大再生产,应对销售情况进行预测,包括短期的和长期的预测,防止一哄而起,一哄而下的情形发生。

习 题

1. 一质量为 m 的物体,从空中自由落下,设空气阻力与速度成正比,试求在运动过程中速度与时间的关系。

2. 在简单的化学反应中,由物质 A 生成物质 B 。如果在 t 时刻 B 的质量为 x ,在温度以及其他一些条件不变的情况下,反应速度 $\frac{dx}{dt}$ 和 A 的剩余质量成正比。如果初始时刻物质 A 的质量为 a ,考虑生成 B 的质量变化规律。

3. 用均匀材料设计一个高为 h 、顶面直径为 a 的旋转体的支柱,如果支柱顶部所受压力为 P ,并要求每一水平截面上的压强都相等,那么应如何设计这个旋转体。

4. 在某池塘内养鱼,该池塘最多能养鱼 1000 尾。在时刻 t , 鱼数 y 是时间 t 的函数,其变化率与鱼群 y 及 $1000 - y$ 成正比。已知在池塘内放养鱼 100 尾,3 个月后池塘内有鱼 250 尾,求放养 t 月后池塘内鱼数 $y(t)$ 的公式。

5. 如果在一个水域中生存两种鱼。第一种鱼以水中的浮游生物为食物,这种食物永远是很充分的;第二种鱼以第一种鱼为食物,在这种共生过程中,试研究两种鱼数量之间的变化规律。

$$6. \text{ 方程组 } \begin{cases} \frac{dx}{dt} = -ay \\ \frac{dy}{dt} = -by - cxy \end{cases}$$

是正规部队对游击队作战的一个 *Lanchester* 模型,其中游击队 y 的非战斗减员率与 $y(t)$ 成正比。

(1) 求方程组的轨线。

(2) 试问哪一方胜利。

作战部队的非战斗减员率为指非战斗的原因(如开小差、疾病等)减员。

7. 考虑一种既不同于 *Malthus* 模型,也不同于 *Logistic* 模型的情况:设人口为 $N(t)$,环境允许的极限人口为 N_m ,设 t 到 $t + \Delta t$ 时间内人口的增长量与 $N_m - N(t)$ 成正比,试建立模型、求解,并作图与 *Malthus* 模型、*Logistic* 模型的结果进行比较。

8. 一车间容积为 $10800m^3$,开始时空气中含有 0.12% 的 CO_2 ,为了保证工人的健康,用一台风量为 $1500m^3/min$ 的鼓风机通入含 CO_2 为 0.04% 的新鲜空气,假定新鲜空气与原空气能迅速混合均匀,并以相同风量排出,试求

(1) 车间 CO_2 含量所遵循的规律。

(2) 鼓风机开动 10 分种后,车间中含 CO_2 的百分比降低到多少。

9. 对于技术革新的推广,在下列几种情况下建立模型。

(1) 推广工作通过已经采用新技术的人进行,推广速度与已采用新技术人数成正比,推广是无限的。

(2) 设人数有限,因而推广速度随着尚未采用新技术人数的减少而降低。

(3) 在上述条件下还要考虑广告等媒介的传播作用。

10. 生活在阿拉斯加海滨的鲑鱼服从 *Malthus* 生物总数增长率

$$\frac{dp(t)}{dt} = 0.003p(t)$$

其中 t 按分计,在时间 $t = 0$,一群鲑鱼定居在这些水域,开始捕食

鲑鱼,鲨鱼捕杀鲑鱼的速率是 $0.001P^2(t)$, 其中 $P(t)$ 是 t 时刻鲑鱼总数。此外, 由于在它们周围出现意外情况, 平均每分钟有 0.002 条鲑鱼离开阿拉斯加水域。

(1) 考虑到这两种因素, 试修正 *Malthus* 生物总数增长律。

(2) 假设在 $t = 0$ 时存在一百万条鲑鱼, 试求鲑鱼总数, 当 $t \rightarrow \infty$ 时会发生什么情况。

第三章 差分方程方法建模

对于实际问题的数学描述,当变量被认为是离散地或间断地变化而不是连续地或瞬时地变化时,这些变化之间的关系可用差分方程来描述。例如,在经济与管理的实际问题中,许多重要的经济量,如收入、储蓄、消费等,通常为了方便而以相等的时间间隔来统计。经济学家就是通过分析这些定义在离散的时间变量上的经济变量来研究经济变化规律的。而差分方程模型在这些分析中起着非常重要的作用。

由于计算机的应用日益广泛,而其只能对离散数值进行计算,为了适应计算机求解需要,也须建立离散型的数学模型。建立系统的差分方程的方法大体有两种:一种是直接建立描述离散系统的差分方程;另一种是将连续变量作离散化处理,从而将连续型模型(微分方程)化为离散型模型(差分方程)。差分方程与微分方程在概念、理论及研究方法上有许多相似之处。本章通过实例介绍差分方程模型的建模方法。

§ 3.1 预备知识

3.1.1 差分的概念

定义 3.1.1 对于函数 $y(t)$ 及给定的常数 h (此常数 h 保证只要 t 属于 $y(t)$ 的定义域,则 $t+h$ 也属于 $y(t)$ 的定义域),因变量的改变量

$$y(t+h) - y(t)$$

称为 $y(t)$ 在 t 处的一阶差分(简称为差分)。记为 $\triangle y(t)$, 即

$$\Delta y(t) = y(t+h) - y(t)$$

其中 Δ 称为差分算子。

由定义可知

$$\Delta y(t+h) = y(t+2h) - y(t+h)$$

$$\Delta y(t+2h) = y(t+3h) - y(t+2h)$$

定义 3.1.2 对于给定的函数 $y(t)$ 及一阶差分 $\Delta y(t)$, $y(t)$ 的一阶差分在 t 处的差分称为 $y(t)$ 的二阶差分, 记为 $\Delta^2 y(t)$, 即

$$\begin{aligned}\Delta^2 y(t) &= \Delta(\Delta y(t)) = \Delta(y(t+h) - y(t)) \\ &= y(t+2h) - 2y(t+h) + y(t)\end{aligned}$$

类似地, $y(t)$ 的二阶差分在 t 处的差分称为 $y(t)$ 的三阶差分, 记为 $\Delta^3 y(t)$, 即

$$\Delta^3 y(t) = \Delta(\Delta^2 y(t))$$

更一般地, $y(t)$ 在 t 处的 n 阶差分定义为 $y(t)$ 的 $n-1$ 阶差分在 t 处的差分。记为 $\Delta^n y(t)$, 即

$$\Delta^n y(t) = \Delta(\Delta^{n-1} y(t)) \quad n = 2, 3, 4, \dots$$

对于差分运算, 容易证明如下性质。

定理 3.1.1 对于 n 个函数 $y_1(t), y_2(t), \dots, y_n(t)$ 及 n 个常数 c_1, c_2, \dots, c_n , 下式成立:

$$\begin{aligned}\Delta[c_1 y_1(t) + c_2 y_2(t) + \dots + c_n y_n(t)] \\ = [c_1 \Delta y_1(t) + c_2 \Delta y_2(t) + \dots + c_n \Delta y_n(t)]\end{aligned}$$

在我们考察的实际问题中的差分, 特别是经济学中的差分, 通常可以认为函数是定义在从 0 起始的连续整数上。为讨论方便, 我们用 k 来表示自变量, k 取 0 及在此之后的整数。并记 $y_k = y(k)$ 。这样

$$\Delta y(k) = y_{k+1} - y_k$$

$$\Delta^2 y(k) = \Delta(\Delta y(k)) = y_{k+2} - 2y_{k+1} + y_k$$

3.1.2 差分方程的基本概念

定义 3.1.3 含有未知函数 y_k 及 $y_{k+1}, \dots, y_{k+n}, \dots$ 中至少两个的函数方程称为(常)差分方程。

定义中的“常”是指未知函数只有一个自变量。本章只讨论常差分方程问题。简称为差分方程,甚至简称为方程。

例如:(1) $y_{k+1} - 3y_k + y_{k-1} = e^k$

$$(2) y_{k+1} = y_k^2$$

定义 3.1.4 差分方程中未知函数的下标之间的最大差数称为差分方程的阶。

例如:(1) 为二阶差分方程,(2) 为一阶差分方程。

定义 3.1.5 若某个函数代入差分方程后能使差分方程成为恒等式,则称此函数为差分方程的解。

由于差分方程的解是函数,将此函数代入方程,是经过差分运算使等式成立的,因此差分方程的解有无穷多个。

定义 3.1.6 若差分方程的解中所含相互独立的任意常数的个数与差分方程的阶数相同,则称此解为差分方程的通解。

例 1 $y_{t+1} - y_t = a$ (a 为已知常数)。

解: 因为 $y_t = at + c$ (c 是任意常数) 代入方程后, $y_{t+1} - y_t = [a(t+1) + c] - (at + c) = a$, 使两端恒等,且此解中含有一个任意常数,所以它是方程的通解。

为了确定差分方程的某一个特定的解,必须给出这个解所满足的条件,数学上称之为定解条件。初始条件是定解条件之一。如对于 n 阶差分方程

$$F(k, y_k, y_{k+1}, \dots, y_{k+n}) = 0$$

任意给定 n 个常数 A_0, A_1, \dots, A_{n-1} , 令

$$y_0 = A_0, y_1 = A_1, \dots, y_{n-1} = A_{n-1}$$

即为其初始条件。

定义 3.1.7 满足初始条件的解称为该差分方程的特解。

一般来讲,特解可通过初始条件的限制,从通解中确定任意常数而得到。

3.1.3 线性差分方程的基本定理

在自动控制,经济学等领域中用得最多的、最简单的一类方程是线性差分方程。从数学观点上看,这也是最容易处理的一类差分方程。下面给出线性差分方程的几个基本定理。

定义 3.1.8 若差分方程能够化成如下形式

$$y_{k+n} + a_1(k)y_{k+n-1} + \cdots + a_{n-1}(k)y_{k+1} + a_n(k)y_k = f(k) \quad (3.1.1)$$

则称该方程为线性差分方程。

其中 $a_1(k), a_2(k), \cdots, a_n(k), f(k)$ 为 k 的已知函数。

当 $f(k) \equiv 0$ 时,方程(3.1.1)变成

$$y_{k+n} + a_1(k)y_{k+n-1} + \cdots + a_{n-1}(k)y_{k+1} + a_n(k)y_k = 0 \quad (3.1.2)$$

称为线性齐次差分方程。

当 $f(k) \neq 0$ 时,称为线性非齐次差分方程。

如果方程(3.1.1)中的系数 $a_1(k), a_2(k), \cdots, a_n(k)$ 均为常数,则称为常系数线性差分方程。

下面给出方程(3.1.1)及(3.1.2)的通解结构定理。

定义 3.1.9 对于函数组 $y_1(k), y_2(k), \cdots, y_n(k)$,若存在不全为零的常数 c_1, c_2, \cdots, c_n ,使得

$$c_1 y_1(k) + c_2 y_2(k) + \cdots + c_n y_n(k) \equiv 0$$

成立,则称此函数组线性相关。若上式仅当 $c_1 = c_2 = \cdots = c_n = 0$ 时才成立,则称此函数组线性无关。

定理 3.1.2 若 $y_1(k), y_2(k), \cdots, y_n(k)$ 是方程(3.1.2)的 n 个线性无关解,则其通解可表示为

$$y(k) = c_1 y_1(k) + c_2 y_2(k) + \cdots + c_n y_n(k) \quad (3.1.3)$$

其中 c_1, c_2, \cdots, c_n 为 n 个任意常数。

定理 3.1.3 若 $\bar{y}(k)$ 是非齐次方程(3.1.1)的一个特解, $y_1(k), y_2(k), \dots, y_n(k)$ 是相应的齐次方程的 n 个线性无关解, 则方程(3.1.1)的通解可表示为

$$y(k) = \bar{y}(k) + c_1 y_1(k) + c_2 y_2(k) + \dots + c_n y_n(k) \quad (3.1.4)$$

如何求方程(3.1.2)的线性无关解以及方程(3.1.1)的特解的具体方法,将在后面讨论中给出。

3.1.4 一阶常系数线性差分方程

一阶常系数线性差分方程的一般形式为

$$y_{k+1} + ay_k = f(k) \quad (a \neq 0) \quad (3.1.5)$$

其中 a 为已知常数, $f(k)$ 是 k 的已知函数。

(3.1.5) 对应的齐次差分方程为

$$y_{k+1} + ay_k = 0 \quad (3.1.6)$$

1. 齐次差分方程的通解

假设在初始时刻(即 $t = 0$ 时), 函数 y_k 取一任意值 c , 由方程(3.1.6), 逐次计算得

$$\begin{aligned} y_1 &= -ay_0 = -ac \\ y_2 &= -ay_1 = (-a)^2 c \\ y_3 &= -ay_2 = (-a)^3 c \\ y_4 &= -ay_3 = (-a)^4 c \end{aligned}$$

于是, 我们猜想

$$y_k = c(-a)^k$$

即齐次方程(3.1.6)的通解为

$$y_k = c(-a)^k, c \text{ 为任意常数。}$$

2. 非齐次方程的通解

为求方程(3.1.5), 可采用迭代法求解。

由方程(3.1.5)得

$$y_1 = (-a)y_0 + f(0)$$

$$y_2 = (-a)y_1 + f(1) = (-a)^2 y_0 + (-a)f(0) + f(1)$$

根据数学归纳法可证

$$y_k = (-a)^k y_0 + (-a)^{k-1} f(0) + (-a)^{k-2} f(1) + \cdots + f(k-1) \quad (3.1.7)$$

其中 y_0 为任意常数。(3.1.7) 式即为方程(3.1.1) 的通解。

例 2 求差分方程 $y_{k+1} - \beta y_k = \alpha$ 的通解。

解: 根据公式(3.1.7), 当 $\beta \neq 1$ 时, 通解为

$$\begin{aligned} y_k &= \beta^k y_0 + \beta^{k-1} \alpha + \beta^{k-2} \alpha + \cdots + \beta \alpha + \alpha \\ &= \beta^k y_0 + \alpha \frac{\beta^k - 1}{\beta - 1} \\ &= (y_0 + \frac{\alpha}{\beta - 1}) \beta^k - \frac{\alpha}{\beta - 1} \\ &= c \beta^k - \frac{\alpha}{\beta - 1} \end{aligned}$$

当 $\beta = 1$ 时, 通解为

$$y_k = c + \alpha k$$

除了迭代法外, 根据 $f(k)$ 的特殊情形, 可用待定系数法确定非齐次方程(3.1.5) 的一个特解。

待定系数法的基本思想是: 若某类型的函数代入方程(3.1.5) 左端后不改变其函数类型, 而方程(3.1.5) 的右端恰好为同一类型的函数, 则就可以设方程(3.1.5) 的一个解为这种类型的、含有待定系数的一般形式的函数, 待定系数由恒等关系确定。

下面介绍几种特殊情况。

情形 1° $f(k)$ 为常数

若 $f(k) = b$, 此时方程(3.1.5) 变为

$$y_{k+1} + a y_k = b \quad (3.1.8)$$

设(3.1.8) 有特解 $\bar{y}_k = \mu$, μ 为待定常数。

代入(3.1.8) 得 $(1 + a)\mu = b$

若 $a \neq -1$, 则特解 $\bar{y}_k = \frac{b}{1+a}$

若 $a = -1$, 改设特解 $\bar{y}_k = \mu k$

代入(3.1.8)得 $\mu = b$, 特解 $\bar{y}_k = \mu k$

综上所述, 由定理 3.1.3, 方程(3.1.8)的通解为

$$y_k = \begin{cases} c(-a)^k + \frac{b}{1+a} & a \neq -1 \\ c + bk & a = -1 \end{cases} \quad (3.1.9)$$

情形 2° $f(k)$ 是指数函数

若 $f(k) = bd^k$, b, d 为给定的非零常数, 此时, 方程(3.1.5)

变为

$$y_{k+1} + ay_k = bd^k \quad (a \neq 0) \quad (3.1.10)$$

设(3.1.10)有特解 $\bar{y}_k = \mu d^k$, μ 为待定常数。代入(3.1.10), 得

$$[\mu(d+a) - b]d^k = 0$$

若 $d+a \neq 0$, 则 $\mu = \frac{b}{d+a}$, (3.1.10) 有特解

$$\bar{y}_k = \frac{b}{d+a}d^k$$

若 $d+a = 0$, 改设特解 $\bar{y}_k = \mu kd^k$ 代入(3.1.10), 得

$$\mu(d+a)kd^k + \mu d^{k+1} = bd^k$$

由 $d^k \neq 0$, 得 $\mu = \frac{b}{d}$, 于是(3.1.10)有特解

$$\bar{y}_k = bkd^{k-1}$$

综上所述, 同情形 1°, 方程(3.1.10)的通解为

$$y_k = \begin{cases} \frac{b}{a+d}d^k + c(-a)^k & a+d \neq 0 \\ bkd^{k-1} + c(-a)^k & a+d = 0 \end{cases} \quad (3.1.11)$$

情形 3° $f(k)$ 为 k 的多项式

为简便,考虑 $f(k) = b_0 + b_1 k$

其中 $b_0 + b_1$ 为给定常数,且 $b_1 \neq 0$,此时方程变为

$$y_{k+1} + ay_k = b_0 + b_1 k \quad (3.1.12)$$

设特解为 $y_k = \alpha + \beta k$, α, β 为待定常数。代入(3.1.12) 得

$$(1+a)\alpha + \beta + (1+a)\beta k = b_0 + b_1 k$$

上式对任何 k 值成立,则有

$$\begin{cases} (1+a)\alpha + \beta = b_0 \\ (1+a)\beta = b_1 \end{cases}$$

若 $a \neq -1$,则有 $\alpha = \frac{b_0}{1+a} - \frac{b_1}{(1+a)^2}$, $\beta = \frac{b_1}{1+a}$ 于是,(3.1.12) 的特解

$$\bar{y} = \frac{b_0}{1+a} - \frac{b_1}{(1+a)^2} + \frac{b_1}{1+a} k$$

若 $a = -1$,改设特解 $\bar{y} = k(\alpha + \beta k)$,代入(3.1.12),同上理由可求得特解

$$\bar{y} = (b_0 - \frac{1}{2}b_1)k + \frac{1}{2}b_1 k^2$$

对于 $f(k)$ 为 k 的高次多项式的情形,可类似地讨论,不过求解过程更为复杂。

情形 4° $f(k)$ 为正弦、余弦函数

若 $f(k) = b_1 \cos \omega k + b_2 \sin \omega k$,其中 b_1, b_2 和 ω 为已知常数,且 $\omega \neq 0, b_1^2 + b_2^2 \neq 0$,此时方程为

$$y_{k+1} + ay_k = b_1 \cos \omega k + b_2 \sin \omega k \quad (a \neq 0) \quad (3.1.13)$$

设特解 $\bar{y}_k = \alpha \cos \omega k + \beta \sin \omega k$, α, β 为待定常数。代入(3.1.13) 后得

$$\begin{aligned} & \alpha \cos \omega(k+1) + \beta \sin \omega(k+1) + a\alpha \cos \omega k + a\beta \sin \omega k \\ & = b_1 \cos \omega k + b_2 \sin \omega k \end{aligned}$$

利用三角公式有

$$(\alpha \cos \omega + \beta \sin \omega + a\alpha) \cos \omega k + (-\alpha \sin \omega + \beta \cos \omega + a\beta) \sin \omega k \\ = b_1 \cos \omega k + b_2 \sin \omega k$$

上式对任何 k 成立, 则

$$\begin{cases} (a + \cos \omega) \alpha + \sin \omega \cdot \beta = b_1 \\ -\sin \omega \cdot \alpha + (a + \cos \omega) \beta = b_2 \end{cases}$$

由于 $D = \begin{vmatrix} a + \cos \omega & \sin \omega \\ -\sin \omega & a + \cos \omega \end{vmatrix} = (a + \cos \omega)^2 + \sin^2 \omega$

若 $D \neq 0$, 则由上式可解出待定系数 α 和 β 。若 $D = 0$, 改设特解 $\bar{y}_k = k(\alpha \cos \omega k + \beta \sin \omega k)$ 。

使用上述类似方法即可求出待定系数 α 和 β 。

3.1.5 二阶常系数线性差分方程

二阶常系数线性差分方程的一般形式为

$$y_{k+2} + ay_{k+1} + by_k = g(k) \quad (b \neq 0) \quad (3.1.14)$$

对应于方程(3.1.14)的齐次方程为

$$y_{k+2} + ay_{k+1} + by_k = 0 \quad (3.1.15)$$

1. 齐次方程的通解

仿照一阶方程, 设方程(3.1.15)有指数型非零特解 $y_k = \lambda^k$, $\lambda \neq 0$ 为待定常数。将 y_k 代入方程(3.1.15), 得

$$\lambda^{k+2} + a\lambda^{k+1} + b\lambda^k = \lambda^k(\lambda^2 + a\lambda + b) = 0$$

消去非零因子 λ^k , 得

$$\lambda^2 + a\lambda + b = 0 \quad (3.1.16)$$

这就是 y_k 为(3.1.15)的特解的充分必要条件。称(3.1.16)为方程(3.1.15)的特征方程, 其解称为特征根。

方程(3.1.15)的解的性质, 依赖于特征根 $\lambda_{1,2}$ 。下面我们按(3.1.16)的判别式 $\Delta = a^2 - 4b$ 的符号进行讨论。

特征方程(3.1.15)的两个特征根是

$$\lambda_{1,2} = \frac{1}{2}[-a \pm (a^2 - 4b)^{\frac{1}{2}}]$$

情形 1° $\Delta > 0$

这时, λ_1 与 λ_2 为互不相同的实根。(3.1.14) 的两个特解为

$$y_1(k) = \lambda_1^k, \quad y_2(k) = \lambda_2^k$$

且 $y_1(k), y_2(k)$ 线性无关, 由定理 3.1.2, (3.1.15) 的通解为

$$y(k) = c_1 y_1(k) + c_2 y_2(k) \quad (3.1.17)$$

其中 c_1, c_2 为任意常数。

情形 2° $\Delta \equiv 0$

这时方程有重根 $\lambda_1 = \lambda_2 = \lambda = -\frac{a}{2}$, 特解为

$$y_1(k) = \left(-\frac{a}{2}\right)^k.$$

不难验证, $k\left(-\frac{a}{2}\right)^k$ 是方程(3.1.15) 的另一个特解。且特解 λ^k 与 $k\lambda^k$ 是线性无关的。

由定理 3.1.2, 方程(3.1.15) 的通解为

$$y(k) = (c_1 + c_2 k) \left(-\frac{a}{2}\right)^k \quad (3.1.18)$$

情形 3° $\Delta < 0$

这时 λ_1 与 λ_2 为一对共轭复根。

设 $\lambda_{1,2} = \alpha \pm i\beta = r(\cos\omega \pm i\sin\omega)$

$$\text{其中} \begin{cases} \alpha = r\cos\omega = -\frac{a}{2}, & \beta = r\sin\omega = \frac{1}{2}\sqrt{-\Delta} \\ r = \sqrt{\alpha^2 + \beta^2} = \sqrt{b}, & \operatorname{tg}\omega = \frac{\beta}{\alpha} \quad (\omega \in (0, \pi)) \end{cases}$$

直接验证可知

$$y_1(k) = r^k \cos\omega k, \quad y_2(k) = r^k \sin\omega k$$

是方程(3.1.15) 的两个线性无关的解。因此, 方程(3.1.15) 的通解为

$$y(k) = r^k (c_1 \cos\omega k + c_2 \sin\omega k) \quad (3.1.19)$$

$$\text{或} \quad y(k) = A r^k \cos(\omega k - \theta) \quad (3.1.19')$$

其中 c_1, c_2, A, θ 均为任意常数。

例 3 求差分方程 $y_{k+2} - 6y_{k+1} + 9y_k = 0$ 的通解。

解: 对应特征方程为

$$\lambda^2 - 6\lambda + 9 = 0$$

特征根为 $\lambda_1 = \lambda_2 = 3$, 因此, 原方程的通解为

$$y_k = (c_1 + c_2 k)3^k \quad (c_1, c_2 \text{ 为任意常数})$$

例 4 求差分方程 $y_{k+2} - 2y_{k+1} + 2y_k = 0$ 的通解。

解: 对应特征方程为

$$\lambda^2 - 2\lambda + 2 = 0$$

特征根为 $\lambda_1 = 1 + i, \lambda_2 = 1 - i, r = \sqrt{2}, \omega = \frac{\pi}{4}$ 。所以, 原差分方程的通解为

$$y_k = 2^{\frac{k}{2}} \left[c_1 \cos \frac{k\pi}{4} + c_2 \sin \frac{k\pi}{4} \right]$$

其中 c_1, c_2 为任意常数。

2. 非齐次方程的通解

由通解结构定理 3.1.3, 为求非齐次方程(3.1.13) 的通解, 关键是求它的一个特解。

与一阶差分方程类似, 我们只讨论 $g(k)$ 的一些特殊形式。在此我们以 $g(k) = B$ 为例(B 为非零常数), 说明怎样用待定系数法求方程(3.1.13) 的特解。对于

$$g(k) = Bd^k, d \neq 0, d \text{ 为常数}$$

$$g(k) = P_m(k), P_m(k) \text{ 为 } k \text{ 的 } m \text{ 次多项式}$$

$$g(k) = B_1 \cos \omega k + B_2 \sin \omega k, B_1, B_2, \omega \text{ 为常数}$$

等情形, 读者可仿照一阶方程, 自己练习。

对于二阶方程

$$y_{k+2} + ay_{k+1} + by_k = B \quad (B \neq 0) \quad (3.1.20)$$

设特解 $\bar{y}_k = \mu, \mu$ 为待定常数。代入(3.1.20), 得

$$\mu + a\mu + b\mu = B$$

若 $1 + a + b \neq 0$, 则特解 $\bar{y}_k = \frac{B}{1 + a + b}$

若 $1 + a + b = 0$, 改设特解 $\bar{y}_k = k\mu$, 代入(3.1.20) 得

$$\mu(k + 2) + a\mu(k + 1) + b\mu k = B$$

特解 $\bar{y}_k = \frac{Bk}{2 + a} \quad (a \neq -2),$

若 $1 + a + b = 0, 2 + a = 0$, 改设特解 $\bar{y}_k = k^2\mu$, 代入(3.1.20) 得

特解 $\bar{y}_k = \frac{1}{2}Bk^2$

说明:从上面求特解的过程中,我们已经看到,把所设的含有待定系数的解代入非齐次方程后,如果待定系数无解,这时可把所设的解乘上 k 后再试一遍。如果仍出现上述情况,就再乘上 k ,直到能把待定的系数确定下来。此方法是可行的,证明从略。

3.1.6 差分方程的稳定性简介

差分方程的平衡点及其稳定性与微分方程的有关概念是一致的。例如,对于一阶线性常系数差分方程

$$y_{k+1} + ay_k = b \quad (3.1.21)$$

其平衡点由 $y + ay = b$ 解得 $y^* = \frac{b}{1+a}$ 。当 $k \rightarrow \infty$ 时,若 $y_k \rightarrow y^*$, 则 y^* 是稳定的,否则是不稳定的。

容易看出,可以用变量替换方法将方程(3.1.21) 的平衡点稳定性问题转换为

$$y_{k+1} + ay_k = 0 \quad (3.1.22)$$

的平衡点 $y^* = 0$ 的稳定性问题。从而不失一般性,我们可对齐次差分方程讨论零解的稳定性问题。

对于一阶方程(3.1.22),其解为

$$y_k = (-a)^k A \quad (3.1.23)$$

由此可知,其零解稳定的充分必要条件为

$$|a| < 1 \quad (3.1.24)$$

对于二阶线性差分方程

$$y_{k+2} + ay_{k+1} + by_k = 0 \quad (3.1.25)$$

由 3.1.5 对二阶方程的讨论可知,其解的性质由特征方程

$$\lambda^2 + a\lambda + b = 0 \quad (3.1.26)$$

的特征根决定。特征根为

$$\lambda_{1,2} = \frac{1}{2}[-a \pm (a^2 - 4b)^{\frac{1}{2}}]$$

根据 $\Delta = a^2 - 4b$ 的符号及解的表示,容易求得二阶方程(3.1.25) 零解稳定的充分必要条件是下列三个条件之一成立。

(1) $a - 1 < b < \frac{1}{4}a^2$, 且 $0 \leq a < 2$ (此时,特征方程有两个互异实根),

(2) $\frac{1}{4}a^2 = b < 1$ (此时,特征方程有两个相等实根),

(3) $\frac{1}{4}a^2 < b < 1$ (此时,特征方程有一对共轭复根)。

对于 n 阶常系数线性差分方程,我们有

定理 3.1.4 对于常系数齐次差分方程

$$a_0 y_{k+n} + a_1 y_{k+n-1} + \cdots + a_{n-1} y_{k+1} + a_n y_k = 0 \quad (3.1.27)$$

其特征方程(由(3.1.27)的系数构成的 n 次代数方程)为

$$a_0 \lambda^n + a_1 \lambda^{n-1} + \cdots + a_{n-1} \lambda + a_n = 0 \quad (a_0 \neq 0) \quad (3.1.28)$$

设 $\lambda_1, \lambda_2, \cdots, \lambda_n$ 是特征方程(3.1.28)的 n 个根(实或复的,重根按重数计算),则方程(3.1.27) 零解稳定的充分必要条件是 n 个特征根的模

$$|\lambda_j| < 1 \quad (j = 1, 2, \cdots, n) \quad (3.1.29)$$

3.1.7 离散量与连续量的对应关系

设连续函数 $y = f(x)$, x 在其定义域 $[a, b]$ 内取 $n - 1$ 个分

点

$$a = x_0 < x_1 < x_2 < \cdots < x_{n-1} < x_n = b$$

相应的函数值为

$$y_0 = f(x_0), y_1 = f(x_1), \cdots y_n = f(x_n)$$

如进一步假定 x 的分点等距离, 即

$$\Delta x_i = x_i - x_{i-1} = \frac{b-a}{n} = h$$

有 $x_i = x_0 + ih, i = 1, 2, \cdots, n$ 。

(1) 导数与差商: 由导数的定义, 在 h 比较小时, 有

$$\left. \frac{dy}{dx} \right|_{x=x_i} \approx \frac{\Delta y_i}{h}$$

于是得到连续函数的导数与差商的对应关系

$$\frac{dy}{dx} \leftrightarrow \frac{\Delta y_i}{h} \quad (3.1.30)$$

其中 $\Delta y_i = y_{i+1} - y_i$ 为一阶向前差分; 一般地, 有

$$\frac{d^n y}{dx^n} \leftrightarrow \frac{\Delta^n y_i}{h^n} \quad (3.1.31)$$

其中 $\Delta^n y_i = \Delta^{n-1} y_{i+1} - \Delta^{n-1} y_i$ 为 n 阶向前差分。

类似地, 可建立导数与向后差分的关系。

(2) 定积分与数值求和: 由定积分定义有

$$\int_a^b f(x) dx = \lim_{n \rightarrow \infty} \sum_{i=0}^{n-1} f(x_i) \Delta x_i = \lim_{n \rightarrow \infty} \sum_{i=0}^{n-1} f_i h$$

其中 $f_i = f(x_i)$, 当 h 很小时, 有

$$\int_a^b f(x) dx \approx \sum_{i=0}^{n-1} f_i h$$

于是得到连续函数的定积分与数值求和之间的对应关系

$$\int_a^b f(x) dx \leftrightarrow \sum_{i=0}^{n-1} f_i h \quad (3.1.32)$$

综上所述, 连续量与离散量之间的对应关系为

$$\left\{ \begin{array}{l} x \leftrightarrow x_i \\ f(x) \leftrightarrow f(x_i) = y_i \\ \frac{dy}{dx} \leftrightarrow \frac{\Delta y_i}{h} \\ \frac{d^2 y}{dx^2} \leftrightarrow \frac{\Delta^2 y_i}{h^2} \\ \int_a^b f(x) dx \leftrightarrow \sum_{i=0}^{n-1} f(x_i) h \end{array} \right. \quad (3.1.33)$$

§ 3.2 建模实例

3.2.1 弦受集中荷载时的位移问题

一条紧张的弦在 n 个等距离点上施加了集中荷载 $f(k)$ ($k = 1, 2, \dots, n$), 如图 3.2.1。试求受力点的微小位移之间的关系式。

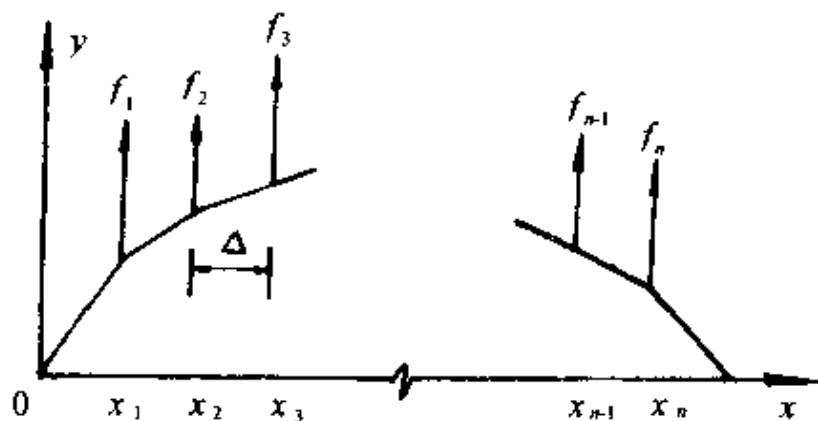


图 3.2.1

设 $y(k)$ 表示弦在第 k 个受力点的偏移量, 在弦中任取 3 点 $k-1, k, k+1$ 进行受力分析, 如图 3.2.2。假设: (1) 弦上的张力都相等, 记张力为 T 。(2) 弦的偏移很小, 记弦的偏移倾角为

$\alpha(k)$ 。

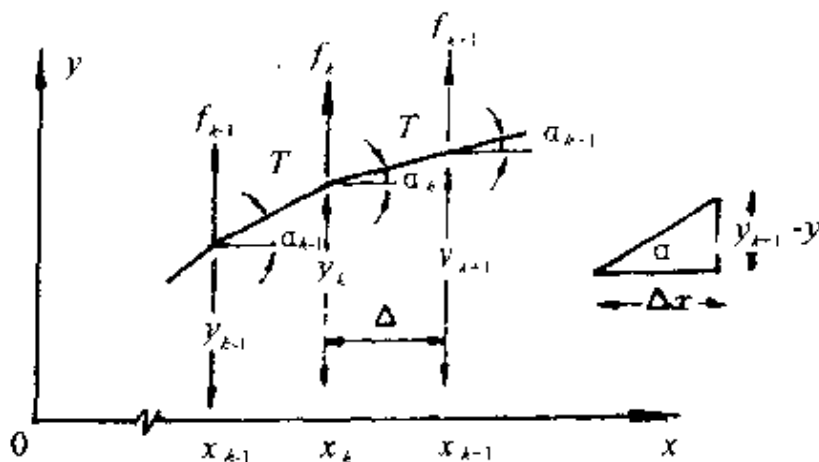


图 3.2.2

由假设, $\alpha(k)$ 是小量, 且有

$$\alpha(k) \approx \sin \alpha(k) \approx \operatorname{tg} \alpha(k) \quad (3.2.1)$$

受力点的距离相等, 设为 Δx , 则有

$$y(k+1) - y(k) = \Delta x \cdot \operatorname{tg} \alpha(k) \quad (3.2.2)$$

因为点是静止的, 所受到的外力合力为零, 所以有

$$T \sin \alpha(k) - T \sin \alpha(k-1) + f(k) = 0 \quad (3.2.3)$$

由(3.2.1) - (3.2.3), 容易得到

$$y(k+1) - 2y(k) + y(k-1) = -\frac{\Delta x}{T} f(k) \quad (k = 1, 2, \dots, n)$$

边界条件为

$$y(0) = 0, y(n+1) = 0$$

故该系统的模型为

$$\begin{cases} y(k+1) - 2y(k) + y(k-1) = -\frac{\Delta x}{T} f(k) \\ y(0) = 0, y(n+1) = 0 \end{cases} \quad (k = 1, 2, \dots, n) \quad (3.2.4)$$

3.2.2 生产和库存系统问题

在市场竞争中,对于一个既是生产型又是销售型的工厂,必须认真对待生产、销售、库存三者之间的关系,做到心中有数。下面给出生产和库存系统的数学模型。

设该工厂某种产品第 i 月的产量是 $u(i)$, 第 i 月的销售量是 $s(i)$, 第 i 月月初的库存量是 $x(i)$ 。那么该产品第 $i+1$ 月月初的库存量 $x(i+1)$ 等于第 i 月的库存量 $x(i)$ 加上第 i 月的生产量 $u(i)$, 再减去第 i 月的销售量 $s(i)$, 因此生产库存系统的数学模型为

$$x(i+1) = x(i) + u(i) - s(i) \quad (3.2.5)$$

3.2.3 银行复利问题

银行关于复利的规定有多种形式,现考虑一个按下述规则进行的储蓄账目:

- (1) 储蓄的年利率为 $r\%$;
- (2) 所付利息一年内复合 n 次,即把一年分为 n 个相等的时间段,而所付利息为每一年时间段的末尾;
- (3) 任何时候都可以存款,但存款利息只从下一时期开始计算,在时间段开始第一天的存款即开始计算利息。

现希望建立一个模型,这个模型可以预测在任意给定时间的帐目余额。

设 $y(k)$ 为第 k 期结束时的总存款, $u(k)$ 为第 k 期内的新存款。显然,第 k 期结束时的总存款等于第一时期余下的本利 $y(k-1)$, 由本利 $y(k-1)$ 得到的利息 $(\frac{r}{n})y(k-1)$ 与第 k 期内新存入的存款 $u(k)$ 之和,即

$$y(k) = y(k-1) + (\frac{r}{n})y(k-1) + u(k)$$

$$\text{或} \quad y(k) = (1 + \frac{r}{n})y(k-1) + u(k) \quad (3.2.6)$$

上式中 $n = 2$ 相当于半年的复利, 而 $n = 365$ 则是相应于逐日计算的复利。

3.2.4 换热器系列的最优设计问题

换热器系列的最优设计问题是要设计有 N 个换热器的系列, 使原来温度为 $a^\circ\text{C}$ 的油料, 连续流经 N 个换热器逐级加温后, 到离开第 N 个换热器时达到预定的温度 $b^\circ\text{C}$ 。如图 3.2.3 所示。问在保证油料由 $a^\circ\text{C}$ 升温到 $b^\circ\text{C}$ 的条件下, 各换热器的换热面积应如何分配, 方能使换热器系列的总面积最小?

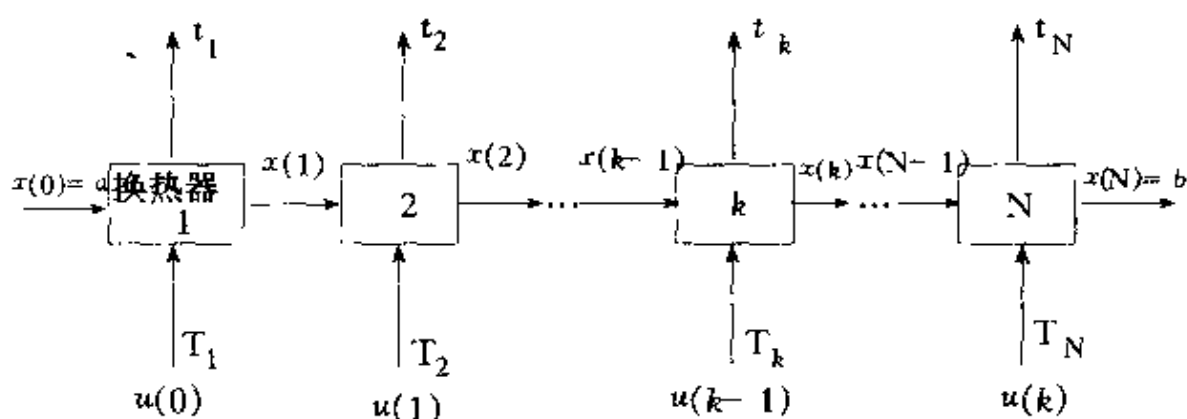


图 3.2.3

为解决上述问题, 我们首先需要建立数学模型。由于上述问题是一个需要从内部来观察和分析的系统, 我们可使用描述系统的方法——状态变量法, 建立状态变量方程模型。

设 a 表示油料初始温度,

b 表示加热后油料要达到的温度;

w 表示油料流量,

c_p 表示油料比热,

w_k 表示第 k 个换热器中热载体流量,

c_{pk} 表示第 k 个换热器中热载体比热,

q_k 表示第 k 个换热器的热传导率,

T_k 表示进入第 k 个换热器的热载体温度,
 t_k 表示流出第 k 个换热器的热载体温度,
 $x(k)$ 表示流出第 k 个换热器的油料温度,
 $u(k-1)$ 表示第 k 个换热器的换热面积。

为使模型简化,假定

$$w_k c_{pk} = w c_p \quad (k = 1, 2, \dots, N) \quad (3.2.7)$$

由热量平衡方程,有

$$w c_p [x(k) - x(k-1)] = w_k c_{pk} (T_k - t_k)$$

由(3.2.7),得

$$T_k - x(k) = t_k - x(k-1) \quad (3.2.8)$$

由第 k 个换热器的热传导方程,有

$$w c_p [x(k) - x(k-1)] = q_k u(k-1) \left(\frac{T_k + t_k}{2} - \frac{x(k) + x(k-1)}{2} \right)$$

应用(3.2.8),可得

$$w c_p (x(k) - x(k-1)) = q_k u(k-1) (T_k - x(k))$$

$$\text{即 } x(k) = \frac{x(k-1) + \frac{q_k}{w c_p} T_k u(k-1)}{1 + \frac{q_k}{w c_p} u(k-1)} \quad (k = 1, 2, \dots, N) \quad (3.2.9)$$

因此,如果将油料温度 $x(k)$ 作为状态变量,而将换热器换热面积

$u(k)$ 看作控制变量,并令 $Q(k) = \frac{q_k}{w c_p}$, 则状态变量的数学模型为

$$x(k+1) = \frac{x(k) + Q(k+1) T_{k+1} u(k)}{1 + Q(k+1) u(k)} \quad (k = 0, 1, \dots, N-1) \quad (3.2.10)$$

其边界条件是

$$x(0) = a, x(N) = b$$

3.2.5 市场价格的形成问题

人们在市场上常会感受到某种商品的价格能影响消费者对这种商品的需求量。例如,电视机降价,买的人就增多;化妆品涨价,用的人就减少,而粮食价格的变动,对需求量的影响很小。这样下一时期就会出现完全相反的局面。在没有外界干预的情况下,这种现象将如此循环下去。

在完全竞争的市场经济中上述现象是不可避免的。许多商品的**市场价格**、供给量和需求量会随着时间的变化而发生变化,呈现出时涨时跌、时增时减、交替变化的规律。经济学家常常用一幅平面网状图来描述这种变化规律。经济学家卡尔多(Kaldor)称之为蛛网定理。下面我们对**市场价格**建立模型,并对结果进行分析。

商品需求随该商品**市场价格**的变化而发生变化的规律,可用需求函数 $D = D(p)$ 来表示。其中 p 表示价格。

最简单的需求函数为

$$D = a + bp$$

其中 a, b 为常数。

某种商品的供给量随**市场价格**变化而变化的规律,可用供给函数 $S = S(p)$ 表示。最简单的供给函数为

$$S = a_1 + b_1 p$$

其中 a_1, b_1 为常数。

应说明的一点是, a, b, a_1, b_1 是反映商品特性的常数,需作长期的调查统计来确定。

在自由竞争的市场条件下,为了确定一种商品的**市场价格**,我们假设供需均衡。

模型 1 静态供需均衡模型(不依赖于时间的平衡问题)

如果不考虑需求量、供给量和价格与时间的关系,可有如下的静态供需均衡模型

$$\begin{cases} D = a + bp & (3.2.11) \\ S = a_1 + b_1 p & (3.2.12) \\ D = S & (3.2.13) \end{cases}$$

将(3.2.11), (3.2.12) 代入(3.2.13), 可解得价格的静态均衡值(称作均衡价格)

$$p_e = \frac{a - a_1}{b - b_1} \quad (3.2.14)$$

模型2 动态供需均衡模型

在实际价格形成过程中, 并不是静态的, 我们将需求量, 供给量和价格都看作时间 t 的函数, 因为生产某种商品, 既不是瞬时的, 也不是连续的, 而是要求有一个固定的生产周期(如粮食之类的农产品, 一年一收成), 所以我们将 t 取整数型离散值, 进一步假设

(1) 现期需求依赖于同期价格,

(2) 现期供给依赖于前期价格,

则传统的基本动态供需均衡模型为

$$\begin{cases} D_t = a + bp_t & (3.2.15) \\ S_t = a_1 + b_1 p_{t-1} & (3.2.16) \\ D_t = S_t & (3.2.17) \end{cases}$$

将(3.2.15), (3.2.16) 代入(3.2.17), 得供需均衡条件

$$a + bp_t = a_1 + b_1 p_{t-1}$$

由此得动态供需模型的等价差分方程

$$p_t - \frac{b_1}{b} p_{t-1} = \frac{a_1 - a}{b} \quad (3.2.18)$$

由 § 3.1 关于一阶差分方程的解法, 可解得通解

$$p_t = c \left(\frac{b_1}{b} \right)^t + p_e \quad (3.2.19)$$

其中 c 为任意常数, p_e 为静态均衡价格。

如果初始价格 p_0 已知, 则任意常数 $c = p_0 - p_e$, 方程(3.2.18) 的特解为

$$p_t = (p_0 - p_e)\left(\frac{b_1}{b}\right)^t + p_e \quad (3.2.20)$$

如果初始价格 $p_0 = p_e$, 那么 $p_t \equiv p_e$ 。这表明价格将固定在常数 p_e 上。即是模型 1 的情形。

但是, 实际的市场价格不会恰好等于均衡价格, 我们自然会关心如果初始价格 $p_0 \neq p_e$, 那么价格 p_t 将会如何变化呢?

对于(3.2.20), 容易看出(或根据 § 3.1), 价格 p_t (渐近) 稳定的充分必要条件为 $|b_1| < |b|$ 。

下面给出价格围绕均衡价格运动的图形表示。对于普通商品(所谓普通商品, 就是需求量随价格反向变化、供给量随价格同向变化的那些商品), $b < 0, b_1 > 0$ 。

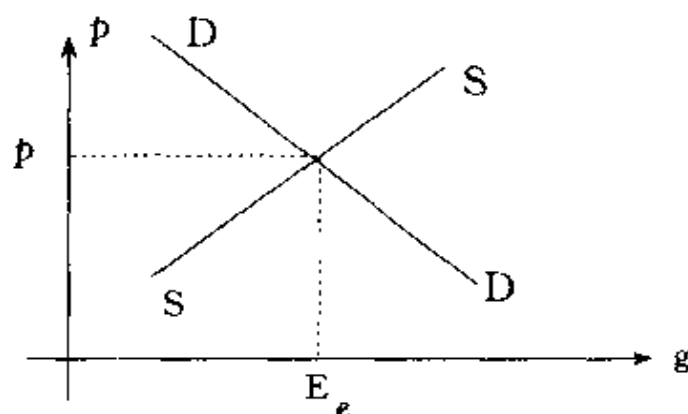


图 3.2.4

对于静态供需模型, 图 3.2.4 给出了供给量和需求量 q 与价格 p 的关系。

其中 DD 表示需求函数曲线, SS 表示供给函数曲线, p_e 表示均衡价格。 (p_e, q_e) 称为静态均衡点。

对于动态供需模型, 图 3.2.5 - 图 3.2.7 分别给出了供给量和

需求量 q_t 与价格 p_t 在 $b_1 < -b$, $b_1 = -b$, $b_1 > -b$ 三种情况下的关系。

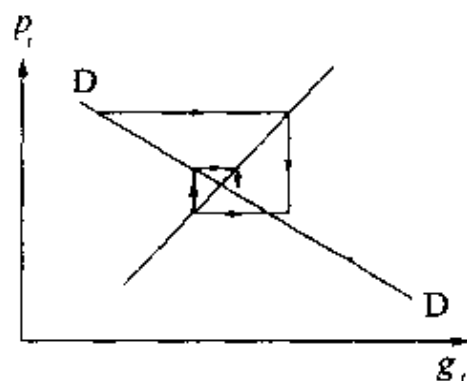


图 3.2.5

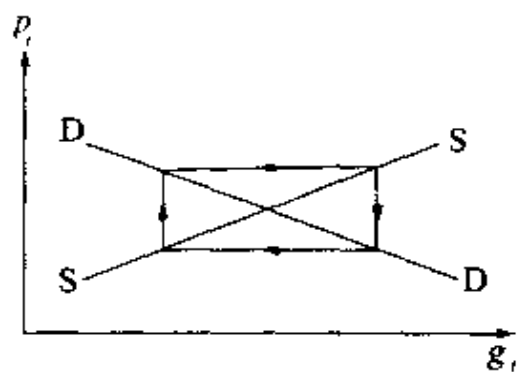


图 3.2.6

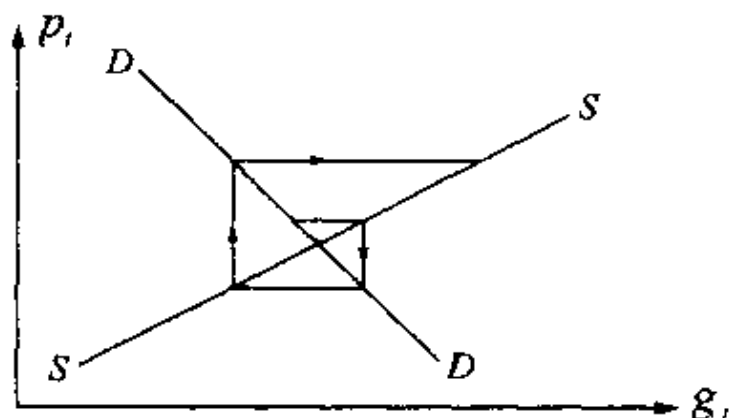


图 3.2.7

由图 3.2.5 也可看出,需求曲线越平,供应曲线越陡,越有利于经济稳定。

模型 3 引入期望价格和正常价格的模型

在传统的蛛网模型 2 中,假设现期供给仅依赖于前期价格,显然,这种假设是不合理的。因为生产者当商品一个周期接一个周期地相继生产时,会从经验中学习很多东西,使得生产者在“开始”

生产某种商品时,能提出产量实现后希望能够得到的价格。为此,可将模型 2 改为如下模型

$$\begin{cases} D_t = a + bp_t & (3.2.21) \end{cases}$$

$$\begin{cases} S_t = a_1 + b_1 \hat{p}_t & (3.2.22) \end{cases}$$

$$\begin{cases} D_t = S_t & (3.2.23) \end{cases}$$

其中 \hat{p}_t 表示 t 期的期望价格。显然,在传统的蛛网模型中,假设了 $\hat{p}_t = p_{t-1}$ 。

为确定期望价格,可以令

$$\hat{p}_t = p_{t-1} + c(p_N - p_{t-1}) \quad (3.2.24)$$

其中 p_N 为正常价格,即生产者认为在市场上迟早会获得的价格。正常价格最简单的方法是,取 $p_N = p_e$,即取静态均衡价格 p_e 为正常价格。

c 为反应系数,用来反映生产者对市场的估计和决策。 $c = 0$ 时, $\hat{p}_t = p_{t-1}$,这就是传统的蛛网模型 2。对于毫无经商知识的人,开始经商时多半抱有这种期望。

将(3.2.21), (3.2.22), (3.2.24) 代入(3.2.23),得供需均衡条件

$$bp_t = b_1(1 - c)p_{t-1} = a_1 + b_1cp_e - a$$

$$\text{或} \quad p_t - \frac{b_1(1 - c)}{b}p_{t-1} = \frac{a_1 + b_1cp_e - a}{b} \quad (3.2.25)$$

由 § 3.1 关于一阶差分方程的解法,可解得通解

$$p_t = A\left(\frac{b_1(1 - c)}{b}\right)^t + p_e \quad (3.2.26)$$

对于(3.2.26),根据 § 3.1 的讨论可知,价格 p_t (渐近)稳定的充分必要条件为 $\left|\frac{b_1(1 - c)}{b}\right| < 1$, 即 $\left|\frac{b_1}{b}\right| < \left|\frac{1}{1 - c}\right|$ 。

下面讨论反应系数 c 对 p_t 稳定性的影响。

情形 1° 当 $c > 0$ 时,称 \hat{p}_t 为内推期价格。当 $p_{t-1} < p_N$ 时,

有 $\hat{p}_t > p_{t-1}$, 反之, 当 $p_{t-1} > p_N$ 时, 有 $\hat{p}_t < p_{t-1}$, 特别地当 $c = 1$ 时, 有 $\hat{p}_t = p_N$ 。由此说明, 生产者认为, 当价格低于正常价格时, 下期价格将上涨; 当价格高于正常价格时, 下期价格将下跌。

当 $0 < c < 2$ 时, 由不等式

$$\left| \frac{b_1(1-c)}{b} \right| < \left| \frac{b_1}{b} \right| < 1$$

可知: 具有内推期望价格的供需均衡模型比传统的蛛网模型更稳定。

当 $c > 2$ 时, 由不等式

$$\left| \frac{b_1}{b} \right| < \left| \frac{b_1(1-c)}{b} \right| < 1$$

可知: 具有内推期望价格的供需均衡模型比传统的蛛网模型更不稳定。

情形 2° 当 $c < 0$ 时, 称 \hat{p}_t 为外推期望价格。当 $p_{t-1} < p_N$ 时, $\hat{p}_t < p_{t-1}$, 生产者认为当价格低于正常价格时, 下期价格会继续下跌; 反之类推。

当 $c < 0$ 时, 由不等式

$$\left| \frac{b_1}{b} \right| < \left| \frac{b_1(1-c)}{b} \right| < 1$$

可知: 外推期望供需均衡模型比传统的蛛网模型更不稳定。

由以上分析可知, 仅当反应系数 c 取不是太大的正数时, 期望模型才能改善传统蛛网模型的稳定性。

说明:

(1) 确定期望价格的方法有多种方式, 例如

$$\hat{p}_t = p_{t-1} + \rho(p_{t-1} + p_{t-2})$$

其中常数 ρ 表示期望系数, 用于反映生产者对价格变化的估计和态度。

当 $\rho > 0$ 时,生产者期望价格会同向变化,称为外推期望。

当 $\rho < 0$ 时,生产者期望价格会继续反向变化,称为内推期望。

当 $\rho = 0$ 时, $\hat{p}_t = p_{t-1}$, 传统的蛛网模型。

由此得到的改进模型(*Goodwin*) 为

$$\begin{cases} D_t = a + bp_t \\ S_t = a_1 + b_1 \hat{p}_t \\ \hat{p}_t = p_{t-1} + \rho(p_{t-1} - p_{t-2}) \\ D_t = S_t \end{cases}$$

利用 § 3.1 关于二阶差分方程的解法和稳定性分析,可类似讨论上述模型。

(2) 可利用差分方程讨论经济学中的其他问题。有兴趣的读者可参阅有关文献。

以上我们给出了几个差分方程建模的例子,对于有些实际问题(例如人口问题)既可建立连续模型,又可建立离散模型,究竟采用哪种模型可视建模目的而定。

习 题

1. 银行关于复利的规定为某一期的利息归入下期的本金。某人在银行存款 1000 元,复利率为 6%,问 k 期后本息合计多少元?

2. 设银行的半年复利是 3.6%,初始存款 1000 元,某人每半年存款数如下:476 元,355 元, - 217 元,727 元,求两年后的余额。

3. 试用差分形式建立类似于第二章中关于人口的 *Malthus* 模型和 *Logistic* 模型。

4. 已知 $D_t = 100 - ap_t$, $S_t = -20 + 3p_{t-1}$, 求价格的均衡值 p_e , 并讨论其稳定性。假设初始价格为 $p_0 = 25$, 计算 p_t 的值,直到

p_4 。

5. 对 *Goodwin* 模型确定价格 p_t 的运动规律, 并讨论 p_t 的稳定性。

6. 对于下列模型

$$\begin{cases} D_t = 100 - 2p_t \\ S_t = -20 + 3\hat{p}_t \\ \hat{p}_t = p_{t-1} + \frac{1}{3}(24 - p_{t-1}) \\ D_t = S_t \end{cases}$$

确定价格 p_t 的运动规律。其中 $c = \frac{1}{3}$ 为反应系数, $p_N = 24$ 为正常价格。

7. 与传统的蛛网模型的推广一段相类似, 如果价格是由生产者根据库存情况决定的, 即当库存积累时, 规定现时价格较以前低, 试建立拥有库存的市场模型, 并讨论价格变化的运动规律。

第四章 概率方法建模

§ 4.1 预备知识

4.1.1 概率的定义及性质

在自然界里,生产实践和科学试验中,人们观察到的现象大体可归为两类:确定性现象(必然现象)和偶然性现象(随机现象)。概率论就是研究偶然现象统计规律性的学科。为了叙述方便,我们针对典型的偶然现象——随机试验展开讨论。

定义 4.1.1 具有下列三个特性的试验称为随机试验:

- 1° 试验可以在相同条件下重复进行;
- 2° 各次试验的结果不一定相同,但能事先明确试验的所有可能结果;
- 3° 进行试验前,不能确知哪个结果发生。

定义 4.1.2 在随机试验中可能出现的结果称为随机事件(简称事件),常以大写字母 A 、 B 、 C 表示。每次试验必然发生的事件称为必然事件,记为 Ω ;必不发生的事件称为不可能事件,记为 Φ ;不可能再分的事件称为基本事件,我们常用仅含单点的集 $\{\omega\}$ 表示。

定义 4.1.3 在随机试验中,由所有基本事件对应的全部元素组成的集合称为样本空间,记为 Ω 。可见, Ω 的某些元素所构成集即为一个具有某特性的事件。

事件有如下的关系和运算,它们类同于集合的关系和运算。

关系:

1° 若两事件 A 与 B 不可能同时发生,称事件 A 与 B 互不相容。

2° 若事件 A 发生必然导致 B 发生,称事件 A 含于事件 B ,记为 $A \subset B$

若 $A \subset B$ 且 $B \subset A$,则称 A 与 B 相等(等价),记为 $A = B$ 。

3° 若 C 表示“事件 A 与 B 中至少有一个发生”这一事件,称为 A 与 B 的和(并),记为 $C = A \cup B$ 。

4° 若 D 表示“事件 A 与 B 同时发生”这一事件,称为 A 与 B 的积(交),记为 $D = A \cap B$ 。

5° 若 E 表示“事件 A 发生而事件 B 不发生”这一事件,称为 A 与 B 的差,记为 $E = A - B$ 。

6° Ω 与 A 之差这一事件称为 A 的逆事件(对立事件),记为 \bar{A} 。

事件的和与事件的积都可以推广到有限或可数无穷多个事件的情形。

运算:

1° 交换律: $A \cup B = B \cup A, A \cap B = B \cap A$

2° 结合律: $A \cup (B \cup C) = (A \cup B) \cup C$

$$A \cap (B \cap C) = (A \cap B) \cap C$$

3° 分配律: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

4° 对偶法则: $\overline{\bigcup_i A_i} = \bigcap_i \bar{A}_i$

$$\overline{\bigcap_i A_i} = \bigcup_i \bar{A}_i$$

5° 不变化: $A \cup B = A \cup (B - A \cap B)$

6° 包含差: $A - B = A - A \cap B$

7° $A - B = A \cap \bar{B}$

我们观察随机试验中的事件,主要是想了解其统计规律性,即它发生可能性的大小,我们把刻画事件发生可能性大小的数量指标称为事件的概率,事件 A 的概率以 $P(A)$ 表示,并且规定 $0 \leq P(A) \leq 1$ 。下面先针对几个典型问题给出定义概率和计算概率的各种方法,最后给出概率的一般定义及性质。

定义 4.1.4 若随机试验的全体基本事件 E_1, E_2, \dots, E_n 是有穷的,且每个基本事件发生的可能性相等,称其为古典型试验。对任事件 A ,定义其概率为

$$P(A) = \frac{\text{事件 } A \text{ 包含的基本事件个数}(k)}{\text{基本事件总数}(n)} \quad (4.1.1)$$

称作古典概率。

计算古典概率主要应用排列组合知识,尤其是加法原理和乘法原理。其关键是要正确区分不同的基本事件,这类问题在数学建模中是经常出现的。

定义 4.1.5 设在同一条件下进行了 n 次试验(n 足够大),事件 A 发生了 m 次,则事件 A 发生的频率 $f(A)$ 定义为

$$f(A) = \frac{A \text{ 出现的次数 } m}{\text{试验的总次数 } n} \quad (4.1.2)$$

用 $f(A)$ 作为事件 A 概率的一个度量,称为统计概率。

定义 4.1.6 若随机试验的样本空间为欧氏空间的某一区域,且样本点具有“均匀分布”的性质:它落入区域中任一子域的可能性相等,称为几何型试验。对于事件 $A, A \subset \Omega$ (样本空间),定义其概率为

$$P(A) = \frac{\mu(A)}{\mu(\Omega)} \quad (4.1.3)$$

这里 $\mu(A)$ 表示其度量(一维:长度,二维:面积,三维:体积,…)称 $P(A)$ 为几何概率。

不论是古典概率,统计概率,还是几何概率,它们都具有如下三个性质:

- 1° 对每一事件 $A, 0 \leq P(A) \leq 1$
- 2° 对必然事件 $\Omega, P(\Omega) = 1$
- 3° 设 $A_i (i = 1, 2, \dots)$ 是一列事件, $A_i \cap A_j = \Phi (i \neq j)$,

有

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

我们直接由上述三条出发, 给出概率的一般定义: 对应于一个随机试验的事件, 若有函数 P 满足上述三条, 则它为一个概率。由概率的定义, 它具有下列重要性质:

- 1° $P(\Phi) = 0$
- 2° 有限可加性: 设 $A_i (i = 1, 2, \dots, n)$ 是一列事件, $A_i \cap A_j = \Phi (i \neq j)$, 有

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i)$$

- 3° 单调性: 若 $A \supset B$, 则 $P(A) \geq P(B)$, 且 $P(A - B) = P(A) - P(B)$

- 4° 求逆: $P(\bar{A}) = 1 - P(A)$

- 5° 多除少补原理: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

- 6° 次可加性: $A_i (i = 1, 2, \dots)$ 为事件,

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} P(A_i)$$

性质 1° ~ 6° 在概率计算中具有重要作用。同时, 为了简化计算, 进一步地引入

定义 4.1.7 若 $P(A) > 0$, 定义条件概率

$$P(B | A) = \frac{P(A \cap B)}{P(A)}$$

为“已知事件 A 发生条件下, 事件 B 发生”的概率。

条件概率也是概率, 一般地 $P(B | A) \neq P(B)$, 我们可以通

过条件概率来简化概率的计算,得到下述有用的公式:

1° 乘法公式:设 $A_i (i = 1, 2, \dots, n)$ 为事件,且 $P(A_1 \cap \dots \cap A_{n-1}) > 0$, 则

$$\begin{aligned} P(A_1 \cap \dots \cap A_n) \\ = P(A_1) \cdot P(A_2 | A_1) \cdot P(A_3 | A_1 \cap A_2) \cdots P(A_n | \\ A_1 \cap \dots \cap A_{n-1}) \end{aligned} \quad (4.1.4)$$

作用:把一个复杂的事件分成几步,逐步针对“简单”事件求解概率,最后得到复杂事件的概率。

2° 全概率公式:设 $A_i (i = 1, 2, \dots)$ 为事件,满足: $\bigcup_i A_i = \Omega, A_i \cap A_j = \Phi (i \neq j), P(A_i) > 0$, 则对事件 B

$$P(B) = \sum_i P(B | A_i) P(A_i) \quad (4.1.5)$$

作用:把一个“多原因”事件 B 化分为若干个“单原因”事件求解其概率,最后得到“多原因”事件概率。

3° 贝叶斯公式:设 $A_i (i = 1, 2, \dots)$ 为事件,满足 $\bigcup_i A_i = \Omega, A_i \cap A_j = \Phi (i \neq j), P(A_i) > 0$, 若事件 B 满足 $P(B) > 0$, 则

$$P(A_i | B) = \frac{P(B | A_i) P(A_i)}{\sum_{j=1}^{\infty} P(B | A_j) P(A_j)} \quad (4.1.6)$$

作用:若把 $P(B | A_i)$ 称为“原因”概率,而 $P(A_i | B)$ 称“事后”概率,则“事后”概率可通过一系列“原因”概率求得,常用于“找原因”的模型。

注意到有事件 A, B 满足 $P(B | A) = P(B)$, 即 A 发生与否不影响 B 发生可能性大小,说明了它们之间具有“独立”性。

定义 4.1.8 设事件 A, B 满足

$$P(A \cap B) = P(A) \cdot P(B)$$

则称事件 A, B 相互独立。

若事件 A, B 独立,则 \bar{A} 与 B, A 与 \bar{B}, \bar{A} 与 \bar{B} 独立,且对于独

立事件 A, B 有下述求解和的概率公式:

$$P(A \cup B) = 1 - (1 - P(A))(1 - P(B)) \quad (4.1.7)$$

定义 4.1.9 设 $A_i (i = 1, 2, \dots, n)$ 为 n 个事件, 若任 $s (1 < s \leq n)$ 及 $i_k \quad k = 1, 2, \dots, s \quad 1 \leq i_1 < i_2 < \dots < i_s \leq n$ 有

$$P(A_{i_1} \cap \dots \cap A_{i_s}) = P(A_{i_1}) \cdot P(A_{i_2}) \cdots P(A_{i_s}) \quad (4.1.8)$$

称 A_1, \dots, A_n 相互独立。

注: 两两独立 $\nRightarrow n$ 个事件相互独立。

应用定义 4.1.9, 我们给出概率论中一类重要的试验——贝努里试验概型。

定义 4.1.10 进行 n 次重复的随机试验, 这些试验具有下述特点:

- 1° 每次条件一样, 可能的结果只有两个 A 及 \bar{A}
- 2° 各次试验结果互不影响, 即相互独立
- 3° $P(A) = p, P(\bar{A}) = 1 - p = q$

称为 n 次贝努里概型。

定理 4.1.1 对于贝努里概型, 事件 A 在 n 次试验中出现 k 次的概率为

$$P_n(k) = C_n^k p^k q^{n-k} \quad (0 \leq k \leq n) \quad (4.1.9)$$

且

$$\sum_{k=0}^n P_n(k) = 1 \quad (4.1.10)$$

贝努里概型不论在理论上, 还是在应用中都具有重要作用, 是概率论的核心内容之一。

4.1.2 随机变量及其分布

前一段, 我们讨论了事件及其概率, 但是由于样本空间千差万别, 一般不具有很好的结构, 因此就考虑把它与实数空间建立对应关系, 应用实数空间良好的分析性质统一研究实数空间上的事件与概率。

定义 4.1.11 设 Ω 是联系于随机试验的样本空间, 若对 $\omega \in \Omega$, 都有实数 $\xi = \xi(\omega)$ 与其对应, 且 $P(\omega; \xi(\omega) < x)$ 存在 (这里 $x \in R^1$ 为任实数), 称 ξ 为随机变量, 简记为 $r.v.$.

针对 ξ 的取值情况, 我们主要讨论两类重要的 $r.v.$ —— 离散型 $r.v.$ 及连续型 $r.v.$.

定义 4.1.12 若 $r.v. \xi$ 的可能取值只有有限或可数个, 则称 ξ 为离散型 $r.v.$.

若 ξ 的所有可能取值为 x_1, \dots, x_i, \dots , 相应地有 $p(\xi = x_i) = p_i (i = 1, 2, \dots)$ 则可有 ξ 的分布列:

$$\begin{array}{c|cccccc} \xi & x_1 & x_2 & \cdots & x_i & \cdots & \cdots \\ \hline P & p_1 & p_2 & \cdots & p_i & \cdots & \cdots \end{array} \quad (4.1.11)$$

分布列具有下述性质:

$$1^\circ \quad 0 \leq p_i \leq 1 \quad (i = 1, 2, \dots)$$

$$2^\circ \quad \sum_i p_i = 1$$

3° $\exists i_0$ 使 $p_{i_0} = \max_i p_i$, 称 $P(\xi = x_{i_0}) = p_{i_0}$ 所对应的 x_{i_0} 为 ξ 的最大可能值, p_{i_0} 为最大可能概率。

定义 4.1.13 若 $r.v. \xi$ 取值于任一区间 (a, b) 的概率可表为非负可积函数 $f(x)$, ($\int_{-\infty}^{\infty} f(x) dx < \infty$) 在 (a, b) 上积分:

$$P(a < \xi < b) = \int_a^b f(x) dx \quad (4.1.12)$$

称 ξ 为连续型 $r.v.$, $f(x)$ 称为 ξ 的分布密度函数。

对于连续型 $r.v. \xi$, $P(\xi = b) = 0$, b 为任实数。

分布密度函数应具有性质:

$$1^\circ \quad f(x) \geq 0$$

$$2^\circ \quad \int_{-\infty}^{\infty} f(x) dx = 1$$

对离散型 $r.v.$ 用分布列描述其概率分布, 对连续型 $r.v.$, 用分布密度函数描述, 为了统一对任一 $r.v. \xi$ 概率分布的描述, 引入

定义 4.1.14 设 $r.v. \xi$, 对 $x \in R^1$, 定义

$$F(x) = P(\xi < x) \quad (4.1.13)$$

作为 x 的函数, 称 $F(x)$ 为 ξ 的分布函数。

若 ξ 是离散型 $r.v.$, $P(\xi = x_i) = p_i$ 则

$$F(x) = \sum_{x_i < x} p_i \quad (4.1.14)$$

为一个阶梯函数。

若 ξ 是连续型 $r.v.$, 分布密度函数为 $f(x)$, 则

$$F(x) = \int_{-\infty}^x f(t) dt \quad (4.1.15)$$

为一个单调不降连续函数。

定理 4.1.2 设 $F(x)$ 为 $r.v. \xi$ 的分布函数, 则

1° $F(x)$ 单调不降;

2° $F(x)$ 左连续;

3° $\lim_{x \rightarrow -\infty} F(x) = 0, \lim_{x \rightarrow +\infty} F(x) = 1$

在实际中, 常遇到需同时用几个 $r.v.$ 才能更好描绘某一现象的情况, 为此引入多元 $r.v.$ 定义, 以下就二维情况进行讨论(高维类同)

定义 4.1.15 设 ξ, η 为二 $r.v.$ 称 (ξ, η) 为二维随机变量, 对 $x, y \in R^1$, 令

$$F(x, y) = P(\omega: \xi(\omega) < x, \eta(\omega) < y) \quad (4.1.16)$$

称 $F(x, y)$ 为 (ξ, η) 的联合分布函数。

定理 4.1.3 设 $F(x, y)$ 为 $r.v. (\xi, \eta)$ 的联合分布函数, 则

1° $F(x, y)$ 分别对每个变元单调不降

2° $F(x, y)$ 对每个变元左连续

$$3^{\circ} \quad \lim_{x \rightarrow -\infty} F(x, y) = 0, \quad \lim_{y \rightarrow -\infty} F(x, y) = 0$$

$$\lim_{\substack{x \rightarrow +\infty \\ y \rightarrow +\infty}} F(x, y) = 1$$

4° 任四实数 $a_1 \leq b_1, a_2 \leq b_2$ 有

$$F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2) \geq 0$$

二维 $r. v. (\xi, \eta)$ 也分最常用的两类:

1° 若 (ξ, η) 的所有可能取值为有限或可数个, 称 $r. v. (\xi, \eta)$ 为离散型的, 记 (ξ, η) 的可能取值为 $(x_i, y_j), P(\xi = x_i, \eta = y_j) = p_{ij} \quad (i = 1, 2, \dots \quad j = 1, 2, \dots)$, 称 p_{ij} 为 (ξ, η) 的联合概率函数, 满足

$$\begin{cases} p_{ij} \geq 0 \\ \sum_i \sum_j p_{ij} = 1 \end{cases} \quad (4.1.17)$$

其分布函数为

$$F(x, y) = \sum_{x_i < x} \sum_{y_j < y} p_{ij} \quad (4.1.18)$$

2° 若存在非负可积函数 $f(x, y) \left(\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) < \infty \right)$ 满足 $\forall x, y \in R^1$

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) du dv \quad (4.1.19)$$

称 $r. v. (\xi, \eta)$ 为二维连续型 $r. v.$, $f(x, y)$ 称为联合分布密度函数, 满足:

$$\begin{cases} f(x, y) \geq 0 \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1 \end{cases} \quad (4.1.20)$$

下面讨论二维 $r. v. (\xi, \eta)$ 的分布函数 $F(x, y)$ 与 ξ, η 的分布函数 $F_{\xi}(x), F_{\eta}(y)$ 之间的关系, 有:

$$\lim_{y \rightarrow \infty} F(x, y) = F_{\xi}(x)$$

$$\lim_{x \rightarrow \infty} F(x, y) = F_{\eta}(y)$$

若 (ξ, η) 是离散型 $r.v.$ p_{ij} 为其联合概率函数,则有:

$$p_{\xi} = \sum_j p_{ij}$$

$$p_{\eta} = \sum_i p_{ij}$$

这里 p_{ξ} 、 p_{η} 、分别为 ξ 、 η 的概率函数。

若 (ξ, η) 为连续型 $r.v.$, $f(x, y)$ 为其联合分布密度函数,则有

$$f_{\xi}(x) = \int_{-\infty}^{\infty} f(x, y) dy$$

$$f_{\eta}(y) = \int_{-\infty}^{\infty} f(x, y) dx$$

这里 $f_{\xi}(x)$ 、 $f_{\eta}(y)$ 分别为 ξ 、 η 的分布密度函数。

由此,若已知 (ξ, η) 的概率分布则可完全确定 ξ 、 η 的概率分布情况,但反之不然,即由 ξ 、 η 的分布不能推知 (ξ, η) 的分布,这主要由于 (ξ, η) 的分布还有赖于 ξ 、 η 结合的紧密程度。

如同事件的独立性,对 $r.v.$ 亦有独立的概念

定义 4.1.16 设 $F(x, y)$ 为 $r.v.$ (ξ, η) 的联合分布函数, $F_{\xi}(x)$ 、 $F_{\eta}(y)$ 分别为 ξ 、 η 的分布函数,若有

$$F(x, y) = F_{\xi}(x)F_{\eta}(y) \quad (4.1.21)$$

称 $r.v.$ ξ, η 相互独立。

若 ξ, η 为离散型 $r.v.$ 、 p_{ij} 、 p_{ξ} 、 p_{η} 分别表示 (ξ, η) 、 ξ 、 η 的概率函数,则

$$\xi, \eta \text{ 独立} \Leftrightarrow p_{ij} = p_{\xi}p_{\eta}$$

若 ξ, η 是连续型 $r.v.$ 、 $f(x, y)$ 、 $f_{\xi}(x)$ 、 $f_{\eta}(y)$ 分别为 (ξ, η) 、 ξ 、 η 的分布密度函数,则

$$\xi, \eta \text{ 独立} \Leftrightarrow f(x, y) = f_{\xi}(x)f_{\eta}(y)$$

对于相互独立 $r.v.$ 而言,若知道了它们各自的分布,则其联

合分布也就随之确定了。

4.1.3 随机变量的数字特征

分布函数反映了 $r.v.$ 的全部概率特征,但是有时我们不需要了解其所有性质,只需了解某些侧面,诸如平均值、差异程度等;有时我们又无法确知 $r.v.$ 的所有概率性质,而不得不退而求其次。对于 $r.v.$,最重要的数字特征有期望、方差、相关系数等。

定义 4.1.17 若 ξ 为离散型 $r.v.$, $P(\xi = x_i) = p_i$ 为其分布列 ($i = 1, 2, \dots$) 若

$$\sum_i |x_i| p_i < \infty$$

记

$$E\xi = \sum_i x_i p_i \quad (4.1.22)$$

称 $E\xi$ 为 ξ 的数学期望

若

$$\sum_i [x_i - E\xi]^2 p_i < \infty$$

记

$$D\xi = \sum_i [x_i - E\xi]^2 p_i \quad (4.1.23)$$

称 $D\xi$ 为 ξ 的方差

定义 4.1.18 若 ξ 是以 $f(x)$ 为密度的连续型 $r.v.$ 若

$$\int_{-\infty}^{\infty} |x| f(x) dx < \infty$$

记

$$E\xi = \int_{-\infty}^{\infty} x f(x) dx \quad (4.1.24)$$

称 $E\xi$ 为 ξ 的数学期望

若

$$\int_{-\infty}^{\infty} [x - E\xi]^2 f(x) dx < \infty$$

记

$$D\xi = \int_{-\infty}^{\infty} [x - E\xi]^2 f(x) dx \quad (4.1.25)$$

称 $D\xi$ 为 ξ 的方差

注:并不是所有 $r.v.$ 都存在期望和方差。

定理 4.1.4 设 $f(x)$ 为连续型 $r.v.$ ξ 的分布密度函数, $g(x)$ 为 R^1 上连续函数。若 $\int_{-\infty}^{\infty} |g(x)| f(x) dx < \infty$ 则 $g(\xi)$ 数学期望存在,且

$$Eg(\xi) = \int_{-\infty}^{\infty} g(x) f(x) dx \quad (4.1.24)$$

注:对离散 $r.v.$ ξ 亦有类似结果。

由定理 4.1.4 可给出 $r.v.$ ξ 期望、方差的性质:

- 1° 期望线性: $E(a\xi + b) = aE\xi + b, Eb = b$
- 2° 方差变换: $D(a\xi + b) = a^2 D\xi, Db = 0$
- 3° $D\xi = E\xi^2 - (E\xi)^2, E\xi^2 \geq (E\xi)^2$
- 4° $f(x) = E(x - \xi)^2$ 在 $x = E\xi$ 达到最小值 $D\xi$
- 5° 切比雪夫不等式:若 $r.v.$ ξ 的方差 $D\xi \exists$ 则 $\forall \epsilon > 0$

$$P(|\xi - E\xi| \geq \epsilon) \leq \frac{D\xi}{\epsilon^2} \quad (4.1.25)$$

- 6° 若 $r.v.$ ξ 方差 $D\xi > 0$, 则 $\exists r.v. \xi^*$

$$\xi^* = \frac{\xi - E\xi}{\sqrt{D\xi}}$$

$E\xi^* = 0, D\xi^* = 1, \xi^*$ 称为 ξ 的标准化。

定义 4.1.19 对 $r.v.$ ξ , 称 $v_k = E\xi^k, \alpha_k = E|\xi|^k$ 为 ξ 的 k 阶原点矩、 k 阶原点绝对矩。称 $\mu_k = E[\xi - E\xi]^k, \beta_k = E|\xi - E\xi|^k$ 为 ξ 的 k 阶中心矩、 k 阶中心绝对矩。

我们指出,原点矩可由中心矩求得,中心矩亦可由原点矩求得,矩在统计学中参数估计问题具有重要作用。

下面我们讨论多维 $r.v. (\xi_1, \dots, \xi_n)$ 数字特征这时除了考虑其各自的特征外, 还要考虑它们关联的紧密程度, 就连续型 $r.v. (\xi_1, \dots, \xi_n)$

定义 4.1.20 设 $f(x_1, \dots, x_n)$ 为 $r.v. (\xi_1, \dots, \xi_n)$ 的联合密度函数, ξ_k 的密度为 $f_k(x_k)$, (ξ_i, ξ_k) 的联合密度为 $f_{ik}(x_i, x_k)$, 设出现的所有积分都绝对收敛,

记

$$a_k = \int_{-\infty}^{\infty} x_k f_k(x_k) dx_k \quad (k = 1, 2, \dots, n) \quad (4.1.26)$$

称 (a_1, \dots, a_n) 为 n 维 $r.v. (\xi_1, \dots, \xi_n)$ 的数学期望,

记

$$\begin{aligned} b_{ik} &= E(\xi_i - E\xi_i)(\xi_k - E\xi_k) \\ &= \begin{cases} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_i - E\xi_i)(x_k - E\xi_k) f_{ik}(x_i, x_k) dx_i dx_k & i \neq k \\ \int_{-\infty}^{\infty} (x_i - E\xi_i)^2 f_i(x_i) dx_i & i = k (i, k = 1, 2, \dots, n) \end{cases} \end{aligned} \quad (4.1.27)$$

称 b_{ik} 为协方差, 方阵

$$\Sigma = \begin{pmatrix} b_{11} & b_{12} \cdots & b_{1n} \\ b_{21} & b_{22} \cdots & b_{2n} \\ \vdots & \vdots & \vdots \\ b_{n1} & b_{n2} \cdots & b_{nn} \end{pmatrix} \quad (4.1.28)$$

为 (ξ_1, \dots, ξ_n) 的协方差阵, 协方差阵具有性质:

$$1^\circ \quad b_{kk} = D\xi_k \quad (k = 1, 2, \dots, n)$$

$$2^\circ \quad b_{ik} = b_{ki}$$

$$3^\circ \quad b_{ik}^2 \leq D\xi_i D\xi_k$$

定理 4.1.5 设 $f(x_1, \dots, x_n)$ 为 (ξ_1, \dots, ξ_n) 的联合分布密度函数, $g(x_1, \dots, x_n)$ 为 n 元连续函数, 则

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} |g(x_1, \dots, x_n)| f(x_1, \dots, x_n) dx_1 \cdots dx_n < \infty$$

$\Leftrightarrow E g(\xi_1, \dots, \xi_n)$ 存在, 且当它存在时有

$$E g(\xi_1, \dots, \xi_n) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, \dots, x_n) f(x_1, \dots, x_n) dx_1 \cdots dx_n \quad (4.1.29)$$

由定理 4.1.5 我们可得到 n 维 $r.v. (\xi_1, \dots, \xi_n)$ 对应函数期望, 方差的若干性质:

$$1^\circ \quad E(\xi_1 + \cdots + \xi_n) = E\xi_1 + \cdots + E\xi_n \quad (4.1.30)$$

2° 若 ξ_1, \dots, ξ_n 独立, 则

$$E(\xi_1 \cdot \xi_2 \cdots \xi_n) = E\xi_1 \cdot E\xi_2 \cdots E\xi_n \quad (4.1.31)$$

$$3^\circ \quad D\left(\sum_{i=1}^n c_i \xi_i\right) = \sum_{i=1}^n c_i^2 D\xi_i + \sum_{\substack{i,k=1 \\ i \neq k}}^n c_i c_k b_{ik} \quad (4.1.32)$$

4° 若 ξ_1, \dots, ξ_n 独立, 则

$$D\left(\sum_{i=1}^n c_i \xi_i\right) = \sum_{i=1}^n c_i^2 D\xi_i \quad (4.1.33)$$

定义 4.1.21 设 $r.v. \xi_1, \xi_2$ 的方差 $D\xi_1, D\xi_2 \exists$ 且大于 0, 记

$$\rho_{12} = \frac{E(\xi_1 - E\xi_1)(\xi_2 - E\xi_2)}{\sqrt{D\xi_1} \sqrt{D\xi_2}} \quad (4.1.34)$$

称为 ξ_1, ξ_2 的相关系数。

相关系数是反映二 $r.v.$ 线性相关程度的量, 容易看出它是标准化了的二 $r.v.$ 的协方差, 因而不受单位及原点选取的影响, 设 ξ_1, ξ_2 的相关系数为 ρ , 则它具有性质

$$1^\circ \quad -1 \leq \rho \leq 1$$

2° $|\rho| = 1 \Leftrightarrow \xi_1, \xi_2$ 以概率 1 线性相关, 即存在常数 α, β 且 $\alpha \neq 0$ 使

$$\rho(\xi_2 = \alpha\xi_1 + \beta) = 1$$

若 $\rho = 0$ 则称 ξ_1, ξ_2 不相关, 它有性质:

$$1^\circ \quad \rho = 0 \Leftrightarrow E\xi_1 \cdot \xi_2 = E\xi_1 \cdot E\xi_2$$

$$2^\circ \quad \rho = 0 \Leftrightarrow D(\xi_1 + \xi_2) = D\xi_1 + D\xi_2$$

3° 若 ξ_1, ξ_2 独立则 $\rho = 0$, 反之不一定成立。

4.1.4 大数定律与中心极限定理

大数定律与中心极限定理是近代概率得以发展的基础,也是
我们进行统计推断的理论基础,因而具有重要理论意义,然而在它
们叙述中需要较多的数学概念,因而本书只给予直观解释而不做
数学推导。

1. 大数定律:反映了随机现象的统计规律性。

一般提法:设 $\{\xi_n\}$ 为一随机变量序列,数学期望 $E\xi_n$ 存在,令

$$\bar{\xi}_n = \frac{1}{n} \sum_{i=1}^n \xi_i \text{ 若 } \forall \varepsilon > 0$$

有

$$\lim_{n \rightarrow \infty} P(|\bar{\xi}_n - E\bar{\xi}_n| < \varepsilon) = 1$$

则称 $\{\xi_n\}$ 服从大数定律。

定理 4.1.6 设 $\{\xi_n\}$ 为相互独立同分布的随机变量序列,且

$$P(\xi_n = 1) = p, P(\xi_n = 0) = q = 1 - p$$

则 $\{\xi_n\}$ 服从大数定律,即 $\forall \varepsilon > 0$

$$\lim_{n \rightarrow \infty} P(|\bar{\xi}_n - E\bar{\xi}_n| < \varepsilon) = 1$$

注意到: $\bar{\xi}_n = \frac{1}{n} \sum_{i=1}^n \xi_i$ 反映的是 n 次独立试验中 A 发生的频率,
 $E\bar{\xi}_n = p$ 为 A 发生概率,即知试验次数足够多时,可用频率替代概
率。

2. 中心极限定理则反映了许多微小的、独立的随机因素的总
后果近似正态分布。

一般提法:设 $\{\xi_n\}$ 为相互独立,有有限期望和方差,且方差均
大于 0,令

$$E\xi_i = a_i \quad D\xi_i = \sigma_i^2$$

$$B_n^2 = \sum_{i=1}^n \sigma_i^2$$

$$Y_n = \frac{\sum_{i=1}^n (\xi_i - a_i)}{B_n}$$

若对任 $x \in (-\infty, \infty)$ 都有

$$\begin{aligned} & \lim_{n \rightarrow \infty} P(Y_n < x) \\ &= \lim_{n \rightarrow \infty} P\left(\frac{1}{B_n} \sum_{i=1}^n (\xi_i - a_i) < x\right) \\ &= \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \end{aligned}$$

则称序列 $\{\xi_n\}$ 服从中心极限定理。

4.1.5 建模中常用的几个分布

1. 两点分布

若 $r.v. \xi$ 分布列为

ξ	0	1
P	q	p

$0 \leq p, q \leq 1, p + q = 1$ 称 ξ 服从两点分布。

概型: 描述“非此即彼”的概率模型

$$E\xi = p, D\xi = pq.$$

2. 二项分布

若 $r.v. \xi$ 分布列为

ξ	0	1	...	k	...	n
P	q^n	$C_n^1 p q^{n-1}$		$C_n^k p^k q^{n-k}$...	p^n

$0 \leq p, q \leq 1, p + q = 1 \quad n > 0$ 为整数

称 ξ 服从二项分布, 记为 $b(n, p)$ 。

概型: 描述贝努里试验概型, 即“在 n 次重复贝努里试验中, 事件 $A: P(A) = p$ 发生次数”的概率模型。

特别 $n = 1$ 时, $b(1, p)$ 即为两点分布

$$E\xi = np, D\xi = npq$$

二项分布的泊松近似:

定理 4.1.7 若 $\lim_{n \rightarrow \infty} np_n = \lambda \geq 0$, 则

$$\lim_{n \rightarrow \infty} b(k; n, p_n) = \lim_{n \rightarrow \infty} C_n^k p_n^k (1 - p_n)^{n-k} = \frac{\lambda^k}{k!} e^{-\lambda} \quad (4.1.35)$$

特别若 $np_n = \lambda \geq 0 (n = 1, 2, \dots)$ 则(4.1.35) 仍成立。

由此可见, 若 n, p_n 满足定理 4.1.7 条件, 则当 n 充分大时, 有

$$b(k, n, p_n) = C_n^k p_n^k (1 - p_n)^{n-k} \approx \frac{\lambda^k}{k!} e^{-\lambda}$$

二项分布的正态近似:

定理 4.1.8 若 $r.v. \xi \sim b(n, p) (0 < p < 1)$ 即 $P(\xi_n = k)$

$$= C_n^k p^k q^{n-k} (k = 0, 1, 2, \dots, n), \text{ 则 } \lim_{n \rightarrow \infty} P(a < \frac{\xi_n - np}{\sqrt{npq}} < b) = \int_a^b$$

$$\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

也就是说, 若满足定理 4.1.8 条件, 则当 n 充分大时, 有

$$P(a < \frac{\xi_n - np}{\sqrt{npq}} < b) \approx \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

二项分布的最大可能出现次数:

若 $(n+1)p$ 为正整数, 则 $b(n, p)$ 最大可能次数为 $(n+1)p - 1$ 及 $(n+1)p$; 若 $(n+1)p$ 不为正整数, 则 $b(n, p)$ 最大可能次数为 $[(n+1)p]$ 。

3. 泊松分布

若 $r.v. \xi$ 分布列为

ξ	0	1	...	k	...
P	$e^{-\lambda}$	$\frac{\lambda}{1!} e^{-\lambda}$...	$\frac{\lambda^k}{k!} e^{-\lambda}$...

