〔美〕 P·Lax　　S·Burstein　　A·Lax 著

微积分及其应用与计算

第一卷　第二册

人民教育出版社

微积分及其应用与计算

第一卷　第二册

P. Lax
[美] S. Burstein　著
A. Lax

唐述钊　黄开斌　滕振寰　黄禄平
卢绮龄　苏煜城　黄　敦　林应举　译

胡祖炽　何旭初　黄　敦　卢绮龄　校

人民教育出版社

微积分及其应用与计算

第一卷 第二册

P. Lax
[美] S. Burstein 著
A. Lax

唐述钊 黄开斌 滕振寰 黄禄平
卢绮龄 苏煜城 黄 敦 林应举 译

胡祖炽 何旭初 黄 敦 卢绮龄 校

*

人民教育出版社出版
新华书店北京发行所发行
人民教育出版社印刷厂印装

*

开本 850×1168 1/32 印张 8.75 字数 204,000
1980 年 11 月第 1 版 1981 年 7 月第 1 次印刷
印数 00,001—20,500
书号 13012·0535 定价 0.77 元

目　录

第六章　概率论及其应用

微积分起源于牛顿力学，在§3.6 中已简单地介绍了牛顿力学. 在地球与月球运动的例子里已看到：一旦指定了质点的初始位置与速度并确定了作用于质点的力，那末质点未来的整个路径便是可预测的了. 这样一种可预测的运动称为是*确定性的*. 在§8.1 中我们将看到在恢复力与摩擦力的合力作用下质点的运动同样也是确定性的. 事实上，指定了作用力而按照牛顿定律运动的任何质点系都作可预测的运行路径. 另一方面，当作用于质点的力不能精确地确定以至近似地确定时，或者当质点的初始位置与速度是我们所不能控制的乃至所不能观测时，那末质点的路径就决不是可预测的. 日常生活中观察到的许多运动，可说几乎所有的运动都是属于这一类的. 典型的一些例子是烟的飘荡、天空中云的飘移、骰子的投掷、纸牌的洗牌与发牌. 这样的一种不可预测的运动称为*非确定性的*或*随机的*.

即使单独掷一次骰子所得结果是不可预测的，但一长串的平均结果是完全可以预测的，至少对标准的骰子而言，大量的投掷次数中每个数字的出现大概是六分之一. 同样，如果一付 52 张的纸牌我们重复洗牌和发出最上的一张牌，每张牌大概出现发牌次数的 1/52 次. 用某些类型的云形成的经验可以指出五次中三次有雨.

概率论是这样的一个数学分支，它讨论的事件其单个的结果是不可预测的，但平均起来事件的结果是可预测的. 在下两节我们将讲述概率的法则，在下三节将对一些特殊的情形应用这些法则. 读者会看到在这些应用中微积分的概念和方法起很重要的作

用，尤其是对数函数与指数函数处处出现．正因为如此所以本书包含了这一章．

6.1 离散概率

我们将考虑一些简单试验，其中最简单的是掷一粒骰子，洗一付牌并发出最上面的一张牌，掷一枚钱币．稍为实际的例子是做一种物理试验．一个试验分两步，即安排试验和观察它的结果．在许多情形下，例如气象学、地质学、海洋学，我们不能安排试验，而只能观察试验．

我们将讨论可重复的和非确定性的试验．可重复的意义是试验可重复安排任意多次；非确定性的意义是任何单独一次试验可产生种种结果．本节开始所提到的那几个简单的例子中种种可能的结果分别是：1 和 6 之间的一个整数，52 张牌中的任一张，钱币的正面与反面．这节我们将讨论的试验，跟上面一样，它们只有有限多个可能的结果．我们用 n 表示可能结果的个数，并把它们由 1 到 n 标号．

最后我们假定在任何单个情况下其结果不可预测的试验平均起来其结果是可预测的，但这是说要重复试验无限多次．用 S_j 表示在前 N 次试验中第 j 个结果出现的次数，那末第 j 个结果出现的频率 S_j/N 当 N 趋于无穷时它趋向一个极限．我们称这个极限为第 j 个结果的概率，并记为 p_j:

(1.1) $$p_j=\lim_{N\to\infty}\frac{S_j}{N}.$$

这些概率具有下列的性质：

(i) 每个概率 p_j 是 0 与 1 之间的一个实数：

(1.2) $$0\leqslant p_j\leqslant 1.$$

(ii) 所有的概率之和等于 1：

(1.3) $$p_1+p_2+\cdots+p_n=1.$$

这两个性质都能由(1.1)推出，因为 S_j/N 是在 0 与 1 之间的，因而它的极限也如此，这就证明了(1.2)．另一方面，总共有 n 个可能的结果，因而所作试验序列的头N个结果中的每一个都是那 n 个情形之一．因为 S_j 是头N次中第 j 个结果出现的次数，所以

$$S_1+S_2+\cdots+S_n=N.$$

用N去除我们得到

(1.4) $$\frac{S_1}{N}+\frac{S_2}{N}+\cdots+\frac{S_n}{N}=1.$$

现令N趋于无穷大，根据(1.1)，S_1/N 的极限是 p_1，S_2/N 的极限是 p_2，等等．在 § 1.4 中我们已经知道 n 个数列之和的极限是各个数列的极限之和．把这个结论应用到(1.4)我们得到

$$\lim_{N\to\infty}\left\{\frac{S_1}{N}+\frac{S_2}{N}+\cdots+\frac{S_n}{N}\right\}=\lim_{N\to\infty}\frac{S_1}{N}+\lim_{N\to\infty}\frac{S_2}{N}+\cdots+\lim_{N\to\infty}\frac{S_n}{N}=1.$$

因而根据(1.1)

$$p_1+p_2+\cdots+p_n=1.$$

这就是结论(1.3)．

我们有时候而实际上是常常不去注意试验结果的详细情况，而只注意它的一个特定方面．例如，在抽取一张牌时可能只注意它属于哪种花色；在掷一粒骰子时我们可能只注意结果是偶数点还是奇数点．象抽得的一张牌是黑桃，或掷出的是偶数点这样的结果称为一事件．一般地，我们定义可能结果的任一集合为一个事件 E．这样抽得一张黑桃便是抽得一张黑桃二、黑桃三，等等一直到抽得一张黑桃A这些结果的集体名字了．同样，掷出一个偶数点是掷出二点、四点或六点的集体名字了．

我们用类似于定义一个结果的概率的方法来定义事件E的概率：

(1.5) $$p(E)=\lim_{N\to\infty}\frac{S(E)}{N},$$

其中 $S(E)$ 是在头 N 次试验中事件 E 出现的次数．容易证明这个极限存在．事实上，容易给出 $p(E)$ 的一个公式．因为，根据定义，当试验结果属于由组成事件 E 的那些可能结果的集合时，事件 E 发生．因此，事件 E 发生的次数 $S(E)$，是当 j 为组成 E 的 j 时所有 S_j 的和：

(1.6) $$S(E)=\sum_{j\text{在}E\text{中}}S_j.$$

用 N 去除：

(1.7) $$\frac{S(E)}{N}=\sum_{j\text{在}E\text{中}}\frac{S_j}{N},$$

这个关系式说明数列 $S(E)/N$ 是 j 在 E 中的那些数列 S_j/N 之和．按照在推导(1.3)已用过的事实，即，一些数列之和所成的数列之极限是各个数列的极限之和．因为根据(1.1)

$$\lim_{N\to\infty}\frac{S_j}{N}=p_j,$$

当 N 趋于∞时取极限我们由(1.7)导出

(1.8) $$p(E)=\sum_{j\text{在}E\text{中}}p_j.$$

事件 E_1 与 E_2 称为*互斥的*，如果它们是不能同时发生的，即，构成事件 E_1 的那些结果所成的集合和构成事件 E_2 的那些结果所成的集合没有公共元素．下面是互斥事件的一些例子：

例 1　E_1 = 抽得一张黑桃；E_2 = 抽得一张红心．

例 2　E_1 = 掷得一偶数点；E_2 = 掷得 3 点．

我们定义*两事件* E_1 *与* E_2 *之和*为或者 E_1 发生或者 E_2 发生那个事件，记为 E_1+E_2，即，构成 E_1+E_2 的那些结果是构成 E_1 的结果或构成 E_2 的那些结果．

下列的结论是既简单又重要：

两个互斥事件之和的概率是两个事件的概率之和：

(1.9) $$p(E_1+E_2)=p(E_1)+p(E_2).$$

这称为互斥事件的加法法则. 这个结果是由事件的概率公式(1.8)推出的. 因为根据和的定义,

$$p(E_1+E_2)=\sum_{j\text{ 在 }E_1\text{ 或 }E_2\text{ 中}} p_j.$$

另一方面, 互斥的意义是说任一个 j 可能属于 E_1 也可能属于 E_2, 但不能同时属于 E_1 和 E_2. 因而