其中 $\lambda > 0$ 为常数

称 ξ 服从参数为 λ 的泊松分布, 记为 $p(\lambda)$ 。

概型:描述“大量随机试验中,稀有事件发生次数”的概率模型。

$$E\xi = D\xi = \lambda$$

泊松分布的最大可能出现次数:

若 $\lambda > 0$ 为正整数,则其最大可能次数为 $\lambda - 1$ 及 λ ,若 λ 不为正整数,则其最大可能次数为 $[\lambda]$ 。

4. 超几何分布

若 $r. v. \xi$ 分布列为

ξ	0	...	k	...	l
P	$\frac{C_M^0 C_{N-M}^n}{C_N^n}$...	$\frac{C_M^k C_{N-M}^{n-k}}{C_N^n}$...	$\frac{C_M^l C_{N-M}^{n-l}}{C_N^n}$

其中 $0 \leq n \leq N, 0 \leq M \leq N, l = \min(M, n)$ 为非负整数。称 ξ 服从超几何分布。

概型:描述“有 N 个产品,其中 M 个废品,从中任取 n 个,其中含有次品个数”的概率模型。

$$E\xi = \frac{nM}{N}, D\xi = \frac{nM(N-n)(N-M)}{N^2(N-1)}$$

超几何分布的二项近似:

定理 4.1.9 若有 $\lim_{N \rightarrow \infty} \frac{M}{N} = p$

则

$$\lim_{N \rightarrow \infty} \frac{C_M^k C_{N-M}^{n-k}}{C_N^n} = C_n^k p^k q^{n-k} \quad (4.1.37)$$

即在满足定理 4.1.9 条件下,当 N 充分大有

$$\frac{C_M^k C_{N-M}^{n-k}}{C_N^n} \approx C_n^k p^k q^{n-k}$$

5. 均匀分布

若 $r. v. \xi$ 分布密度函数为

$$f(x) = \begin{cases} \frac{1}{b-a} & x \in [a, b] \\ 0 & x \notin [a, b] \end{cases}$$

称 ξ 服从 $[a, b]$ 上的均匀分布。

概型:描述“ $r.v. \xi$ 取值 $[a, b]$ 之外的概率为 0, 而取值于 $[c, d] \subset [a, b]$ 的概率仅依赖于 $[c, d]$ 长度而与位置无关”的概率模型, 即具有“均匀性”概型。

$$E\xi = \frac{a+b}{2} \quad D\xi = \frac{(b-a)^2}{12}$$

6. 正态分布

若 $r.v. \xi$ 分布密度函数为

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}} \quad \sigma > 0, x \in R^1$$

称 ξ 服从参数为 a, σ 的正态分布, 记为 $\xi \sim N(a, \sigma)$

概型:描述“可看作许多微小的, 独立的随机因素的总后果”的概率模型。

$E\xi = a, D\xi = \sigma^2$, 可见 $N(a, \sigma)$ 中二参数具有典型意义: a 反映 ξ 的集中位置, σ^2 反映 ξ 的离散程度。

正态分布密度函数 $f(x)$ 的性质:

1° $f(x) > 0$, 具有各阶导数。

2° $f(x)$ 在 $(-\infty, a]$ 上 ↗, $[a, \infty)$ 上 ↘ 在 $x = a$ 达到极大值

$f(a) = \frac{1}{\sqrt{2\pi}\sigma}$, 即 ξ 的取值密集在 a 附近。

3° $f(x)$ 关于 $x = a$ 对称, 即 ξ 落入 $x < a - \sigma$ 与 $x > a + \sigma$ 的相应等长区间上概率相等。

4° 3σ 原则, $\xi \sim N(a, \sigma)$ 则

$$p(|\xi - a| < 3\sigma) = 0.9973 \quad (4.1.38)$$

即 $\xi \sim N(a, \sigma)$ 则有 99.73% 的质量聚集在 $(a - 3\sigma, a + 3\sigma)$ 内。从而, σ 越小, 质量聚集越紧密, 图形越陡。

$\xi \sim N(a, \sigma)$ 的标准化:

注意到 $E\xi = a, D\xi = \sigma^2$, 则 $\eta = \frac{\xi - a}{\sigma}$ 为其标准化, $E\eta = 0, D\eta = 1$, 从而 $\eta \sim N(0, 1)$ 其密度为

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

称 $N(0, 1)$ 为标准正态 *r. v.*, 它在理论与实际中占有重要地位。

r. v. $\xi \sim N(a, \sigma)$ 的概率 $P(x_1 < \xi < x_2)$ 计算。

1) 首先指出对于 $\eta \sim N(0, 1)$, 形如 $P(\eta < x) (x \geq 0)$ 的概率在任何一本概率统计教本中有表可查, 记

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

而对于 $P(\eta < x) (x < 0)$ 情形, 则利用对称性有

$$\begin{aligned} P(\eta < x) &= P(\eta > -x) \\ &= 1 - P(\eta < -x) \\ &= 1 - \Phi(-x) \quad (\text{因 } -x > 0, \Phi(-x) \text{ 有表可查}) \end{aligned}$$

2) 一般正态 *r. v.* $\xi \sim N(a, \sigma)$, 要先标准化后求值

$$\begin{aligned} P(\xi < x) &= P\left(\frac{\xi - a}{\sigma} < \frac{x - a}{\sigma}\right) \\ &= \Phi\left(\frac{x - a}{\sigma}\right) \end{aligned}$$

$$\begin{aligned} P(x_1 < \xi < x_2) &= P\left(\frac{x_1 - a}{\sigma} < \frac{\xi - a}{\sigma} < \frac{x_2 - a}{\sigma}\right) \\ &= \Phi\left(\frac{x_2 - a}{\sigma}\right) - \Phi\left(\frac{x_1 - a}{\sigma}\right) \end{aligned}$$

7. Γ 分布

若 *r. v.* ξ 的分布密度函数为

$$f(x) = \begin{cases} \frac{1}{\beta^{\alpha+1} \Gamma(\alpha+1)} x^{\alpha} e^{-\frac{x}{\beta}} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

其中 $\alpha > -1, \beta > 0$

则称 ξ 服从 Γ - 分布, 记为 $\xi \sim \Gamma(\alpha, \beta)$ 。

概型: 描述“水文统计, 最大风速、最大风压”的概率模型。

$$E\xi = \beta(\alpha + 1), D\xi = \beta^2(\alpha + 1)$$

特别若 $\alpha = 0, \beta = \frac{1}{\lambda}$ 称 $\Gamma(0, \frac{1}{\lambda})$ 为参数为 $\lambda (\lambda > 0)$ 的指数分布, 其密度为

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

指数分布主要用于研究寿命分布, 排队论等概率模型。 $E\xi = \frac{1}{\lambda}, D\xi = \frac{1}{\lambda^2}$ 。

8. 二维正态 $r. v.$

若 (ξ, η) 的分布密度函数为

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\lambda^2}} \exp\left\{-\frac{1}{2(1-\lambda^2)}\left[\frac{(x-m_1)^2}{\sigma_1^2} - 2\lambda\frac{(x-m_1)(y-m_2)}{\sigma_1\sigma_2} + \frac{(y-m_2)^2}{\sigma_2^2}\right]\right\}$$

其中 $\sigma_1, \sigma_2 > 0, m_1, m_2 \in R_1, |\lambda| < 1$

称 (ξ, η) 服从二维正态分布, 记为 $(\xi, \eta) \sim N(m_1, \sigma_1; m_2, \sigma_2; \lambda)$

二维正态分布的有关性质

1° 若 $(\xi, \eta) \sim N(m_1, \sigma_1; m_2, \sigma_2; \lambda)$ 则 $\xi \sim N(m_1, \sigma_1), \eta \sim N(m_2, \sigma_2)$

2° (ξ, η) 的期望为 (m_1, m_2) , 协方差阵为:

$$\begin{bmatrix} \sigma_1^2 & \lambda\sigma_1\sigma_2 \\ \lambda\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

3° 若 $(\xi, \eta) \sim N(m_1, \sigma_1; m_2, \sigma_2; \lambda)$, 则 λ 表示 ξ, η 的相关系数, 我们有

$\lambda = 0 \Leftrightarrow \xi, \eta$ 相互独立。

即对于正态分布不相关性与独立性等价。

4° (ξ, η) 服从二维正态 $\Leftrightarrow \xi, \eta$ 的任意线性组合 $C_0 + C_1\xi + C_2\eta$ 服从一维正态分布。

§ 4.2 建模实例

4.2.1 粒子运动问题

设有 n 个粒子,都以概率 $\frac{1}{N}$ 落于 $N(N \geq n)$ 个盒子中的每一个里。试问“某预先给定的 n 个盒中各含一个粒子”的概率 p 是多少?

由于这里粒子与盒子没有明确的规定,因而对它们的进一步假设的不同,就会导致问题的不同解答。由此可见“假设”在数学建模中的重要作用。

解法 1(麦克斯威尔——波尔兹曼)

假设: n 个粒子是不同的,即可辨别;每个盒子中能容纳的粒子数是没有限制的。

设 E 为随机试验:“把一个粒子放入 N 个盒子中之一”,于是每盒子对应于一基本事件,而 E 共有 N 个不同的基本事件。本问题中的试验为 $\tilde{E} = E^n$,即知它共有 N^n 个不同的基本事件。

今预先指定 n 个盒子,第一个粒子可分配到其中任一个,有几种分法,第二个可分配到余下的 $n-1$ 个盒子中去,有 $n-1$ 种分法……依次类推,因而由乘法原理,事件 A :“某预先给定的 n 个盒子中各含一个粒子”中含有 $n \cdot (n-1) \cdot \cdots \cdot 2 \cdot 1$ 种不同的基本事件,从而有

$$p = \frac{n!}{N^n}$$

解法 2(玻色——爱因斯坦)

假设： n 个粒子完全相同，因而不可辨别；每个盒子中能容纳的粒子数是没有限制的。

在解法 1 中，分布法不仅依赖于每个盒子中粒子个数，而且还依赖于那几个粒子；而现在则只依赖盒中粒子个数，因而计算基本事件个数的方法不同。

为清楚起见，就 $N = 3, n = 2$ 写出这二种解法中的全体基本事件。以 a, b 表这二粒子，记号 $(ab, -, -)$ 表第一盒中含二个粒子，其余皆空，其他记号解释类似，则在解 1° 中共有 $3^2 = 9$ 个基本事件：

- | | | |
|-----------------|-----------------|-----------------|
| 1. $(ab, -, -)$ | 2. $(-, ab, -)$ | 3. $(-, -, ab)$ |
| 4. $(a, b, -)$ | 5. $(b, a, -)$ | 6. $(a, -, b)$ |
| 7. $(b, -, a)$ | 8. $(-, a, b)$ | 9. $(-, b, a)$ |

但在解法 2 假设中，由于粒子不可辨别，则不区别 4, 5 而把它们看成相同基本事件，同样 6, 7 与 8, 9 也如此，故共有 6 个基本事件，而 A 则只含一个（设预先指定的是第一、二个盒子，则 A 中含的基本事件由 4, 5 表示）于是 $N = 3, n = 2$ 时 $p = \frac{1}{6}$ 。

一般情况的求解采用下列直观想法，把 N 个盒子排成一行， n 个粒子的一种分布法可表为下图

| * | * * | * * * * | - | ... | * * * |

即表第一盒中有 1 个粒子，第二盒中有 2 个，第三盒 4 个，第四盒空，…，最后一盒有 3 个，* 表粒子，“|”表盒壁。显见最外两端各有一壁。现把每个壁及每个粒子都看成一个位置，则其一种分配法相当于 $n + N - 1$ 个位置（两端不在内）被 n 个粒子占领的一种占位法。故共有 C_{n+N-1}^n 种分布法，即共有 C_{n+N-1}^n 个基本事件。而 A 中只含 1 个故

$$p = \frac{1}{C_{n+N-1}^n}$$

实际中确有许多粒子是无法辨别的，例如微观世界的基本粒

子(电子、光子等),解法 2 就是 1924 年物理学家玻色——爱因斯坦在研究玻色子时给出的。

解法 3(费米——迪拉克)

假设:粒子是不可辩认的,每个盒子至多只能容纳 1 个粒子。

此时任一种分布法都必须占用 n 个盒子,由于粒子不可辨别,分布法只依赖于各盒中的落点个数,而不依赖于落的是那几个点(粒子),因而共有 C_N^n 种分布即共有 C_N^n 个基本事件,而 A 只含其中 1 个,故

$$p = \frac{1}{C_N^n}$$

解法 3 是物理学家费米——迪拉克在 1925 年提出的,适于费米子的计算。

由此,针对数学建模问题选取正确的假设是成功的开端,只有在合理假设下才能很好地完成题目。

4.2.2 随机徘徊问题

设质点 M 在整数点集 $(0, 1, 2, \dots, a)$ 上作随机徘徊,就是说,每经一单位时间按下列规则改变一次位置:如果它现在在 z ($0 < z < a$) 上,下一步以概率 p ($0 < p < 1$) 转移到 $z + 1$,以概率 $q = 1 - p$ 转移到 $z - 1$ 。

1° 若它现在在 0(或 a),它就永远停留在 0(a)(称具有吸引壁的随机徘徊)。

2° 若它现在在 0,它就永远停留在 0;若它现在在 a ,则下一步以概率 1 转移到 $a - 1$, (称具有单侧反射壁的随机徘徊)

试求自 z 出发,终于要到达 0 的概率 q_z 。

为了在 1°、2° 求解问题,我们引入差分方程法,通过求解一个差分方程来得到该问题的解,这是这类问题的一个典型解法。

1° 质点 M 自 z ($0 < z < a$) 出发,由于 $p + q = 1$,即它下一步到 $z + 1$ 或 $z - 1$ 的概率为 1,由全概率公式有

$$q_z = pq_{z+1} + qq_{z-1} \quad (0 < z < a) \quad (4.2.1)$$

$$q_0 = 1, q_a = 0 \quad (4.2.2)$$

可见 (q_0, q_1, \dots, q_a) 为方程组 (4.2.1), (4.2.2) 的解, 下证其解惟一, 并具体求出此解, 由惟一性此解即为所求。

首先注意 (4.2.1) 有二特解

$$q_z = 1; q_z = \left(\frac{q}{p}\right)^z \quad (4.2.3)$$

由 (4.2.1) 的齐次性, 对任二常数 A, B

$$q_z = A + B\left(\frac{q}{p}\right)^z \quad (4.2.4)$$

也是 (4.2.1) 的解, 为使边值条件 (4.2.2) 成立, 有

$$A + B = 1 \quad A + B\left(\frac{q}{p}\right)^a = 0$$

由此求出 A, B 代入 (4.2.4) 有

$$q_z = \frac{\left(\frac{q}{p}\right)^a - \left(\frac{q}{p}\right)^z}{\left(\frac{q}{p}\right)^a - 1} \quad (p \neq q) \quad (4.2.5)$$

是 (4.2.1)、(4.2.2) 解, 为证其惟一性, 只证 (4.2.4) 为 (4.2.1) 解的一般形式。设 $\{q_z\}$ 为 (4.2.1) 的任一解, 在 (4.2.4) 中可选 A, B 使 $q_0 = q_0, q_1 = q_1$, 则 (4.2.1) 知 q_{z+1} 由 q_{z-1}, q_z 所惟一决定, 从而由 $q_0 = q_0$ 及 $q_1 = q_1$ 知 $q_i = q_i (i = 0, 1, 2, \dots)$

若 $p = q$, 则 (4.2.3) 有两重复特解。但它有如下二特解:

$$q_z = 1, q_z = z$$

因而仿上述知当 $p = q$ 时有

$$q_z = 1 - \frac{z}{a} \quad (4.2.6)$$

进一步地可求自 z 出发到 a 的概率 p_z , 由下述方程

$$p_z = pp_{z+1} + qp_{z-1} \quad (0 < z < a) \quad (4.2.7)$$

$$p_0 = 0, p_a = 1 \quad (4.2.8)$$

的解给出,用同样的方法可得

$$p_z = 1 - q_z \quad (0 \leq z \leq a) \quad (4.2.9)$$

即:带有双吸引壁的随机徘徊,迟早总要被边点吸引,不可能永远在内部徘徊。

2° 在此情形下,仍由全概率公式有

$$q_z = pq_{z+1} + qq_{z-1} \quad (0 < z < a) \quad (4.2.10)$$

$$q_0 = 1, q_a = q_{a-1} \quad (4.2.11)$$

用类似的方法解差分方程组(4.2.10)、(4.2.11),即有

$$q_z = 1 \quad (0 \leq z \leq a) \quad (4.2.12)$$

由此可见:质点在边界点 $0, a$ 上转移法则的改变对应于边值条件的改变。

注意:有时直观的概率意义可以帮助我们检验结果是否正确,容易想象, z 越近于 a , 质点被 0 吸引可能性就越小,故 q_z 为 z 的不增函数,(4.2.5)、(4.2.6) 恰为该性质。又若 $p = q, a = 2n, z = n$ 是 $(0, 2n)$ 中点,则易见 $q_n = \frac{1}{2}$,这正是(4.2.6)的结果。因而,在完成建模问题后,进行简单情形的验算、回代是必要的,有时会带来意想不到的好结果。

4.2.3 彩票问题

现有 50000 张 1 元的彩票出售,有头奖一个,奖金 10000 元;二奖一个,奖金 5000 元;三奖一个,奖金 2000 元;四奖三个,奖金各为 500 元。问购一张彩票的利润期望值是多少?

设 r, v, ξ 表示一张彩票可能获得的利润,则有

$$\xi(\text{头奖}) = 9999(\text{元})$$

$$\xi(\text{二奖}) = 4999(\text{元})$$

$$\xi(\text{三奖}) = 1999(\text{元})$$

$$\xi(\text{四奖}) = 499(\text{元})$$

$$\xi(\text{未中奖}) = -1(\text{元})$$

进一步假设全部 50000 张彩票中奖的机会相等, 获各种利润的概率为

$$\text{头奖: } \frac{1}{50000}$$

$$\text{二奖: } \frac{49999}{50000} \cdot \frac{1}{49999} = \frac{1}{50000}$$

$$\text{三奖: } \frac{49999}{50000} \cdot \frac{49998}{49999} \cdot \frac{1}{49998} = \frac{1}{50000}$$

$$\text{四奖: } \frac{49999}{50000} \cdot \frac{49998}{49999} \cdot \frac{49997}{49998} \cdot \frac{3}{49997} = \frac{3}{50000}$$

$$\text{未中奖: } \frac{49994}{50000}$$

值得注意的是: 一张彩票如在第 n 次被抽到, 则它在前 $n - 1$ 次都未被抽到。同时, 不论 n 为多少 ($1 \leq n \leq 50000$), 彩票在第 n 次被抽到的概率都为 $\frac{1}{50000}$ 即利润的概率分布如下

利润(元)	概率
9999	$\frac{1}{50000}$
4999	$\frac{1}{50000}$
1999	$\frac{1}{50000}$
499	$\frac{3}{50000}$
- 1	$\frac{49994}{50000}$

这样可得利润的期望值

$$E\xi = (9999 + 4999 + 1999 + 3 \times 499) \cdot \frac{1}{50000} - \frac{49994}{50000}$$

$$= -0.63(\text{元})$$

由此即说明,对于购买彩票者毕竟是不合算的,那么人们为何要购买彩票呢?答案在于效用函数的概念。购买一张彩票最大损失仅为1元,若购买者中了任何一种奖,利润都是非常巨大的。于是,当人们购买彩票时,所作的决策是出于对效用函数的主观估计而不是出于利润的概率结构。应该注意的是,效用函数因人而异,为了更好地理解效用函数的性质,人们必须去研究人类自身的特性,这归于深一步的决策论方面的知识。

4.2.4 企鵝后代个数问题

设企鵝生下 r 个蛋的概率是服从参数为 λ 的泊松分布的。而每个蛋能发育成企鵝的概率是 p , 且各蛋是否能发育成企鵝是彼此独立的。试求企鵝后代个数的期望值。

针对这一问题,直观的考虑就是要给出企鵝后代个数的分布。

令 ξ 表企鵝后代的个数,显见 ξ 取非负整值 $0, 1, 2, \dots, k, \dots$, 又设企鵝生下蛋的个数为 η , 则 η 亦是取非负整值的 $r. v.$, 且 $\eta \geq \xi$, 由题

$$P(\eta = r) = \frac{\lambda^r}{r!} e^{-\lambda} \quad (4.2.13)$$

则对于 $P(\xi = k)$, 我们应用全概率公式

$$\begin{aligned} P(\xi = k) &= \sum_{r=k}^{\infty} P(\xi = k, \eta = r) \\ &= \sum_{r=k}^{\infty} P(\xi = k | \eta = r) P(\eta = r) \end{aligned}$$

应用到(2.13)可见只需知道 $P(\xi = k | \eta = r)$ 即可求得 $P(\xi = k)$ 。注意到蛋是否发育成小企鵝是相互独立的, 且发育成小企鵝的概率为 p , 从而 $P(\xi = k | \eta = r)$ 实际上反映了有 r 个蛋, 每蛋发育成概率为 p 且发育独立, 恰好发育成 k 个企鵝的概率, 显见它是一个贝努里试验, 因而

$$P(\xi = k | \eta = r) = C_r^k p^k q^{r-k} \quad (4.2.14)$$

于是有

$$\begin{aligned} P(\xi = k) &= \sum_{r=k}^{\infty} C_r^k p^k q^{r-k} \cdot \frac{\lambda^r}{r!} e^{-\lambda} \\ &= \sum_{r=k}^{\infty} \frac{r!}{k!(r-k)!} p^k q^{r-k} \cdot \frac{\lambda^r}{r!} e^{-\lambda} \\ &= \sum_{r=0}^{\infty} \frac{p^k q^r}{k! r!} \cdot \lambda^{r+k} e^{-\lambda} \\ &= \frac{(\lambda p)^k}{k!} e^{-\lambda} \sum_{r=0}^{\infty} \frac{\lambda^r q^r}{r!} \\ &= \frac{(\lambda p)^k}{k!} e^{-\lambda} e^{\lambda q} \\ &= \frac{(\lambda p)^k}{k!} e^{-\lambda p} \end{aligned} \quad (4.2.15)$$

即知 $\xi \sim P(\lambda p)$, 从而知 $E\xi = \lambda p$, 企鹅后代的个数期望值为 λp 。这是一个很自然的结果, 它说明了后代数与蛋的个数及发育成功概率成正比。

4.2.5 花店老板的进货问题

某花店对鲜花需求量的概率分布如下:

商品数量(百打)	0	1	2	3	4	5	6	7	8
概率	0.05	0.1	0.1	0.25	0.2	0.15	0.05	0.05	0.05

每出售百打鲜花, 老板可获利 250 元, 若当天不能出售, 每百打老板将损失 325 元, 在这种价格和需求结构下, 要求确定店主每天进多少百打鲜花方能获取最大利润?

这属于一个存贮问题, 我们首先根据这一简单模型应用枚举法求解, 然后引入边际分析的方法进行讨论, 最后给出这类存贮问题的一般解法。

1. 枚举法

假设不满足需求时, 对店主没有任何损失。由需求量的概率分

布中,我们可以计算出需求量的期望值为 3.65 百打。由此可以排除进货量 ≤ 1 或 ≥ 5 的可能性,因这样做不如进货 2,3,4 件那样有利可图(若从这三个数字还不能得到最优结果,我们不妨也可检查其余数字)。

为此我们定义纯利利润为出售的获利中减去因未出售而受的损失,负的利润即表示损失的值。

由表 4.2.1 显见,每天进 3 百打可获最大利润。

表 4.2.1

需求量 (百打)	概率	三种数量的纯利润		
		2	3	4
0	0.05	- 650	- 975	- 1300
1	0.10	- 75	- 400	- 725
2	0.10	500	175	- 150
3	0.25	500	750	425
4	0.20	500	750	1000
5	0.15	500	750	1000
6	0.05	500	750	1000
7	0.05	500	750	1000
8	0.05	500	750	1000
利润期望值		385	491.25	453.75

2. 边际分析

边际分析就是人们将利用价格结构方面的信息来检查和判断再多增加一个单位的需求量是否合算。

假设已知持有 n 件商品合理,要求通过分析再增加一件商品所增加的利润和导致的损失,来判断持有 $n + 1$ 件商品是否合算。最后增加的单位称边际单位。我们亦定义边际利润(MP)为由增

加的那件商品所得的纯利润。类似地可定义边际损失(ML)为由增加的那件商品所导致的损失。最后增加那件商品出售的概率为

$$p = P(\text{需求量} \geq n + 1) \quad (4.2.16)$$

它未被出售的概率为 $1 - p$ 。如果边际利润的期望值大于边际损失的期望值,就断定持有 $n + 1$ 件商品是合理的。注意到这时期望值即为价值与相应的概率的乘积。因而我们可把上述条件改写成

$$p(MP) \geq (1 - p)(ML) = (ML) - p(ML)$$

整理后,可得

$$p \geq \frac{ML}{ML + MP} \quad (4.2.17)$$

由(4.2.16)注意到,概率 p 是由需求量的概率分布的右尾部给出的(1 减去累计分布函数)。

有了上述准备,我们再回到我们的花店,有 $MP = 250$, $ML = 325$ 于是由(4.2.17)

$$p \geq \frac{325}{575} = 0.5652$$

即仅当出售概率 ≥ 0.5652 时,老板再增加 1 百打才是合算的。

从需求量的概率分布中,我们有

$$P(\text{需求量} \geq 3) = 0.75$$

$$P(\text{需求量} \geq 4) = 0.50$$

说明第 4 个商品出售的概率小于 0.5652,即进货量最好是 3 百打。

最后,我们给出边际分析法的步骤:

- 1) 确定边际损失
- 2) 确定边际利润
- 3) 计算比值 $p = ML / (ML + MP)$
- 4) 从概率分布上界开始依次逐个计算累计概率,当累计概率超过 p 时就停止,相应的需求量就是应持有的最优数量。

3. 一般解法

假设每天购进量为 n ,因为需求量为 $r.v. \xi$ 可能大于 n 、小于

n 、等于 n ,从而老板的收入也随机。因而作为优化的目标函数,不能是其每天的收入,而应是其日平均收入即每天收入的期望值

记花店老板每天购进 n 百打花的平均收入为 $G(n)$,需求量为 r 的概率为 $f(r)$,有

$$G(n) = \sum_{r=0}^n [250r - 325(n-r)]f(r) + \sum_{r=n+1}^{\infty} 250nf(r) \quad (4.2.18)$$

则我们的问题归为求 n ,使 $G(n)$ 最大

把问题中的数值代入,可见当 $n = 3$ (百打)时可使 $G(3)$ 达最大,即每天进 3 百打花可使花店老板获取最大利润。

4.2.6 病人排队时间问题

某医院手术室根据病人来诊和完成手术时间的记录,任意抽取 100 个工作小时,每小时来就诊的病人数 n 的出现次数如表 4.2.2。又任抽取 100 个完成手术的病历,所用时间 v (小时)出现次数如表 4.2.3,试就医院的繁忙度、病人的等待时间等作出评价。

表 4.2.2

到达病人数 n	出现次数 f_n
0	10
1	28
2	29
3	16
4	10
5	6
6 以上	1
合计	100

表 4.2.3

为病人完成手术 时间 v (小时)	出现次数 f_v
0.0 ~ 0.2	38
0.2 ~ 0.4	25
0.4 ~ 0.6	17
0.6 ~ 0.8	9
0.8 ~ 1.0	6
1.0 ~ 1.2	5
1.2 以上	0
合计	100

该问题可以当作一个标准的单服务台排队模型 $M/M/1$ 处

理,即假设病人到达的规律服从参数为 λ 的泊松分布,服务时间服从参数为 μ 的指数分布,顾客(病人)单个到来,相互独立,只有一个服务台(手术室),先到先服务。

很自然地假设 $\rho = \frac{\lambda}{\mu} < 1$ (否则队列将无限长) 令 $P_n(t)$ 表 t 时刻系统中有 n 个病人的概率,为简单令 $P_n(t) \equiv P_n$ 不依赖于 t , 则有

$$\begin{cases} \lambda P_{n-1} + \mu P_{n+1} - (\lambda + \mu) P_n = 0 & n \geq 1 \\ -\lambda P_0 + \mu P_1 = 0 \end{cases} \quad (4.2.19)$$

解差分方程(4.2.19) 有

$$\begin{cases} P_0 = 1 - \rho & \rho < 1 \\ P_n = (1 - \rho) \rho^n & n \geq 1 \end{cases} \quad (4.2.20)$$

由此,我们得到以下结果:

(1) 系统中平均病人数(队长的期望值)

$$\begin{aligned} L_s &= \sum_{n=0}^{\infty} n P_n \\ &= \sum_{n=1}^{\infty} n (1 - \rho) \rho^n \\ &= \frac{\rho}{1 - \rho} \quad 0 < \rho < 1 \end{aligned} \quad (4.1.21)$$

(2) 在队列中等待的平均病人数(队列长期期望值)

$$L_q = \sum_{n=1}^{\infty} (n - 1) P_n = L_s - \rho = \frac{\rho \lambda}{\mu - \lambda} \quad (4.2.22)$$

病人在系统中的逗留时间 W 为 *r. v.*, 在 $M/M/1$ 情形,它服从参数为 $\mu - \lambda$ 的指数分布。故

(3) 系统中病人逗留时间的期望值

$$W_s = EW = \frac{1}{\mu - \lambda} \quad (4.2.23)$$

(4) 在队列中顾客等待时间的期望值

$$W_q = W_s - \frac{1}{\mu} = \frac{\rho}{\mu - \lambda} \quad (4.2.24)$$

现在回到我们所给的问题,首先要确定系统的两参数 λ 、 μ ,注意到:

$$\text{每小时病人平均到达率} = \frac{\sum n f_n}{100} = 2.1 (\text{人/小时})$$

$$\text{每次手术平均时间} = \frac{\sum v f_v}{100} = 0.4 (\text{小时/人})$$

$$\text{每小时完成手术人数(平均服务率)} = \frac{1}{0.4} = 2.5 (\text{人/小时})$$

于是可知该系统可看作到达规律服从参数为 $\lambda = 2.1$ 的泊松分布,服务时间服从参数为 $\mu = 2.5$ 的指数分布的 $M/M/1$ 模型。于是有:

(1) $\rho = \frac{\lambda}{\mu} = 0.84$, 即说明手术室有84%的时间是繁忙(被利用),16%时间是空闲的。

(2) 病人在病房中人数(期望)

$$L_s = \frac{0.84}{1 - 0.84} = 5.25 (\text{人})$$

(3) 排队等待的病人数(期望)

$$L_q = 0.84 \cdot 5.25 = 4.41 (\text{人})$$

(4) 病人在病房中逗留时间(期望)

$$W_s = \frac{1}{2.5 - 2.1} = 2.5 (\text{小时})$$

(5) 病人排队等待时间(期望)

$$W_q = \frac{0.84}{2.5 - 2.1} = 2.1 (\text{小时})$$

注:排队论的模型在数学建模中占有重要地位,我们仅讨论了最简单的 $M/M/1$ 情形,对于多服务台的情形 $M/M/s$,队长有限制的情形以及任意分布的情形 $M/G/s$ 可以参阅有关排队论的书

籍,例如徐光辉著《随机服务系统》等。

4.2.7 一个更换问题

某办公楼有 1000 只电灯泡,它们都是办公楼工作时所必需的,已知灯泡损坏的百分比如下(灯泡寿命单位:月):

5	在第一个月月底前损坏
20	在第二个月月底前损坏
45	在第三个月月底前损坏
70	在第四个月月底前损坏
85	在第五个月月底前损坏
95	在第六个月月底前损坏
100	在第七个月月底前损坏

又设灯泡单个更换时,每只费用为 1.50 元,但当 1000 只灯泡一起更换时每只平均费用为 0.40 元,试给出一个最经济的更换方案。

由题设,首先我们可得到灯泡的寿命 ξ 的分布:

月数	1	2	3	4	5	6	7
概率	0.05	0.15	0.25	0.25	0.15	0.10	0.05

即知其平均寿命为

$$E\xi = 3.8(\text{月})$$

于是得到灯泡的损坏率为 $1/3.8 = 0.263$,即 1000 只灯泡平均每月约需更换 263 只。单个更换平均每月费用为 394.50 元。

现考虑一种更换策略,每过 $M(M = 1, 2, \dots)$ 个月更换全部灯泡,同时在两次总更换之间所有损坏的灯泡也单个更换,这种策略的费用分为两类:

- a 每过 M 个月更换一次的总更换费用为 400 元;
- b 在两次总更换之间的更换费用。

对于第二类费用,需了解在问题期间单个更换灯泡的平均数,这些数字由表 4.2.4 中得到。表内方框中数字即为每月更换的数字,它为表中相应列元素之和。表内各元素是这样确定的:

开始时,1000 只灯泡都更新了,这些灯泡的损坏情况是按前面所给的寿命分布给定的,填在第一列中 1000 这个数字之下而且在相应月份旁边的数字就是原有灯泡分别在第一、第二、… ,第七个月中更换的平均数。从第二个月起,更换灯泡时也要考虑第一个月中更新的 50 个灯泡,它们的损失情况亦应符合原有的寿命分布。用相应的概率乘 50 即得到第二、第三、… ,第八个月更换这种灯泡的平均数,即在第二列 50 之下的那些数字。在第二个月更换灯泡的总数为 152.5(150 个是更换原有的 1000 个中的灯泡,而 2.5 个是更换第一个月所更换过的 50 个中的灯泡)。表内所有其他元素都是用类似方式得到的。

表 4.2.4

月										
0	1000									
1	50	50								
2	150	2.5	152.5							
3	250	7.5	7.6	265.1						
4	250	12.5	22.9	13.3	298.7					
5	150	12.5	38.1	39.8	14.9	255.3				
6	100	7.5	38.1	66.3	44.8	12.8	269.5			
7	50	5.0	22.9	66.3	74.7	38.3	13.5	270.7		
8		2.5	15.3	39.8	74.7	63.8	40.4	13.5	250	
9			7.6	26.6	44.8	63.8	67.4	40.6	12.5	263.3
10				13.3	29.8	38.3	67.4	67.7	37.5	13.2
11					14.9	25.6	40.4	67.7	62.5	39.5
12						12.8	27.0	40.6	62.5	65.8
13							13.5	27.0	37.5	65.8
14								13.5	25.0	39.5
15									12.5	26.4
16										13.2

表 4.2.5 灯泡更换策略

全部更换一次时间(月)	全部更换一次费用(元)	中间更换费用(元)	总费用(元)	平均每月费用
1	400	$50 \times 1.5 = 75.0$	475.0	475.0
2	400	$202.2 \times 1.5 = 303.5$	703.5	351.9
3	400	$467.6 \times 1.5 = 701.4$	1101.4	367.1
4	400	$766.3 \times 1.5 = 1149.4$	1549.4	387.4
5	400	$1021.6 \times 1.5 = 1532.4$	1932.4	386.5
6	400	$1291.1 \times 1.5 = 1936.6$	2336.6	389.4
7	400	$1561.8 \times 1.5 = 2342.7$	2742.7	391.8
8	400	$1811.8 \times 1.5 = 2717.7$	3117.7	389.7
9	400	$2074.9 \times 1.5 = 3112.4$	3512.4	390.3

假定采用的策略是灯泡三个月全部更换一次,则两次总更换之间更换的灯泡数为 $50 + 152.5 + 265.1 = 467.6$ 。为精确起见,在记录每月更换的平均数时取到小数点后面一位。

表 4.2.5 给出了在不同的更换策略下每月费用的期望值。从中我们可知每两月总更换一次的策略所得的期望值最小。应注意,仅采用单个更换的策略平均每月费用为 394.5 元,由表 4.2.5 亦可看出:若采用超过 4 个月总更换一次的策略,平均每月费用将趋向于总更换期为最长(即仅采用单个更换策略)的数字。这由于从表 4.2.4 可见每月更换的平均数在四个月后就与先前得到的 263 相差不多。

4.2.8 建厂问题

为生产某产品,设计了两个基建方案:一是建大厂,二是建小厂。大厂需投资 300 万元,小厂需投资 160 万元,两者的使用期均为 10 年。估计在此期间,产品销路好的可能性是 0.7,两个方案的年度益损值如下:

表 4.2.6 (单位:万元)

自然状态	概率	建大厂	建小厂
销路好	0.7	100	40
销路差	0.3	-20	10

你作为顾问,试给出决策方案。

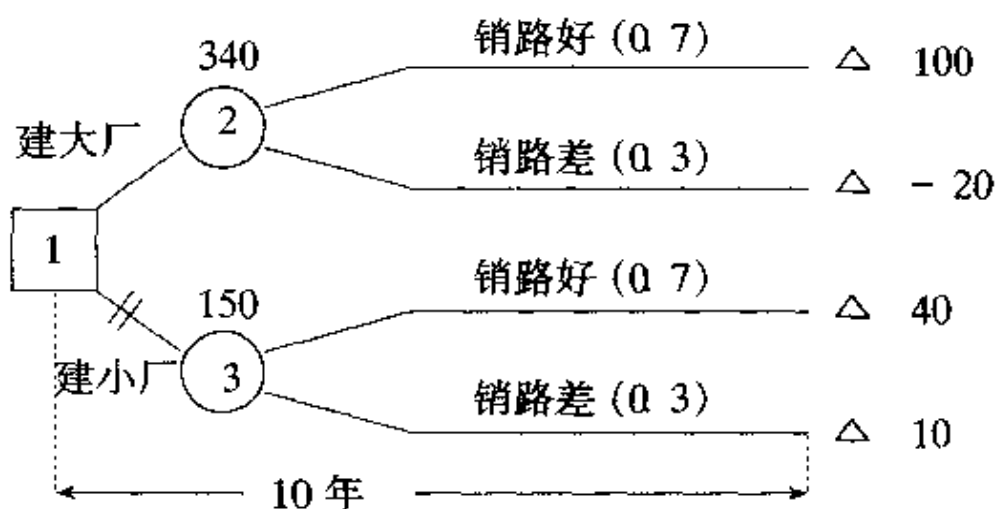


图 4.2.1

针对该问题,我们采用决策树来处理,其基本思想是以损益期望值为基础,按“三思而后行”,“走一步看几步”的想法,划出树形图——即决策树进行决策。为此引入记号:

□:表示决策点,从它引出的分枝叫方案分枝,分枝数反映可能的行动方案数。

○:表示方案节点,其上方数字表示该方案的效益期望值,由它引出的分枝叫概率分枝,每条分枝上写明了自然状态及其出现的概率,分枝数反映可能的自然状态数。

△:表示结果节点,它旁边的数字是每一方案在相应状态下的效益值。

将各方案节点上期望值进行比较,选取效益期望值最大者作为选定方案,而对于放弃的方案则在方案分枝中打“#”号以示删除。

回到我们的问题,首先画决策树图 4.2.1,然后计算各点的损益期望值

点 2: $0.7 \times 100 \times 10 + 0.3 \times (-20) \times 10 - 300 = 340$ (万元)

点 3: $0.7 \times 40 \times 10 + 0.3 \times 10 \times 10 - 160 = 150$ (万元)

两者比较,建大厂的方案是合理的。

考虑另外二种进一步的情况,完成我们的问题:

1° 假定把问题分为前三年和后七年两期考虑。由市场预测,前三年销路好的概率为 0.7 而如果前三年销路好,则后七年销路好的概率为 0.9,而若前三年销路差,则后七年销路肯定差,在这种情况下进行决策。

首先画出决策树图 4.2.2

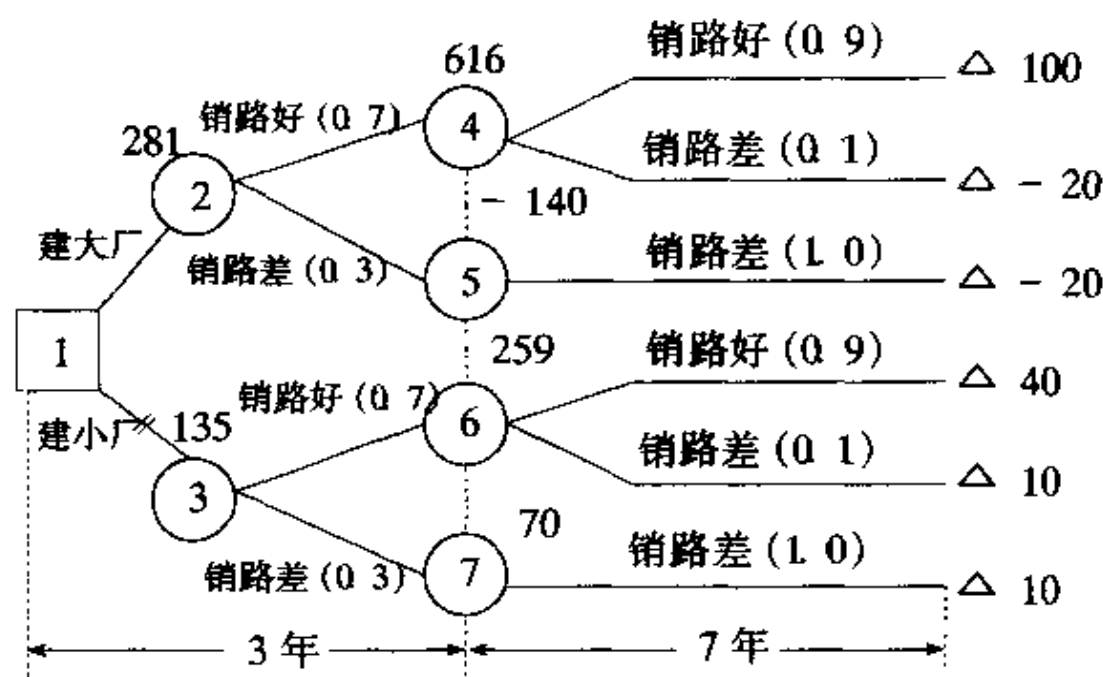


图 4.2.2

接着计算各点的益损期望

$$\text{点 4: } 0.9 \times 100 \times 7 + 0.1 \times (-20) \times 7 = 616$$

$$\text{点 5: } 1.0 \times (-20) \times 7 = -140$$

$$\begin{aligned} \text{点 2: } & 0.7 \times 100 \times 3 + 0.7 \times 616 + 0.3 \times (-20) \times 3 \\ & + 0.3 \times (140) - 300 = 281(\text{万元}) \end{aligned}$$

$$\text{点 6: } 0.9 \times 40 \times 7 + 0.1 \times 10 \times 7 = 259$$

$$\text{点 7: } 1.0 \times 10 \times 7 = 70$$

$$\begin{aligned} \text{点 3: } & 0.7 \times 40 \times 3 + 0.7 \times 259 + 0.3 \times 10 \times 3 + 0.3 \\ & \times 70 - 160 = 135(\text{万元}) \end{aligned}$$

注意到 $281 > 135$, 即知建大厂是合理方案。

2° 现考虑先建小工厂, 如销路好, 三年后考虑扩建。扩建投资需 140 万元, 扩建后可使用七年, 每年的益损值与大厂相同。试评价这个方案与直接建大厂的优劣。

首先画决策图 4.2.3, 接着计算各点的益损期望

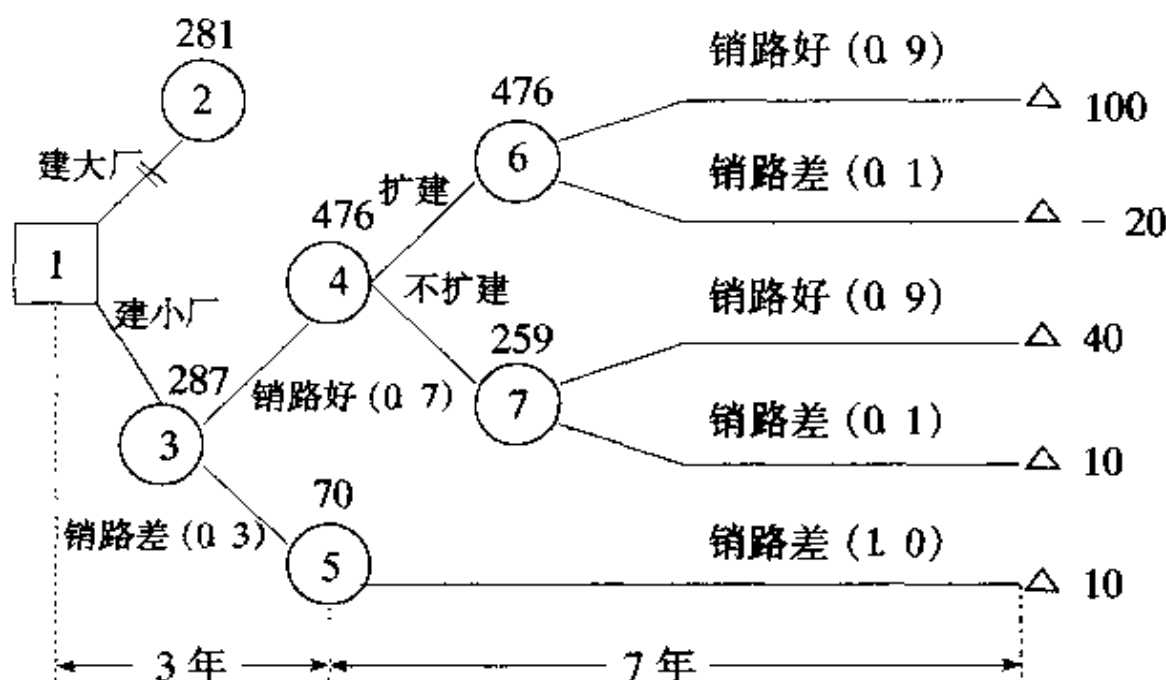


图 4.2.3

点 2: 同图 4.2.2 一样,建大厂的益损期望为 281 万元。

点 6: $0.9 \times 100 \times 7 + 0.1 \times (-20) \times 7 - 140 = 476$ (万元)

点 7: $0.9 \times 40 \times 7 + 0.1 \times 10 \times 7 = 259$ (万元)

由于 $476 > 259$,说明扩建方案较好,将点 6 的 476 转移到点 4。

点 5: $1.0 \times 10 \times 7 = 70$ (万元)

点 3: $0.7 \times 40 \times 3 + 0.7 \times 476 + 0.3 \times 10 \times 3 + 0.3$
 $\times 7 - 160 = 287$ (万元)

由于 $287 > 281$,故得我们的决策方案:先建小厂,三年后扩建为大厂。

决策树法构成了一个简单的决策过程,使决策人有顺序有步骤地进行决策;同时,它又较直观,可使决策人以科学的推理步骤去周密思考问题;对较复杂的问题,尤其是多级决策问题来说尤感方便简捷。因而,可以说它是一种直观、高效的决策手段,具有重要作用。

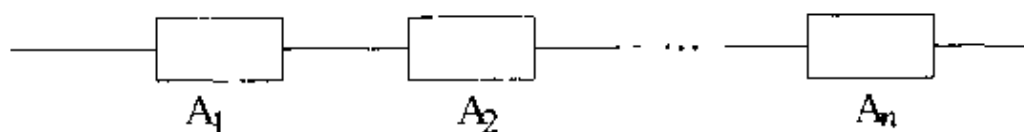
习 题

1. (贝特朗奇论) 在一半径为 r 的圆 c 内“任意”作一弦,试求此弦长度 l 大于圆内接等边三角形的边长 $r\sqrt{3}$ 的概率 p 。

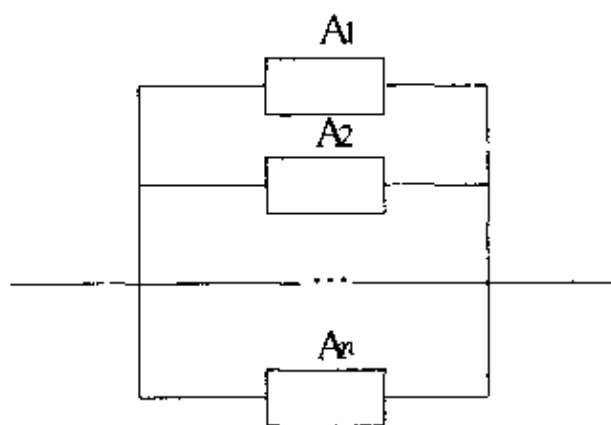
提示:注意对题设中“任意”的假设。

2. (可靠性分析) 在某系统中元件的连接方式通常有二种:并联和串联,一般地并联意味着可靠性的增强而串联则反映功能的增多。今设系统由 n 个元件按下述方式连接,每个元件正常工作的概率为 p ,且它们是独立的,试求系统正常工作的概率,并且由此说明系统并得越多越可靠,串得越多越易坏。又你能由此说明生活中的实际问题吗(比如多功能彩电等)?

(1) 串联系统



(2) 并联系统



3. (供电问题) 某车间有 200 台车床互相独立工作, 由于经常需检修、测量等原因, 即使在生产期间, 各台车床开动时亦需停车。若每台车床有 60% 的时间在开动, 而每台车床在开动时需耗电 1 千瓦, 问应供给这个车间至少多少电力才能保证在 8 小时生产中因电力不足而影响生产的时间不超过半分钟?

4. 12 月 23 日 A 先生想到要给他妻子买圣诞礼物, 他所选中的礼品有两种型式, 普通型售价 279.5 元, 高级型售价 399.5 元。一般地, 到最后一天很可能会削价。A 先生估计在 12 月 24 日普通型式削价为 259.5 元的可能性有 50%, 然而若他等到 24 日再买, 那么他也可能买不到价廉的那一种, 这时将迫使他购买价高的型式。他估计这种可能占 20%。A 先生应采取的正确态度是什么?

5. 某汽车修理站的一个工人一天(8 小时)可修理 12 辆汽车。设修理时间服从指数分布, 汽车来到修理站是一个泊松过程, 每小

时指派一辆汽车给这位工人修理。要是一顾客决定等候汽车修好后交货,那么他要取回汽车平均需等多长时间?

6. 某报的售出份数总介于 80000 到 120000, 每售出一份报收入为 0.30 元, 报纸印刷费等为 0.08 元, 存贮费为 0, 据以往经验, 估计售出份数的概率分布如下表:

可能售出份数	概率
80000	0.1
90000	0.2
100000	0.3
110000	0.3
120000	0.1

1° 如果印刷 100000 份, 收入的期望值是多少? 利润的期望值又是多少?

2° 若印 x 份, 利润的期望值是多少?

3° 为了获取最大利润, 应印刷的最佳份数是多少?

第五章 统计方法建模

§ 5.1 预备知识

5.1.1 基本概念及统计学中的重要分布

前章我们介绍了概率论的基本内容及其建模方法。本章运用概率知识,对随机现象进行多次观察或实验,研究如何合理地获得数据资料,提出有效的数学方法,分析所获资料,对所关心问题作出估计与检验。具体讲,就是依中心极限定理,大数定律,通过抽样的分析研究,来推断研究对象的整体性质,用所获结果来指导生产、实验。

定义 5.1.1 称某一问题的研究对象的全体为总体(母体),其中的一个单位叫做个体。从总体中随机抽取的 n 个个体(单位)称为容量为 n 的子样。

今后,我们常用“总体 ζ ”表示总体的某一侧面的特征,而用“由总体 ζ 抽取子样 ζ_1, \dots, ζ_n ”表示与 ζ 相对应的这一个侧面的特征。我们总是以总体的一个或几个(多元总体情形)侧面的特征作为研究对象,以后,对总体与总体对应的某一侧面的特征不加区别。

定义 5.1.2 设 $\zeta_1, \zeta_2, \dots, \zeta_n$ 为由总体 ζ 抽取的容量为 n 的子样,如果 ζ_1, \dots, ζ_n 相互独立且每一个都是与 ζ 同分布的 *r. v.*, 则称 ζ_1, \dots, ζ_n 为总体 ζ 的容量为 n 的简单随机子样,简称为子样。

一般地,我们用有返回抽样来获取简单随机子样。引入它的目的就是要从子样所获得的概率统计特性来对总体的概率统计特性

作估计和推断。

定义 5.1.3 子样可由随机向量 $(\zeta_1, \dots, \zeta_n)$ 表示。记 x_i 为 ζ_i 的一次观察值, (x_1, \dots, x_n) 为子样的一次观察值。子样的所有可能观察值的全体称为子样空间,记为 X ,它为 n 维空间,而子样的一次观察值 (x_1, \dots, x_n) 为 X 中一点。

定义 5.1.4 设 ζ_1, \dots, ζ_n 为总体 ζ 的子样, T 为子样空间 X 中点 (x_1, \dots, x_n) 的实值函数,作子样函数 $T = T(\zeta_1, \dots, \zeta_n)$, T 的取值记为 $t = T(x_1, \dots, x_n)$ 。若 $T(\zeta_1, \dots, \zeta_n)$ 也为 $r.v.$,且不带任何参数,则称 T 或 $T(\zeta_1, \dots, \zeta_n)$ 为统计量。

统计量可以说是统计学中最重要的概念,为了完成某统计目的,我们就要构造相应的统计量——子样不含未知参数的函数,以此来完成我们的估计、推断。下面给出几个常用的统计量。

1° 设 ζ_1, \dots, ζ_n 为总体 ζ 的子样,容量为 n ,记

$$\begin{aligned}\bar{\xi} &= \frac{1}{n} \sum_{i=1}^n \xi_i, S^2 = \frac{1}{n} \sum_{i=1}^n (\xi_i - \bar{\xi})^2 \\ A_k &= \frac{1}{n} \sum_{i=1}^n \xi_i^k\end{aligned}\quad (5.1.1)$$

则 $\bar{\xi}$ 、 S^2 及 A_k 都是统计量,分别称为子样 ξ_1, \dots, ξ_n 的平均值、方差及 k 阶矩。子样观察值为 x_1, \dots, x_n , $\bar{\xi}$ 、 S^2 及 A_k 的观察值记为

$$\begin{aligned}\bar{x} &= \frac{1}{n} \sum_{i=1}^n x_i, S^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \\ a_k &= \frac{1}{n} \sum_{i=1}^n x_i^k\end{aligned}$$

2° 设 ξ_1, \dots, ξ_n 为总体 ξ 的子样,今由子样建立 n 个函数:

$$\xi_k^* = \xi_k^*(\xi_1, \dots, \xi_n) \quad k = 1, 2, \dots, n$$

其中 ξ_k^* 为这样的统计量,它的观察值为 x_k^* , x_k^* 为子样 ξ_1, \dots, ξ_n 的观察值 x_1, \dots, x_n 中由小至大排列(即 $x_1^* \leq \dots \leq x_n^*$)后的第 k 个数值,则称 ξ_1^*, \dots, ξ_n^* 为顺序统计量。

$\xi_1^* = \min(\xi_1, \dots, \xi_n)$ 称最小项统计量, $\xi_n^* = \max(\xi_1, \dots, \xi_n)$ 称最大项统计量, $D_n^* = \xi_n^* - \xi_1^*$ 为子样的极差。

为了进行统计推断, 现引入统计学中重要的分布及抽样分布定理:

1° χ^2 分布: 设 $\xi_i \sim N(0, 1) (i = 1, 2, \dots, n)$, 且 ξ_1, \dots, ξ_n 相互独立, 令

$$\chi^2(n) = \sum_{i=1}^n \xi_i^2 \quad (5.1.2)$$

称 $\chi^2(n)$ 服从自由度为 n 的 χ^2 分布, 可以通过查表的方式来对 $\chi^2(n)$ 进行计算, 图示如下。

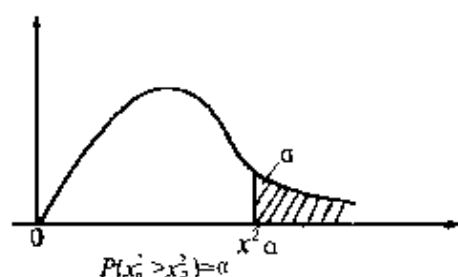


图 5.1.1

2° t 分布: 设 $\xi \sim N(0, 1)$, $\eta \sim \chi^2(n)$ 且有 ξ, η 相互独立, 令

$$t(n) = \frac{\xi}{\sqrt{\eta/n}} \quad (5.1.3)$$

称 $t(n)$ 服从自由度为 n 的 t 分布, 可以通过查表的方式来对 $t(n)$ 计算, 图示如下。

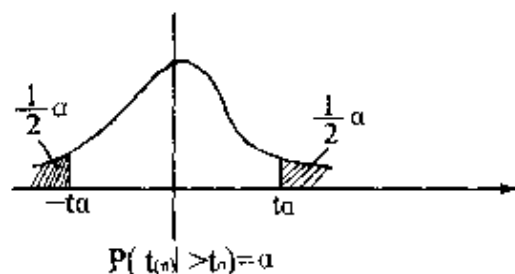


图 5.1.2

3° F 分布: 设 $\xi \sim \chi^2(m)$, $\eta \sim \chi^2(n)$, ξ, η 相互独立, 令

$$F(m, n) = \frac{\xi/m}{\eta/n} \quad (5.1.4)$$

称 $F(m, n)$ 服从第一个自由度为 m 、第二个自由度为 n 的 F 分布, 也可以由查表的方式来计算 $F(m, n)$, 图示如下。

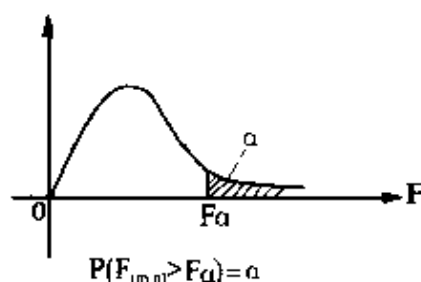


图 5.1.3

χ^2 分布、 t 分布、 F 分布是统计学中占有重要地位的三个分布, 许多问题都是通过它们加以解决的。

4° 设总体 $\xi \sim N(a, \sigma)$, ξ_1, \dots, ξ_n 为其子样, 子样平均值与方差分别记为 $\bar{\xi}$ 和 S^2 , 则 $\bar{\xi} \sim N(a, \frac{\sigma}{\sqrt{n}})$, $\frac{nS^2}{\sigma^2} \sim \chi^2_{(n-1)}$ 且 $\bar{\xi}$ 与 S^2 相互独立。

5° 设总体 $\xi \sim N(a, \sigma)$, ξ_1, \dots, ξ_n 为子样则

$$T = \sqrt{n-1} \frac{\bar{\xi} - a}{S} \sim t(n-1) \quad (5.1.5)$$

6° 设总体 $\xi \sim N(a_1, \sigma)$, ξ_1, \dots, ξ_{n_1} 为其子样, 总体 $\eta \sim N(a_2, \sigma)$, $\eta_1, \dots, \eta_{n_2}$ 为其子样, 而且这两子样是相互独立的, 记

$$\begin{aligned} \bar{\xi} &= \frac{1}{n_1} \sum_{i=1}^{n_1} \xi_i, & S_1^2 &= \frac{1}{n_1} \sum_{i=1}^{n_1} (\xi_i - \bar{\xi})^2 \\ \bar{\eta} &= \frac{1}{n_2} \sum_{j=1}^{n_2} \eta_j, & S_2^2 &= \frac{1}{n_2} \sum_{j=1}^{n_2} (\eta_j - \bar{\eta})^2 \end{aligned}$$

则有

$$(1) \quad \frac{(n_2 - 1)n_1 S_1^2}{(n_1 - 1)n_2 S_2^2} \sim F_{(n_1-1, n_2-1)} \quad (5.1.6)$$

$$(2) \quad \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}} \frac{(\bar{\xi} - \bar{\eta}) - (a_1 - a_2)}{\sqrt{n_1 S_1^2 + n_2 S_2^2}} \sim t_{(n_1+n_2-2)} \quad (5.1.7)$$

7° 设 ξ_1, \dots, ξ_k 为 k 个相互独立的且每个都服从正态 $N(0, 1)$, 今从 k 个总体中分别抽取容量为 n_i 的子样 $\xi_{i1}, \dots, \xi_{in_i} (i = 1, 2, \dots, k)$, 记

$$\begin{aligned} \bar{\xi}_i &= \frac{1}{n_i} \sum_{j=1}^{n_i} \xi_{ij} \\ S_i^2 &= \frac{1}{n} \sum_{j=1}^{n_i} (\xi_{ij} - \bar{\xi}_i)^2 \\ n &= \sum_{i=1}^k n_i \\ \bar{\xi} &= \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^{n_i} \xi_{ij} = \frac{1}{n} \sum_{i=1}^k n_i \bar{\xi}_i \end{aligned}$$

则

$$(1) \quad S_{\text{总}}^2 = \sum_{i=1}^k \sum_{j=1}^{n_i} (\xi_{ij} - \bar{\xi}_i)^2 = \sum_{i=1}^k n_i S_i^2 + \sum_{i=1}^k n_i (\bar{\xi}_i - \bar{\xi})^2 \quad (5.1.8)$$

$$(2) \quad \text{记 } Q_e = \sum_{i=1}^k n_i S_i^2, \quad U = \sum_{i=1}^k n_i (\bar{\xi}_i - \bar{\xi})^2$$

有

$$\frac{U/(k-1)}{Q_e/(n-k)} \sim F(k-1, n-k) \quad (5.1.9)$$

5.1.2 参数的估计理论

参数估计主要包含点估计和区间估计两方面的内容。本段主要介绍点估计的两个重要的方法——矩法和极大似然法, 估计量

的优良性准则以及区间估计。

1. 参数的点估计

定义 5.1.5 设总体 ξ 的分布函数 $F(x, \theta_1, \dots, \theta_l)$ 中有 l 个未知参数 $\theta_1, \dots, \theta_l$, 假定总体 ξ 的 l 阶原点矩存在, 记

$$V_k(\theta_1, \dots, \theta_l) = E_{\theta_1, \dots, \theta_l}(\xi^k) (k = 1, 2, \dots, l)$$

由下列方程组

$$\begin{cases} \frac{1}{n} \sum_{i=1}^n \xi_i = \hat{v}_1(\theta_1, \dots, \theta_l) \\ \frac{1}{n} \sum_{i=1}^n \xi_i^2 = \hat{v}_2(\theta_1, \dots, \theta_l) \\ \dots \\ \dots \\ \frac{1}{n} \sum_{i=1}^n \xi_i^l = \hat{v}_l(\theta_1, \dots, \theta_l) \end{cases} \quad (5.1.10)$$

解得 $\hat{\theta}_k = \hat{\theta}_k(\xi_1, \dots, \xi_n) (k = 1, 2, \dots, l)$, 并以 $\hat{\theta}_k$ 作为参数 θ_k 的估计量, 则称 $\hat{\theta}_k$ 为未知参数 θ_k 的矩法估计量。

理论依据: 大数定律, 用子样矩估计总体矩。特别地, 若 ξ 的分布中只含两个未知参数即 $\xi \sim F(x, \theta_1, \theta_2)$, 则以子样均值估计总体均值, 子样方差估计总体方差建立方程。

由矩法估计易知:

(1) 事件发生频率是事件发生概率的矩法估计量。

(2) 若总体 ξ 服从 $[\theta_1, \theta_2]$ 上均匀分布, θ_1, θ_2 未知参数, 则其矩法估计为:

$$\hat{\theta}_1 = \bar{\xi} - \sqrt{3}S, \hat{\theta}_2 = \bar{\xi} + \sqrt{3}S \quad (5.1.11)$$

矩法估计的优缺点:

优点: 不需知道总体的分布, 只要依据数字特征——矩即可完成估计。

缺点:由于没有用到总体的分布,即没有充分到用上所提供的一切信息。

定义 5.1.6 设总体 ξ 的密度函数为 $f(x, \theta_1, \dots, \theta_l)$ 其中 $\theta_1, \dots, \theta_l$ 为未知参数,参数空间 Ω 是 l 维的 ξ_1, \dots, ξ_n 为子样,它的联合分布密度函数表示为 $f(x_1, \dots, x_n; \theta_1, \dots, \theta_l)$ 。称

$$L(\theta_1, \dots, \theta_l) = \prod_{i=1}^n f(\xi_i; \theta_1, \dots, \theta_l) \quad (5.1.12)$$

为 $\theta_1, \dots, \theta_l$ 的似然函数,若有 $\hat{\theta}_1, \dots, \hat{\theta}_l$ 使

$$L(\hat{\theta}_1, \dots, \hat{\theta}_l) = \max_{(\theta_1, \dots, \theta_l) \in \Omega} L(\theta_1, \dots, \theta_l) \quad (5.1.13)$$

称 $\hat{\theta}_j = \hat{\theta}_j(\xi_1, \dots, \xi_n)$ 为 θ_j 的极大似然法估计量($j = 1, 2, \dots, l$)。

基本思想:某事情发生了,我们找最容易使它发生的原因,认为该事情就是由这个原因导致的。

极大似然估计量的求法:

注意到

$$\ln L(\theta_1, \dots, \theta_l) = \sum_{i=1}^n \ln f(\xi_i; \theta_1, \dots, \theta_l)$$

由于 $\ln x$ 是 x 的单增函数,因而 $\ln L$ 与 L 有相同的极大值点。 $\hat{\theta}_j$ 为极大似然估计量的必要条件为

$$\frac{\partial \ln L(\theta_1, \dots, \theta_l)}{\partial \theta_j} \Big|_{\theta_j = \hat{\theta}_j} = 0 \quad j = 1, 2, \dots, l \quad (5.1.14)$$

称它为似然方程,由它解得 $\hat{\theta}_j = \hat{\theta}_j(\xi_1, \dots, \xi_n)$ 当满足(5.1.13)时,则 $\hat{\theta}_j$ 为 θ_j 的极大似然估计量, $j = 1, 2, \dots, l$ 。

若 ξ 是离散型 $r. v.$, 则将其概率函数 $P(x, \theta_1, \dots, \theta_l)$ 代替定义 5.1.6 中连续型 $r. v.$ 的分布密度函数,其他不变,仍按上述方法可得到 θ_j 的极大似然估计量 $\hat{\theta}_j (j = 1, 2, \dots, l)$

由极大似然法易知

(1) 设总体 ξ 服从参数为 λ 的泊松分布,则 $\hat{\lambda} = \bar{\xi}$

(2) 设总体 ξ 服从 $[\theta_1, \theta_2]$ 上均匀分布, θ_1, θ_2 为未知参数, 则 $\hat{\theta}_1 = \xi_1^*, \hat{\theta}_2 = \xi_n^*$

(3) 设 $\xi \sim N(a, \sigma)$, a, σ 为未知参数, 则有 $\hat{a} = \bar{\xi}, \hat{\sigma}^2 = S^2$

由上述可见, 不论是矩法估计量还是极大似然法估计量, 都是寻找统计量作为未知参数的估计量。应注意的是: 由于存在期望不存在的 r, v (总体), 因而有的总体分布中的未知参数不存在矩法估计量; 由极大似然法和矩法对总体 ξ 服从 $[\theta_1, \theta_2]$ 上的估计结果不同, 即相同的未知参数可能有不同的估计量, 这就提示我们应考虑估计量的好坏, 好坏的标准如何确定?

一般地, 我们从如下几个方面来刻划估计量的好坏:

(1) 无偏性: 若参数 θ 的估计量 $T(\xi_1, \dots, \xi_n)$ 对一切 n 及 $\theta \in \Omega$ 有

$$E_{\theta}[T(\xi_1, \dots, \xi_n)] = \theta \quad (5.1.15)$$

则称 $T(\xi_1, \dots, \xi_n)$ 为参数 θ 的无偏估计量。

我们有:

(i) 子样均值 $\bar{\xi}$ 是总体均值 $E\xi$ 的无偏估计。

(ii) 若 ξ_1, \dots, ξ_n 为由总体 ξ 抽取的容量为 n 的子样, 则

$\sum_{i=1}^n C_i \xi_i$ (其中 $C_i \geq 0$ 常数有 $\sum_{i=1}^n C_i = 1$) 是总体均值 $E\xi$ 的无偏估计。

(iii) 子样方差 S^2 是总体方差 $D\xi$ 的有偏估计。

即

$$ES^2 = \frac{n-1}{n} D\xi \neq D\xi$$

对子样方差, 我们可进行无偏化运算 $S^{*2} = \frac{n}{n-1} S^2$, 则 S^{*2} 为 $D\xi$ 的无偏估计量。

由 (i)、(ii) 知, 一般地未知参数的无偏估计不惟一, 由此有

(2) 最小方差无偏估计量(最优无偏估计量):

设 $T(\xi_1, \dots, \xi_n)$ 为未知参数 θ 的无偏估计, 若对于 θ 的任一无偏估计 T' 有

$$D_{\theta}T \leq D_{\theta}T'$$

称 T 为 θ 的最优无偏估计; 我们有: 对未知参数 θ 而言至多存在一个最优无偏估计。

若 $\xi \sim N(a, \sigma)$ a, σ 为未知参数, 则 $\bar{\xi}, S^{*2}$ 分别为 a, σ^2 的最优无偏估计。

为了兼顾方差和偏差绝对值都较小, 还有

(3) 相合性估计量: 设 $T(\xi_1, \dots, \xi_n)$ 为未知参数 θ 的估计量, 若对任 $\epsilon > 0$, 有

$$\lim_{n \rightarrow \infty} P(|T(\xi_1, \dots, \xi_n) - \theta| \leq \epsilon) = 1 \quad (5.1.16)$$

称 $T(\xi_1, \dots, \xi_n)$ 为 θ 的弱相合估计量。若

$$P\left(\lim_{n \rightarrow \infty} T(\xi_1, \dots, \xi_n) = \theta\right) = 1 \quad (5.1.17)$$

称 $T(\xi_1, \dots, \xi_n)$ 为 θ 的强相合估计量。

2. 参数的区间估计

定义 5.1.7 设总体 ξ 分布函数为 $F(x, \theta)$, θ 为未知参数, ξ_1, \dots, ξ_n 为其子样, 建立二统计量 $T_1(\xi_1, \dots, \xi_n)$ 及 $T_2(\xi_1, \dots, \xi_n)$, 满足

$$T_1(\xi_1, \dots, \xi_n) < T_2(\xi_1, \dots, \xi_n)$$

则称 $[T_1, T_2]$ 为随机区间。

定义 5.1.8 设 α 为一给定常数, 满足 $0 < \alpha < 1$ 若有

$$P(T_1 \leq \theta \leq T_2) = 1 - \alpha \quad (5.1.18)$$

并用随机区间 $[T_1, T_2]$ 作为参数 θ 的估计, 称其为参数 θ 的置信水平为 $1 - \alpha$ 的区间估计, 称 α 为显著性水平, T_1, T_2 分别为下、上置信限。

实质:区间估计实际上就是寻找一个范围,使这个范围以充分大的可能(概率)包含了未知参数。其关键在于求上、下置信限 T_2 、 T_1 时估计变量的选取。这个估计变量必须是一个含且只含一个未知参数(待估参数)的 $r.v.$ 。

下面我们就总体 $\xi \sim N(a, \sigma)$ (这一类总体所代表的实际模型相当丰富,这可由中心极限定理保证),讨论未知参数 a 及 σ^2 的区间估计。

先考虑 a 的区间估计。

(1) 若 $\sigma^2 = \sigma_0^2$ 已知,则由 $\bar{\xi} \sim N(a, \frac{\sigma_0}{\sqrt{n}})$, 有

$$U = \frac{\bar{\xi} - a}{\sigma_0 / \sqrt{n}} \sim N(0, 1)$$

给定置信水平 $1 - \alpha$, 在标准正态表中查临界值 u_α 使

$$P\left(\left|\frac{\bar{\xi} - a}{\sigma_0 / \sqrt{n}}\right| \leq u_\alpha\right) = 1 - \alpha$$

$$P\left(\bar{\xi} - u_\alpha \frac{\sigma_0}{\sqrt{n}} \leq a \leq \bar{\xi} + u_\alpha \frac{\sigma_0}{\sqrt{n}}\right) = 1 - \alpha \quad (1.19)$$

即得随机区间 $[\bar{\xi} - u_\alpha \frac{\sigma_0}{\sqrt{n}}, \bar{\xi} + u_\alpha \frac{\sigma_0}{\sqrt{n}}]$ 为数学期望 a 的置信水平为 $1 - \alpha$ 的区间估计。

(2) 若 σ^2 未知,则由(5.1.5)

$$\sqrt{n-1} \frac{\bar{\xi} - a}{S} \sim t_{(n-1)}$$

给定置信水平 $1 - \alpha$, $0 < \alpha < 1$, 在 t 分布表中查自由度为 $n - 1$ 对应的临界值 $t_{(n-1)}(\alpha)$ 使

$$P\left(\left|\sqrt{n-1} \frac{\bar{\xi} - a}{S}\right| \geq t_{(n-1)}(\alpha)\right) = \alpha$$

$$P\left(\bar{\xi} - t_{(n-1)}(\alpha) \frac{S}{\sqrt{n-1}} \leq a \leq \bar{\xi} + t_{(n-1)}(\alpha) \frac{S}{\sqrt{n-1}}\right) = 1 - \alpha$$

(5.1.20)

即得随机区间 $[\bar{\xi} - t_{(n-1)}(\alpha) \frac{S}{\sqrt{n-1}}, \bar{\xi} + t_{(n-1)}(\alpha) \frac{S}{\sqrt{n-1}}]$ 为 a 的置信水平为 $1 - \alpha$ 的区间估计。

注:在(2)中,我们不能用(1)中给的

$$U = \frac{\bar{\xi} - a}{\sigma / \sqrt{n}} \sim N(0, 1)$$

原因在于 U 中含二未知参数 a, σ , 若选它则在对应它而得的 a 的随机区间上、下限中必含参数 σ , 即它们不是统计量, 实质上你没有给出 a 的区间估计。

(3) 考虑 σ^2 的区间估计, 由于

$$\frac{nS^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

给定置信水平 $1 - \alpha, 0 < \alpha < 1$, 在 χ^2 分布临界值表中查

$\chi^2_{(n-1)}(\frac{\alpha}{2})$ 及 $\chi^2_{(n-1)}(1 - \frac{\alpha}{2})$ 有

$$P\left(\chi^2_{(n-1)}(1 - \frac{\alpha}{2}) \leq \frac{nS^2}{\sigma^2} \leq \chi^2_{(n-1)}(\frac{\alpha}{2})\right) = 1 - \alpha$$

$$P\left(\frac{nS^2}{\chi^2_{(n-1)}(\frac{\alpha}{2})} \leq \sigma^2 \leq \frac{nS^2}{\chi^2_{(n-1)}(1 - \frac{\alpha}{2})}\right) = 1 - \alpha$$

(5.1.20)

即随机区间 $\left[\frac{nS^2}{\chi^2_{(n-1)}(\frac{\alpha}{2})}, \frac{nS^2}{\chi^2_{(n-1)}(1 - \frac{\alpha}{2})}\right]$ 为 σ^2 的置信水平为 $1 - \alpha$ 的区间估计。

在进行区间估计时, 我们是冒着犯“错误”的风险的, “犯错”的概率为显著性水平 α 。

5.1.3 参数及模型的假设检验

假设检验实质上是对人们实践经验所提出的结果或通过估计

理论所求出结果进行验证的一种方法。它所依据的是小概率事件原理：小概率事件（或概率很小的事件）在一次试验（观察）中是几乎不可能发生的。假设检验一般是针对参数或模型服从分布提出的，进行假设检验的一般步骤是：

第一步：提出原假设及备选假设

$$H_0: a = a_0, \quad H_1: a \neq a_0 \quad (\text{参数})$$

或

$$H_0: F(x) = F_0(x), H_1: F(x) \neq F_0(x) \quad (\text{模型分布})$$

第二步：建立检验统计量，以此来构造 H_0 成立条件下小概率事件。检验统计量是进行假设检验的核心步骤，它是一个除了待检量之外不含任何其他未知参数的 $r.v.$ 。

第三步：确定否定域，进行判断，实质上否定域就是 H_0 成立下的一个小概率事件，通过对总体进行一次抽样，发现其值落入了否定域即一次试验小概率事件发生了，这就使我们有理由对原假设产生怀疑，否定原假设。

下面就参数的假设检验及模型的假设检验分别予以讨论。

1. 参数的假设检验

首先我们就总体 $\xi \sim N(a, \sigma_0)$, σ_0 已知，讨论未知参数 a 的假设检验。

$$H_0: a = a_0, \quad H_1: a \neq a_0$$

注意到 $\bar{\xi} \sim N(a, \frac{\sigma_0}{\sqrt{n}})$

$$U = \frac{\bar{\xi} - a}{\sigma_0 / \sqrt{n}} \sim N(0, 1)$$

若 H_0 成立，则统计量 $U = \frac{\bar{\xi} - a_0}{\sigma_0 / \sqrt{n}} \sim N(0, 1)$

对给定的显著性水平 α , $0 < \alpha < 1$, 查表有 u_α

$$P(|U| \leq u_\alpha) = 1 - \alpha, P(|U| > u_\alpha) = \alpha$$

从而得 H_0 的否定域为 $(-\infty, -u_\alpha) \cup (u_\alpha, \infty)$ 。

对总体 ξ 进行一次抽样, 得一组观察值为 x_1, \dots, x_n , 由此可求得 U 的观察值 u ——它为一个数! 若 $u \in (-\infty, -u_\alpha) \cup (u_\alpha, \infty)$, 即落入了 H_0 的否定域, 则否定 H_0 ; 若 $u \in (-\infty, u_\alpha) \cup (u_\alpha, \infty)$, 则不否定 H_0 , 接受 H_0 。

进行假设检验的步骤均同上述, 值得注意的是 α 的选取, 它是所构造的小概率事件发生的概率, 因而一般地 $\alpha \leq 0.1$, 经常选用的 α 值有 0.01, 0.05, 0.1, 不同的 α 值可能得到不同的检验结论。下面就常见的参数检验问题, 给出它们的检验统计量, 请读者按上述步骤完成假设检验。

单总体的情形: 设总体 $\xi \sim N(a, \sigma)$, ξ_1, \dots, ξ_n 为容量为 n 的子样。

(1) $\sigma^2 = \sigma_0^2$ 已知时, 数学期望 a 的检验:

$$H_0: a = a_0 \quad H_1: a \neq a_0$$

检验量

$$U = \frac{\bar{\xi} - a}{\sigma_0 / \sqrt{n}} \sim N(0, 1) \quad (5.1.22)$$

(2) σ^2 未知时, 数学期望 a 的检验

检验量

$$T = \sqrt{n-1} \frac{\bar{\xi} - a}{S} \sim t(n-1) \quad (5.1.23)$$

(3) 方差 σ^2 的检验

$$H_0: \sigma^2 = \sigma_0^2,$$

$$H_1: \sigma \neq \sigma_0^2$$

检验量

$$\chi^2 = \frac{ns^2}{\sigma^2} \sim \chi^2(n-1)$$

二总体的情形: 设 $\xi \sim N(a_1, \sigma_1)$, $\eta \sim N(a_2, \sigma_2)$, $a_1, \sigma_1, a_2, \sigma_2$ 都

是未知的,现独立地分别从总体 ξ 及 η 中抽取子样 ξ_1, \dots, ξ_n 及 $\eta_1, \dots, \eta_{n_2}$, 其子样均值、子样方差记为 $\bar{\xi}, S_1^2$ 及 $\bar{\eta}, S_2^2$, 下首先分三种情况进行数学期望的比较, 然后进行方差的比较

$$(4) \text{ 设 } \sigma_1^2 = \sigma_2^2 = \sigma^2$$

$$H_0: a_1 = a_2, \quad H_1: a_1 \neq a_2$$

检验量(式(5.1.7))

$$\sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}} \frac{(\bar{\xi} - \bar{\eta}) - (a_1 - a_2)}{\sqrt{n_1 S_1^2 + n_2 S_2^2}} \sim t(n_1 + n_2 - 2) \quad (5.1.26)$$

$$\text{特别当 } n_1 = n_2 = n \text{ 时, 有 } \sqrt{n-1} \frac{(\bar{\xi} - \bar{\eta}) - (a_1 - a_2)}{\sqrt{S_1^2 + S_2^2}} \sim$$

$$t_{(2n-2)}$$

$$(5) \sigma_1^2 \neq \sigma_2^2, n_1 = n_2 = n$$

$$H_0: a_1 = a_2, \quad H_1: a_1 \neq a_2$$

采用“配对法”令

$$Z_i = \xi_i - \eta_i$$

有

$$EZ_i = a_1 - a_2 = d$$

$$DZ_i = \sigma_1^2 + \sigma_2^2 = \sigma^2$$

故 Z_1, \dots, Z_n 看作为总体 $Z \sim N(d, \sigma)$ 抽取子样, 于是原假设检验问题化成了单总体时情形:

$$H_0: d = 0, H_1: d \neq 0$$

应用(5.1.23) 给出的检验量即可。

$$(6) \sigma_1^2 \neq \sigma_2^2, n_1 \neq n_2$$

$$H_0: a_1 = a_2 \quad H_1: a_1 \neq a_2$$

采用斯切非解法, 不妨设 $n_1 < n_2$, 令

$$Z_i = \xi_i - \sqrt{\frac{n_1}{n_2}} \eta_i + \frac{1}{\sqrt{n_1 \cdot n_2}} \sum_{k=1}^{n_1} \eta_k - \frac{1}{n_2} \sum_{k=1}^{n_2} \eta_k \quad (i = 1, 2, \dots, n_1) \quad (5.1.27)$$

则

$$EZ_i = a_1 - a_2 = d$$

$$DZ_i = \sigma_1^2 + \frac{n_1}{n_2} \sigma_2^2$$

$$\text{cov}(Z_i, z_j) = 0 \quad (i \neq j, i, j = 1, 2, \dots, n_1)$$

即 Z_1, \dots, Z_{n_1} 可看作由总体 $Z \sim N(d, \sqrt{\sigma_1^2 + \frac{n_1}{n_2} \sigma_2^2})$ 抽取的子样, 于是原问题化作单总体情形:

$$H_0: d = 0 \quad H_1: d \neq 0$$

应用(5.1.23) 给出的检验量。

(7) 考虑方差的比较

$$H_0: \sigma_1^2 = \sigma_2^2 \quad H_1: \sigma_1^2 \neq \sigma_2^2$$

检验量(式(5.1.6))

$$F = \frac{\frac{n_1 S_1^2}{\sigma_1^2} / (n_1 - 1)}{\frac{n_2 S_2^2}{\sigma_2^2} / (n_2 - 1)} = \frac{n_1 (n_2 - 1) S_1^2}{n_2 (n_1 - 1) S_2^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} \sim F_{(n_1-1, n_2-1)} \quad (5.1.28)$$

值得注意的是, 我们在作出判断时是冒着犯二类错误的风险的:

第一类错误: 把原本正确的原假设 H_0 由于观察值落入了其否定域而错误地否定掉, 也称为“弃真”的错误, 其概率为 α 。

第二类错误: 原本错误的原假设 H_0 由于观察值没有落入其否定域而错误地接收下来, 也称为“存伪”的错误, 其概率为 β 。

这里 $\beta \neq 1 - \alpha$, 一般地 α, β 不能同时减小(当子样容量 n 固

定时), 而 α, β 又反映了两种不同性质的错误, 因此通常情况下我们采取这样的原则: 在控制第一类错误 α 的前提条件下, 使犯第二类错误的概率 β 尽可能地小。

2. 模型的假设检验

考虑下面的假设检验问题:

$$H_0: F(x) = F_0(x), H_1: F(x) \neq F_0(x)$$

我们针对 $F_0(x)$ 为正态分布及一般情况讨论

(1) 若 $F_0(x)$ 为正态分布, 即检验总体 ξ 是否服从正态, 则有一个简便方法——正态概率纸法。只要把你的观察结果给在正态概率纸上, 若所得点近似地位于一条直线上, 则可认为它服从正态分布。

(2) 若 $F_0(x)$ 为一般分布(含正态分布), 则我们用皮尔逊 χ^2 检验法。考虑

$$H_0: F(x) = F_0(x) \quad H_1: F(x) \neq F_0(x)$$

设由总体 ξ 抽取子样 ξ_1, \dots, ξ_n , 分下面几步进行讨论:

(i) 分组: 把 R^1 (或区间 $[y_0, y_m]$ 满足 $y_0 \leq \xi_i \leq y_m$) 分组几个互不相交的小区间(一般 10 ~ 20 个)

(ii) 记 v_i 为子样 ξ_1, \dots, ξ_n 落入第 i 个小区间的个数(频数), 称 v_i 为观察频数

(iii) 考虑

$$p_i = F_0(y_i) - F_0(y_{i-1})$$

则 $0 < p_i < 1$, $\sum_{i=1}^m p_i = 1$, y_{i-1}, y_i 为小区间分点, 则有 np_i 为子样 ξ_1, \dots, ξ_n 落入第 i 个小区间的理论频数

(iv) 令

$$\eta = \sum_{i=1}^m \frac{(v_i - np_i)^2}{np_i} = \sum_{i=1}^m \frac{v_i^2}{np_i} - n \quad (5.1.29)$$

我们有皮尔逊定理:

定理 5.1.1 不论 $F_0(x)$ 是什么分布, 当 H_0 正确时, 则由 (5.1.29) 建立的统计量 η 以自由度 $m-1$ 的 χ^2 分布为极限分布, 其中 $F_0(x)$ 不带有未知参数。

推证 5.1.1 若 $F_0(x, \theta_1, \dots, \theta_l)$ 中含有 l 个未知参数 $\theta_1, \dots, \theta_l$ ($m > 1 + l$), 则首先用极大似然法估计 l 个未知参数 $\hat{\theta}_1, \dots, \hat{\theta}_l$, 则有 $\eta(\hat{\theta}_1, \dots, \hat{\theta}_l)$ 以 $\chi^2(m-1-l)$ 为其极限分布。

这样应用定理 5.1.1 (或推论 5.1.1) 就可以完成模型的假设检验。

5.1.4 回归分析方法

所谓回归分析就是建立一个随机变数同另一个或一组非随机变数的关系。若由一个或一组非随机变数来估计或预测另一个随机变数期望值时建立的是线性关系, 称之为线性回归分析。回归分析在预测、控制、工艺优化等方面都具有显著的效果。

我们就简单的一元线性模型展开讨论。设 x 是一个非随机的量, y 是一个随机变量, 它与 x 有关, 现抽取 n 个 x 值进行观察, 得到了 y 的 n 个相应的观察值, 记为 $(x_1, y_1), \dots, (x_n, y_n)$ 。首先假设 x, y 之间具有线性关系

$$y = \beta_0 + \beta_1 x + \varepsilon$$

其中 β_0, β_1 未知参数, $\varepsilon \sim N(0, \sigma)$ 。

则若求得 β_0, β_1 的估计值, 则也就确定了 x, y 之间的关系, 为此令

$$Q = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2 \quad (5.1.30)$$

则使 Q 达到最小的 $\hat{\beta}_0, \hat{\beta}_1$ 能够最好地反映 x, y 关系, 为此我们来求 Q 的极小点获得 β_0, β_1 的估计 $\hat{\beta}_0, \hat{\beta}_1$ ——称为最小二乘估计量。由

$$\frac{\partial Q}{\partial \beta_i} = 0 \quad i = 0, 1$$

求得

$$\begin{cases} \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \\ \hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2} \\ \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \end{cases}$$

由此得到了 y 与 x 之间的回归直线

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \quad (5.1.32)$$

下面就要对我们提出的假设进行检验,看用线性关系描述 x, y 之间联系是否合理,提出如下假设检验

$$H_0: \beta_1 = 0 \quad H_1: \beta_1 \neq 0$$

我们选取检验量

$$\sqrt{n-2} \frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}/S_1} \sim t(n-2) \quad (5.1.33)$$

其中

$$\hat{\sigma}^2 = S_2^2 - \hat{\beta}_1^2 S_1^2$$

$$S_2^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2, S_1^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

若接受了 H_0 ,则认为 x, y 之间不具有线性关系;若否定了 H_0 则认为 x, y 之间可以用线性关系描述。

对于 x, y 之间某些特殊的关系,我们可以通过数据处理的方法把它们转化成线性关系予以处理,称之为曲线问题直线化,常见的有:

1° 双曲线方程: $\frac{1}{y} = a + \frac{b}{X}$

2° 幂函数方程: $y = ax^b$

3° 指数方程: $y = ae^{bx}$

4° 指数曲线方程: $y = ae^{\frac{b}{x}}$

5° 对数曲线方程: $y = a + b \log x$

6° S 型曲线方程: $y = \frac{1}{a + be^{-x}}$

7° 抛物线方程: $y = b_0 + b_1x + b_2x^2$

但是在实际中,我们一般不能确定 x, y 之间的关系,为此我们给出解决这类问题的一般步骤。

设 $(x_1, y_1), \dots, (x_n, y_n)$ 为一组观察值

1° 描点:把它们描在直角坐标系上,大约估计一下 x, y 之间的函数关系,例粗略地认为服从幂函数方程: $y = ax^b$

2° 数据处理:曲线问题直线化,注意到

$$\ln y = \ln a + b \ln x$$

这样令 $x' = \ln x, y' = \ln y$ 可见 x', y' 之间大约具有线性关系,为此处理观察值为 $(\ln x_i, \ln y_i)$ 。

3° 最小二乘法估计参数:针对 x', y' 应用最小二乘法,可求得估计值 $\hat{\ln a}, \hat{b}$,进一步地即求得了 $\hat{a} = e^{\hat{\ln a}}, \hat{b}$ 。

4° 模型假设检验:对于 x', y' 应用(5.1.33)若否定了 H_0 即可认为 x', y' 之间具有线性关系,从而也就可认为 x, y 之间具有幂函数关系。

对于多元回归模型(即一个 r, v 依赖于几个非随机变数),处理方法大致相同,可参阅一般的数理统计教本。

5.1.5 单因子方差分析方法

单因子方差分析实质上是一个多总体的假设检验问题。所谓因子联系于实验的目的,而水平则反映了与之相关的不同状态。我们进行单因子方差分析时,主要是将试验的总误差进行分解,用 F 检验,来判断各水平间是否具有显著差异。

设 $\xi_i \sim N(a_i, \sigma)$ $i = 1, 2, \dots, r$ 相互独立,现从总体 ξ_i 中抽

取容量为 n_i 的子样 $\xi_{i1}, \dots, \xi_{ini} (i = 1, 2, \dots, r)$ 考虑如下假设检验

$$H_0: a_1 = \dots = a_r \quad H_1: \text{否}$$

记

$$\bar{\xi}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \xi_{ij} \quad i = 1, 2, \dots, r$$

$$S_i^2 = \frac{1}{n_i} \sum_{j=1}^{n_i} (\xi_{ij} - \bar{\xi}_i)^2 \quad i = 1, 2, \dots, r$$

则 $\bar{\xi}_i$ 及 S_i^2 分别为第 i 个总体的子样均值及子样方差。记

$$\bar{\xi} = \frac{1}{n} \sum_{i=1}^r \sum_{j=1}^{n_i} \xi_{ij} = \frac{1}{n} \sum_{i=1}^r n_i \bar{\xi}_i$$

$$S^2 = \frac{1}{n} \sum_{i=1}^r \sum_{j=1}^{n_i} (\xi_{ij} - \bar{\xi})^2$$

$$n = \sum_{i=1}^r n_i$$

称 $\bar{\xi}$ 、 S^2 为总平均及总方差,有

$$S_{\text{总}}^2 = nS^2 = \sum_{i=1}^r n_i S_i^2 + \sum_{i=1}^r n_i (\bar{\xi}_i - \bar{\xi})^2$$

记

$$Q_e = \sum_{i=1}^r n_i S_i^2 \quad U_1 = \sum_{i=1}^r n_i (\bar{\xi}_i - \bar{\xi})^2$$

则有

$$S_{\text{总}}^2 = Q_e + U_1 \quad (5.1.34)$$

其中 $S_{\text{总}}^2$ 称总离差平方和, Q_e 为误差项平方和, U_1 为因子平方和, (5.1.34) 为总离差平方和的分解式。若 H_0 真, 则可认为所有 ξ_{ij} 出自相同总体, 故误差项平方和应占主导作用, 以下我们就通过对其大小比较, 来作出判断, 这也正是方差分析名称的由来。我们有

$$F = \frac{U_1/(r-1)}{Q_e/(n-r)} \sim F_{(r-1, n-r)} \quad (5.1.35)$$

故对显著性水平 $\alpha, 0 < \alpha < 1$ 有 $F_{(r-1, n-r)}(\alpha)$ 满足

$$P(F > F_{(r-1, n-r)}(\alpha)) = \alpha$$

否定域为 $(F_{(r-1, n-r)}(\alpha), \infty)$ 。若进行一次观察, 算得 F 值满足 $F \in (F_{(r-1, n-r)}(\alpha), \infty)$, 则说明因子误差远大于误差项误差(试验误差) 故否定 H_0 , 认为水平间有明显差异。否则, 即可认为水平间没有显著差异。这些结论在实际中都具有实际的意义: 若得到了各水平间有显著差异, 则有选优的必要, 把性能最好的水平选出; 若各水平没有显著的差异, 则有经济上的价值, 把最廉价的水平选出用于生产。

最后, 我们指出方差分析的过程, 可以通过如下的方差分析表进行:

表 5.1.1 方差分析表

项目	离差平方和	自由度	方差	F 值	临界值	结论
因子	U_1	$r - 1$	$U_1 / (r - 1)$	$U_1 / (r - 1)$	$F_{(r-1, n-r)}(\alpha)$	
误差	Q_e	$n - r$	$Q_e / (n - r)$	$Q_e / (n - r)$		
总和	$S_{\text{总}}^2$	$n - 1$				

§ 5.2 建模实例

5.2.1 钓鱼问题

为了估计湖中鱼的数目 N , 采用如下方法, 自湖中随机取出 r 条, 鱼做上标记后放回湖中然后再自湖中随机取出 S 条鱼, 结果发现这 S 条中有 x 条带有标记。试问应如何估计 N ? 又对于 $r = 300, S = 180, x = 40$ 估计湖中鱼的数目。

这是一个典型的用极大似然法思想求解的问题。设第二次取出的标有记号的鱼数为 ξ 条, ξ 是一个随机变数, 且其分布为超几

何分布:

$$P(\xi = x) = \frac{C_r^x C_{N-r}^{s-x}}{C_N^s} \quad (5.2.1)$$

其中 x 为整数, 且 $\max[0, s - (N - r)] \leq x \leq \min(r, s)$.

令

$$L(x; N) = \frac{C_r^x C_{N-r}^{s-x}}{C_N^s}$$

则据极大似然法的思想, 取使 $L(x, N)$ 达到极大值的 \hat{N} 作为 N 的估计, 为此考虑比值:

$$A(x; N) = \frac{L(x; N)}{L(x; N-1)} = \frac{N^2 - (r+s)N + rs}{N^2 - (r+s)N + Nx} \quad (5.2.2)$$

由(5.2.2)可见: 当且仅当 $N < \frac{rs}{x}$, $L(x; N) < L(x; N-1)$, 当

且仅当 $N > \frac{rs}{x}$, $L(x; N) < L(x; N-1)$ 因此, $L(x; N)$ 在 $\frac{rs}{x}$ 附近取极大值, 于是 N 的估计量为:

$$\hat{N} = \left[\frac{rs}{x} \right] \quad (5.2.3)$$

这里 $[\cdot]$ 表取整数部分

应用(5.2.3), 可以算得

$$\hat{n} = \left[\frac{300 \cdot 180}{40} \right] = 1350 (\text{条})$$

即若出现了题设的实验结果, 则湖中有 1350 条鱼的可能性最大。

5.2.2 新药药效问题

为了检验某新药的疗效, 某药厂进行了这样的试验, 让 10 个病人首先服用原来的药, 记录服用后某指标的增幅为 ξ ; 然后再让这 10 个病人服用新药, 记录某指标增幅为 η , 试验结果如表 5.2.1, 试问这种新药与原药疗效是否有显著差异? (显著性水平 $\alpha = 0.05$)

表 5.2.1

增 幅 人 药 物	1	2	3	4	5	6	7	8	9	10
ξ	0.7	-1.6	-0.2	-1.2	-0.1	3.4	3.7	0.8	0.0	2.0
η	1.9	0.8	1.1	0.1	0.1	4.4	5.5	1.6	4.6	3.4

很自然地,我们可设 $\xi \sim N(a_1, \sigma)$, $\eta \sim N(a_2, \sigma)$ 。这样原问题就化成了对其期望值的比较,即如下的假设检验问题:

$$H_0: a_1 = a_2, \quad H_1: a_1 \neq a_2$$

注意到 $n_1 = n_2 = 10$, σ^2 亦相同,故既可进行(5.1.26) 为检验量的检验,又可进行“配对法”检验,下面我们分别两种情况进行讨论。

(1) 试验数据不配对分析的结论:

选检验量

$$T = \sqrt{n-1} \frac{\bar{\xi} - \bar{\eta} - (a_1 - a_2)}{\sqrt{S_1^2 + S_2^2}} \sim t_{(2n-2)}$$

若 H_0 成立,则统计量

$$T = \sqrt{n-1} \frac{\bar{\xi} - \bar{\eta}}{\sqrt{S_1^2 + S_2^2}} \sim t_{(2n-2)}$$

对于 $\alpha = 0.05$, 有 $t_{(2n-2)}(0.05)$ 使

$$P(|T| > t_{(2n-2)}(0.05)) = 0.05$$

故其否定域为 $(-\infty, -t_{(2n-2)}(0.05)) \cup (t_{(2n-2)}(0.05), \infty)$

现知 $n = 10$, 由题可算得:

$$\bar{x} = 0.75, S_1^2 = 2.88$$

$$\bar{y} = 2.35, S_2^2 = 3.52 \quad t_{18}(0.05) = 2.10$$

从而知 T 的观察值为

$$t = \sqrt{10-1} \frac{0.75 - 2.35}{\sqrt{2.88 + 3.52}} = -1.90$$

由于 $-1.90 \in (-\infty, -2.10) \cup (2.10, \infty)$ 故不否定 H_0 , 即在显著性水平为 $\alpha = 0.05$ 条件下, 认为新药与原药没有显著差异。

(2) 试验数据配对分析的结论。

令 $Z_i = \xi_i - \eta_i$ 则

$$EZ_i = a_1 - a_2 = d$$

$$DZ_i = 2\sigma^2$$

故 Z_1, \dots, Z_n 可看作由 $Z \sim N(d, \sqrt{2}\sigma)$ 抽取子样, 于是原问题化为

$$H_0: d = 0 \quad H_1: d \neq 0$$

由于 $\sqrt{2}\sigma$ 未知, 故检验量

$$T = \sqrt{n-1} \frac{\bar{Z} - d}{S} \sim t_{(n-1)}$$

若 H_0 真, 则统计量

$$T = \sqrt{n-1} \frac{\bar{Z}}{S} \sim t_{(n-1)}$$

对于 $\alpha = 0.05$, 有 $t_{(n-1)}(0.05)$ 使

$$P(|T| > t_{(n-1)}(0.05)) = 0.05$$

故其否定域为 $(-\infty, -t_{(n-1)}(0.05)) \cup (t_{(n-1)}(0.05), \infty)$

现知 $n = 10$, 由表 5.2.1 可算得

$$\bar{\xi} = -1.6 \quad S^2 = 1.302 \quad t_9(0.05) = 2.262$$

于是可算出 T 的观察值

$$t = \sqrt{10-1} \frac{-1.60}{\sqrt{1.302}} = -4.21$$

由于 $-4.21 \in (-\infty, -2.262) \cup (2.262, \infty)$ 故否定原假设 H_0 即认为新药与原药有显著差异。

由(1)、(2)可见, 采用配对、不配对法得到了不同的检验结论, 这主要由于当采用配对法时, 突出了两种药效的差异, 消除了

不同人对药效的干扰,只要两种药有一定差异,就可能否定 H_0 ;而当数据不配对分析时,则把人之间与药之间的影响交织在一起,不易分辨药效的差异。总之,何时采取配对法检验,何时采取不配对法检验,要按处理问题的具体要求而确定。

5.2.3 婴儿出生时刻问题

为了研究婴儿出生时刻在一天内是否均匀问题,某医院 2880 名婴儿出生时刻进行了观察,记录观察值如表 5.2.2,试据此说明出生时刻均匀性问题。

表 5.2.2

小时	0	1	2	3	4	5	6	7
人数	127	139	143	138	134	115	127	113
人数	8	9	10	11	12	13	14	15
小时	126	122	121	119	130	125	112	97
人数	16	17	18	19	20	21	22	23
小时	115	94	99	97	100	119	127	139

这是一个假设检验问题,提出原假设

H_0 : 婴儿出生时刻在一天内分布均匀

H_1 : 否

则我们可得每小时婴儿出生数的理论频数为:

$$np_i = 2880 \cdot \frac{1}{24} = 120$$

由皮尔逊定理

$$\eta = \sum_{i=1}^{24} \frac{(v_i - np_i)^2}{np_i} \sim \chi^2_{(23)}$$

对于给定显著性水平 $\alpha = 0.05$, 有 $\chi^2_{(23)}^{(0.05)} = 35.2$, 从而得其否定域为 $(35.2, \infty)$

由表 2.2 可算得 η 的观察值为

$$y = 40.467$$

由于 $40.467 \in (35.2, \infty)$, 故否定原假设 H_0 , 即不能认为婴儿在天内出生时刻服从均匀分布。

若给定显著性水平 $\alpha = 0.01$, 有 $\chi_{(23)}^2(0.01) = 41.6$ 从而其否定域为 $(41.6, \infty)$, 由于 $40.467 \in (41.6, \infty)$, 故在显著性水平 $\alpha = 0.01$ 下不否定 H_0 , 即可以认为婴儿在一天内出生时刻均匀分布。

由此我们看出: 随着显著性水平 α 的减小, 否定 H_0 的可能性也就随之减小, 这主要是由于 α 越小, 你所构造的小概率事件发生的概率越小, 故这样的事情就越难发生, 从而就越不容易落入 H_0 的否定域。

5.2.4 质量控制问题

在生产过程中, 如生产正常质量稳定则质量特征的统计量应表现出稳定状态, 否则表现出不稳定状态, 根据这一原则常采用统计方法来检验和控制产品的质量。

首先考虑生产过程中的产品。若生产过程稳定, 则产品质量不会发生重大差异, 虽然如此, 即使生产过程稳定, 影响质量的因素也是会发生波动, 不过波动的幅度较小。现设期望制造直径为 a 的钢条, 由于随机波动, 制成钢条直径 ξ 为一随机变量。在假设 H_0 : “生产过程稳定” 条件下, 由中心极限定理, 可近似地假定 $\xi \sim N(a, \sigma)$

运用统计方法, 可在废品尚未出现前, 就能预告生产稳定性即将破坏, 事故即将出现, 并且在一定程度上指出毛病所在, 以便生产者及时采取行动, 防止废品出现, 这是积极的质量控制, 称为预先检查。

预先检查的基本思想: 在假设 H_0 : “生产过程稳定” 条件下, 某产品特征 ξ (例如钢条直径) 可认为服从 $N(a, \sigma)$, 对显著性水平

$\alpha, 0 < \alpha < 1$, 有 u_α 使

$$P(-u_\alpha\sigma < \xi - a < u_\alpha\sigma) = 1 - \alpha \quad (5.2.4)$$

即

$$|\xi - a| \geq u_\alpha\sigma \quad (5.2.5)$$

为一小概率事件(概率为 α), 现对一系列产品进行观察, 得子样 ξ_1, \dots, ξ_n , 以 ξ_1 代(5.2.5)中 ξ 若(5.2.5)不成立, 即可认为制成第一个抽样时 H_0 正确, 即生产在显著性水平为 α 下稳定, 同样考虑 ξ_2, \dots , 若某 ξ_k (5.2.5) 成立, 则认为在制第 k 个抽样产品时应否定 H_0 , 即在显著水平 α 下认为生产不稳定, 应立即采取修理行动。

一般地, 常数 a 预先给定, σ^2 可通过以往生产记录找到, 或以子样方差 S^2 代替, 一般取 $1 - \alpha = 95.5\%$ 或 99.73% , 相应地 $u_\alpha = 2$ 或 3 。

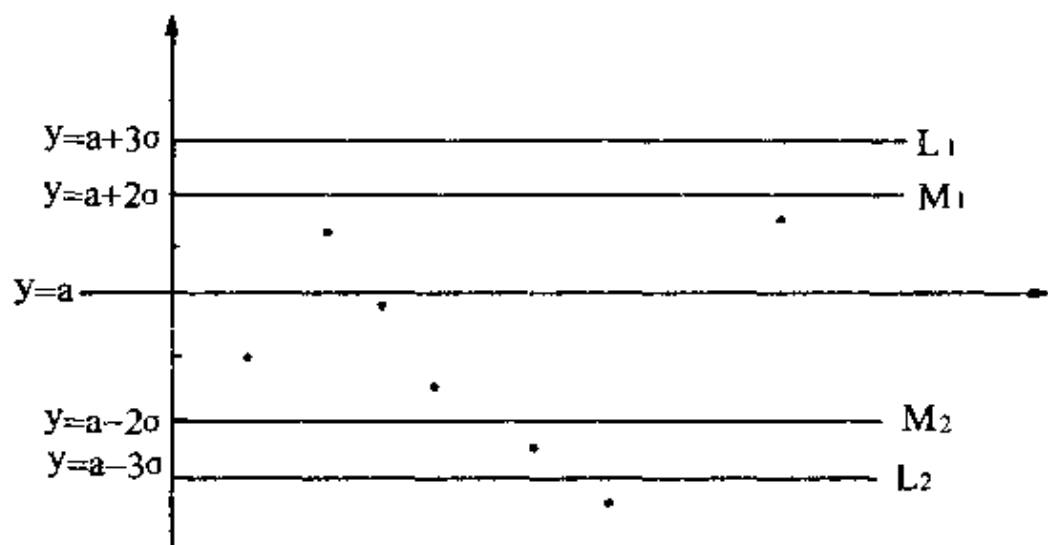


图 5.2.1

图 5.2.1 给出了质量控制图; 设 $1 - \alpha = 99.73\%$, 以 $y = a$ 为中心线, 作

$$L_1: y = a + 3\sigma$$

$$L_2: y = a - 3\sigma$$

$$M_1: y = a + 2\sigma$$

$$M_2: y = a - 2\sigma$$

称 $L_1(L_2)$ 为上(下)废品线, $M_1(M_2)$ 为上(下)控制线

当 ξ_k 跑出 L_1 与 L_2 所围地带, 生产已不稳定, 当 ξ_k 跑出 M_1 与 M_2 所围地带, 警告生产稳定有破坏的危险, 即 M_1, M_2 起警告作用。

利用控制图, 不必等到点落入控制区外, 往往就能看出问题。
例如:

若点总落在中心线以上(下), 说明可能机器安装不适, 机件受损或零件安装太紧。

若点在控制区内波动很大, 说明可能某因素(原料等)有较大波动或发动机故障。

若控制图点的连线呈周期变化, 说明可能是机器有系统故障。

若点有单调变化趋势, 说明可能是工具磨损的结果。

有时若不能认为 ξ 服从正态, 则可改用

$$\bar{\xi} = \frac{1}{n} \sum_{i=1}^n \xi_i \quad (5.2.6)$$

代替 ξ , 由中心极限定理知 $\bar{\xi}$ 近似服从 $N(a, \frac{\sigma}{\sqrt{n}})$ 在用 $\bar{\xi}$ 进行控制时, 应观察子样 $\xi_{i1}, \dots, \xi_{in} (i = 1, 2, \dots)$ 此外, 我们还考虑极差

$$R = \xi_n^* - \xi_1^*$$

的控制, 有时进行 $\bar{\xi} - R$ 控制。

预先检查在生产中是非常有用的, 它能帮助我们及时发现生产异常, 避免大量不合格品的出现, 减少浪费。

再来考虑成品的质量控制, 设有 N 个产品, 验收人员需检查每个产品是否合格, 若 N 很大, 逐一检查很难做到(事实上, 对某些产品例如炸弹的爆破强度, 灯泡的使用寿命无法做全额检验), 于是利用抽样检验, 选取 $n (n < N)$ 个个体, 以它的检验结果代替全面检验, 其优点是检验数减少, 然而它是否可真正反映总体情况呢? 易见: n 越大, 越能保证质量, 同时费用增多。因此, 选取合理的

n 是重要问题。

母体(总体)质量好坏标准为

废品率 $p = \frac{1}{N} \times \text{总体中废品个数}$

由质检要求,常给二个数 $p_1, p_2, 0 < p_1 < p_2 < 1$,若有 $p \leq p_1$,说总体合格,若 $p \geq p_2$,说总体不合格。

一般的验收方案,由总体抽取容量 n 的子样,子样废品数 d ,若 $d \leq r$,接收总体, $d > r$ 拒收总体。验收人自然希望,接收合格总体的概率不小于 $1 - \epsilon_1$ (充分大),接收不合格总体的概率不大于 ϵ_2 (充分小), $0 < \epsilon_1, \epsilon_2 < 1$ 。本着这种要求确定 n, r ,称为 (n, r) 方案。

以 $h(p)$ 表废品率为 p 时接收总体的概率, $q = 1 - p$, 则

$$h(p) = \sum_{i=0}^r \frac{C_{Np}^i C_{Nq}^{n-i}}{C_N^n} \quad (5.2.7)$$

若相对于 n 来说 N 充分大,则由超几何分布的二项近似有 $h(p)$ 近似于 $f(p)$

$$f(p) = \sum_{i=0}^r C_n^i p^i q^{n-i}$$

当 p 由 0 到 1 时, $f(p)$ 不增,连续 $f(0) = 1, f(1) = 0$ 按验收要求,有方程

$$f(p_1) = 1 - \epsilon_1 \quad (5.2.8)$$

$$f(p_2) = \epsilon_2 \quad (5.2.9)$$

由此即可解出 n, r , 执行这一方案可保证接收合格总体的概率不小于 $1 - \epsilon_1$, 接收不合格总体的概率不大于 ϵ_2 。

5.2.5 运动成绩预测问题

表 5.2.3 给出了男子 400 米跑的世界纪录变化情况,试找一个合适的公式刻划年度(Y) 与纪录(T 秒) 之间的关系,并且由此可以对将来的运动成绩进行预测。

表 5.2.3 男子 400 米世界纪录

年 Y	1910	1924	1933	1948	1960	1980
记录 T(秒)	48"3	47"5	47"	46"1	45"6	44"7

这显然是一个回归分析问题,因而我们处理的第一步就是描点、确立 Y, T 之间大致函数关系:

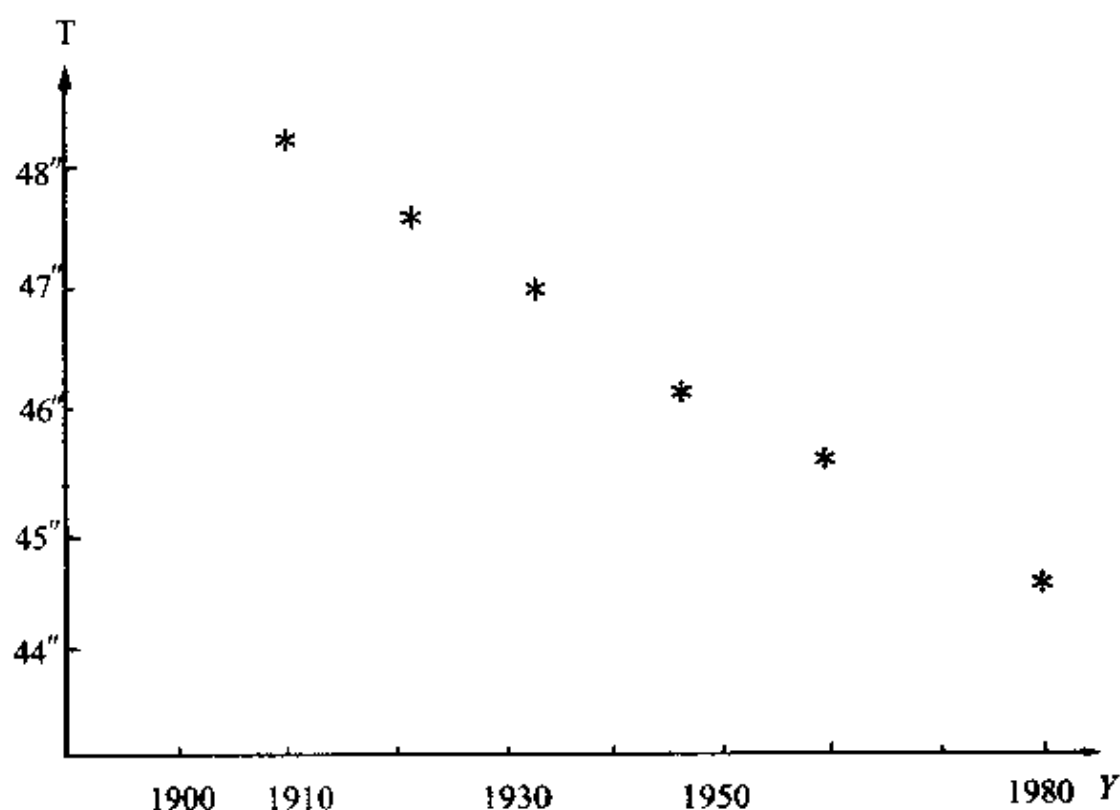


图 5.2.2

由描点可看出它近似地是一条直线,故设

$$T = \beta_0 + \beta_1 Y + \varepsilon \quad \varepsilon \sim N(0, \sigma)$$

下面就用最小二乘估计来确定未知参数估计,由(5.1.31)式可得

$$\hat{\beta}_0 = \bar{T} - \hat{\beta}_1 \bar{y} \doteq 146.63$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n y_i T_i - n \bar{y} \bar{T}}{\sum_{i=1}^n y_i^2 - n \bar{y}^2} \doteq -0.0515$$

下面我们对模型进行检验,提出原假设及备选假设

$$H_0: \beta_1 = 0 \quad H_1: \beta_1 \neq 0$$

采用(5.1.33)式

$$\sqrt{6-2} \frac{\hat{\beta}_1 - \beta_1}{\sigma/S_1} \sim t_{(4)}.$$

若 H_0 成立,则统计量

$$2 \frac{\hat{\beta}_1}{\sigma/S_1} \sim t_{(4)}$$

经查表计算知观察值落入了 H_0 的否定域 $(-\infty, -2.776) \cup (2.776, \infty)$ ($\alpha = 0.05$),故否定原假设,即认为 T, Y 之间具有线性关系,其回归直线为:

$$T = 146.63 - 0.0515Y \quad (5.2.10)$$

我们用这样的回归直接来预测当年的运动成绩,马上就会发现明显的弊病,例可算得出 2400 年男子 400 米的运动成绩大约为 23"03;3000 年的成绩竟会出现负值。这显然与实际不符,同时观察已给值我们还会发现运动成绩提高有变慢的趋势,为此我们想到应对模型进行修正,一个明显的曲线为负指数曲线:

$$T = ae^{-bY} \quad (5.2.11)$$

欲求(5.2.11)中未知参数 a, b 的估计,我们首先要采用曲线问题直线化技术:两边取对数

$$\ln T = \ln a - bY \quad (5.2.12)$$

令 $T' = \ln T$, 可见 T', Y 具有线性关系,为此进行相应的数据处理,如表 5.2.4。

表 5.2.4

Y	1910	1924	1933	1948	1960	1980
$T' = \ln T$	3.877	3.861	3.850	3.831	3.820	3.800

下面我们首先应用(5.1.31)式求得(5.2.12)式中的参数最
 α 二乘估计:

$$\ln \hat{a} = 3.886, \hat{b} \approx 0.011$$

下面我们仍用(5.1.33)式对 (T', Y) 进行检验, 得到可认为 T' , Y 具有线性关系, 回归直线为

$$T' = 3.886 - 0.011Y$$

由此得到

$$T = 48.76e^{-0.011Y} \quad (5.2.13)$$

应用这个模型得到的预测结果: $Y = 2000, T = 43.64$, 而 $Y = 2200$, 则有 $T \approx 35$ 。这比原来的模型要切合实际得多。

由此可知: 对于近期的预测, 可通过(5.2.10)来完成; 而对于较远期的预测, 则应通过(5.2.13)才会获得较切合实际的预测结果。这一点亦可留作以后去进行观察。

5.2.6 灯丝材料问题

某灯泡厂用四种不同配料方案制成灯丝, 生产了四批灯泡, 在每批灯泡中随机地抽取若干个灯泡测其使用寿命(单位: 小时), 数据列于表 5.2.5

试问这四种灯丝生产的使用寿命有无显著性差异?

设灯泡的配料为因子, 四种配料方案为水平, 则这实际给出了一个单因子四水平试验。

表 5.2.5

使用寿命 灯丝 \ 灯泡	1	2	3	4	5	6	7	8
甲	1600	1610	1650	1680	1700	1700	1780	
乙	1500	1640	1400	1700	1750			
丙	1640	1550	1600	1620	1640	1600	1740	1800
丁	1510	1520	1530	1570	1640	1680		

我们将以每一种配料方案制成灯丝后生产的灯泡视为一总体,灯泡使用寿命用 ξ 表示,则 ξ 为 r, v , 现用 $\xi_1, \xi_2, \xi_3, \xi_4$ 分别表示这四批灯泡的使用寿命,即有四个总体,若由第 i 个总体抽取容量为 n_i 的子样 $\xi_{i1}, \dots, \xi_{in_i} (i = 1, 2, 3, 4)$, 我们将应用单因子方差分析方法由这四个子样 ξ_{ij} 来推断四个总体之间有无显著性差异。

设 $\xi_i \sim N(a_i, \sigma)$, 则问题归结为判断原假设

$$H_0: a_1 = a_2 = a_3 = a_4$$

是否成立, σ^2 为未知参数。

为此我们应用(5.1.34)把总离差平方和化成二部分:误差项平方和 Q_e 及因子平方和 u_1 , 通过在一定意义下的比较,若认为误差项平方和占总离差平方和的主要部分,则接收 H_0 , 即认为材料对灯丝所生产灯泡寿命没有显著影响,否则,则认为材料配方对灯泡寿命有显著影响,

$$\text{记 } \bar{\xi}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \xi_{ij}$$

$$S_i^2 = \frac{1}{n_i} \sum_{j=1}^{n_i} (\xi_{ij} - \bar{\xi}_i)^2 \quad i = 1, 2, 3, 4$$

$$\bar{\xi} = \frac{1}{n} \sum_{i=1}^4 \sum_{j=1}^{n_i} \xi_{ij} = \frac{1}{n} \sum_{i=1}^4 n_i \bar{\xi}_i$$

$$S^2 = \frac{1}{n} \sum_{i=1}^4 \sum_{j=1}^{n_i} (\xi_{ij} - \bar{\xi})^2$$

$$n = \sum_{i=1}^4 n_i = 26$$

则有

$$S_{\text{总}}^2 = nS^2 = \sum_{i=1}^4 n_i S_i^2 + \sum_{i=1}^4 n_i (\bar{\xi}_i - \bar{\xi})^2$$

令

$$Q_e = \sum_{i=1}^4 n_i S_i^2 \quad u_1 = \sum_{i=1}^4 n_i (\bar{\xi}_i - \bar{\xi})^2$$

若 H_0 成立则有统计量

$$F = \frac{u_1/(4-1)}{Q_e/(26-4)} \sim F(3, 22)$$

于是对于显著性水平 $\alpha = 0.05$, 有 $F_{(3,22)}^{(0.05)} = 3.05$, 即

$$P(F > 3.05) = 0.05$$

得 H_0 的否定域为 $(3.05, \infty)$, 现由观察值可计算得 F 值为 2.15, 由于 $2.15 \notin (3.05, \infty)$, 故接受 H_0 , 即认为四种材料配方对灯泡寿命没有显著影响, 于是我们就可在四种配方中选取最经济的为生产所有。

下面给出针对上述问题的方差分析表 5.2.6。

表 5.2.6 方差分析表

方差来源	平方和	自由度	方差	F 值	$F_{(3,22)}^{(0.05)}$	判断
因子	44361.2	3	14787.1	2.15	3.05	不显著
误差	151350.8	22	6879.2			
总和	195712	25				

5.2.7 蠓虫识别问题

生物学家试图对两类蠓虫(A_f 与 A_{pf})进行鉴别,依据的资料是蠓虫的触角和翅膀的长度,已经测得9只 A_f 和6只 A_{pf} 的数据如下:

A_f : (1.24, 1.27), (1.36, 1.74), (1.38, 1.64), (1.38, 1.90),
(1.40, 1.70), (1.48, 1.82), (1.38, 1.82), (1.54, 1.82),
(1.56, 2.08)

A_{pf} : (1.14, 1.78), (1.18, 1.96), (1.20, 1.86), (1.26, 2.00),
(1.28, 2.00), (1.30, 1.96)

试用上述数据确定一种方法,正确区分二类蠓虫。

首先我们假设:

(1) 如果以 ξ (触角长度) 和 η (翅膀长度) 为随机变量, 则 $r.v.(\xi, \eta)$ 服从二维正态分布。

(2) A_f 类与 A_{pf} 类群体的自然总值比率为 r , 此值为已知数。

针对该问题,很自然的想法就是分别针对 A_f 与 A_{pf} 类蠓虫建立回归关系, 然后进行识别, 但这样首先遇到问题是 ξ, η 都为 $r.v.$, 而回归是针对一个非随机变量与一个 $r.v.$ 之间关系而言的, 究竟把哪一个当作非随机的呢? 另一个问题是它们之间究竟由什么样函数关系联系? 对于第二个问题, 由生物学中比率效应可知

$$\xi = a\eta^{2/3}$$

经数据处理, 可分别求得两条回归曲线, 但经过检验发现不能接受所设模型, 因此这种想法是行不通的。为此想通过未经变换数据, 应用贝叶斯公式等建模。

由设 (ξ, η) 服从正态, 其密度为:

$$A_{pf}: f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r_1^2}} \exp \left\{ -\frac{1}{2(1-r_1^2)} \left[\left(\frac{x-a_1}{\sigma_1} \right)^2 - 2r_1 \left(\frac{x-a_1}{\sigma_1} \right) \left(\frac{y-a_2}{\sigma_2} \right) + \left(\frac{y-a_2}{\sigma_2} \right)^2 \right] \right\}$$

$$A_f: g(x, y) = \frac{1}{2\pi\sigma_3\sigma_4\sqrt{1-r_2^2}} \exp \left\{ -\frac{1}{2(1-r_2^2)} \left[\left(\frac{x-a_3}{\sigma_3} \right)^2 - 2r_2 \left(\frac{x-a_3}{\sigma_3} \right) \left(\frac{y-a_4}{\sigma_4} \right) + \left(\frac{y-a_4}{\sigma_4} \right)^2 \right] \right\}$$

其中 $a_1, a_2, \sigma_1, \sigma_2, r_1; a_3, a_4, \sigma_3, \sigma_4, r_2$ 为未知参数, 因此处理的第一步是应用所提供的 A_{pf}, A_f 观察值应用矩法估计 $f(x, y), g(x, y)$ 中的 10 个未知参数。

表 5.2.7

	a_1	a_2	σ_1	σ_2	r_1
A_{pf}	1.227	1.927	0.0628	0.0882	0.653
	a_3	a_4	σ_3	σ_4	r_2
A_f	1.413	1.804	0.0990	0.130	0.593

下面我们应用贝叶斯公式求一个值为 (x, y) 的蠓虫属 A_{pf} 类的概率 P 。令

$A_1: A_{pf}$ 类蠓虫, $A_2: A_f$ 类蠓虫,

B : 值为 (x, y) 的蠓虫 (实际考虑值为落入小邻域: $[x - \Delta x, y - \Delta y] \times [x + \Delta x, y + \Delta y]$)

则有

$$P(A_1) = \frac{1}{1+r}, \quad P(A_2) = \frac{r}{1+r}$$

$$P(B | A_1) = \int_{x-\Delta x}^{x+\Delta x} \int_{y-\Delta y}^{y+\Delta y} f(u, v) du dv \approx f(x, y) 4\Delta x \Delta y$$

$$P(B | A_2) = \int_{x-\Delta x}^{x+\Delta x} \int_{y-\Delta y}^{y+\Delta y} g(u, v) du dv \approx g(x, y) 4\Delta x \Delta y$$

这样应用贝叶斯公式有

$$P = P(A_1 | B) = \frac{P(B | A_1)P(A_1)}{P(B | A_1)P(A_1) + P(B | A_2)P(A_2)}$$

$$= \frac{f(x, y)}{f(x, y) + rg(x, y)} \quad (5.2.14)$$

下面给定一个阈值 α ($0 < \alpha \leq 1$), 用(5.2.14) 进行判断:

(1) 若 $p \geq \alpha$, 则就认为蠓虫属于 A_{pf} 类

(2) 若 $p < \alpha$, 则认为属于 A_f 类

当我们进行上述类判断时, 是冒着两类错误的风险的:

(1) A_{pf} 类负误判: 误把一个 A_{pf} 类蠓虫识别为 A_f 类, 其概率为

$$\iint_{(p < \alpha)} f(x, y) dx dy \quad (5.2.15)$$

(2) A_{pf} 类正误判: 误把一个 A_f 类蠓虫识别为 A_{pf} 类, 其概率为

$$\iint_{(p \geq \alpha)} g(x, y) dx dy \quad (5.2.16)$$

易见, A_{pf} 类负误判与 A_{pf} 类正误判受到两个条件的影响, 阈值 α 的选取及自然总比率值 r 的制约, 下表给出了 $r = 1, 1.5$ 及不同 α 值下正、负误判的概率。

表 5.2.8

α	负误判(A_{pf})		正误判(A_{pf})	
	$r = 1$	$r = 1.5$	$r = 1$	$r = 1.5$
0.01	0.0	0.0	0.121	0.109
0.05	0.001	0.002	0.078	0.066
0.1	0.002	0.004	0.065	0.053
0.5	0.015	0.021	0.028	0.023
0.9	0.073	0.092	0.010	0.008

由表可见: 当选择阈值 $\alpha = 0.5$ 时, 不论 $r = 1$ 还是 $r = 1.5$

其正、负误判的概率都小于 3%,确可作为判断 A_{pf} 类或 A_f 类的准则。由于只是对两类蚊虫进行区分,选择 0.5 作为阈值是合理的。

同时,对于不同的试验目的,可以通过阈值的不同选取来实现:若 A_{pf} 是害虫,危害极大,为了不让它漏网,我们宁可把 A_f 误判为 A_{pf} ,此时就可通过减小阈值 α 的方法来实现;若 A_f 是一种珍稀物种,需加以重点保护,这是就要宁可把 A_{pf} 判为 A_f ,它可通过增大阈值的方法来达到目的。

注:这是美国 1989 年大学生建模竞赛题的主体部分。

习 题

1. 对任一地区,地震的震级 y 与其发生的次数 n 之间有经验公式 $n(y) = e^{\alpha - \beta y}$, $\alpha > 0, \beta > 0, y \geq 0$ 试按频率估计概率的原理,求震级 ξ 的分布。

2. 若从自动车床加工的一批零件中随机抽取 10 个,测得其尺寸与规定尺寸的偏差(单位:微米)分别为:2,1,-2,3,4,2,-2,5,3,4,零件尺寸的偏差记作 ξ ,设 $\xi \sim N(a, \sigma)$

(1) 试求 a, a^2 的无偏估计值

(2) 求 a 的置信水平为 0.9 的区间估计

3. 某大城市 60 小时内每小时犯罪案件数的统计如下:

10,11,10,14,11,9,12,16,13,11,10,15,11,17,10,14,11,7,21,12,13,8,10,18,12,9,15,12,12,15,10,13,8,16,12,6,13,10,8,11,9,13,14,10,22,13,11,17,15,13,8,10,15,12,19,11,9,12,12,9

估计处理每件案件平均需要 3.5 个值勤警察。警察局想保证每一时间报的案中无人分管的案件的比例不超过 2%,而不是简单地平均进行工作,为了做到这一点至少需多少警察?

提示:首先要确定一种合适的概率模型,然后确定 98% 的可

能发生的案件数。

4. 我们给某些对照组的病人服用一种看上去与要受检验的某新药相类似但并无疗效的药物,以此来衡量新药的治疗效果。由于心理作用,甚至对照组对药物亦有所反应。就某种受检药物来说,在对照组的 45 个病例中观察到药物对两个人有效。而在实际检验组的 42 个病例中观察到药物对三个人有效。若你假定新药也没疗效,那么 42 个病例中对三个人或更多人有效的概率是多少?基于这个值,你能得出什么样的结论?

5. 今要验收一批水泥:如果这种水泥制成混凝土后,断裂强度为 5000(单位),验收者希望 100 次中有 95 次被“接收”,若断裂强度为 4600(单位),验收者希望 100 次中有 10 次通过验收。已知混凝土断裂强度的均方差 $\sigma = 600$ (单位),试为验收者制定验收抽样方案。

6. 某汽车零件厂的车床加工汽缸螺栓,为了对螺栓口径(单位:厘米)进行质量控制,采取这样方案:每天抽取 8 次,每隔一小时抽取一次 10 只,测得结果如下:

(1) 8.37, 6.49, 5.39, 8.26, 7.54, 7.38, 8.00, 9.10, 7.34, 8.21

(2) 9.10, 7.28, 7.39, 5.49, 6.55, 8.10, 8.30, 6.84, 7.12, 9.00

(3) 6.57, 6.43, 5.39, 6.21, 6.10, 5.38, 7.10, 6.20, 5.48, 5.90

(4) 8.12, 8.30, 7.94, 7.32, 8.15, 8.40, 8.21, 7.00, 9.00, 6.81

(5) 9.21, 9.18, 9.30, 7.42, 9.00, 8.59, 8.80, 9.10, 7.68, 6.00

(6) 8.10, 7.48, 7.39, 6.92, 6.38, 6.54, 6.94, 9.00, 8.30, 7.10

(7) 7.21, 6.58, 8.34, 9.12, 7.66, 7.59, 8.32, 8.17, 9.00, 8.33

(8) 3.45, 7.92, 9.64, 9.82, 4.65, 6.68, 9.12, 8.00, 8.39, 7.15

设螺栓口径 $\xi \sim N(a, \sigma)$, σ, a 为未知参数,试应用 $L - M$ 质量控制线来就当天的生产进行评述。

7. 某医院用光电比色计检尿汞时,获得尿汞含量(mg/l)与消光系数的数据如下:

尿汞含量	2	4	6	8	10
消光系数	64	138	205	285	360

- (1) 试用回归分析法建立尿汞含量与消光系数的关系。
- (2) 对你所描述的模型进行检验。
- (3) 若有某人的尿汞含量为 6.8, 其消光系数的读数可能为多少?

8. 某医院应用克硒平治疗硒肺, 治疗前、中、后期患者血液中粘蛋白含量($mg\%$) 观察结果如下:

编号	1	2	3	4	5	6	7	8
治疗前	6.5	7.3	7.3	3.0	7.3	5.6	7.3	6.5
治疗中	4.5	4.4	5.9	3.6	5.5	4.5	5.2	3.8
治疗后	3.5	3.6	3.7	2.6	4.3	3.7	5.0	3.2

试问用克硒平治疗硒肺对降低血液中粘蛋白含量是否有作用($\alpha = 0.05$)?

第六章 规划方法建模

人们在经济活动中,经常会遇到在有限的资源情况下,如何合理安排,使之产值或利润最大;或在任务给定后,如何统筹安排,使之以最小成本或最小代价完成任务等决策问题。规划论就是解决这类问题的重要数学方法。本章主要介绍线性规划、整数规划、非线性规划的基本内容及在数学建模中的一些应用。

§ 6.1 预备知识

规划中的数学模型由两部分组成,一是目标函数,一是约束条件,其一般形式为:

$$\begin{aligned} & \max(\min) f(x) \\ \text{s.t.} \quad & \begin{cases} g_i(x) \geq (=, \leq) 0 \\ x_i \geq 0 \end{cases} \end{aligned}$$

根据目标函数和约束条件中函数的性质分以下几类:

6.1.1 线性规划

定义 6.1.1 规划的数学模型中如果满足:① 目标函数是决策变量的线性函数;② 约束条件都是决策变量的线性等式或不等式,则称该规划为线性规划。

1. 线性规划的一般形式和标准形式

(1) 线性规划的一般形式为

目标函数 $\max(\min) S = c_1x_1 + c_2x_2 + \cdots + c_nx_n$

$$\text{约束条件} \begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \geq (=, \leq) b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \geq (=, \leq) b_2 \\ \cdots \cdots \cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \geq (=, \leq) b_m \\ x_1, x_2, \cdots, x_n \geq 0 \end{cases}$$

可以缩写为 $\max(\min) S = \sum_{j=1}^n c_j x_j$

$$\begin{cases} \sum_{j=1}^n a_{ij}x_j \geq (=, \leq) b_i (i = 1, 2, \cdots, m) \\ x_j \geq 0 (j = 1, 2, \cdots, n) \end{cases}$$

(2) 线性规划的标准形式

从线性规划数学模型的一般形式可以看出, 目标函数可以是实现最大化, 也可以实现最小化, 约束条件可以是不等式, 也可以是等式, 这种模型形式上的多样性势必给求解带来不便, 为了便于讨论线性规划的求解方法, 我们给出规划问题的标准形式。

$$\begin{aligned} \min S &= c_1x_1 + c_2x_2 + \cdots + c_nx_n \\ \begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \\ x_1, x_2, \cdots, x_n \geq 0 \end{cases} \end{aligned}$$

可缩写为

$$\begin{aligned} \min S &= \sum_{j=1}^n c_j x_j \\ \begin{cases} \sum_{j=1}^n a_{ij}x_j = b_i (i = 1, 2, \cdots, m) \\ x_j \geq 0 (j = 1, 2, \cdots, n) \end{cases} \end{aligned}$$

标准形式的矩阵形式为

$$\min S = CX \quad \text{其中 } C = (c_1 \cdots c_n) \quad X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$\begin{cases} AX = b \\ X \geq 0 \end{cases} \quad A = \begin{bmatrix} a_{11}, a_{12} \cdots a_{1n} \\ a_{21}, a_{22} \cdots a_{2n} \\ \cdots \cdots \cdots \\ a_{m1}, a_{m2} \cdots a_{mn} \end{bmatrix} \quad b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

标准形式的向量形式为:

$$\min S = CX$$

$$\begin{cases} \sum_{j=1}^n p_j x_j = b \\ x_j \geq 0 (j = 1, 2, \cdots, n) \end{cases} \quad \text{其中 } p_j = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \cdots \\ a_{mj} \end{pmatrix} (j = 1, 2, \cdots, n)$$

(3) 如何把线性规划的一般形式化为标准形式

1° 目标函数为最大化类型即 $\max S = CX$, 把目标函数化为最小化类型, 只需令 $S' = -S$, 则 $\max S = -\min(-S) = -\min S'$ 即 $\min S' = -CX$.

2° 化约束不等式为约束等式只需引入新的非负变量, 以表示不等式左、右两端间的差额即可, 这些新变量称为松弛变量(或剩余变量)。

如: 当第 i 个约束条件为

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n \leq b_i$$

此时在不等式两端加上松弛变量 x_{n+i} 使之成为 $a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n + x_{n+i} = b_i$

当第 k 个约束为 $a_{k1}x_1 + a_{k2}x_2 + \cdots + a_{kn}x_n \geq b_k$ 时, 减去松弛变量 x_{n+k} , 使之成为 $a_{k1}x_1 + a_{k2}x_2 + \cdots + a_{kn}x_n - x_{n+k} = b_k$

注意,松弛变量与线性规划中的原变量一样,是问题的一部分,在计算过程中也始终保持非负。

3° 某变量 x_k 无符号限制,即可以取正值也可以取负值时,为满足标准形式对变量的非负要求,通常可用两个变量之差替换,即令 $x_k = x_k' - x_k''$,其中 $x_k'', x_k' \geq 0$ 。

4° 约束条件右端常数为负数时,对于等式约束,只需两边同乘 -1 ,对于不等式约束,同乘 -1 后,同时改变不等号方向即可。

例 试将下列线性规划化成标准形式

$\max S = x_1 - 3x_2 + 2x_3$ 其标准形式为

$$S. t. \begin{cases} 2x_1 + x_2 + x_3 \leq 10 \\ x_1 - 4x_2 + 2x_3 \geq 3 \\ 3x_1 - x_2 - x_3 = -5 \\ x_1, x_2 \geq 0, x_3 \text{ 无限制} \end{cases} \quad \min S' = -x_1 + 3x_2 - 2x_3' + 2x_3''$$

$$S. t. \begin{cases} 2x_1 + x_2 + x_3' - x_3'' + x_4 = 10 \\ x_1 - 4x_2 + 2x_3' - 2x_3'' - x_5 = 3 \\ -3x_1 - x_2 + x_3' - x_3'' = 5 \\ x_1, x_2, x_3', x_3'', x_4, x_5 \geq 0 \end{cases}$$

2. 线性规划解的概念

对于线性规划问题的标准形式

$$\min S = \sum_{j=1}^n c_j x_j \quad (1)$$

$$S. t. \begin{cases} \sum_{j=1}^n a_{ij} x_j = b_i (i = 1, 2, \dots, m) \\ x_j \geq 0 (j = 1, 2, \dots, n) \end{cases} \quad (2) \quad (3)$$

可行解:满足约束条件(2),(3)的解 $X = (x_1, x_2 \dots x_n)^T$ 称为线性规划的可行解,所有可行解的集合称为可行域。

最优解,满足(1)的可行解称为线性规划的最优解,对应于最优解的目标函数值称为最优值。线性规划的实质就是要从许许多多的可行解中,找出一个可行解,使目标函数达到最优值。

基:设 A 是约束方程组(2)的系数矩阵。其秩 $r(A) = m < n$,如果 B 是矩阵 A 的 m 阶非奇异的子矩阵,即 $|B| \neq 0$,称为矩

阵为线性规划问题的一个基。基 B 是由 m 个线性无关的列向量组成的,不妨设 A 中前 m 个列向量线性无关,则

$$B = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mm} \end{pmatrix} = (p_1 p_2 \cdots p_m)$$

基向量和基变量:组成基 B 的列向量 $p_j (j = 1, 2, \cdots, m)$ 称为基向量。与基向量对应的决策变量 $x_j (j = 1, 2, \cdots, m)$ 称为基变量,其余 $n - m$ 个变量称为非基变量。

设基变量用 x_B 表示,非基变量用 x_N 表示,则 $X = (x_B, x_N)^T$

基本解:在满足(2)的解中如果取非基变量 $x_N = 0$,则称此解 $X = (x_B, 0)^T$ 为对应于基 B 的基本解。

一个基就有一个基本解与之对应。一个线性规划问题的基本解是有限的,它不超过 C_n^m 个。由于系数矩阵 A 的秩为 m ,所以每一个基本解的非零分量不超过 m 。

基可行解和可行基:满足非负约束条件(3)的基本解,称为基可行解。相应的基 B 称为可行基。一个线性规划问题的基可行解的个数不超过基本解的个数。

基本最优解和最优基:使目标函数(1)达到最优值的基可行解,称为基本最优解。相应的基 B 称为最优基。

下图进一步说明了线性规划问题的解之间的关系。

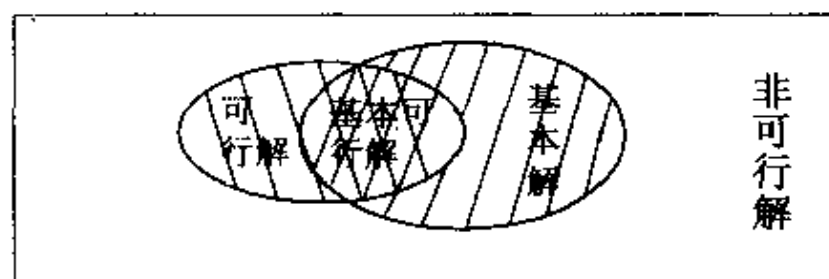


图 6.1.1

3. 线性规划的基本解法 —— 单纯形法

单纯形法是求解线性规划的基本方法。其基本思路为：首先求出线性规划的一个可行基及该基所对应的单纯形表，其次判定该基是否为最优基，若是，该基所对应的解即为最优解，求解终止。若不是，通过换基迭代，求出线性规划的另一个可行基，再重复上述步骤，如此继续，直至求出最优基为止。

对于线性规划问题 $\min S = CX$

$$s.t. \begin{cases} AX = b \\ X \geq 0 \end{cases}$$

如果有一个可行基为 B ，对应的基变量为 x_B ，基变量对应的目标函数中的系数为 C_B ，称 $C_B B^{-1}A - C$ 为该线性规划的检验数。称由

$$\begin{bmatrix} C_B B^{-1}b & C_B B^{-1}A - C \\ B^{-1}b & B^{-1}A \end{bmatrix}$$

所组成的表为基 B 对应的单纯形表，一般用 $T(B)$ 表示。

定理 6.1.1 (最优解判定定理)

若 x^0 为基 B 对应的基本可行解，而且检验数 $C_B B^{-1}A - C \leq 0$ ，则 x^0 为线性规划问题的最优解。此时基 B 称为最优基。

单纯形法的计算步骤：

(1) 找出线性规划问题的一个初始可行基，如果系数矩阵 A 中有单位矩阵的基，则该基一定是初始可行基，如果没有单位矩阵的基，可以用两阶段法求人造基，关于两阶段法下边再介绍。

(2) 求出初始可行基，比如 B 所对应的初始单纯形表 $T(B) =$

$$\begin{bmatrix} C_B B^{-1}b & C_B B^{-1}A - C \\ B^{-1}b & B^{-1}A \end{bmatrix} \text{ 详细为表 6.1 所示。}$$

表 6.1.1

		x_1	x_2	$x_3 \cdots$	$x_S \cdots$	x_n
S	b_{00}	b_{01}	b_{02}	$b_{03} \cdots$	$b_{0S} \cdots$	b_{0n}
x_{j1}	b_{10}	b_{11}	b_{12}	$b_{13} \cdots$	$b_{1S} \cdots$	b_{1n}
x_{j2}	b_{20}	b_{21}	b_{22}	$b_{23} \cdots$	$b_{2S} \cdots$	b_{2n}
\vdots	\vdots	\vdots		\vdots		\vdots
x_{jr}	b_{r0}	b_{r1}	b_{r2}	$b_{r3} \cdots$	$b_{rS} \cdots$	b_{rn}
\vdots	\vdots	\vdots		\vdots	\vdots	\vdots
x_{jm}	b_{m0}	b_{m1}	b_{m2}	$b_{m3} \cdots$	$b_{mS} \cdots$	b_{mn}

其中 $x_{j1} \cdots x_{jm}$ 为基变量, $C_B B^{-1} b = b_{00}$

$C_B B^{-1} A - C = (b_{01} b_{02} \cdots b_{0n})$ 为检验数

$$B^{-1} b = \begin{bmatrix} b_{10} \\ \vdots \\ b_{m0} \end{bmatrix} \text{ 为基变量的取值, } B^{-1} A = \begin{bmatrix} b_{11} b_{12} \cdots b_{1n} \\ \cdots \cdots \cdots \\ b_{m1} b_{m2} \cdots b_{mn} \end{bmatrix}$$

(3) 用最优解判定定理判定该基是否为最优基,若是最优基,求解终止,若不是最优基,即检验数中有正分量时,要换基迭代。

(4) 换基迭代

1° 为方便起见,从检验数中最左边的一个正分量开始,比如为 b_{0s} ,此时 x_s 要由原来的非基变量变为基变量,称为调入变量,通过轴心项确定调出变量,即由原来的基变量变为非基变量。

2° 求轴心项:由 b_{0s} 所在的列中所有的正分量 b_{js} 去除其对应的基变量的值 b_{j0} ,取其中最小者,即 $\theta_{rs} = \min \left\{ \frac{b_{j0}}{b_{js}} \mid b_{js} > 0 \right\}$,此时 b_{rs} 为轴心项, x_{jr} 为调出变量,如果出现几个同时为最小时,

可取最上边的一个。

3° 换基迭代:把轴心项 b_{rs} 用“□”框起来,记为 $\boxed{b_{rs}}$,在原单纯形表上将第 r 行乘以 $\frac{1}{b_{rs}}$,使 b_{rs} 变为 1,然后将第 r 行乘以 $(-b_{is})$ 加到第 i 行($i = 0, 1, \dots, m, i \neq r$),于是第 s 列中的其他元素变成 0,使得第 s 列成为单位列向量,此时所得到的单纯形表为新基 B_1 所对应的 $T(B_1)$ 。

4° 对新基 B_1 再判定是否最优,即重复(3)以后的步骤,直到求出最优解为止。

在最优基所对应的单纯形表中,基变量所对应的检验数全部为零,如果非基变量所对应的检验数全部为负,此时线性规划有惟一最优解。如果非基变量所对应的检验数有等于零的,此时线性规划有无穷多最优解。如果可行基对应的单纯形表中正检验数所在列中无正分量,此时线性规划问题无有界最优解。

例 1 $\max S = 4x_1 + 9x_2 - 10x_3$

$$S. t. \begin{cases} -x_1 + x_2 + 3x_3 \leq 6 \\ x_1 + 2x_2 + 7x_3 \leq 21 \\ x_1, x_2, x_3 \geq 0 \end{cases}$$

解:首先将问题标准化

$$\min S' = -4x_1 - 9x_2 + 10x_3$$

$$S. t. \begin{cases} -x_1 + x_2 + 3x_3 + x_4 = 6 \\ x_1 + 2x_2 + 7x_3 + x_5 = 21 \\ x_1, x_2, x_3, x_4, x_5 \geq 0 \end{cases}$$

取初始可行基

$$B = (p_4 p_5) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

求初始单纯形表 $T(B)$,并进行换基迭代。

表 6.1.2

		x_1	x_2	x_3	x_4	x_5
S'	0	4	9	-10	0	0
x_4	6	-1	1	3	1	0
x_5	21	$\boxed{1}$	2	7	0	1
S'	-84	0	1	-38	0	-4
x_4	27	0	$\boxed{3}$	10	1	1
x_1	21	1	2	7	0	1
S'	-93	0	0	$-\frac{124}{3}$	$-\frac{1}{3}$	$-\frac{13}{3}$
x_2	9	0	1	$\frac{10}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
x_1	3	1	0	$\frac{1}{3}$	$-\frac{2}{3}$	$\frac{1}{3}$

由此可知原问题的最优解为 $x = (3, 9, 0)$, 最优值 $S = 93$, 最优基 $B_1 = (p_2 p_1)$, 有惟一最优解。

例 2 $\max S = 3x_1 + x_2$

$$S. t. \begin{cases} x_1 + x_2 \leq 4 \\ -x_1 + x_2 \leq 2 \\ 6x_1 + 2x_2 \leq 18 \\ x_1, x_2 \geq 0 \end{cases}$$

解: 将问题标准化

$$\min S' = -3x_1 - x_2$$

$$S. t. \begin{cases} x_1 + x_2 + x_3 = 4 \\ -x_1 + x_2 + x_4 = 2 \\ 6x_1 + 2x_2 + x_5 = 18 \\ x_1, x_2, x_3, x_4, x_5 \geq 0 \end{cases}$$

取初始可行基为 $B = (p_3, p_4, p_5)$, 求初始单纯形表 $T(B)$ 并进行换基迭代。

表 6.1.3

		x_1	x_2	x_3	x_4	x_5
S'		3	1	0	0	0
x_3	4	1	1	1	0	0
x_4	2	-1	1	0	1	0
x_5	18	6	2	0	0	1
S'	-9	0	0	0	0	$-\frac{1}{2}$
x_3	1	0	$\frac{2}{3}$	1	0	$-\frac{1}{6}$
x_4	5	0	$\frac{4}{3}$	0	1	$\frac{1}{6}$
x_1	3	1	$\frac{1}{3}$	0	0	$\frac{1}{6}$

由此可知原问题的最优解 $x = (3, 0, 1, 5, 0)$, 最优值为 $S = 9$, 最优基为 $B_1 = (p_1, p_4, p_5)$, 因为非基变量 x_2 的检验数为 0, \therefore 该问题有无限多个最优解, 比如以 $\frac{2}{3}$ 为轴心项, 再换基迭代又得最优解, $x_1 = (\frac{5}{2}, \frac{3}{2}, 0, 3, 0)$ 最优值仍为 9, 由线性规划的理论, 点 $(3, 0)$ 与 $(\frac{5}{2}, \frac{3}{2})$ 连线上的所有点都是该问题的最优解。

4. 两阶段法

线性规划问题中如果有单位矩阵的初始可行基, 直接用上述的单纯形法即可, 如果没有单位矩阵的基, 如何求第一个初始可行基呢? 一般可用两阶段法。

两阶段法的基本思想为: 首先作出一个辅助线性规划问题, 使辅助问题中有单位初始可行基, 也称为人造基。由该基出发求出辅助问题的最优解, 然后再过渡到第二阶段, 求原问题的最优解。

设有线性规划问题

$$\begin{aligned} \min S &= c_1x_1 + c_2x_2 + \dots + c_nx_n \\ \text{s.t. } \left\{ \begin{array}{l} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n = b_m \\ x_1, x_2, \dots, x_n \geq 0 \end{array} \right. \end{aligned}$$

首先假设 $b_i \geq 0, (i = 1, 2, \dots, m)$ (若 $b_i \leq 0$, 两边乘 -1 即可), 且无单位可行基。

作辅助问题

$$\begin{aligned} \min Z &= y_1 + y_2 + \dots + y_m \\ S.t. \quad &\left\{ \begin{array}{l} S - c_1x_1 - c_2x_2 \cdots - c_n x_n = 0 \\ y_1 + a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ y_2 + a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \vdots \\ y_m + a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \\ x_1 \cdots x_n \geqslant 0, y_1 \cdots y_m \geqslant 0 \end{array} \right. \end{aligned}$$

该问题有单位初始可行基 $B = (p_1 p_2 \cdots p_{m+1})$ 该基称为人造基, 变量 $y_i (i = 1, 2, \cdots, m)$ 称为人工变量。

人造基 B 对应的单纯形表如下。

表 6.1.4

		s	y_1	$y_2 \cdots y_m$	x_1	$x_2 \cdots x_n$
Z		0	0	$0 \cdots 0$	$\sum_{i=1}^m a_{i1}$	$\sum_{i=1}^m a_{i2} \cdots \sum_{i=1}^m a_{in}$
S	0	1	0	$0 \cdots 0$	$-c_1$	$-c_2 \cdots -c_n$
y_1	b_2	0	1	$0 \cdots 0$	a_{11}	$a_{12} \cdots a_{1n}$
y_2	b_2	0	0	$1 \cdots 0$	a_{21}	$a_{22} \cdots a_{2n}$
\vdots	\vdots	\vdots	\vdots	$\vdots \cdots \vdots$	\vdots	$\vdots \cdots \cdots$
y_m	b_m	0	0	$0 \cdots 1$	a_{m1}	$a_{m2} \cdots a_{mn}$

对此用上述的单纯形法换基迭代求出辅助问题的最优基 B^* , 此时第一阶段完成, 然后向第二阶段转化。转化时分以下几种情况:

(1) 如果对应于最优基 B^* 的目标函数值 $\min Z > 0$, 则原线性规划问题无可行解。

(2) 若对应于最优基 B^* 的目标函数值 $\min Z = 0$, 显然 $y_i = 0 (i = 1, 2, \dots, m)$, 这时如果 B^* 对应的基变量中不含人工变量, 此时将 B^* 对应的单纯形表中前 $(m+1)$ 列和含有 Z 的第一行划掉, 剩余的表即为原问题的一个可行基所对应的单纯形表, 然后按照上述的单纯形法求原问题的最优解。

(3) 若对应于最优基 B^* 的目标函数值 $\min Z = 0$, 而 B^* 对应的基变量中含有人工变量, 比如 y_r , 且 $T(B^*)$ 中第 r 行对应的方程为 $y_r + \sum_{k \in K} b_{r,k+1} y_k = 0$, 其中 y_r 为基变量, $y_k (k \in K)$ 为非基变量, 这表明原问题的约束方程中第 r 个方程是多余的, 可以去掉, 辅助问题中含有 y_r 的方程可以去掉, 此时在 $T(B^*)$ 去掉 y_r 为所在的行, 成为 2 的情况, 由 2 的方法求解即可。

(4) 若对应于最优基 B^* 的目标函数值 $\min Z = 0$, B^* 对应的基变量中有人工变量比如 y_r , 且 $T(B^*)$ 的第 r 行对应的方程为 $y_r + \sum_{k \in K} b_{r,k+1} y_k + \sum_{j \in J} b_{r,j+m+1} x_j = 0$, 其中 y_r 为基变量, $y_k (k \in K)$, $x_j (j \in J)$ 为非基变量, 且 $b_{r,j+m+1} (j \in J)$ 中至少有某 $b_{rs} \neq 0$, 则以 b_{rs} 为轴心项进行换基迭代 (用 x_s 代替 y_r), 得新基 \bar{B}^* 及 $T(\bar{B}^*)$ 。这时新基中减少了一个人工变量, 经过有限次迭代, 一定可以去掉基底中所有人工变量, 化成情形 2, 再由 2 的方法求原问题的最优解。

例 3 $\min S = x_1 - x_2$

$$S. t. \begin{cases} -x_1 + x_2 - x_3 = 2 \\ -x_1 - x_2 - x_4 = 1 \\ x_j \geq 0, j = 1, 2, 3, 4 \end{cases}$$

解:作辅助问题

$$\min Z = y_1 + y_2$$

$$S. t. \begin{cases} S - x_1 + x_2 = 0 \\ y_1 - x_1 + x_2 - x_3 = 2 \\ y_2 - x_1 - x_2 - x_4 = 1 \\ x_1 \cdots x_4 \geq 0, y_1, y_2 \geq 0 \end{cases}$$

取人造基 $B = (p_1 p_2 p_3)$, 求 $T(B)$ 如下。

表 6.1.5

		S	y_1	y_2	x_1	x_2	x_3	x_4
Z	3	0	0	0	-2		-1	-1
S		1	0	0	-1	1	0	0
y_1	2	0	1	0	-1	1	-1	0
y_2	1	0	0	1	-1	-1	0	-1

由此表知 B 为辅助问题的最优基, 且 $\min Z = 3 > 0$, \therefore 原问题无可行解。

例 4 $\min S = -x_1 + 2x_2 + x_3$

$$S. t. \begin{cases} -2x_1 + x_2 - x_3 + x_4 = 4 \\ x_1 + 2x_2 = 6 \\ x_j \geq 0, j = 1, 2, 3, 4 \end{cases}$$

解:作辅助问题

$$\min Z = y_1 + y_2$$

$$S. t. \begin{cases} s + x_1 - 2x_2 - x_3 = 0 \\ y_1 - 2x_1 + x_2 - x_3 + x_4 = 4 \\ y_2 + x_1 + 2x_2 = 6 \\ x_j \geq 0, j = 1, 2, 3, 4, y_1, y_2 \geq 0 \end{cases}$$

取基 $B = (p_1 p_2 p_3)$ 用单纯形法, 可得辅助问题的最优基 B^* 及 $T(B^*)$ 。

表 6.1.6

		S	y_1	y_2	x_1	x_2	x_3	x_4
Z	0	0	-1	-1	0	0	0	0
S	6	1	0	1	2	0	-1	0
x_4	1	0	1	$\frac{1}{2}$	$-\frac{2}{5}$	0	-1	0
x_2	3	0	1	$\frac{1}{2}$	$\frac{1}{2}$	1	0	0

此时 $\min Z = 0$, 且基变量中不含人工变量, 因此 x_4, x_2 对应的列向量为原问题的第一个可行基 B' , 去掉上表的前三列和含 Z 的第一行, 可得相应的单纯形表 $T(B')$, 如表 6.1.7。

表 6.1.7

		x_1	x_2	x_3	x_4
S	6	2	0	-1	0
x_4	1	$-\frac{5}{2}$	0	-1	0
x_2	3	$\frac{1}{2}$	1	0	0

$\because b_{01} = 2 > 0, \therefore B'$ 不是原问题的最优基, 再用单纯形法求得最优解为 $x^* = (6, 0, 0, 16)$ 。

例 5 $\min S = 4x_1 + 3x_3$

$$S. t. \begin{cases} \frac{1}{2}x_1 + x_2 + \frac{1}{2}x_3 - \frac{2}{3}x_4 = 2 \\ \frac{3}{2}x_1 + \frac{3}{4}x_3 = 3 \\ 3x_1 - 6x_2 + 4x_4 = 0 \\ x_j \geq 0, j = 1, 2, 3, 4 \end{cases}$$

解:作辅助问题

$$\min Z = y_1 + y_2 + y_3$$

$$S. t. \begin{cases} s - 4x_1 - 3x_3 = 0 \\ y_1 + \frac{1}{2}x_1 + x_2 + \frac{1}{2}x_3 - \frac{2}{3}x_4 = 2 \\ y_2 + \frac{3}{2}x_1 + \frac{3}{4}x_3 = 3 \\ y_3 + 3x_1 - 6x_2 + 4x_4 = 0 \\ x_j \geq 0, j = 1, 2, 3, 4, y_i \geq 0, i = 1, 2, 3 \end{cases}$$

取基 $B = (p_1 p_2 p_3 p_4)$ 用单纯形法求辅助问题得最优基 B^* 及 $T(B^*)$ 。

表 6.1.8

		S	y_1	y_2	y_3	x_1	x_2	x_3	x_4
Z	0	0	$-\frac{5}{2}$	0	$-\frac{5}{4}$	0	0	0	0
S	8	1	4	0	$\frac{2}{3}$	0	0	-1	0
x_2	1	0	$\frac{1}{2}$	0	$-\frac{1}{12}$	0	1	$\frac{1}{4}$	$-\frac{2}{3}$
y_2	0	0	$-\frac{2}{3}$	1	$-\frac{1}{4}$	0	0	0	0
x_1	2	0	1	0	$\frac{1}{6}$	1	0	$\frac{1}{2}$	0

因为此时 $\min Z = 0$, B^* 的基变量中含有人工变量 y_2 , 由

$T(B^*)$ 可得 $y_2 = \frac{2}{3}y_1 + \frac{1}{4}y_3$, 这表明原问题中第二个约束方程是多余的可以去掉, 因此 x_2x_1 所对应的列向量构成原问题的可行基 $B_1 = (p_2p_1)$, 去掉 $T(B^*)$ 中的前 4 列及含 Z 的第一行, 可得 $T(B_1)$ 。

表 6.1.9

		x_1	x_2	x_3	x_4
S	8	0	0	-1	0
x_2	1	0	1	$\frac{1}{4}$	$-\frac{2}{3}$
x_1	2	1	0	$\frac{1}{2}$	0

由此可得原问题的最优解为 $x^* = (2, 1, 0, 0)$

例 6 $\min S = 4x_1 + 3x_3$

$$S. t. \begin{cases} \frac{1}{2}x_1 + x_2 + \frac{1}{2}x_3 - \frac{2}{3}x_4 = 2 \\ \frac{3}{2}x_1 - \frac{1}{2}x_3 = 3 \\ 3x_1 - 6x_2 + 4x_4 = 0 \\ x_j \geq 0, j = 1, 2, 3, 4 \end{cases}$$

解: 作辅助问题

$$\min Z = y_1 + y_2 + y_3$$

$$S. t. \begin{cases} s - 4x_1 - 3x_3 = 0 \\ y_1 + \frac{1}{2}x_1 + x_2 + \frac{1}{2}x_3 - \frac{2}{3}x_4 = 2 \\ y_2 + \frac{3}{2}x_1 - \frac{1}{2}x_3 = 3 \\ y_3 + 3x_1 - 6x_2 + 4x_4 = 0 \\ x_j \geq 0, j = 1, 2, 3, 4, y_i \geq 0, i = 1, 2, 3 \end{cases}$$

取基 $B = (p_1 p_2 p_3 p_4)$ 用单纯形法求出辅助问题的最优基 B^* 及 $T(B^*)$ 。

表 6.1.10

		S	y_1	y_2	y_3	x_1	x_2	x_3	x_4
Z	0	0	$-\frac{5}{2}$	0	$-\frac{5}{4}$	0	0	$-\frac{5}{4}$	0
S	8	1	4	0	$\frac{2}{3}$	0	0	-1	0
x_2	1	0	$\frac{1}{2}$	0	$-\frac{1}{12}$	0	1	$\frac{1}{4}$	$-\frac{2}{3}$
y_2	0	0	$-\frac{3}{2}$	1	$-\frac{1}{4}$	0		$-\frac{5}{4}$	0
x_1	2	0	1	0	$\frac{1}{6}$	1	0	$\frac{1}{2}$	0

$\because \min Z = 0, B^*$ 的基变量中有人工变量 y_2 , 对应表中的方程为 $y_1 = \frac{2}{3}y_1 + \frac{1}{4}y_3 + \frac{5}{4}x_3$, 因为 $b_{37} \neq 0$, \therefore 再以 b_{37} 为轴心项换基迭代得新基 \bar{B}^* 及 $T(\bar{B}^*)$ 。

表 6.1.11

		S	y_1	y_2	y_3	x_1	x_2	x_3	x_4
Z	0	0	-1	-1	-1	0	0	0	0
S	8	1	$\frac{26}{5}$	$-\frac{4}{5}$	$\frac{13}{5}$	0	0	0	0
x_2	1	0	$\frac{1}{5}$	$\frac{1}{5}$	$-\frac{2}{15}$	0	1	0	$-\frac{2}{3}$
x_3	0	0	$\frac{6}{5}$	$-\frac{4}{5}$	$\frac{1}{5}$	0	0	1	0
x_1	2	0	$\frac{2}{5}$	$\frac{2}{5}$	$\frac{1}{15}$	1	0	0	0

此时已变成 2 中情况, 去掉表中前 4 列与含 Z 的第一行, 得原问题的最优解为 $x^* = (2, 1, 0, 0)$

把上述的两阶段法用框图表示如下:

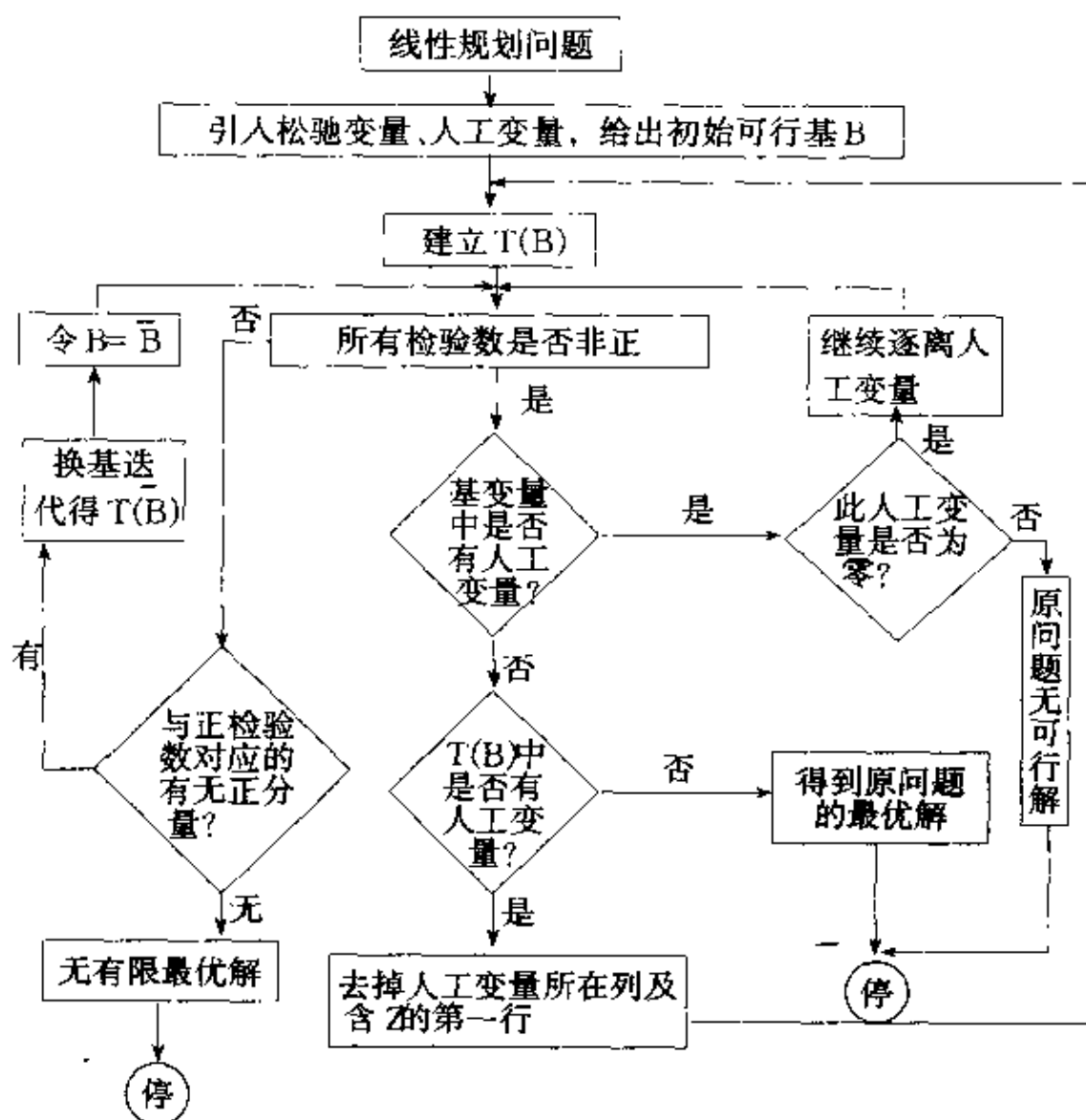


图 6.1.2

5. 对偶线性规划

定义 6.1.2 设线性规划问题

$$\begin{aligned} \min S &= cx \\ S. t. \quad &\begin{cases} Ax \geq b \\ x \geq 0 \end{cases} \end{aligned}$$

由此可得线性规划

$$\begin{aligned} \max g &= b^T y, \\ \text{s.t. } \begin{cases} A^T y \leq c^T \\ y \geq 0 \end{cases} \end{aligned}$$

则称前二者互为对偶线性规划,

$$\text{其中 } A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \quad c = (c_1, c_2, \cdots, c_n)$$

$$b = (b_1, b_2, \cdots, b_m)^T \quad x = (x_1, x_2, \cdots, x_n)^T \quad y = (y_1, y_2, \cdots, y_m)^T$$

对偶线性规划有以下性质:

性质 1 线性规划问题的对偶问题的对偶即为原线性规划。

性质 2 若线性规划(I)的第 k 个约束条件是等式,则它的对偶规划(II)的第 k 个变量无非负限制,反之也成立。

性质 3 设 x, y 分别为互为对偶问题(I)和(II)的可行解,则对应的目标函数之间的关系为 $s \geq g$,即(I)的最小值的目标函数值不小于(II)的最大的目标函数值。

性质 4 如果 x^*, y^* 分别是互为对偶问题(I)和(II)的可行解,且 $cx^* = b^T y^*$,则 x^*, y^* 分别为(I)和(II)的最优解。

性质 5 如果原问题(I)有最优解,则对偶问题(II)也有最优解,且最优值相等,反之亦然。

对偶单纯形法:

定义 6.1.3 设 B 为线性规划 $\text{s.t. } \begin{cases} \min S = cx \\ Ax = b \\ x \geq 0 \end{cases}$ 的一个基, $x = \begin{pmatrix} B^{-1}b \\ 0 \end{pmatrix}$ 为相应于基 B 的基本解,如果满足 $C_B B^{-1} A - C \leq 0$,则称 $x = \begin{pmatrix} B^{-1}b \\ 0 \end{pmatrix}$ 为线性规划的基本对偶解,相应的基 B 称为对偶可行基。

对偶单纯形法的一般步骤:

(1) 取初始对偶基 $B(C_B B^{-1}A - C \leq 0)$, 计算 $T(B)$ 。

(2) 判断 $B^{-1}b$ 是否非负。若 $B^{-1}b \geq 0$, 则 $x = \begin{pmatrix} B^{-1}b \\ 0 \end{pmatrix}$ 是最优解, 算法终止, 否则转(3)。

(3) 若 $B^{-1}b$ 中有负值, 且负值对应的行向量中的所有分量均非负, 则线性规划无解。否则, 定指标 r 使 $b_{r0} = \min\{b_{i0} \mid b_{i0} < 0\}$ 转(4)。

(4) 让 b_{r0} 对应的行向量中各负的分量分别去除它对应的检验数 b_{0j} , 取 $\frac{b_{0s}}{b_{rs}} = \min\left\{\frac{b_{0j}}{b_{rj}} \mid b_{rj} < 0\right\}$ 称 b_{rs} 为对偶轴心项, 以 b_{rs} 为轴心项进行换基迭代得新基 \bar{B} 及 $T(\bar{B})$ 转(2)。

实际问题中的线性规划模型一般决策变量和约束条件的个数都比较多。关于线性规划的求解问题现已有不同类别的软件包可供使用, 现在线性规划的求解都是用软件包在计算机上直接实现。

6. 应用时注意的问题

线性规划是有关规划、计划等最优决策工作中最常用的建模方法之一, 下面仅从实际应用的角度, 对建模、求解、方案评价等问题进行初步的综合讨论。

(1) “静态” 最优决策

以线性规划的数学问题本身而言, 其变量和系数不是时间的函数, 它是用来描述“静态” 优化问题的, 但是客观世界中绝对“静态” 的事件是没有的, 只是有时候事件的变化周期远远长于所研究的时间范围。这时, 可以用“静态” 模型去描述, 有时也可以采用分段的多个“静态” 优化问题去逼近一个“动态” 问题。

(2) 避免有用变量成为“0”

以线性规划方法建模, 用以确定最优产品结构或企业生产计划等类问题时, 各种资源限制及生产工艺限制等约束条件数一般

会远远小于决策变量(待确定产量的产品种类数),取得最优解时,很多决策变量等于“0”,这显然是不符合实际的。这一问题的解决,应当是引入各决策变量的上、下界约束(各类产品的最低或最高需求量),这时,变量有界的单纯形算法将显示较大的优越性。

(3) 变无可行解为有可行解

多数关于线性规划的论著中,都是在有最优解的前提条件下来讨论解的性质和求解方法的。在各种实际应用中,系统分析后所建立的模型,难于—建就成,很多情况是所建的模型无可行解,需经多次调整才能成为有可行解的满意的适用的模型。所以,在实际应用中更为关心的是:当所建立的模型无可行解时,应对模型参数的哪些调整,方可使模型有可行解。经分析,造成无可行解的原因在于所引入的各种约束限制太苛刻以至于相互矛盾。解决这类问题的有效办法是对各个约束条件的限制量给出其调整方向及松动量的大小,然后据实际可能按此进行调整。各约束条件松动量的值可以通过求解对应的目标规划来求取。

6.1.2 整数规划

在 6.1.1 所论的线性规划问题中,变量的取值可以是分数和小数,但在有的问题中,变量的要求取整数,如马匹数、人数、列车车厢数等。如果所有变量都取整数,称为纯整数规划,如果有一部分取整数,称为混合整数规划。如果对变量作进一步限制只取“0—1”,称为“0—1”型整数规划。以下所讨论的是整数线性规划问题,简称为整数规划,其一般数学模型为

$$\begin{aligned} \max(\text{或 } \min) S &= \sum_{j=1}^n c_j x_j \\ \text{s. t. } \begin{cases} \sum_{j=1}^n a_{ij} x_j \leq b_i & i = 1, 2, \dots, m \\ x_j \geq 0 & j = 1, 2, \dots, n \text{ 且为整数(或部分为整数)} \end{cases} \end{aligned}$$

例如古老的百马百瓦问题就是一个整数规划。它是这样的:—

百匹马运一百块瓦,每匹大马运三块瓦,每匹小马运二块瓦,每两匹马驹运一块瓦,要求马和瓦都不能剩余,问这百匹马中有大马、小马、马驹各多少匹?

以 x_1, x_2, x_3 分别为大马、小马和马驹数,由题意知

$$\begin{cases} x_1 + x_2 + x_3 = 100 \\ 3x_1 + 2x_2 + \frac{1}{2}x_3 = 100 \end{cases}$$

如果不加其他条件,该问题有许多组解,如果使小马驹最少,则该问题的数学模型为

$$\begin{aligned} \min & x_3 \\ \text{s.t.} & \begin{cases} x_1 + x_2 + x_3 = 100 \\ 3x_1 + 2x_2 + \frac{1}{2}x_3 = 100 \\ x_1, x_2, x_3 \geq 0 \text{ 且为整数} \end{cases} \end{aligned}$$

本节主要介绍整数规划的解法。既然整数规划与线性规划只有变量上要求的差异,能否用四舍五入的方法或穷举的方法计算呢?回答是不可行的。因为四舍五入之后的解可能不是可行解或不是最优解。穷举法对于大的规划问题是不可取的。整数规划的解法是从其解集中的部分解中通过搜索来实现。

1. 分枝定界法

分枝定界法是一类搜索方法,用它解题的过程就是对可行解空间系统地进行搜索。

整数规划与和它对应的线性规划相比,多了整数限制,其可行域是相应线性规划可行域的一部分。因此,在极大化问题中,线性规划最优解的目标函数值总是相应整数规划目标函数值的上界。

分枝定界法的思想是把全部可行域分割为越来越小的子集(称为分枝),对于每个子集解相应的线性规划,其目标函数最优值是该子集所有整数可行点的目标函数值的上界。如果这个上界比

某个已知可行整数点对应的目标函数值小,那么该子集称为剪枝,属于已探明,因为最优整数解不可能在该子集中。当子集对应的线性规划没有可行解或其最优解符合整数限制时,该子集也属于已探明,这样当每个子集都属于探明时,求出的即为满足整数限制的最优解。

$$\begin{aligned} \text{例 7} \quad \max S &= 6x_1 + 5x_2 \\ [0] \quad \text{s.t.} \quad &\begin{cases} 3x_1 + 6x_2 \leq 18 \\ 5x_1 + 4x_2 \leq 20 \\ x_1, x_2 \geq 0 \text{ 且为整数} \end{cases} \end{aligned}$$

解:首先不考虑整数限制,解对应的线性规划(称为[0]的松弛问题的最优解为 $x_1 = 2\frac{2}{3}, x_2 = 1\frac{2}{3}, S = 24\frac{1}{3}$ 如图 6.1.3

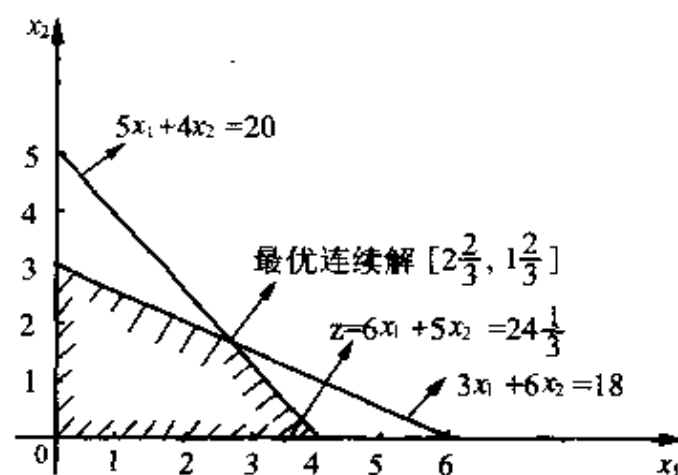


图 6.1.3

因为 x_1, x_2 都不符合整数限制,对其中任一个分数,如对 x_1 , 令 $x_1 \leq 2$ 或 $x_1 \geq 3$ 加在[0]上将[0]分成[1],[2]

$$\max S = 6x_1 + 5x_2$$

$$[1] S. t. \begin{cases} 3x_1 + 6x_2 \leq 18 \\ 5x_1 + 4x_2 \leq 20 \\ x_1 \leq 2 \\ x_1 x_2 \geq 0 \text{ 且为整数} \end{cases}$$

$$\max S = 6x_1 + 5x_2$$

$$[2] S. t. \begin{cases} 3x_1 + 6x_2 \leq 18 \\ 5x_1 + 4x_2 \leq 20 \\ x_1 \geq 3 \\ x_1 x_2 \geq 0 \text{ 且为整数} \end{cases}$$

解[1]的松弛问题的最优解, $x_1 = 2, x_2 = 2, S = 22$ 该子集已求出满足整数条件的解, 属于已探明并且找到一个新界 $S = 22$, 将原来整数解为空集时下界为 $-\infty$ 改为新界即

可行解集	对应下界
\emptyset	$-\infty$
$x^1 = (2, 2)$	22

解[2]的松弛问题的最优解 $x_1 = 3, x_2 = 1\frac{1}{4}, S = 24\frac{1}{4}$, 如

图 6.1.4

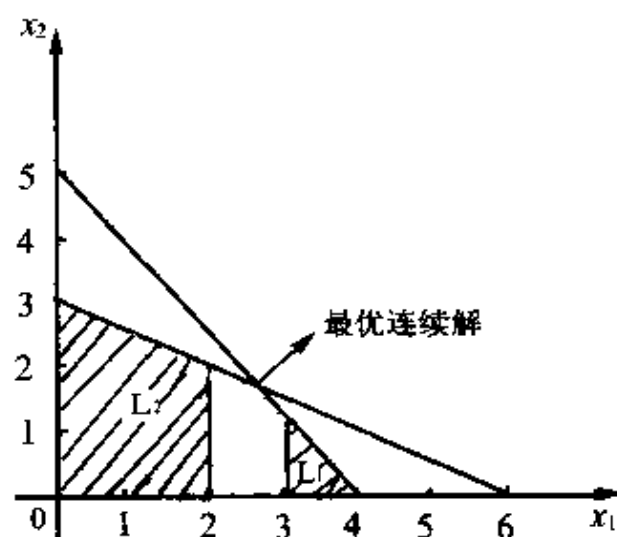


图 6.1.4

$\because x_2$ 不满足整数限制, 且 $24\frac{1}{4} > 22$, \therefore 将[2]按 $x_2 \leq 1$ 或 $x_2 \geq 2$ 再分成[3],[4]

$$\begin{aligned} \max S &= 6x_1 + 5x_2 \\ [3] \text{ S.t. } &\begin{cases} 3x_1 + 6x_2 \leq 18 \\ 5x_1 + 4x_2 \leq 20 \\ x_1 \geq 3, x_2 \leq 1 \\ x_1, x_2 \geq 0 \text{ 且为整数} \end{cases} \end{aligned}$$

$$\begin{aligned} \max S &= 6x_1 + 5x_2 \\ [4] \text{ S.t. } &\begin{cases} 3x_1 + 6x_2 \leq 18 \\ 5x_1 + 4x_2 \leq 20 \\ x_1 \geq 3, x_2 \geq 2 \\ x_1, x_2 \geq 0 \text{ 且为整数} \end{cases} \end{aligned}$$

如下图 6.1.5

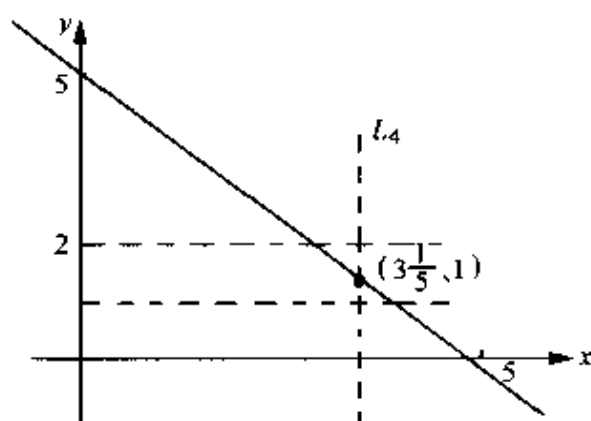


图 6.1.5

由图 6.1.5 可知 L_4 是空集, 即无可行解, 所以该子集也属于探明。

解[3] 的松弛问题的最优解为 $x_1 = 3 \frac{1}{5}, x_2 = 1, S = 24 \frac{1}{5}$. $\because x_1$ 不满足整数限制, 且 $24 \frac{1}{5} > 22$, \therefore 将[3] 按 $x_1 \leq 3$ 或 $x_1 \geq 4$ 分成[5][6], 如图 6.1.6

$$\begin{aligned} \max S &= 6x_1 + 5x_2 \\ [5] \text{ S.t. } &\begin{cases} 3x_1 + 6x_2 \leq 18 \\ 5x_1 + 4x_2 \leq 20 \\ x_1 \geq 3, x_1 \leq 3, x_2 \leq 1 \\ x_1, x_2 \geq 0 \text{ 且为整数} \end{cases} \end{aligned}$$

$$\begin{aligned} \max S &= 6x_1 + 5x_2 \\ [6] \text{ S.t. } &\begin{cases} 3x_1 + 6x_2 \leq 18 \\ 5x_1 + 4x_2 \leq 20 \\ x_1 \geq 4, x_1 \leq 1 \\ x_1, x_2 \geq 0 \text{ 且为整数} \end{cases} \end{aligned}$$

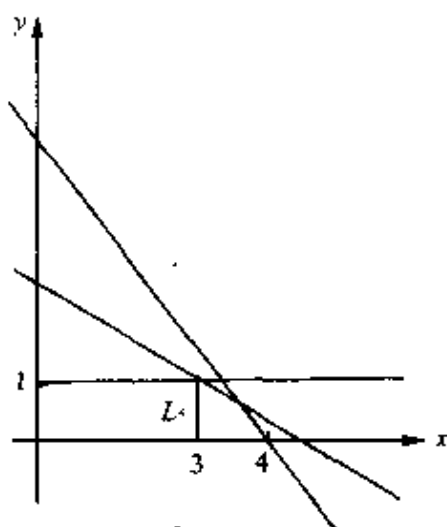


图 6.1.6

解由图 6.1.6 可见, L_5 为一线段 \therefore 最优解为 $x_1 = 3, x_2 = 1$, $S = 23$, L_6 为一点 $(4, 0)$ \therefore 最优解为 $x_1 = 4, x_2 = 0, S = 24$, 此时所有子集都属于探明。原 $[0]$ 的最优解为 $x_1 = 4, x_2 = 0$, 对应的目标函数值为 $S = 24$, 而 L_1 和 L_5 显然也有整数解, 但因为对应的目标函数分别为 22 和 23, 都小于 24, 所以该子集应剪枝。上述解题过程也可用树形图表示出来。(见图 6.1.7)

现将分枝定界法的原则归纳如下:

(1) 每个相应的线性规划问题的最优值均是它的整数规划问题最优值的上界。

(2) 在求解子问题相应的线性规划时, 会出现以下几种情况:

1° 若相应的线性规划问题无可行解, 则该子问题也无可行解, 此时属于探明。

2° 若相应的线性规划问题的解满足整数限制, 则此解为该子问题的最优解, 此时属于探明。如果目标函数值大于目前的下界值, 则修改下界值。

3° 若相应的线性规划问题的解不满足整数限制,但对应的目标函数值不大于目前的下界值,应剪枝,该子问题也属探明。

4° 若相应的线性规划问题的解不满足整数限制,而对应的目标函数值大于目前的下界值,应分枝,比如 $x_j = x_j^0$ 为非整数,则按 $x_j \leq [x_j^0]$ 或 $x_j \geq [x_j^0] + 1$ 将该子问题再分成两个子问题(其中 $[x_j^0]$ 表示不大于 x_j^0 的最大正数)。

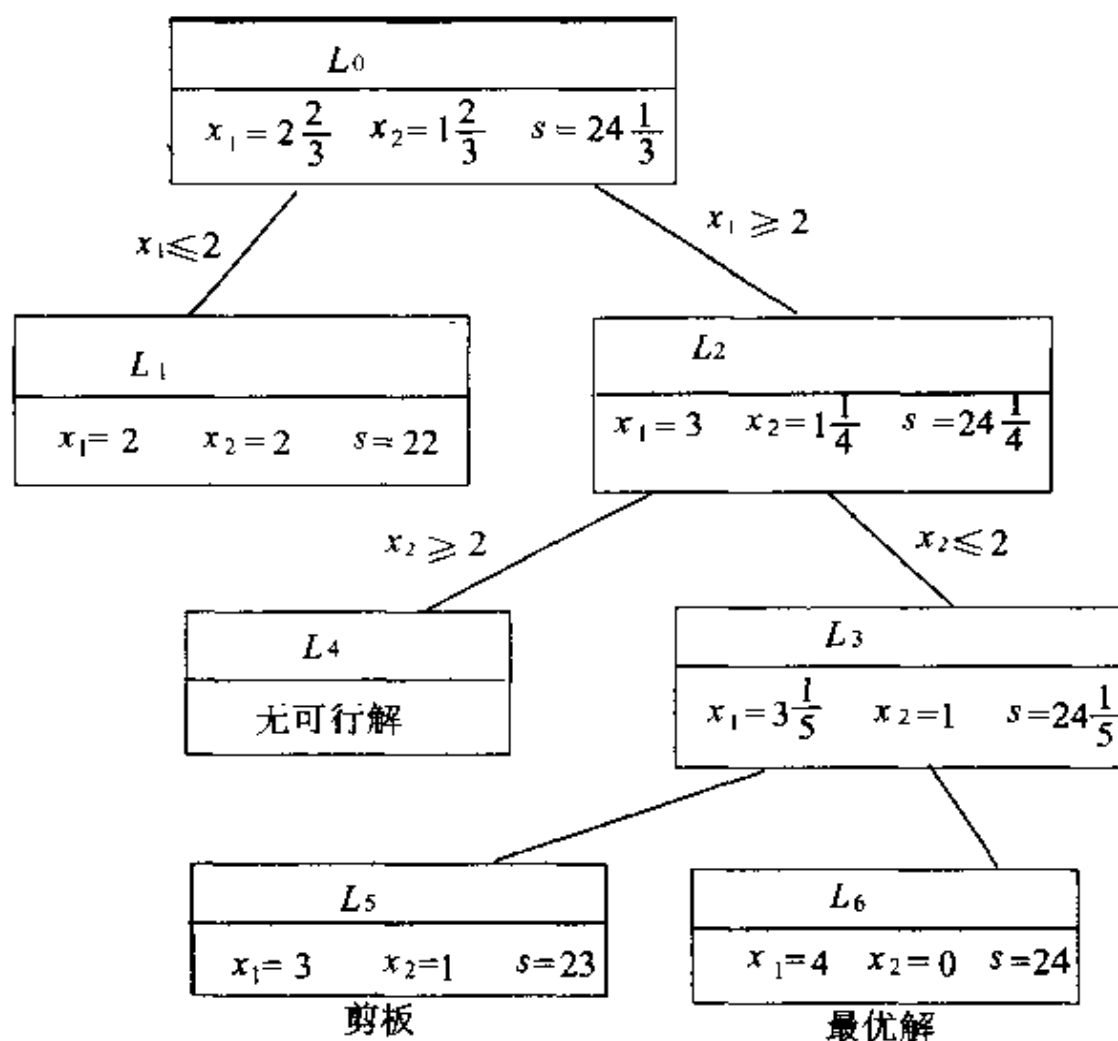


图 6.1.7

2. 隐枚举法

0-1 规划是很重要的一类整数规划,这不仅因为许多实际问题可以用 0-1 规划来表达,而且还因为一般整数规划问题可以化成 0-1 规划问题。

因为 0-1 变量的 0-1 组合是有限的,所以可以通过对这些组合进行系统搜索来求解。隐枚举法是在搜索时力图避免穷举法,而只通过枚举部分的 0-1 组合来实现的。例如满足不等式 $3x_1 - 8x_2 + 5x_3 \leq -6$ $x_j \in \{0,1\}, j = 1,2,3$ 的 0-1 变量有三个,组合数为 8,若用穷举法将这 8 组一一列出检查。但因为 $x_2 = 0$ 时不等式 $3x_1 + 5x_3 \leq -6$ 显然不成立,这样 $x_2 = 0$ 的四种组合自动被丢掉,称这四种组合被隐枚举了,而接下去对其余四种(对应于 $x_2 = 1$)进行搜索就行了。

下面解一个例子,以说明隐枚举法

$$\max S = -x_1 - x_2 - 2x_3 - x_4$$

$$S.t. \begin{cases} x_1 + 2x_2 + x_3 - x_4 \geq 3 \\ x_j \in \{0,1\} \quad j = 1,2,3,4 \end{cases}$$

为便于叙述,将约束不等式改成

$$\theta(I) = x_1 + 2x_2 + x_3 - x_4 - 3 \geq 0$$

现在还不知每个变量必须怎样取值才能满足约束,都可以取 0 或 1,所以称之为自由变量,由目标函数知,若自由变量全部取 0,则 S 一定最大,但这个解不是可行的。为寻找可行解,用固定一些变量的取值的方法进行分枝。选一个自由变量为 1,与原来已有固定值的变量(现在还没有)一起,称为一个“部分解”,有固定值的变量称为固定变量,用集合 J 表示部分解的变量及其固定值,任给各自由变量一个值,则与固定变量一起构成问题的一个“完全解”。另外为了尽快找到可行解,应从不满足的约束中,找出带有正系数且能使所有约束距离可行性的总距离最小的自由变量,将其固定为 1。所谓离可行性的距离,对满足约束的其距离为 0,对不满足

的,使此 $\theta(I)$ 增加到 0 的数量为其距离。如自由变量 x_1, x_2, x_3 的系数均为正,且 $x_1 = 1$ 时 $\theta(I) = 1 - 3 = -2$ 离可行性的总距离为 2; $x_2 = 1$ 时 $\theta(I) = 2 - 3 = -1$ 离可行性的总距离为 1; $x_3 = 1$ 时 $\theta(I) = 1 - 3 = -2$ 离可行性的总距离为 2。所以,选 x_2 固定为 1。这时 $J = \{2\}$, 这里 J 中的 2 表示下标为 2 的变量 x_2 固定为 1,若 $x_2 = 0$ 则 $J = \{-2\}$,用枚举树表示:现在的部分解对应点 ①(如图 6.1.8)。

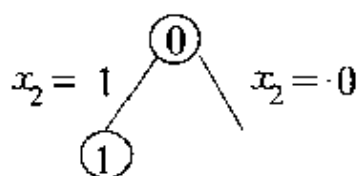


图 6.1.8

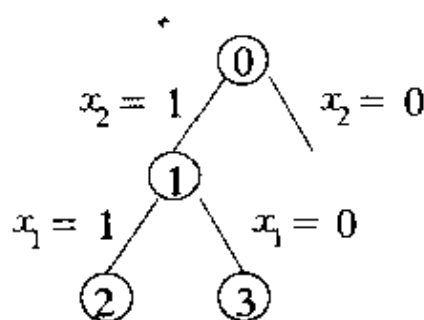


图 6.1.9

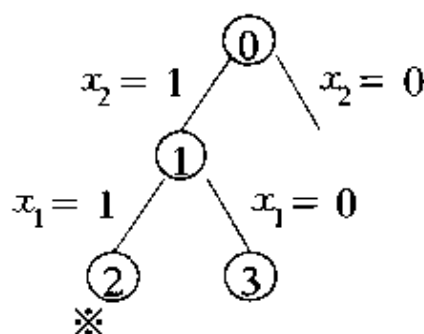


图 6.1.10

若取自由变量 x_1, x_3, x_4 均为 0, 则 $\theta(I) = 2 - 3 = -1$, 所以点 ① 的完全解不可行。这时再找一个自由变量固定为 1, 因在不满足的约束中带正系数的变量 x_1 和 x_3 使离可行性的距离均为 0, 所以任选一个如 x_1 , 这时 $J = \{2, 1\}$ 这个部分解对应枚举树的点 ②(如图 6.1.9)。

现在自由变量为 x_3, x_4 令其为 0, 得到完全解 $(1, 1, 0, 0)$ 代入约束 $\theta(I) = 1 + 2 - 3 = 0$ 是可行解, 从点 ② 不必再分枝。因为这时的自由变量均为 0, 而目标函数中的系数全为负且问题是求极大的, 所以从这点继续分枝不会得到更好的解, 这点上 $S = -2$, 所以 $S^* \geq -2$ 。

若 J 中的 x_1 取 0, 则 $J = \{2, -1\}$, 对应的点 ③(如图 6.1.9), 图中 ② 下的星号表示该分枝搜索完(已经探明)(如图 6.1.10)。

这时取 $x_3 = 1, J = \{2, -1, 3\}$ 则在点 ④ 完全解 $(0, 1, 1, 0)$ 可行, 对应的 $S = -3$, 由前边已知 $S^* \geq -2$ 所以这个解不可能是最优解, 此时点 ④ 已经探明。

若取 $x_3 = 0, J = \{2, -1, -3\}$, 点 ⑤ 的完全解 $(0, 1, 0, 0)$ 不可行, 但在这一点上, 自由变量在不满足的约束条件中的系数全是负的, 所以不需要从这点再分枝(如图 6.1.11)。

现在又要从 J 中取一个固定值为 1 的变量改为 0, 因为 x_1, x_3 固定 1 和 0 的两种状态都枚举过, 所以选 $x_2 = 0$, 并把 x_1, x_3 从 J 中拿掉, 重新成为自由变量, $J = \{-2\}$ 得点 ⑥, 该点的完全解 $(0, 0, 0, 0)$ 是不行的(如图 6.1.12)。在不满足的约束中带正系数的自由变量为 x_1, x_3 , 令它们为 1 则 $\theta(I) = 1 + 1 - 3 = -1$ 不满足, 所以从这点不可能再产生可行解。

所有分枝都枚举完了, J 中不再有正元素, 求解过程结束。 $x^* = (1, 1, 0, 0), S^* = -2$ 。

将隐枚举算法归纳如下:

应用这个算法前要先将问题化为标准形式:

$$\max S = \sum_{j=1}^n c_j x_j \quad \text{其中 } c_j \leq 0$$

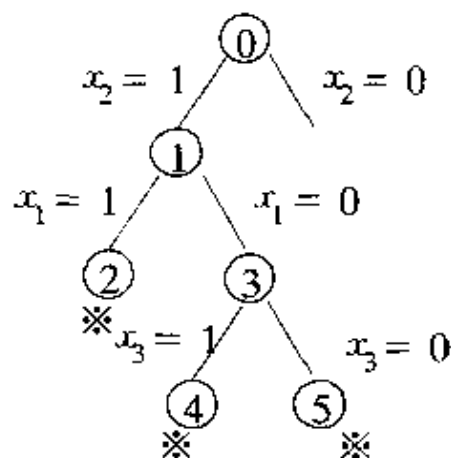


图 6.1.11

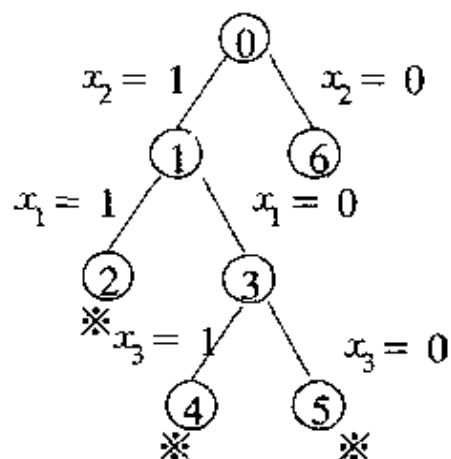


图 6.1.12

$$S.t. \begin{cases} \theta_i = \sum_{j=1}^n a_{ij}x_j - b_i \geq 0 & i = 1, 2, \dots, m \\ x_j \in \{0, 1\} & j = 1, 2, \dots, n \end{cases}$$

在算法中,用 J 表示固定变量的集合, $j \in J$ 表示 x_j 是固定变量。 J 中 j 的前面没有符号表示 x_j 的固定值为 1, j 的前面有负号,表示 x_j 的固定值为 0,用 \underline{S} 表示当前已知最好的可行解对应的目标函数值。

(1) 令 $J = \phi, \underline{S} = -\infty, x_0 = (0, 0, \dots, 0)$

(2) 取 J 中的变量为其固定值,其余变量(自由变量)为 0,求每个约束的值 $\theta_i, i = 1, 2, \dots, m$ 。如果对所有 i 都成立 $\theta_i \geq 0$,则转(6) 否则。

(3) 找有希望使所有约束满足的自由变量,以 T 表示它们的集合,这些变量有下列特点:

① 在某些不满足的约束中,其系数为正,

② 在目标函数中,其系数 $> \underline{S} - \sum_{j \in J} c_j x_j$ 。

条件②的意义在于,如果现在把不满足这个条件的变量固定为 1,即使得到可行解,也不会比已知的可行解好。

若 T 是空集,转(7) 否则

(4) 对每个不满足的约束,令 J 中变量为其固定值,令在该约束中有正系数且属于 T 的自由变量为 1 如果仍有约束不满足转(7) 否则

(5) 从 T 中选出可以使所有离可行性的总距离最小的变量,加到 J 中,固定它的值为 1,转(2)

这里所谓 T 中某变量对应的所有约束离可行性的总距离为:令该变量为 1 且令 J 中变量为其固定值时,仍不满足的那些约束 θ_i 的总和的绝对值

(6) 令 J 中变量为其固定值,自由变量为 0,组成一个完全解,

若对应目标函数值 $S > \underline{S}$, 则令 $\underline{S} = S$, 并保留这个完全解, 以 x_0 表示

(7) 若 J 为空集, 转(10) 否则

(8) 如 J 中的元素全为负, 转(10) 否则

(9) 使 J 中最后面的正元素为负, 并从 J 中消去它后面的元素, 使它们重新成为自由变量, 转(2)

(10) 隐枚举过程结束, 如果 $\underline{S} = -\infty$, 则问题无可行解; 否则最优解 $x^* = x_0, S^* = \underline{S}$ 。

6.1.3 非线性规划

在规划问题的目标函数或约束条件中, 如果存在非线性函数, 则该规划称为非线性规划。其一般数学模型为

$$\begin{cases} \min(\text{或 max}) f(x_1, x_2, \dots, x_n) \\ g_i(x_1, x_2, \dots, x_n) \geq 0 \quad i = 1, 2, \dots, m \end{cases}$$

非线性规划的有关理论要比线性规划复杂的多, 到目前为止还没有一个关于非线性规划的理想算法, 本节中我们简要介绍的几种算法都有一定的局限性。

1. 梯度、凸函数

定义 6.1.4 设 $f(x) = f(x_1, x_2, \dots, x_n)$ 具有一阶连续偏导数, 令 $f_r(x)^T = \left(\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \dots, \frac{\partial f(x)}{\partial x_n} \right)^T$ 则 $f_r(x)^T$ 称为 n 元函数 $f(x)$ 的梯度。

$f(x)$ 在点 x^0 的梯度方向 $f_r(x^0)^T$ 是函数 $f(x)$ 在点 x^0 处增加最快的方向。

定义 6.1.5 若对任意 $x^0, y^0, x^0 \neq y^0$, 及 $\lambda \in (0, 1)$ 都有 $f(\lambda x^0 + (1 - \lambda)y^0) \leq \lambda f(x^0) + (1 - \lambda)f(y^0)$ 则称 $f(x)$ 为凸函数。

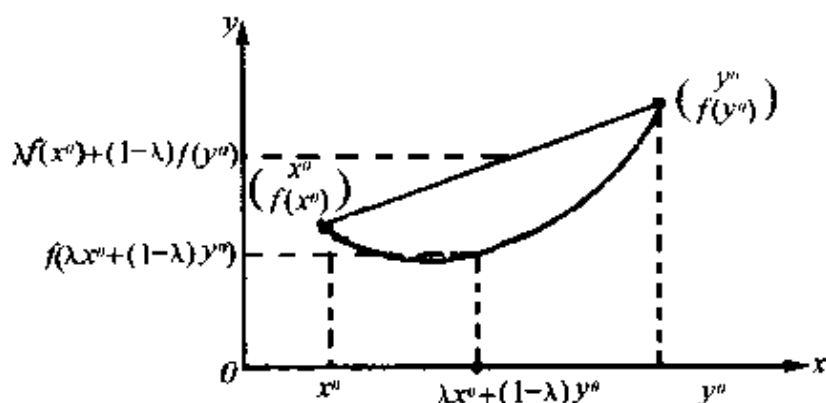


图 6.1.13

由图 6.1.13 可知凸函数的曲线都在其割线之下。或者说整个曲线都在其任一点切线的上方。

即:对任意 x 及 $x^0, x \neq x^0$ 都有 $f(x) \geq f(x^0) + f_x(x^0)(x - x^0)$

2. 非线性规划的几种解法

(1) 最速下降法

非线性规划 $\min_{x \in E_n} f(x)$, 其中 $x = (x_1, x_2, \dots, x_n)^T$ 称为无约束的极值问题。

给定初始点 $x^1 \in E_n$ 。若 $f_x(x^1) = 0$, 则当 $f(x)$ 为凸函数时, x^1 即为该问题的最优解, 因为对于任意 $x \in E_n$, 都有 $f(x) \geq f(x^1) + f_x(x^1)(x - x^1) = f(x^1)$ 。当 $f_x(x^1) \neq 0$ 时, 由于 $f(x)$ 在点 x^1 的负梯度 $-f_x(x^1)^T$ 是 $f(x)$ 在点 x^1 处下降最快的方向, 所以取方向 $Z^1 = -f_x(x^1)^T$ 。在此方向上求出使 $f(x)$ 取最小值的步长 λ_1 , 即

$$\min_{\lambda \geq 0} f(x^1 + \lambda Z^1) = f(x^1 + \lambda_1 Z^1) \text{ 令 } x^2 = x^1 + \lambda_1 Z^1$$

得 $f(x^2) < f(x^1)$, 也就是说, 在 $f_x(x^1) \neq 0$ 的情况下, 找到

一个比初始点 x^1 要好的新点 x^2 。如果 $f_x(x^2) \neq 0$, 还可以重复上述步骤, 得到比 x^2 还好的点 $x^3 \cdots$ 这就是最速下降法的基本思路。

迭代步骤(最速下降法):

- ① 取初始点 $x^1 \in E_n$, 允许误差 $\epsilon > 0$, 令 $k := 1$
- ② 计算 $Z^1 = -f_x(x^1)^T$
- ③ 检验是否满足条件($\|Z^k\| \leq \epsilon$), 若满足则迭代停止, 得到点 x^k , 否则转(4)

④ 求单变量极值问题的最优解 $\lambda_k \in E_1$, 使得

$$\min_{\lambda \geq 0} f(x^k + \lambda Z^k) = f(x^k + \lambda_k Z^k)$$

⑤ 令 $x^{k+1} = x^k + \lambda_k Z^k$ $k := k + 1$ 转到(2)

最速下降法的迭代步骤用框图表示如下:

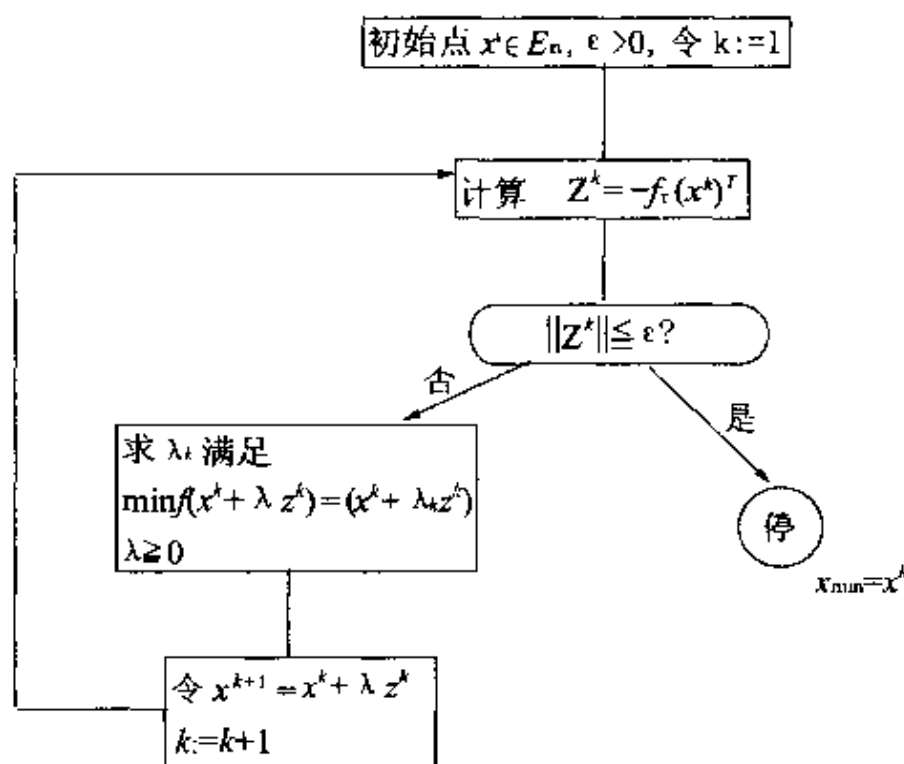


图 6.1.14

例1 求下列无约束极值问题的最优解。

$$\min_{x \in E_2} [x_1^2 + 25x_2^2]$$

(1) 取初始点 $x^1 = (2, 2)^T$ $\epsilon = 0.1$, $f(x^1) = 104$

(2) 由 $f(x) = x_1^2 + 25x_2^2$ 所以 $f_x(x) = (2x_1, 50x_2)$

$$Z = -f_x(x^1)^T = (-4, -100)^T$$

(3) $\|Z^1\| = \sqrt{(-4)^2 + (-100)^2} > \epsilon$

(4) 由 $x^1 + \lambda Z^1 = (2, 2)^T + \lambda(-4, -100)^T$
 $= (2 - 4\lambda, 2 - 100\lambda)^T$,

$$\therefore f(x^1 + \lambda Z^1) = (2 - 4\lambda)^2 + 25(2 - 100\lambda)^2$$

$\min_{\lambda \geq 0} f(x^1 + \lambda Z^1) = \min_{\lambda \geq 0} [(2 - 4\lambda)^2 + 25(2 - 100\lambda)^2]$ 的最优

解为 $\lambda_1 = 0.02$

(5) $x^2 = x^1 + \lambda_1 Z^1 = (1.92, 0)^T$ $f(x^2) = 3.69$

由于 $Z^2 = -f_x(x^2)^T = (-3.84, 0)^T$ $\|Z^2\| = 3.84 > \epsilon$

\therefore 还需要由 x^2 开始重复上述步骤得到下一点 x^3 (以后留给读者完成)

(2) 罚函数方法

$$\text{非线性规划} \begin{cases} \min f(x) & x = (x_1, x_2, \dots, x_n)^T \in E_n \\ g_i(x) \geq 0 & i = 1, 2, \dots, m \end{cases}$$

称为有约束的极值。

对于上述形式的规划问题做函数

$$T(x, M_k) = f(x) + M_k \sum_{i=1}^m [\min(0, g_i(x))]^2 \text{ 称该函数为罚函数,}$$

其中 M_k 是一个充分大的数, 于是 $T(x, M_k)$ 作为无约束极值问题, 是和原问题同解。这样我们就可以利用上述的最速下降法来解这一类问题了。

罚函数法的迭代步骤:

① 取 $M_1 > 0$ (例如取 $M_1 = 1$) 允许误差 $\epsilon > 0$ 令 $K := 1$

② 求无约束极值问题 $\min_{x \in E_n} T(x, M_k)$ 的最优解 x^k

其中 $T(x, M_k) = f(x) + M_k \sum_{i=1}^m [\min(0, g_i(x))]^2$

③ 若存在 $i (1 \leq i \leq m)$ 使 $g_i(x_k) \leq -\varepsilon$ 则取 $M_{k+1} > M_k$ (例如 $M_{k+1} = 10M_k$), 令 $k := k + 1$ 转到(2)

否则迭代停止, 得到点 x_k 。

罚函数法的迭代步骤用框图表示如下。

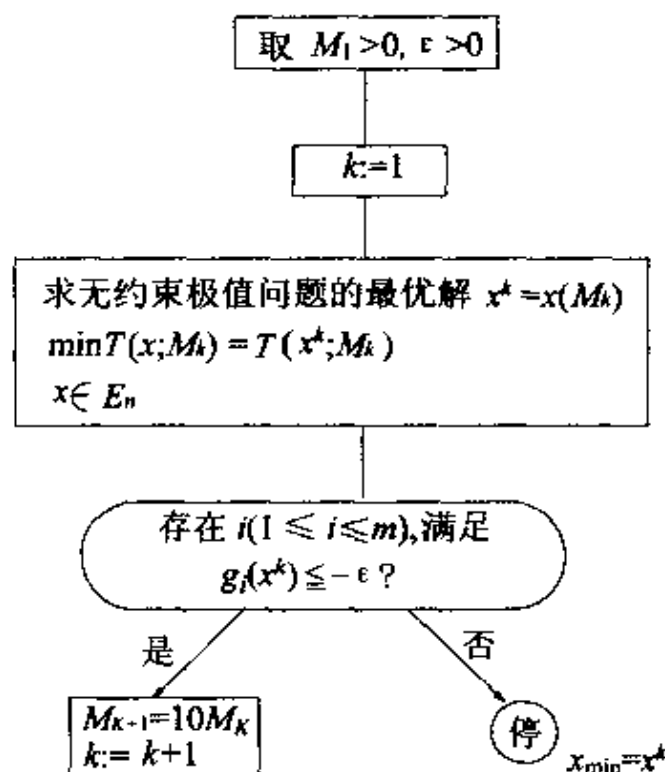


图 6.1.15

例 2
$$\begin{cases} \min [x_1^2 + 2x_2^2] \\ 1 - x_1 - x_2 = 0 \end{cases}$$

此时 $T(x, M_k) = (x_1^2 + 2x_2^2) + M_k(1 - x_1 - x_2)^2$

求 $T(x, M_k)$ 的梯度 $T_x(x, M_k) = 0$, 即

$$\begin{cases} \frac{\partial T(x, M_k)}{\partial x_1} = 2x_1 - 2M_k(1 - x_1 - x_2) = 0 \\ \frac{\partial T(x, M_k)}{\partial x_2} = 4x_2 - 2M_k(1 - x_1 - x_2) = 0 \end{cases} \quad \text{可见 } x_1 = 2x_2$$

$$\text{解上面方程组得到} \begin{cases} x_1^k = \frac{2M_k}{2 + 3M_k} & \lim_{k \rightarrow \infty} x_1^k = \frac{2}{3} \\ x_2^k = \frac{M_k}{2 + 3M_k} & \lim_{k \rightarrow \infty} x_2^k = \frac{1}{3} \end{cases}$$

所以最优解为 $x_{\min} = (\frac{2}{3}, \frac{1}{3})^T$

(3) 线性约束条件下线性逼近的方法

具有线性约束条件下的非线性规划问题

$$\begin{cases} \min f(x) & x = (x_1, x_2, \dots, x_n)^T \in E_n \\ Ax \geq b \end{cases}$$

其中 $f(x)$ 具有一阶连续偏导数, A 为 $m \times n$ 矩阵, b 为 $m \times 1$ 矩阵

在该规划的可行解集合内任取一个可行解 x^0 , 在 x^0 点把目标函数线性化, 即取泰勒展开式的前两项

$$L(x) = f(x^0) + f_x(x^0)(x - x^0)。$$

我们求 $f(x)$ 在 x^0 点线性化后函数 $L(x)$ 在线性约束下的最

$$\text{优解。即} \begin{cases} \min [f(x^0) + f_x(x^0)(x - x^0)] \\ Ax \geq b \end{cases}$$

$\because L(x) = f(x^0) + f_x(x^0) \cdot x - f_x(x^0) \cdot x^0$, 所以上面的问

题等价于线性规划 $\begin{cases} \min f_x(x^0)x \\ Ax \geq b \end{cases}$ 用单纯形法求解即可。

迭代步骤:

① 取允许误差 $\epsilon > 0$, 初始点 $x^0 \in R$ 令 $k := 0$

② 求线性规划问题的最优解 $y^k \in R$

$$\begin{cases} \min f_x(x^0)x = f_x(x^0)y^k \\ Ax \geq b \end{cases}$$

③ 检验是否满足 $|f_x(x^k)(y^k - x^k)| \leq \epsilon$ 若满足则迭代停止, 得到点 $x^k \in R$, 否则转到(4)

④ 求单变量极值问题的最优解 $\lambda_k \in [0, 1]$

$$\min_{0 \leq \lambda \leq 1} f(x^k + \lambda(y^k - x^k))$$

⑤ 令 $x^{k+1} = x^k + \lambda_k(y^k - x^k)$, $k := k + 1$ 转到(2)

迭代步骤用框图表示如下

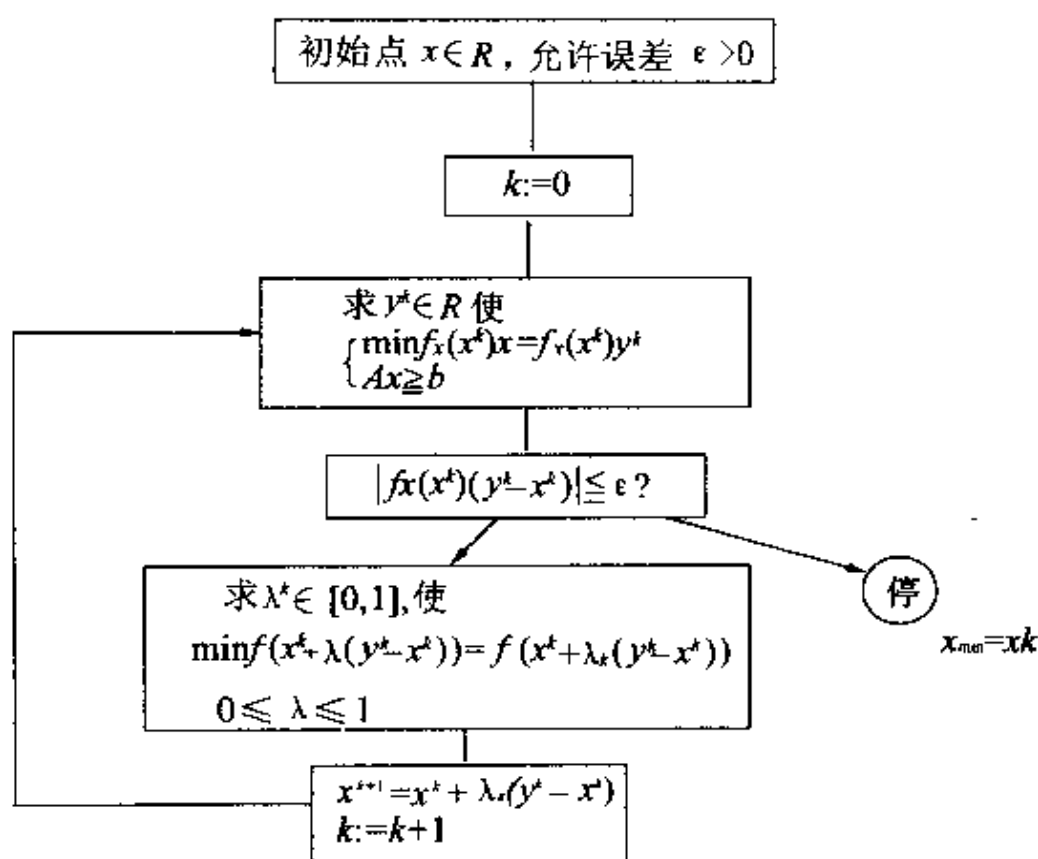


图 6.1.16

例 3 求非线性规划问题

$$\min [(x_1 - 10)^2 + x_2^2]$$

$$S.t. \begin{cases} x_1 - x_2 + x_3 = 4 \\ x_1 + x_2 + x_4 = 8 \\ x_1, x_2, x_3, x_4 \geq 0 \end{cases}$$

$$\because f(x) = (x_1 - 10)^2 + x_2^2 \quad x = (x_1, x_2, x_3, x_4)^T$$

$$\therefore f_x(x) = [2(x_1 - 10), 2x_2, 0, 0]$$

$$\text{取初始点 } x^0 = (0, 0, 4, 8)^T \therefore f_x(x^0) = (-20, 0, 0, 0)$$

$f_x(x^0)x = -20x_1$, 此时的线性规划为

$$\min(-20x_1)$$

$$S.t. \begin{cases} x_1 - x_2 + x_3 = 4 \\ x_1 + x_2 + x_4 = 8 \\ x_1, x_2, x_3, x_4 \geq 0 \end{cases}$$

用单纯形法求得最优解

$$y^0 = (6, 2, 0, 0)^T$$

$$\because f_x(x^0)(y^0 - x^0) = (-20, 0, 0, 0) \begin{bmatrix} 6 \\ 2 \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ 4 \\ 0 \end{bmatrix} \neq 0 \text{ 不满足}$$

判别条件, 于是, 求一个单变量极值 $\min_{0 \leq \lambda \leq 1} f[(x^0 + \lambda(y^0 - x^0))]$

$$\because x^0 + \lambda(y^0 - x^0) = (6\lambda, 2\lambda, 4 - 4\lambda, 8 - 8\lambda)^T$$

$$\therefore f(x^0 + \lambda(y^0 - x^0)) = (6\lambda - 10)^2 + (2\lambda)^2$$

$$\text{求 } \lambda_0 \text{ 使 } \min_{0 \leq \lambda \leq 1} [(6\lambda - 10)^2 + (2\lambda)^2]$$

$$\text{令 } \frac{d[(6\lambda - 10)^2 + (2\lambda)^2]}{d\lambda} = 0 \text{ 得 } \bar{\lambda} = \frac{3}{2} > 1 \therefore \text{取 } \lambda_0 = 1$$

得 $x^1 = x^0 + \lambda_0(y^0 - x^0) = (6, 2, 0, 0)^T \quad f_x(x^1) = (-8, 4)$ 对应于 x^1 的线性规划为

$$\min(-8x_1 + 4x_2)$$

$$\text{S. t. } \begin{cases} x_1 - x_2 + x_3 = 4 \\ x_1 + x_2 + x_4 = 8 \\ x_1, x_2, x_3, x_4 \geq 0 \end{cases}$$

用单纯形法求得最优解, 为 $y^1 = (6, 2, 0, 0)^T$, $\therefore f_x(x^1)(y^1 - x^1) = 0$ 满足判别条件, 即 $x^1 = (6, 2, 0, 0)$ 为对应的线性规划的最优解, 也就是非线性规划的最优解 $x_{\min} = (6, 2, 0, 0)^T$

§ 6.2 建模实例

6.2.1 资源最优利用问题

资源最优利用问题是一个国家、一个地区经济发展中的最基本的问题。资源包括三大类: (1) 自然资源, 如土地、可供开采的矿藏、森林等; (2) 各种可以直接或间接用于再生产过程的人类劳动的产品, 如电、煤、原油、厂房、设备等; (3) 劳动力, 它可按熟练程度或专业加以分类。在现有资源条件下, 如何使经济发展速度最快, 是很重要的问题, 在很多情况下, 可以利用线性规划来帮助解决这个问题。

1. 生产计划问题

一个企业拥有 m 种资源, 已知每种资源的数量为 $b_i (i = 1, 2, \dots, m)$ 这个企业能生产 n 种产品, 生产一个单位的每种产品所消耗资源的数量已知, 以 a_{ij} 表示生产一个单位第 j 种产品对第 i 种资源的消耗。各种产品的单价也是已知的, 用 c_j 表示第 j 种产品的单价, 如何在企业现有资源的条件下(厂房、设备、劳力、原材料等)生产出产值最大(或效益最高)的产品。

数学模型如下

设 x_j 为各种产品的生产量, 选用总产值为目标函数

$$\max Z = c_1 x_1 + c_2 x_2 + \dots + c_n x_n$$

$$S. t. \begin{cases} a_{i1}x_1 + a_{i2}x_2 + \cdots a_{in}x_n \leq b_i (i = 1, 2, \cdots m) \\ x_j \geq 0 (j = 1, 2, \cdots n) \end{cases} \begin{matrix} \text{(对各种资源的需求量不超过其拥有量)} \\ \text{(变量的非负要求)} \end{matrix}$$

如果以企业所创造的净产值或利润作为目标函数,那么 c_j 就是第 j 种产品每单位的净产值或利润。

2. 机床负荷问题

某企业有 m 种机床,生产 n 种产品。第 i 种机床的台数为 b_i , 每种机床加工各种产品的生产效率已知,以 a_{ij} 表示第 i 种机床生产第 j 种产品的效率。这 n 种产品要求保持一定的配套比例,其配套比例为 p_1, p_2, \cdots, p_n 。问如何分配生产任务,使生产产品最多?

数学模型如下:

设 x_{ij} 表示第 i 类机床用来生产第 j 种产品的机床台数,因为产品数量配套按比例,选取其中一个比例为目标函数。

$$\max Z = \sum_{i=1}^n a_{i1}x_{i1}/p_1$$

$$S. t. \begin{cases} \sum_{j=1}^n x_{ij} = b_i (i = 1, 2, \cdots m) \text{(机床分配方程)} \\ \frac{\sum_{i=1}^m a_{i1}x_{i1}}{p_1} = \frac{\sum_{i=1}^m a_{i2}x_{i2}}{p_2} = \cdots = \frac{\sum_{i=1}^m a_{in}x_{in}}{p_n} \\ \text{(产品配套要求)} \end{cases}$$

$x_{ij} \geq 0 (i = 1, 2, \cdots m) (j = 1, 2, \cdots n)$ 变量非负限制且为整数,该问题是整数规划。

3. 合理下料问题

合理下料问题是许多工业部门特别是机械制造部门,缝纫工业部门,制鞋工业部门经常遇到的问题。要求在长度一定的条型材料上,或面积、形状一定的材料上切割若干具有一定形状,尺寸的毛坯,在一般情况下,材料不可能被完全利用,会有一部分残料(下

脚料)。如何最大限度地减少残料,使得切割规定数量的毛坯所用的材料最少,就是合理下料问题所需要研究的。

假设需要切割 n 种毛坯,其数量分别以 b_i 表示 ($i = 1, 2 \cdots n$)。设有 m 种下料方式,分别以 $a_{1j}, a_{2j} \cdots a_{nj}$ 表示第 j 种下料方式每根原材料(或每块板料)所切割出来的各种毛坯的数量。问如何下料,才能使原料最省?

数学模型如下:

设 x_j 表示用第 j 种下料方式所切割的原材料的根数(或板料的块数),选所用原材料之和为目标函数

$$\begin{aligned} \min Z &= x_1 + x_2 + \cdots + x_m = \sum_{j=1}^m x_j \\ S. t. &\begin{cases} a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{im}x_m \geq b_i (i = 1, 2 \cdots n) \\ \text{(各种下料方式切割出的毛坯数量满足要求)} \\ x_j \geq 0 \text{ 且为整数, } j = 1, 2 \cdots m \text{ (变量非负要求)} \end{cases} \end{aligned}$$

该问题也是整数规划

下料问题还可以有另一种提法,在很多原材料供应紧张的情况下可考虑采用。这时问题归结为在一定数量的原材料上,如何切割出数量最多的配套毛坯来,以 $k_1, k_2 \cdots k_n$ 表示所要求的各种毛坯的配套比例,原材料的数量为 c ,其他符号同上。

数学模型如下:

$$\begin{aligned} \max Z &= \frac{\sum_{j=1}^m a_{1j}x_j}{k_1} \\ S. t. &\begin{cases} x_1 + x_2 + \cdots + x_m = c \\ \frac{\sum_{j=1}^m a_{1j}x_j}{k_1} = \frac{\sum_{j=1}^m a_{2j}x_j}{k_2} = \cdots = \frac{\sum_{j=1}^m a_{nj}x_j}{k_n} \\ x_j \geq 0 \text{ 且为整数 } (j = 1, 2 \cdots m) \end{cases} \end{aligned}$$

4. 合理配料问题:

某饲养场制订牲畜饲料的配料方案。

已知:(1) 各种饲料的最大供应量;(2) 各种饲料中所含有的营养成分的数量;(3) 配料方案中所含的各种营养成分的数量不能少于规定的数量;(4) 各种饲料的单价。问如何配料才能即满足最低的营养要求,又使价格最便宜?

引入以下符号:

m — 营养成分的种类

n — 现有饲料的种类

b_i — 第 i 种营养成分的最低需求数量

d_j — 第 j 种饲料的最大供应量

a_{ij} — 第 j 种饲料一个单位含有的第 i 种营养成分的数量

c_j — 第 j 种饲料的单价

数学模型如下:

设 x_j 为配料方案中所含有的第 j 种饲料的数量,选总费用为目标函数

$$\begin{aligned} \min Z &= c_1x_1 + c_2x_2 + \cdots + c_nx_n \\ S. t. &\begin{cases} a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n \geq b_i & i = (1, 2, \cdots, m) \\ \text{(配料方案中所含各种营养成分达到最低要求)} \\ x_j \leq d_j & (j = 1, 2, \cdots, n) \\ \text{(各种饲料的使用量不超过规定的数量)} \\ x_j \geq 0 & (j = 1, 2, \cdots, n) \text{(变量非负要求)} \end{cases} \end{aligned}$$

6.2.2 生产任务的分配问题

1. 在不超过设备生产能力的条件下,在各企业之间分配生产任务

某一个主管部门有 m 个工厂,各个工厂主要设备的能力以 $d_i (i = 1, 2, \cdots, m)$ 表示。这个主管部门在计划期需要生产 n 种产

品,每种产品的计划产量为 $b_j (j = 1, 2 \cdots n)$ 。已知第 i 个工厂生产第 j 种产品所需要的主要设备台时数为 a_{ij} , 其单位产品成本为 c_{ij} , 问如何合理分配生产任务, 在完成各种产品产量的前提下, 使这个主管部门的总成本最小?

数学模型如下:

设 x_{ij} 表示第 i 个工厂生产第 j 种产品的数量。选取总成本为目标函数

$$\begin{aligned} \min Z &= \sum_{i=1}^m \sum_{j=1}^n c_{ij} x_{ij} \\ \text{s.t.} \quad &\begin{cases} \sum_{i=1}^m x_{ij} = b_j (j = 1, 2 \cdots n) \\ \text{(保证主管部门生产任务的完成)} \\ \sum_{j=1}^n a_{ij} x_{ij} \leq d_i (i = 1, 2 \cdots m) \\ \text{(各企业所承担的生产任务不超过主要设备的能力)} \\ x_{ij} \geq 0 (i = 1, 2 \cdots m, j = 1, 2 \cdots n) \text{(变量非负要求)} \end{cases} \end{aligned}$$

2. 在设备超负荷情况下, 在各企业之间分配生产任务

当某一主管部门所承担的生产任务很重时, 为了完成任务, 各企业有可能超负荷工作。问如何合理分配生产任务, 使得整个部门超负荷的数量达到最小?(或使生产费用达到最小)

设 u_i 表示第 i 个工厂设备负荷不足的数值; v_i 表示第 i 个工厂设备超负荷的数量。在求解过程中保证 u_i 和 v_i 中至少有一个其值为零。其他符号意义同上。

数学模型如下:

选取每工厂设备能力与实际负荷之差为目标函数: $\min Z =$

$$\sum_{i=1}^m (u_i + v_i)。$$

为使规模大小不同的企业超负荷情况得到合理的分配。也可选目标函数如下。 $\min Z = \sum_{i=1}^m \left(\frac{u_i + v_i}{d_i} \right)$, 第三种目标函数

是以整个部门的生产费用最小作为目标。以 c_{ij} 表示在设备未超负荷前第 i 个工厂生产第 j 种产品的单位成本,以 p_i 表示设备超负荷一个台时所带来的各种额外的费用和损失,则目标函数为

$$\min Z = \sum_{i=1}^m \sum_{j=1}^n c_{ij} x_{ij} + \sum_{i=1}^m p_i v_i。$$

该问题的约束条件有以下三组

$$(1) \sum_{i=1}^m x_{ij} = b_j (j = 1, 2 \cdots n)$$

(保证主管部门生产任务的完成)

$$(2) \sum_{j=1}^n a_{ij} x_{ij} + u_i = d_i + v_i (i = 1, 2 \cdots m)$$

(设备的负荷方程)

$$(3) x_{ij} \geq 0 \quad u_i \geq 0 \quad v_i \geq 0 (i = 1, 2 \cdots m, j = 1, 2 \cdots n)$$

(变量非负要求)

3. 年度任务的按月分配问题

一个企业在制订了年度计划后,尚需编制月度计划。某些企业,对编制月度计划有如下要求:(1) 从数量和品种上保证年度生产任务的完成;(2) 某些产品在上半年交货;(3) 另一些产品由于某些原因(如生产准备工作的进度,原材料,外协的供应等)只能在下半年生产;(4) 某些批量很小的产品集中在一个月内生产完毕,不要分散到几个月;(5) 由于各种产品所需要的加工设备种类不同,要求各个月份几种主要设备负荷尽可能均衡。这个企业计划生产 n 种产品,第 j 种产品全年的计划产量以 b_j 表示;这个企业共有 m 种主要设备,以 d_{is} 表示第 s 个月份第 i 种设备的能力。问如何按月分配生产任务,使各个月份各种设备的生产能力与实际负荷之差达到最小,或主要设备在各月份工作时间比较均衡。

数学模型如下:(使用符号意义同上)

设以 x_{js} 表示第 s 个月份第 j 种产品的计划产量。目标函数选

取为 $\min Z = \sum_{i=1}^m \sum_{s=1}^{12} (u_{is} + v_{is})$ 其中 u_{is}, v_{is} 分别表示第 s 月份第 i 种设备负荷不足数或超负荷数。

该问题的约束条件有以下六组:

(1) 保证各种产品完成全年生产任务

$$\sum_{s=1}^{12} x_{js} = b_j (j = 1, 2, \dots, n)$$

(2) 保证第 j_1 种产品上半年交货

$$\sum_{s=1}^6 x_{j_1 s} = b_{j_1} \text{ (或 } \sum_{s=7}^{12} x_{j_1 s} = 0 \text{)}$$

(3) 保证第 j_2 种产品下半年进行生产

$$\sum_{s=1}^6 x_{j_2 s} = 0 \text{ (或 } \sum_{s=7}^{12} x_{j_2 s} = b_{j_2} \text{)}$$

(4) 保证第 j_0 种产品要求集中在某月生产

$$\sum_{s=1}^{12} x_{j_0 s}^2 = b_{j_0}^2 \text{ 这个约束条件与约束条件(1) } \left(\sum_{s=1}^{12} x_{j_0 s} = b_{j_0} \right) \text{ 同}$$

时考虑, 就可以保证第 j_0 种产品集中在一个月生产。这个条件也称为保证“集中生产”约束条件组。例如第一种产品第一季度的计划产量为 10 件, 要求集中在某一个月份生产, 则我们可以建立以下约束条件组

$$\begin{cases} x_{11} + x_{12} + x_{13} + x_{14} = 10 \\ x_{11}^2 + x_{12}^2 + x_{13}^2 + x_{14}^2 = 100 \end{cases} \text{ 来保证达到这一}$$

要求。这个建立模型的技巧也可用于其他场合。

(5) 设备负荷方程

$$\sum_{j=1}^n a_{ij} x_{js} + u_{is} = d_{is} + v_{is} (i = 1, 2, \dots, m, s = 1, 2, \dots, 12)$$

$$(6) x_{js} \geq 0 \quad u_{is} \geq 0 \quad v_{is} \geq 0 (i = 1, 2, \dots, m, j = 1, 2, \dots, n, s = 1, 2, \dots, 12)$$

变量的非负要求

这个问题中包含非线性约束条件, 所以是一个非线性规划问

题。可以把这个问题转化为求一系列无约束的极值求解。如罚函数法等。

6.2.3 生产力布局问题

1. 生产地接近消费地的生产力布局模型

这个模型适用于重量大,运费贵,产地比较集中而消费地却相对地说比较分散的很多产品。例如煤炭、木材,食盐等。由于这个模型主要根据消费地的需求来确定产品的生产地及产量,所以它从生产布局上消除了不合理运输现象。

某种产品有若干个消费地,有若干个可能的生产地。各个消费地对各种产品的需要量已知。每个生产地的现有生产能力,可能达到的最大生产能力,单位产品生产成本、单位产品的资金占用量(基建投资额)都已知。从每个生产地到各个消费地的单位产品运输费用,以及为增加生产能力所可能使用的资金总额(基建投资总额)也已知。要求从消费地的需要出发,确定该种产品的生产布局。使得到达消费地的产品劳动消耗量和资金占用量(基建投资额)达到最小。

模型中假设有 n 个消费地, m 个生产地。有关的已知数据用以下符号表示:

B_j —第 j 个消费地的需要量。

A_i —第 i 个生产地原有的生产能力,如 $A_i = 0$ 则表示该地点目前没有生产这种产品的企业。

λ_i —第 i 个生产地新建或扩建的生产能力在计划期的动用系数

H_i —根据当地条件第 i 个生产地最大可能达到的生产能力。

g_i —第 i 个地点新建(扩建)单位生产能力所需要的资金数量(基本建设投资额)。

G —为增加生产能力而可能使用的资金数量。

b_{ij} —从第 i 个生产地到第 j 个消费地的单位产品运输费用。

c_i —第 i 个生产地的单位产品生产成本。

E —资金补偿期的定额(投资回收额的定额)。

模型中的变量含义如下:

x_i —第 i 个生产地的产量。

W_i —第 i 个生产地新建或扩建的生产能力。

x_{ij} —第 i 个生产地供应第 j 个消费地的该种产品数量。

该问题数学模型如下:

选取的目标函数是求在定额的资金补偿期(定额的投资回收期)内所消耗的产品生产成本和运费,以及占用的资金总量(投资

额)之和最小。即
$$\min Z = E \left(\sum_{i=1}^m \sum_{j=1}^n b_{ij} x_{ij} + \sum_{i=1}^m c_i x_i \right) + \sum_{i=1}^m g_i W_i$$

其约束条件如下:

(1) 各个消费地从所有生产地得到的该种产品数量等于其需要量

$$\sum_{i=1}^m x_{ij} = B_j (j = 1, 2, \dots, n)$$

(2) 各个生产地对所有消费地的供应总量应等于其产量

$$\sum_{j=1}^n x_{ij} = x_i (i = 1, 2, \dots, m)$$

(3) 各个生产地的产量不能超过其生产能力

$$x_i \leq A_i + \lambda_i W_i (i = 1, 2, \dots, m)$$

$\lambda_i W_i$ 表示第 i 个生产地计划期可以动用的新增生产能力、 $A_i + \lambda_i W_i$ 表示第 i 个生产地计划期的实际生产能力。

(4) 各个生产地的建设规模不能超过规定的数额

$$A_i + W_i \leq H_i (i = 1, 2, \dots, m)$$

(5) 新增生产能力所耗用的资金总额(基建投资总额)不超过规定的数量

$$\sum_{i=1}^m g_i W_i \leq G (i = 1, 2 \cdots m)$$

(6) 变量的非负要求

$$x_i \geq 0 \quad W_i \geq 0 \quad x_{ij} \geq 0 (i = 1, 2 \cdots m) (j = 1, 2 \cdots n)$$

这个问题是一个标准的线性规划问题。利用单纯形法在电子计算机上求解,就可能找到一个满足上述六组约束条件,使目标函数达到最小的方案。但是为使这个问题有解,在确定约束条件中有关系数时,至少必须满足以下两个条件:

第一,各个生产地对建设规模的总限额应大于等于该种产品的需要量的总和。即 $\sum_{i=1}^m H_i \geq \sum_{j=1}^n B_j$ 一般情况下 $\sum_{i=1}^m H_i$ 会超过 $\sum_{j=1}^n B_j$ 很多。如这个条件不满足就表示无法通过国内生产满足需要。

第二,资金总额应大于最低限度的需要量。

$$G \geq \epsilon \left(\sum_{j=1}^n B_j - \sum_{i=1}^m A_i \right) \text{ 这里 } \left(\sum_{j=1}^n B_j - \sum_{i=1}^m A_i \right)$$

表示为了满足消费地的需要,计划期最低限度需要增加的可动用的生产能力。 $\epsilon = \min \left\{ \frac{g_i}{\lambda_i} \right\}$ 表示为得到一个单位可动用的生产能力,所需要的资金数量的最小值。

2. 生产地接近原料地的生产力布局模型

某些部门在考虑生产力布局问题时,首先应当考虑生产地接近原料地的问题。食品工业中的很多企业就属于这种情况。如奶制品企业,其原料牛奶是一种易坏的、不适宜于长途运输而且产地又比较分散的产品。又如某些冶金企业,其原料矿石的金属含量很低,长途运输的费用很高,对于这些产品可采用以下模型。

假设这些产品共有若干个原料产地,已知每个地点的原料供应量。这种产品有若干个可能的生产地,要求从原料供应情况出发

决定各个生产地的产量及基建规模,使得总的生产费用及资金占用量达到最小。

在模型中假设有 n 个原料产地和 m 个生产地,除了使用上一模型中的有关符号外,再补充以下符号:

y_{ki} —第 k 个原料地运给第 i 个生产地的原料数量。这是一个未知数。

D_k —第 k 个原料地的原料供应量。

r —生产单位产品所消耗的原料数量

a_{ki} —第 k 个原料地到第 i 个生产地原料的单位运价

c_i —第 i 个生产地生产单位产品的成本(不包括原料运费)

该问题的数学模型如下:

选取的目标函数是求定额的资金补偿期内产品的原材料运费和生产成本,以及新建生产能力的资金占用量的总和达到最小。即

$$\min Z = E \left(\sum_{k=1}^n \sum_{i=1}^m (a_{ki} y_{ki} + \sum_{i=1}^m c_i x_i) + \sum_{i=1}^m g_i W_i \right)$$

其约束条件有以下六组:

(1) 各个原料地运出的原料数量等于其供应量(保证所有原料得到加工)

$$\sum_{i=1}^m y_{ki} = D_k (k = 1, 2, \dots, n)$$

(2) 各个生产地所得到的原料数量等于其生产消耗量

$$\sum_{k=1}^n y_{ki} = r x_i (i = 1, 2, \dots, n)$$

(3) 各个生产地的产量不超过现有的生产能力和新动用的生产能力之和

$$x_i \leq A_i + \lambda_i W_i (i = 1, 2, \dots, m)$$

(4) 各个生产地的建设规模不超过规定限额

$$A_i + W_i \leq H_i (i = 1, 2, \dots, m)$$

(5) 新建生产能力所占用的资金总额(基建设资额)不超过规定的数量

$$\sum_{i=1}^m g_i W_i \leq G$$

(6) 变量的非负要求

$$y_{ki} \geq 0 \quad x_i \geq 0 \quad W_i \geq 0 (i = 1, 2 \cdots m, k = 1, 2 \cdots n)$$

为使这个问题有可能得到解,至少必须满足以下条件:

第一,所有生产地建设规模的总限额,应大于等于该种产品的产量,即 $\sum_{i=1}^m H_i \geq \sum_{k=1}^n D_k / r$ 一般说 $\sum_{i=1}^m H_i$ 应比 $\sum_{k=1}^n D_k / r$ 大很多。

第二,资金总额应大于等于最低限度的需要量, $G \geq \epsilon (\sum_{k=1}^n D_k / r - \sum_{i=1}^m H_i)$, 该式中括号内表示计划期需要新增加的可动用的生产能力, ϵ 为增加一个单位可动用的生产能力所需要的最低限度的资金数量,故不等式右端表示对资金的最低限度的需要量。

3. 生产地接近原料地和消费地的生产力布局模型

问题:某种产品的生产过程要消耗一种(或数种)运量很大的原料,原材料的供应地及各个地点的最大供应量已确定。该种产品有若干个可能的生产地,每个生产地的现有生产能力,可能建设的最大生产能力,单位产品的生产成本以及新建单位生产能力的资金需要量(基建投资额)已知。该种产品有若干个可能的消费地,其需要量已确定。从各个原料地到生产地的原料的单位运价和从各个生产地到各个消费地的产品单位运价都是已知数,要求从满足消费地对该产品的需要量出发,确定各个生产地的产量及建设规模,确定原材料和产品的合理、运输方案,使得在定额资金补偿期(定额的投资回收期)内产品的成本,原材料和成品的运费,以及新增生产能力所占用的资金总量(基建设投资额)达到最小。

数学模型如下:(所用符号含义同上)

选取目标函数为求在定额的资金补偿期内该种产品的生产成本(不包括运费),原材料运输费用和产品的运输费用,以及所占用的资金总量达到最小。即

$$\min Z = E \left(\sum_{k=1}^n \sum_{i=1}^m a_{ki} y_{ki} + \sum_{i=1}^m c_i x_i + \sum_{i=1}^m \sum_{j=1}^n b_{ij} x_{ij} \right) + \sum_{i=1}^m g_i w_i$$

其约束条件有以下八组:

(1) 各个原料地运出的原材料不能超过其供应量

$$\sum_{i=1}^m y_{ki} \leq D_k (k = 1, 2, \dots, n)$$

(2) 每个生产地从所有原料地所得到的原料的供应量应等于其需要量

$$\sum_{k=1}^n y_{ki} = r x_i (i = 1, 2, \dots, m)$$

(3) 各个生产地的产量不超过其生产能力

$$x_i \leq A_i + \lambda_i W_i (i = 1, 2, \dots, m)$$

(4) 各个生产地的生产能力不超过规定的限额

$$A_i + W_i \leq H_i (i = 1, 2, \dots, m)$$

(5) 各个生产地该种产品的运出量等于其产量

$$\sum_{j=1}^n x_{ij} = x_i (i = 1, 2, \dots, m)$$

(6) 新增生产能力所占用的资金总额(基建投资额)不超过限额。

$$\sum_{i=1}^m g_i w_i \leq G$$

(7) 每个消费地从所有生产地得到的该种产品供应量等于其需要量。

$$\sum_{i=1}^m x_{ij} = B_j (j = 1, 2 \cdots n)$$

$$(8) \text{ 变量的非负要求 } y_{ki} \geq 0 \quad x_i \geq 0 \quad W_i \geq 0 \\ x_{ij} \geq 0 (i = 1, 2 \cdots m, j = 1, 2 \cdots n, k = 1, 2 \cdots n)$$

这是一个比较全面的生产力布局模型,考虑了原料供应,产品的生产、销售和基建等方面的因素和条件,它适用于原料和成品的运输费用都比较大的情况。

为了使这个模型有解,必须满足以下条件:

第一,所有生产地的建设规模的总限额应大于一定的数量,

$$\sum_{i=1}^m H_i \geq \sum_{j=1}^n B_j$$

第二,对资金数量有一定要求, $G \geq \epsilon (\sum_{j=1}^n B_j - \sum_{i>1}^m A_i)$

第三,原材料的供应总量应达到一定数量

$$\sum_{k=1}^R D_k \geq r \sum_{j=1}^n B_j$$

4. 布局实例:黑龙江食用豆油厂的合理布局问题

作为全国大豆生产基地的黑龙江省,它的油厂规模及布局规划,对有效地开发经济和利用我国大豆资源来说,具有重大意义。下面就黑龙江粮食局管辖的 113 个油厂来讨论其布局问题。该省按地区和省辖市考虑,划分为 12 片,每片可供的大豆数量,所需油粕、豆油的数量及加工成本皆列入表 6.2.1。

表 6.2.1

序号	地区名称(省辖市)	原料 $B_j(t)$	需粕 $D_j(t)$	需油 $G_j(t)$	成本 $H_j(\text{元}/t)$
1	哈尔滨市、松花江	72465	147970	35765	124
2	齐齐哈尔、嫩江	109700	99920	19765	276
3	绥化	79000	30840	10460	211

续表

序号	地区名称(省辖市)	原料 $B_j(t)$	需粕 $D_j(t)$	需油 $G_j(t)$	成本 $H_j(\text{元}/t)$
4	牡丹江	171050	24920	9125	137
5	合 江	322920	17655	8815	170
6	黑 河	139670	10820	3220	379
7	大兴安岭	2230	2135	2195	394
8	伊 春	2880	14400	6250	289
9	鹤 岗	13920	13237	3265	160
10	鸡 西	6180	15417	4435	139
11	双鸭山	13020	18256	2115	258
12	大 庆	4660	18012	3200	270

全省油厂布局数学模型从总体最优效益出发。

引入以下符号：

x_{ij} —表示由第 i 号地区运到第 j 号地区的大豆数量。

y_j —表示第 j 号产地设厂年产油数量。

Z_{ij}, U_{ij} —分别表示由第 i 号地区运送到第 j 号地区的豆油、豆粕数量。

a_{ij} —表示两地区间的距离。

B_j —表示各地可供的大豆数量。

G_j, D_j —分别表示各地需要的豆油和豆粕数量。

E_{ij}, F_{ij}, G_{ij} —分别表示两地间大豆、豆油、豆粕的单位运价。

H_j —表示各地油厂的加工费用。

M_j, N_j —分别表示各地生产单位数量的豆油和豆粕所消耗的大豆数量。

数学模型如下：

目标函数是使总的调运费用和加工成本之和最小。

$$\min S = \sum_{i=1}^{12} \sum_{j=1}^{12} E_{ij} a_{ij} x_{ij} + \sum_{i=1}^{12} \sum_{j=1}^{12} F_{ij} a_{ij} Z_{ij} + \sum_{i=1}^{12} \sum_{j=1}^{12} G_{ij} a_{ij} U_{ij} + \sum_{j=1}^{12} H_j y_j$$

约束条件为以下 6 组。

(1) 各地大豆数及调入调出差,其总和应满足本地油厂的需要:

$$B_j + \sum_{i=1}^{12} x_{ij} - \sum_{i=1}^{12} x_{ji} \geq M_j y_j$$

(2) 某一地区自用的及调出的豆油数量,不超过本地区油厂的生产数量:

$$y_j \geq \sum_{i=1}^{12} Z_{ij}$$

(3) 从各地调往 j 地区的豆油数量,满足该地区的需要量:

$$\sum_{i=1}^{12} Z_{ij} \geq G_j$$

(4) 从 j 地区调出的豆粕的数量,应不超过该地区的豆粕生产量:

$$\frac{M_j y_j}{N_j} \geq \sum_{i=1}^{12} U_{ji}$$

(5) 从各地调往 j 地区的豆粕数量,应满足该地区对豆粕的需要量:

$$\sum_{i=1}^{12} U_{ij} \geq D_j$$

(6) 变量的非负限制

$$x_{ij} \geq 0 (i \neq j) \quad Z_{ij} \geq 0 \quad U_{ij} \geq 0 (i, j = 1, 2, \dots, 12)$$

如一号地区的约束方程为

$$72465 + \sum_{i=1}^{12} x_{i1} \geq \sum_{j=1}^{12} x_{1j} + 5.86 y_1$$

$$y_1 \geq \sum_{j=1}^{12} Z_{1j}$$

$$\sum_{i=1}^{12} Z_{i1} \geq 35765$$

$$4.51y_1 \geq \sum_{j=1}^{12} U_{1j}$$

$$\sum_{i=1}^{12} U_{i1} \geq 147970$$

这样在1号地区共得5个约束方程,同理对2-12号地区分别列出类似的约束方程(从略)再加上 $x_{ij} \geq 0 (i, j = 1, 2 \cdots 12, i \neq j)$

$$y_j \geq 0 (j = 1, 2 \cdots 12) \quad Z_{ij} \geq 0 \quad U_{ij} \geq 0 (i, j = 1, 2 \cdots 12)$$

等共同作为优化问题的约束方程。目标函数是全省油厂的总成本最小。总成本是加工成本与运输成本之和。运输成本又分为原料运费和两种成品运费等3项。这样,可以把全省油厂整体布局问题归结为上述约束条件下,使目标函数为最小值的较大规模的线性规划问题。

这个模型中共有555个变量,492个约束条件,系数形成大规模稀疏矩阵。在电子计算机上经135次迭代,得出全省油厂统一规划的最佳方案(计算结果从略)。按此方案:(1)绥化、大兴安岭二片不需建厂;(2)齐齐哈尔、大庆二片在考虑需粕量时才建厂(仅考虑油量不需建厂);(3)哈尔滨、牡丹江、鹤岗三片要建造超过该片需要,即有调出任务的大厂。例如,哈尔滨每年需分别向齐齐哈尔、绥化、大兴安岭、大庆调整19765t,10460t,2195t和3200t油,牡丹江则需分别向黑河、伊春、鸡西和双鸭山调3220t,6230t,2174t,780t油等;(4)黑河、伊春、鸡西三片只建部分满足该片需油量的油厂,不足部分由外片调入;(5)合江、双鸭山需建厂,基本自给自足。

6.2.4 赛事举例

1. AMCM-88B 题:两辆铁路平板车的装货问题

要把七种规格的包装箱装到两辆平板车上去,包装箱的宽和高都是相同的,但厚度(t ,以厘米计)及重量(w ,以千克计)却不同,下表给出了它们的厚度、重量及数量。

	C_1	C_2	C_3	C_4	C_5	C_6	C_7
t (厘米)	48.7	52.0	61.3	72.0	48.7	52.0	64.0
W (千克)	2000	3000	1000	500	4000	2000	1000
箱 数	8	7	9	6	6	4	8

每辆平板车有 10.2 米长的地方可以用来装箱(像面包片那样),载重为 40 吨,由于当地货运的限制,对 C_5, C_6, C_7 三类包装箱的总数有如下特殊约束:它们所占的空间(厚度)不得超过 302.7 厘米,试把这些包装箱装到平板车上去,而浪费的空间最小。

(1) 问题的分析

题中所有的包装箱共重 89 吨,而两辆平板车只能载 $2 \times 40 = 80$ 吨,因此不能全装下,究竟在两辆车上装哪些种,各多少个箱子才合适,必须有评价的标准,这标准是遵守题中说明的重量、厚度方面的约束条件,并且体现出尽可能多装。

由题意,只考虑象面包片重叠那样的装法,把问题简化为:两辆车上装箱总厚度之和尽可能大,这是一个典型的整数线性规划问题。

(2) 数学模型

设 x_{ij} 为在第 i 辆平板车装 C_j 箱的个数

$i = 1, 2, j = 1, 2, 3, 4, 5, 6, 7$

目标函数为
$$\max S = \sum_{i=1}^2 (0.487x_{i1} + 0.52x_{i2} + 0.613x_{i3} + 0.72x_{i4} + 0.487x_{i5} + 0.52x_{i6} + 0.64x_{i7})$$

约束条件为以下 5 组

① 箱数约束

$$x_{11} + x_{21} \leq 8$$

$$x_{12} + x_{22} \leq 7$$

$$x_{13} + x_{23} \leq 9$$

$$x_{14} + x_{24} \leq 6$$

$$x_{15} + x_{25} \leq 6$$

$$x_{16} + x_{26} \leq 4$$

$$x_{17} + x_{27} \leq 8$$

② 重量约束

$$2x_{i1} + 3x_{i2} + x_{i3} + 0.5x_{i4} + 4x_{i5} + 2x_{i6} + x_{i7} \leq 40$$

$$i = 1, 2$$

③ 厚度约束

$$0.487x_{i1} + 0.52x_{i2} + 0.613x_{i3} + 0.72x_{i4} + 0.487x_{i5} + 0.52x_{i6} + 0.64x_{i7} \leq 10.2 \quad i = 1, 2$$

④ 对 C_5, C_6, C_7 的特殊约束

$$0.487x_{i5} + 0.52x_{i6} + 0.64x_{i7} \leq 3.027 \quad i = 1, 2$$

⑤ 变量的非负限制

$$x_{ij} \geq 0 \text{ 且取整数} \quad i = 1, 2 \quad j = 1, 2, 3, 4, 5, 6, 7$$

该问题可以用分枝定界法求最优解(略)

2. AMCM - 89B 题 飞机排队

机场通常按“先来先走”的原则来分配飞机跑道,即当飞机准备好离开登机口时,驾驶员电告地面控制中心,加入等候跑道的队伍。

假设控制中心可以从快速联机数据库中得到每架飞机的如下方面信息:

① 预定离开登机口的时间

② 实际离开登机口的时间

© 机上乘客人数

④ 预定在下一站转机的人数和转机的时间

⑤ 到达下一站的预定时间

又设共有七种飞机,载客量从 100 人起以 50 人递增,载客最多的一种达 400 人

试开发和分析一种能使乘客和航空公司双方满意的数学模型。

① 问题的分析

不妨设机场只有一条跑道供起飞用,又设任何飞机起飞占用此跑道的时间为 d ,于是把从 $t = 0$ 开始的时间离散化为长 d 的若干间隔,这些时间间隔是一个接一个,它们是

$$[0, d], [d, 2d] \cdots [(n-1)d, nd]$$

其中 n 为 $t = 0$ 时刻请求起飞的飞机数

此题要求安排的起飞顺序使乘客和航空公司双方都满意,这个满意的准则需要探讨,乘客最不满意的事(除不安全以外)莫过于起飞过晚影响了下一程旅行,而航空公司最不满意的事也可以说是由于飞机迟飞而支付的费用总和过大,基于这样的分析,可以设想每架飞机的起飞时间有其约束条件,而在满足这些约束的前提下使总费用最小,就不失为解决此问题的合理思路。

② 数学模型

$$\text{设 } x_{ij} = \begin{cases} 1, & \text{第 } i \text{ 架飞机被安排在 } [(j-1)d, jd] \text{ 起飞} \\ 0, & \text{其他情况} \end{cases} \quad j \leq j(i), i = 1, 2, \cdots, n$$

这里的 $j(i)$ 是这样的,若第 i 架飞机安排起飞的间隔时序数 $j > j(i)$,由于影响旅客下一程旅行而绝不能被他们容许,比方说所有旅客都会耽误准点搭乘下一程的飞机、轮船、火车之类,这 $j(i)$ 需根据 ③④⑤ 等方面信息来确定。

设 C_{ij} 为 $x_{ij} = 1$ 时航空公司支付的费用,它由 ③④⑤ 等

方面信息来确定。于是 $\sum_{i=1}^n \sum_{j=1}^{j(i)} c_{ij} x_{ij}$ 是所支付费用的总和

$$\text{目标函数 } \min S = \sum_{i=1}^n \sum_{j=1}^{j(i)} c_{ij} x_{ij}$$

约束条件为以下 3 组

1° 每次只能安排一架飞机起飞

$$\sum_{i=1}^n x_{ij} = 1 \quad j = 1, 2, \dots, n$$

2° 每架飞机只能占一个时间间隔

$$\sum_{j=1}^{j(i)} x_{ij} = 1 \quad i = 1, 2, \dots, n$$

3° 变量的 0-1 限制

$$x_{ij} \in \{0, 1\} \quad j \leq j(i) \quad i = 1, 2, \dots, n$$

这是 0-1 型整数规划, 可用隐枚举法求解。

3. CMCM-95A 题 飞机管理问题

原题参看 1995 年全国大学生数学建模竞赛 A 题。

为使问题简化, 作出如下假设

- ① 不碰撞的标准是任意两架飞机的距离大于 $8km$
- ② 飞机飞行方向角调整幅度不应超过 30 度, 而且要尽可能小
- ③ 所有飞机飞行速度为 $800km/h$ 不受其他因素影响
- ④ 进入该区域的飞机在到达区域边缘时, 与区域内的飞机的

距离应在 $60km$ 以上

⑤ 不考虑飞机离开区域后的情况

⑥ 建模时暂时考虑 6 架飞机

以上是题中所给假设, 补充以下假设:

① 飞机在所定区域内作直线飞行, 不偏离航向

② 飞行管理系统发出的转向指令应被飞机立即执行, 即认为转向是在瞬间完成

③ 飞机在区域内不发生意外,如发动机失灵,或其他意外原因迫使飞机改变航向

设 6 架飞机在调整时的方向角为 θ_i , 调整后的方向角为 $\theta_i' = \theta_i + \Delta\theta_i (i = 1, 2, 3, 4, 5, 6)$ 又设任意两架飞机在区域内的最短距离为 $d_{ij}(\theta_i, \theta_j)$

$$d_{ij}^2 = (x_{i0} + \theta_i \cos \theta_{i0} - \theta_j \cos \theta_{j0})^2 + (y_{i0} + \theta_i \sin \theta_{i0} - y_{j0} - \theta_j \sin \theta_{j0})^2$$

于是,有目标函数

$$\min S = \sum_{i=1}^6 |\Delta\theta_i|$$

$$S. t. \quad \begin{aligned} d_{ij}(\theta_i + \Delta\theta_i, \theta_j + \Delta\theta_j) &> 8 \\ |\Delta\theta_i| &\leq 30^\circ \end{aligned}$$

$$i = 1, 2, 3, 4, 5, 6 \quad j = 1, 2, 3, 4, 5, 6 \quad i \neq j$$

这是一个非线性规划问题,可以用最速下降法求解但很困难,根据要求可以以 0.01 为单位搜索,但必须与其他方法相配合。比如二分法等,否则所用时间太长。

习 题

1. 某昼夜服务的公交线路每天各时间区段内所需司机和乘务人员数如下:

班次	时 间	所需人数
1	6 : 00—10 : 00	60
2	10 : 00—14 : 00	70
3	14 : 00—18 : 00	60

4	18:00—22:00	50
5	22:00—2:00	20
6	2:00—6:00	30

设司机和乘务人员分别在各时间区段一开始时上班,并连续工作八小时,问该公交线路至少配备多少名司机和乘务人员,写出这个问题的线性规划模型。

2. 某糖果厂用原料 A、B、C 加工成三种不同牌号的糖甲、乙、丙,已知各种牌号糖果中 A、B、C 含量,原料成本,各种原料的每月限制用量,三种牌号糖果的单位加工费及售价如下表所示

	甲	乙	丙	原料成本(元/千克)	每月限制用量(千克)
A	$\geq 60\%$	$\geq 15\%$		2.00	2000
B				1.50	2500
C	$\leq 20\%$	$\leq 60\%$	$\leq 50\%$	1.00	1200
加工费(元/千克)	0.50	0.40	0.30		
售 价	3.40	2.85	2.25		

问该厂每月应生产这三种牌号糖果各多少千克,使该厂获利最大?试建立这个问题的线性规划的数学模型。

3. 一项食品加工业,为使几种粗油精炼,然后加以混合成为成品油。原料油有两大类,共 5 种:其中植物油 2 种,分别记作 V_1 和 V_2 ;非植物油 3 种,记为 O_1 , O_2 和 O_3 。各种原料油均从市场采购。现在(一月份)和未来半年中,市场价格(元/吨)如下表所示:

油 月份	V_1	V_2	O_1	O_2	O_3
一	1100	1200	1300	1100	1150
二	1300	1300	1100	900	1150
三	1100	1400	1300	1000	950
四	1200	1100	1200	1200	1250

油 月份	V_1	V_2	O_1	O_2	O_3
五	1000	1200	1500	1100	1050
六	900	1000	1400	800	1350

成品油售价 1500 元 / 吨

植物油和非植物油要在不同的生产线精炼。每个月最多可精炼植物油 200 吨, 非植物油 250 吨。精炼过程中没有重量损失, 精炼费用可以忽略。

每种原料油量最多可存贮 1000 吨备用。存贮费为每吨每月 50 元。成品油和经过精炼的原料油不能存贮。

对成品油限定其硬度在 3 至 6 单位之间。各种原料油的硬度如下表所示:

油	V_1	V_2	O_1	O_2	O_3
硬度	8.8	6.1	2.0	4.2	5.0

假设硬度是线性地混合的。

为使公司获得最大利润, 应取什么样的采购和加工方案。

现存有 5 种原料油每种 500 吨, 要求在 6 月底仍然有这样多存货。

研究总利润和采购与加工方案适应不同的未来市场价格应如何变化。考虑如下的价格变化方式: 2 月份植物油价上升 $x\%$, 非植物油价上升 $2x\%$; 3 月份植物油价升 $2x\%$, 非植物油上升 $4x\%$; 其余月份保持这种线性的上升势头。对不同的 x 值 (直到 20), 就方案的必要的变化及对总利润的影响, 作出全面计划。

4. 某厂拥有 4 台磨床、2 台立式钻床、一台镗床和一台刨床, 用以生产 7 种产品, 记作 p_1 至 p_7 。工厂收益规定为产品售价减去原材料费用之剩余。每种产品单件的收益及所需各机床的加工工时 (以小时计) 列于下表:

产 品	P_1	P_2	P_3	P_4	P_5	P_6	P_7
收 益	10	6	8	4	11	9	3
磨	0.5	0.7	0	0	0.3	0.2	0.5
垂直钻孔	0.1	0.2	0	0.3	0	0.6	0
水平钻孔	0.2	0.8	0.8	0	0	0	0.6
镗 孔	0.05	0.03	0	0.07	0.1	0	0.08
刨	0	0	0.01	0	0.05	0	0.05

本月(一月)和随后的5个月中,下列机床停工维修:

一月:磨床一台

二月:卧式钻床二台

三月:镗床一台

四月:立式钻床一台

五月:磨床一台,立式钻床一台

六月:刨床一台,卧式钻床一台

各种产品各月份的市场容量如下表:

产品	P_1	P_2	P_3	P_4	P_5	P_6	P_7
一月	500	1000	300	300	800	200	100
二月	600	500	200	0	400	300	150
三月	300	600	0	0	500	400	100
四月	200	300	400	500	200	0	100
五月	0	100	500	100	1000	300	0
六月	500	500	100	300	1100	500	60

每种产品有货最多可到100件。存费每件每月为0.5。现在无存货。要求到6月底每种产品有存货50件。

工厂每周工作 6 天,每天 2 班,每班 8 小时。不需要考虑排队等待加工的问题。

为使收益最大,工厂应如何安排各月份各种产品的产量?考虑价格的某种变化及引入新机床对计划和收益的影响。

注意,可假设每月仅有 24 个工作日。

5. 某炼油厂购买两种原油(C_1 和 C_2),经过蒸馏、重整、分裂和混合等 4 种加工过程,生成汽油和燃料出售。

蒸馏:该过程将两种原油按沸点不同分离成下列分馏物;轻石油精、中石油精、重石油精、轻油、重油和残渣。轻、中和重石油精的辛烷值分别为 90、80 和 70。一桶原油生成分馏物的桶数如下表所示

原 油 \ 分 馏 物	石 油 精			油		残 渣
	轻	中	重	轻	重	
C_1	0.1	0.2	0.2	0.12	0.12	0.13
C_2	0.15	0.25	0.18	0.08	0.19	0.12

注:蒸馏过程中有少量损耗。

重整:石油精可直接用于混合产生各种品级的汽油,也可以再经过一种称作重整的加工过程,生成一种称为重整汽油的产品,其甲烷值为 115。不同石油精每桶产重整汽油如下表:

石油精	轻	中	重	
重整汽油	0.6	0.52	0.45	

裂解:轻油和重油可以直接混合生产飞机燃料和燃料油,也可以经过催化裂解,生成裂化油和裂化汽油,后者辛烷值为 105。一桶轻油经裂解生成裂化油 0.68 桶和裂化汽油 0.28 桶;一桶重油相应的生成量为 0.75 和 0.2 桶。裂化油用于混合生成飞机燃料和

燃料油；裂化汽油用于混合生成汽油。

残渣可用于生产润滑油，或混合入飞机燃料或燃料中。一桶残渣产生 0.5 桶润滑油混合。

汽油(发动机燃料)

汽油有两种：普通汽油和特级汽油，通过将石油精，重整汽油和裂化汽油得到。普通汽油必须有不小于 84 的辛烷值。特级汽油必须有不小于 94 的辛烷值。假设辛烷值是按体积线性地混合的，即混合物的辛烷值是各组分的辛烷值的组分的体积为权的加权平均。

飞机燃料：飞机燃料的蒸发压不能超过 $1\text{kg}/\text{cm}^2$ 。轻油、重油、裂化油和残渣的蒸发压依次为 1.0, 0.6, 1.5 和 $0.05\text{kg}/\text{cm}^2$ 。同样假设蒸发压是按体积线性地混合的。

燃料油：轻油、裂化油、重油和残渣按 10 : 4 : 3 : 1 的比例混合生成燃料油。

关于原料供应和加工能力的限制条件有：

(a) C_1 日供 20000 桶

(b) C_2 日供 30000 桶

(c) 日蒸馏原油最多 45000 桶

(d) 日重整石油精最多 10000 桶

(e) 日裂解油最多 8000 桶

(f) 日产润滑油必须左 500 桶至 1000 桶之间

(g) 特级汽油产量必须是普通汽油产量的 40%

售出一桶成品，利润数如表所示：

产品	特级汽油	普通汽油	飞机燃料	燃料油	润滑油
利润	70 元	60 元	40 元	35 元	15 元

炼油厂应如何安排生产，总利润为最大？

6. 某地区有 4 个矿，产同一种矿石。某采矿公司获得了这些矿在未来连续 5 年中的开采权。但在每一年度中，该公司最多有能力

开 3 个矿,而有一矿闲置。对于闲置的矿,如果在这 5 年期内随后的某年还要开采,则不能关闭;如果从闲置起在这 5 年内不再开采,就关闭。对开采和保持不关闭的矿,公司应交付土地使用费。各矿每年土地使用费额见表 1 第 2 行。各矿每年矿砂产量上限如表 1 第 3 行。不同矿所产矿砂质量不同。矿砂质量用一种质量指数表示,见表 1 第 4 行。将不同矿的矿砂混合所成的矿砂,其质量指数为各组份的质量指数的线性组合,组合系数为各组份在混成矿砂中所占的重量百分数。例如,等量的二矿砂混合,混成矿砂的质量指数为二组份指数的平均值。每一年度公司将各矿全年产出的矿砂混合,要生成具有约定指数的矿砂。不同年度的约定质量指数如表 2 所示。各年度产品矿砂售价每吨 10 元。年度总收入和费用开支,为扣除物价上涨因素,以逐年 9 折计入 5 年总收入和总费用中。

表 1

矿	1	2	3	4
土地使用费(万元)	500	400	400	500
产量上限(万吨)	200	250	130	300
质量指数	1.0	0.7	1.5	0.5

表 2

年度	1	2	3	4	5
质量指标	0.9	0.8	1.2	0.6	1.0

试问各年度应开采哪几个矿?产量应各为多少?

7. 英国某农场主有 200 英亩土地的农场,用来饲养奶牛。现在要为未来五年制订生产计划。

现在他有 120 头母牛,其中 20 头为不到 2 岁的幼牛,100 头为

产奶牛。每头幼牛需用 $2/3$ 英亩土地供养,每头产奶牛需用一英亩。产奶牛平均每头每年生 1.1 头牛,其中一半为公牛,生出后不久即卖掉,平均每头卖 30 英磅;另一半为母牛,可以在生出后不久卖掉,平均每头卖 40 英磅,也可以留下饲养,养至 2 岁成为产奶牛。幼牛年损失 5%;产奶牛年损失 2%。产奶牛养到满 12 岁就卖掉,平均每头卖 120 英磅。现有的 20 头幼牛,0 岁和 1 岁各 10 头;100 头产奶牛,从 2 岁至 11 岁,每一年龄的都有 10 头。应该卖掉的小母牛都已卖掉,所有 20 头是要饲养成产奶牛的。

一头牛所产的奶提供年收入 370 英磅。现在最多只能饲养 130 头牛。超过此数每多养一头要投资 200 英磅。每头产奶牛每年消耗 0.6 吨粮食和 0.7 吨甜菜。粮食和甜菜可以由农场种植出来。每英亩产甜菜 1.5 吨。只有 80 英亩的土地适于种粮食,且产量不同。按产量可分作 4 组:第一组 20 英亩,亩产 1.1 吨;第二组 30 英亩,亩产 0.9 吨;第三组 20 英亩,0.8 吨;第四组 10 英亩,0.65 吨。从市场购粮食每吨 90 英磅,卖粮食每吨 75 英磅。买甜菜每吨 70 英磅,卖出 50 英磅。

养牛和种植所需劳动量为:每头牛每年 10 小时;每头产奶牛每年 42 小时;种一英亩粮食每年需 4 小时;种一英亩甜菜需 14 小时。

其他费用:每头幼牛每年 50 英磅;产奶牛每头每年 100 英磅;种粮食每英亩每年 15 英磅;种甜菜每年 10 英磅;劳动费用现在每年为 4000 英磅,提供 5500 小时的劳动量。超过此数的劳动量每小时费用为 1.2 英磅。

任何投资支出都从 10 年期贷款得到。贷款年利率 15%,每年偿还本息总和的 $1/10$,十年还清。每年货币的收支之差不能为负值。此外农场主不希望产奶牛的数目在五年末较现在减少超过 50%,也不希望增加超过 75%。

应如何安排 5 年的生产,使收益为最大?

第七章 层次分析法建模

层次分析法(Analytic Hierarchy Process 简称 AHP) 是美国运筹学家 T. L. Saaty 教授于 70 年代初期提出的一种简便、灵活而又实用的多准则决策方法,它把一个复杂问题分解成一些因素,并按支配关系形成层次结构,然后应用两两比较的方法确定决策方案的相对重要性,选出最优方案。层次分析法特别适用于非结构化问题的建模。本章将简要介绍层次分析法的基本方法和步骤以及在实际建模中的应用。

§ 7.1 预备知识

7.1.1 AHP 的基本方法和步骤

运用 AHP 解决问题,大体可以分为四个步骤。

第一,分析系统中各因素之间的关系,建立系统的递阶层次结构

第二,对同一层次的各元素关于上一层次中某一准则的重要性进行两两比较,构造两两比较判断矩阵。

第三,由判断矩阵计算被比较元素关于该准则的相对权重。

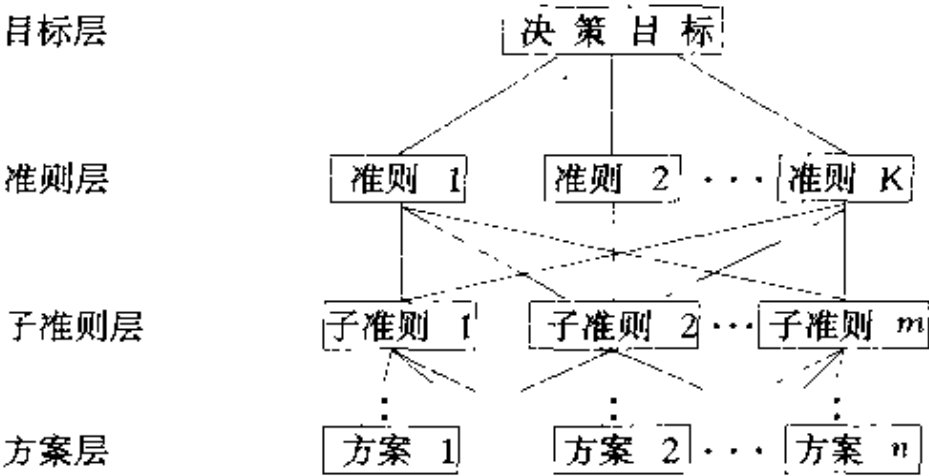
第四,计算各层元素对系统目标的合成权重,并进行排序。

下面分别说明这四个步骤的实现方法。

1. 建立递阶层次结构

这是 AHP 中最重要的一步。首先,把复杂问题分解为称之为元素的各组成部分,把这些元素按属性不同分成若干组,以形成不同层次。同一层次的元素作为准则,对下一层次的某些元素起支配

作用,同时它又受到上一层次元素的支配。这种从上至下的支配关系形成了一个递阶层次。处于最上面的层次通常只有一个元素,一般是分析问题的预定目标,或理想结果。中间的层次一般是准则,子准则。最低一层包括决策的方案。层次之间元素的支配关系不一定是完全的,即可以存在这样的元素,它并不支配下一层次的所有元素。一个典型的层次可以用下图表示出来。



层次数与问题的复杂程度和所需要分析的详尽程度有关。一个好的层次结构对于解决问题是极为重要的。层次结构建立在决策者对所面临的问题具有全面深入的认识基础上,如果在层次的划分和确定层次之间的支配关系上举棋不定,最好重新分析问题,弄懂问题各部分相互之间的关系。有时一个复杂问题仅仅用递阶层次形式表示是不够的,需要采用更复杂结构形式,如循环层次结构、反馈层次结构等,这些结构是在递阶结构基础上的扩展形式。

2. 构造两两比较判断矩阵

在建立递阶层次结构以后,上下层次之间元素的隶属关系就被确立了。假定上一层次元素 C_k 作为准则,对下一层次元素 A_1, A_2, \dots, A_n 有支配关系,我们的目的是在准则 C_k 之下按它们

相对重要性赋予 A_1, A_2, \dots, A_n 相应的权重。对于大多数社会经济问题,特别是对于人的判断起重要作用的问题,直接得到这些元素的权重并不容易,往往需要通过适当的方法导出它们的权重。AHP 所用的是两两比较的方法。

在这一步中,决策者要反复回答问题:针对准则 C_k ,两个元素 A_i 和 A_j 哪一个更重要些,重要多少,需要对重要多少赋予一定数值。这里使用 1—9 的比较标度,它们的意义见表 7.1。例如,准则是社会经济效益,子准则可分为经济、社会和环境效益。如果认为经济效益比社会效益明显重要,它们的比例标度取 5。而社会效益对于经济效益的比例标度则取 $1/5$ 。对于 n 个元素来说,得到两两比较判断矩阵 A 。

$A = (a_{ij})_{n \times n}$ 判断矩阵具有如下性质:

$$(1) a_{ij} > 0 \quad (2) a_{ij} = \frac{1}{a_{ji}} \quad (3) a_{ii} = 1$$

我们称 A 为正的互反矩阵,由于性质(2)(3)事实上,对于 n 阶判断矩阵仅需对其上(下)三角元素共 $\frac{n(n-1)}{2}$ 个给出判断。 A 的元素不一定具有传递性,即未必成立等式:

$$a_{ij} \cdot a_{jk} = a_{ik} \quad (7.1.1)$$

但如果满足此式,则称 A 为一致性矩阵。在说明由判断矩阵导出元素排序权值时,一致性矩阵有重要意义。

表 7.1.1 标度的含义

1	表示两个元素相比,具有同样重要性
3	表示两个元素相比,一个元素比另一个元素稍微重要
5	表示两个元素相比,一个元素比另一个元素明显重要
7	表示两个元素相比,一个元素比另一个元素强烈重要
9	表示两个元素相比,一个元素比另一个元素极端重要

2、4、6、8 为上述相邻判断的中值

1—9 的标度方法是将思维判断数量化的一种好方法。首先在区分事物质的差别时,人们总是用相同、较强、强、很强、极端强的语言,再进一步细分,可以在相邻的两级中插入折衷的提法,因此对于大多数决策判断来说 1—9 级的标度是适用的。其次,心理学实验表明,大多数人对不同事物在相同属性上差别的分辨能力在 5—9 级之间,采用 1—9 的标度反映多数人的判断能力。第三,当被比较的元素其属性处于不同的数量级,一般需要将较高数量级的元素进一步分解,这可以保证被比较元素在所考虑的属性上有同一个数量级或比较接近,从而适用于 1—9 的标度。当然根据问题的特点也可以采用别的类型标度方法。如 0—1 的标度,指数型的标度等。

若因素 i 与 j 比较得 a_{ij} , 则因素 j 与 i 比较的判断为 $1/a_{ij}$ 。

3. 计算单一准则下元素的相对权重

这一步要解决在准则 C_k 下, n 个元素 A_1, A_2, \dots, A_n 排序权重的计算问题, 并进行一致性检验。对于 A_1, A_2, \dots, A_n 通过两两比较得到判断矩阵 A , 解特征根问题 $AW = \lambda_{\max} W$ 所得到的 W 经正规化后作为元素 A_1, A_2, \dots, A_n 在准则 C_k 下排序权重, 这种方法称排序权重向量的特征根方法。 λ_{\max} 存在且惟一, W 可以由正分量组成, 除了差一个常数倍数外, W 是惟一的。 λ_{\max} 和 W 的计算方法一般采用幂法, 其步骤为:

(1) 设初值向量 W 例如 $W_0 = (\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n})^T$

(2) 对于 $k = 1, 2, 3, \dots$ 计算 $\bar{W}_k = AW_{k-1}$

式中 W_{k-1} 为经归一化所得到的向量。

(3) 对于事先给定的计算精度, 若 $\max | \bar{W}_{ki} - \bar{W}_{(k-1)i} | < \epsilon$ (式中 \bar{W}_{ki} 表示 \bar{W}_k 的第 i 个分量), 则计算停止, 否则继续(2)

(4) 计算 $\lambda_{\max} = \frac{1}{n} \sum_{i=1}^n \frac{\bar{W}_{ki}}{\bar{W}_{(k-1)i}}$ (7.1.2)

$$W_{kt} = \frac{W_{kt}}{\sum_{j=1}^n \bar{W}_{kj}} \quad (7.1.3)$$

在精度要求不高的情况下,可以用近似方法计算 λ_{\max} 和 W , 这里介绍两种方法:

(1) 和法

步骤为:

$$1^\circ \quad A \text{ 的元素按列归一化: } \bar{a}_{ij} = \frac{a_{ij}}{\sum_{k=1}^n a_{kj}} \quad (i, j = 1, 2, \dots, n)$$

$$2^\circ \quad \text{将归一化的元素各列相加: } \bar{W}_i = \sum_{j=1}^n \bar{a}_{ij} \quad (i, j = 1, 2, \dots, n)$$

$$3^\circ \quad \text{将所得向量归一化得排序权向量: } W_i = \frac{\bar{W}_i}{\sum_{i=1}^n \bar{W}_i}$$

$$4^\circ \quad \text{按下列公式计算 } \lambda_{\max}: \lambda_{\max} = \sum_{i=1}^n \frac{(AW_i)_i}{nW_i} \quad (7.1.4)$$

式中 $(AW_i)_i$ 表示 AW_i 的第 i 个元素。

(2) 根法

步骤为:

$$1^\circ \quad A \text{ 的元素按行相乘: } M_i = \prod_{j=1}^n a_{ij} \quad (i, j = 1, 2, \dots, n)$$

$$2^\circ \quad \text{所得到的乘积分别开 } n \text{ 次方: } \bar{W}_i = \sqrt[n]{M_i}$$

$$3^\circ \quad \text{将方根向量归一化即得排序权向量: } W_i = \frac{\bar{W}_i}{\sum_{i=1}^n \bar{W}_i}$$

$$4^\circ \quad \text{由下式计算 } \lambda_{\max}: \lambda_{\max} = \sum_{i=1}^n \frac{(AW_i)_i}{nW_i} \quad (7.1.5)$$

特征根方法是 AHP 中最早提出的排序权向量计算方法,使用广泛。近年来,不少学者提出了排序权向量计算的其他方法,如最小二乘法,对数最小二乘法,上三角元素法等等,这些方法在不同场合下运用各有优点,由于篇幅所限,这里不再介绍。另外上述两种计算现在都有软件包可供使用。

在判断矩阵的构造中,并不要求判断具有一致性,即不要求 $a_{ij} \cdot a_{jk} = a_{ik}$ 成立,这是由于客观事物的复杂性与人的认识多样性所决定的。但要求判断有大体的一致性却是应该的,出现甲比乙极端重要,乙比丙极端重要,而丙比甲极端重要的情况一般是违反常识的。而且,当判断偏离一致性过大时,排序权向量计算结果作为决策依据将出现某些问题。因此在得到 λ_{\max} 后,需要进行一致性检验,其步骤为:

① 计算一致性指标

$$C.I = \frac{\lambda_{\max} - n}{n - 1} \quad (7.1.6)$$

式中 n 为判断矩阵的阶数。

② 平均随机一致性指标 $R.I$

表 2 给出 3~15 阶正互反矩阵计算 1000 次得到的平均随机一致性指标。

表 7.1.2 平均随机一致性指标 $R.I$

矩阶数	1	2	3	4	5	6	7	8
$R.I$	0	0	0.52	0.89	1.12	1.26	1.36	1.41
矩阶数	9	10	11	12	13	14	15	
$R.I$	1.46	1.49	1.52	1.54	1.56	1.58	1.59	

③ 计算一致性比例 $C.R$

$$C.R = \frac{C.I}{R.I} \quad (7.1.7)$$

当 $C.R < 0.1$ 时,一般认为判断矩阵的一致性是可以接受的。

4. 计算各层元素的组合权重

为了得到递阶层次结构中每一层次中所有元素对于总目标的相对权重,需要把第三步的计算结果进行适当的组合,并进行总的判断一致性检验。这一步骤是由上而下逐层进行的。最终计算结果得出最低层次元素,即决策方案优先顺序的相对权重和整个递阶层次模型的判断一致性检验。

假定已经计算出第 $k-1$ 层元素相对于总目标的组合排序权重向量 $a^{k-1} = (a_1^{k-1}, a_2^{k-1}, \dots, a_m^{k-1})^T$, 第 k 层在第 $k-1$ 层第 j 个元素作为准则下元素的排序权重向量为 $b_j^k = (b_{1j}^k, b_{2j}^k, \dots, b_{nj}^k)^T$, 其中不受支配(即与 $k-1$ 层第 j 个元素无关)的元素权重为零。 $B^k = (b_1^k, \dots, b_m^k)$, 则第 k 层元素相对于总目标的组合排序权重向量由下式给出: $a^k = B^k a^{k-1}$, 更一般地, 有排序的组合权重公式为 $a^k = B^k \cdots B^3 a^2$, 式中 a^2 为第二层次元素的排序向量 $3 \leq k \leq h$ (h 为层次数)。

对于递阶层次组合判断的一致性检验,需要类似地逐层计算 CI 。若分别得到了第 $k-1$ 层次的计算结果 CI_{k-1} , RI_{k-1} 和 CR_{k-1} , 则第 k 层的相应指标为

$$CI_k = (CI_k^1, \dots, CI_k^m) a^{k-1} \quad (7.1.8)$$

$$RI_k = (RI_k^1, \dots, RI_k^m) a^{k-1} \quad (7.1.9)$$

$$CR_k = CR_{k-1} + \frac{CI_k}{RI_k} \quad (7.1.10)$$

这里 CI_k^i 和 RI_k^i 分别为在 $k-1$ 层第 i 个准则下判断矩阵的一致性指标和平均随机一致性指标。当 $CR_k < 0.1$ 时,认为递阶层次在 k 层水平上整个判断有满意的一致性。

AHP 的最终结果是得到相对于总的目标各决策方案的优先顺序权重,并给出这一组合排序权重所依据的整个递阶层次结构所有判断的总的一致性指标,据此可以做出决策。

7.1.2 不完全判断信息下的排序问题

应用 AHP 进行决策时,人们对于每个准则都需要填写一个判

断矩阵。每个判断矩阵需要进行 $n(n-1)/2$ 次两两比较。当层次很多,因素复杂时,总的判断量有时是很大的。很可能出现参与决策的某个专家对某些判断缺少把握、不感兴趣,或对某些比较敏感的问题不想发表意见的情形。这种情形是应当允许的,否则,勉强他们发表意见反而可能掩盖事物的本质。但这时得到的判断矩阵的元素必定会有空缺,这种判断称为残缺判断,由残缺判断矩阵导出的元素相对重要性的排序问题,称为不完全信息下的排序问题。

显然一个判断矩阵的残缺程度对排序正确性是有明显影响的。信息越少,排序的随意性越大。要能够进行排序,必须对残缺程度及其位置有一些限制,也就是要研究什么样的残缺矩阵是“可接受的”。一个可接受的矩阵至少应有多少个元素?其次,当残缺矩阵可接受时,应如何计算排序向量及如何进行一致性检验等。

1. 残缺判断矩阵可接受的条件

对于残缺判断矩阵,暂且用“0”表示其残缺元素,这里的“0”只是一个残缺元素的记号,并非元素值为零。显然当 $a_{ij} = 0$ 时,亦必有 $a_{ji} = 0$ 。

处理残缺判断的出发点是,当缺少某个元素的直接信息时,希望最终能通过间接的判断比较获得该元素的信息。例如元素 a_{ij} 可以通过 a_{ik} 、 a_{kj} 获得。如果这种间接获取信息的渠道比较多,那么,就有可能比较准确地对残缺元素作出估计。

如果 $(i, j) \cap (k, l) \neq \emptyset$, 元素 a_{ij}, a_{kl} 称为是相邻接的,对于残缺元素 a_{ij} , 如果存在互相邻接的非残缺元素 $a_{ik}, a_{k_1 k_2}, \dots, a_{k_j}$, 那么称 a_{ij} 是可以间接获得的。

定义 7.1.1 一个残缺判断矩阵是可接受的,如果它的任一残缺元素都可通过已给出的元素间接获得,否则就是不可接受的。

可以证明一个残缺判断矩阵是可接受的必要条件是除对角元素外,每行每列至少有一个给定元素,或者说至少要作 $(n-1)$ 个判断。

定义 7.1.2 方阵 A 若能用行列同时调换化成 $\begin{bmatrix} A_1 & 0 \\ A_2 & A_4 \end{bmatrix}$ 形式, 则 A 称为可约矩阵, 否则 A 称为不可约矩阵, 这里 A_1, A_4 都是方阵。

例如矩阵

$$\begin{pmatrix} 1 & 2 & 0 \\ \frac{1}{2} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 & 3 & 2 \\ 0 & 1 & 0 & 1 \\ \frac{1}{3} & 0 & 1 & \frac{1}{3} \\ \frac{1}{2} & 0 & \frac{1}{3} & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 2 & 0 & 0 \\ \frac{1}{2} & 1 & 0 & 0 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & \frac{1}{2} & 1 \end{pmatrix}$$

都是可约矩阵。

定理 7.1.1 设 A 为非负方阵, 且其阶数为 n , 则 A 为不可约的充分必要条件是存在正整数 $S \leq n-1$, 使得 $A^S > 0$ (证明略)

定理 7.1.2 一个残缺矩阵 A 是可接受的充分必要条件是 A 为不可约的。(证明略)

2. 残缺矩阵排序向量计算方法

这里我们只介绍特征根方法

对残缺矩阵 A 构造辅助矩阵 C , 使得

$$C = (C_{ij}) = \begin{cases} a_{ij} & a_{ij} \text{ 非残缺元素 } i \neq j \\ 1 & i = j \\ W_i/W_j & a_{ij} \text{ 为残缺元素, } i \neq j \end{cases}$$

求 C 的特征根问题 $CW = \lambda_{\max} W$

$$\text{例如: } A = \begin{pmatrix} 1 & 2 & 0 \\ \frac{1}{2} & 1 & 2 \\ 0 & \frac{1}{2} & 1 \end{pmatrix} \quad W = (W_1 W_2 W_3)^T$$

作辅助矩阵 C

$$C = \begin{bmatrix} 1 & 2 & W_1/W_3 \\ \frac{1}{2} & 1 & 2 \\ W_3/W_1 & \frac{1}{2} & 1 \end{bmatrix}$$

解特征根问题 $CW = \lambda_{\max} W$

即

$$\begin{bmatrix} 1 & 2 & W_1/W_3 \\ \frac{1}{2} & 1 & 2 \\ W_3/W_1 & \frac{1}{2} & 1 \end{bmatrix} \begin{bmatrix} W_1 \\ W_2 \\ W_3 \end{bmatrix}$$

$$= (2W_1 + 2W_2, \frac{1}{2}W_1 + W_2 + 2W_3, \frac{1}{2}W_2 + 2W_3)$$

$$= \lambda_{\max}(W_1, W_2, W_3) \text{ 由此解得 } \lambda_{\max} = 3$$

$$W = (0.5714, 0.2857, 0.1429)^T$$

对于残缺矩阵 A , 作矩阵 \bar{A}

$$\bar{A} = (\bar{a}_{ij}) = \begin{cases} a_{ij} & \text{当 } a_{ij} \text{ 为非残缺元素 } i \neq j \\ 0 & \text{当 } a_{ij} \text{ 为残缺元素} \\ m_i + 1 & \text{当 } i = j, i = 1, 2, \dots, n \end{cases}$$

其中 m_i 为 A 的第 i 行中残缺元素个数。

如对上述的 A 有 $\bar{A} = \begin{bmatrix} 2 & 2 & 0 \\ \frac{1}{2} & 1 & 2 \\ 0 & \frac{1}{2} & 2 \end{bmatrix}$

由此可见求解特征根问题 $CW = \lambda_{\max} W$ 与 $\bar{A}W = \lambda_{\max} W$ 完全等价, 称 \bar{A} 为 A 的等价矩阵。直接求 \bar{A} 的特征根问题即可求得不完全信息下的排序向量。

对于一般的 A , 当 A 有一个残缺元素 $a_{ij} = 0$ 时, 它的辅助矩阵 C 及等价矩阵 \bar{A} 分别为:

$$C = \begin{matrix} & & & i & & j & & \\ \begin{pmatrix} 1 & a_{12} & \cdots & \cdots & \cdots & \cdots & \cdots & a_{1n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{21} & 1 & & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \cdots & \cdots & \cdots & 1 & \cdots & \frac{W_i}{W_j} & \cdots & \cdots \\ \vdots & & & \ddots & \vdots & \vdots & \vdots & \vdots \\ \cdots & \cdots & \cdots & \frac{W_j}{W_i} & 1 & \cdots & \cdots & \vdots \\ \vdots & & & \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{n1} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 1 \end{pmatrix} & \begin{matrix} i \\ j \end{matrix} \end{matrix}$$

$$\bar{A} = \begin{matrix} & & & i & & j & & \\ \begin{pmatrix} 1 & a_{12} & \cdots & \cdots & \cdots & \cdots & \cdots & a_{1n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{21} & 1 & & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \cdots & \cdots & \cdots & 2 & \cdots & 0 & \cdots & \cdots \\ \vdots & & & \ddots & \vdots & \vdots & \vdots & \vdots \\ \cdots & \cdots & \cdots & 0 & \cdots & 2 & \cdots & \vdots \\ \vdots & & & \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{n1} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 1 \end{pmatrix} & \begin{matrix} i \\ j \end{matrix} \end{matrix}$$

7.1.3 群组决策

为使决策科学化,民主化,一个复杂系统通常总是有多个决策者(即专家)或决策部门参与决策的。由于决策者的地位、立场、知识水平以及个人偏好的差异,对同一个问题会有不同的判断。这些个体判断是否合理,如果把它们综合成一个较合理的结果,就是所

谓的群组决策问题。关于群组决策应做好以下工作：

1. 重视并做好专家咨询工作

层次分析法是一种定量与定性相结合的决策方法。许多定量分析是建立在专家判断基础上的，因而做好专家咨询工作至关重要，专家咨询工作必须注意以下四个方面。

(1) 合理选择咨询对象，要充分了解专家专长及其熟悉的领域，可针对不同准则请有关专家填写判断矩阵。

(2) 创造适合于咨询工作的良好环境。咨询过程中要向专家简明而准确介绍 AHP，提供可靠的资料和信息，并创造既有集体讨论又有个人独立思考的环境。

(3) 掌握正确的咨询方法，首先要通过咨询确定递阶层次结构，为节省时间避免误解，要设计好咨询表格。

(4) 及时分析专家咨询信息，必要时要进行反馈及多轮次咨询。

2. 群组决策综合方法

这里我们只介绍特征根法。

方法 1 加权几何平均综合排序向量法

对 S 个专家的判断矩阵 $A_k = (a_{ij_k})$ ，分别求出它们的排序向量 $W_k = (W_{1k}, W_{2k}, \dots, W_{nk})^T, k = 1, 2, \dots, S$ ，然后求出它们的加权几何平均综合向量 $W = (W_1, W_2, \dots, W_n)^T$ ，其中

$$\begin{cases} W_j = \frac{\bar{W}_j}{\sum_{i=1}^n \bar{W}_i} \\ \bar{W}_j = (W_{j1})^{\lambda_1} (W_{j2})^{\lambda_2} \cdots (W_{js})^{\lambda_s} \quad j = 1, 2, \dots, n \\ \sum_{k=1}^s \lambda_k = 1 \end{cases}$$

λ_k 为第 k 个专家的权重，它是对专家能力水平的一个综合的数量表示。当对专家的能力水平的高低难以获得先验信息或不易作出比较时，可取 $\lambda_i = 1/S \quad i = 1, 2, \dots, S$ ，此时 $\bar{W}_j = (\bar{W}_{j1},$

$W_{j2}, \dots, W_{jm})^{1/S}$

计算 \bar{W}_j 的标准差 σ_j

$$\sigma_j = \sqrt{\frac{1}{S-1} \sum_{k=1}^S (W_{jk} - \bar{W}_j)^2} \quad (7.1.11)$$

以及相应于新的总体判断矩阵 $A = (a_{ij} = \frac{W_i}{W_j})$ 的总体标准差

$$\sigma_{ij} = \sqrt{\frac{1}{S-1} \sum_{k=1}^S (a_{ijk} - a_{ij})^2} \quad (7.1.12)$$

以及个体标准差

$$\sigma^{(k)} = \sqrt{\frac{1}{n-1} \sum_{j=1}^n (W_{jk} - a_j)^2} \quad (7.1.13)$$

当总体标准差满足要求时,这个群组判断可采用,当个体 $\sigma^{(k)} < \epsilon$ 时认为第 k 个专家通过,否则将信息反馈给有关专家,供修改时参考。

方法2 加权算术平均综合向量法

将各专家判断矩阵得到的排序向量的加权算术平均作为综合排序向量 $W = (W_1, W_2, \dots, W_n)^T$, 即

$$\begin{cases} W_j = \lambda_1 W_{j1} + \lambda_2 W_{j2} + \dots + \lambda_s W_{js} & j = 1, 2, \dots, n \\ \sum_{k=1}^s \lambda_k = 1 \\ \text{当 } \lambda_1 = \lambda_2 = \dots = \lambda_s = \frac{1}{s} \text{ 时有 } W_j = \frac{1}{s} (W_{j1} + W_{j2} + \dots + W_{js}) \end{cases}$$

同样根据方法1中的公式计算相应的标准差。

§ 7.2 建模实例

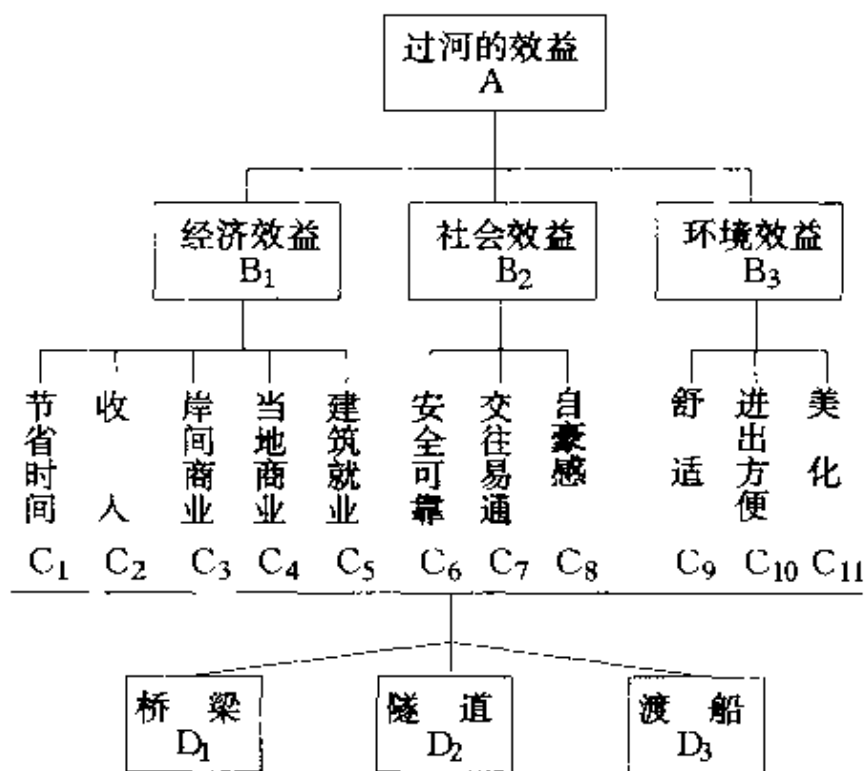
当我们要确定是否对某个项目进行投资,或者要进行方案选择时,就需要对它们的效益与代价(即成本)进行分析。当需要在

若干方案中选择一个最优方案时,我们可以通过层次分析法确定选择一个效益代价比最大的方案。下面通过一个例子来说明成本效益分析法。

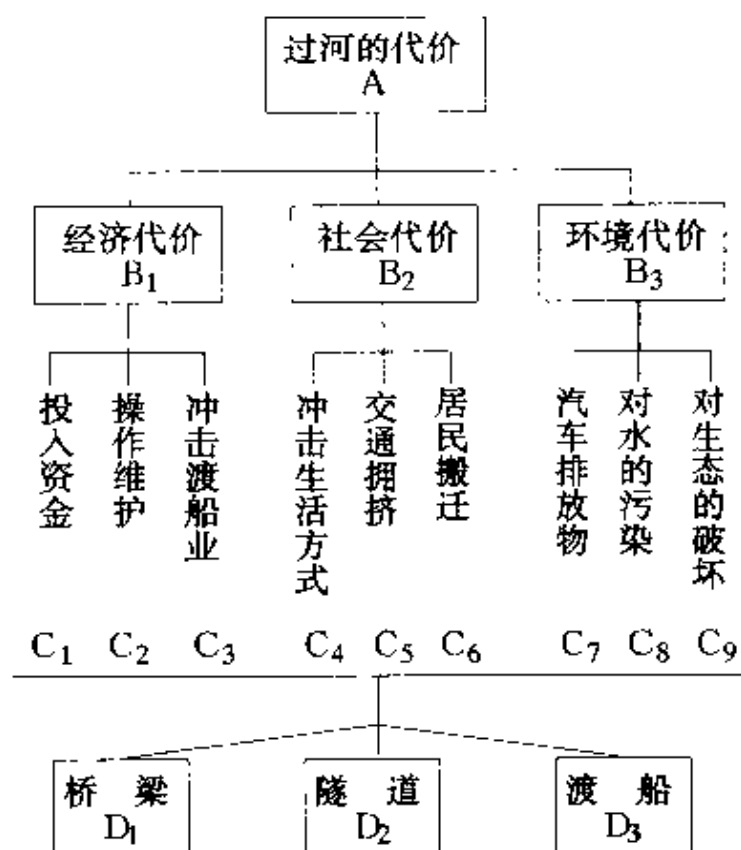
7.2.1 过河的代价与效益分析

设某港务局要改善一条河道的过河运输条件,为此要确定是否要建立桥梁或隧道以代替现存的轮渡。

首先我们在两个层次结构中给出影响过河的效益和代价的因素。这些因素分为3种类型,经济的,社会的和环境的。决策的制定将取决于效益代价比。图7.2.1(1)和(2)分别给出了这两个层次结构。



(1) 过河效益层次结构



(2) 过河的代价层次结构

图 7.2.1

在效益层次结构中,对影响渡河方法选择的经济因素来说,桥梁或隧道显然具有明显的优越性。这一方面有节省时间带来的效益。另外由于交通量的增加,可使航运费增收,从而增加地方政府的财政收入。交通的发达又引起岸间商业的繁荣,从而有助于本地商业(如加油站、饭店、服务业等)的发展,同时建筑施工又创造了就业机会。这部分效益一般都可以进行数量计算,它们的判断矩阵可以由货币效益直接比较获得。但社会效益和环境效益则难以用货币表示,此时就用两两比较的方法进行。从整体看,桥梁和隧道比轮渡更安全,更有助于旅行与交往,也可激发市民对自己城市建设的自豪感。从环境效益看,桥梁和隧道可以给人们更大的舒适

性、方便性,但渡船更具有美感。由此得到关于效益的各个判断矩阵如表 7.1.3 ~ 7.1.17。

表 7.1.3

A	B ₁	B ₂	B ₃	W ⁽²⁾
B ₁	1	3	6	0.61
B ₂	$\frac{1}{3}$	1	2	0.22
B ₃	$\frac{1}{6}$	$\frac{1}{2}$	1	0.11

$$C.I. = 0$$

表 7.1.4

B ₁	C ₁	C ₂	C ₃	C ₄	C ₅	W ₁ ⁽³⁾
C ₁	1	$\frac{1}{3}$	$\frac{1}{7}$	$\frac{1}{5}$	$\frac{1}{6}$	0.04
C ₂		1	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{2}$	0.09
C ₃			1	7	5	0.54
C ₄				1	$\frac{1}{5}$	0.11
C ₅					1	0.23

$$C.I. = 0.14$$

表 7.1.5

B ₂	C ₆	C ₇	C ₈	W ₂ ⁽³⁾
C ₆	1	6	9	0.76
C ₇		1	4	0.18
C ₈			1	0.06

$$C.I. = 0.05$$

表 7.1.6

B ₃	C ₉	C ₁₀	C ₁₁	W ₃ ⁽³⁾
C ₉	1	$\frac{1}{4}$	6	0.25
C ₁₀		1	8	0.69
C ₁₁			1	0.06

$$C.I. = 0.07$$

表 7.1.7

C ₁	D ₁	D ₂	D ₃	W ₁ ⁽⁴⁾
D ₁	1	2	7	0.58
D ₂		1	6	0.35
D ₃			1	0.07

$$C.I. = 0.02$$

表 7.1.8

C ₂	D ₁	D ₂	D ₃	W ₂ ⁽⁴⁾
D ₁	1	$\frac{1}{2}$	8	0.36
D ₂		1	9	0.59
D ₃			1	0.05

$$C.I. = 0.02$$

表 7.1.9

C_3	D_1	D_2	D_3	$W_3^{(4)}$
D_1	1	4	8	0.69
D_2		1	6	0.25
D_3			1	0.06

$$C.I = 0.07$$

表 7.1.10

C_4	D_1	D_2	D_3	$W_4^{(4)}$
D_1	1	1	6	0.46
D_2		1	6	0.46
D_3			1	0.08

$$C.I = 0$$

表 7.1.11

C_5	D_1	D_2	D_3	$W_5^{(4)}$
D_1	1	$\frac{1}{4}$	9	0.41
D_2		1	9	0.54
D_3			1	0.05

$$C.I = 0.11$$

表 7.1.12

C_6	D_1	D_2	D_3	$W_6^{(4)}$
D_1	1	4	7	0.59
D_2		1	6	0.35
D_3			1	0.06

$$C.I = 0.09$$

表 7.1.13

C_7	D_1	D_2	D_3	$W_7^{(4)}$
D_1	1	1	5	0.46
D_2		1	5	0.46
D_3			1	0.09

$$C.I = 0$$

表 7.1.14

C_8	D_1	D_2	D_3	$W_8^{(4)}$
D_1	1	5	3	0.64
D_2		1	$\frac{1}{3}$	0.11
D_3			1	0.26

$$C.I = 0.02$$

表 7.1.15

C_9	D_1	D_2	D_3	$W_9^{(4)}$
D_1	1	5	8	0.73
D_2		1	5	0.21
D_3			1	0.06

$$C.I. = 0.07$$

表 7.1.16

C_{10}	D_1	D_2	D_3	$W_{10}^{(4)}$
D_1	1	3	7	0.64
D_2		1	6	0.29
D_3			1	0.07

$$C.I. = 0.05$$

表 7.1.17

C_{11}	D_1	D_2	D_3	$W_{11}^{(4)}$
D_1	1	6	$\frac{1}{5}$	0.27
D_2		1	$\frac{1}{3}$	0.10
D_3			1	0.63

$$C.I. = 0.31$$

这样我们得方案关于效益的合成排序为

$$W_{\text{益}}^{(4)} = (0.57, 0.36, 0.07)^T$$

效益层次模型的总一致性比例 $C.R^{(4)} < 0.1$ (最后一个矩阵一致性较差, 但因 C_{11} 的排序权重很低, 故不影响最后结果)。从效益看造桥梁方案最佳。

与效益分析相同, 在代价分析的经济代价中, 包括资本耗费, 运行与维护耗费以及由于取消渡轮带来的经济后果。社会代价仍表示社会整体付出的代价, 其中人民生活方式的改变被认为十分重要; 其次, 是过河方式可能带来的交通拥挤; 最后是不同过河方式导致居民迁移对社会的影响。环境代价与环境效益相反, 它表示各种过河方案导致的对环境所造成的损害。代价分析的判断矩阵如表 7.1.18 ~ 7.1.30。

表 7.1.18

A	B ₁	B ₂	B ₃	W ⁽²⁾
B ₁	1	5	7	0.74
B ₂		1	2	0.16
B ₃			1	0.09

$$C.I = 0.01$$

表 7.1.19

B ₁	C ₁	C ₂	C ₃	W ₁ ⁽³⁾
C ₁	1	7	9	0.77
C ₂		1	5	0.17
C ₃			1	0.06

$$C.I = 0.01$$

表 7.1.20

B ₂	C ₄	C ₅	C ₆	W ₂ ⁽³⁾
C ₄	1	$\frac{1}{3}$	$\frac{1}{5}$	0.11
C ₅		1	$\frac{1}{5}$	0.26
C ₆			1	0.64

$$C.I = 0.02$$

表 7.1.21

B ₃	C ₇	C ₈	C ₉	W ₃ ⁽³⁾
C ₇	1	3	4	0.62
C ₈		1	$\frac{1}{3}$	0.13
C ₉			1	0.25

$$C.I = 0.11$$

表 7.1.22

C ₁	D ₁	D ₂	D ₃	W ₁ ⁽⁴⁾
D ₁	1	7	9	0.77
D ₂		1	5	0.17
D ₃			1	0.06

$$C.I = 0.05$$

表 7.1.23

C ₂	D ₁	D ₂	D ₃	W ₂ ⁽⁴⁾
D ₁	1	$\frac{1}{3}$	8	0.30
D ₂		1	9	0.65
D ₃			1	0.05

$$C.I = 0.05$$

表 7.1.24

C_3	D_1	D_2	D_3	$W_3^{(4)}$
D_1	1	1	9	0.47
D_2		1	9	0.47
D_3			1	0.05

$$C.I. = 0$$

表 7.1.25

C_4	D_1	D_2	D_3	$W_4^{(4)}$
D_1	1	4	9	0.69
D_2		1	8	0.26
D_3			1	0.05

$$C.I. = 0.09$$

表 7.1.26

C_5	D_1	D_2	D_3	$W_5^{(4)}$
D_1	1	1	9	0.47
D_2		1	9	0.47
D_3			1	0.05

$$C.I. = 0$$

表 7.1.27

C_6	D_1	D_2	D_3	$W_6^{(4)}$
D_1	1	1	9	0.47
D_2		1	9	0.47
D_3			1	0.05

$$C.I. = 0$$

表 7.1.28

C_7	D_1	D_2	D_3	$W_7^{(4)}$
D_1	1	3	8	0.65
D_2		1	6	0.29
D_3			1	0.06

$$C.I. = 0.04$$

表 7.1.29

C_8	D_1	D_2	D_3	$W_8^{(4)}$
D_1	1	3	7	0.65
D_2		1	5	0.28
D_3			1	0.07

$$C.I. = 0.03$$

表 7.1.30

C_9	D_1	D_2	D_3	$W_9^{(4)}$
D_1	1	$\frac{1}{6}$	7	0.21
D_2		1	8	0.73
D_3			1	0.05

$$C.I. = 0.16$$

得到方案关于代价的合成排序为 $W_{\text{代}}^{(4)} = (0.36, 0.58, 0.05)^T$, 总一致性比例 $C.R^{(4)} < 0.1$ 。

各方案的效益/代价如下:

$$\text{桥梁} \quad " \quad = 1.58$$

$$\text{隧道} \quad " \quad = 0.62$$

$$\text{轮渡} \quad " \quad = 1.28$$

方案选择的准则是应使效益代价比最大, 因此应选择建造桥梁方案。

7.2.2 企业留成的分配问题

某厂在扩大企业自主权后, 有一笔留成利润, 要由厂领导和职代会来决定如何使用, 可供选择的方案有: P_1 —发奖金, P_2 —扩建集体福利事业, P_3 —办职工工业余技校, P_4 —建图书馆、俱乐部, P_5 —引进新设备, 这些方案都各有其合理的因素, 因此如何对这些方案进行综合评价, 并由此进行方案排序及选优是厂领导和职代会面临的实际问题。

上述问题属于方案排序与选优问题, 且各待选方案的具体内容已经确定, 所以可采用 AHP 方法来解决。

(1) 建立方案评价的层次结构模型, 该模型的最高层次(第一层次)为总目标 A —合理使用企业利润, 以促进企业发展; 中间层(第二层次)为方案评价的准则层, 它包括: B_1 —进一步调动职工

的劳动积极性, B_2 —提高企业技术水平; B_3 —改善职工物质和文化生活三个准则。最低层为方案层, 它包含前述 P_1 — P_5 五种方案。其层次结构模型如图 7.2.2。

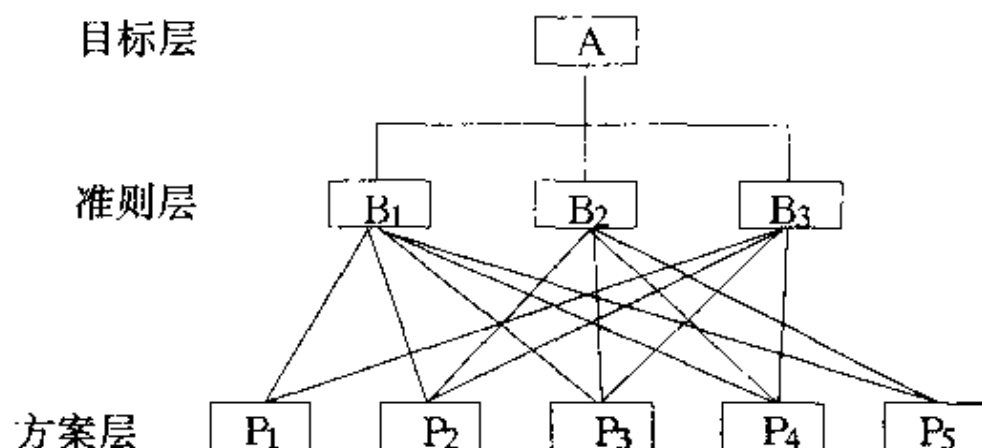


图 7.2.2

(2) 构造比较判断矩阵, 计算层次单排序及其一致性检验。注意到以 A 为比较准则, B 层次各因素的两两比较判断矩阵为 $A-B$, 类似的以每一个 B_i 为准则, P 层次各因素两两比较可构成三个比较判断矩阵 $B_i-P, i=1,2,3$ 。故本模型构造四个比较判断矩阵如表 7.1.31 ~ 7.1.34。

表 7.1.31

A	B_1	B_2	B_3	$W^{(2)}$
B_1	1	$\frac{1}{5}$	$\frac{1}{3}$	0.1047
B_2		1	3	0.6370
B_3			1	0.2582

$$C.I = 0.0192$$

$$R.I = 0.58$$

$$C.R = 0.332$$

表 7.1.32

B_1	P_1	P_2	P_3	P_4	P_5	$W_1^{(3)}$
P_1	1	2	3	4	7	0.439
P_2		1	3	2	5	0.2636
P_3			1	$\frac{1}{2}$	1	0.0894
P_4				1	3	0.1465
P_5					1	0.0615

$$C.I = 0.0318$$

$$R.I = 1.12$$

$$C.R = 0.0284$$

表 7.1.33

B_2	P_2	P_3	P_4	P_5	$W_2^{(3)}$
P_2	1	$\frac{1}{7}$	$\frac{1}{3}$	$\frac{1}{5}$	0.0553
P_3		1	5	3	0.565
P_4			1	$\frac{1}{3}$	0.1175
P_5				1	0.2622

$$C.I = 0.0389$$

$$R.I = 0.9$$

$$C.R = 0.0433$$

表 7.1.34

B_3	P_1	P_2	P_3	P_4	$W_3^{(3)}$
P_1	1	1	3	3	0.375
P_2		1	3	3	0.375
P_3			1	1	0.125
P_4				1	0.125

$$C.I = 0$$

$$R.I = 0$$

$$C.R = 0$$

由上表所知四个层次单排序的 $C.R$ 值均小于 0.1, 符合满意一致性要求。

(3) 层次总排序: 由于第二层(B 层) 相对于总目标 A 的排序权重向量 $W^{(2)} = (0.1047 \ 0.637 \ 0.2582)^T$ 已得到, 而第三层(P 层) 以第二层第 i 个因素 B_i 为比较准则时的重要性标度 $W_i^{(3)}$ 分别为

$$W_1^{(3)} = (0.439, 0.2636, 0.0894, 0.1465, 0.0615)^T$$

$$W_2^{(3)} = (0, 0.0553, 0.565, 0.1175, 0.2622)^T$$

$$W_3^{(3)} = (0.375, 0.375, 0.125, 0.125, 0)^T$$

则第三层(P 层) 相对于总目标 A 的排序权重向量 $W^{(3)} = B^{(3)} W^{(2)} = (W_1^{(3)} W_2^{(3)} W_3^{(3)}) W^{(2)} = (0.1428, 0.1597, 0.4015, 0.1225, 0.1735)^T$ 注意到前述评价模型总共只有 3 层, 因此上述计算的 $W^{(3)}$ 即为各方案相对于总目标的排序权重向量。

(4) 层次总排序的一致性检验。由(2) 的计算已得到第二层第 j 个因素 B_j 为比较准则时, 第三层(P 层) 次单排序的一致性指标

与平均随机一致性指标 $C.I_j^{(2)}$ 与 $R.I_j^{(2)}$ 并有

$$C.I^{(2)} = (C.I_1^{(2)} \ C.I_2^{(2)} \ C.I_3^{(2)}) = (0.0318, 0.0837, 0)$$

$$R.I^{(2)} = (R.I_1^{(2)} \ R.I_2^{(2)} \ R.I_3^{(2)}) = (1.12, 0.9, 0.9)$$

则第三层次总排序的一致性检验有

$$C.I^{(3)} = C.I^{(2)} W^{(2)} = (0.0318, 0.0837, 0)(0.1047, 0.637, 0.2582)^T = 0.0281$$

$$R.I^{(3)} = R.I^{(2)} W^{(2)} = (1.12, 0.9, 0.9)(0.1047, 0.637, 0.2582)^T = 0.9229$$

$$C.R^{(3)} = C.R^{(2)} + \frac{C.I^{(3)}}{R.I^{(3)}} = 0.0332 + \frac{0.0281}{0.9229} = 0.0332 + 0.0304 = 0.0636$$

注意到层次总排序的随机一致性指标 $C.R^{(3)} < 0.1$, 故可认为上述计算得到的层次总排序 $W^{(3)}$ 是可信的。因此对于某工厂合理使用企业留成利润这一总目标所考虑的五种方案的相对优先排序为:

- 1° P_3 (开办职工业务技校) 权重为 0.4015
- 2° P_5 (引进新技术设备) 权重为 0.1735
- 3° P_2 (扩建集体福利事业) 权重为 0.1597
- 4° P_1 (发奖金) 权重为 0.1428
- 5° P_4 (建图书馆、俱乐部) 权重为 0.1225

厂领导和职代会可根据上述分析结果, 决定各种考虑方案的实施先后次序, 或决定分配使用企业留成利润的比例。

在第五章中我们介绍了规划的方法建模。规模的方法是一类最优化方法, 近年来在处理各类问题中得到了广泛应用。一个最优化问题总是一个带有目标的活动, 问题的目标是根据人们的感觉和偏好确定的。一个典型的优化模型要包括目标函数和一些等式或不等式约束。优化的目的是确定一个变量的最优解集。由于实际问题的复杂性, 大量问题只有经过大大简化, 化为这种典型形式后

才能求解。在有些情形下,一方面由于简化后的优化模型的有效性
与可行性受到了严重影响,另一方面由于实际问题本身要涉及到
一些重要的但难确定的、模糊的概念。例如,讨论军事问题,商业问
题时经常遇到的“军事价值”、“生活损害”以及“商业信益”等等,
这就使标准的最优化方法难以应用。而在这方面,层次分析法提供
了一种人们根据自己的知识、感觉和偏好进行判断从而得到决策
问题最优解的方法。下面仅用两个简单实例加以说明。

7.2.3 农民卖水果问题

有一个农民要出售他种植的一种水果,若现在收摘可得到水
果 120 公斤,每公斤价格为 1 元。由于果实仍在继续生长,因此若
晚一些时候收摘,每过一周可增加产量 20 公斤,但价格要下降 0.1
元,试问他应怎样安排收摘使收益最大?

这是一个简单的优化问题:设 x 周后收摘

$$\max(120 + 20x)(1 - 0.1x)$$

解之得最优解 $x = 2$,即在两周后收摘卖出获利最大。

假定这位农民对上述算法并不了解,并且不认为产量和价格
会以上述方式有规律地变化,此时问题的复杂性在于,即使他会进
行一些计算,也难以获得一个可以求得精确解的简单的数学公式。
本例的目的是要帮助他处理这一不确定问题,选择一个出售水果
的最佳时机,并说明复杂问题的处理方法。

首先建立一个层次结构,其目标是此农民的最大收益,此收益
即与收获时的产量有关。又与当时的价格有关,而归根到底是取决
于农民出售水果还需等待的周数。假设此时间最大不超过 5 周。那
么在第四层上是(0,1,2,3,4,5)6 个方案。在目标层下第二层是产
量,有高、中、低三种情形。第三层是每公斤的价格,也有低中高三
种,图 7.2.3 是这个问题的层次结构图。

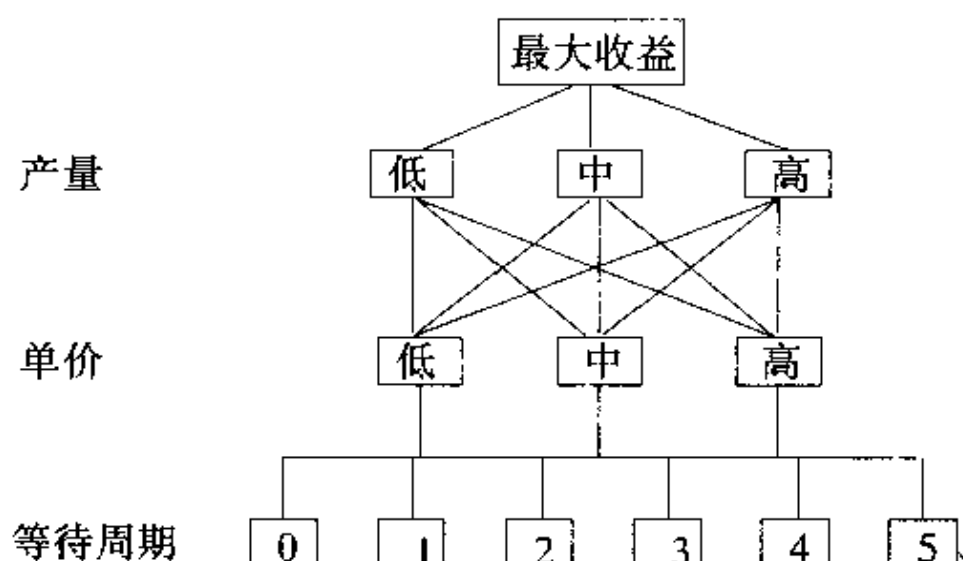


图 7.2.3 最大收益层次图

显然,为了获得最大收益,该果实相对地比较大,但是太大的果实,由于运输,贮存中容易损坏以及收获早晚等原因,往往实际上不是最值钱的。这样,通过回答比较所生产的果实大小对收益的贡献哪一个强的问题,可以得到第一个判断矩阵(表 7.1.35)。

表 7.1.35

收益	(产量)	低	中	高	$W^{(2)}$
低		1	$\frac{1}{3}$	3	0.258
中		3	1	5	0.637
高		$\frac{1}{3}$	$\frac{1}{5}$	1	0.105

$$C.R = 0.033$$

为了简便起见,不再说明给出判断的理由,也可能有人会不同意其中一些判断,但事实上对层次分析法来说,除非某一步的判断有重大出入,其结果是相当稳定的。下面对在各种产量(果实大小)下比较哪一种价格更有可能达到最优得到的判断矩阵如表 7.1.36、7.1.37、7.1.38 所示。

表 7.1.36

低(产量)	(价格)	低	中	高	$W_1^{(3)}$
低		1	$\frac{1}{3}$	$\frac{1}{5}$	0.109
中		3	1	$\frac{1}{2}$	0.309
高		5	2	1	0.582

$$C.R = 0.003$$

表 7.1.37

中(产量)	(价格)	低	中	高	$W_2^{(3)}$
低		1	$\frac{1}{5}$	$\frac{1}{2}$	0.128
中		5	1	2	0.595
高		2	$\frac{1}{2}$	1	0.276

$$C.R = 0.005$$

表 7.1.38

高(产量)	(价格)	低	中	高	$W_3^{(3)}$
低		1	4	5	0.691
中		$\frac{1}{4}$	1	1	0.160
高		$\frac{1}{5}$	1	1	0.149

$$C.R = 0.005$$

对第四层上时间的比较需要回答的问题是,对于指定的价格,哪一个出售时间更有可能达到最优?这三个判断矩阵如表 7.1.39、7.1.40、7.1.41。

表 7.1.39

低(价格)	0	1	2	3	4	5	$W_1^{(4)}$
0	1	1					0.064
1	1						0.084
2	2	1	1				0.12
3	3	2	1	1			0.173
4	4	3	2	1	1		0.243
5	5	4	3	2	1	1	0.315

$$C.R = 0.015$$

表 7.1.40

中(价格)	0	1	2	3	4	5	$W_2^{(4)}$
0	1						0.053
1	3	1					0.122
2	5	3	1				0.321
3	5	3	1	1			0.321
4	3	1	$\frac{1}{3}$	$\frac{1}{3}$	1		0.122
5	1	$\frac{1}{2}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{2}$	1	0.060

$$C.R = 0.008$$

表 7.1.41

高(价格)	0	1	2	3	4	5	$W_3^{(4)}$
0	1	3	4	5	6	7	0.426
1		1	3	4	5	6	0.254
2			1	3	4	5	0.150
3				1	3	4	0.088
4					1	3	0.052
5						1	0.031

$$C.R = 0.076$$

最后得到各周出售水果的权重可解释为在各周出售水果时取得最大收益的概率(表 7.1.42)

表 7.1.42

周 数	0	1	2	3	4	5
出售概率	0.183	0.16	0.226	0.214	0.12	0.097

该农民有两种方案可供选择。第一种是选择概率最高的第二周出售,这和前面最优解是一致的。第二种方案是计算期望值

$$0 \times 0.183 + 1 \times 0.16 + 2 \times 0.226 + 3 \times 0.214 + 4 \times 0.12 + 5 \times 0.097 = 2.22$$

即在过 2.22 周后出售水果,这种方法可能更合理。

7.2.4 最佳食品搭配问题

假设某人有 3 种食品可供选择:肉、面包、蔬菜。它们所含的营养成分及单价如表 7.1.43 所示。

表 7.1.43

食品	维生素 $A(1. V/g)$	维生素 $B_2(mg/g)$	热 量 $Kcal/g$	单 价 (元/g)	搭配量
肉	0.3527	0.0021	2.86	0.0055	x_1
面包	0.0000	0.0006	2.76	0.0012	x_2
蔬菜	25	0.002	0.25	0.0014	x_3

该人体重 55 公斤,每天对各种营养的最小需求为:维生素 A 7500 个国际单位,维生素 B_2 1.6338 毫克,热量为 2050 千卡。问应如何搭配食品使在保证营养下支出最小?

这是线性规划问题,其模型为:

$$\begin{aligned}
 \min S &= 0.0055x_1 + 0.0012x_2 + 0.0014x_3 \\
 s, t \quad &\begin{cases} 0.3527x_1 + 25x_3 \geq 7500 \\ 0.0021x_1 + 0.0006x_2 + 0.002x_3 \geq 1.6338 \\ 2.86x_1 + 2.76x_2 + 0.25x_3 \geq 2050 \\ x_1, x_2, x_3 \geq 0 \end{cases} \quad (1)
 \end{aligned}$$

解之得到最优解为 $x_1 = 0, x_2 = 687.44g, x_3 = 610.67g$, 每天支出为 $S = 1.67$ 元。

这个最优方案显然是行不通的,其问题首先在于没有考虑到本人对食品的偏好。根据偏好应再加上约束: $x_1 \geq 140g, x_2 \leq 560g, x_3$ 无限制。其次,各种营养成分在这里被看作是同样重要的,而支出成了支配一切的因素。但实际上即要考虑支出,也要考虑营养价值。不过它们的地位可能不一样。此时要写出一个目标函数是很困难的。可以利用 AHP 来处理。首先构造层次结构如图 7.2.4 所示。

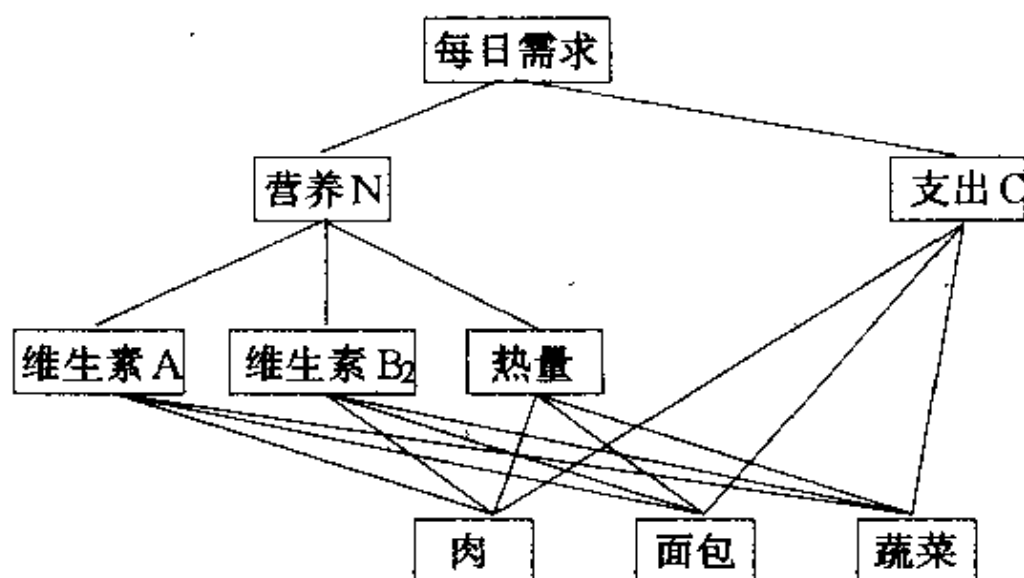


图 7.2.4 选择食品的层次结构

对于一个具有中等收入的人,满足营养要求(而不是起码要求)比支出更重要,于是有表 7.1.44、7.1.45。

表 7.1.44

每日需求	营养	支出	权重
营 养	1	3	0.75
支 出	1/3	1	0.25

表 7.1.45

营 养	维生素 A	维生素 B ₂	热量	权重
维生素 A	1	1	2	0.4
维生素 B ₂	1	1	2	0.4
热 量	1/2	1/2	1	0.2

将 7.1.43 中 1 克食品所含的维生素 A、维生素 B₂ 以及热量归一化后得到表 7.1.46。

表 7.1.46

	维生素 A	维生素 B ₂	热量(千卡)
肉	0.0139	0.4468	0.4872
面包	0.0000	0.1277	0.4702
蔬菜	0.9861	0.4255	0.0426

考虑到应使支出最小,故将各食品单价倒数归一化作为对支出的权重,得表 7.1.47。合成后得到三种食品的权重如表 7.1.48。将上述食品比例代入方程(1),得到最优食品搭配如表 7.1.49。

表 7.1.47

支出	权重
肉	0.105
面包	0.476
蔬菜	0.419

表 7.1.48

食品	肉	面包	蔬菜
合成权重	0.24	0.23	0.53

表 7.1.49

食品	肉	面包	蔬菜	总支出
量	338 克	324.29 克	747.4 克	3.15 元

此时各营养成分含量为维生素 A = 15058.8 国际单位,维生素 B₂ = 2.40(毫克),热量 = 2050(千卡),均满足方程(1)及(2)的要求。

如果认为总支出太大,那么可以适当地降低第二层中营养的权重。如果将营养及支出的权重均取为 0.5,此时的食品搭配如表 7.1.50。

表 7.1.50

食品	肉	面包	蔬菜	总支出
量	258.39 克	414.97 克	662.76 克	2.85 元

此时的营养成分为维生素 A = 16660 国际单位,维生素 B₂ = 2.18(毫克),热量 = 2050 千卡,均满足要求。

从以上两个例子可以看出,层次分析法在处理最优化问题时的优越性在于可以反映决策者的偏好,并在很多方面都能与传统的严格的数学方法相吻合。应用 AHP 的另一个现实意义在于既可考虑采用大量数据,又可避免复杂的运算。

习 题

1. 城市能源供应系统改造方案的比较。

某市能源供应方面存在着设施落后、供应分散、利用率低、污染严重等问题,经有关部门分析,可供选择的改造方案有:

P_1 :制作高效煤制品,取代目前居民用的普通蜂窝煤;

P_2 :将分散供热方式改造为区域供热;

P_3 :建造热电联供设施,利用电厂余热;

P_4 :实现城市煤气化;

P_5 :郊县推广沼气池,作为农村主要生活能源。

这些措施在节约能源、降低污染等方面的效益,及在投资、原有设施闲置等方面的代价各不相同。为了综合考虑各种因素,确定这些方案的优劣顺序,以效益和代价为两个目标,建立两个层次结构模型。(在考虑准则时均应考虑社会效益和社会因素)。

2. 科技成果的综合评价。

科技成果涉及的领域很广,种类很多,这里指的是直接应用于国民经济的某个生活部门后,可迅速转化为生产力,带来可定量计算的经济效益的那一类成果。评价准则先分为效益 C_1 ,水平 C_2 、规模 C_3 等三类。 C_1 中分直接经济效益、间接经济效益、社会效益等指标; C_2 中分学识水平、学术创新、技术水平、技术创新等指标。待评价的科研成果若干,试建立评价科研成果的层次结构图。

3. 资源开发的综合判断。

某发展中国家有 7 种可供开发的矿产资源：铁、铜、磷酸盐、铀、铝、金、金钢石。它们开发后对国家经济发展的贡献可以通过两两比较得到，另一方面，制定开发计划要考虑 6 个因素：潜在经济价值，开采费用，风险，需求，战略重要性，交通条件。综合判断要求即要给出各种资源在国家发展中的优先程度，又要确定各种资源开发中各个因素的相对重要性。试建立该问题的层次结构模型。

4. 工作选择。

一个刚获得学位的大学毕业生面临选择工作岗位，他将要考虑的准则有：能够发挥自己的才干为国家作贡献；丰厚的收入；适合个人的兴趣与发展；良好的声誉；人际关系；地理位置等，可供选择的工作有若干，试建立该问题的层次分析模型，并给出有关判断矩阵。

第八章 图论方法建模

图论是近年来发展十分迅速,应用比较广泛的一个新兴的数学分支。在许多领域,诸如物理、化学、生物学、地理学、计算机科学、信息论、系统论、控制论、社会科学、军事科学以及经济管理等方面,图论都有着广泛的应用。因此,图论受到了全世界数学界和工程技术界乃至经营决策管理者越来越广泛的重视。同时也成了数学建模中一种十分重要的方法。本章由于篇幅所限,只就图论中的几个方面作简要介绍。

§ 8.1 图的基本概念

定义 8.1.1 集合 $V = \{v_1, v_2, \dots, v_n\}$ 叫做一个“顶点集合”,其中的元素 $v_i (i = 1, 2, \dots, n)$ 叫做顶点。集合 $E = \{e_1, e_2, \dots, e_m\}$ 叫做一个边集合,其中

元素 $e_j (j = 1, 2, \dots, m)$ 叫做边。如果对任意的边 $e_j \in E$ 都有且仅有 V 中的一对顶点 v_k, v_h 与之对应,这时记 $e_j = (v_k, v_h)$,那末我们就说顶点集 V 与边集 E 一起构成了一个图 G ,记为 $G = (V, E)$ 。

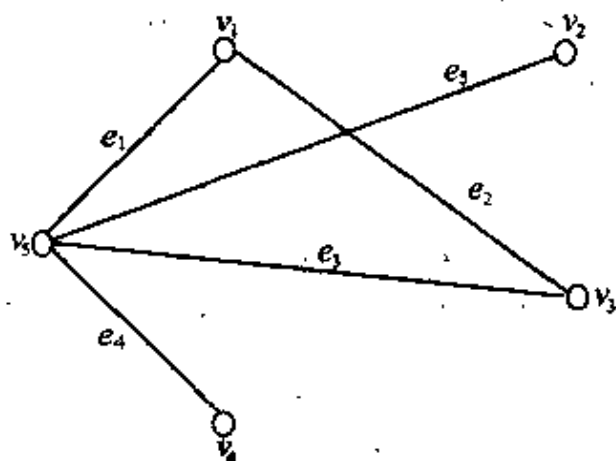


图 8.1.1

如图 8.1.1 中边集为 $E = \{e_1, e_2, e_3, e_4, e_5\}$, 顶点集为

$$V = \{v_1, v_2, v_3, v_4, v_5\}.$$

图论中的图有以下特点:

① 图中的顶点与几何学中的点不同,它可以随着所讨论的实际问题的不同,代表城镇、人或团体等不同的各种适当事物。

② 图中的边与几何学中的线有本质的区别,它也可以随着所讨论的实际问题的不同而代表联系城镇的公路、代表人与人之间认识或不认识的关系,代表团体与团体之间有无联系等。总之,它可以代表事物与事物之间有没有某种特定性质的关系。

③ 一个图中各顶点之间相互位置的摆法对我们来说并不重要,我们只关心一个图有哪些顶点。其次,图中所划出的边的长短、曲直我们也不关心,我们只关心哪些顶点之间有或没有相连。

④ 在一个图中,边与边之间若有除顶点之外的交点,我们不予承认,即这不是顶点。

⑤ 图中不允许出现没有端点或只有一个端点的边。

由于图论中的图有以上这些特点,使得许多实际问题都可以抽象成图,用图论的有关理论和方法加以解决。

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 $p > 0, q \geq 0$, 且均为整数。

一条边 e_j 若是连接 v_k 与 v_h 两个顶点,记为 (v_k, v_h) , 即有 $e_j = (v_k, v_h)$, 称 v_k 和 v_h 为 e_j 的端点,也称 e_j 是点 v_k (或 v_h) 的关联边。

如果 $v_k, v_h \in V(G)$, 并且 $(v_k, v_h) \in E(G)$, 称顶点 v_k 与 v_h 为相邻的。与同一个顶点关联的边 e_1, e_2 称为邻接边。

对于边 $e \in E(G)$, 如果它的两个端点相同,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 8.1.2 中 $p = 5, q = 9, e_2, e_3$ 为邻接边, e_1 为环, e_8, e_9 为多重边。

无环、无多重边的图称为简单图。

如果图 G 中没有任何边,即 $q = 0$, 称这个图为空图。

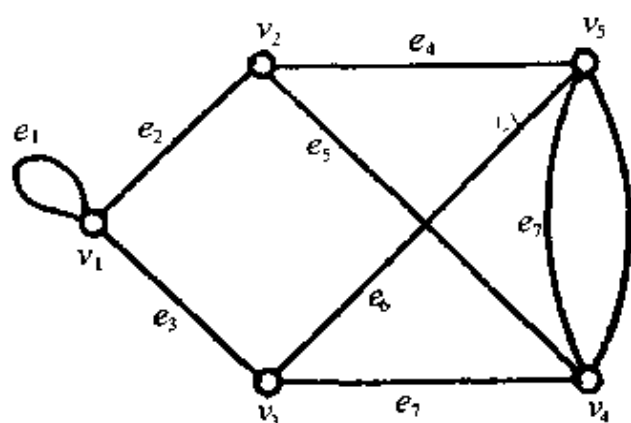


图 8.1.2

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 为完全图。 n 个顶点的完全图记为 K_n 。如果在图的边上加数字, 该图称为赋权图, 该数字称为权数(图 8.1.3(a))。

如果在图 $G = (V, E)$ 中, V 可以剖分成两个集合 X, Y , 满足 $V = X \cup Y, X \cap Y = \phi$, 使得 G 的每条边的一个端点在 X 中, 另一端点在 Y 中, 就称图 G 为二分图, 也称为偶图, 记为 $G = (X, Y, E)$ (图 8.1.3(b))。

使得 G 的每条边的一个端点在 X 中, 另一端点在 Y 中, 就称图 G 为二分图, 也称为偶图, 记为 $G = (X, Y, E)$ (图 8.1.3(b))。

$$x = \{v_1, v_2, v_4, v_5, v_7, v_9\}$$

$$y = \{v_3, v_6, v_8\}$$

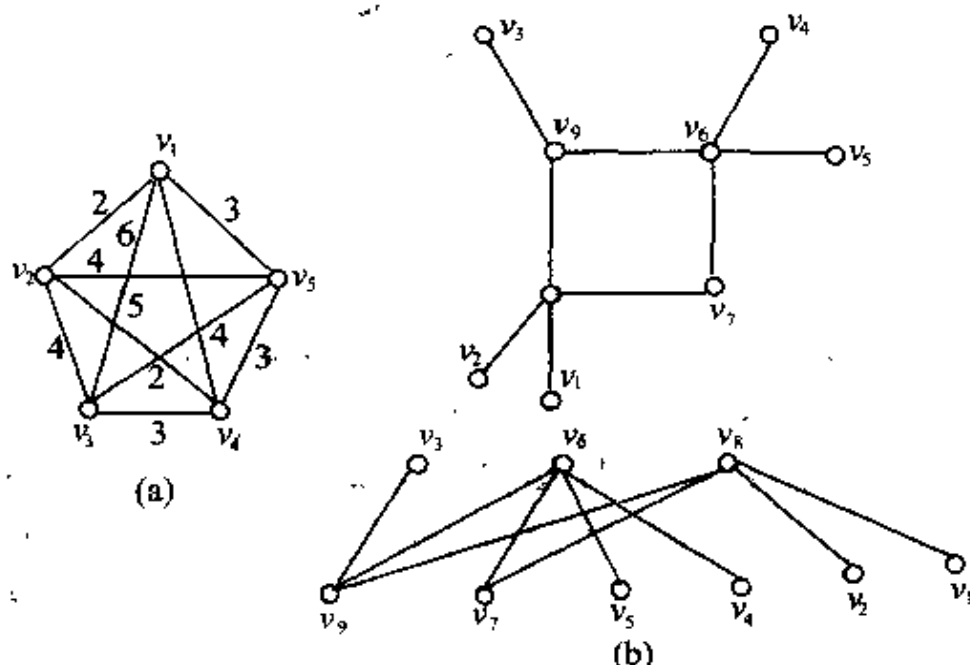


图 8.1.3

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) 记为 $d(v)$ 。 $\delta = \min_{v \in V} d(v)$ 称为 G 的最小次, $\Delta = \max_{v \in V} d(v)$ 称为 G 的最大次。将次为奇数的顶点称为奇顶点, 反之, 称为偶顶点。

定理 8.1.1 $\sum_{v \in V} d(v) = 2q$

证: 因图 G 中每条边均与两个顶点关联, 所以各点的次之和等于边数的 2 倍。

推论 在任何图中, 奇顶点的个数是偶数。

一个图的图形表示不是惟一的。如图 8.1.4 中的三个图形表示的都是同一个图 G 。

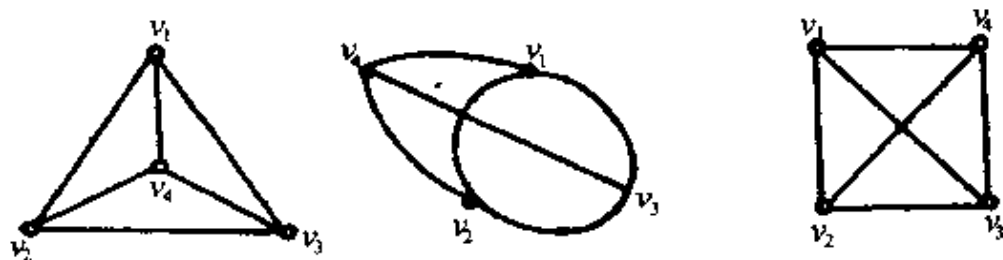


图 8.1.4

一个图除了用图形表示外, 还可以用矩阵来表示。

在图 $G = (V, E)$ 中, $V = \{v_1, v_2, \dots, v_p\}$, $E = \{e_1, e_2, \dots, e_q\}$, 如果用 m_{ij} 表示点 v_i 与边 e_j 相关联的次数 $(0, 1, 2, \dots)$, 则矩阵 $M = (m_{ij})_{p \times q}$, 称为图 G 的关联矩阵。如果用 a_{ij} 来表示连结点 v_i 和 v_j 的边的数目, 则矩阵 $A = (a_{ij})_{p \times p}$ 称为 G 的邻接矩阵。

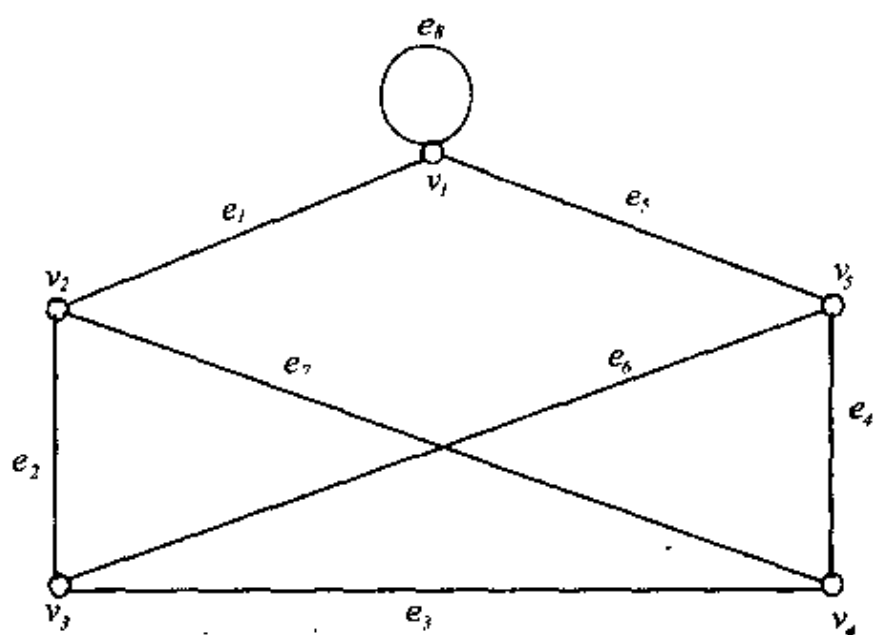


图 8.1.5

例 1 若图 G 为图 8.1.5。图 G 的关联矩阵为

	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 G 的邻接矩阵为

	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

一般来说,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一个图,关联矩阵和邻接矩阵在以后的有关内容中多处使用,要很好掌握。

定义 8.1.2 设 $G_1 = (V_1, E_1), G_2 = (V_2, E_2)$, 如果 $V_1 \subseteq V_2, E_1 \subseteq E_2$, 则称 G_1 是 G_2 的子图。记作 $G_1 \subseteq G_2$

特别,当 $V_1 = V_2, E_1 \subseteq E_2$ 时,称 G_1 为 G_2 的支撑子图(图 8.1.6)。

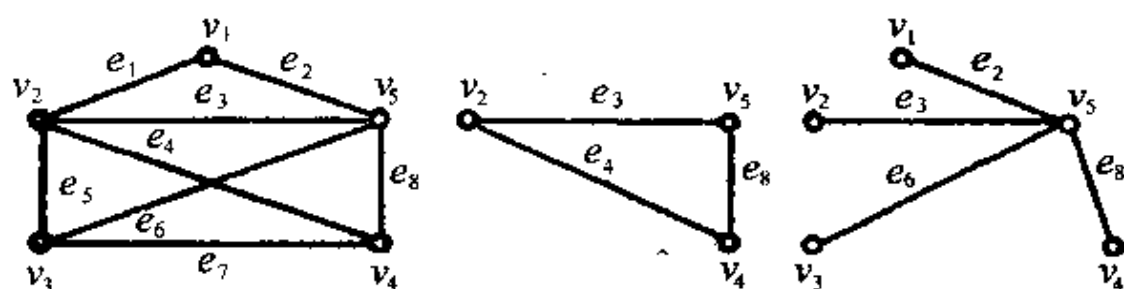


图 G

子图 G'

支撑子图

图 8.1.6

定义 8.1.3 在图 $G = (V, E)$ 中,顶点和边交替的有限非零序列 $W = v_0 e_1 v_1 e_2 v_2 \cdots e_k v_k$, 若对 $1 \leq i \leq k$, 边 e_i 的端点是 v_{i-1} 和 v_i , 则称 W 是从 v_0 到 v_k 的一条链。 v_0 和 v_k 分别称为链 W 的起点和终点, k 称为链 W 的长。如果 $v_0 = v_k$, 则称 W 为闭链, 否则称之为开链。

在简单图中,链也可以由顶点序列来表示。

如果链 W 中所有的 e_i 都不相同, 则称 W 为简单链。如果链 W 中所有的 v_i 都不相同, 则称 W 为初等链(或路)。一条开的初等链称为路。一条闭的初等链称为圈(或回路)。如图 8.1.7 中, 图 G 是

简单图。

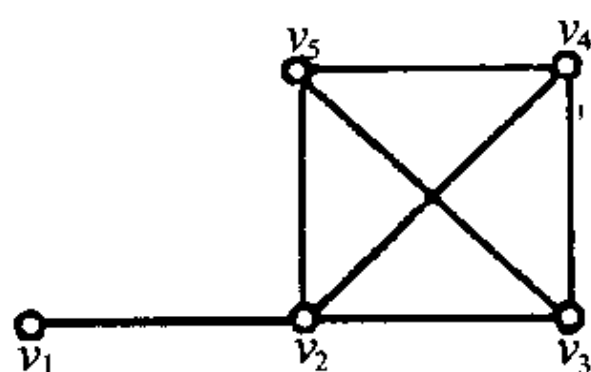


图 8.1.7

$W_1 = v_1 v_2 v_5 v_4 v_2 v_3 v_5 v_2$ 是一条链但不是简单链

$W_2 = v_1 v_2 v_5 v_4 v_2 v_3$ 是一条简单链,但不是初等链

$W_3 = v_1 v_2 v_5 v_3 v_4$ 是一条初等链,也是路。

$C = v_2 v_3 v_4 v_5 v_2$ 是一个圈(回路)。

定义 8.1.4 如果在图 G 的任意两点间,都存在一条路,则称图 G 是连通图,否则图 G 是不连通的。

在图 8.1.8 中, G_1 是连通图, G_2 是不连通的。

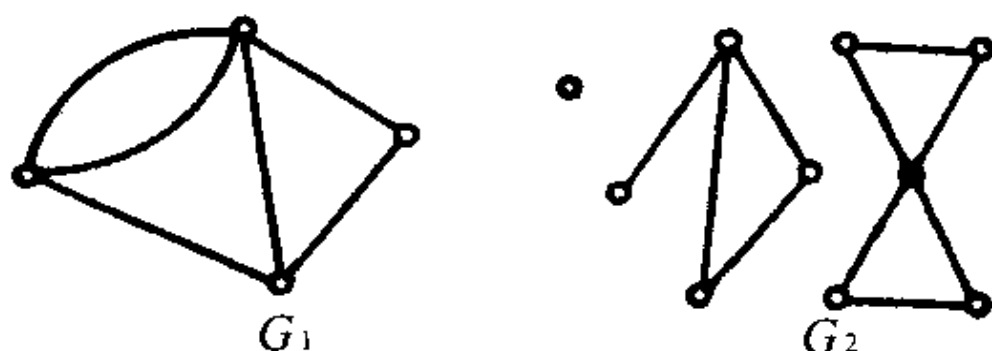


图 8.1.8

图 G 的任意一个极大连通子图称为 G 的一个连通分支。

在图 8.1.8 中, G_2 是不连通图,但它有三个连通分支,这里孤立点本身也算一个分支。而连通图 G_1 只包含惟一的一个连通分支,即它自身。

§ 8.2 E 图和 H 图

8.2.1 E 图

定义 8.2.1 在图 G 中一个简单链, 经过 G 每条边一次且仅一次称为欧拉链。

如果图 G 中有一条闭的欧拉链, 则称图 G 为欧拉图 (Euler) 简称为 E 图。

如图 8.2.1(1) 中, $W_1 = v_1 v_2 v_3 v_4 v_2 v_5 v_4$ 是一条欧拉链, 但不是欧拉图, 在 (2) 中存在闭的欧拉链, 所以是 E 图。

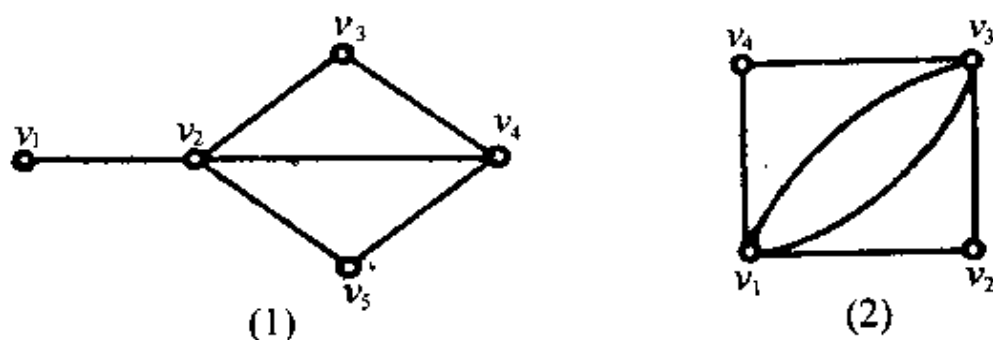


图 8.2.1

定理 8.2.1 非空连通图 G 有欧拉链的充分必要条件是图 G 的奇顶点个数为 0 或 2, 并且, 当且仅当奇顶点个数为 0 时, 图 G 是 E 图。

由上述定理可知, 要检验一个图是否 E 图或图中有无欧拉链是比较容易的。但要实际地来找出这个闭欧拉链或欧拉链, 当图比较复杂时, 一般来说就不太容易了。我们只知道, 在寻找闭欧拉链时, 出发点可以是图中任何一点, 在图中有两个奇顶点而又要找出其欧拉链时, 出发点必须是这两个奇顶点中的一个。在这种认识的基础上, 我们给出寻找闭欧拉链和欧拉链的一种具体算法

——Fleury 算法。步骤如下：

- (1) 选初始点 v_0 , 并令 $u_0 = [v_0]$, $i = 0$
- (2) 假设链 $u = [v_0 e_1 v_1 \cdots e_i v_i]$ 已选出, 那末用如下方法从 $E \setminus \{e_1 e_2 \cdots e_i\}$ 中选边 e_{i+1}
 - (a) e_{i+1} 与 e_i 关联
 - (b) e_{i+1} 关联一个与 v_i 不同的顶点 v_{i+1}
 - (c) 令 $u = u \cup [v_i e_{i+1} v_{i+1}]$, $i = i + 1$, 返回(2)
- (3) 第(2)步不能进行时即停止。

8.2.2 H 图

定义 8.2.2 若图 $G = (V, E)$ 中存在一条包含 G 的所有顶点的路, 则此路称为哈密尔顿路。

如果图 G 中存在包括它的每一顶点的圈, 则这个圈称为 G 的哈密尔顿圈。则 G 称为哈密尔顿图(Hamilton) 简称为 H 图。

容易看出, 完全图 k_n 一定是 H 图(图 8.2.2(1)), 而图 8.2.2 中的(2)、(3)不是 H 图。

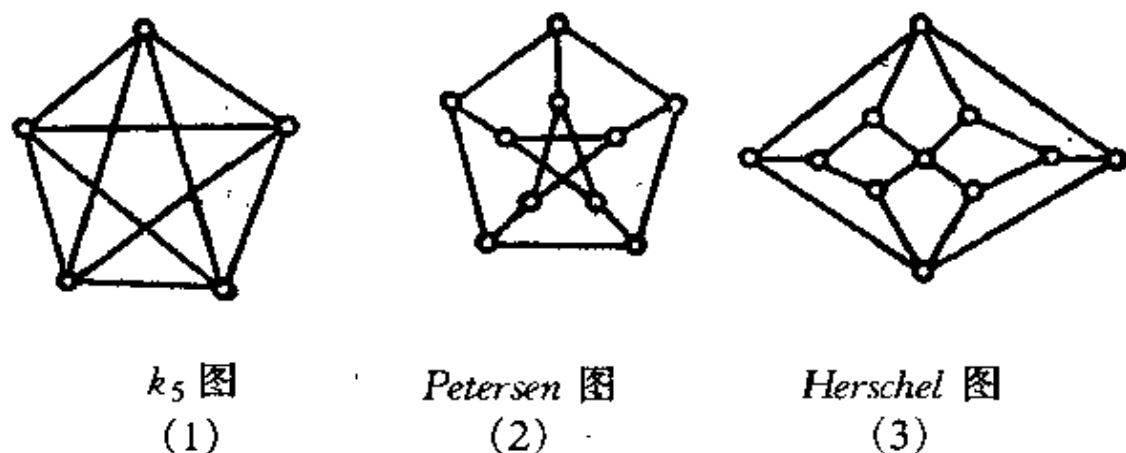


图 8.2.2

一百多年来,世界上许多数学家对于 H 图进行了研究,取得了很大成绩,引出很多问题,导致很多新的概念,得到很多有趣的结果,但其中一个最基本的问题,判定一个图 G 是否为 H 图的充分必要条件是什么?至今没有找到。这是图论里一个著名的未解决的问题,所谓哈密尔顿圈的问题。在人们的长期研究中,已经得到很多充分条件和必要条件,这里仅介绍其中之一。

定理 8.2.1(Ore) 若 G 为简单图, $P \geq 3$, 并且对于每一对不相邻顶点 v, v' 有 $d(v) + d(v') \geq P$, 则图 G 有哈密尔顿圈。

8.2.3 建模实例

1. 七桥问题

七桥问题是图论中一个古老的著名问题。著名的数学家欧拉(Euler)在1736年成功地解决了这个问题,从而成为图论的创始人。

问题:穿过哥尼斯堡的一条河中有两个小岛,连结河中两岛及两岸陆地有七座桥,某人从某地出发散步,经过每座桥恰好一次,再回到出发地点,问是否有这样的行走路线?

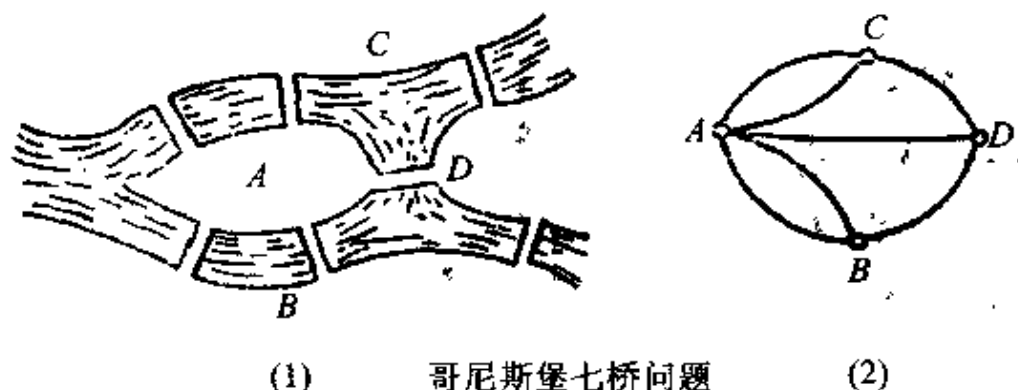


图 8.2.3

欧拉巧妙地将陆地分别用点来表示,每座桥用连接两点的边

来表示,这样将七桥问题抽象成一个图 G (如图 8.2.3(2)),于是问题就变成,在图 G 中,从某点出发,每条边恰好经过一次再回到原处,这样的走法是否存在?也就是判定图 G 是否为 E 图。由定理 8.2.1,显然这不是 E 图。所以上述的走法是不存在的。

2. 周游世界问题

所谓哈密尔顿问题,最初是由数学家哈密尔顿(Hamilton)在 1859 年向他的一个朋友提出的一个数学游戏产生的。

问题:世界上有二十个大城市,某人从某一城市出发到其他城市旅游,每个城市经过一次且仅经过一次,再回到原出发地,问这样的旅行路线是否存在?这就是所谓的周游世界问题。

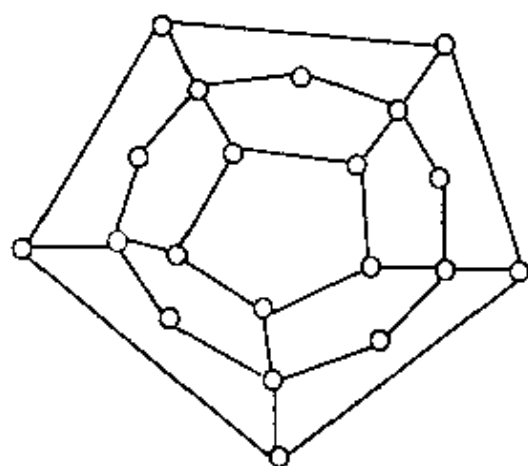


图 8.2.4

哈密尔顿将这二十个城市用一个正十二面体的二十个顶点表示,旅行者沿着多面体的棱旅行。将这个正十二面体“拉平”成为平面图(图 8.2.4),由图可见,上述要求的旅行路线是存在的,且不止一条,如黑线标出的就是其中一条。这就是一个哈密尔顿圈。

3. 中国邮递员问题

自从欧拉解决了七桥问题以后,长期以来它都是数学爱好者手中的一个趣味数学问题——一笔画问题,或是数学工作者手中的一个纯理论性的问题。自从中国邮递员问题提出来以后,这个问题才具有了强烈的实际意义。

问题:邮递员在他沿着邮路出发之前,必须先从邮局取出他所应分发的邮件,然后沿着邮路的每一个街段分送邮件,最后又返回邮局。为了节省时间,每一位邮递员都愿以尽可能少的行程走完他

所必走的所有路程。用非图论的话说,所谓中国邮递员问题是这样一类问题:如何以尽可能少的行程遍历邮路上的所有各条街道而又回到他的出发点。很明显,不仅邮递员,而且很多其他实际问题都具有这类现象。

这类问题是由中国数学家管梅谷教授在 1962 年提出的,因而在国际上称这类问题为中国邮递员问题。

如果用顶点表示邮局和街道的交叉路口,用边表示邮递员所负责的街道,则中国邮递员问题就可以抽象成一个赋权图。用图论的话说,就是在一个赋权图 $G = (V, E, L)$ 上,找一条回路 W , 使 W 包含 G 中每条边至少一次,且具有最小长度。

我们知道,如果图 G 是欧拉图,则 W 就是欧拉回路,可以通过 *Hleury* 算法寻找。但若 G 不是欧拉图,则问题的解决就比较困难了。下面给出解决这类问题的方法。

我们知道,不是欧拉图的图中一定有偶数个奇顶点,于是我们可以用奇偶点作业法解决这个问题。计算步骤如下:

(1) 找出图 G 中所有奇顶点(必有偶数个),将它们两两配对。每对奇顶点间必有一条通路 P (因 G 是连通图),将通路 P 上所有的边都重复一次加到图 G 上,使所得到的新图中的顶点全是偶顶点。

(2) 如果边 $e = (v, u)$ 上重复边数多于一条,则可从 e 的重复边中去掉偶数条,使得图中顶点仍全部是偶顶点。

(3) 检查图中的每个圈,如果每一个圈的重复边的总长不大于该圈总长的一半时,则已求得最优方案。

如果存在一个圈,重复边的总长大于该圈总长的一半时,就进行调整,将这个圈中的重复边去掉,而将该圈中原来没有重复边的各边加上重复边,其他各圈的边不变,返回(2)。

以上过程可以总结为口诀:先分奇偶点,奇点对对连,连线不重迭,重迭需改变,圈上连线长,不得过半圈。

例 1 如图 8.2.5, 图中各边所标注的数字, 表示该边的长度, 邮递员从邮局 A 出发, 求他的最优投递路线。

解: 图中的奇顶点有 v_1, v_2, v_5, v_6 , 将 $v_1 v_5$ 配成一对, $v_2 v_6$ 配成一对。

在图中直接求得 v_1-v_5 的通路 $v_1 v_6 v_5$, 求得 v_2-v_6 通路 $v_2 v_1 v_6$, 并将通路上各边都重复一次得图 8.2.6。

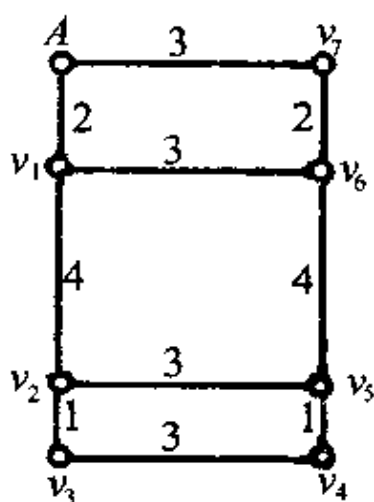


图 8.2.5

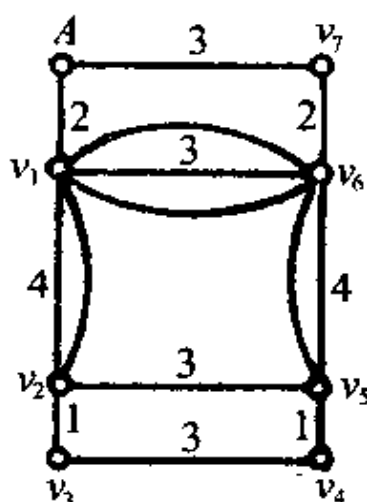


图 8.2.6

在图 8.2.6 中, 边 $(v_1 v_6)$ 上重复边多于一条, 因此可以去掉偶数条重复边, 就得到图 8.2.7。这图中, 共有六个圈, 要逐圈检查。显然, 在图 $v_2 v_5 v_6 v_1 v_2$ 中, 重复边总长 $= 4 + 4 = 8$, 而该圈总长 $= 4 + 3 + 4 + 3 = 14$, 而 $8 > \frac{14}{2}$, 需要进行调整, 将该圈中的重复边去掉, 原来圈中没有重复边的加上重复边, 得到图 8.2.8, 而不在该圈中的其他边均不做改变。

然后, 再重新逐圈进行检查, 可知已得到最优投递路线。根据图 8.2.8 可作出最优投递路线的具体行走路线 (图 8.2.9 中虚线所示)。

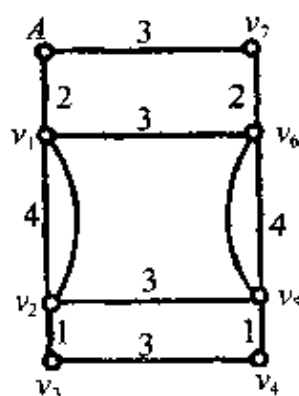


图 8.2.7

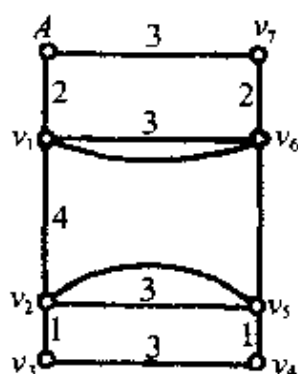


图 8.2.8

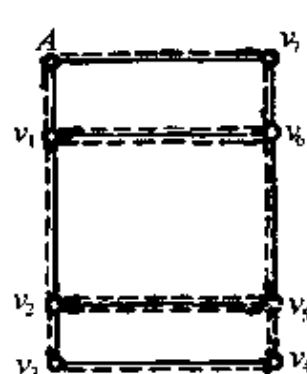


图 8.2.9

由此例可见,当重复边所在的圈有多个时,计算量是相当大的,所以很不方便,Edmonds 和 Johnson,1973 年提出一个比较有效的方法,有兴趣者可参见 *Math Programming* 5 (1973) 88 - 124。

4. 货郎担问题

货郎担问题是哈密尔顿问题的一个有名的应用问题,提法是这样的:一个货郎要去若干城镇售货,已经知道各城镇之间的距离(设各城镇之间都有交通线相连接),那么货郎应如何选择行走路线,使每个城镇恰好经过一次,并回到原来出发地,使得总行程为最短。这个问题也称为旅行推销员问题。

抽象成图的问题,就是要在一个每边有非负权的完全图中,找一条总权最小的哈密尔顿圈,或简称为最小 H - 圈。

这个问题直到现在还没有一个完全令人满意的解法,这里介绍的“改良圈”解法只是一种比较满意的解法。这种解法是 Lin(1965) 和 Held, Karp(1970) 提出的,大致是这样进行的:由于该图是完全图,因而其必有哈密尔顿圈,而且在图中的顶点数大于 3 时,其哈密尔顿圈的数目并不惟一。任取其中一个,然后再逐步改进,直到不能再改进了为止,最后所得就是一个较好的解。

具体步骤如下:

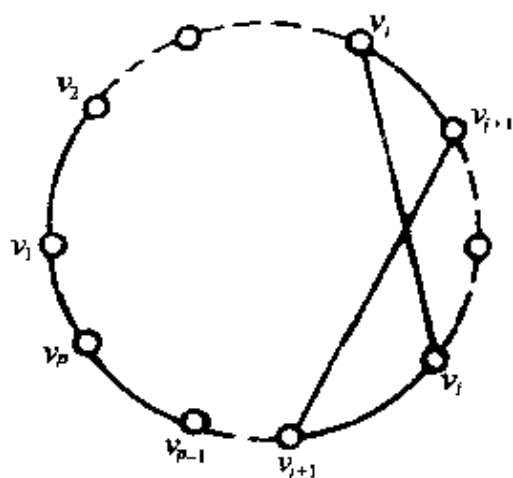


图 8.2.10

设 G 是一个赋权完全图。边 (v_i, v_j) 上的权, 记为 $L(v_i, v_j)$,

在图 G 中, 首先任取一个 H - 圈, 不妨设为 $C = v_1 v_2 \cdots v_p v_1$ (图 8.2.10)。如果存在 i, j 满足

① $1 < i + 1 < j < p$ (记 $v_{p+1} = v_1$)

② $L(v_i, v_j) + L(v_{i+1}, v_{j+1}) < L(v_i, v_{i+1}) + L(v_j, v_{j+1})$ 。

则从圈 C 中去掉边 (v_i, v_{i+1}) 和 (v_j, v_{j+1}) , 而添加边 (v_i, v_j)

和 (v_{i+1}, v_{j+1}) , 得到一个新 H - 圈: $C_1 = v_1 \cdots v_i v_j v_{j-1} \cdots v_{i+1} v_{j+1} v_{j+2} \cdots v_p v_1$ 。

显然 $L(C_1) < L(C)$, 把圈 C_1 的各顶点重新标号记作 $v_1 v_2 \cdots v_p v_1$, 继续这样做下去, 直到不存在那样的 i, j 为止。

我们可以看出, 利用上面的算法, 最后结果依赖于刚开始所找的 H - 圈, 所以用这种方法所求得解不一定是最好的。为了使结果更加接近于最优解, 我们可以从不同的 H - 圈开始, 重复做几次, 从中求得一个最好的 H - 圈。

例 2 某食品厂, 每天要用汽车将该厂的产品送到若干零售点, 再回到工厂继续送下一批货物。由于该车的送货地点每天均有变化, 因此, 希望找到一个方法, 使得对任意需求该厂产品的若干零售点都确定一条最好的行驶路线。假设某一天, 送货车要将货物送到五个零售点, A, B, C, D, E 。如果各零售点装卸货物的时间是固定的, 可以不加考虑。而且汽车在工厂与各零售点之间, 以及在各零售点之间的行驶时间均是已知的, 我们就可以画出一个有

六个顶点的赋权完全图(图 8.2.11)。图中工厂用 v 表示, 每条边旁的数字表示相应两点间汽车所需的行驶时间(单位: 分), 求花费总时间最少的一条行驶路线, 显然, 这个问题就是求该图的最小 H -圈。

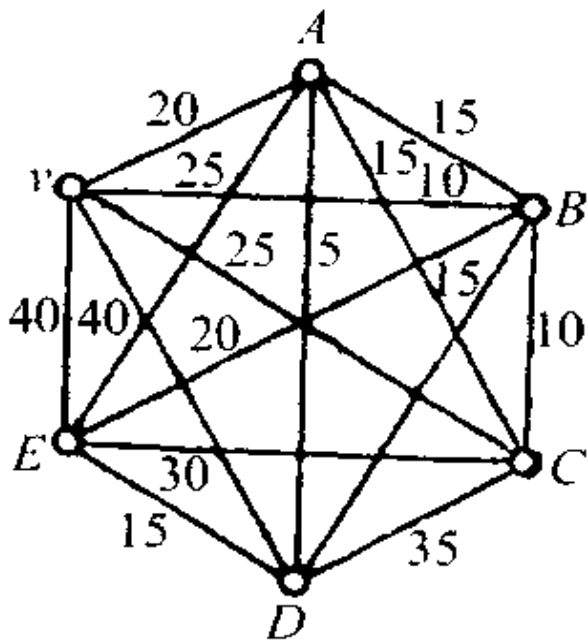


图 8.2.11

开始任意取一个 H -圈(图 8.2.2(1)) 该圈为 $v \rightarrow A \rightarrow B \rightarrow C \rightarrow D \rightarrow E \rightarrow v$, 总行驶时间为 $T = 135$ 分钟, 然后依图 8.2.12 的(1)至(4)作三次修改, 就得到一条总行驶时间为 $T = 90$ 分钟的行驶路线, 即送货路线为 $v \rightarrow B \rightarrow E \rightarrow D \rightarrow A \rightarrow C \rightarrow v$ 。

如果我们对于其他可能的各组零售点都求出相应的最优行驶路线, 就可以对于任何变化的情况作出较合理的决策。

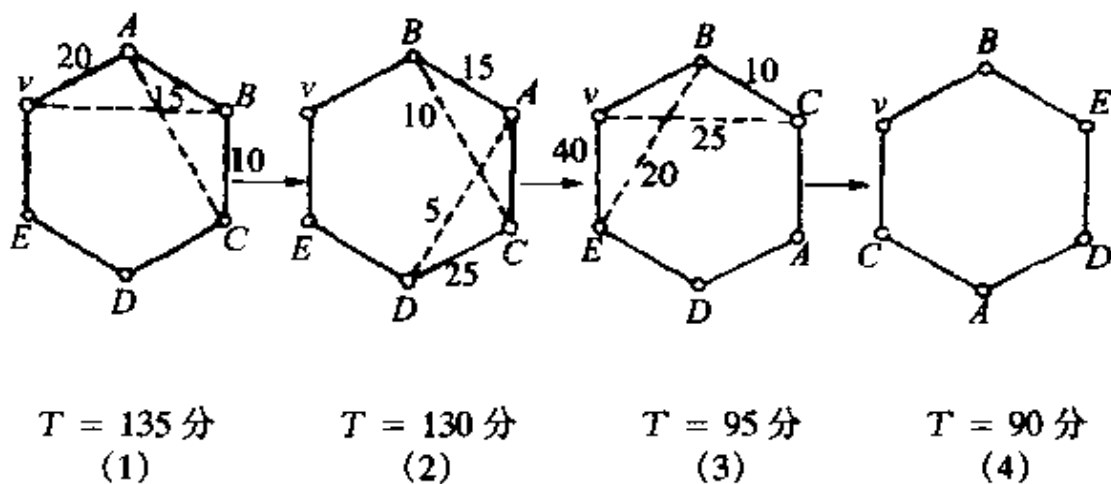


图 8.2.12

§ 8.3 树

树是一类特殊的图,早在 1847 年克希霍夫研究电网络时,便发展了有关树的理论。树在分子结构、电网络分析、计算机科学等领域有广泛的应用。

定义 8.3.1 不含圈的连通图称为树。一个无圈的图,如果它的每一分支都是树,则称为森林(图 8.3.1)。显然,树也是一个特殊的森林。

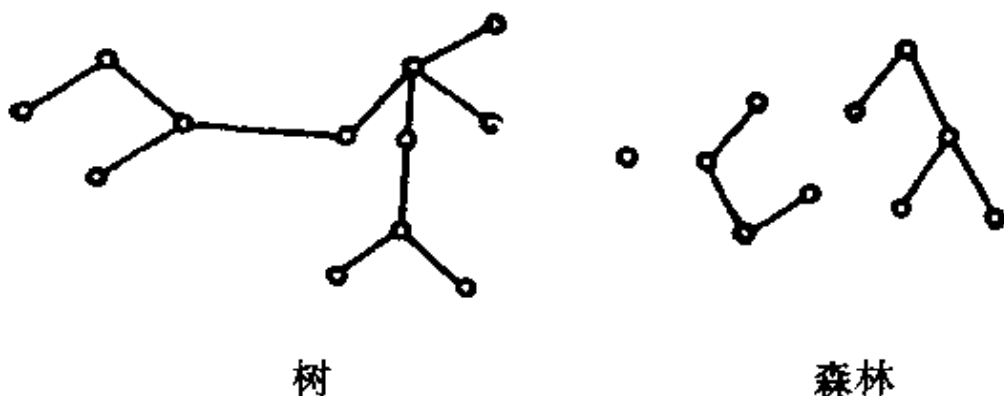


图 8.3.1

定理 8.3.1 如果图 G 是棵树, 则

- (1) G 的任意两个顶点间有惟一的一条路。
- (2) $q = p - 1$
- (3) 当 $p \geq 2$ 时, G 至少有两个次数为 1 的顶点。
- (4) 任意去掉 G 的一条边, 则 G 不连通。
- (5) 不相邻的两顶点间添加一条边, 可得到一个且仅得一个圈。

如果图 G 的支撑子图是一棵树 T , 则 T 称为图 G 的支撑树、

也称为部分树。(图 8.3.2)

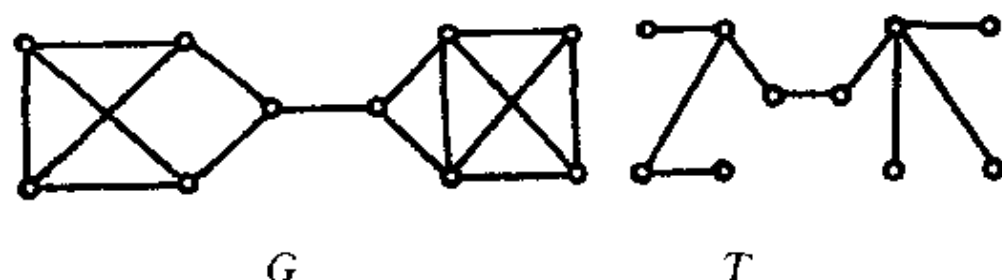


图 8.3.2

显然,任何连通图都有部分树,并且有时不只一个。

8.3.1 最小树

定义 8.3.2 在一个赋权的图 G 中(权数均非负), T 为 G 的一个部分树,并且 T 的边权和为 G 的所有部分树中最小者,称 T 为 G 的一棵最小树。

最小树在实际中有广泛应用,例如,要在若干城镇之间架设电话线,使任何两个城镇间都可以直通电话。设城镇 i, j 间架设电话线的费用是 ω_{ij} 。这时就需要确定一个总费用最小的架设方案。这就是一个求最小树的问题。

在这里我们介绍两种求最小树的方法。

1. 破圈法

对连通赋权图 $G = (V, E, L)$, 将其边按权之大小, 从大到小顺序排列。不妨设有 $L(e_1) \geq L(e_2) \geq \cdots L(e_m)$ 。首先看 e_1 是否在 G 的一个圈中, 若是将 e_1 舍去, 同时令 $G_1 = G - e_1$, 否则, 将 e_1 留下, 同时亦令 $G_1 = G$ 。

其次, 再看 e_2 是否在 G_1 的一个圈中, 若是将 e_2 舍去, 同时令 $G_2 = G_1 - e_2$, 否则, 将 e_2 留下, 同时亦令 $G_2 = G_1$, ……

如此下去, 一般地到第 i 步, 看 e_i 是否在 G_{i-1} 的一个圈中, 若是将 e_i 舍去, 同时令 $G_i = G_{i-1} - e_i$, 否则将 e_i 留下, 同时亦令

$$G_i = G_{i-1}, \dots$$

最后得到图 T , 显然是 G 的一个部分树, 而且是最小树。

例 1 用破圈法求图 8.3.3 的最小树。

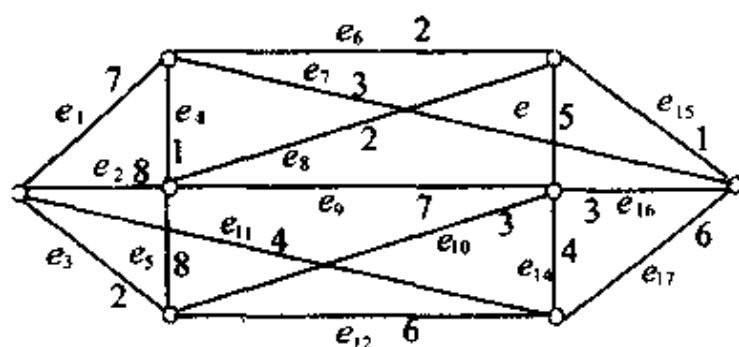


图 8.3.3

解: 首先将各边按权的大小排列如下:

8	8	7	7	6	6	5	4	4
e_2	e_5	e_1	e_9	e_{17}	e_{12}	e_{13}	e_{14}	e_{11}
3	3	3	2	2	2	1	1	
e_{16}	e_{10}	e_7	e_6	e_8	e_3	e_{15}	e_4	

按上述方法得 $E' = \{e_{11}, e_{16}, e_{10}, e_8, e_3, e_{15}, e_4\}$, 最小树 $T = G(E')$ (图 8.3.4) 中双线所示, 其权为 $T = 4 + 3 + 3 + 2 + 2 + 1 + 1 = 16$

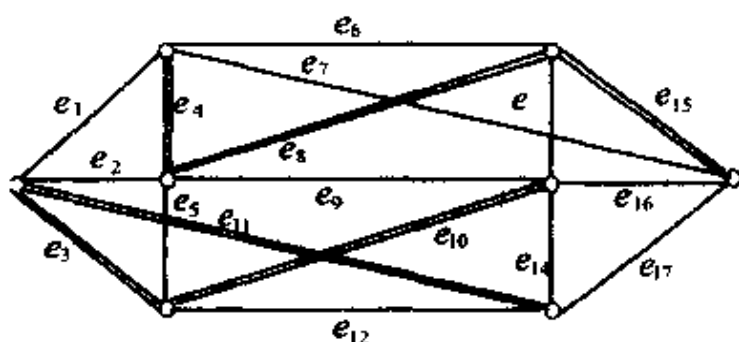


图 8.3.4

2. 加边法(Kruskal 算法)

对连通赋权图 $G = (V, E, L)$, 将其边按权的大小, 从小到大进行排序, 不妨设为

$$l(e_1) \leq l(e_2) \leq \dots \leq l(e_m)$$

首先选 $E_2 = \{e_1, e_2\}$

其次, 选 e_3 , 若 e_3 与 E_2 构成圈则舍去 e_3 , 令 $E_3 = E_2$, 否则, 令 $E_3 = \{e_3\} \cup E_2 \dots$

再选 e_4 , 若 e_4 与 E_3 构成圈则舍去 e_4 , 令 $E_4 = E_3$, 否则, 令 $E_4 = \{e_4\} \cup E_3, \dots$

如此下去, 一直到某一步 $k (k \geq n - 1)$, 使得边子集 E_k 满足 $|E_k| = n - 1$, 则所得到的 E_k 即为所求的最小树。

例 2 用加边法求图 8.3.5 的最小树。

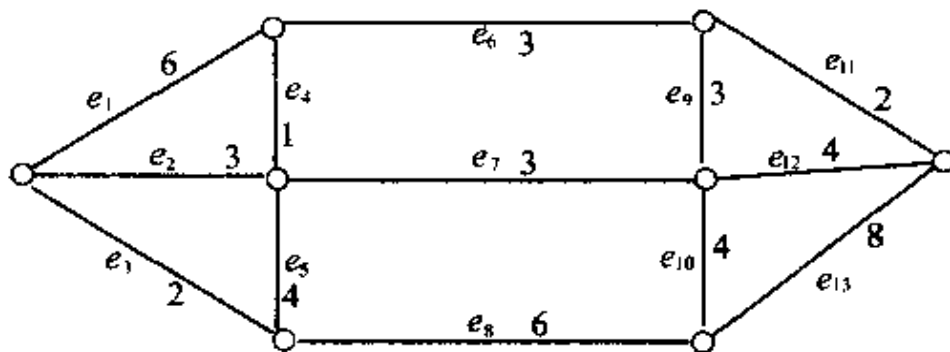


图 8.3.5

将边按权大小, 从小到大排列如下:

1	2	2	3	3	3	3	4	4	4	6	6	8
e_4	e_3	e_{11}	e_2	e_6	e_7	e_9	e_5	e_{10}	e_{12}	e_1	e_8	e_{13}

按上述方法得 $E' = \{e_4, e_3, e_{11}, e_2, e_7, e_9, e_{10}\}$, 则最小树 $T = G(E')$, 其权为 $T = 1 + 2 + 2 + 3 + 3 + 3 + 4 = 18$ 。如图 8.3.6 中双线所示。

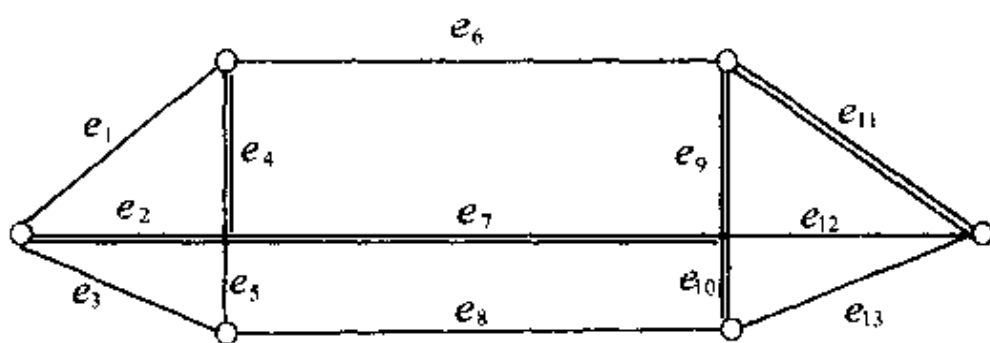


图 8.3.6

8.3.2 有向图

定义 8.3.3 集合 $V = \{v_1, v_2, \dots, v_n\}$ 叫做一个顶点集合, 其中的元素 $v_i (i = 1, 2, \dots, n)$ 叫做顶点, 集合 $A = \{a_1, a_2, \dots, a_m\}$ 叫做一个弧集合, 其中的元素 $a_j (j = 1, 2, \dots, m)$ 叫做弧。如果对任意的弧 $a_j \in A$, 都有且仅有 A 的一对有序的顶点 v_k, v_h 与之对应, (这时记 $a_j = (v_k, v_h)$), 那末我们就说顶点集合 V 与弧集合 A 一起构成了一个有向图 $D = (V, A)$ 。

对 $a_j(v_k, v_h)$, 称弧 a_j 为以“顶点 v_k 的出弧”“顶点 v_h 的入弧”, 同时又称顶点 v_k 为“弧 a_j 的始点”, 称 v_h 为“弧 a_j 的终点”。

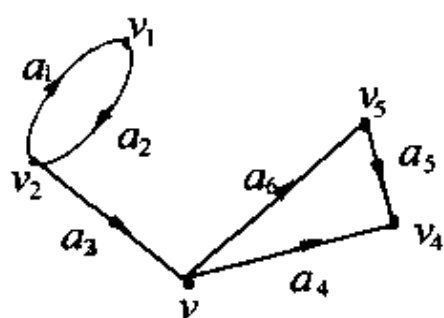
定义 8.3.4 对有向图 $D = (V, A)$, v 为其一个顶点, v 的出弧的条数叫做“顶点 v 的出次”, 并记其为 $d_D^+(v)$, v 的入弧的条数叫做“顶点 v 的入次”, 并记其为 $d_D^-(v)$; v 的出次与入次之和为 $d_D^+(v) + d_D^-(v)$, 称其为该顶点 v 的次数, 记其为 $d_D(v)$ 。

定义 8.3.5 一个有向图 $D(V, A)$, 其中每条弧 $a_j = (v_k, v_h)$ 都变为一条边 $e_j = (v_k, v_h)$, 这样得到的无向图 $G = (V, E)$ 为有向图 $D = (V, A)$ 的基础图, 如图 8.3.7。

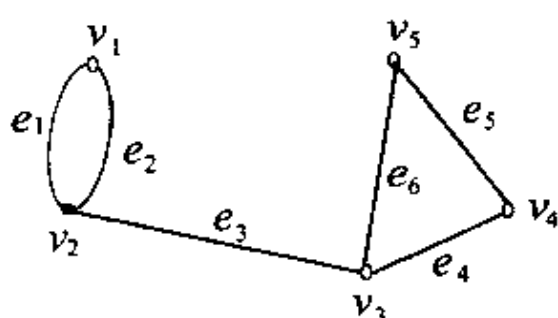
由上例可见, 一个有向图的基础图可是一个多重图。

8.3.3 有根树

定义 8.3.6 对有向图 $D = (V, A)$, 若其中有一点 v , 使对任



$$D = (V, A)$$



$$G = (V, E)$$

图 8.3.7

意点 v' , 都有从 v 到 v' (或从 v' 到 v) 的有向路, 则称 v 是 D 的一个“外(内)根”, 同时又称 D 是“有外(或内)根”的有向图。有外根或内根的有向图, 统称为“有根图”。

例如下面的图 8.2.8 中, 有向图 D_2, D_3 都是有外根的有向图, v_2 和 v_3 点分别为其外根, D_4 是有内根的有向图, v_4 是其内根, D_2, D_3, D_4 都是有根图, 而 D_1 不是有根图。

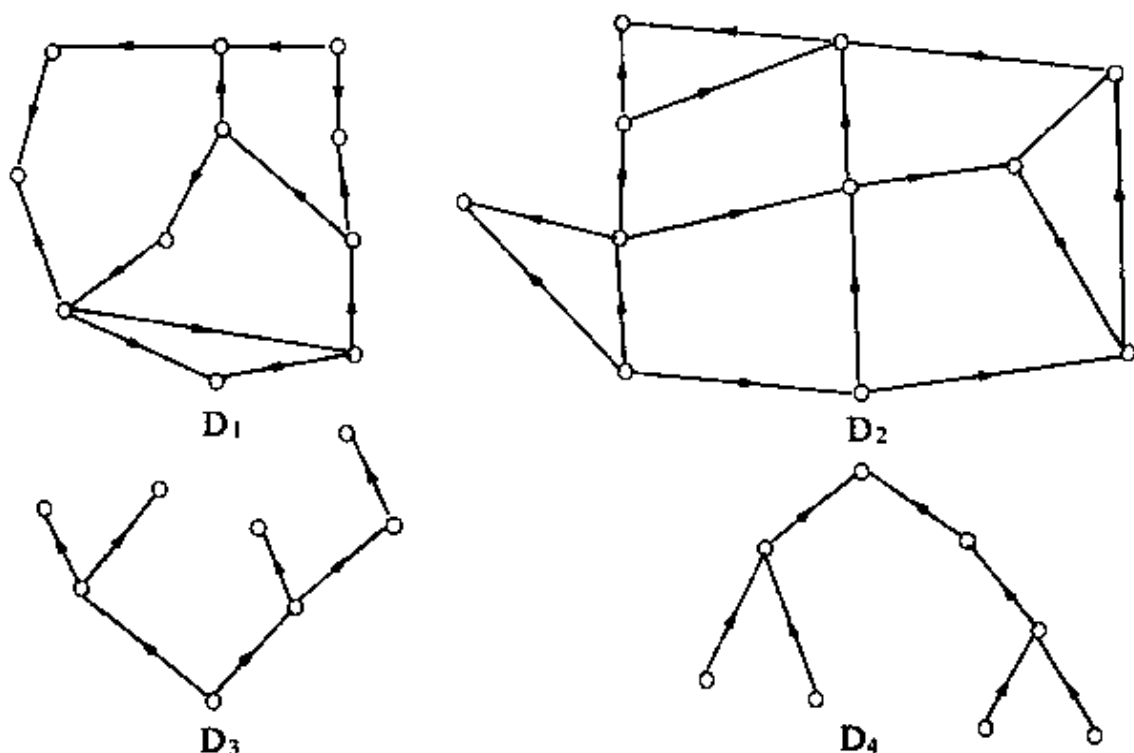


图 8.3.8

定义 8.3.7 对以其一点 v 为其外(或内)根的有外根(或内根)的有向图 $D = (V, A)$, 若其基础图 G 是一个树, 则称 D 是一个“有外(或内)根”的有向树或称为“外(或内)向树”, 点 v 称为外(或内)根, 有外根的有向树和有内根的有向树统称为“有根树”或“有向树”。

定义 8.3.8 对有向图 $D = (V, A)$, $T = (V, A')$ 为 D 的一个部分树, 如果 T 是一个有根树, 那末称 T 是 D 的一个“有根部分树”。

当然, 这里的有根部分树要分有外根的部分树和有内根的部分树两种, 亦可称其是有向(外向或内向)部分树。

在连通的有向赋权图中, 边权之和最小的有向部分树, 称为最小有向部分树。最小有向部分树在许多实际问题中都很有用。

8.3.4 建模实例

灌溉渠道的设计

例如图 8.3.9 所示的有向图 D 中, v_1 点表示水源, 其他各顶点表示需要灌溉的地块, 各条弧表示可以选来修建渠道的线路, 从

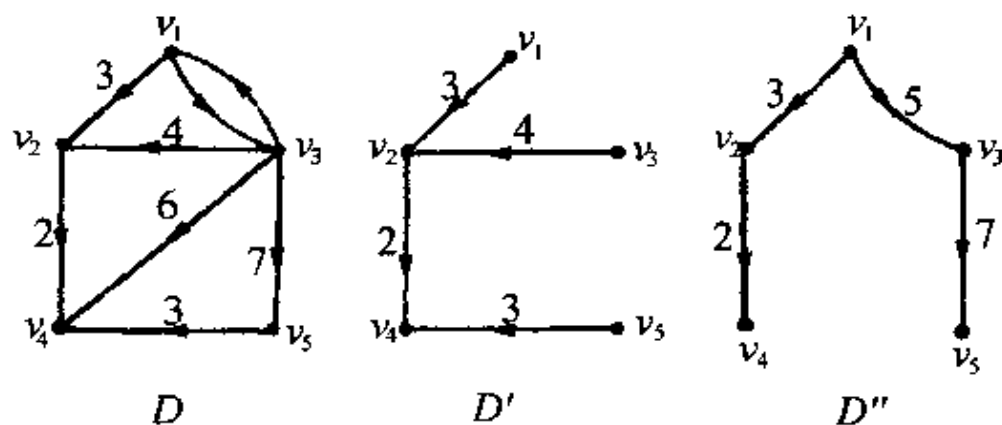


图 8.3.9

弧上箭头可以看出地势的高低。例如有一条弧从 v_1 到 v_2 , 这就表示 v_1 比 v_2 高, 水可以从 v_1 流到 v_2 去。又比如, 即有从 v_1 到 v_3 的

弧又有从 v_3 到 v_1 的弧,这说明 v_1 和 v_3 的地势一样高,水即可以从 v_1 流向 v_3 ,也可以从 v_3 流向 v_1 。现在需要解决的问题是:应该选择哪几条线路来修建渠道才能最节省?

如果不考虑弧的方向,用求最小树的方法可得到 D' ,它的各条弧长的总和是 12,但是按这个方案来修建渠道是不行的。因为 v_1 的水流不到 v_3 和 v_5 去。实际上,要使各地块都能得到灌溉而且总长度又最小的设计方案应该是 D'' 。虽然 D'' 的总长为 17,但这样设计才能使 v_1 的水流到各地块上去。这是求最合理的设计方案问题,就是求一个连通的有向赋权图的最小有向部分树问题。

关于求最小有向部分树的算法问题,因为较为烦琐,这里不再介绍,有兴趣者可以查看图论的专门书籍。

§ 8.4 最短路

定义 8.4.1 对赋权图 $D = (V, A, L)$, v_0 与 v_k 为其两顶点, $u = [v_0, v_1 \cdots v_k]$ 为其一条路, u 上各边权之和 $L(u) = \sum_{i=0}^{k-1} l(v_i, v_{i+1})$, 称为“路 u 之权(或长)”。若路 u 之权是 G 中以 v_0 和 v_k 为端点的所有路的权中最小者,则称 u 是“ D 中以 v_0 与 v_k 为两端点的最短路”。

所谓最短路问题就是对一个赋权图 G 及其中两顶点 v_i 和 v_j , 求以它们两为端点的最短路求法问题。

最短路问题是一个有着广泛应用价值的问题,例如,各种管道的铺设,线路的安排,输送网络费用等问题,都可以用到这里的最短路的求法。

在实际运用时,我们问题中的“边权”可以有各种解释。例如,在运输网络中,从 v_i 运送一批货物到 v_j ,若“边权”视为是通常意义的路程,则最短路问题就是使运输总路程最短的路线,若“边权”

表示输送时间,则最短路就是使运输总时间最短的路线,“边权”也可以代表费用,这时相对应的就是总费用最省的路线。

8.4.1 图中某点到其他顶点的最短路

Dijkstra 算法:

设在赋权有向图 $D = (V, A, L)$ 中,所有弧上的权均非负。如果 $(v_i, v_j) \in A$, 则令权 $l(v_i, v_j) = +\infty$ 。我们可以把权 $l(v_i, v_j)$ 看做弧 a_{ij} 的长度。将起点 v_1 到点 v_k 的最短通路的长记为 $L(v_1, v_k)$ 或 L_k , 记 $v_1 \rightarrow v_p$ 的单向路为 p 。

Dijkstra 算法的基本思想是很简单的:

设 S 是 v 的真子集, 并且 $v_1 \in S$, 记 $\bar{S} = V \setminus S$ 。如果 $p = v_1 v_2 \cdots v_k v$ 是从起点 v_1 到 \bar{S} 的最短路, 我们用 $L(v_1, \bar{S})$ 来表示此路的长, 则显然有 $v_k \in S, v \in \bar{S}$, 并且 p 上的路 $v_1 \rightarrow v_k$ 必然是最短的 $v_1 \rightarrow v_k$ 路 (见图 8.4.1), 所以有 $L(v_1, v) = L(v_1, v_k) + l(v_k, v)$, $L(v_1, \bar{S}) = \min_{\substack{u \in S \\ v \in \bar{S}}} \{L(v_1, u) + l(u, v)\}$ 。

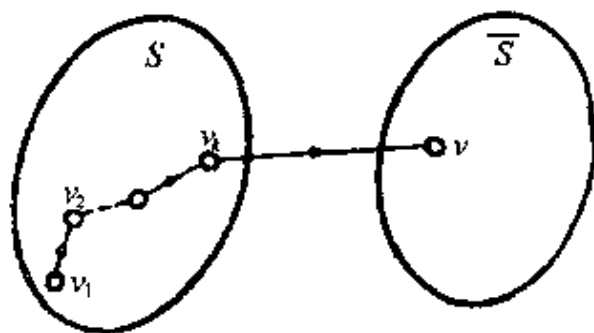


图 8.4.1

根据上式, 得到以下标号算法:

① 开始给顶点 v_1 标上(永久)标号 $L_1 = 0$, 其他各点标上临时标号 $L_k^{(1)} = l_{1k}$, ($k = 2, 3, \cdots, p$)。令 $S_1 = \{v_1\}$, $\bar{S}_1 = \{v_2, \cdots, v_p\}$ 。如果设 $\min\{l(a_{ij})\} = l(a_{12})$, 则 v_2 是与 v_1 距离最近的点, 即求出了最短 $v_1 - v_2$ 通路, 其长记为 L_2 。这时 v_2 获得永久标号

L_2 。

令 $S_2 = \{v_1 v_2\}$, $\bar{S}_2 = \{v_3, \dots, v_p\} = \bar{S}_1 \setminus \{v_2\}$

下面我们总是用 S_m 记所有得到永久标号的顶点集, 用 \bar{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 $v_1 \in S_m$, $\bar{S}_m = V \setminus S_m$ 。在 \bar{S}_m 中求一点 v_k , 使得 $L_k = L_k^{(m)} = \min_{v_j \in \bar{S}_m} \{L_j^{(m)}\}$, 若 $L_k = +\infty$, 则说明不存在 v_1 到 v_k 的最短路。否则令 $S_{m+1} = S_m \cup \{v_k\}$, $\bar{S}_{m+1} = \bar{S}_m \setminus \{v_k\}$ 。

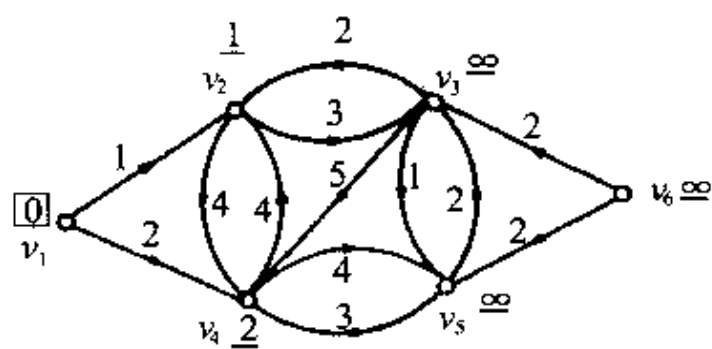
③再修改临时标号; 对 \bar{S}_{m+1} 中的每一点 v_j , 令 $L_j^{(m+1)} = \min\{L_j^{(m)}, L_k + l_{kj}\}$, 其中 L_k 是在②中已经求出的。

再返回步骤②, 直到 $m = p - 1$ 为止。

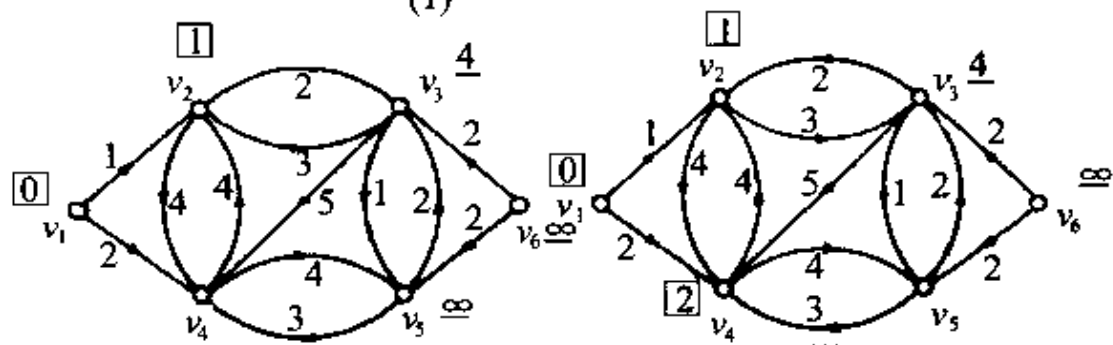
例1 求下图中, v_1 到其他各点的最短路, 首先给 v_1 赋以永久标号 $L_1 = 0$, 其他顶点标上临时标号 $L_j^{(1)} = l_{1j}$ ($j = 2, \dots, 6$)。如图 8.4.2 中(1), 图中永久标号用顶点旁边的 \square 中的数字表示, 临时标号用顶点旁边的 $_$ 上数字表示。这时 $S_1 = \{v_1\}$, $\bar{S}_1 = \{v_2, v_3, \dots, v_6\}$, $v_j \in \bar{S}_1$, 然后, 我们求得 $\min_{v_j \in \bar{S}_1} \{L_j^{(1)}\} = L_2^{(1)} = 1$, 于是 v_2 得到永久标号 $L_2 = L_2^{(1)} = 1$ 。再修改其他顶点的临时标号。例如 $L_3^{(2)} = \min\{L_3^{(1)}, L_2 + l_{23}\} = \min\{\infty, 1 + 3\}$, 所以 $L_3^{(2)} = 4$ 。类似地, 修改其他的临时标号, 得到图 8.4.2(2), 重复以上过程, 求解的各步可见图 8.4.2 的(1)~(5)。最后得到从 v_1 到 v_2, v_3, v_4, v_5 的最短通路长分别为 1、4、2、5, v_1 到 v_6 无单向路。如果还需求出具体的最短路线, 我们只要在各点得到永久标号的同时, 记录其先前的顶点, 就很容易得到具体的最短路了。

一个赋权的有向图, 也可以用其长度矩阵表示, 如图 8.4.2(1) 的长度矩阵如下:

$$L = \begin{pmatrix} 0 & 1 & \infty & 2 & \infty & \infty \\ \infty & 0 & 3 & 4 & \infty & \infty \\ \infty & 2 & 0 & 5 & 1 & \infty \\ \infty & 4 & \infty & 0 & 4 & \infty \\ \infty & \infty & 2 & 3 & 0 & \infty \\ \infty & \infty & 2 & \infty & 2 & 0 \end{pmatrix}$$

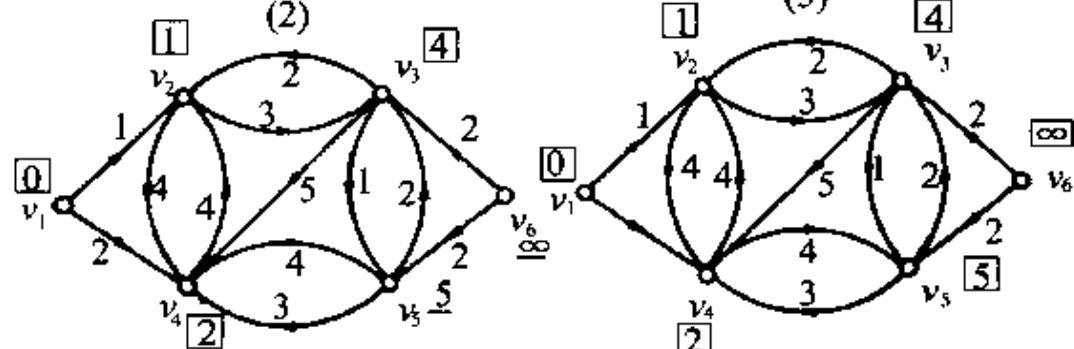


(1)



(2)

(3)



(4)

(5)

图 8.4.2

Dijkstra 算法还可以在长度矩阵上直接实现,其步骤如下:

① 划掉矩阵 L 的第一列,并给第一行标号 $\boxed{0}$ 。

② 在已标号的行中的未被划去的元素中,找一个最小的元素并圈起来,不妨设是 L_{ij} ,那么就把第 j 列划掉,同时给第 j 行以标号 \boxed{i} ,并把第 j 行各元素都加上 L_{ij} 。

③ 如果所有各行都已获得标号,则步骤终止。由标号返回追踪,则得到 v_1 到各点的最短路。

利用这一方法求解前面的例子,过程如下:

$$\begin{array}{c} \boxed{0} \\ \left[\begin{array}{cccccc} 0 & \textcircled{1} & \infty & 2 & \infty & \infty \\ \infty & 0 & 3 & 4 & \infty & \infty \\ \infty & 2 & 0 & 5 & 1 & \infty \\ \infty & 4 & \infty & 0 & 4 & \infty \\ \infty & \infty & 2 & 3 & 0 & \infty \\ \infty & \infty & 2 & \infty & 2 & 0 \end{array} \right] \end{array} \rightarrow \begin{array}{c} \boxed{0} \\ \boxed{1} \\ \left[\begin{array}{cccccc} 0 & \textcircled{1} & \infty & \textcircled{2} & \infty & \infty \\ \infty & 0+1 & 3+1 & 4+1 & \infty & \infty \\ \infty & 2 & 0 & 5 & 1 & \infty \\ \infty & 4 & \infty & 0 & 4 & \infty \\ \infty & \infty & 2 & 3 & 0 & \infty \\ \infty & \infty & 2 & \infty & 2 & 0 \end{array} \right] \end{array} \rightarrow$$

$$\begin{array}{c} \boxed{0} \\ \boxed{1} \\ \left[\begin{array}{cccccc} 0 & \textcircled{1} & \infty & \textcircled{2} & \infty & \infty \\ \infty & 1 & \textcircled{4} & 5 & \infty & \infty \\ \infty & 2 & 0 & 5 & 1 & \infty \\ \infty & 4+2 & \infty & 0+2 & 4+2 & \infty \\ \infty & \infty & 2 & 3 & 0 & \infty \\ \infty & \infty & 2 & \infty & 2 & 0 \end{array} \right] \end{array} \rightarrow \begin{array}{c} \boxed{0} \\ \boxed{1} \\ \boxed{2} \\ \boxed{1} \\ \left[\begin{array}{cccccc} 0 & \textcircled{1} & \infty & \textcircled{2} & \infty & \infty \\ \infty & 1 & \textcircled{4} & 5 & \infty & \infty \\ \infty & 2+4 & 0+4 & 5+4 & \textcircled{1+4} & \infty \\ \infty & 6 & \infty & 2 & 6 & \infty \\ \infty & \infty & 2 & 3 & 0 & \infty \\ \infty & \infty & 2 & \infty & 2 & 0 \end{array} \right] \end{array} \rightarrow$$

$$\begin{array}{c} \boxed{0} \\ \boxed{1} \\ \boxed{2} \\ \boxed{1} \\ \boxed{3} \\ \left[\begin{array}{cccccc} 0 & \textcircled{1} & \infty & \textcircled{2} & \infty & \infty \\ \infty & 1 & \textcircled{4} & 5 & \infty & \infty \\ \infty & 6 & 0 & 9 & \textcircled{5} & \infty \\ \infty & 6 & \infty & 2 & 6 & \infty \\ \infty & \infty & 2+5 & 3+5 & 0+5 & \infty \\ \infty & \infty & 2 & \infty & 2 & 0 \end{array} \right] \end{array} \rightarrow \begin{array}{c} \boxed{0} \\ \boxed{1} \\ \boxed{2} \\ \boxed{1} \\ \boxed{3} \\ \boxed{6} \\ \left[\begin{array}{cccccc} 0 & \textcircled{1} & \infty & \textcircled{2} & \infty & \infty \\ \infty & 1 & \textcircled{4} & 5 & \infty & \infty \\ \infty & 6 & 0 & 9 & \textcircled{5} & \infty \\ \infty & 6 & \infty & 2 & 6 & \infty \\ \infty & \infty & 7 & 8 & 5 & \infty \\ \infty & \infty & 2 & \infty & 2 & 0 \end{array} \right] \end{array}$$

注意在计算过程中的最后两个矩阵中,由于在已标号的行中的未划去的元素都是 ∞ ,这时不存在最小元素(即矩阵的第6列中的前5个元素都是 ∞),这时可划去第6列,而第6列得标号⑥,实际上表示没有从 v_1 到 v_6 的单向路。

由短阵最左边的标号返回追踪,得到 v_1 到各点的最短路(图8.4.3)。而且在最后一个短阵中圈起来的元素恰好是 v_1 至各点的最短路的长,即 $L_1 = 1, L_3 = 4, L_4 = 2, L_5 = 5, L_6 = \infty$ 。

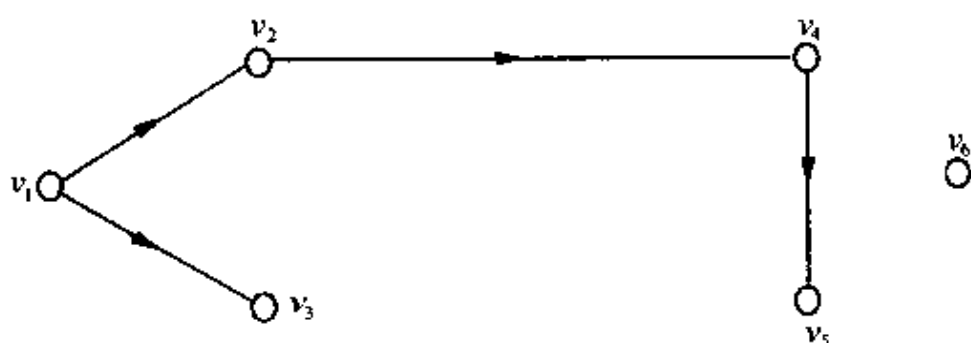


图 8.4.3

在无向图 $G = (V, E)$ 中,也可以利用 Dijkstra 算法求指定两点间的最短通路,这只要把图 G 中的每条边都看成是具有相同权的两条方向相反的弧,则 G 也可以看做是有向图了。

8.4.2 图中任意两点间的最短路

图中任意两点之间的最短路问题,也称为距离表问题,在实际生产中也有着很广泛的应用。例如,航空公司每天都要运送很多旅客来往于全国许多城市之间。由于节省的原因,每条航线应使各个旅客的总行程达到最小。这就需要编制一个全国每一对城市之间的空中最短路航行里程表。这就是这里所讲的距离表。

其实,要把一个图的任意两点间的最短路都求出来,对我们并不困难。可以将图中每一个顶点依次视为计算的始点,用 Dijkstra 算法就行了。不过那就要用到大量的计算。这里介绍另一个较好一

些的算法, - Floyd 算法。

R. W. Floyd (1962) 算法:

首先取 $D^{(0)} = (L_{ij})$ ——边权矩阵

然后计算出 n 个矩阵 $D^{(1)} D^{(2)} \dots D^{(n)}$, 其中 $D^{(k)} = (l_{ij}^{(k)})$,
 $k = 1, 2, \dots, n, l_{ij}^{(k)} = \min\{l_{ij}^{(k-1)}, l_{jk}^{(k-1)} + l_{ki}^{(k-1)}\}$. 这时, $D^{(n)}$ 中
 的元素 $l_{ij}^{(n)}$ 之值就是顶点 v_i 与 v_j 之间最短路之值。

例 2 求下图的距离表。

该图为无向图, 其长度矩阵为

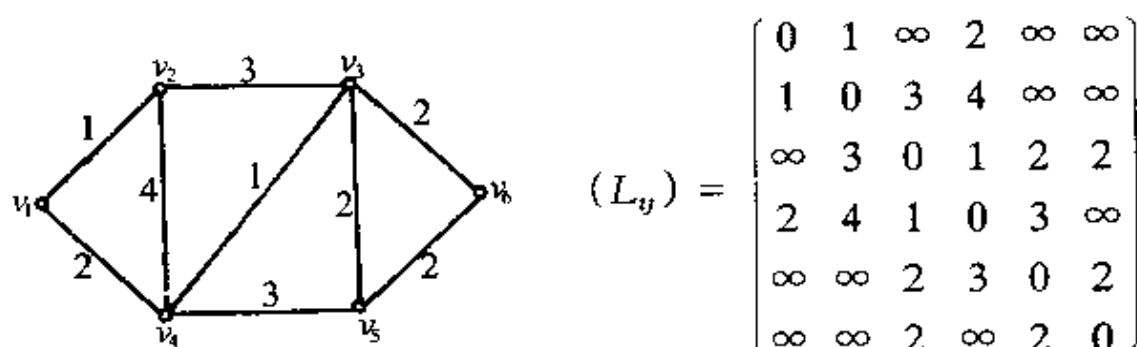


图 8.4.4

由 R. W. Floyd 算法有, 取 $D^0 = (L_{ij})$

$$D^{(1)} = \begin{bmatrix} 0 & 1 & \infty & 2 & \infty & \infty \\ 1 & 0 & 3 & 3 & \infty & \infty \\ \infty & 3 & 0 & 1 & 2 & 2 \\ 2 & 3 & 1 & 0 & 3 & \infty \\ \infty & \infty & 2 & 3 & 0 & 2 \\ \infty & \infty & 2 & \infty & 2 & 0 \end{bmatrix} \quad D^{(2)} = \begin{bmatrix} 0 & 1 & 4 & 2 & \infty & \infty \\ 1 & 0 & 3 & 3 & \infty & \infty \\ 4 & 3 & 0 & 1 & 2 & 2 \\ 2 & 3 & 1 & 0 & 3 & \infty \\ \infty & \infty & 2 & 3 & 0 & 2 \\ \infty & \infty & 2 & \infty & 2 & 0 \end{bmatrix}$$

$$D^{(3)} = \begin{bmatrix} 0 & 1 & 4 & 2 & 6 & 6 \\ 1 & 0 & 3 & 3 & 5 & 5 \\ 4 & 3 & 0 & 1 & 2 & 2 \\ 2 & 3 & 1 & 0 & 3 & 3 \\ 6 & 5 & 2 & 3 & 0 & 2 \\ 6 & 5 & 2 & 3 & 2 & 0 \end{bmatrix} \quad D^{(4)} = \begin{bmatrix} 0 & 1 & 3 & 2 & 5 & 5 \\ 1 & 0 & 3 & 3 & 5 & 5 \\ 3 & 3 & 0 & 1 & 2 & 2 \\ 2 & 3 & 1 & 0 & 3 & 3 \\ 5 & 5 & 2 & 3 & 0 & 2 \\ 5 & 5 & 2 & 3 & 2 & 0 \end{bmatrix}$$

$$D^{(5)} = \begin{bmatrix} 0 & 1 & 3 & 2 & 5 & 5 \\ 1 & 0 & 3 & 3 & 5 & 5 \\ 3 & 3 & 0 & 1 & 2 & 2 \\ 2 & 3 & 1 & 0 & 3 & 3 \\ 5 & 5 & 2 & 3 & 0 & 2 \\ 5 & 5 & 2 & 3 & 2 & 0 \end{bmatrix} \quad D^{(6)} = \begin{bmatrix} 0 & 1 & 3 & 2 & 5 & 5 \\ 1 & 0 & 3 & 3 & 5 & 5 \\ 3 & 3 & 0 & 1 & 2 & 2 \\ 2 & 3 & 1 & 0 & 3 & 3 \\ 5 & 5 & 2 & 3 & 0 & 2 \\ 5 & 5 & 2 & 3 & 2 & 0 \end{bmatrix}$$

于是得距离表:

	v_1	v_2	v_3	v_4	v_5	v_6
v_1	0	1	3	2	5	5
v_2	1	0	3	3	5	5
v_3	3	3	0	1	2	2
v_4	2	3	1	0	3	3
v_5	5	5	2	3	0	2
v_6	5	5	2	3	2	0

上面已经看出,事实上到 $D^{(4)}$ 就已经得到这距离表了,因而知 $D^{(6)}$ 是最大的数目。

8.4.3 建模实例

1. 设备更新问题

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。现假设该公司在第一年开始时必须购置一台这种设备。为简便起见,假设计划使用这台设备的时间为五年,估计这台设备的购买费用和维修费(单位:万元)如表 8.4.1、8.4.2。

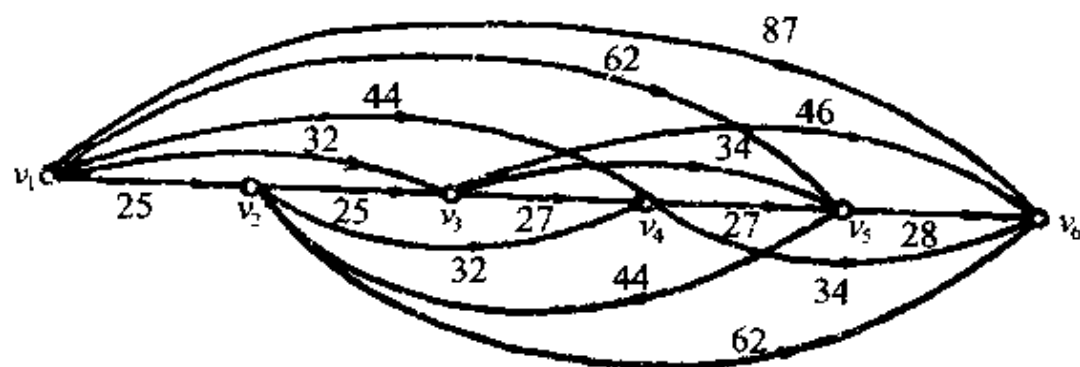
表 8.4.1 第一年到第五年的购买价格

年号	1	2	3	4	5
价格	20	20	22	22	23

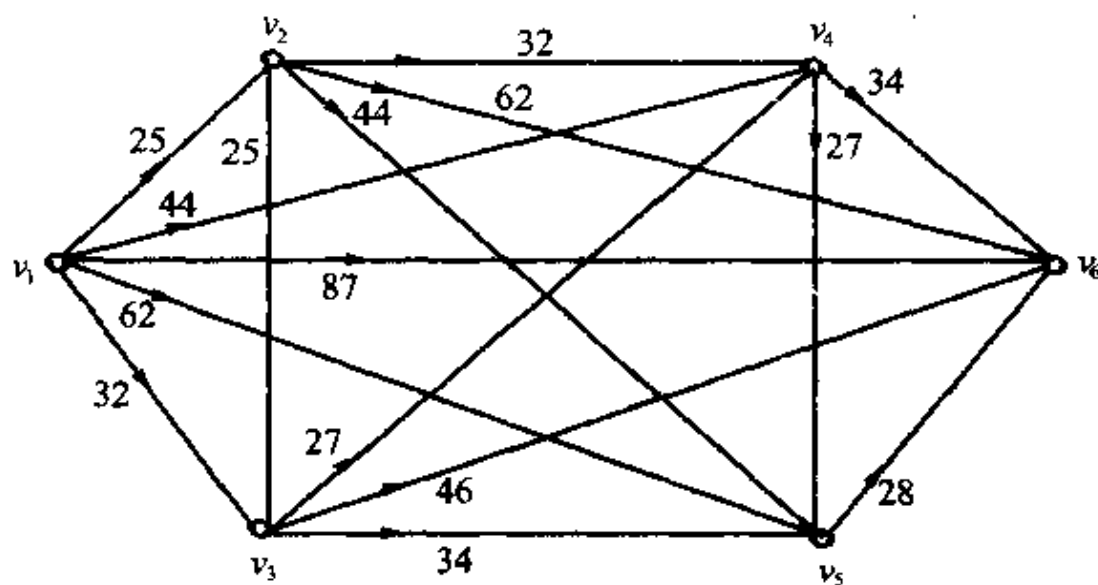
表 8.4.2 不同使用年龄的设备的维修费

使用年龄	0 - 1	1 - 2	2 - 3	3 - 4	4 - 5
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。



(1)



(2)

图 8.4.5

这个问题化成图的问题：考虑六个点 $v_1, v_2, v_3, v_4, v_5, v_6$ ，其中 $v_i (i = 1, 2, 3, 4, 5)$ 表示在第 i 年初要购买新设备。 v_6 是虚设点，表示第五年的年底才购置新设备。再从点 $v_i (i = 1, 2, 3, 4, 5)$ 引出指向点 $v_{i+1}, v_{i+2}, \dots, v_6$ 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 $(j = 2, \dots, 6)$ 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 年初这段时间的维修总费用。例如，弧 (v_1, v_4) 上所赋的权 $L(v_1, v_4) = 20 + (5 + 7 + 12) = 44$ 万元，等等。这样，就可以得到一个赋权有向图 D (图 8.4.5)。于是问题就变成：要在一个赋权有向图 D 中，求一条总权最小的 v_1 至 v_6 的单向路问题。这就是最短路问题。

由刚才介绍的 Dijkstra 算法，最优解为 $v_1 - v_3 - v_6$ 或 $v_1 - v_4 - v_6$ ，即第一年购置新设备，用到第 3 年或第 4 年更换，总费用最小，为 78 万元。

2. 服务点的选址问题

某县拟建立一个消防站，为该县辖区内七个城镇服务，问需设在哪个城镇上，才能使它至最远城镇的路程达到最小？

该问题抽象成图，如图 8.4.6 所示。

类似本例中的消防站设置问题有：医院、商店、电影院、仓库等的设置，总之可称为“服务点选址问题”。这类问题的一般提法是：

对一个连通赋权图 $G = (V, E, L)$ 现取其中一点 v_i ，然后考虑此点与 G 中

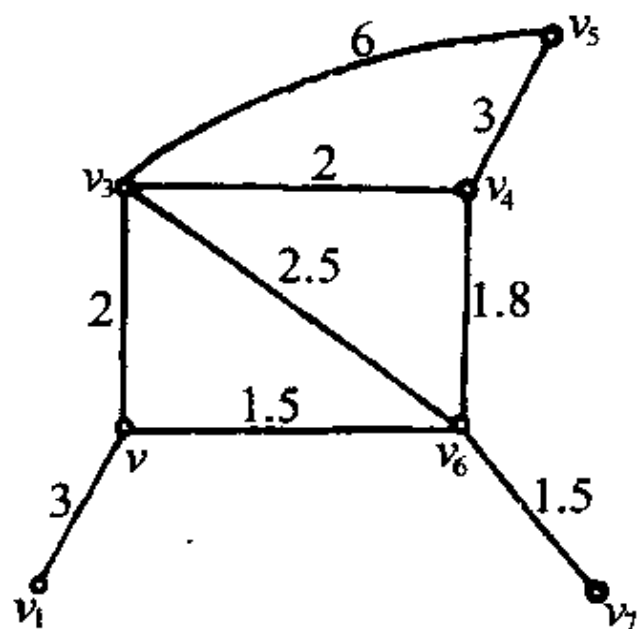


图 8.4.6

其他各点间最短路的长: $d(v_i v_1) \ d(v_i v_2) \cdots d(v_i v_n)$, 我们称这几个距离中最大的数为顶点 v_i 的最大服务距离, 记其为 $S(v_i)$, 即有 $S(v_i) = \max_{v_j \in V} \{d(v_i v_j)\}$ 。

我们现在的问题就是要求出: $S(v_1) S(v_2) \cdots S(v_n)$ 中的最小者的值及其所对应的点。设 v_k 就是所求的点, 即有 $S(v_k) = \min_{v_i \in V} \{S(v_i)\}$, 我们称 v_k 为图 G 的中心点。

求中心点的算法步骤如下:

- ① 用 Floyd 算法求出图 G 的距离表,
- ② 对每个点 v_i , 求出数 $S(v_i)$,
- ③ 求出中心点。

对本例计算结果如下表。

	v_1	v_2	v_3	v_4	v_5	v_6	v_7	$S(v_i)$
v_1	0	3	5	6.3	9.3	4.5	6	9.3
v_2	3	0	2	3.3	6.3	1.5	3	6.3
v_3	5	2	0	2	5	2.5	4	5
v_4	6.3	3.3	2	0	3	1.8	3.3	6.3
v_5	9.3	6.3	5	3	0	4.8	6.3	9.3
v_6	4.5	1.5	2.5	1.8	4.8	0	1.5	4.8
v_7	6	3	4	3.3	6.3	1.5	0	6.5

由该表可知 v_6 为所求的中心点

$S(v_6) = \min\{S(v_i)\} = 4.8$ 即消防站应设在 v_6 。

§ 8.5 最大流

现在考虑这样的问题: 在战争期间, 要把后方基地 $x_1, x_2 \cdots x_n$ 处的物资, 通过铁道运到前方的 $y_1, y_2 \cdots y_n$ 处去。把铁路网上的车

站看做顶点,两个车站间的铁路线看做弧。可以认为在某条铁路线上可以运送的物资总量是有限的,我们把某条线路上的最大运送量称为它的容量。这时要考虑如何安排运输方案,使得由后方运到前方的物资总运量达到最大。

这就是要在这个铁路网中,求最大流的问题。这里的“流”是指弧上的实际运量。

8.5.1 最大流与最小割集

定义 8.5.1 在有向图 $D = (V, A)$ 中,指出两类顶点 V_1, V_2 , 分别称为起点集和终点集,其余的顶点 $(V \setminus \{V_1 \cup V_2\})$ 称为中间点,记为 M 。对于 D 的每一条弧 $a_{ij} = (v_i, v_j)$ 都对应一个数 $C(a_{ij}) > 0$ 称为弧 a_{ij} 的容量,简记为 c_{ij} ,将这样的赋权有向图 D 叫做一个容量网络,如图 8.5.1。

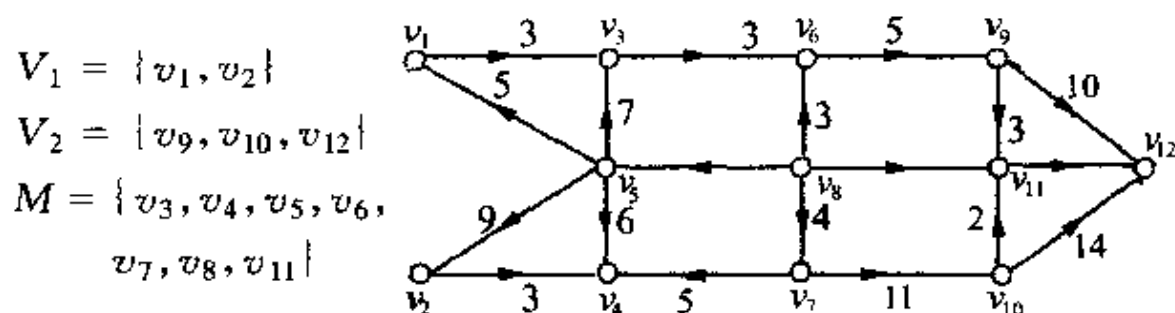


图 8.5.1

如果一个网络 D 中,有多个起点或多个终点,可以将 D 化成只有一个起点,一个终点的另一个网络 D' ,这只需在 D 中增加两个新顶点 x, y 。从 x 到起点集 V_1 中的每一点各连结一条容量为 $+\infty$ 的弧,从终点集 V_2 中的每一点到 y 各连结一条容量为 $+\infty$ 的弧,就得到一个新的网络 D' ,在 D' 中指定 x 为起点(发点), y 为终点(收点)如图 8.5.1 中的网络就可以化为图 8.5.2 中只有一个

起点 x , 一个终点 y 的网络。

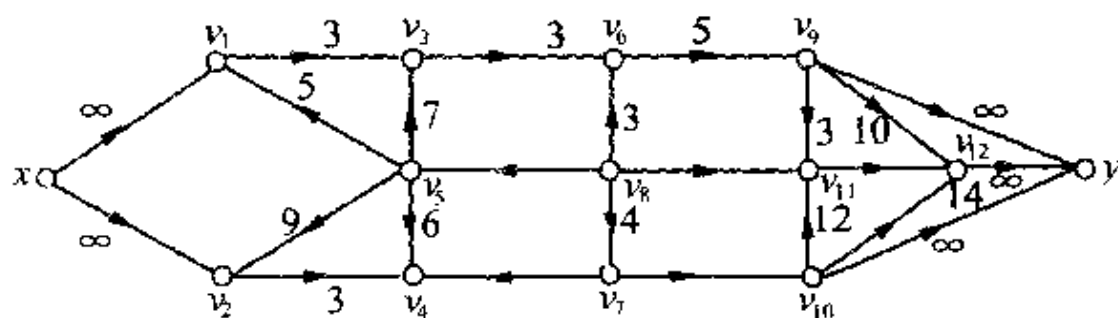


图 8.5.2

所以,以后只讨论有一个起点和一个终点的网络 D 。

在网络 D 中,把通过弧 a_{ij} 的运量记作 f_{ij} 并称之为弧 a_{ij} 上的流。显然有 $0 \leq f_{ij} \leq c_{ij}$ 。

为方便起见,我们将起点记为 v_1 , 终点记为 v_p , 显然,从起点 v_1 运出的货物总量 v 等于起点运出的货物总量减去其他点运往起点的货物总量。即 $\sum_{a_{1j} \in A} f_{1j} - \sum_{a_{j1} \in A} f_{j1} = v$ 。同时,终点 v_p 净收到的货物总量应等于 v , 也等于从其他各点运进 v_p 的货物总重量减去从 v_p 运往其他点的货物总量, 即 $\sum_{a_{jp} \in A} f_{jp} - \sum_{a_{pj} \in A} f_{pj} = v$ 。对于中间点,运进该点的货物总量应等于从该点运出的总量。

定义 8.5.2 设 $D = (V, A)$ 中,如果每条弧 a_{ij} 都对应一个数 f_{ij} 且满足:

(1) $0 \leq f_{ij} \leq c_{ij}$ (容量限制)

$$(2) \sum_{a_{ij} \in A} f_{ij} - \sum_{a_{ji} \in A} f_{ji} = \begin{cases} v & \text{当 } i = 1 \text{ 时} \\ 0 & \text{当 } i \neq 1, p \text{ 时} \\ -v & \text{当 } i = p \text{ 时} \end{cases}$$

则称 $|f_{ij}|$ 为 D 的一个可行流,记为 f , 并记 $v(f)$ 为对应于流 f 的流量(或流值)。

网络上的最大流问题,就是要在给定的网络上,求一个可行流 $f = \{f_{ij}\}$,使得流量 $v(f)$,取得最大值。显然,可行流是存在的。例如,令所有的 $f_{ij} = 0$,就是一个可行流。这时 $v(f) = 0$

定义 8.5.3 对于网络 $D = (V, A)$, 如果将 V 分成两个非空集合 S, \bar{S} , 使 $v_1 \in S, v_p \in \bar{S}$ 则所有起点属于 S 而终点属于 \bar{S} 的弧的集合,称为由 S 决定的割集,记作 (S, \bar{S}) 。割集 (S, \bar{S}) 中所有弧的容量之和,称为这个割集的容量,记为 $C(S, \bar{S})$ 。如图 8.5.3 中的网络 D 中,如果取 $S = \{v_1, v_3\}$, 则 $\bar{S} = \{v_2, v_4, v_5, v_6\}$, 于是割集 $(S, \bar{S}) = \{(v_1, v_2), (v_3, v_5), (v_3, v_4)\}$, 而 $C(S, \bar{S}) = C_{12} + C_{35} + C_{34} = 9 + 7 + 8 = 24$ 。

由于对 V 的分法不同,割集也就不同,对应的割集容量也不同,其中容量最小的割集称为最小割集。

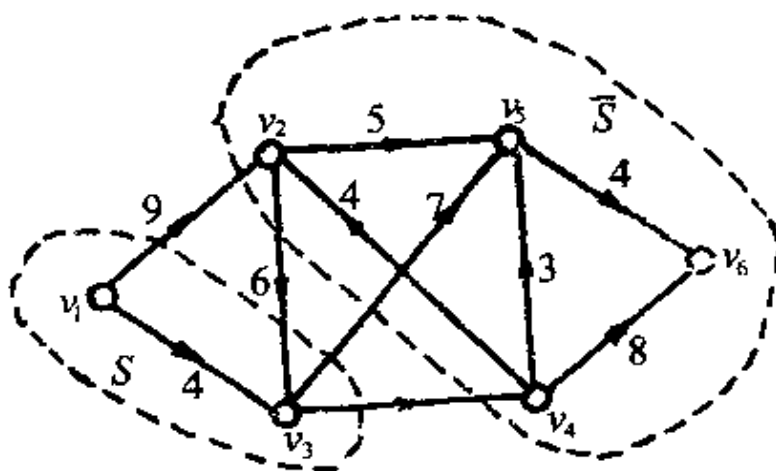


图 8.5.3

定理 8.5.1 (Ford - Fulkerson) 在任何网络中最大流的流值等于最小割集的容量。

(证明略)

在网络 D 中,设 P 是一条从 v_1 到 v_k 的无向路,规定路 P 的方向是从 v_1 到 v_k ,则在路 P 上并且与 P 的方向一致的弧,称为前向

弧,在路 P 上并与 P 的方向相反的弧称为后向弧。

定义 8.5.4 设 $\{f_{ij}\}$ 是 D 的一个可行流,如果存在一条从起点 v_1 到终点 v_p 的(无向)路 P ,满足:

- (1) 在 P 上的所有前向弧上, $f_{ij} < c_{ij}$ 。
- (2) 在 P 上的所有后向弧上 $f_{ij} > 0$

则称 P 是一条关于流 $\{f_{ij}\}$ 的可扩充路。

如图 8.5.4 中,可行流 $\{f_{ij}\}$ 用方框里的数字来表示,没有框起来的数字表示对应弧的容量。

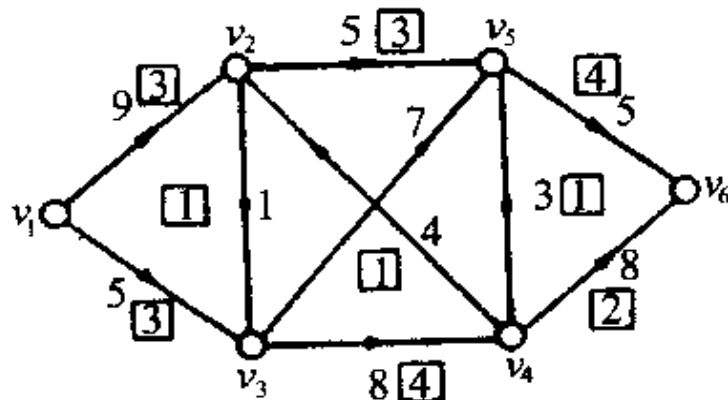


图 8.5.4

对于(无向)路 $P_1 = v_1 v_2 v_5 v_4 v_6$, (v_1, v_2) , (v_2, v_5) , (v_4, v_6) 为前向弧 (v_4, v_5) 为后向弧, 由于前向弧上 $f_{12} = 3 < 9 = C_{12}$ $f_{25} = 3 < 5 = C_{25}$ $f_{46} = 2 < 8 = C_{46}$, 同时后向弧上 $f_{45} = 1 > 0$ 。所以, P_1 是一条可扩充路。类似地, 可验证(无向)路 $P_2 = v_1 v_2 v_5 v_6$ 也是一条可扩充路。

对于网络 $D = (V, A)$ 的一个可行流 $\{f_{ij}\}$, 如果能找到可扩充路, 就可以把 $\{f_{ij}\}$ 调整为流值更大的另一可行流 $\{f'_{ij}\}$ 。方法如下:

令 $\epsilon_1 = \min \{C_{ij} - f_{ij} \mid (v_i, v_j) \text{ 是 } P \text{ 上的前向弧}\}$,

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

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

(如果 P 上无后向弧, 则令 $\varepsilon_2 = +\infty$)

求得调整量 $\varepsilon = \min\{\varepsilon_1, \varepsilon_2\} > 0$ 之后, 可定义流 $\{\bar{f}_{ij}\}$ 如下:

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

容易看出, $\{\bar{f}_{ij}\}$ 仍是可行流, 并且其流值比 $\{f_{ij}\}$ 的流值大 ε 。

例如, 在图 8.5.4 中, 对于可行流 $\{f_{ij}\}$ 我们前面已得到了可扩充路 $P_1 = v_1 v_2 v_5 v_4 v_6$, 可行流 $\{f_{ij}\}$ 的流值是 $v(f) = 6$ 。对于可扩充路 P_1 , 利用上述调整方法, 可得:

$$\varepsilon = \min\{9 - 3, 5 - 3, 8 - 2\} = 2 \quad \varepsilon_2 = \min\{1\} = 1$$

$$\text{所以 } \varepsilon = \min\{\varepsilon_1, \varepsilon_2\} = \min\{2, 1\} = 1$$

在 P_1 上的所有前向弧上的流均加上 1, 而后向弧上的流均减去 1, 其他弧上的流不变, 就得到流 $\{\bar{f}_{ij}\}$ (如图 8.5.5) 这时对应的流值 $v(\bar{f}) = 7$ 。

由此例可以看出
对于一个可行流 $\{f_{ij}\}$
可以通过求出网络中的可扩充路, 经调整, 就可求得一个新的可行流, 并且使流值增大。反复这样做下去, 直到网络中不存在可扩充路, 就求得了网

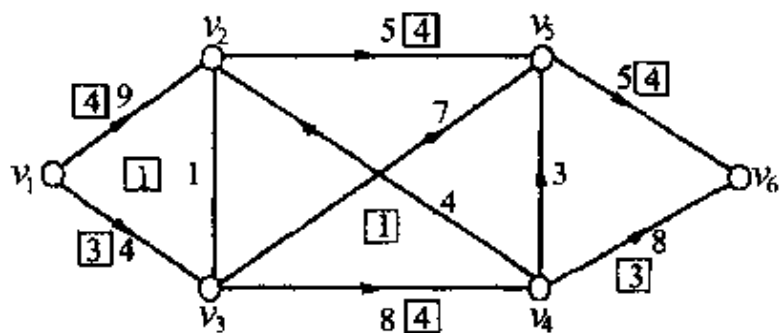


图 8.5.5

络的最大流。实际上, 有如下定理:

定理 8.5.2 设 $\{f_{ij}\}$ 是网络 D 的一个可行流并且对于 $\{f_{ij}\}$

来说,不存在可扩充路,则 $\{f_{ij}\}$ 是网络 D 的最大流(证明略)。

8.5.2 Ford - Fulkerson 标号法

下面介绍的标号法,就是一种系统地寻找可扩充路的方法。它是由 Ford - Fulkerson 于 1956 年提出的,步骤如下:

设已有初始可行流 $\{f_{ij}\}$ (也可以是零流)则

(1) 给起点 v_1 以标号 $(0, +)$, v_1 成为已标号但未检查的点:

(2) 按下面的方法对网络中的各顶点进行检查和标号,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 (i, j) ,如果有 $f_{ij} < c_{ij}$,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 为终点的所有的弧 (k, i) ,若 (k, i) 满足 $f_{ki} > 0$,且 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,在标号下面划一横线。

(3) 检查 v_p 是否已得到标号。如果未标号,则重复进行(2),如果 v_p 得到标号,就转步骤(4)

(4) 用“倒向追踪”的方法找出从 v_1 到 v_p 的可扩充路:设 v_p 的标号是 $(i, +)$,则可扩充路上 v_p 前面的点就是 v_i ,弧是 (i, p) 。一般,考虑点 v_i 。如果 v_i 的标号是 $(j, +)$,则 v_i 的前面的点是 v_j ,弧是 (j, i) ;如果 v_i 的标号是 $(j, -)$,则 v_i 的前一点是 v_j ,而弧是 (i, j) 。如此进行,直至追踪到出现标号 $(0, +)$ 的顶点 v_1 出现为止。这样,就得到一条从 v_1 到 v_p 的可扩充路,转步骤(5)。如果找不到 $\{f_{ij}\}$ 的可扩充路,则终止计算, $\{f_{ij}\}$ 已是最大流。

(5) 按前面所讲的方法进行调整,得到新的可行流 $\{f_{ij}\}$ 。抹去所有标号,重新返回(1)。

例 1 我们对图 8.5.6 中的可行流开始(一般可以从零流开始)计算。

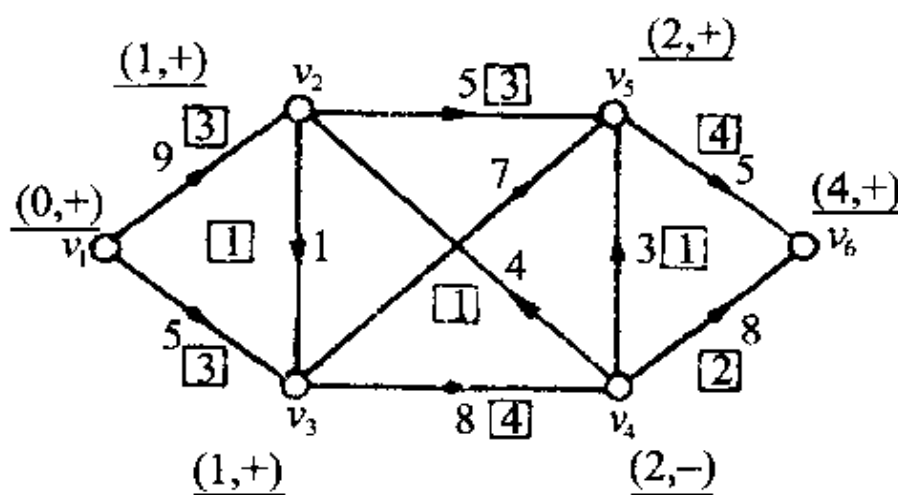


图 8.5.6

先给 v_1 标号 $(0, +)$ 对 v_1 进行检查, 得知 v_2, v_3 可得标号 $(1, +)$ 。 v_1 成为已检查过的点。取已标号而未检查的点 v_2 (或 v_3), 知 v_5 可得标号

$(2, +)$, 而 v_4 可得标号 $(2, -)$, v_2 成为已检查过的点。这时 v_3, v_4, v_5 都是已标号而未检查的点。

我们可对先得标号的点先检查, 于是检查 v_3 , 而知它不能再使其他点获得标号, v_3 也成为已检查点。再检查 v_4 (或 v_5)。这时 v_6 获得标号 $(4, +)$ 。由于终点已获得标号, 立刻“倒向追踪”。因 v_6 的标号 $(4, +)$, 得知可扩充路的最后一段是 $|v_4(4, 6), v_6|$ 。而 v_4 的标号是 $(2, -)$, 所以其前点是 v_2 , 而弧是 (v_4, v_2) 。于是可扩充路再往前是 $|v_2(4, 2)v_4(4, 6)v_6|$ 。 v_2 的标号是 $(1, +)$, 得知 v_2 的前点是 v_1 , 弧是 $(1, 2)$ 。而 v_1 的标号是 $(0, +)$, 所以求得可扩充路 $P = \{v_1(1, 2)v_2(4, 2)v_4(4, 6)v_6\}$

$$\text{而 } \epsilon_1 = \min\{9 - 3, 8 - 2\} = 6 \quad \epsilon_2 = \min\{1\} = 1$$

$$\text{所以 } \epsilon = \min\{\epsilon_1, \epsilon_2\} = 1$$

用前述方法进行调整, 得到新的可行流 (如图 8.5.7), 抹去旧的标号, 对新的可行流再用标号法找可扩充路 (图 8.5.7)。又可得可扩充路 $P = \{v_1(1, 2)v_2(2, 5)v_5(5, 6)v_6\}$, 调整量为 1。调整后的可行流如图 8.5.8。

再利用上述标号法, 又可得可扩充路 $P = \{v_1(1,3)v_3(3,4)v_4(4,6)v_6\}$ 调整量为 2, 调整后的可行流如图 8.5.9。

再重新标号, 又可得可扩充路: $P = \{v_1(1,2)v_2(2,5)v_5(4,5)v_4(4,6)v_6\}$ 取调整量 $\varepsilon = 1$, 调整后得图 8.5.10 所示的可行流。

再对此可行流进行标号, v_6 已不可能得到标号, 所以找不到可扩充路了, 这时图 8.5.10 中的可行流已是最大流, 其流值为 $v = 11$ 。

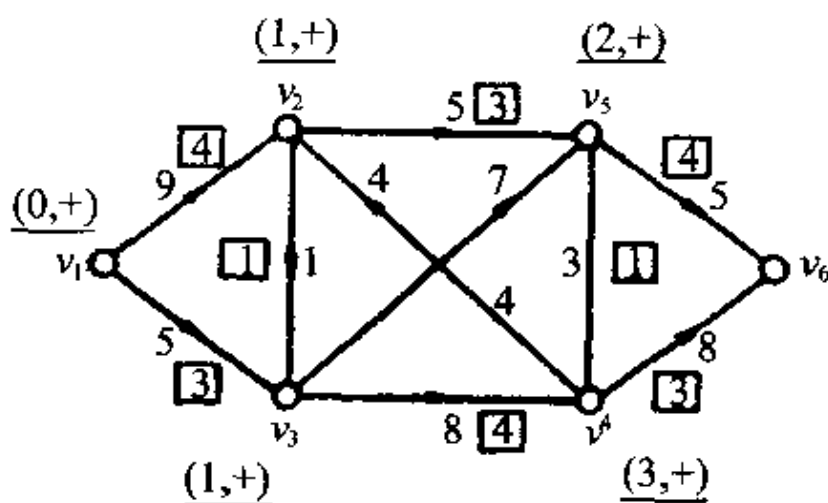


图 8.5.7

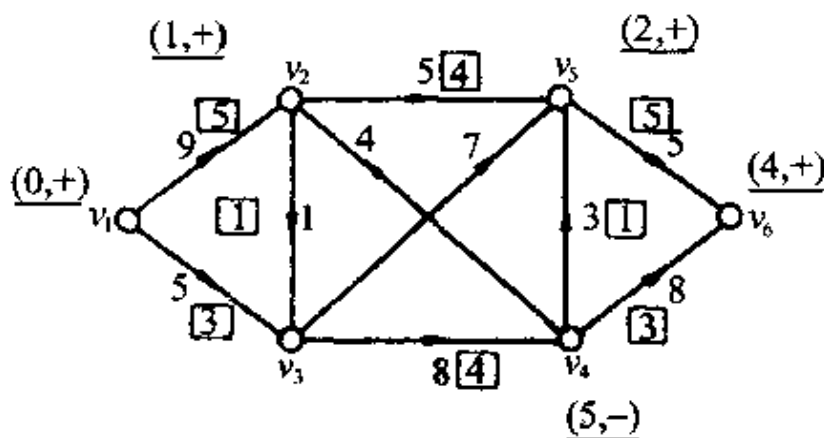


图 8.5.8

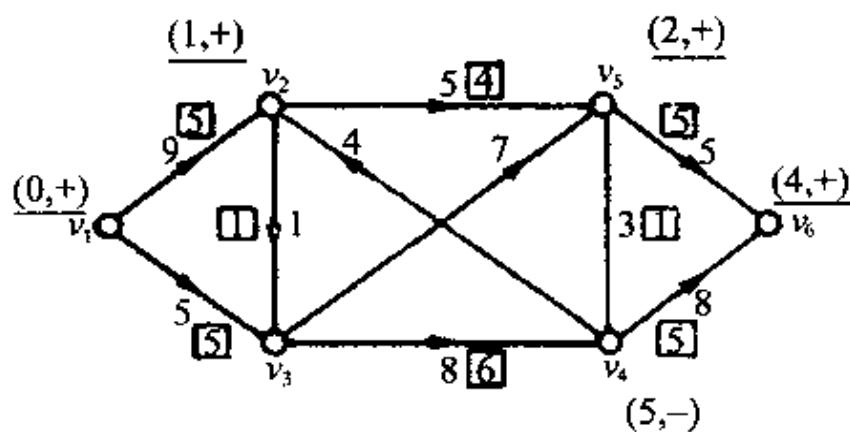


图 8.5.9

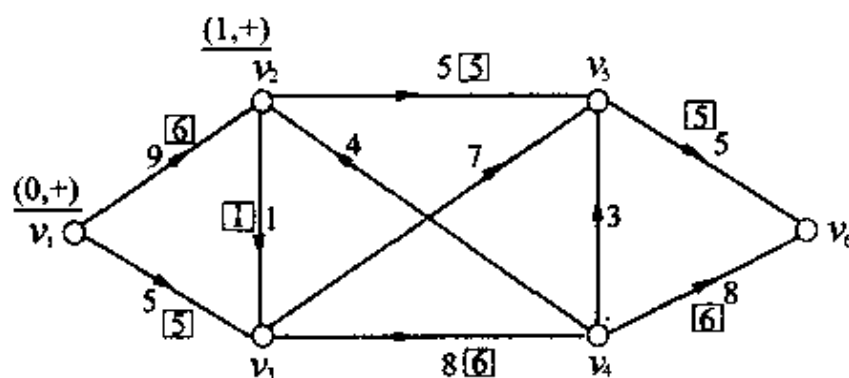


图 8.5.10

在计算过程中,我们应注意采用“先标号的先检验”的原则,从而保证以尽可能少的调整次数,求得网络最大流。

§ 8.6 匹配

8.6.1 最大匹配

定义 8.6.1 对于一个图 $G = (V, E)$, 若有边子集 $M \subset E$, 使 M 中任意两边 e', e'' 之间均无公共端点, 则称边子集 M 为图 G

的一个“匹配”。如图 8.6.1 中, (a) 表示图 G , (b) 中用双线所示的边子集 $M_1 = \{(v_2, v_7)(v_3, v_5)\}$ 为 G 的一个匹配; (c) 中用双线所示的边子集 $M_2 = \{(v_1, v_7)(v_3, v_4)(v_5, v_6)\}$ 为 G 的另一个匹配。

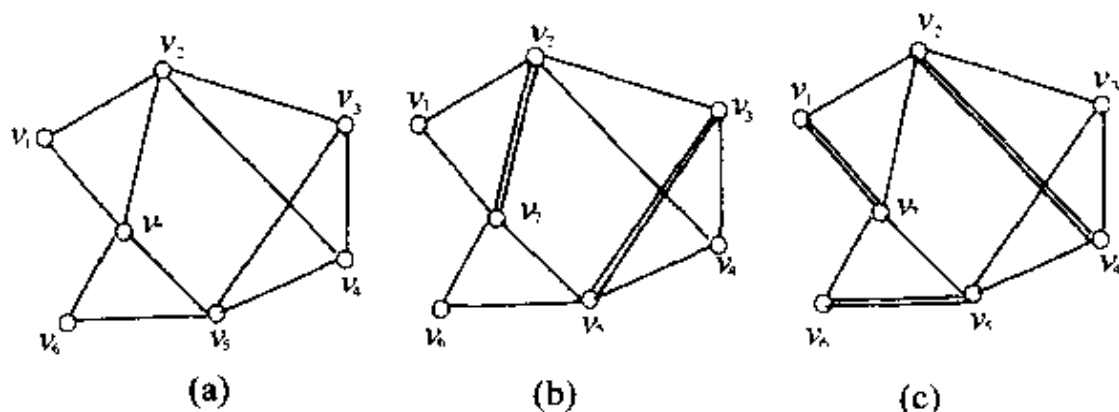


图 8.6.1

定义 8.6.2 设 M 是图 G 的一个匹配, 若对 G 中任意匹配 M' , 均有 M 中的边数大于或等于 M' 中的边数, 即 $|M| \geq |M'|$, 则称 M 是 G 的一个最大匹配。

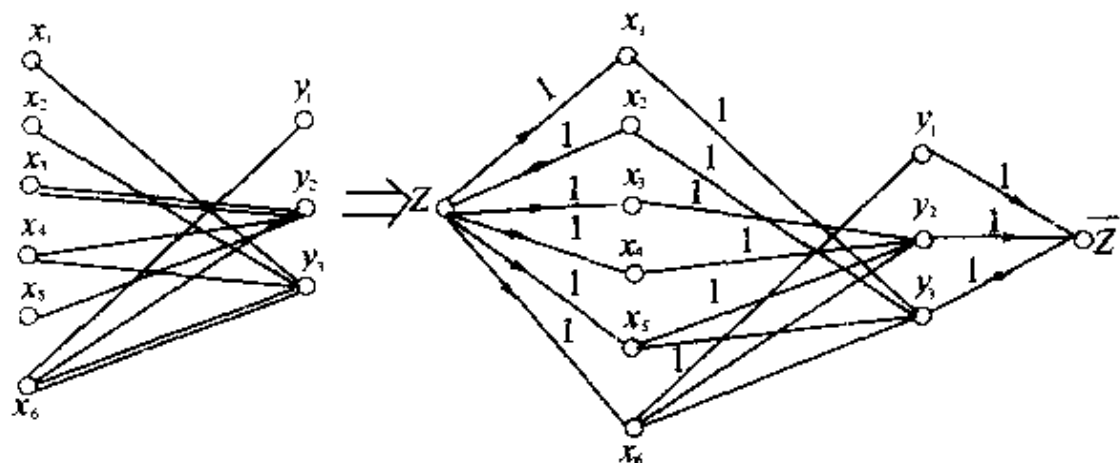


图 8.6.2

8.6.3

关于一个图上最大匹配的求法分两种情况,一是二分图,一是一般图,因为一般图上最大匹配的求法比较麻烦,所以这里不再介绍,有兴趣者可参阅图论的有关书籍。我们这里仅介绍二分图上最大匹配的求法。

二分图上的最大匹配是容易求出的。这种求法是以求网络图上的最大流为基础的。

对于一个二分图 $G = (x, y, E)$ 图 8.6.2, 我们如下构造一个有向赋数网络 D , 图 8.6.3

(1) 增加两个顶点: 发点 Z 和收点 \bar{Z} ;

(2) 从 Z 向 X 中每一点都分别连一条容量为 1 的弧, 从 Y 中每一点都分别向 \bar{Z} 连一条容量为 1 的弧; 把二分图中原有的边均分别换为由 X 中的点为始点; Y 中的点为终点的容量为 1 的弧。

二分图 G 与用上述方法构造的网络 D 之间有如下关系:

1. 如果 G 有一个由 P 条边组成的匹配 M , 那末 D 就有一个流 f , 它的值是 P ,

2. 若 G 的最大匹配包含 m 条边, D 的最大流的值是 u , 那末有 $m \geq u$ 。

定义 8.6.3 网络 D 上一个流 f , 如果对 D 上的任一弧 a , f 在 a 上只可取值 0 和 1, 那末就称 f 是 D 上的一个“0,1”——整数流。

3. 如果 D 上有一个值为 P 的 (0,1)——整数流 f , 那末 G 就有一个由 P 条边构成的匹配。

4. 若 G 的最大匹配包含 m 条边, D 的最大 (0,1)——整数流的值为 v , 则有 $m = v$ 。

由以上结论, 可得二分图的最大匹配的一个求法如下:

1° 从二分图 G 出发构造出其对应的网络 D 。

2° 求出 D 上的最大 (0,1)——整数流, 其值为 v 。

3° 如果组成 (0,1)——整数流的所有弧即无共同起点, 又无共同终点, 将这些弧去掉箭头后构成的边就组成了 G 的匹配

M , 且 M 就是所要求的 G 的最大匹配。

8.6.2 建模实例

1. 工作安排问题

要给 n 个工作人员安排 m 项工作, 已知每个工作人员能够胜任这 m 项工作中的一项或几项。现在要问, 如何分配才能使尽可能多的工作人员得到满意的工作?

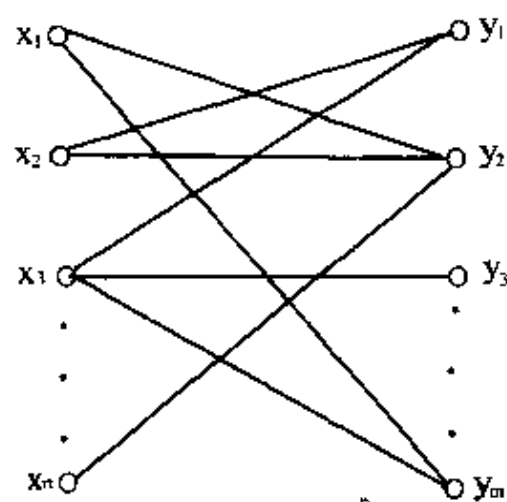


图 8.6.4

能多的工作人员得到满意的工作?

这个问题抽象成图的问题是: 用点集 $x = \{x_1, x_2, \dots, x_n\}$ 来代表 n 个工作人员, 用点集 $y = \{y_1, y_2, \dots, y_m\}$ 来代表 m 项工作。当工作人员 x_i 能胜任工作 y_j 时, 就把顶点 x_i 和 y_j 用一条边连接起来。这样就得到一个二分图 (图 8.6.4)。现在的问题就变成如何才能从所得的二分图中选出两两无公共点的尽可能多的边?

这个问题就是求一个二分图上的最大匹配问题, 用上述介绍的方法可以很容易解决。

2. 飞行员的搭配问题

第二次世界大战时, 欧洲许多飞行员从沦陷国应

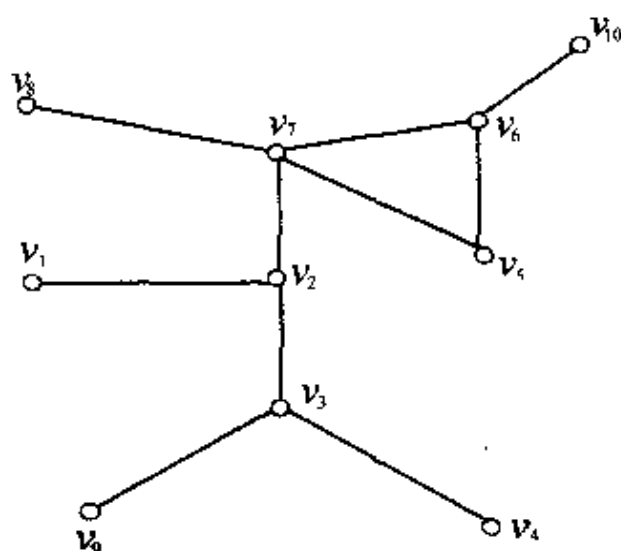


图 8.6.5

募来到英国加入英国皇家空军。由皇家空军送出的每一架飞机上都需要配备有航行技能与语言上能互相搭配的两名飞行员。皇家空军希望能同时送出尽可能多的飞机。

该问题抽象成如图8.6.5。图中每一顶点代表一个皇家飞行员。用一条边把任意两个顶点连接起来,这两个顶点代表可以在一起飞行的飞行员。送出一架飞机,就是从图中选出一条边,边两端的两个顶点所对应的两个飞行员同机飞出。于是上面的问题就变成了如何才能从图中选出两两无公共端点的尽可能多的边的问题?

这是一个求一般图上的最大匹配问题。

§ 8.7 边染色与顶点染色

图的边染色与图的顶点染色是图论中两个很令人感兴趣的问题,而且它们的应用也比较广泛。

8.7.1 基本知识

1. 边染色

定义 8.7.1 对于图 $G = (V, E)$ 的边着上色,使邻接的边色均互不相同,若用 K 种色可以作到这一点,则称 G 是 K -可边上色的。这种着色又叫做 K -正常边上色。能做到这一点的颜色最少数目 $q(G)$,叫做 G 的边色数。

一个图 $G = (V, E)$ 的 K -可边上色的,这就是说其边集 E 必有一种划分,使之成为 K 个边子集 M_1, M_2, \dots, M_k 。其中每一个边子集 M_i 都是 G 的一个匹配。 G 的边色数是 $q(G)$,这就是说, E 至少要划分为 $q(G)$ 个匹配。

例如图 8.7.1(a) 中 $q(G) = 4$, (b) 中 $q(G) = 5$, (c), $q(G) = 5$

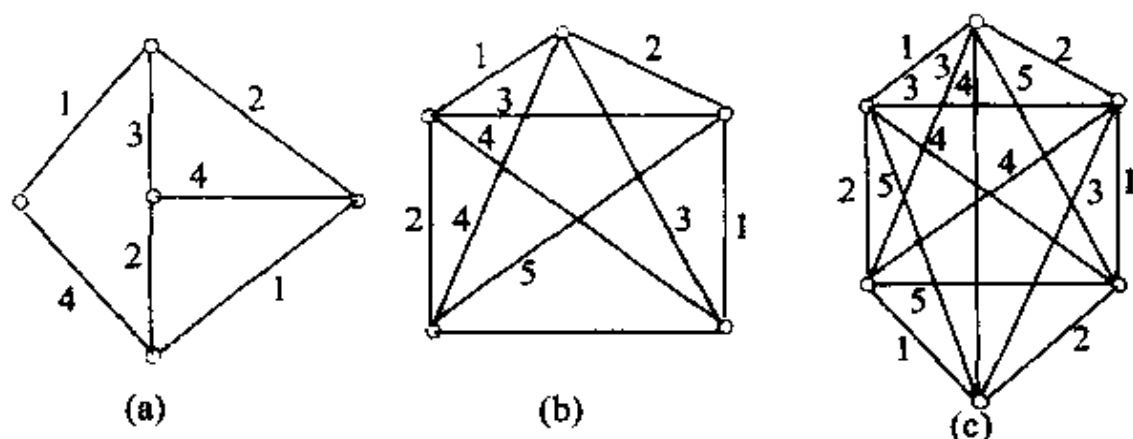


图 8.7.1

图中的数字表示颜色的种类。对图 G 的边集给出一个正常着色, 也就是说对 G 中任一顶点 v , 与其关联的边都着上了不同的颜色, 假如 $\Delta(G)$ 是 G 中顶点的最大次, 显然可知有 $q(G) \geq \Delta(G)$ 。

对于简单图若按其边色数来划分, 可分为两大类, 一类图 G , 其边色数满足 $q(G) = \Delta(G)$, 这类图叫做第一型图, 如上图 (C)。另一类图 G , 其边色数满足 $q(G) = \Delta(G) + 1$, 这类图叫做第二型图。如图 8.7.1(a), (b)

二分图 $G = (x, y, E)$ 为第一型图。

n 阶完全图 K_n , 当 n 为奇数时, 其为第二型图, 当 n 为偶数时, 其为第一型图。

2. 顶点染色

定义 8.7.2 对于图 $G = (V, E)$ 的顶点着上色, 使相邻的顶点色均互不相同, 若用 K 种色可以作到这一点, 则称 G 是 K -可点上色的。能做到这一点的颜色最少数目 $P(G)$, 叫做 G 的点色数。

一个图 $G = (V, E)$ 是 K -可点上色的, 就是说其点集 V 有一种划分, 使之成为 K 个顶点子集 V_1, V_2, \dots, V_k 。其中每一个点子

集 V_i 都是 G 的一个点独立集(即 V_i 中任何两点均不相邻)。 G 的点色数是 $P(G)$, 就是说 V 至少要划分为 $P(G)$ 个点独立集。

如图 8.7.2(a) 中 $P(G) = 3$, (b) 中 $P(G) = 5$

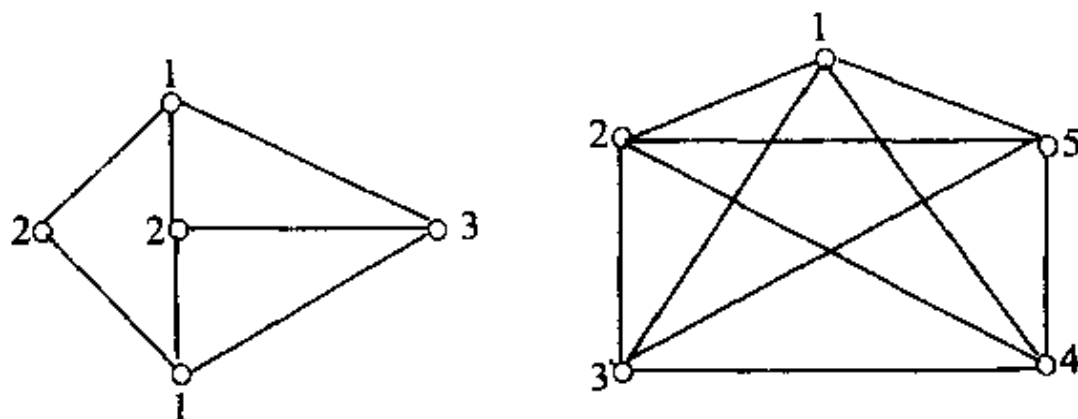


图 8.7.2

图中的数字表示顶点的颜色种类。

对于图的点色数, 有下面几个简单结果:

(1) 图 G 是空图(即只有孤立点)时, $P(G) = 1$

(2) n 个顶点的完全图 K_n 有 $P(K_n) = n$

(3) n 个顶点的圈 C_n , 有 $P(C_n) = \begin{cases} 2 & n \text{ 为偶数} \\ 3 & n \text{ 为奇数} \end{cases}$

(4) 若连通图 G 中无奇圈(由奇数条边组成的圈), 则有 $P(G) = 2$

若图 G 的点色数为 2, 则 G 的点集可划分为两个点独立集, 这就是说 G 必为二分图。

一个树 T , 当其顶点数大于或等于 2 时, 其点色数 $P(T)$ 等于 2。但反之, 一个点色数为 2 的图, 却不一定是树。

8.7.2 建模实例

1. 交易会的时间安排问题

有一个交易会, 有 n 家企业参加, 其中每一家企业都希望与其

他一些企业进行商谈,假设每次商谈持续一天,并且每次商谈都恰有两家参加,问这次交易会至少多少天才能结束?

我们取 $v_1 v_2 \cdots v_n$ 分别代表这 n 家企业,假若企业 v_i 与企业 v_j 要进行商谈,那末就取一条边 $[v_i, v_j]$,这样就得到一个 n 阶的完全图 $G = [V, E]$ 。然后再对 G 的边着色,使邻接的边色均不相同。这样可得到 G 的边色数 $q(G)$,这就是交易会结束的最少天数。

2. 选修课考试的安排问题

某个学校设有几门选修课,学期末要进行考试,当然每个学生每场只能参加一门课的考试。试问这次考试最少要进行几场?

显然没有共同的学生选修的课可以同时进行考试。

令 $v_1 v_2 \cdots v_n$ 分别代表 n 门课程,若课 v_i 与 v_j 有学生同时选修,则在其间连一条边。这样就得到一个简单图 $G = (V, E)$ 。现在我们对 G 的顶点着色,使相邻的点着上的色互不相同。这样就得到图 G 的点色数 $P(G)$ 。这就是要进行的考试的最少场次。

§ 8.8 最小复盖、最小控制集、极大独立集

8.8.1 基本知识

1. 最小复盖

定义 8.8.1 若图 G 的每条边都至少有一个端点在顶点集 V 的一个子集 K 之中,则 K 称为 G 的一个复盖。含顶点个数最少的复盖称为最小复盖。最小复盖中顶点的个数称为复盖数。

$$K_1 = \{v_1 v_2 v_3 v_4\}$$

$$K_2 = \{v_1 v_2 v_3 v_5\}$$

$$K_3 = \{v_2 v_3 v_4\}$$

$$K_4 = \{v_1 v_3 v_5\}$$

$K_5 = \{v_2 v_4 v_5\}$ 等都是其复盖, 其中 K_3, K_4, K_5 都是其最小复盖。其复盖数为 3。由此可知一个图的最小复盖不是惟一的。

因为关联矩阵表示的是顶点与边之间的关系, 所以关联矩阵与复盖密切相关。

顶点集 V 的子集 K 是图 G 的一个复盖, 当且仅当 G 的关联矩阵 M 中 K 的各顶点所对应的行内, 每列至少存在一个元素 1。

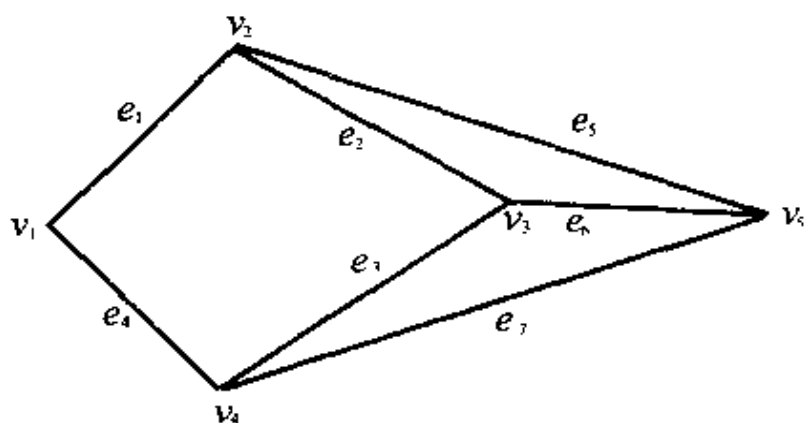


图 8.8.1

从图的关联矩阵可以找出一个最小复盖。如图 8.8.1 的关联矩阵为

$$M = \begin{matrix} & e_1 & e_2 & e_3 & e_4 & e_5 & e_6 & e_7 \\ \begin{matrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \end{matrix} & \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix} \end{matrix}$$

求 M 的最小复盖的具体步骤如下:

(1) 在 M 中先找出含 1 最多的行, (若有几行相同随便取其中之一), 如 v_3 行, 令 $v_3 \in K$, 划去 v_3 行及 v_3 行中元素为 1 所在的

列 e_2, e_3, e_6 得 M_1

$$M_1 = \begin{matrix} & e_1 & e_4 & e_5 & e_7 \\ \begin{matrix} v_1 \\ v_2 \\ v_4 \\ v_5 \end{matrix} & \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix} \end{matrix}$$

$$M_2 = \begin{matrix} & e_1 & e_4 \\ \begin{matrix} v_1 \\ v_2 \\ v_4 \end{matrix} & \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \end{matrix}$$

(2) 在 M_1 中重复(1)的过程, 如 v_5 行, 令 $v_5 \in K$, 划去 v_5 行及 v_5 行中元素为 1 所在的列 e_5, e_7 得 M_2

(3) 因为 $v_1 > v_2, v_1 > v_4$ (若对所有的 $j, r_{kj} = 1 \Rightarrow r_{ij} = 1$ 记 $v_i > v_k$) 划去 v_2, v_4 行。令 $v_1 \in K$, 过程结束得最小复盖 $K = \{v_1, v_3, v_5\}$

如果满足上述条件选取其他顶点, 就可以得到另一个最小复盖。

2. 最小控制集

定义 8.8.2 若图 G 的每个顶点, 或者直接属于顶点集 V 的某个子集 C , 或者它的邻边的另一端点属于 C , 则 C 称为 G 的控制集。含顶点个数最少的控制集称为最小控制集, 最小控制集中顶点的个数称为控制数。

如图 8.8.1 中 $(v_1, v_3, v_5), (v_1, v_5), (v_2, v_3)$ 等都该图的控制集, 其中 $(v_1, v_5), (v_2, v_3)$ 是最小控制集, 显然一个图的最小控制集也不是惟一的。

由定义可知, 控制集是研究图中顶点与顶点之间的关系的。图中的邻接矩阵表示的是顶点之间的联系, 所以邻接矩阵与它的控制集之间有密切联系, 一个图的最小控制集可以通过其邻接矩阵求取。

如图 8.8.1, 其邻接矩阵为

$$A = \begin{matrix} & v_1 & v_2 & v_3 & v_4 & v_5 \\ \begin{matrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \end{matrix} & \begin{pmatrix} 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 \end{pmatrix} \end{matrix}$$

求 A 的最小控制集的具体步骤如下:

(1) 先由 A 作出 $A + I$ 设为 A_1 (其中 I 为与 A 同阶单位矩阵)。

$$A_1 = \begin{matrix} & v_1 & v_2 & v_3 & v_4 & v_5 \\ \begin{matrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \end{matrix} & \begin{pmatrix} 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \end{pmatrix} \end{matrix}$$

(2) 从 A_1 选取含 1 最多的行 (若几行相同, 随便选其一) 比如 v_2 , 令 $v_2 \in C$ 从 A_1 划去 v_2 行及 v_2 行中元素 1 所在的 v_1, v_2, v_3, v_5 列得 A_2 。

$$A_2 = \begin{matrix} & v_4 \\ \begin{matrix} v_1 \\ v_3 \\ v_4 \\ v_5 \end{matrix} & \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \end{matrix}$$

(3) 从 A_2 中选取 v_4 令 $v_4 \in C$ 则 $(v_2 v_4) = C$ 即为所求的最小控制集。

如果满足上述条件选取其他顶点, 就可以得到另一些最小控

制集。

3. 极大独立集

定义 8.8.3 在图 G 中, 不包含相邻顶点的 V 的子集 S 称为独立集。如果在 S 中添加任何顶点都会使 S 不再是独立集, 那么 S 称为极大独立集。如图 8.8.2。

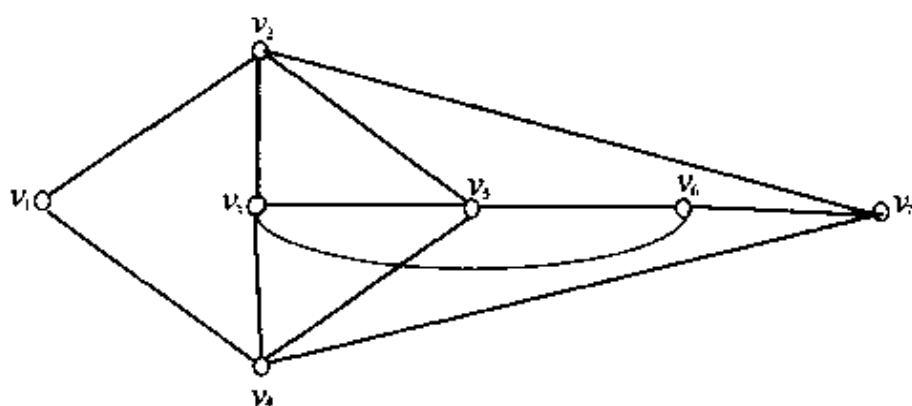


图 8.8.2

$(v_1), (v_1 v_3), (v_1 v_5 v_7)$ 都是独立集, 其中 $(v_1 v_5 v_7)$ 是极大独立集。还可以找到其他的极大独立集, 如 $(v_1 v_6), (v_1 v_3 v_7), (v_2 v_4 v_6)$ 等。

独立集与复盖之间有密切关系。用 $\bar{S} (\subset V)$ 表示 S 对 V 的补集, 有如下结论

若 S 是独立集, 则 \bar{S} 是复盖, 反之亦然。如图 8.8.2 中, $S = \{v_1 v_5\}$ 是独立集, $\bar{S} = \{v_2 v_3 v_4 v_6 v_7\}$ 是复盖。

与极大独立集形成这种互补关系的是极小复盖。如果从复盖 K 中去掉任何顶点都会使 K 不再是复盖的话, 那么 K 称为极小复盖。如图 8.8.2 中 $S = (v_1 v_5 v_7)$ 是极大独立集, $\bar{S} = \{v_2 v_3 v_4 v_6\}$ 是极小复盖。这样, 求全部极大独立集又归结为找出所有的极小复盖。

求所有极小复盖有一种逻辑算法, 这种算法的基础是建立在

极小复盖如下性质上。

当且仅当对于每个顶点 v , 或者 v 属于 K , 或者 v 的所有相邻顶点属于 K , 并且二者不能同时成立时, K 是极小复盖。

这种算法的程序是: 对于每个顶点 v , 选择 v 或者选择 v 的所有相邻顶点。

利用逻辑代数方法可以有效地执行上述程序, 逻辑代数中的“和”(+) 运算和“积”(·) 运算分别相当于集合中的“并”(∪) 运算和“交”(∩) 运算。

用这种方法求图 8.8.2 的全部极小复盖如下:

$$\begin{aligned}& (v_1 + v_2v_4)(v_2 + v_1v_3v_5v_7)(v_3 + v_2v_4v_5v_6)(v_4 + v_1v_3v_5v_7) \\& (v_5 + v_2v_3v_4v_6)(v_6 + v_3v_5v_7)(v_7 + v_2v_4v_6) \\& = (v_1v_2 + v_2v_4 + v_1v_3v_5v_7)(v_3 + v_2v_4v_5v_6) \cdots \cdots \\& = (v_1v_2v_3 + v_2v_3v_4 + v_1v_3v_5v_7 + v_2v_4v_5v_6)(v_4 + v_1v_3v_5v_7) \\& \quad \cdots \cdots \\& = v_1v_3v_5v_7 + v_2v_4v_5v_6 + v_2v_3v_4v_6 + v_2v_3v_4v_5v_7\end{aligned}$$

所以得图 8.8.2 的四个极小复盖, 而其所有的极大独立集即 $(v_2v_4v_6), (v_1v_3v_7), (v_1v_5v_7), (v_1v_6)$ 。

找出全部极大独立集也是解决顶点染色问题的关键。在所有的极大独立集中, 含顶点数目最多的独立集, 称为最大独立集, 其中的顶点数目称为独立数。

现在用极大独立集对图 8.8.2 进行顶点染色, 任取一最大独立集, 如 $S_1 = (v_2v_4v_6)$, 给它染上第一种颜色, 再利用上述方法找出 $G \setminus S_1$ 的全部极大独立集, 可见为 $(v_1v_3v_7), (v_1v_5v_7)$, 取其中之一, 如 $S_2 = (v_1v_3v_7)$, 再染上第二种颜色, 于是在 $G \setminus S_1 \setminus S_2$ 中就只有 v_5 了。给 v_5 染第三种颜色, 可见图 8.8.2 的顶点色数为 3。

8.8.2 建模实例

1. 消防设施的安置问题

某城镇有若干条街道构成居民小区。如图 8.8.1 所示,其中 e_1, e_2, \dots, e_7 表示街道, v_1, v_2, \dots, v_5 表示交叉路口。现计划在某些路口安置消防设施,只有与路口直接相连的街道才能使用它们。为使所有街道必要时都有消防设施可用,在哪些路口安置设施才最节省呢?

所谓最节省就是用最少的设施达到目的,这就是求图 8.8.1 的一个最小复盖问题。由上述介绍可知,最少要在三处安置消防设施。

2. 监狱看守问题

一座监狱的几间牢房有道路相连,不妨如图 8.8.1 所示,其中 v_1, v_2, \dots, v_5 表示牢房, e_1, e_2, \dots, e_7 表示道路。监狱看守要设在通过道路能直接监视所有牢室的地方,如果看守不得走动,那么他们应呆在哪些牢房(即路口)所在地,问至少需要几名看守才能完成监视任务?

该问题就是求图中的一个最小控制集的问题。由上述介绍可知最少需要二名看守。

3. 化学制品的存放问题

某家公司生产若干种化学制品,其中某些制品是互不相容的,如果存放在一起,则可能发生化学反应,引起危险。因此公司必须把仓库分成相互隔离的若干区,以便把不相容的制品分开存放,问至少要划分多少小区,怎样存放才能保证安全?

假设有 7 种化学制品,用 $v_1 v_2 v_3 v_4 v_5 v_6 v_7$ 表示,其中不能存放在一起的是 $(v_1 v_2), (v_1 v_4), (v_2 v_3), (v_2 v_5), (v_2 v_7), (v_3 v_4), (v_3 v_5), (v_3 v_6), (v_4 v_5), (v_4 v_7), (v_5 v_6), (v_6 v_7)$, 用图来表示这些制品和它们之间的关系很方便。把化学制品作为图的顶点,把不能放在一起的制品之间连成一条边。上述问题就可以用图 8.8.2 来表示。我们用顶点染色的方法就可以解决这个问题。由上边的介绍,已经较好

的解决了这个问题,只要分成三个小区即可安全存放。

用图论方法建模是数学建模的一种重要方法,它可以解决相当广泛的一些实际问题。这种方法的关键就是将研究的对象用图的两个主要元素顶点和边表示出来,形成一个图,然后再将要解决的问题归结为图论中的某些概念。

习 题

1. 九名数学家在一次国际会议上相遇,发现他们中的任意三个人中至少有两个人可以用同一种语言对话。如果每个数学家至多会说三种语言,证明至少有三个数学家可以用同一种语言对话。

2. 有八种化学药品 A 、 B 、 C 、 D 、 P 、 R 、 S 、 T 要放进贮藏室保管。出于安全原因,下列各组药品不能贮在同一室内: $A-R$, $A-C$, $A-T$, $R-P$, $P-S$, $S-T$, $T-B$, $B-D$, $D-C$, $R-S$, $R-B$, $P-D$, $S-C$, $S-D$ 问贮藏这八种药品至少需多少房间。

3. 有四个立方体,在各自的六个面上分别涂上红、白、兰、绿四种颜色,如把四个立方体一个压一个堆起来,如何做到使堆成的四侧每一侧面恰好包含红、白、兰、绿四种颜色。已知四个立方体四个面颜色如下表所示

立方体	北	南	东	西	上	下
1	红	红	红	白	白	白
2	兰	白	兰	绿	绿	绿
3	红	白	白	绿	红	兰
4	兰	白	兰	绿	兰	绿

4. 十名研究生参加六门课程的考试。由于选修内容不同,考试门数也不一样,下表中给出了每个研究生应参加考试的课程(打

※的)

考 试 课 程 研 究 生	A	B	C	D	E	F
1	※	※		※		
2	※		※			
3	※					※
4		※			※	※
5	※		※	※		
6			※		※	※
7			※		※	
8		※		※		
9	※	※				※
10	※		※			※

规定考试在三天内结束,每天上、下午各安排一门。研究生提出希望每人每天最多考一门,又课程 A 必须安排在第一天上午考,课程 F 安排在最后一天上午考,课程 B 只能安排在下午考,试列一张满足各方面要求的考试日程表。

5. 有下图表示某两个生产队的水稻田,用堤埂分隔成很多小块。为了用水灌溉,需要挖开一些堤埂。问两个队各最少挖开多少堤埂,才能使水浇灌到每小块稻田。

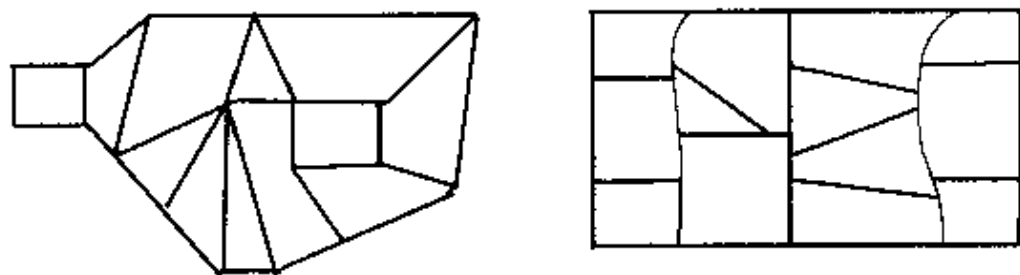


图 8.8.3

6. 某公司在六个城市 C_1, C_2, \dots, C_6 中有分公司,从 C_i 到 C_j 的直

接航程票价记在下述矩阵中的 (i, j) 位置上。 (x) 表示无直接航路), 请帮助该公司设计一张任意两城市间的票价最便宜的路线表。

$$\begin{bmatrix} 0 & 50 & x & 40 & 25 & 10 \\ 50 & 0 & 15 & 20 & x & 25 \\ x & 15 & 0 & 10 & 20 & x \\ 40 & 20 & 10 & 0 & 10 & 25 \\ 25 & x & 20 & 10 & 0 & 55 \\ 10 & 25 & x & 25 & 55 & 0 \end{bmatrix}$$

7. 如下图所示。一个人从 C_3 骑自行车出发去图中各处送紧急文件, 然后回到 C_3 , 试帮助设计一条最短路线。如图中数字单位为百米, 自行车速度 15 公里 / 时, 送文件时每处耽误 5 分钟, 问从出发算起半小时内该人能否回到出发地点?

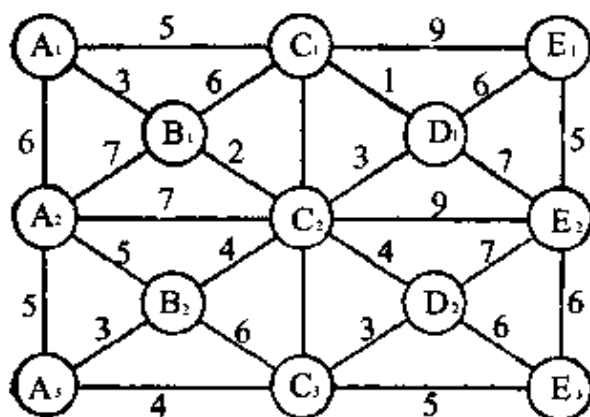


图 8.8.4

8. 图 8.8.5 中 A, B, C, D, E, F 分别表示陆地和岛屿, ①, ②……⑭表示桥梁及其编号。若河两岸分别为互为敌对的双方部队占领, 问至少应切断几座桥梁(具体指出编号)才能达到阻止对方部队过河的目的。试用图论方法进行分析。

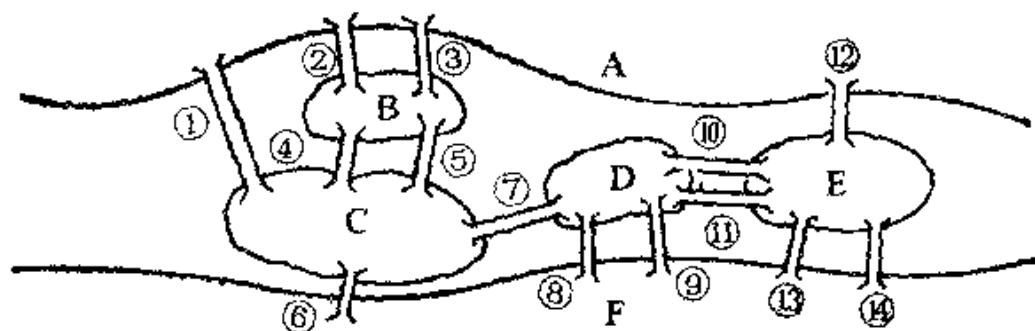


图 8.8.5

9. 图 8.8.6 中 A、B、C、D、E、F 分别代表岛和陆地, 它们之间有桥相连, 问一个人能否从某地出发经过图中的每座桥恰好一次又回到原出发地?

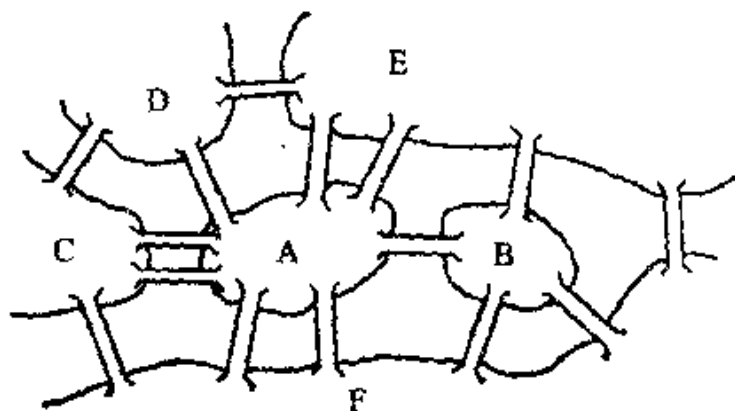


图 8.8.6

10. 在甲、乙两城之间有一条繁忙的双轨铁路, 运行旅客快车和货车。客车时间是预先排定的, 运行上优先于货车。当有一列客车按排定时间在两个运行站之间会超越货车时, 货车应在原站等候。为了增加货运业务, 需要排定货车运行时刻表, 使每天可以发出的列车数最大而不干扰客车行车时刻表。

现研究从甲城到乙城方向线路,设区间内有 S 个车站,每个车站从甲城到乙城方向有 $n_i (i = 1, 2 \cdots S)$ 条线路,货车从 i 站到 $(i + 1)$ 站间运行需 t_i 个时间单位(以 0.1 小时为一个时间单位,不是一个单位时取整),其中 t_0 为从甲城到第 1 站运行时间单位, t_S 为从第 S 站到乙城运行的时间单位。相继各次货车必须保持 0.1 小时的排定时差,一列货车只有在到达下一站 $(i + 1)$ 站之前不会被一列排定的客车追上时才准许在时刻 $j (j = 0.0, 0.1 \cdots 23.9)$ 驶过或离开第 i 站。又一列货车如在被一列客车追上之前在随后将到达的各站上都无线路可停车,也只好停留在原车站。试将这个问题转化为一个求网络最大流的问题。

第九章 量纲分析法建模

量纲分析法是在物理领域中建立数学模型的一种重要方法。其根据是一个物理定律的数学描述,等式两端必须保持量纲一致,或称量纲齐次性。因此,所谓量纲分析法,就是利用量纲齐次原则,来寻求各物理量之间关系的一种方法。用这种方法可从单一的前提条件对某一物理现象推论出许多关于所涉及的物理量的相关信息,有时,在某些复杂现象的开始研究阶段,量纲分析法是惟一可行的理论方法。所以,该方法对物理问题尤其是复杂问题的数学建模,具有十分重要的意义。

本章首先介绍量纲分析法的基本知识,第二节列举几个量纲分析法建模的实例,给出该方法建模的一般步骤,并于第三节指出应用量纲分析法建立数学模型时的注意事项及其局限性。本章的定理证明中用到了线性代数知识,但定理证明并非本章重点。

§ 9.1 预备知识

9.1.1 量纲

在物理问题中,各物理量之间均存在着规律性的联系,因此,我们可以选定几个物理量作为基本量,通过定义或定律导出其他各物理量。由基本量导出的量称为导出量。

导出量由基本量导出,也就是可以用基本量的某种组合来表示。这种反映某物理量与基本量之间幂次关系的公式称为该物理量的量纲。基本量的量纲称为基本量纲,导出量的量纲称为导出量纲。

在研究力学问题时,我们常把长度 l 、质量 m 、时间 t 作为基本量,其量纲分别用 L 、 M 、 T 表示,则其他的各物理量的量纲就可用这三个字母的某种组合表示,例如,速度 v 、加速度 a 的量纲按照其定义分别用 LT^{-1} 和 LT^{-2} 表示,力 f 的量纲根据牛顿第二定律用质量和加速度量纲的乘积 LMT^{-2} 表示。也有的物理量是无量纲的,即纯数字,例如,角度(弧度)的量纲为 $LL^{-1} = L^{\circ} = 1$,即为无量纲量。

某些物理常数是有量纲的,如万有引力定律 $f = k \frac{m_1 m_2}{r^2}$ 中的引力常数 k ,由 $k = \frac{fr^2}{m_1 m_2}$ 可知,其量纲应从力 f 、距离 r 和质量 m 的量纲求出,为 $LMT^{-2} \cdot L^2 \cdot M^{-2} = L^3 M^{-1} T^{-2}$ 。再如一个关于运动物体受到空气阻力的模型,假设力 f 的大小与速度的平方成正比, $f = \mu v^2$,由 $\mu = \frac{f}{v^2}$,导出系数 μ 的量纲为 $LMT^{-2}(LT^{-1})^{-2} = ML^{-1}$ 。另外,如果在表达式中含有 $\exp(at)$ 或 $\sin(at)$, t 表示时间,则参数 a 就有量纲 T^{-1} ,以保证 at 是一个无量纲数。但有些常数是无量纲的,如 $\pi = 3.14159$ 等。

通常,一个物理量 q 的量纲记为 $[q]$,于是前述各物理量的量纲记为

$$[l] = L, [m] = M, [t] = T, [v] = LT^{-1}$$

$$[a] = LT^{-2}, [f] = LMT^{-2},$$

对于无量纲量,记为 α ,其量纲 $[\alpha] = 1$ (可视为 $[\alpha] = L^{\circ} M^{\circ} T^{\circ}$)。

如果方程中含有导数运算,则导数的量纲即为相应物理量的量纲之比。例如, p 表示流体中任一点的压强,压强在 Z 方向上的梯度为 $\frac{dp}{dz}$,则

$$\left[\frac{dp}{dz}\right] = \frac{[p]}{[z]} = \frac{[ML^{-1}T^{-2}]}{[L]} = ML^{-2}T^{-2}$$

9.1.2 单位制与物理定律

为用数量表示物理量的大小,必须先规定一个计量标准,用此标准与被测物理量进行比较,所得倍数就是被测物理量的数值,这个计量标准叫做单位.只要规定了几个基本量的单位后,就可根据物理公式得到其他物理量的单位,这几个基本量的单位称做基本单位.由基本单位通过数值系数为1的只包含乘、除运算的代数式导出的其他物理量的单位叫导出单位.选择不同的基本单位,可得不同的导出单位,从而形成不同的单位制.如在力学中,以厘米、克、秒作为基本量长度、质量、时间的单位时,称厘米—克—秒制;以米、千克、秒作为基本单位时,秒米—千克—秒制。

每个量纲都可以有几种不同的单位,如长度量纲有米、厘米等,质量量纲有千克、克等,速度量纲可以有每秒米或每小时公里等。用量纲可以定出同一物理量不同单位制之间的换算因素。例如力的量纲为 MLT^{-2} , 力的单位从米—千克—秒制换算为厘米—克—秒制时,可以写成:

$$\begin{aligned} 1 \text{ 牛顿} &= 1 \text{ 千克} \cdot \text{米} / \text{秒}^2 = 1000 \text{ 克} \times 100 \text{ 厘米} / \text{秒}^2 \\ &= 10^5 \text{ 克} \cdot \text{厘米} / \text{秒}^2 = 10^5 \text{ 达因} \end{aligned}$$

一般地,设 X_1, X_2, \dots, X_m 是一组基本量纲,相应各基本量在原单位制下取值 x_1, x_2, \dots, x_m , 在新单位制下取值将变为 $\bar{x}_1 = \lambda_1 x_1, \bar{x}_2 = \lambda_2 x_2, \dots, \bar{x}_m = \lambda_m x_m$, 其中 $\lambda_1, \lambda_2, \dots, \lambda_m > 0$ 为一组常数。若某物理量 q 的导出量纲为 $\{q\} = \prod_{i=1}^m X_i^{C_i}$, 则在原单位制下取值 q 的物理量在新单位制下的取值将为 $\bar{q} = \prod_{i=1}^m \lambda_i^{C_i} \cdot q$ 。

由理论或直接由实验建立的物理定律,都是表征所研究现象的诸量之间的函数关系的。这些有量纲的物理量的数值,依赖于单位制的选取,而单位制是与现象的本质无关的。因此,描述物理定律的数学公式,即各物理量之间的数量关系通常是与单位制的选取无关的,具体地讲,这种无关性即是上述的物理量 q 与 \bar{q} 满足相

同的物理定律。对此,有正式定义如下:

设 q_1, q_2, \dots, q_n 满足物理定律

$$\varphi(q_1, q_2, \dots, q_n) = 0 \quad (9.1.1)$$

X_1, X_2, \dots, X_m 是一组基本量纲, q_1, q_2, \dots, q_n 的量纲可表示为:

$$[q_j] = \prod_{i=1}^m X_i^{C_{ij}}, j = 1, 2, \dots, n, \quad (9.1.2)$$

若对于任意常数 $\lambda_1, \lambda_2, \dots, \lambda_m > 0$, 令

$$\bar{q}_j = \prod_{i=1}^m \lambda_i^{C_{ij}} q_j, j = 1, 2, \dots, n \quad (9.1.3)$$

而

$$\varphi(\bar{q}_1, \bar{q}_2, \dots, \bar{q}_n) = 0 \quad (9.1.4)$$

与(9.1.1)式等价,则称物理定律(9.1.1)与量纲 X_1, X_2, \dots, X_m 的单位的选取无关。

9.1.3 量纲相互独立的充要条件

设有 m 个物理量 q_1, q_2, \dots, q_m , 对于无量纲量 α , 任取一组数 k_1, k_2, \dots, k_m , 若

$$q_1^{k_1} \cdot q_2^{k_2} \cdots q_m^{k_m} = \alpha \quad (9.1.5)$$

必有 $k_1 = k_2 = \dots = k_m = 0$, 则称这 m 个物理量的量纲相互独立。即, 若几个物理量的量纲是相互独立的, 则无法用这些物理量的非零幂次的乘积组成无量纲量。显然, 由于基本量的非零幂次的乘积无法组成无量纲量, 基本量纲一定是相互独立的。

定理 9.1.1 设 X_1, X_2, \dots, X_m 是一组基本量纲, 另有 m 个物理量 q_1, q_2, \dots, q_m 的导出量纲为:

$$[q_j] = \prod_{i=1}^m X_i^{C_{ij}}, j = 1, 2, \dots, m \quad (9.1.6)$$

则 $[q_1], [q_2], \dots, [q_m]$ 相互独立的充要条件是:

$$\begin{vmatrix} C_{11} & C_{12} & \cdots & C_{1m} \\ C_{21} & C_{22} & \cdots & C_{2m} \\ \vdots & \vdots & & \vdots \\ C_{m1} & C_{m2} & \cdots & C_{mm} \end{vmatrix} \neq 0 \quad (9.1.7)$$

证明如下:

任取一组数 k_1, k_2, \dots, k_m , 则乘积 $q_1^{k_1} \cdot q_2^{k_2} \cdots q_m^{k_m}$ 的量纲式为:

$$\begin{aligned} [q_1]^{k_1} [q_2]^{k_2} \cdots [q_m]^{k_m} &= \left(\prod_{i=1}^m X_i^{C_{i1}} \right)^{k_1} \left(\prod_{i=1}^m X_i^{C_{i2}} \right)^{k_2} \cdots \left(\prod_{i=1}^m X_i^{C_{im}} \right)^{k_m} \\ &= X_1^{\sum_{j=1}^m C_{1j} k_j} \cdot X_2^{\sum_{j=1}^m C_{2j} k_j} \cdots X_m^{\sum_{j=1}^m C_{mj} k_j} \end{aligned} \quad (9.1.8)$$

若 $q_1^{k_1} \cdot q_2^{k_2} \cdots q_m^{k_m} = \alpha$ (α 为无量纲量, $[\alpha] = X_1^0 X_2^0 \cdots X_m^0$) 即, 由量纲齐次原则

$$X_1^{\sum_{j=1}^m C_{1j} k_j} \cdot X_2^{\sum_{j=1}^m C_{2j} k_j} \cdots X_m^{\sum_{j=1}^m C_{mj} k_j} = X_1^0 X_2^0 \cdots X_m^0 \quad (9.1.9)$$

则必有:

$$\sum_{j=1}^m C_{1j} k_j = \sum_{j=1}^m C_{2j} k_j = \cdots = \sum_{j=1}^m C_{mj} k_j = 0 \quad (9.1.10)$$

亦即

$$\begin{cases} C_{11} k_1 + C_{12} k_2 + \cdots + C_{1m} k_m = 0 \\ C_{21} k_1 + C_{22} k_2 + \cdots + C_{2m} k_m = 0 \\ \vdots \\ C_{m1} k_1 + C_{m2} k_2 + \cdots + C_{mm} k_m = 0 \end{cases} \quad (9.1.11)$$

此线性方程组有惟一零解 $k_1 = k_2 = \cdots = k_m = 0$ 的充要条件是:

$$\begin{vmatrix} C_{11} & C_{12} & \cdots & C_{1m} \\ C_{21} & C_{22} & \cdots & C_{2m} \\ \vdots & \vdots & & \vdots \\ C_{m1} & C_{m2} & \cdots & C_{mm} \end{vmatrix} \neq 0$$

即(9.1.7)式, 证毕。

推论 设 X_1, X_2, \dots, X_m 是一组基本量纲, 若有 n 个物理量

($n \geq m$), 其导出量纲分别为:

$$[q_j] = \prod_{i=1}^m X_i^{C_{ij}} \quad j = 1, 2, \dots, n \quad (9.1.12)$$

则在 q_1, q_2, \dots, q_n 中有 r 个物理量的量纲相互独立的充要条件是矩阵 $C = (C_{ij})_{m \times n}$ 的秩为 r 。其中, 矩阵 $C = (C_{ij})_{m \times n}$ 称为量纲矩阵。

因矩阵 $C = (C_{ij})_{m \times n}$ 的秩最大为 m , 所以, $r \leq m$ 。

9.1.4 量纲分析建模的一般方法

用量纲分析法解决问题的关键是确定好该问题中所涉及到的物理量。这里, 首先介绍一个简单的例子, 然后将推导过程一般化, 引入量纲分析建模的一般方法。

单摆的周期

单摆运动是大家熟悉的物理现象, 在这里我们取相关的物理量为摆动周期时间 t , 摆球质量 m , 摆长 l 及重力加速度 g 。

设上述各物理量之间有关系:

$$t = \lambda m^{y_1} l^{y_2} g^{y_3} \quad (9.1.13)$$

其中 y_1, y_2, y_3 是待定常数, λ 是无量纲的比例系数。(9.1.13) 式的量纲表达式为:

$$[t] = [m]^{y_1} [l]^{y_2} [g]^{y_3} \quad (9.1.14)$$

单摆运动是力学问题, 取基本量纲为 L, M, T , 则上述各物理量的量纲用基本量纲表示为:

$$[t] = T, [m] = M, [l] = L, [g] = LT^{-2}$$

代入(9.1.14) 式得:

$$T = M^{y_1} L^{y_2+y_3} T^{-2y_3} \quad (9.1.15)$$

根据量纲齐次原则, 应有:

$$\begin{cases} y_1 = 0 \\ y_2 + y_3 = 0 \\ -2y_3 = 1 \end{cases} \quad (9.1.16)$$

由此解得 $y_1 = 0, y_2 = \frac{1}{2}, y_3 = -\frac{1}{2}$, 代入(9.1.13) 式得:

$$t = \lambda \sqrt{\frac{l}{g}} \quad (9.1.17)$$

与力学定律得到的结果一致。

这里的系数 λ 不能用量纲分析进一步求出, 因 λ 是无量纲量。 λ 与单摆的初始摆角 θ (无量纲量) 有关, 当 θ 很小时, λ 的值近似为 2π 。

将上例中各物理量之间的关系一般化, 写为:

$$\varphi(l, t, g, m) = 0 \quad (9.1.18)$$

并将各物理量的量纲用基本量纲 L, M, T 表示为:

$$\begin{cases} [l] = L^1 M^0 T^0 & [t] = L^0 M^0 T^1 \\ [g] = L^1 M^0 T^{-2} & [m] = L^0 M^1 T^0 \end{cases} \quad (9.1.19)$$

将上述各式中基本量纲的幂次排列成矩阵的形式, 且每一纵列表示一个物理量关于各个基本量纲的幂次, 得:

	l	t	g	m
L	1	0	1	0
M	0	0	0	1
T	0	1	-2	0

其中, 矩阵

$$C = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & -2 & 0 \end{pmatrix} \quad (9.1.20)$$

即为量纲矩阵。

不难验证矩阵 C 的秩为 $3 (< 4)$, 由定理 9.1.1 的推论, 物理量 l, t, g, m 中仅有三个物理量的量纲是相互独立的。因此, 必可找到至少一组不全为零的数 y_1, y_2, y_3, y_4 , 使得:

$$l^{y_1} t^{y_2} g^{y_3} m^{y_4} = \pi \quad (9.1.21)$$

其中 y_1, y_2, y_3, y_4 是待定常数, π 是无量纲常数, 注意到 $[\pi] = L^0 M^0 T^0$, 则(9.1.21)式的量纲式为:

$$[l]^{y_1} [t]^{y_2} [g]^{y_3} [m]^{y_4} = L^0 M^0 T^0 \quad (9.1.22)$$

将(9.1.19)中各式代入式(9.1.22), 根据量纲齐次原则, 整理得:

$$\begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & -2 & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \text{ 或 } Cy = 0 \quad (9.1.23)$$

其中 $y = (y_1, y_2, y_3, y_4)^T$ 。(9.1.23)式可写为:

$$\begin{cases} y_1 + y_3 = 0 \\ y_4 = 0 \\ y_2 - 2y_3 = 0 \end{cases} \quad (9.1.23)$$

此方程组有一个基本解:

$$y = (y_1, y_2, y_3, y_4)^T = (-1, 2, 1, 0)^T$$

代入(9.1.21)式得:

$$l^{-1} t^2 g = \pi \quad (9.1.24)$$

将(9.1.24)式中的一个物理量解出, 代入(9.1.18)式中, 通过单位的变化(见下述定理 9.1.2 的证明), 可得(9.1.18)式等价于(Φ 表示一个未定的函数关系)

$$\Phi(\pi) = 0 \quad (9.1.25)$$

此式即为单摆运动中各物理量间关系的一般表达式, (9.1.17)式只是它的特殊形式。

由此我们将(9.1.18)式到(9.1.25)式的推导过程一般化, 就是量纲分析建模的一般方法所依据的 *Buckingham* π 定理。

定理 9.1.2 设有 n 个物理量 q_1, q_2, \dots, q_n 满足如下与单位选取无关的物理定律

$$\varphi(q_1, q_2, \dots, q_n) = 0 \quad (9.1.26)$$

且设 X_1, X_2, \dots, X_m ($m < n$) 是基本量纲, q_1, q_2, \dots, q_n 的量纲可表示为:

$$[q_j] = \prod_{i=1}^m X_i^{C_{ij}}, j = 1, 2, \dots, n \quad (9.1.27)$$

矩阵 $C = (C_{ij})_{m \times n}$ 为量纲矩阵。若矩阵 C 的秩为 r , 则(9.1.26)式可等价地表示为(Φ 为一未定的函数):

$$\Phi(\pi_1, \pi_2, \dots, \pi_{n-r}) = 0 \quad (9.1.28)$$

其中 π_s ($s = 1, 2, \dots, n-r$) 为 $n-r$ 个相互独立的无量纲量, 为:

$$\pi_s = \prod_{j=1}^n q_j^{y_{sj}}, s = 1, 2, \dots, n-r \quad (9.1.29)$$

而 $y_s = (y_{s1}, y_{s2}, \dots, y_{sn})^T, s = 1, 2, \dots, n-r$, 是齐次线性方程组

$$Cy = 0 \quad (9.1.30)$$

的 $n-r$ 个基本解。(证明略)

作为定理 9.1.2 ($r = m$) 的特殊情形, 我们有:

定理 9.1.3 设有 n 个物理量 q_1, q_2, \dots, q_n 满足关系

$$\varphi(q_1, q_2, \dots, q_n) = 0 \quad (9.1.31)$$

其中 q_1, q_2, \dots, q_m ($m < n$) 有基本量纲, 而 $q_{m+1}, q_{m+2}, \dots, q_n$ 的量纲可由这些基本量纲表示。则(9.1.44)式可以表示为 $n-m$ 个无量纲量 $\pi_1, \pi_2, \dots, \pi_{n-m}$ 的关系(Φ 是一个未定的函数):

$$\Phi(\pi_1, \pi_2, \dots, \pi_{n-m}) = 0 \quad (9.1.32)$$

这里 π_s ($s = 1, 2, \dots, n-m$) 同(9.1.29)式。

上述 π 定理(定理 9.1.2 或定理 9.1.3) 可以表示成另一种等价形式, 这一形式在很多场合往往更便于使用。在一定问题中物体系的发展和演化往往由其量纲相互独立的若干个物理量决定, 不妨把它们叫做“主定参量”。在上面的推演中, 曾假设 q_1, q_2, \dots, q_r 的量纲是相互独立的, 则取它们为主定参量, 如果我们此时感兴趣的物理量是 $q_{r+1}, q_{r+2}, \dots, q_n$ 中的某一个, 例如 q_{r+k} ($1 \leq k \leq n-r$)

r), 则由定理 9.1.2 式, 我们有:

$$q_{r+k} = \pi_k \prod_{j=1}^r q_j^{\beta_{kj}}, 1 \leq k \leq n-r \quad (9.1.33)$$

其中 $\beta_{kj} (j = 1, 2, \dots, r)$ 为(9.1.30) 的第 k 个基本解的元素, 由(9.1.45) 式解出 π_k :

$$\pi_k = \Psi(\pi_1, \pi_2, \dots, \pi_{k-1}, \pi_{k+1}, \dots, \pi_{n-r}), 1 \leq k \leq n-r \quad (9.1.34)$$

代入(9.1.34) 式, 于是有:

$$q_{r+k} = \prod_{j=1}^r q_j^{\beta_{kj}} \Psi(\pi_1, \dots, \pi_{k-1}, \pi_{k+1}, \dots, \pi_{n-r}), 1 \leq k \leq n-r \quad (9.1.35)$$

这便是 π 定理另一种等价形式的表述。

§ 9.2 建模实例

在上节的基本知识中, 我们介绍了有量纲的一些概念以及量纲分析法建模所依据的定理, 本节我们列举几个实例, 并给出量纲分析法建模的一般步骤, 以阐明该方法的具体应用。

9.2.1 考虑空气阻力的单摆周期问题

若考虑空气阻力, 则在单摆运动过程中与周期 t 可能有关的物理量可取为: 摆长 l , 质量 m , 重力加速度 g , 空气阻力 R 。

显然这是个力学问题, 基本量纲选为 L, M, T 。

上述各物理量的量纲用基本量纲表示为:

$$\begin{cases} [t] = L^0 M^0 T^1, [l] = L^1 M^0 T^0, [m] = L^0 M^1 T^0 \\ [g] = L^1 M^0 T^{-2}, [R] = L^1 M^1 T^{-2} \end{cases} \quad (9.2.1)$$

由此得量纲矩阵为:

$$C_{3 \times 5} = \begin{matrix} (t)(l)(m)(g)(R) \\ \begin{bmatrix} 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & -2 & 2 \end{bmatrix} \end{matrix} \begin{matrix} (L) \\ (M) \\ (T) \end{matrix} \quad (9.2.2)$$

不难计算出, 矩阵 C 的秩为 3。

因此方程组 $Cy = 0$ 有 $n - r = 5 - 3 = 2$ 个基本解, 令 $y = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5)^T$, 则

$$\begin{bmatrix} 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & -2 & 2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad (9.2.3)$$

即

$$\begin{aligned} \beta_2 + \beta_4 + \beta_5 &= 0 \\ \beta_3 + \beta_5 &= 0 \\ \beta_1 - 2\beta_4 + 2\beta_5 &= 0 \end{aligned}$$

解方程组可得其两个基本解为:

$$\begin{cases} y_1 = (2, -1, 0, 1, 0)^T \\ y_2 = (0, 0, 1, 1, -1)^T \end{cases} \quad (9.2.4)$$

因此, 得两个相互独立的无量纲量(见(9.1.29)式):

$$\begin{cases} \pi_1 = t^2 l^{-1} g \\ \pi_2 = mgR^{-1} \end{cases} \quad (9.2.5)$$

由 π 定理, 得出考虑空气阻力时单摆问题的各物理量间的一般关系为:

$$\Phi(\pi_1, \pi_2) = 0 \quad (9.2.6)$$

其中, Φ 为一个未定的函数。

由(9.2.5)中第一式, 得:

$$t = \pi_1^{\frac{1}{2}} \sqrt{\frac{l}{g}} \quad (9.2.7)$$

又由(9.2.6)中解出 π_1 :

$$\pi_1^{\frac{1}{2}} = \Psi(\pi_2) \quad (9.2.8)$$

代入(9.2.7)式,则:

$$t = \sqrt{\frac{l}{g}} \cdot \Psi(\pi_2) = \sqrt{\frac{l}{g}} \Psi\left(\frac{mg}{R}\right) \quad (9.2.9)$$

在上式中 Ψ 是一个未定的函数,反映着单摆问题中各个无量纲量对单摆周期的影响,例如,反映着空气阻力造成的衰减及单摆初始摆角 θ 的影响等。函数 Ψ 是不能由量纲分析法确定的。

9.2.2 航船的阻力问题

长 l , 吃水深度为 h 的船以速度 v 航行,若不考虑风的影响,那么航船的阻力 f 除依赖于诸变量 l, h, v 外,还与水的参数——密度 ρ , 粘性系数 μ , 及重力加速度 g 有关。

因此,航船阻力问题中涉及的物理量有:

阻力 f , 船长 l , 吃水深度 h , 速度 v , 水的密度 ρ , 水的粘性系数 μ , 重力加速度 g 。

这是一个力学问题,取基本量纲为 L, M, T 。

根据上述各物理量的物理意义,可写出它们的量纲为:

$$\begin{cases} [f] = LMT^{-2}, [l] = L, [h] = L, \\ [v] = LT^{-1}, [\rho] = L^{-3}M, \\ [u] = L^{-1}MT^{-1}, [g] = LT^{-2} \end{cases} \quad (9.2.10)$$

其中 μ 的量纲由基本关系 $p = \mu \frac{\partial v}{\partial x}$ 得到,这里 p 是压强(单位面积受的力),故此 $[p] = LMT^{-2} \cdot L^{-2} = L^{-1}MT^{-2}$, v 是流速, x 是尺度, $[\frac{\partial v}{\partial x}] = LT^{-1} \cdot L^{-1} = T^{-1}$, 所以, $[\mu] = L^{-1}MT^{-2} \cdot T = L^{-1}MT^{-1}$ 。

由(9.2.10)中各式可写出量纲矩阵为:

$$C_{3 \times 7} = \begin{matrix} & (f) & (l) & (h) & (v) & (\rho) & (\mu) & (g) \\ \left[\begin{array}{cccccc} 1 & 1 & 1 & 1 & -3 & -1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ -2 & 0 & 0 & -1 & 0 & -1 & -2 \end{array} \right] & \begin{matrix} (L) \\ (M) \\ (T) \end{matrix} \end{matrix} \quad (9.2.11)$$

容易看出,量纲矩阵 C 的秩为 3。

因此,线性齐次方程组

$$Cy = 0 \quad (9.2.12)$$

有 $7 - 3 = 4$ 个基本解,可解出为:

$$\begin{cases} y_1 = (0, 1, -1, 0, 0, 0, 0)^T \\ y_2 = (0, 1, 0, -2, 0, 0, 1)^T \\ y_3 = (0, 1, 0, 1, 1, -1, 0)^T \\ y_4 = (1, -2, 0, -2, -1, 0, 0)^T \end{cases} \quad (9.2.13)$$

由(9.2.13)中各式,我们得到 4 个相互独立的无量纲量(见(9.1.29)式):

$$\begin{cases} \pi_1 = lh^{-1}, \pi_2 = lv^{-2}g \\ \pi_3 = lv\rho\mu^{-1}, \pi_4 = fl^{-2}v^{-2}\rho^{-1} \end{cases} \quad (9.2.14)$$

根据 π 定理,航船问题中各物理量间的全部关系可表示为:

$$\Phi(\pi_1, \pi_2, \pi_3, \pi_4) = 0 \quad (9.2.15)$$

其中 Φ 为未定的函数。

为了得到阻力 f 的显式表示式,由(9.2.14)中的 π_4 可解出:

$$f = l^2 v^2 \rho \pi_4 \quad (9.2.16)$$

将 π_4 由(9.2.15)中解出

$$\pi_4 = \Psi(\pi_1, \pi_2, \pi_3) \quad (9.2.17)$$

代入(9.2.16)式,则

$$f = l^2 v^2 \rho \Psi(\pi_1, \pi_2, \pi_3) \quad (9.2.18)$$

其中 Ψ 是一个未定的函数。

在流体力学中,无量纲量 $\pi_2^{-\frac{1}{2}} = \frac{v}{\sqrt{lg}}$ 称为 *Froude* 数, $\pi_3 = \frac{lv\rho}{\mu}$ 称为 *Reynold* 数,分别记为:

$$F_r = \frac{v}{\sqrt{lg}}, \quad R_e = \frac{lv\rho}{\mu} \quad (9.2.19)$$

因此(9.2.18)式又可写为:

$$f = l^2 v^2 \rho \Psi\left(\frac{l}{h}, F_r, R_e\right) \quad (9.2.20)$$

上式即为用量纲分析法确定的航船阻力与各物理量之间的关系,这个结果用通常的机理分析是难以得到的。虽然这里的函数 Ψ 的具体形式无从知道,但表示式(9.2.20)在物理模拟实验中有十分重要的意义。

9.2.3 点热源的热扩散问题

观察集中在一点的瞬时热源在无穷空间引起的热扩散。设 $t = 0$ 时在坐标原点 $r = 0$ 有一热量为 e 的瞬时热源,在热量向无穷空间扩散的过程中,空间任一点的温度 u 是时刻 t 和该点到原点的距离 r 的函数,而且与热扩散过程中空间介质的体积比热 c (单位体积的介质温度升高 1 度所需的热量)、介质的扩散系数 k 及热量 e 有关。另外,我们还假定在这个过程中介质的初始温度为零。

因此,此问题中涉及的物理量有:

温度 u , 距离 r , 时间 t , 热量 e , 体积比热 c , 扩散系数 k 。

这是一个热学问题,基本量纲除 L, M, T 外,还应加上温度量纲 Θ , 上述各物理量的导出量纲为:

$$[u] = \Theta, [r] = L, [t] = T$$

$$[e] = L^2 M T^{-2}$$

(热量与功的量纲相同,而功是力与距离的乘积)

$$[c] = L^{-1} M T^{-2} \Theta^{-1}$$

(由体积比热的定义, $[c] = [e] \cdot L^{-3} \cdot \Theta^{-1}$)

$$[k] = LMT^{-3}\Theta^{-1}$$

其中 k 的量纲由基本关系 $q = -k \frac{\partial u}{\partial x}$ 得到, 这里 q 是单位时间内通过单位面积的热量, $[q] = [e] \cdot T^{-1} \cdot L^{-2} = L^2 MT^{-2} \cdot T^{-1} \cdot L^{-2} = MT^{-3}$, u 是温度, x 是尺度, $[\frac{\partial u}{\partial x}] = \Theta L^{-1}$, 所以, $[k] = MT^{-3} \cdot \Theta^{-1} \cdot L = LMT^{-3}\Theta^{-1}$

由上述各量纲式得量纲矩阵为:

$$A = \begin{matrix} & (u) & (r) & (t) & (e) & (c) & (k) \\ \left\{ \begin{array}{l} 0 & 1 & 0 & 2 & -1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & -2 & -2 & -3 \\ 1 & 0 & 0 & 0 & -1 & -1 \end{array} \right. & \begin{matrix} (L) \\ (M) \\ (T) \\ (\Theta) \end{matrix} \end{matrix} \quad (9.2.21)$$

通过计算, 易知矩阵 A 的秩为 4。

因此, 方程组 $Ay = 0$ 有 $6 - 4 = 2$ 个基本解, 可取为:

$$\begin{cases} y_1 = (0, 2, -1, 0, 1, -1)^T \\ y_2 = (-2, 0, -3, 2, 1, -3)^T \end{cases} \quad (9.2.22)$$

进而, 由(9.2.22) 中两式我们得到两个相互独立的无量纲量 (见(9.1.29) 式):

$$\begin{cases} \pi_1 = r^2 t^{-1} c k^{-1} \\ \pi_2 = u^{-2} t^{-3} e^2 c k^{-3} \end{cases} \quad (9.2.23)$$

由 π 定理, 此热扩散问题中各物理量间的全部关系可表示为:

$$\Phi(\pi_1, \pi_2) = 0 \quad (9.2.24)$$

为了求得温度函数并与已知结果比较, 由(9.2.23), (9.2.24) 式, 通过运算可以写出:

$$\begin{aligned} u &= t^{-\frac{3}{2}} e c^{\frac{1}{2}} k^{-\frac{3}{2}} \pi_2^{-\frac{1}{2}} \\ &= \frac{e}{c} (a^2 t)^{-\frac{3}{2}} g\left(\frac{r^2}{a^2 t}\right), \quad a^2 = \frac{k}{c} \end{aligned} \quad (9.2.25)$$

其中 g 表示一个未定的函数。上式即为用量纲分析法得到的温度函数的表达式。

另一方面,瞬时点源的热扩散过程可以用热传导方程的方法解出,其结果为:

$$u = \frac{e}{c} \left(\frac{1}{2\sqrt{\pi a^2 t}} \right)^3 e^{-\frac{r^2}{4a^2 t}}, a^2 = \frac{k}{c} \quad (9.2.26)$$

比较(9.2.25)式与(9.2.26)式可以看出量纲分析法在建立物理问题的数学模型中所起的作用,以及能解决问题的程度(如(9.2.25)式中函数 g 的具体形式无法用量纲分析法确定)。

基本量纲的选取并不是惟一的,采用合适的基本量纲可以简化问题的求解。如在上述热扩散问题中,以热量量纲 E 代替质量量纲 M ,即取基本量纲为 Θ, L, T, E ,则该问题中各物理量的导出量纲为:

$$\begin{aligned} [u] &= \Theta, [r] = L, [t] = T, [e] = E \\ [c] &= E\Theta^{-1}L^{-3}, [k] = \Theta^{-1}L^{-1}T^{-1}E \end{aligned}$$

量纲矩阵

$$A = \begin{matrix} & (u)(r)(t)(e) & (c) & (k) \\ \begin{pmatrix} 1 & 0 & 0 & 0 & -1 & -1 \\ 0 & 1 & 0 & 0 & -3 & -1 \\ 0 & 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix} & \begin{matrix} (\Theta) \\ (L) \\ (T) \\ (E) \end{matrix} \end{matrix}$$

由于矩阵 A 的前4列构成单位矩阵,简化了方程组 $Ay = 0$ 的求解。

9.2.4 量纲分析法建模的一般步骤

由2.1~2.3中的几个建模实例,我们可总结出,根据 π 定理,量纲分析法建立数学模型的一般步骤如下:

(1) 确定问题中有关的 n 个有量纲物理量(包括变量和常

数),确定这个问题的 m 个基本量纲。

(2) 根据定义或定律,将各物理量的量纲用基本量纲表示。

(3) 写出量纲矩阵 $C = (c_{ij})_{m \times n}$,并确定量纲矩阵的秩 r 。

(4) 求解齐次线性方程组 $Cy = 0$,得到其 $n - r$ 个基本解 $y_s = (y_{s1}, y_{s2}, \dots, y_{sn})$, $s = 1, 2, \dots, n - r$ 。

(5) 由第一节(9.1.29)式求出 $n - r$ 个相互独立的无量纲量。

(6) 根据 π 定理,写出该问题中各物理量间的一般关系式即第一节(9.1.28)式。

(7) 根据 π 定理的等价形式即第一节(9.1.48)式,求出所寻求的物理量的表达式。此即为用量纲分析法建立的相应物理量的数学模型。

(8) 对模型进行分析、评价或根据其他知识进一步求解。

§ 9.3 方法总结

从上节的例子可以看出,量纲分析法在建立物理问题的数学模型时,常常很简便,甚至不需知道具体的定律和机制,便可得到一些重要的概念、有用的结果和信息。但该方法也有局限性,正确运用不总是那么容易,这要求对所解决的问题有相当的经验和对物理现象本质的透彻了解,因此,在应用量纲分析法建立数学模型时,应注意以下几点:

(1) 正确确定各物理量

面对一个实际问题,应该考虑哪些物理量或应建立哪些物理量的函数关系,对所得结果的合理性是至关重要的。例如,对上节的航船问题,如果忽略了水的密度 ρ 或粘性系数 μ ,则不可能得到正确的结果。各物理量的确定主要靠经验和物理知识,无法绝对保证所得结果是正确或有用的。

(2) 合理选取基本量纲

基本量纲的合理选取影响着问题能否解决及解决问题的难易程度。基本量纲选少了,无法表示各物理量,选多了也会使问题复杂化。在上节的热扩散问题中,若选 L, M, T, Θ, E , 五个基本量纲,则它们之间的相关性使量纲矩阵的秩不变,但方程组求解复杂了。另一方面,即使选取的基本量纲的数目相同,合适的选择也会使问题的求解简化,如上节的热扩散问题。在一般情况下,力学问题选取 L, M, T 为基本量纲即可,热学问题加上温度量纲 Θ ,电学问题加上电量量纲 Q 。

(3) 恰当构造基本解

线性齐次方程组 $Cy = 0$ 的基本解可以有許多不同的构造方法,虽然基本解组能够相互线性表示,但是为了特定的建模目的恰当地构造基本解,能够更直接地得到我们所期望的结果。如上节热扩散问题中若构造异于(9.2.22)式的基本解,则不能直接得到(9.2.25)式,而需要作必要的线性组合。

(4) 结果的效用和局限性

量纲齐次原则和 π 定理是具有普遍意义的又是相当初等的方法,不需要非常专门的物理知识及高等的数学方法,就可以得到用其他方法难以得到的结果,如上节的(9.2.20)式,或类似于用其他复杂方法得到的结果,如上节的(9.2.25)式。一般地说,从未知定律 $\varphi(q_1, q_2, \dots, q_n) = 0$ 到用量纲分析法得到的等价形式 $\Phi(\pi_1, \pi_2, \dots, \pi_{n-r}) = 0$ 。不仅物理量的个数减少了 r 个,而且,原始物理量 q_1, q_2, \dots, q_n 组合成了一些有用的无量纲量 $\pi_1, \pi_2, \dots, \pi_{n-r}$ 。另一方面,用这个方法得到的结果是有局限的,“不彻底”的。 $\Phi(\cdot) = 0$ 中仍然包含着一些未定函数和常数(无量纲量),诸如物理定律中经常出现的三角函数 $\sin(\cdot)$, 指数函数 $\exp(\cdot)$ 不可能用量纲分析法得到,因为这些函数的自变量和函数值都是无量纲的。

习 题

1. 速度为 v 的风吹在迎风面积为 s 的风车上, 空气密度是 ρ 。用量纲分析法确定风车获得的功率 p 与 v, s, ρ 的关系。

2. 深水中的波速 v 与波长 λ 、水深 d 、水的密度 ρ 和重力加速度 g 有关。用量纲分析法证明它们之间的关系可以表示为 $v = \sqrt{\lambda g} \varphi(d/\lambda)$, φ 是未知函数。

3. 原子弹爆炸时巨大的能量从爆炸点以冲击波形式向四周传播。据分析在时刻 t 冲击波达到的半径与释放能量 e 、大气密度 ρ 、大气压强 p 有关(设 $t = 0$ 时, $r = 0$), 用量纲分析方法证明 $r = (\frac{et^2}{\rho})^{1/5} \varphi(\frac{p^5 t^6}{e^2 \rho^3})$, φ 是未知函数。

4. 用量纲分析方法研究人体浸在匀速流动的水里时损失的热量, 记水的流速 v , 密度 ρ , 比热 c , 粘性系数 μ , 热传导系数 k , 人体尺寸 d 。证明人体与水的换热系数 h 与上述各物理量的关系可表为

$$h = \frac{k}{d} \cdot \varphi\left(\frac{v\rho d}{\mu}, \frac{\mu c}{k}\right),$$

φ 是未知函数, h 定义为单位时间内人体的单位面积在人体与水的温差为 1°C 时的热量交换。

5. 用量纲分析方法研究两带电平行板间的引力。两板面积为 s , 间距为 d , 电位差为 v , 板间介质的介电常数为 ϵ 。证明两板之间的引力 $f = \epsilon v^2 \varphi(s/d^2)$ 。如果又知道 f 与 s 成正比。写出 f 的表达式。这里介电常数 ϵ 的定义是 $f = \frac{q_1 q_2}{\epsilon d^2}$, 其中 q_1, q_2 是两个点电荷的电量, d 是点电荷的距离, f 是点电荷间的引力。

第十章 其他方法建模

前面各章已分别介绍了一些数学建模的方法,但由于现实现象本身以及与之相应的数学模型的多样性和复杂性,本章再简要介绍几种建立数学模型的方法,以作为进一步的补充。

§ 10.1 变分法建模

10.1.1 变分法的基本知识

变分法是研究泛函极值问题的数学方法。下面简要给出变分法的一些基本概念。

1. 泛函

设 S 为一函数集合,若对于每一个函数 $x(t) \in S$ 有一个实数 J 与之对应,则称 J 是定义在 S 上的泛函,记作 $J(x(t))$ 。 S 称为 J 的容许函数集。

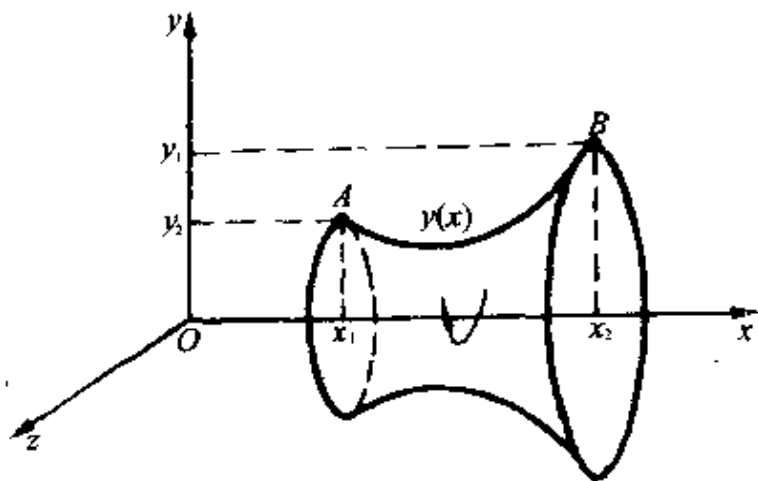


图 10.1.1

例 1 对于 xy 平面上过点 $A(x_1, y_1)$ 和 $B(x_2, y_2)$ 的每一条光滑曲线 $y(x)$, 绕 x 轴旋转得一旋转体, 如图 10.1.1, 旋转体的侧面积是曲线 $y(x)$ 的泛函 $J(y(x))$ 。由微积分知识不难写出

$$J(y(x)) = \int_{x_1}^{x_2} 2\pi y(x) \sqrt{1 + (y'(x))^2} \quad (10.1.1)$$

容许函数集可表示为:

$$S = \{y(x) \mid y(x) \in C^1[x_1, x_2], y(x_1) = y_1, y(x_2) = y_2\} \quad (10.1.2)$$

一般地,泛函的最简单形式可表为

$$J(x(t)) = \int_{t_1}^{t_2} F(t, x, x') dt \quad (10.1.3)$$

被积函数 F 包含自变量 t , 未知函数 x 及导数 x' 。

2. 泛函的极值

泛函 $J(x(t))$ 在 $x_0(t) \in S$ 取得极小值是指, 对于任意一个与 $x_0(t)$ 接近的 $x(t) \in S$, 都有 $J(x(t)) \geq J(x_0(t))$ 。所谓接近, 可以用距离 $d(x(t), x_0(t)) < \varepsilon$ 度量, 而距离定义为:

$$d(x(t), x_0(t)) = \max_{t_1 \leq t \leq t_2} \{|x(t) - x_0(t)|, |x'(t) - x'_0(t)|\}$$

泛函的极大值可以类似地定义。 $x_0(t)$ 称为泛函的极值函数或极值曲线。

3. 泛函的变分

如同函数的微分是增量的线性主部一样, 泛函的变分是泛函增量的线性主部。作为泛函的自变量, 函数 $x(t)$ 在 $x_0(t)$ 的增量记作

$$\delta x(t) = x(t) - x_0(t) \quad (10.1.4)$$

也称函数的变分。由它引起的泛函的增量为

$$\Delta J = J(x_0(t) + \delta x(t)) - J(x_0(t))$$

如果 ΔJ 可以表示为

$$\Delta J = L(x_0(t), \delta x(t)) + r(x_0(t), \delta x(t))$$

其中 L 是 δx 的线性项, 而 r 是 δx 的高阶项, 则 L 称为泛函在 $x_0(t)$ 的变分, 记作 $\delta J(x_0(t))$ 。用变动的 $x(t)$ 代替 $x_0(t)$, 就有 $\delta J(x(t))$ 。

泛函变分的一个重要形式是它可以表示为对参数 α 的导数:

$$\delta J(x(t)) = \left. \frac{\partial}{\partial \alpha} J(x(t) + \alpha \delta x(t)) \right|_{\alpha=0} \quad (10.1.5)$$

这是因为当变分存在时,增量

$$\Delta J = J(x(t) + \alpha \delta x) - J(x(t)) = L(x(t), \alpha \delta x) + r(x(t), \alpha \delta x)$$

根据 L 和 r 的性质有

$$L(x(t), \alpha \delta x) = \alpha L(x(t), \delta x)$$

$$\lim_{\alpha \rightarrow 0} \frac{r(x(t), \alpha \delta x)}{\alpha} = \lim_{\alpha \rightarrow 0} \frac{r(x(t), \alpha \delta x)}{\alpha \delta x} \cdot \delta x = 0$$

所以

$$\begin{aligned} \left. \frac{\partial}{\partial \alpha} J(x + \alpha \delta x) \right|_{\alpha=0} &= \lim_{\alpha \rightarrow 0} \frac{J(x + \alpha \delta x) - J(x)}{\alpha} \\ &= \lim_{\alpha \rightarrow 0} \frac{L(x, \alpha \delta x) + r(x, \alpha \delta x)}{\alpha} = L(x, \delta x) = \delta J(x) \end{aligned}$$

4. 极值与变分

利用变分的表达式(10.1.5)可以得到泛函极值与变分的关系:

若 $J(x(t))$ 在 $x_0(t)$ 达到极值(极大或极小)则

$$\delta J(x_0(t)) = 0 \quad (10.1.6)$$

这是因为对任意给定的 δx , $J(x_0 + \alpha \delta x)$ 是变量 α 的函数,该函数在 $\alpha = 0$ 处达到极值。根据函数极值的必要条件知

$$\left. \frac{\partial}{\partial \alpha} J(x_0 + \alpha \delta x) \right|_{\alpha=0} = 0$$

于是由(10.1.5)式直接得到(10.1.6)式。

5. 泛函极值的必要条件——欧拉方程

讨论最简泛函在固定端点条件下取得极值的必要条件。泛函和端点条件表示为

$$J(x(t)) = \int_{t_1}^{t_2} F(t, x(t), x'(t)) dt \quad (10.1.7)$$

$$x(t_1) = x_1, x(t_2) = x_2 \quad (10.1.8)$$

其中 F 具有二阶连续偏导数, 容许函数集 S 为满足(10.1.8)的二阶可微函数集合。

引理 设 $g(t)$ 是 $[t_1, t_2]$ 内的连续函数, 若对于任意的充分光滑函数 $\varphi(t)$ (满足 $\varphi(t_1) = 0, \varphi(t_2) = 0$) 有

$$\int_{t_1}^{t_2} g(t) \varphi(t) dt \equiv 0$$

则在 $[t_1, t_2]$ 内 $g(t) \equiv 0$ 。证明略。

设泛函(10.1.7)在 $x(t)$ 取得极值, $x(t)$ 满足(10.1.8)。记 $\varphi(t) = \delta x(t)$, 满足 $\varphi(t_1) = \varphi(t_2) = 0$, 按照泛函极值与变分的关系(10.1.5)、(10.1.6)式有

$$\left. \frac{\partial}{\partial \alpha} J(x(t) + \alpha \varphi(t)) \right|_{\alpha=0} = 0 \quad (10.1.9)$$

对于(10.1.7)式表示的 $J(x(t) + \alpha \varphi(t))$ 计算

$$\begin{aligned} \left. \frac{\partial}{\partial \alpha} J(x + \alpha \varphi) \right|_{\alpha=0} &= \int_{t_1}^{t_2} \left. \frac{\partial}{\partial \alpha} F(t, x + \alpha \varphi, x' + \alpha \varphi') \right|_{\alpha=0} dt \\ &= \int_{t_1}^{t_2} [F_x(t, x, x') \varphi + F_{x'}(t, x, x') \varphi'] dt \end{aligned} \quad (10.1.10)$$

对右端第二项作分部积分并利用 $\varphi(t_1) = \varphi(t_2) = 0$ 得

$$\int_{t_1}^{t_2} F_{x'} \varphi' dt = - \int_{t_1}^{t_2} \frac{d}{dt} F_{x'} \varphi dt$$

代入(10.1.10)式并根据(10.1.9)式有

$$\int_{t_1}^{t_2} (F_x - \frac{d}{dt} F_{x'}) \varphi dt = 0$$

因为 φ 是任意的, 由引理立刻得到

$$F_x - \frac{d}{dt} F_{x'} = 0 \quad (10.1.11)$$

(10.1.11)称为欧拉方程, 是 $J(x(t))$ 在 $x(t)$ 取得极值的必要条件, (10.1.11)又可记作

$$F_x - F_{xx'} - F_{xx'}x' - F_{x'x'}x'' = 0 \quad (10.1.12)$$

通常这是 $x(t)$ 的二阶微分方程, 其通解的两个任意常数由(10.1.8) 式中的两个端点条件确定。

例2 最速降线问题 求连接定点 A 、 B 的光滑曲线, 使质点在重力作用下沿该曲线以最短时间从 A 点滑到 B 点(摩擦力不计)。

将 A 点取为坐标原点, 在图 10.1.2 所示坐标系中 B 点为 $B(x_1, y_1)$ 。根据能量守恒定律, 质点在曲线 $g(x)$ 上任一点处的速度 $\frac{ds}{dt}$ (s 为弧长) 满足

$$\frac{1}{2} m \left(\frac{ds}{dt} \right)^2 = mgy$$

将 $ds = \sqrt{1 + (y'(x))^2} dx$ 代入上式得

$$dt = \sqrt{\frac{1 + (y')^2}{2gy}} dx$$

于是质点滑行时间应表示 $y(x)$ 的泛函

$$J(y(x)) = \int_0^{x_1} \sqrt{\frac{1 + (y')^2}{2gy}} dx \quad (10.1.13)$$

端点条件为

$$y(0) = 0, y(x_1) = y_1 \quad (10.1.14)$$

因为

$$F(y, y') = \sqrt{\frac{1 + (y')^2}{y}}$$

不含自变量 x , 所以方程(10.1.12) 可写作

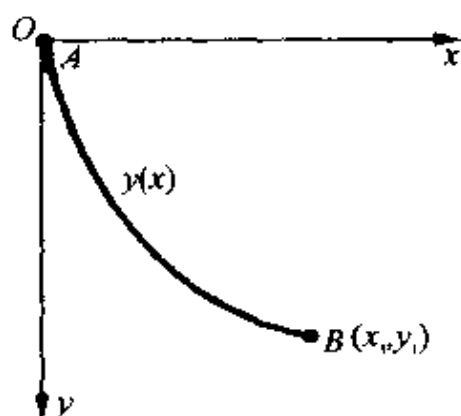


图 10.1.2 最速降线

$$F_y - F_{yy}y' - F_{y'y}y'' = 0$$

等价于

$$\frac{d}{dx}(F - y'F_{y'}) = 0$$

积分得

$$F - y'F_{y'} = c$$

c 为任意常数。将 F 代入得

$$\sqrt{\frac{1 + (y')^2}{y}} - \frac{(y')^2}{\sqrt{y(1 + (y')^2)}} = c$$

即

$$y(1 + (y')^2) = \frac{1}{c^2}$$

化为参数方程可表为

$$\begin{cases} x = c_1(t - \sin t) + c_2 \\ y = c_1(1 - \cos t) \end{cases} \quad (10.1.15)$$

这是圆弁线方程, 由 $y(0) = 0$ 可知 $c_2 = 0$, c_1 由另一端点条件 $y(x_1) = y_1$ 确定。

虽然欧拉方程的解(10.1.15)只是泛函(10.1.13)的达到板值的必要条件, 但如同求函数极植一样, 对于实际问题通常不用检验它是否满足泛函达到极小值的充分条件。因极小值曲线一定存在, 它满足欧拉方程, 即(10.1.15)式即为所求。

泛函极值的必要条件可以推广到含有两个及两个以上未知函数的情况, 如

$$J(x(t), u(t)) = \int_{t_1}^{t_2} F(t, x, x', u, u') dt \quad (10.1.16)$$

的欧拉方程是

$$\begin{cases} f_x - \frac{d}{dt}F_{x'} = 0 \\ F_u - \frac{d}{dt}F_{u'} = 0 \end{cases} \quad (10.1.17)$$

6. 可动边界的泛函极值

如果容许函数 $x(t)$ 的一个端点如 $t = t_2$ 不固定,而是在一条给定的曲线 $x = \psi(t)$ 上变动,如图 10.1.3,于是端点条件表示为

$$x(t_1) = x_1, x(t_2) = \psi(t_2) (t_2 \text{ 可变}) \quad (10.1.18)$$

讨论泛函(10.1.7) 在条件(10.1.18) 下极值曲线的求法。

首先,如果 $x(t)$ 使(10.1.7)、(10.1.18) 取得极值,那么当 t_2 固定时 $x(t)$ 也使(10.1.7)、(10.1.18) 取得极值,因为前者的容许函数集包含了后者。也即使(10.1.7)、(10.1.18) 取得极值的 $x(t)$ 也满足欧拉方程(10.1.11)。

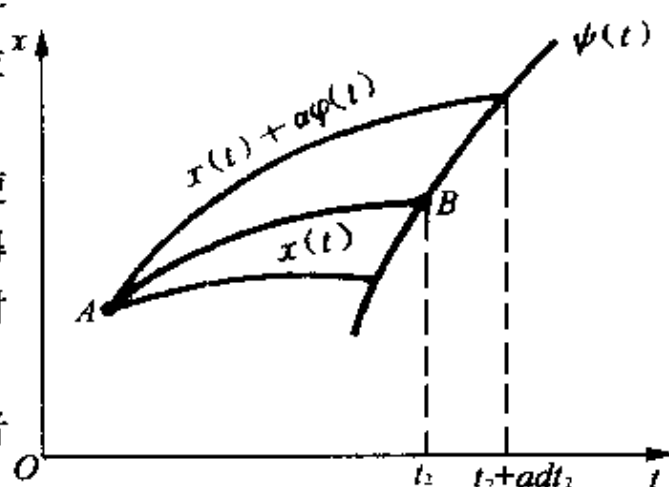


图 10.1.3

其次,欧拉方程的另一个定解条件要由(10.1.18) 中的 $x(t_2) = \psi(t_2)$ 确定。其推导与欧拉方程的推导类似。不过因 t_2 可变,故容许曲线 $x(t)$ 的末端时刻可表为 $t_2 + \alpha dt_2$,于是从(10.1.9) 式开始的推导过程为

$$0 = \frac{\partial}{\partial \alpha} \int_{t_1}^{t_2 + \alpha dt_2} F(t, x + t\varphi, x' + \alpha\varphi') dt \Big|_{\alpha=0}$$

$$\int_{t_1}^{t_2 + \alpha dt_2} (F_x \varphi + F_{x'} \varphi') dt \Big|_{\alpha=0}$$

$$\begin{aligned}
& + F(t_2 + \alpha dt_2, x + \alpha \varphi, x' + \alpha \varphi') dt_2 \Big|_{t=t_2}^{\alpha=0} \\
& = \int_{t_1}^{t_2} (F_x - \frac{d}{dx} F_{x'}) \varphi dt + F_x \varphi \Big|_{t=t_2} + F \Big|_{t=t_2} dt_2 \quad (10.1.19)
\end{aligned}$$

因 $x(t)$ 满足欧拉方程,所以上式右端第一项为零,为得到 $\varphi(t_2)$ 与 dt_2 的关系考察末端时刻 $t_2 + \alpha dt_2$,应有

$$x(t_2 + \alpha dt_2) + \alpha \varphi(t_2 + \alpha dt_2) = \psi(t_2 + \alpha dt_2)$$

对 α 求导并令 $\alpha = 0$ 得

$$x'(t_2) dt_2 + \varphi(t_2) = \psi'(t_2) dt_2$$

$$\text{即 } \varphi(t_2) = [\psi'(t_2) - x'(t_2)] dt_2. \quad (10.1.20)$$

将(10.1.20)代入(10.1.19)式并利用 dt_2 的任意性,我们得到

$$[F + (\psi' - x') f_{x'}] \Big|_{t=t_2} = 0 \quad (10.1.21)$$

(10.1.21) 就是确定欧拉方程中另一常数的定解条件,称为横截条件。

横截条件有两种常用的特殊情况:

当 $x = \psi(t)$ 垂直于横轴时, t_2 固定但 $x'(t_2)$ 自由,称 t_2 为自由端点,因为在(10.1.19)中 $dt_2 = 0$ 并注意到 φ 的任意性得

$$F_{x'} \Big|_{t=t_2} = 0 \quad (10.1.22)$$

这是自由端点的定解条件

当 $x = \psi(t)$ 平行于横轴时, $\psi' = 0$, (10.1.21) 变为

$$(F - x' F_{x'}) \Big|_{t=t_2} = 0 \quad (10.1.23)$$

7. 条件极值

前面讨论的是无条件泛函极值问题,最优控制涉及的通常是条件极值,典型形式为

$$J(u(t)) = \int_{t_1}^{t_2} F(t, x(t), u(t)) dt \quad (10.1.24)$$

$$\dot{x}(t) = f(t, x(t), u(t)) \quad (10.1.25)$$

其中 $u(t)$ 是控制函数, $x(t)$ 是状态函数或称轨线, 问题的提法是求 $u(t) \in U$ 使泛函(10.1.24) 在条件(10.1.25) 下达到极值, 并且 $x(t) \in S$, 这里 U 和 S 分别是 $u(t)$ 和 $x(t)$ 的容许集。

用拉格朗日乘子法化条件极值为无条件极值问题, 引入乘子函数 $\lambda(t)$ 构造泛函

$$I(x(t), u(t)) = \int_{t_1}^{t_2} [F(t, x, u) + \lambda(t)(f(t, x, u) - x')] dt \quad (10.1.26)$$

记 $H(t, x, u) = F(t, x, u) + \lambda(t)f(t, x, u)$ (10.1.27)

H 称哈密尔顿函数, (10.1.26) 式可写作

$$I(x(t), u(t)) = \int_{t_1}^{t_2} (H - \lambda x') dt \quad (10.1.28)$$

条件极值(10.1.24)、(10.1.25) 与含有两个函数 $x(t), u(t)$ 的无条件极值问题(10.1.28) 等价, 根据(10.1.17) 式, (10.1.28) 的欧拉方程为

$$\begin{cases} (H - \lambda x')_x - \frac{d}{dt}(H - \lambda x')_{x'} = 0 \\ (H - \lambda x')_u - \frac{d}{dt}(H - \lambda x')_{u'} = 0 \end{cases} \quad (10.1.29)$$

将(10.1.27) 给出的 H 的表达式代入(10.1.29)

$$\begin{cases} \frac{\partial H}{\partial x} + \lambda'(t) = 0 \\ \frac{\partial H}{\partial u} = 0 \end{cases}$$

于是条件极值问题归结为求解未知函数 $x(t), u(t), \lambda(t)$ 的微分方程组

$$\begin{cases} \lambda'(t) = -\frac{\partial H}{\partial x} \\ \frac{\partial H}{\partial u} = 0 \\ x' = f(t, x, u) \end{cases} \quad (10.1.30)$$

最优控制函数 $u(t)$ 和最优状态函数(最优轨线) $x(t)$ 由方程组 (10.1.30) 在 $x(t)$ 的端点条件下解出。

10.1.2 建模实例

1. 生产计划的制订问题

工厂与客户签订了一项在某时刻提交一定数量产品的合同, 在制定生产计划时要考虑生产和贮存两种费用。生产费用通常取决于生产率(单位时间的产量), 生产率越高费用越大; 贮存费用由已经生产出来的产品数量决定, 数量越多费用越大, 所谓生产计划这里简单地看作是到每一时刻为止的累积产量, 它与每单位时间(如每天)的产量可以互相推算。建模目的是寻求最优的生产计划, 使完成合同所需的总费用(生产与贮存费用之和)最小。

假设开始生产时刻记为 $t = 0$, 按照合同应在 $t = T$ 提交数量为 Q 的产品, 到时刻 t 为止的产量记作 $x(t)$, $x(t)$ 即生产计划。因为时刻 t 的生产率表示为 $x'(t)$, 所以单位时间的生产费用可以一般地记作 $f(x'(t))$, 而单位时间的贮存费用则应记为 $g(x(t))$, 于是从 $t = 0$ 到 $t = T$ 的总费用 $C(x(t))$ 是

$$C(x(t)) = \int_0^T [f(x'(t)) + g(x(t))] dt \quad (10.1.31)$$

为了确定 f 和 g 的具体形式作如下假设。

(1) 单位时间内生产率提高一个单位所需的生费用与这时的生产率成正比

(2) 贮存费与贮存量(即累积产量)成正比

由假设(1), 生产费用 f 对生产率 x' 的变化率与 x' 成正比, 即

$$\frac{df}{dx'} \propto x'$$

于是,可令

$$f(x'(t)) = k_1 x'^2(t) \quad (10.1.32)$$

k_1 是比例系数,由假设(2) 则可直接写出

$$g(x(t)) = k_2 x(t) \quad (10.1.33)$$

k_2 是单位数量产品单位时间的贮存费。

将(10.1.32)、(10.1.33) 代入(10.1.31) 式并注意到 $x(t)$ 在 $t = 0$ 和 $t = T$ 时的值,我们有

$$C(x(t)) = \int_0^T [k_1 x'^2(t) + k_2 x(t)] dt \quad (10.1.34)$$

$$x(0) = 0, \quad x(T) = Q \quad (10.1.35)$$

制订最优生产计划归结为在固定端点条件(10.1.35) 下,求 $x(t)$ 使(10.1.34) 式定义的泛函 $C(x(t))$ 取得最小值。

用变分法求解。

记 $F(t, x, x') = k_1 x'^2 + k_2 x$, 根据欧拉方程

$$F_x(t, x, x') - \frac{d}{dt} F_{x'}(t, x, x') = 0$$

可得关于 $x(t)$ 的二阶微分方程

$$k_2 - 2k_1 x''(t) = 0 \quad (10.1.36)$$

方程(10.1.36) 在端点条件(10.1.35) 下的解为

$$x(t) = \frac{k_2}{4k_1} t^2 + \frac{4k_1 Q - k_2 T^2}{4k_1 T} \quad (10.1.37)$$

这就是使总费用 $C(x(t))$ 达到最小的生产计划。

由(10.1.37) 式不难画出 $x(t)$ 的示意图(图 10.1.4),它是过 $x(0) = 0, x(T) = Q$ 两点的抛物线,且因 $x''(t) > 0$ 而呈下凸状,随着参数 k_1, k_2, T, Q 的不同,曲线 $x(t)$ 可能有 S_1 和 S_2 两种形状,但是对于生产计划 $x(t)$ 应该有明显的限制条件

$$x(t) \geq 0, 0 \leq t \leq T \quad (10.1.38)$$

亦即, 只有当 $x(t)$ 呈 S_1 形状时才有实际意义。

容易看出, 对于(10.1.37) 式表示的 $x(t)$, 条件(10.1.38) 等价于 $x'(0) \geq 0$ (10.1.39)

由(10.1.37) 式算出 $x'(0)$, 可知(10.1.39) 式又表示为

$$Q \geq \frac{k_2 T^2}{4k_1} \quad (10.1.40)$$

于是仅当(10.1.40) 式成立时(10.1.37) 式确定的 $x(t)$ 才是最优生产计划。当 k_1, k_2 固定时条件(10.1.40) 表明, 在一定交货期 T 内要完成的产量 Q 相当大, 需要从 $t = 0$ 就开始生产, 如图10.1.4 中曲线 S_1 。

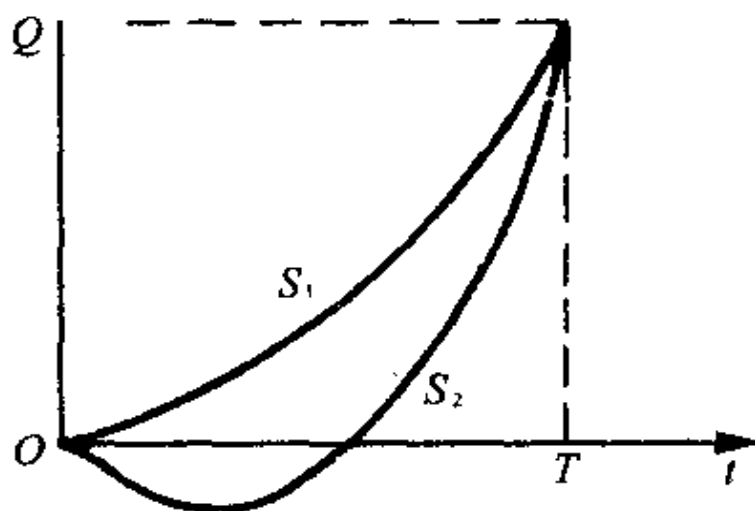


图 10.1.4

但是, 当

$$Q < \frac{k_2 T^2}{4k_1} \quad (10.1.41)$$

即在 T 内要完成的产量 Q 较小时最优生产计划是什么呢? 直观的

想法是为了节省贮存费用, 到 $t = t_1$ 才开始生产, 如图 10.1.5 的图线 S_3 。此时, t_1 的确定可视为曲线 S_3 在直线 $x = 0$ 上变动的泛函极值问题, 这里不再叙述求解过程。

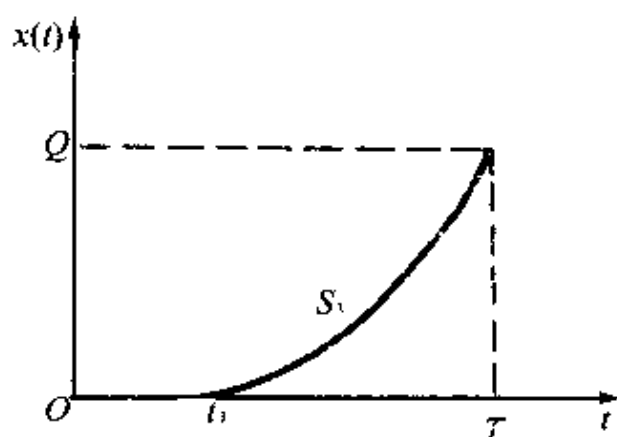


图 10.1.5

说明 为了对最优生产计划作出解释, 考察它满足的方程(10.1.36)式, 即

$$k_2 = \frac{d}{dt} \left(\frac{d}{dx'} (k_1 x'^2) \right) = \frac{d}{dt} \left(\frac{df}{dx'} \right) \quad (10.1.42)$$

式中 $\frac{df}{dx'}$ 是单位时间内生产率提高一个单位所需的生产费用, 经济理论中称为边际成本, 而 k_2 (单位时间单位数量产品的贮存费) 称为边际贮存。于是(10.1.42)式表明, 使边际成本的变化率等于边际贮存的生产计划是最优的。

2. 国民收入的增长问题

国民经济收入主要用于两个方面: 扩大再生产的积累资金和满足人民生活需要的消费资金。如何安排积累和消费资金的比例使国民收入得到最快的增长, 是一个重大的理论和实践问题, 这里仅从最优控制的角度介绍一个十分简化的模型。

一般模型, 将时刻 t 的国民经济收入记作 $x(t)$, 其中用于积累资金的部分记作 $y(t)$, 积累资金在国民经济收入中占的比例 $u(t) = \frac{y(t)}{x(t)}$ 称为积累率。建模的目的是寻求最优积累率 $u(t)$ 使国民收入增长最快, 下面给出模型的一般形式。

国民收入的增长率 $x'(t)$ 取决于当时的收入总值 $x(t)$ 和积累率 $u(t)$, 可以表示为

$$x'(t) = f(t, x, u) \quad (10.1.43)$$

考虑一段时间 T (一个五年或十年计划), 使 $x(t)$ 从初值 x_0 达到尽可能大的 $x(T)$, 即

$$x(0) = x_0 \quad (10.1.44)$$

$$\max x(T) \quad (10.1.45)$$

问题归结为在条件(10.1.43)、(10.1.44)下求 $u(t)$ 满足(10.1.45)。而这又等价于它的对偶问题: 在固定端点条件 $x(T) = x_1$ 下使 T 最小。于是若记

$$x(0) = x_0, x(T) = x_1 \quad (10.1.46)$$

$$J(u(t)) = \int_0^T dt \quad (10.1.47)$$

则模型的一般提法是在条件(10.1.43)、(10.1.46)下求 $u(t)$ 使(10.1.47)式的 J 达到最小。

按照泛函的条件极值问题的解法, 构造哈密尔顿函数

$$H = 1 + \lambda f(t, x, u) \quad (10.1.48)$$

并得到方程组

$$\begin{cases} \lambda'(t) = -\lambda f_x(t, x, u) \\ \lambda f_u(t, x, u) = 0 \\ x'(t) = f(t, x, u) \\ x(0) = x_0, x(T) = x_1 \end{cases} \quad (10.1.49)$$

由此求解最优控制函数 $u(t)$ 和最优状态 $x(t)$ 。

简化模型 为了确定函数 f 具体的、简化的形式我们粗略地考虑这样的经济规律: 当积累率 u 较小时, 国民收入的相对增长率 $\frac{x'}{x}$ 随着 u 的增加而增加, 这是因为积累资金的扩大再生产起着促进作用; 随着 u 的增大, $\frac{x'}{x}$ 增加变得越来越慢; 当 u 增到一定程度后 $\frac{x'}{x}$ 反而会减少, 这是由于消费资金比例太少, 各种制约因素导

致国民收入下降,描述这种规律的最简单的模型是

$$\frac{x'}{x} = u(a - bu) \quad (10.1.50)$$

其中 a, b 为常数,可以根据统计数据或经验估计得出。

比较(10.1.43)、(10.1.50)式得到 $f = u(a - bu)x$,代入方程(10.1.49)得

$$\begin{cases} \lambda' = -\lambda u(a - bu) \\ \lambda(a - 2bu)x = 0 \\ x' = u(a - bu)x \\ x(0) = x_0, x(T) = x_1 \end{cases} \quad (10.1.51)$$

方程(10.1.51)的解为

$$u(t) = \frac{a}{2b} \quad (10.1.52)$$

$$x(t) = x_0 e^{\frac{a^2}{4b}t} \quad (10.1.53)$$

即最优积累率是常数 $\frac{a}{2b}$,这时国民收入以指数形式(10.1.53)增长,用最短时间

$$T = \frac{4b}{a^2} \ln \frac{x_1}{x_0} \quad (10.1.54)$$

从 x_0 上升到 x_1 。

说明 一般情况下国民收入依赖于积累率的最优增长是典型的快速控制问题(10.1.43)、(10.1.46)、(10.1.47),最优解要从非线性方程组(10.1.49)得到。而在我们对国民收入的简化假设(10.1.50)下,得到的是平凡解——常数 $\frac{a}{2b}$ 。实际上,对于这种简单情况不必求解泛函极值问题。由(10.1.50)式可以直接看出,当 $u = \frac{a}{2b}$ 时 x' 最大,所 $u(t) = \frac{a}{2b}$ 必定是模型的最优控制函数。

3. 赛跑速度的安排问题

参加赛跑的运动员要根据自己的生理状况对赛程中各阶段的速度作出最恰当的安排,以期获得最好的成绩。寻求速度安排的最佳策略是一个涉及生物力学的复杂问题。*T. B. Keller* 提出了一个简单模型,根据 4 个生理参数从最优控制的角度确定赛程中的速度函数,并可以预测比赛成绩。

按照 *Keller* 的模型,短跑比赛应该用最大冲力跑完全程,对于中长跑则要将距离分为 3 段,先用最大冲力起跑,然后匀速跑过大部分赛程,最后把贮存在体内的能量用完,靠惯性冲过终点。

问题分析 运动员在赛跑过程中克服体内外的阻力以达到或保持一定速度,需要发挥出向前的冲力,为冲力做功提供能量有两个来源:一是呼吸和循环系统通过氧的新陈代谢作用产生的与吸入氧量等价的能量,二是赛跑前贮存在身体内的供赛跑用的能量。对于前者可以合理地假设在赛跑过程中保持常数,而后者则有一个如何将贮存能量分配到赛程的各个阶段,并恰好在到达终点前将其用完的问题。

模型需要确定三个关系:一是冲力与速度的关系,二是冲力做功与上述两个能量来源的关系,三是速度与比赛成绩的关系。虽然比赛成绩指的是一定赛程所用的时间,但是在数学处理时可以等价为一时间所跑的距离,于是最佳成绩归结为以距离为目标泛函的,与速度、冲力、贮存能量等函数有关的极值问题。这个问题一般解过于复杂,*Keller* 把它简化了。

由以上分析还可看出模型需要 4 个生理参数:运动员能发挥的最大冲力,体内外的阻力系数,氧的新陈代谢作用单位时间所提供的能量,体内贮存能量的初值,在 *Keller* 的模型中这些参数是用世界记录创造者的成绩估计出来的。

模型假设 需要对赛跑中的阻力作出假设,以确定冲力与速度的关系,还要对氧的代谢作用提供的能量作出假设以建立能量供给与消耗间的平衡。

(1) 赛跑时体内外的阻力与速度成正比, 比例系数 τ^{-1} 。运动员能发挥的最大冲力为 F , 初速为零。

(2) 呼吸与循环系统在氧的代谢作用下单位时间提供的能量是常数 σ , 初始时刻体内贮存的能量是 E 。

上述参数因人而异, 特别与运动员体重有关, 我们对运动员的单位质量建模, 即在下面各式中均设质量 $m = 1$ 。

一般模型 设运动员以速度函数 $v(t)$ 跑完赛程 D 的时间为 T , 则

$$D(v(t)) = \int_0^T v(t) dt \quad (10.1.55)$$

D 已知时求 $v(t)$ 使 T 最小, 等价于 T 固定求 $v(t)$ 使 D 达到最大。

运动员的冲力记作 $f(t)$, 由假设(1) 根据牛顿第二定律得

$$v'(t) + \frac{v}{\tau} = f(t) \quad (10.1.56)$$

$$v(0) = 0 \quad (10.1.57)$$

$$0 \leq f(t) \leq F \quad (10.1.58)$$

运动员体内贮存的能量记作 $E(t)$, 则单位时间提供给赛跑用的为 $-E'(t)$ (贮存能量的减少率)。因赛跑中单位时间消耗的能量是 $f(t)v(t)$, 由假设(2) 这个能量被 σ 和 $-E'(t)$ 所平衡, 所以

$$\sigma - E'(t) = fv \quad (10.1.59)$$

$$E(0) = E_0 \quad (10.1.60)$$

$$E(t) \geq 0 \quad (10.1.61)$$

这样构成了在条件(10.1.56) – (10.1.61) 下以(10.1.55) 式 $D(v(t))$ 为目标函数的泛函极值问题, 其中 F, τ, σ, E_0 视为已知参数。

将(10.1.55)、(10.1.56)、(10.1.59) 合并, 得到泛函

$$D(E(t)) = \sqrt{2} \int_0^T e^{-\frac{t}{\tau}} \left[\int_0^T (\sigma - E'(t)) e^{\frac{2s}{\tau}} ds \right]^{\frac{1}{2}} dt \quad (10.1.62)$$

这个一般模型的求解是困难的,下面将 *Keller* 提出的简化方法分两个模型叙述。

短跑模型 当赛程较短时可以用最大冲力 F 跑完全程,至于多长赛程用这种方法跑,应以 $E(t)$ 不小于零为标准,由参数 F 、 τ 、 σ 、 E_0 决定。

将 $f(t) = F$ 代入(10.1.56),在初始条件(10.1.57)下的解为

$$v(t) = F\tau(1 - e^{-\frac{t}{\tau}}) \quad (10.1.63)$$

可知速度是递增的,将(10.1.63)代入(10.1.59)得

$$E'(t) = \sigma - F^2\tau(1 - e^{-\frac{t}{\tau}}) \quad (10.1.64)$$

方程(10.1.64)在初始条件(10.1.60)下的解为

$$E(t) = E_0 - (F^2\tau - \sigma)t + F^2\tau^2(1 - e^{-\frac{t}{\tau}}) \quad (10.1.65)$$

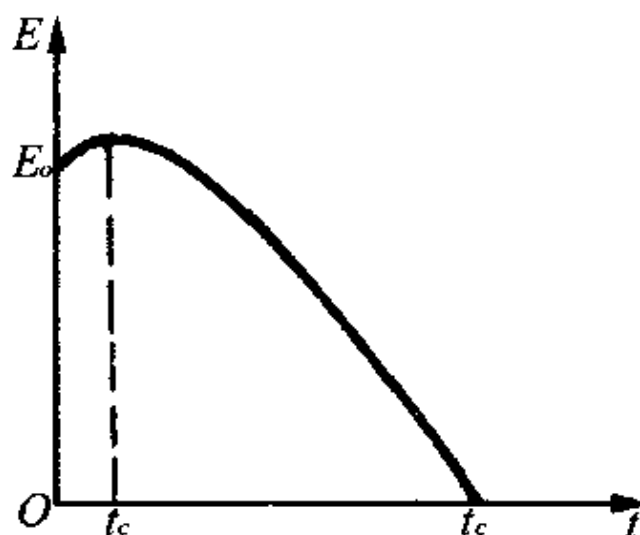


图 10.1.6 体内能量 $E(t)$ 的示意图

由(10.1.64)、(10.1.65)可画出 $E(t)$ 的示意图(图10.1.6),起跑后在很短的时间 $0 \leq t \leq t_c$ 内,由于速度 v 很小, σ 的一部分补充给 $E(t)$,所以 $E(t)$ 增加,随着 v 的迅速变大, $E(t)$ 下降,当 $t = t_c$ 时 $E(t) = 0$ 。即

$$t_c = t |_{E(t)=0} \quad (10.1.66)$$

$$t_c = t |_{E(t)=0} = \tau \ln \frac{F^2\tau}{F^2\tau - \sigma} \quad (10.1.67)$$

在这种情况下所能跑的最远距离为

$$D_c = \int_0^{t_c} v(t) dt = F\tau^2 \left(e^{-\frac{t_c}{\tau}} + \frac{t_c}{\tau} - 1 \right) \quad (10.1.68)$$

利用各个参数的估计值(见本节参数估计一段)可以算出 $t_c \doteq 27.63$ 秒, $t_p \doteq 0.33$ 秒, $D_c \doteq 291$ 米。这就是说,当赛程小于 291 米时用最大冲力跑完全程是可行的,并且能够取得最佳成绩(因为参数是用世界纪录创造者的成绩估计出来的,所以算出的 t_c 、 t_p 、 D_c 对于一般运动员不一定合适)。

中长跑模型 当赛程大于 D_c 时,将全程分为 3 段:初始阶段 $0 \leq t \leq t_1$ 以最大冲力 $f(t) = F$ 跑,以便在短时间内获得尽可能的高速度,中间阶段 $t_1 \leq t \leq t_2$ 保持匀速,最后阶段 $t_2 \leq t \leq T$ 把体内贮存能量用完,即 $E(t) = 0$,可以看出,由于在第 1、3 段分别把控制函数 $f(t)$ 、 $E(t)$ 确定在约束条件(10.1.58)、(10.1.61)的边界上,且取常值,从而这两阶段的速度 $v(t)$ 也被确定。由此不难进一步求出中间阶段的 $v(t)$ 。

第 1 段 $0 \leq t \leq t_1$ (t_1 待定) $v(t)$ 由(10.1.63) 式表示。

第 3 段 $t_2 \leq t \leq T$ (t_2 待定) 将 $E(t)$ 代入(10.1.56)、(10.1.59) 式得

$$\frac{1}{2} \frac{d}{dt} v^2 + \frac{v^2}{\tau} = \sigma \quad (10.1.69)$$

其解为

$$v(t) = \left[(v^2(t_2) - \sigma\tau) e^{-\frac{2(t-t_2)}{\tau}} + \sigma\tau \right]^{\frac{1}{2}} \quad (10.1.70)$$

$v(t)$ 是单调减和下凸的 ($v'(t) < 0$, $v''(t) > 0$)

第 2 段 $t_1 \leq t \leq t_2$ $v(t)$ 由目标泛函(10.1.55) 式 $D(v(t))$ 达到最大来确定。将(10.1.63)、(10.1.70) 代入(10.1.55) 式得

$$\begin{aligned} D(v(t)) = & \int_0^{t_1} F\tau(1 - e^{-\frac{t}{\tau}}) dt + \int_{t_1}^{t_2} v(t) dt \\ & + \int_{t_2}^T \left[(v^2(t_2) - \sigma\tau) e^{-\frac{2(t-t_2)}{\tau}} + \sigma\tau \right]^{\frac{1}{2}} dt \end{aligned} \quad (10.1.71)$$

其中 $v(t_2)$ 可由 $E(t_2) = 0$ 确定, 而 $E(t)$ 满足方程(10.1.56)、(10.1.59), 即

$$E' = \sigma - vv' - \frac{v^2}{\tau} \quad (10.1.72)$$

(10.1.72) 在初始条件(10.1.60) 下的解为

$$E(t) = E_0 + \sigma t - \frac{v^2(t)}{2} - \frac{1}{\tau} \int_0^t v^2(t) dt \quad (10.1.73)$$

令 $t = t_2$ 并将(10.1.63) 式代入得

$$\begin{aligned} E(t_2) = E_0 + \sigma t_2 - \frac{v^2(t_2)}{2} - \frac{1}{\tau} \int_0^{t_1} F^2 \tau^2 (1 - e^{-\frac{t}{\tau}})^2 dt \\ - \frac{1}{\tau} \int_{t_1}^{t_2} v^2(t) dt \end{aligned} \quad (10.1.74)$$

则问题归结为在条件

$$E(t_2) = 0 \quad (10.1.75)$$

下求 $v(t)$ ($t_1 \leq t \leq t_2$, t_2 自由) 使(10.1.71) 式的 $D(v(t))$ 最大。

因为(10.1.71)、(10.1.74) 中均不含 $v'(t)$, 引入拉格朗日乘子化为无条件极值问题后的欧拉方程将蜕化为代数方程, 最优解 $v(t)$ 是常数, 所以可以用待定常数 λ 构造函数

$$I(v(t), t_2) = D(v(t)) + \frac{\lambda}{2} E(t_2) \quad (10.1.76)$$

不妨将(10.1.76) 式右端与 $v(t)$ 和 t_2 无关的项略去, 写成

$$\begin{aligned} I(v(t), t_2) &= \int_{t_1}^{t_2} [v(t) - \frac{\lambda}{2\tau} v^2(t)] dt \\ &+ \int_{t_2}^T [(v^2(t_2) - \sigma\tau) e^{-\frac{2(t-t_2)}{\tau}} + \sigma\tau]^{\frac{1}{2}} dt + \frac{\lambda\sigma}{2} t_2 - \frac{\lambda}{4} v^2(t_2) \\ &= \int_{t_1}^{t_2} F(v(t)) dt + G(t_2) \end{aligned} \quad (10.1.77)$$

由(10.1.77) 式的变分为零, 可得 $v(t)$ 、 t_2 为最优解的必要条件是

$$F_v = 0, (F + G_t)|_{t=t_2} = 0 \quad (10.1.78)$$

以(10.1.77)代入(10.1.78)式得

$$v(t) = \frac{\tau}{\lambda} \quad (10.1.79)$$

$$2 \int_{t_2}^T [(v^2(t_2) - \sigma\tau)e^{-\frac{2(t-t_2)}{\tau}} + \sigma\tau] \cdot \frac{1}{2} e^{-\frac{2(t-t_2)}{\tau}} dt = \lambda \quad (10.1.80)$$

至此,3个阶段的 $v(t)$ 分别由(10.1.63)、(10.1.70)、(10.1.79)给出,其中 $v(t_2) = v(t) (t_1 \leq t \leq t_2)$ 。下面确定 t_1 、 t_2 和 λ 。

t_1 、 t_2 和 λ 的确定 利用 $v(t)$ 在 $t = t_1$ 的连续性,由(10.1.63)和(10.1.79)式得

$$\lambda F(1 - e^{-\frac{t_1}{\tau}}) = 1 \quad (10.1.81)$$

将(10.1.79)代入(10.1.74)、(10.1.75)得

$$E_0 + \sigma t_2 - \frac{\tau^2}{2\lambda^2} - F^2\tau(t_1 - \frac{3\tau}{2} + 2\tau e^{-\frac{t_1}{\tau}} - \frac{\tau}{2} e^{-\frac{2t_1}{\tau}}) - \frac{\tau}{\lambda^2}(t_2 - t_1) = 0 \quad (10.1.82)$$

将(10.1.79)式的 $v(t)$ 代入(10.1.80)式的 $v(t_2)$,作出积分得

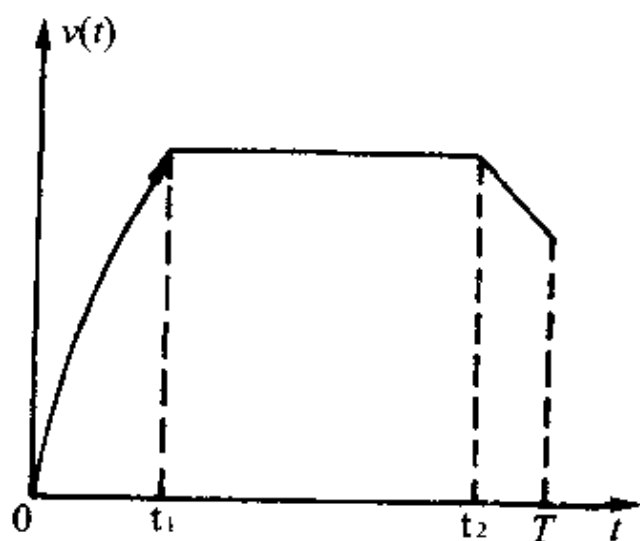
$$2[(\tau^2 - \lambda^2\sigma\tau)e^{-\frac{2(T-t_2)}{\tau}} + \lambda^2\sigma\tau]^{\frac{1}{2}} - 2\tau = \lambda^2\sigma - \tau \quad (10.1.83)$$

t_1 、 t_2 、 λ 由(10.1.81) - (10.1.83)确定。

参数估计 为了估计4个生理参数 F 、 τ 、 σ 、 E_0 , Keller 先用若干个短跑(赛程 $D < D_c$) 世界记录拟合由(10.1.55)、(10.1.63)式计算的理论结果,用最小二乘法得到 F 、 τ 的估计值,再用一些中长跑(赛程 $D > D_c$) 世界记录拟合(10.1.63)、(10.1.70)、(10.1.79)、(10.1.81) - (10.1.83) 和(10.1.55)计算的理论结果,用最小二乘法得到 σ 和 E_0 计算的理论结果,用最小二乘法得

到 σ 和 E_0 的估计值。他得到的结果是(对质量 $m = 1$ 千克而言): $F = 12.2$ 米/秒², $\tau = 0.892$ 秒, $\sigma = 41.5$ 米²/秒³, $E_0 = 2403.5$ 米²/秒²。

模型解释 中长跑模型的速度函数由 3 段组成, 如图 10.1.7。对于最后一段(通常有一两秒)速度的下降, Keller 的解释是: 象汽车比赛到终点前将燃料用完, 靠惯性冲进终点一样, 赛跑的最佳策略应该是把贮存在体内供赛跑用的能量全部耗尽, 借助惯性冲刺, 这必导致短暂的速度下降。单从赛跑所用的时间来看, 譬如一名运动员测验自己的成绩, 这样做是最优的, 而在实际比赛中当运动员与对手势均力敌时, 从击败对手取得好名次的目的出发, 需要按照实际情况巧妙地安排自己的速度, 这已不是本模型讨论的范围了。



说明: Keller 对赛跑的一般模型(10.1.55) ~ (10.1.62) 提出了分段解法, 虽然没有严格证明它的解就是(10.1.55) ~ (10.1.62) 的最优解, 但从

图 10.1.7 中长跑模型的最佳速度示意图

分析过程和实际检验可以看出这种简化方法是合理的。另一方面, 模型本身也存在着一些不合适的地方, 例如对短跑和中长跑的所有赛程用了同一组生理参数, 而短跑和中长跑运动员的这些参数可能有较大差别, 以及运动员达到一定的高速度后不可能持续地发挥自己的最大冲力, 等等。尽管如此, 这个模型在把动力学与生

理学相结合,用建模方法讨论体育运动问题上,为人们做出了示范。

§ 10.2 对策方法建模

10.2.1 对策论的基本知识

在现实世界中,存在着许多具有竞争性质的现象,如政治论战、武装冲突、企业兼并、商业竞争、体育竞赛等等。所有这些现象都有个共同的特点,就是都反映了集团之间或者个人之间的利害冲突,这些关于利害冲突现象的理论模型可称为对策模型,建立对策模型的目的,就是要把利害冲突现象中的本质问题抽象出来。对策论就是研究这些问题的数学理论和方法。限于篇幅,本节只讨论最基本的对策模型即矩阵对策。下面简要介绍一些对策论的基本概念及基本定理。

1. 矩阵对策

对策:对策是决策者在某种竞争场合下作出的决策,或者说是参加竞争的各方为了自己获胜采取的对付对方的策略。一场竞争或斗争简称一局对策。

对策三要素:(1) 局中人 对策的参与者叫局中人,它是一个有限集合,一般局中人数不少于两个。(2) 策略 局中人可采取的行动方案称为策略,每个局中人的策略全体,称为其策略集合。(3) 赢得函数 一局对策结束时,每个局中人的“得失”是全体局中人所取定的一组策略的函数,通常称为赢得函数。

局势:每个局中人在其策略集合中各取一个策略,组成的策略组,称作局势。于是,赢得函数实际上是局势集合上的函数,是一个局势的数量表示。

矩阵对策:如果局中人仅有两个,且每个局中人的策略集合是有限的,两个人的赢得之和为零,称此对策 G 为二人有限零和对

策,也即是矩阵对策。

若设局中人 I 有 m 个策略(即纯策略,以区别下文的混合策略) $\alpha_1, \alpha_2, \dots, \alpha_m$, 用 S_1 表示它们的集合:

$$S_1 = \{\alpha_1, \alpha_2, \dots, \alpha_m\} \quad (10.2.1)$$

同样,对局中人 II, 其纯策略集合为

$$S_2 = \{\beta_1, \beta_2, \dots, \beta_n\} \quad (10.2.2)$$

则,矩阵对策可记为:

$$G = \{S_1, S_2, A\} \quad (10.2.3)$$

其中

$$A = (a_{ij})_{m \times n} \quad (10.2.4)$$

称为局中人 I 的赢得矩阵(或局中人 II 的支付矩阵), a_{ij} 对应于纯局势 (α_i, β_j) 为局中人 I 在此纯局势(即纯策略组成的局势)下的赢得值。

对应于 S_1 和 S_2 , 我们共有 $m \times n$ 个纯局势, 只有给出赢得矩阵 $A = (a_{ij})_{m \times n}$ 之后, 才能进行对策。

2. 最优纯策略

对于矩阵对策 G :

$$G = \{S_1, S_2, A\} \quad (10.2.5)$$

若存在一局势 (α_i^*, β_j^*) , 使得对一切 $i = 1, 2, \dots, m; j = 1, 2, \dots, n$ 都有

$$a_{ij}^* \leq a_{i^*j^*} \leq a_{i^*j} \quad (10.2.6)$$

成立, 则称局势 (α_i^*, β_j^*) 为对策 G 在纯策略中的解, 也称为 G 的鞍点, 而 α_i^*, β_j^* 分别称为局中人 I 及局中人 II 的最优纯策略, $a_{i^*j^*}$ 称为对策 G 的值, 记为 V_G 。

定理 10.2.1 矩阵对策 $G = \{S_1, S_2, A\}$ 有最优纯策略的充分必要条件是

$$\max_i \min_j a_{ij} = \min_j \max_i a_{ij} = a_{i^*j^*} = V_G \quad (10.2.7)$$

并且满足上式的 a_i^*, j^* 的相应策略 α_i^* 和 β_j^* 必为局中 I 与局中人 II 的最优纯策略。

由此,我们得到求解最优纯策略的方法如下:

对于局中人 I 的赢得矩阵 A:

①局中人 I 对 A 的每一行取最小值 $\min_j a_{ij} (i = 1, 2, \dots, m)$, 再从这些最小值中取最大值, 记为 $V_1 = \max_i \min_j a_{ij}$ 。

②局中人 II 对 A 的每一列取最大值 $\max_i a_{ij} (j = 1, 2, \dots, n)$, 再从中取最小值, 记为 $V_2 = \min_j \max_i a_{ij}$ 。

由定理 10.2.1, 若 $V_1 = V_2$, 则该值即为对策的值, 其对应的行和列的策略即分别为局中人 I 和局中人 II 的最优纯策略。

应该指出的是, 若对策有鞍点, 则其在纯策略中有解且惟一, 若对策无鞍点, 则会有两种情况:

① $V_1 = V_2$, 即对策值存在, 则对策在纯策略中有解但不惟一, 此情形的解有两条性质: 第一, 无差别性, 即若 $(\alpha_k, \beta_l), (\alpha_t, \beta_r)$ 是两个解, 那么 $a_{kl} = a_{tr}$; 第二, 可交换性, 若 $(\alpha_k, \beta_l), (\alpha_t, \beta_r)$ 是两个解, 那么 $(\alpha_k, \beta_r), (\alpha_t, \beta_l)$ 也都是对策的解。

② $V_1 \neq V_2$, 则对策在纯策略中无解, 这是下文讨论的内容。

3. 混合策略与混合扩充

我们知道, 有些矩阵对策在纯策略中无解, 两个局中人也就没有最优纯策略, 面对这种情况, 局中人应如何选择纯策略参加对策呢? 这时, 就得估计选取各个纯策略可能性的大小来进行对策, 也就是用多大概率选取各个纯策略。

一般情况下有:

给定一个矩阵对策

$$G = \{S_1, S_2, A\} \quad (10.2.8)$$

$$S_1 = \{\alpha_1, \alpha_2, \dots, \alpha_m\}, S_2 = \{\beta_1, \beta_2, \dots, \beta_n\} \quad (10.2.9)$$

$$A = \{a_{ij}\}_{m \times n}$$

我们把局中人 I 及局中人 II 选取其各自纯策略集合中各纯策略的概率组成的概率向量

$$X = (x_1, x_2, \dots, x_m) \quad (x_i \geq 0, i = 1, 2, \dots, m, \sum_{i=1}^m x_i = 1) \quad (10.2.10)$$

与 $Y = (y_1, y_2, \dots, y_n) \quad (y_j \geq 0, j = 1, 2, \dots, n, \sum_{j=1}^n y_j = 1)$ (10.2.11)

分别称为局中人 I 与 II 的混合策略,也简称策略。这里 x_i 是局中人 I 选取 α_i 的概率, y_j 是局中人 II 选取 β_j 的概率。

特殊情况:

$$x_i = \begin{cases} 1 & i = k \\ 0 & i \neq k \end{cases} \quad y_j = \begin{cases} 1 & j = s \\ 0 & j \neq s \end{cases} \quad (10.2.12)$$

就是前面提到的纯策略情况。

为了定义局中人的赢得情况,我们称数学期望

$$E(X, Y) = \sum_{i=1}^m \sum_{j=1}^n a_{ij} x_i y_j \quad (10.2.13)$$

为局中人 I 的赢得, $-E(X, Y)$ 为局中人 II 的赢得,而 (X, Y) 称为混合局势。

局中人 I 的所有混合策略的集合记为 S_1^* ,

$$S_1^* = \{X\} \quad (10.2.14)$$

局中人 II 的所有混合策略的集合记为 S_2^* ,

$$S_2^* = \{Y\} \quad (10.2.15)$$

给定一个对策, $G = \{S_1, S_2, A\}$ 则

$$G^* = \{S_1^*, S_2^*, E\} \quad (10.2.16)$$

叫做 G 的混合扩充。

局中人采取混合策略 X 时,他只希望获得

$$\min_{Y \in S_2^*} E(X, Y) \quad (10.2.17)$$

因此局中人 I 应选取 $X \in S_1^*$ 使(10.2.17)式取最大,即局中人 I 保证自己的赢得不少于

$$\max_{X \in S_1^*} \min_{Y \in S_2^*} E(X, Y) = V_1 \quad (10.2.18)$$

同样局中人 II 的输出至多是

$$\min_{Y \in S_2^*} \max_{X \in S_1^*} E(X, Y) = V_2 \quad (10.2.19)$$

设 $G^* = \{S_1^*, S_2^*, E\}$ 是 $G = \{S_1, S_2, A\}$ 的混合扩充,如果

$$\min_{Y \in S_2^*} E(X^*, Y) = \max_{X \in S_1^*} E(X, Y^*) = V \quad (10.2.20)$$

则称 V 这个公共值为对策 G 的值,而混合局势 (X^*, Y^*) 称为 G 在混合策略下的解,而 X^* 与 Y^* 分别称为局中人 I 与局中人 II 的最优策略。

定理 10.2.2 对策的基本定理——任何一个给定的矩阵对策 G 一定有解(在混合扩充中的解)。

如果矩阵对策 G 的值是 V ,则以下两组不等式的解就是局中人 I 与 II 的最优策略:

$$\sum_{i=1}^m a_{ij} x_i \geq V \quad (j = 1, 2, \dots, n) \quad (10.2.21)$$

$$x_i \geq 0, \sum_{i=1}^m x_i = 1$$

$$\sum_{j=1}^n a_{ij} y_j \leq V \quad (i = 1, 2, \dots, m) \quad (10.2.22)$$

$$y_j \geq 0, \sum_{j=1}^n y_j = 1$$

定理 10.2.3 如果 (X^*, Y^*) 是对策 G 的最优混合局势,则对某一个 i 或 j 来说

$$\textcircled{1} \text{ 若 } x_i^* \neq 0, \text{ 则 } \sum_{j=1}^n a_{ij}y_j^* = V \quad (10.2.23)$$

$$\textcircled{2} \text{ 若 } y_j^* \neq 0, \text{ 则 } \sum_{i=1}^m a_{ij}x_i^* = V \quad (10.2.24)$$

$$\textcircled{3} \text{ 若 } \sum_{j=1}^n a_{ij}y_j^* < V, \text{ 则 } x_i^* = 0 \quad (10.2.25)$$

$$\textcircled{4} \text{ 若 } \sum_{i=1}^m a_{ij}x_i^* > V, \text{ 则 } y_j^* = 0 \quad (10.2.26)$$

这里 $V = E(X^*, Y^*)$

4. 矩阵对策的解法

(1) 特殊情况下的解法

如果已查明对策 G 没有鞍点, 根据定理 10.2.2 和定理 10.2.3, 若事先可判断出 x_i^*, y_j^* 均不为零, 就可以把定理 10.2.2 中的两组不等式化成线性方程组来求解。

赢得矩阵的阶数越高, 则求解越困难, 但是对某些特殊结构的矩阵, 可以把它降低阶数来求解。为此, 我们给出如下定义及定理。

设给定矩阵对策

$$G = \{S_1, S_2, A\} \quad \begin{aligned} S_1 &= \{\alpha_1, \alpha_2, \dots, \alpha_m\} \\ S_2 &= \{\beta_1, \beta_2, \dots, \beta_n\} \end{aligned}$$

$$A = (a_{ij})_{m \times n}$$

如果对于一切 $j, 1 \leq j \leq n$ 均有

$$a_{ij} \geq a_{kj} \quad (10.2.27)$$

则称局中人 I 的纯策略 α_i 优越于 α_k 。同样, 对于一切的 $i, 1 \leq i \leq m$ 均有

$$a_{ij} \leq a_{il} \quad (10.2.28)$$

则称局中人 II 的纯策略 β_j 优越于 β_l 。

定理 10.2.4 设 $G = \{S_1, S_2, A\}$ 是一个矩阵对策, 其中 $S_1 = \{\alpha_1, \alpha_2, \dots, \alpha_m\}, S_2 = \{\beta_1, \beta_2, \dots, \beta_n\}, A = (a_{ij})$, 如果 α_1 为其

余的纯策略 $\alpha_i (1 < i \leq m)$ 这一所优超, 由 G 可得到一个新的对策 G' 。

$$G' = \{S_1', S_2, A'\} \quad (10.2.29)$$

$$S_1' = \{\alpha_2, \alpha_3, \dots, \alpha_m\}, a_{ij}' = a_{ij}, \begin{cases} i = 2, 3, \dots, m \\ j = 1, 2, \dots, n \end{cases} \quad (10.2.30)$$

则 ① $V_{G'} = V_G$;

② G' 中局中人 II 的最优策略, 便是 G 中局中人 II 的最优策略;

③ 若 $(x_2^*, x_3^*, \dots, x_m^*)$ 是 G' 中局中人 I 的最优策略, 则 $(0, x_2^*, \dots, x_m^*)$ 是 G 中局中人 I 的最优策略。

对于某些矩阵对策, 把它的元素尽可能多地变成零, 如果对策的解集合不变, 则求解就比较简单。对此, 我们有如下定理:

定理 10.2.5 给定两个矩阵对策

$$G_1 = \{S_1, S_2, A_1 = (a_{ij})\} \quad (10.2.31)$$

$$G_2 = \{S_1, S_2, A_2 = (a_{ij} + d)\} \quad (10.1.32)$$

其中 d 是常数, 则两个对策的解集合不变, 其对策值相差一个 d , 即 $V_2 = V_1 + d$ (V_2 与 V_1 分别为 G_2 与 G_1 的对策值)。

(2) 线性规划解法

一般情况下, 我们知道, 对于扩充后的矩阵对策来说, 求最优解就是解下列两个不等式组:

$$\begin{cases} \sum_{i=1}^m a_{ij} x_i \geq V & (j = 1, 2, \dots, n) \\ \sum_{i=1}^m x_i = 1 \\ x_i \geq 0 & (i = 1, 2, \dots, m) \end{cases} \quad (10.2.33)$$

$$\begin{cases} \sum_{j=1}^n a_{ij}y_j \leq V & (i = 1, 2, \dots, m) \\ \sum_{j=1}^n y_j = 1 \\ y_j \geq 0 & (j = 1, 2, \dots, n) \end{cases} \quad (10.2.34)$$

这里 V 是:

$$V = \max_{x \in S_1^*} \min_{1 \leq j \leq n} \sum_{i=1}^m a_{ij}x_i \quad (10.2.35)$$

$$V = \min_{y \in S_2^*} \max_{1 \leq i \leq m} \sum_{j=1}^n a_{ij}y_j \quad (10.2.36)$$

作如下变换(不妨设 $V > 0$)

$$x_i' = \frac{x_i}{V} \quad (i = 1, 2, \dots, m) \quad (10.2.37)$$

于是(10.2.33) 式变成

$$\begin{cases} \sum_{i=1}^m a_{ij}x_i' \geq 1 & (j = 1, 2, \dots, n) \\ \sum_{i=1}^m x_i' = \frac{1}{V} \\ x_i' \geq 0 & (i = 1, 2, \dots, m) \end{cases} \quad (10.2.38)$$

这样,对局中人 I 来说,就把问题归结为求一组满足约束条件

$$\begin{cases} \sum_{i=1}^m a_{ij}x_i' \geq 1 & (j = 1, 2, \dots, n) \\ x_i' \geq 0 & (i = 1, 2, \dots, m) \end{cases} \quad (10.2.39)$$

的解 $x_i' (i = 1, 2, \dots, m)$ 使得目标函数

$$S(X') = \sum_{i=1}^m x_i' \quad (10.2.40)$$

达到最小。

同样,对于局中人 II 来说有

$$\begin{cases} \sum_{j=1}^n a_{ij}y_j' \leq 1 & (i = 1, 2, \dots, m) \\ y_j' \geq 0 & (j = 1, 2, \dots, n) \end{cases} \quad (10.2.41)$$

的一组解 $y_j' (j = 1, 2, \dots, n)$ 使得目标函数

$$S'(Y') = \sum_{j=1}^n y_j' \quad (10.2.42)$$

达到最大。这里

$$y_j' = \frac{y_j}{V} \quad (j = 1, 2, \dots, n) \quad (10.2.43)$$

显然, (10.2.39) 与 (10.2.40) 和 (10.2.41) 与 (10.2.42) 是线性规划的典型问题。

10.2.2 建模实例

1. 齐王与田忌赛马问题

战国时期, 齐国的国王有一天提出要与田忌赛马。双方约定: ① 各自出三匹马, ② 从上、中、下三个等级的马中各出一匹, ③ 每匹马都得参加比赛, 而且只参加一次, ④ 每次比赛各出一匹马, 一共比赛三次, ⑤ 每次比赛后负者要付给胜者千金。当时的情况是: 三种不同等级的马, 齐王的马均比田忌同等级的马强一些, 看来田忌要输掉三千金了。但是, 田忌手下一个谋士给田忌出了个主意: 每次比赛先让齐王说出他要出哪一匹马, 叫田忌用下马对齐王的上马(负), 用中马对齐王的下马(胜), 用上马对齐王的中马(胜)。比赛结果, 田忌二胜一负反得千金。

以齐王为局中人 I, 田忌为局中人 II, 双方对自己一方每匹马都规定一个上场参赛的次序, 也就是确定一种对策的策略即纯策略。显然, 每一方都有 $3! = 6$ 种纯策略。

齐王的纯策略集合

$$S_1 = \{\alpha_1, \alpha_2, \dots, \alpha_6\}$$

其中 $\alpha_1 = (\text{上}, \text{中}, \text{下}), \alpha_2 = (\text{上}, \text{下}, \text{中}), \alpha_3 = (\text{中}, \text{上}, \text{下}), \alpha_4 =$

(中,下,上), $\alpha_5 = (\text{下}, \text{中}, \text{上})$, $\alpha_6 = (\text{下}, \text{上}, \text{中})$

田忌的纯策略集合

$$S_2 = \{\beta_1, \beta_2, \dots, \beta_6\}$$

其中 $\beta_1 = (\text{上}, \text{中}, \text{下})$, $\beta_2 = (\text{上}, \text{下}, \text{中})$, $\beta_3 = (\text{中}, \text{上}, \text{下})$, $\beta_4 = (\text{中}, \text{下}, \text{上})$, $\beta_5 = (\text{下}, \text{中}, \text{上})$, $\beta_6 = (\text{下}, \text{上}, \text{中})$

由于在任一局势中,齐王所得总是等于田忌所失,显然,此赛马对策为一有限零和对策。此时,齐王的赢得表为:

表 10.2.1

齐王的赢得 齐王的策略 \ 田忌的策略		β_1	β_2	β_3	β_4	β_5	β_6
α_1		3	1	1	1	1	-1
α_2		1	3	1	1	-1	1
α_3		1	-1	3	1	1	1
α_4		-1	1	1	3	1	1
α_5		1	1	-1	1	3	1
α_6		1	1	1	-1	1	3

表中的 1 和 3 表示齐王得到的千金数,但也是田忌应交付的千金数。齐王的赢得表即田忌的交付表。由此,我们得到齐王的赢得矩阵。

$$A = \begin{matrix} & \begin{matrix} \beta_1 & \beta_2 & \beta_3 & \beta_4 & \beta_5 & \beta_6 \end{matrix} \\ \begin{matrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \\ \alpha_6 \end{matrix} & \begin{bmatrix} 3 & 1 & 1 & 1 & 1 & -1 \\ 1 & 3 & 1 & 1 & -1 & 1 \\ 1 & -1 & 3 & 1 & 1 & 1 \\ -1 & 1 & 1 & 3 & 1 & 1 \\ 1 & 1 & -1 & 1 & 3 & 1 \\ 1 & 1 & 1 & -1 & 1 & 3 \end{bmatrix} \end{matrix}$$

不难看出 $\max_i \min_j a_{ij} = -1 \neq \min_j \max_i a_{ij} = 3$, 即这个矩阵对策没有在纯策略中的解, 双方没有最优纯策略。现在求混合策略中的解, 由定理 10.2.2 知, 有下列两组不等式:

$$\begin{cases} 3x_1 + x_2 + x_3 - x_4 + x_5 + x_6 \geq V \\ x_1 + 3x_2 - x_3 + x_4 + x_5 + x_6 \geq V \\ x_1 + x_2 + 3x_3 + x_4 - x_5 + x_6 \geq V \\ x_1 + x_2 + x_3 + 3x_4 + x_5 - x_6 \geq V \\ x_1 - x_2 + x_3 + x_4 + 3x_5 + x_6 \geq V \\ -x_1 + x_2 + x_3 + x_4 + x_5 + 3x_6 \geq V \\ x_1 + x_2 + x_3 + x_4 + x_5 + x_6 = 1 \\ x_i \geq 0 \quad i = 1, 2, \dots, 6 \end{cases}$$

$$\begin{cases} 3y_1 + y_2 + y_3 + y_4 + y_5 - y_6 \leq V \\ y_1 + 3y_2 + y_3 + y_4 - y_5 + y_6 \leq V \\ y_1 - y_2 + 3y_3 + y_4 + y_5 + y_6 \leq V \\ -y_1 + y_2 + y_3 + 3y_4 + y_5 + y_6 \leq V \\ y_1 + y_2 - y_3 + y_4 + 3y_5 + y_6 \leq V \\ y_1 + y_2 + y_3 - y_4 + y_5 + 3y_6 \leq V \\ y_1 + y_2 + y_3 + y_4 + y_5 + y_6 = 1 \\ y_j \geq 0 \quad j = 1, 2, \dots, 6 \end{cases}$$

我们知道, 齐王与田忌各有六个纯策略。观察矩阵 A 的行和列各元素的值差异不大, 因而每个局中人选取每个纯策略的可能性都存在。也就是可以设想 x_i, y_j 均不为零, 所以对上述两组不等式都取等号, 并各自将所有式子相加得:

$$6(x_1 + x_2 + x_3 + x_4 + x_5 + x_6) = 6V$$

$$6(y_1 + y_2 + y_3 + y_4 + y_5 + y_6) = 6V$$

由 $\sum_{i=1}^6 x_i = 1, \sum_{j=1}^6 y_j = 1$, 故得 $V = 1$, 代入其他各式得:

$$x_i = \frac{1}{6} \quad y_j = \frac{1}{6}, \quad i, j = 1, 2, \dots, 6$$

齐王的最优混合策略是:

$$X^* = (\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6})$$

田忌的最优混合策略是:

$$Y^* = (\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6})$$

这与我们的设想相符合, 即双方都以 $\frac{1}{6}$ 的概率选取每个纯策略。换句话说, 每个纯策略被选中的机会是相等的。对策值 $V = 1$, 说明在整个比赛过程中, 双方都很“理智”, 总的结局仍然是齐王赢一千金。

当然只有齐王把某个纯策略选定之后, 而且告诉了田忌, 而田忌有针对性的选取自己的纯策略才能赢得一千金。因此, 给定一个对策如果有鞍点, 那么每个局中人都很“理智”, 则可以事先公开告诉对方自己选取的纯策略, 而结局仍不会改变。如果给定的对策没有鞍点, 则局中人双方要互相保密, 否则, 不保密的一方是要吃亏的。

2. 百货商店的位置问题

假设有一条街道, 街道居民对各种日用百货的需要是均匀分布的, 且总需要量为一常数。设有两家百货商店, 各欲沿街选定各自的店铺位置, 以竞争这一街道的市场, 问应如何选择商店位置才对己方商店最为有利?

以甲商店为局中人 I, 乙商店为局中人 II, 两商店都选定一个商店位置, 即确定一种对策的策略(纯策略)。为了便于讨论, 将整个街道分成五个等段(如图 10.2.1 所示), 并记各段的中点为

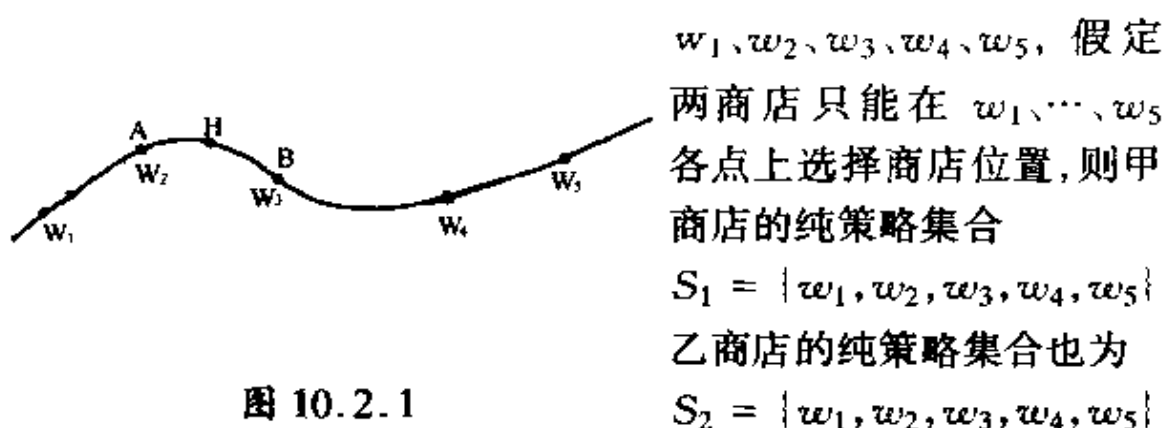


图 10.2.1

考虑到顾客一般愿意就近购买, 所以, 如果甲商店选择在 A 点, 乙商店选择在 B 点, 则两商店就各自独占了点 H (A, B 的中点) 向它一边的全部市场。若两家商店选择同一点, 它们就平分市场。如果设全街对百货的总的购买力为 10, 则利用初等数学知识, 容易算得甲店的赢得 (这里指甲店所占据的购买力) 如下表 (表 10.2.2)。表中的数字表示甲店所占据的购买力, 同时也是乙店在总购买力中所失去的购买力。如第二行、第一列交叉处“8”, 是指当甲店位于 w_2 , 且乙店位于 w_1 时, 甲店占据的购买力。

表 10.2.2

甲店占据的购买力 甲店地点		乙店地点				
		w_1	w_2	w_3	w_4	w_5
w_1		5	2	3	4	5
w_2		8	5	4	5	6
w_3		7	6	5	6	7
w_4		6	5	4	5	8
w_5		5	4	3	2	5

这也是一个二人零和对策问题 (这里按零和考虑是不十分确切的, 实际二人所得之和为 10, 但这并不妨碍我们按零和的情况

去处理,因为,甲店占据的购买力即为乙店失掉的购买力),甲店的赢得矩阵为

$$A = \begin{array}{ccccc} & w_1 & w_2 & w_3 & w_4 & w_5 \\ \begin{array}{c} w_1 \\ w_2 \\ w_3 \\ w_4 \\ w_5 \end{array} & \begin{bmatrix} 5 & 2 & 3 & 4 & 5 \\ 8 & 5 & 4 & 5 & 6 \\ 7 & 6 & 5 & 6 & 7 \\ 6 & 5 & 4 & 5 & 8 \\ 5 & 4 & 3 & 2 & 5 \end{bmatrix} \end{array}$$

不论对甲店,还是对乙店, w_2 、 w_3 、 w_4 三个策略都分别优越于 w_1 、 w_5 两个策略,由定理 10.2.4,上面的 5×5 对策可转化为以下的 3×3 对策。

$$A' = \begin{array}{ccc} & w_2 & w_3 & w_4 \\ \begin{array}{c} w_2 \\ w_3 \\ w_4 \end{array} & \begin{bmatrix} 5 & 4 & 5 \\ 6 & 5 & 6 \\ 5 & 4 & 5 \end{bmatrix} \end{array}$$

不难看出,上述对策有一个鞍点 $a_{33} = a'_{22} = 5$ 。因此从竞争的观点看,两家百货商店的位置都选择 w_3 ,即选择店位于街道中部,以平分市场秋色,这就是两商店在竞争中的最优策略。若这两家商店达成某种协议,共同垄断市场,如一家选位于 w_2 ,一家选位于 w_4 ,也可平分市场秋色,这种结盟的现象是可能的。但它不属于我们这里讨论的范畴。我们讨论的是竞争现象,在竞争中任何一方都想尽量取得最大的利益。我们得出的最优策略是竞争中的平衡。这个结论,无疑是符合实际的。只有如此,才可使消费者就近有不同商店挑选的余地,这是符合消费者的心理和习惯的。

3. 防空问题

甲、乙两个国家正处在一场战争的前夜。假定甲国有三种不同的防空系统: r_1 、 r_2 、 r_3 ;敌对的乙国有三种不同的军用机群: b_1 、

b_2, b_3 。根据情报和以往资料分析, 预计当乙国使用军用机群 b_j ($j = 1, 2, 3$), 而甲国动用防空系统 r_i ($i = 1, 2, 3$) 时, 甲国击落乙国飞机的概率如下面的矩阵 A 所示。

$$A = \begin{array}{cc} & \begin{array}{ccc} \text{乙} & & \text{国} \end{array} \\ \begin{array}{c} \left[\begin{array}{ccc} 0.1 & 0.4 & 0.4 \\ 0.4 & 0.1 & 0.6 \\ 0.3 & 0.2 & 0.5 \end{array} \right] \begin{array}{l} r_1 \\ r_2 \\ r_3 \end{array} \end{array} & \begin{array}{c} \text{甲} \\ \text{国} \end{array} \end{array}$$

例如, 若乙国使用 b_1 机群, 甲国使用 r_1 防空系统, 则甲国击落敌国飞机的概率为 0.1; 若乙国使用 b_3 机群, 甲国使用 r_2 系统, 则击落机群的概率为 0.6 等。

假定甲国在同一时间只能使用一种防空系统, 乙国的空军力量在同一时间也只能使用一种机群, 试确定两国各自的最优策略及平均击落飞机的概率。

显然, 这是矩阵对策问题。甲国为局中人 I, 乙国为局中人 II, 甲国的赢得矩阵即为上述的 A 矩阵, 甲国的纯策略集合为

$$S_1 = \{r_1, r_2, r_3\}$$

乙国的纯策略集合为

$$S_2 = \{b_1, b_2, b_3\}$$

分析矩阵 A , 可以看出, 对乙国而言, 纯策略 b_1, b_2 优越于纯策略 b_3 , 这样, 矩阵 A 的第 3 列可以从矩阵中删去, 简化成下面的 A_1 。

$$A_1 = \begin{array}{cc} & \begin{array}{cc} b_1 & b_2 \end{array} \\ \begin{array}{c} \left[\begin{array}{cc} 0.1 & 0.4 \\ 0.4 & 0.1 \\ 0.3 & 0.2 \end{array} \right] \begin{array}{l} r_1 \\ r_2 \\ r_3 \end{array} \end{array} \end{array}$$

不难看出, A_1 没有鞍点, 即此对策不存在最优纯策略。为了确定甲、乙双方的最优混合策略, 我们用线性规划求解。

设甲国的最优混合策略为 $P^* = (p_1, p_2, p_3)$, 乙国的最优混合策略为 $Q^* = (q_1, q_2, q_3)$, 又设 $p_i' = \frac{p_i}{V} (i = 1, 2, 3)$, $q_j' = \frac{q_j}{V} (j = 1, 2)$, 则此矩阵对策 (A_1) 可归结为以下的两组线性规划问题。

第一组:

$$\text{约束条件} \begin{cases} 0.1p_1' + 0.4p_2' + 0.3p_3' \geq 1 \\ 0.4p_1' + 0.1p_2' + 0.2p_3' \geq 1 \\ p_1', p_2', p_3' \geq 0 \end{cases}$$

$$\text{目标函数: } \min \frac{1}{V} = p_1' + p_2' + p_3'$$

第二组

$$\text{约束条件} \begin{cases} 0.1q_1' + 0.4q_2' \leq 1 \\ 0.4q_1' + 0.1q_2' \leq 1 \\ 0.3q_1' + 0.2q_2' \leq 1 \\ q_1', q_2' \geq 0 \end{cases}$$

$$\text{目标函数: } \max \frac{1}{V} = q_1' + q_2'$$

下面用表解法(请参阅有关线性规划的书籍)解第二组线性规划问题。

表 10.2.3(初表)

	q_1'	q_2'	z_1	z_2	z_3	b_i
①	0.1	0.4	1	0	0	1
②	0.4*	0.1	0	1	0	1
③	0.3	0.2	0	0	1	1
④	-1	-1	0	0	0	0

因为检验数(最后一行)中有负数,所以表 10.2.3 所对应的解不是最优解,需要变换。在第一列中找到轴心数为 0.4,作行初等变换($② \div 0.4 = ②'$; $① - ②' \times \frac{1}{10}$; $③ - ②' \times \frac{3}{10}$; $④ + ②'$)得表 10.2.4。

表 10.2.4

	q_1'	q_2'	z_1	z_2	z_3	b_i
①	0	$\frac{3}{8}^*$	1	$-\frac{1}{4}$	0	$\frac{3}{4}$
②	1	$\frac{1}{4}$	0	$\frac{5}{2}$	0	$\frac{5}{2}$
③	0	$\frac{1}{8}$	0	$\frac{3}{4}$	1	$\frac{1}{4}$
④	0	$-\frac{3}{4}$	0	$-\frac{5}{2}$	0	$\frac{5}{2}$

在第二列中取 $\frac{3}{8}$ 为轴心数,作初等变换($① \times \frac{8}{3} = ①'$; $② - ①' \times \frac{1}{4}$; $③ - ①' \times \frac{1}{8}$; $④ + ①' \times \frac{3}{4}$)得表 10.2.5。

表 10.2.5(终表)

	q_1'	q_2'	z_1	z_2	z_3	b_i
①	0	1	$\frac{8}{3}$	$-\frac{2}{3}$	0	2
②	1	0	$-\frac{2}{3}$	$\frac{8}{3}$	0	2
③	0	0	$-\frac{1}{3}$	$-\frac{2}{3}$	1	0
④	0	0	2	2	0	4

因表中检验数全部非负,所以得第一、二两组规划问题的最优解:

第二组: $q_1' = 2, q_2' = 2, \max \frac{1}{V} = 4$

第一组: $p_1' = 2, p_2' = 2, p_3' = 0$

从而得出对策问题(A_1)的最优解为

$$V = \frac{1}{4} = 0.25$$

$$P^* = (\frac{1}{2}, \frac{1}{2}, 0), Q^* = (\frac{1}{2}, \frac{1}{2})$$

原对策问题(A)的最优解为

$$V = 0.25$$

$$P^* = (\frac{1}{2}, \frac{1}{2}, 0), Q^* = (\frac{1}{2}, \frac{1}{2}, 0)$$

即甲国的策略是以对半机会使用防空系统 r_1, r_2 , 而不使用 r_3 系统; 乙国以对半机会使用 b_1, b_2 军用机群, 而不使用 b_3 机群, 这样甲国平均击落乙国飞机的概率为 0.25, 也就是将有 $\frac{1}{4}$ 的飞机被击落。

§ 10.3 动态规划方法建模

10.3.1 动态规划的基本知识

有不少动态过程可抽象成状态转移问题, 特别是多阶段决策过程的最优化, 如最短路问题, 资源的最优分配问题, 设备更新问题, 生产计划与存贮问题等。

动态规划所研究的对象是多阶段决策问题, 它的目的就是求一个策略, 使各阶段的效益总和达到最优。1951 年, 美国数学家 R. Bellman 等人, 提出了解决多阶段决策问题的最优化原理, 并研究了许多实际问题, 从而创建了动态规划。

动态规划方法的基本思想是: 将一个复杂系统分解成若干阶段, 每一个阶段系统的状态都有一个决策集合, 从中选择一个决

策,从而决定整个过程的策略。阶段往往用时间划分,这就具有“动态”的含义,然而,一些与时间无关的静态规划中的最优化决策,也可以人为地把问题分成若干阶段,作为一个多阶段决策问题来处理,求解过程分为两大步骤,先按整体最优化思想逆序地求出各个可能状态的最优化决策,再顺序地求出整个问题的最优策略和最优路线。

下面,结合一个求最短路的例子,来说明动态规划的一些基本概念。

例 1 求从 A 到 E 的最短距离。

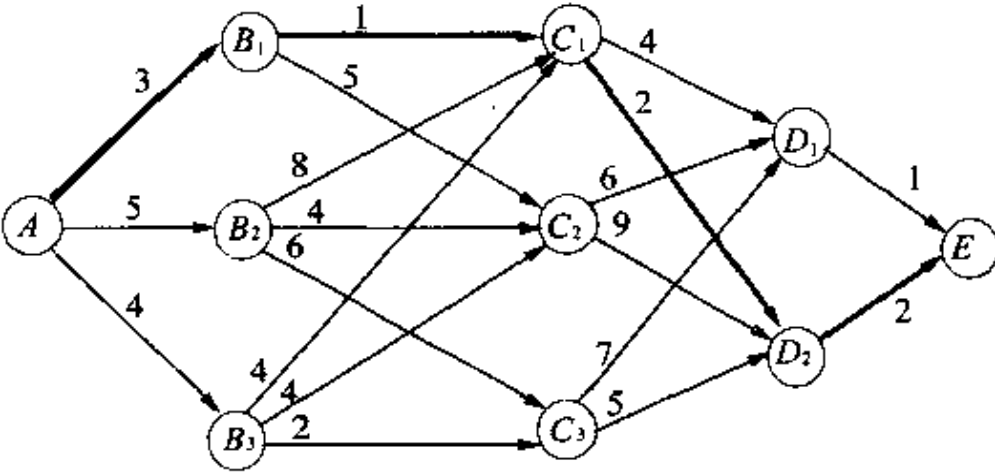


图 10.3.1

1. 多阶段决策过程

在上例中我们把最优路线问题,看作是能够分成若干阶段的一个过程,在过程的每一阶段都需要作出选择,决定究竟走哪一条支路,这些决定不同,则从 A 点到 E 点的路线也不同,即过程的发展不同。我们把这样的过程称为多阶段决策过程。

2. 阶段和阶段变量

在多阶段决策过程中,为了表示决策和过程的发展顺序,要引入阶段,一个阶段就是需要作出一个决策的子问题。阶段一般用阶段变量 k (或 n) 表示,它通常按照决策进行的时间或空间上的先后顺序划分。如上例 $k = 1, 2, 3, 4 \dots$

3. 状态和状态变量

状态是动态规划中最关键的一个参数,它即反映前面各阶段决策的结局,又是本阶段作出决策的出发点和依据,状态是动态规划问题各阶段信息的传递点和结合点。各阶段的状态通常用状态变量 x 来描述,如第 k 阶段的状态可用 x_k 表示。 x_k 可以是一个数,一组数或一个向量。如果状态为非数量表示,则可以给各个阶段的可能状态编号。如上例中的状态为点,第二个阶段的状态有三个点即 B_1, B_2, B_3 组成,可用 $x_2^{(1)} = B_1, x_2^{(2)} = B_2, x_2^{(3)} = B_3$ 表示。

4. 决策和决策变量

决策实质是关于状态的选择。阶段决策就是决策者从本阶段状态出发对下一阶段状态所做出的选择。即确定系统发展的方案。描述决策的变量为决策变量,用 $u_k(x_k)$ 表示第 k 阶段初始状态为 x_k 的决策变量,如上例中如果选择从 B_1 到 C_2 ,即用 $u_2(x_2^{(1)}) = C_2$ 表示。

5. 状态转移方程

系统在阶段 k 处于状态 x_k , 执行决策 $u_k(x_k)$ 的结果是系统状态的转移,即由阶段 k 的状态 x_k 转移到阶段 $k+1$ 的状态 x_{k+1} , 适用于动态规划方法求解的是一类具有无后效性的多段决策过程。无后效性又称马尔科夫性,指系统从某个阶段往后的发展,完全由本阶段所处的状态及其往后的决策决定,与系统以前的状态及决策无关,对于具有无后效性的多阶段决策过程,系统由阶段 k 到阶段 $k+1$ 的状态转移方程为 $x_{k+1} = T_k(x_k, u_k(x_k))$ (或称为转移规律) 由此式可见, x_{k+1} 只与 $x_k, u_k(x_k)$ 有关,而与前面状态无

关。

6. 阶段效应

在阶段 k 的状态 x_k 处,执行了决策 $u_k(x_k)$ 后,不仅带来系统状态的转移,而且也必然要对决策目标有确定的活动效果。衡量这种效果的形式,即为阶段效应,或称为阶段效益值。可用 $v_k(x_k, u_k(x_k))$ 表示,由此可见,阶段效应也只与 $x_k, u_k(x_k)$ 有关,而与以前的状态和决策无关。如上例中 $v_2(B_1, u_2(B_1) = C_1) = 1$ (或 $v_2(x_2^{(1)}, u_2(x_2^{(1)})) = C_1 = 1$)。

7. 目标函数

从阶段 k 到过程终止的效应即为目标函数,目标函数即是阶段 k 以后的各阶段效应的累积,这种累积可能是以后各阶段效应的加,减,乘,除,乘方,开方等不同形式,一般用 $f_k(x_k)$ 表示,至于目标函数是求最大或最小,由问题本身决定,如上例中即求 $f_k(x_k)$ 的最小值,即 $\min f_k(x_k)$ 。

8. 动态规划最优化原理

R. Bellman 指出:“作为整个过程的最优策略具有这样的性质:即无论过去的状态和决策如何,对前面的决策所形成的状态而言,余下的诸决策必须构成最优策略。”其含意即为:最优策略的任何状态后的部分子策略,都是相应于这个状态作为初始状态的最优策略,或每个最优策略只能由最优子策略构成。这个简单而直观的“最优化原理”是动态规划的理论基础。

9. 动态规划的数学模型

利用最优化原理,动态规划的数学模型可表示为:

$$\begin{cases} f_k(x_k) = \max(\text{或} \min) \{ v_k(x_k, u_k(x_k)) + f_{k+1}(x_{k+1}) \} \\ f_{k+1}(x_{k+1}) = 0 \quad k = n, n-1, \dots, 1 \end{cases} \quad (10.3.1)$$

这是一个由后向前的递推方程。

应用动态规划方法解决问题的关键是根据所给问题建立具体的动态规划模型,建立动态规划模型时的主要困难在于:如何将所遇到的最优化问题解释为合适的多阶段决策过程问题。划分阶段,选择状态变量和决策变量,写出状态转移方程,写出阶段效应和目标函数的递推方程,是动态规划模型化时的主要工作,其合适性决定应用动态规划的成败。

下面用动态规划方法具体求解上例中从 A 到 E 的最短路问题。

解:① 阶段: $k = 1, 2, 3, 4, n = 4$;

② 状态变量用 x_k 表示,表示节点;

③ 决策变量用 $u_k(x_k)$ 表示,表示从 k 阶段的某一节点到达下阶段的某节点;

④ 状态转移方程 $x_{k+1} = u_k(x_k)$;

⑤ 阶段效应 $v_k(x_k, u_k(x_k))$,表示从 k 阶段的某点到下阶段某点间的距离;

⑥ 目标函数的基本方程:

$$\begin{cases} f_k(x_k) = \min\{v(x_k, u_k(x_k)) + f_{k+1}(x_{k+1})\} & k = 1, 2, \dots, 4 \\ f_5(x_5) = 0 \end{cases}$$

具体计算时按逆序求解。

$$f_4(x_4^{(1)}) = f_4(D_1) = 1 \quad f_4(x_4^{(2)}) = f_4(D_2) = 2$$

$$f_3(x_3^{(1)}) = f_3(C_1)$$

$$\begin{aligned} &= \min \begin{cases} v_3(x_3^{(1)}, u_3(x_3^{(1)}) = D_1) + f_4(x_4^{(1)}) \\ v_3(x_3^{(1)}, u_3(x_3^{(1)}) = D_2) + f_4(x_4^{(2)}) \end{cases} \\ &= \min \begin{Bmatrix} 4 + 1 \\ 2 + 2 \end{Bmatrix} = 4 \end{aligned}$$

$$f_3(x_3^{(2)}) = \min \begin{cases} v_3(x_3^{(2)}, u_3(x_3^{(2)}) = D_1) + f_4(x_4^{(1)}) \\ v_3(x_3^{(2)}, u_3(x_3^{(2)}) = D_2) + f_4(x_4^{(2)}) \end{cases}$$

$$\begin{aligned}
&= \min \left\{ \frac{6+1}{9+2} \right\} = 7 \\
f_3(x_3^{(3)}) &= \min \left\{ \begin{array}{l} v_3(x_3^{(3)}, u_3(x_3^{(3)}) = D_1) + f_4(x_4^{(1)}) \\ v_3(x_3^{(1)}, u_3(x_3^{(3)}) = D_2) + f_4(x_4^{(2)}) \end{array} \right\} \\
&= \min \left\{ \frac{7+1}{5+2} \right\} = 7 \\
f_2(x_2^{(1)}) &= \min \left\{ \begin{array}{l} v_2(x_2^{(1)}, u_2(x_2^{(1)}) = C_1) + f_3(x_3^{(1)}) \\ v_2(x_2^{(1)}, u_2(x_2^{(1)}) = C_2) + f_3(x_3^{(2)}) \end{array} \right\} \\
&= \min \left\{ \frac{1+4}{5+7} \right\} = 5 \\
f_2(x_2^{(2)}) &= \min \left\{ \begin{array}{l} v_2(x_2^{(2)}, u_2(x_2^{(2)}) = C_1) + f_3(x_3^{(1)}) \\ v_2(x_2^{(2)}, u_2(x_2^{(2)}) = C_2) + f_3(x_3^{(2)}) \\ v_2(x_2^{(2)}, u_2(x_2^{(2)}) = C_3) + f_3(x_3^{(3)}) \end{array} \right\} \\
&= \min \left\{ \begin{array}{l} 8+4 \\ 4+7 \\ 6+7 \end{array} \right\} = 11 \\
f_2(x_2^{(3)}) &= \min \left\{ \begin{array}{l} v_2(x_2^{(3)}, u_2(x_3^{(3)}) = C_1) + f_3(x_3^{(1)}) \\ v_2(x_2^{(2)}, u_2(x_3^{(3)}) = C_2) + f_3(x_3^{(2)}) \\ v_2(x_2^{(2)}, u_2(x_3^{(3)}) = C_3) + f_3(x_3^{(3)}) \end{array} \right\} \\
&= \min \left\{ \begin{array}{l} 4+4 \\ 4+7 \\ 2+7 \end{array} \right\} = 8 \\
f_1(x_1) = f(A) &= \min \left\{ \begin{array}{l} v_1(x_1, u_1(x_1) = B_1) + f_2(x_2^{(1)}) \\ v_1(x_1, u_1(x_1) = B_2) + f_2(x_2^{(2)}) \\ v_1(x_1, u_1(x_1) = B_3) + f_2(x_2^{(3)}) \end{array} \right\} \\
&= \min \left\{ \begin{array}{l} 3+5 \\ 5+1 \\ 4+8 \end{array} \right\} = 8
\end{aligned}$$

计算结束后再按顺序决策。由上述计算可知,最优路线为 $A \rightarrow B_1 \rightarrow C_1 \rightarrow D_2 \rightarrow E$,最短路线为 8。

10.3.2 建模实例

1. 资源分配问题 I:

有某种资源,总数量为 x_1 吨,这些资源可用于两种形式的生产 I 和 II。并且已知收益函数分别为 $g(x)$ 和 $h(x)$,其中 x 为资源的投入量,且满足 $g(0) = h(0) = 0$ (即投入量为 0 时,收益也为 0)。假设这种资源用于生产后还可以回收一部分再用于生产。这两种生产方式 I 和 II 的回收率分别为 a 和 b ($0 \leq a \leq 1, 0 \leq b \leq 1$),我们的问题是:对总数量为 x_1 的资源进行 n 个阶段的生产,每一阶段如何分配投入 I 和 II 的资源数量,使总收益最大?

这个问题可用规划模型求解。设第 i 个阶段生产对第 I 种生产的投入量为 y_i ,第 i 个阶段生产后资源的回收总量为 x_{i+1} 。则有

$$\begin{aligned} \max \sum_{i=1}^n [g(y_i) + h(x_i - y_i)] \\ \begin{cases} x_{i+1} = ay_i + b(x_i - y_i) & i = 1, 2, 3, \dots, (n-1) \\ 0 \leq y_i \leq x_i \end{cases} \end{aligned} \quad (10.3.2)$$

对于该规划问题,当阶段 n 较大,变量个数较多,特别是收益函数为非线性函数时,求解将带来很大困难。如果用动态规划方法处理就很方便。

- ① 阶段按生产阶段划分, $k = 1, 2, 3, \dots, n$;
- ② 状态变量 x_k 表示 k 阶段初的资源拥有量;
- ③ 决策变量 $u_k(x_k) = y_k$ 为阶段 k 对第 I 种生产方式的投入量, $0 \leq y_k \leq x_k$;
- ④ 状态转移方程: $x_{k+1} = ay_k + b(x_k - y_k)$;
- ⑤ 阶段效应 $v_k(x_k, u_k) = g(y_k) + h(x_k - y_k)$;
- ⑥ 目标函数的基本方程:

$$f_k(x_k) = \max\{v_k(x_k, u_k(x_k)) + f_{k+1}(x_{k+1})\}$$

2. 资源分配问题 II

上面的资源分配问题是按生产阶段分配的。另一类资源分配问题是按生产单位,或产品,或生产方式分配资源的。这类问题中没有阶段,时间等因素,本属于静态问题。这时我们按生产单位,或产品,或生产方式来划分阶段,把静态问题化成动态问题来处理。

某工业部门,拟将某种高效率的设备五台分配给所属的甲、乙、丙三个工厂,各工厂获得这种设备之后,可以为国家提供利益如表 10.3.1。问如何分配这些设备,使国家获利最大?

表 10.3.1

利 益 设 备 权	工 厂	甲	乙	丙
0		0	0	0
1		3	5	4
2		7	10	6
3		9	11	11
4		12	11	12
5		13	11	12

解:① 按甲、乙、丙分成三个阶段, $k = 1, 2, 3$;

② 状态变量 x_k 表示第 k 阶段拥有的设备台数;

③ 决策变量 $u_k(x_k)$ 表示第 k 阶段分配给 k 的设备台数, 令 $u_k(x_k) = y_k, 0 \leq y_k \leq x_k$;

④ 状态转移方程 $x_{k+1} = x_k - y_k$;

⑤ 阶段效应, $v_k(x_k, u_k(x_k))$ 为第 k 阶段的利益, 即上表中的数据;

⑥ 目标函数的基本方程:

$$f_k(x_k) = \max\{v_k(x_k, y_k) + f_{k+1}(x_{k+1})\} \quad 0 \leq y_k \leq x_k$$

具体计算过程略,经计算得到两种最优分配方案:

1. 甲,0台,乙,2台,丙,3台
2. 甲,2台,乙,2台,丙,1台

最大利益为21。

以上两例中都是只有一种资源的分配问题称为一维资源分配,如果有两种资源同样有上述的类似分配问题,这类问题称为二维资源分配。

3. 二维资源分配问题

某部门现有300吨钢和30000吨煤,要分配给三个工厂,各厂得到原料后的利润如表10.3.2。

表 10.3.2 (单位 煤:万吨 钢:百吨 利润:十万元)

第一工厂	利 润	煤	0	1	2	3	第二工厂	利 润	煤	0	1	2	3	第三工厂	利 润	煤	0	1	2	3
	钢							钢							钢					
	0		0	1	3	6		0		0	2	4	6		0		0	3	5	8
	1		4	5	6	7		1		1	4	6	7		1		2	5	7	9
	2		5	6	7	8		2		4	6	8	9		2		4	7	9	11
	3		6	7	8	9		3		6	8	10	11		3		6	9	11	13

问如何分配这些资源,使该部门获利最大?

由于资源的数量较大,所以上表中将煤按万吨,钢按百吨为单位予以离散化。

- ① 按工厂的顺序号划分为三个阶段 $k = 1, 2, 3$;
- ② 状态变量 s_k, t_k 分别表示第 k 阶段钢和煤的拥有总量;
- ③ 决策变量 x_k, y_k 表示第 k 阶段分配给第 k 工厂钢和煤的数量;
- ④ 状态转移方程: $s_{k+1} = s_k - x_k, t_{k+1} = t_k - y_k$;
- ⑤ 阶段效应 $v_k(x_k, y_k, s_k, t_k)$ 表示第 k 阶段的利润,如表 10.3.2;
- ⑥ 目标函数的基本方程:

$$f_k(s_k, t_k) = \max \{ v_k(x_k, y_k, s_k, t_k) + f_{k+1}(s_{k+1}, t_{k+1}) \}$$

$$0 \leq x_k \leq s_k, 0 \leq y_k \leq t_k$$

经计算,最优方案为

$$\begin{cases} x_1 = 1 \\ y_1 = 0 \end{cases}, \begin{cases} x_2 = 2 \\ y_2 = 0 \end{cases}, \begin{cases} x_3 = 0 \\ y_3 = 3 \end{cases} \quad (10.3.3)$$

即分配给第一工厂钢 100 吨,煤 0 吨。

第二工厂钢 200 吨,煤 0 吨。

第三工厂钢 0 吨,煤 30000 吨。

最大利润为 160 万元。

4. 设备更新问题

建立满足下述条件的设备更新的最优策略,使总回收额最大。

条件如下:

(1) 设备的使用年限为 10 年。

(2) 役令为 t 时的设备年使用效益 $r(t)$ 和使用费用 $u(t)$ 如表 10.3.3。

表 10.3.3

t	0	1	2	3	4	5	6	7	8	9	10
$r(f)$ 万元	27	26	26	25	24	23	23	22	21	21	20
$u(t)$ 万元	15	15	16	16	16	17	18	18	19	20	20

(3) 设备的处理价格与役令无关,为 4 万元,新设备的价格为 13 万元。

① 按年划分阶段 $k = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10$

② 状态变量 t_k 表示第 k 阶段设备的役令为 t

③ 决策变量 $u_k(t_k)$ 表示只有 k (保留), p (更新) 两种。

④ 状态转移方程 $t_{k+1} = \begin{cases} t_k + 1 & u_k = k \\ 1 & u_k = p \end{cases}$

⑤ 阶段效应为阶段回收额 $k: r(t) - u(t)$

$$p: s(t) (\text{处理价格}) - p (\text{新设备价格}) + r(0) - u(0) = 3$$

⑥ 目标函数的基本方程

$$f_k(t_k) = \max \begin{cases} k: r(t) - u(t) + f_{k+1}(t_{k+1}) \\ p: 3 + f_{k+1}(1) \end{cases} \quad (10.3.4)$$

按逆序计算求解方程可得下表:

表 10.3.4

$f_k(t_k)$	t	0	1	2	3	4	5	6	7	8	9	10
$f_{10}(t_{10})$		12	11	10	9	8	6	5	4	3	3	3
$f_9(t_9)$		23	21	19	17	14	14	14	14	14	14	14
$f_8(t_8)$		33	30	27	24	24	24	24	24	24	24	24
$f_7(t_7)$		42	38	34	33	33	33	33	33	33	33	33
$f_6(t_6)$		50	45	43	42	41	41	41	41	41	41	41
$f_5(t_5)$		57	54	52	50	49	48	48	48	48	48	48
$f_4(t_4)$		66	63	60	58	57	57	57	57	57	57	57
$f_3(t_3)$		75	71	68	66	66	66	66	66	66	66	66
$f_2(t_2)$		83	79	76	75	74	74	74	74	74	74	74
$f_1(t_1)$		91	87	85	83	82	82	82	82	82	82	82

每期的保留与更新用折线隔开。折线左边的为保留,右边的为更新。

假设在阶段 1 有役令为 4 年的设备,制定以后 10 年的设备更新决策。

首先在 $f_1(t_1)$ 行与 $t = 4$ 列的交点上找到数 82,这表示总计

划期的最大回收额为 82 万元。第一年的最优决策是保留。到第二年设备役令为 5, 在 $f_2(t_2)$ 行与 $t = 5$ 列的交点是 74, 它所在的位置表明, 最优策略是更新, 如此继续, 直到 $f_{10}(t_{10})$ 行为止, 可得计划期间的最优策略是:

$(k, p, k, k, k, k, p, k, k, k)$

最优役令为

$(4, 5, 1, 2, 3, 4, 5, 1, 2, 3)$

最大回收额为 82 万元

如果第一年设备的役令为其他, 用上述方法可以得与其相应的最优策略。

5. 生产——存贮问题

用动态规划解决这类问题是比较成功的。它的中心思想就是在满足不同时期确定性的需求量的条件下, 在扩大生产批量或订货量所引起的生产成本的降低和存贮费用的上升之间作出合理选择。

问题: 某公司按协议在预定的四个月内售出某种产品。生产批量为 10 个单位, 最大生产能力为 100 个单位, 每单位存贮费用为 2 元, 生产成本和按协议月供货量, 如表 10.3.5。

表 10.3.5

月	生产成本 (元 / 单位)	销量 (单位)
1	70	60
2	72	70
3	80	120
4	76	60

现要求我们制定一个生产计划

① 满足协议要求;

② 在生产能力范围内;

③ 使生产成本最低。

解:① 按月划分四个阶段 $k = 1, 2, 3, 4$;

② 状态变量 x_k 为第 k 阶段初的存贮量;

③ 决策变量 $u_k(x_k)$ 为第 k 阶段产品生产量。令 $u_k(x_k) = y_k$, 令 e_k 表示第 k 阶段的产品成本, d_k 为第 k 阶段的协议销售量;

④ 状态转移方程: $x_{k+1} = x_k + y_k - d_k$;

⑤ 阶段效应: $v_k(x_k, y_k) = e_k y_k + 2x_k$;

⑥ 目标函数的基本方程:

$$f_k(x_k) = \min \{v_k(x_k, y_k) + f_{k+1}(x_{k+1})\}$$

经计算结果如下, 表 10.3.6。

表 10.3.6

月	月初存贮量	生产量	销售量	生产成本	存贮成本	总成本
1	0	100	60	7000	0	7000
2	40	100	70	7200	80	7280
3	70	50	120	4000	140	4140
4	0	60	60	4560	0	4560
合计						22980

习 题

1. 讨论泛函

$$J(y(x)) = \int_0^1 (xy + y^2 - 2y^2 y') dx$$

$$y(0) = 1, y(1) = 2$$

的极值。

2. 求泛函

$$J(y(x)) = \int_{x_0}^{x_1} A(x, y) e^{\arctg y'} \sqrt{1 + (y')^2} dx$$

的横截条件。

3. 在 10.1 节建模实例生产计划的制订这一问题中, 当 $Q < \frac{K_2 T^2}{4K_1}$ 时求最优解, 图 10.5 中 t_1 的确定可视为曲线 S_3 始端在直线 $x = 0$ 上变的泛函极值问题。

4. 求给定矩阵对策的最优策略和对策的值, 已知赢得矩阵为:

$$(1) A = \begin{pmatrix} 11 & 12 & 12 \\ 10 & 32 & 9 \end{pmatrix} \quad (2) A = \begin{pmatrix} -6 & 2 & 0 & 19 \\ 4 & 4 & 3 & 5 \\ -5 & -3 & -1 & -6 \end{pmatrix}$$

5. 求给定矩阵对策的最优混合策略和对策的值, 已知赢得矩阵为

$$(1) A = \begin{pmatrix} 10 & 0 & -8 \\ 18 & 6 & -12 \\ -2 & 6 & 4 \end{pmatrix} \quad (2) A = \begin{pmatrix} 1 & 0 & 1 & 3 & 2 \\ 6 & 4 & 2 & 13 & 7 \\ 5 & 6 & 1 & 8 & 6 \end{pmatrix}$$
$$(3) A = \begin{pmatrix} 4 & 8 & 6 & 0 \\ 8 & 4 & 6 & 8 \\ 4 & 8 & 8 & 0 \\ 8 & 0 & 0 & 16 \end{pmatrix} \quad (4) A = \begin{pmatrix} 0 & 4 & 1 & 3 \\ -1 & 3 & 0 & 2 \\ -1 & -1 & 4 & 1 \end{pmatrix}$$

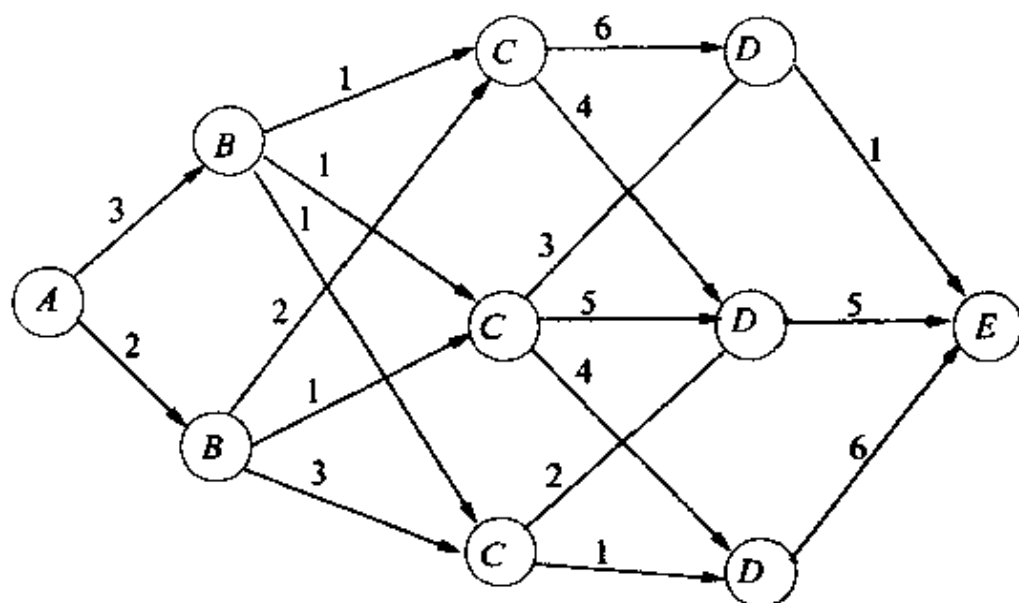
6. 甲方派两架轰炸机去袭击乙方的某个设施, 因每架飞机都带有巨大的杀伤武器, 所以只要一架飞机飞到目的地, 该设施肯定被摧毁。轰炸机可以从三个方面任选一方面接近目标。乙方可以将高射炮配置在三个方面中的任何一个方面。甲方有两个策略: k_1 ——两轰炸机各从一方面接近目标; k_2 ——两轰炸机从同一方面接近目标。乙方有三个策略: c_1 ——三个方面各配置一门炮;

c_2 ——一个方面配置两门炮,另一个方面配置一门炮; c_1 三门炮全配置在一个方面。其对策矩阵如下:

$$\begin{matrix} & c_1 & c_2 & c_3 \\ \begin{matrix} k_1 \\ k_2 \end{matrix} & \begin{bmatrix} 0 & \frac{2}{3} & 1 \\ 1 & \frac{2}{3} & \frac{2}{3} \end{bmatrix} \end{matrix}$$

求双方的最优策略。

7. 用动态规划求从 A 到 E 的最短路。



8. 某种机器可以在高低两种不同负荷下进行生产,在高负荷下进行生产时,产品的年产量由如下函数给出: $g(y) = 8y$,而机器的年折损率为 $a = 0.7$ (即如果年初时的完好机器数为 x 时,则到年终时完好机器数为 ax),如果在低负荷下进行生产时,产品的年产量由如下函数(其中 y 为投入生产的机器数)给出: $h(y) = 5y$,机器的年折损率为 $b = 0.9$ 。假如开始生产时机器总数为 $x_1 = 1000$ 台,要求制定一个五年计划,问如何确定每年按两种负荷投

人生产的机器数量,使得五年内产品的总产量最高。

9. 今拟将五名推销员安排到 A, B, C 三个地区中去, 已知对各地分配推销员后的经济效益如下表所示。用动态规划方法求推销员的最优配置计划。当推销员为四人时的情况是什么?

人 地区	0	1	2	3	4	5
A	38	56	83	100	82	75
B	40	58	82	99	100	100
C	60	78	109	125	125	120

10. 设备更新问题, 已知: ① 设备的使用年限是 10 年; ② 役令为 t 时的设备年使用效益 $r(t)$ 和使用费用 $u(t)$ 如下表所示, ③ 设备的处理价格 $s(t)$ 与役令无关为 0。新设备的价格为 8 万元。问:

(1) 关于役令为 7 的设备的十年最优更新策略。

(2) 关于役令为 6 的设备的九年最优更新策略是什么? 最大收益是多少?

t	0	1	2	3	4	5	6	7	8	9	10
$r(t)$	24	24	24	23	23	22	21	21	21	20	20
$u(t)$	13	14	15	15	17	17	17	18	19	19	19

11. 某厂已知其产品在未来四个月内的需求量如下表所示, 每批生产的准备费用是 500 元, 每件的生产费用是 1 元, 每件产品的库存费用是每月 1 元。假定 1 月初的库存量是 100 件, 4 月末的库存量必须为 0。建立这四个月的最优生产计划。

月份	需求量
1	400
2	500
3	300
4	200

参考文献

- [1] 龚得恩等译, 动态经济学、方法和模型, 中国人民大学出版社, 1992。
- [2] 任善强, 数学模型, 重庆大学出版社, 1987。
- [3] 陈义华, 数学模型, 重庆大学出版社, 1995。
- [4] 姜启源, 数学模型, 高等教育出版社, 1993。
- [5] 周概容, 概率论与数理统计, 高等教育出版社, 1984。
- [6] 李德, 钱颂迪, 运筹学, 清华大学出版社, 1985。
- [7] 许树柏, 层次分析原理, 天津大学出版社, 1988。
- [8] 王连芬, 层次分析法引论, 中国人民大学出版社, 1990。
- [9] 李修睦, 图论导引, 华中工学院出版社, 1982。
- [10] 赵宏量, 图论基础教程, 西南师范大学出版社, 1988。
- [11] (英) B. 布鲁巴斯, 图论导引教程, 黑龙江科技出版社, 1988。
- [12] 叶其孝, 大学生数学建模竞赛辅导教材, 湖南教育出版社, 1994。
- [13] 叶其孝, 数学建模教育与国际数学建模竞赛“工科教学”专辑, 1995。
- [14] 赵凯华, 定性与半定量物理学, 高等教育出版社, 1991。
- [15] 王梦光, 刘继印, 线性规划与整数规划, 东北大学出版社, 1993。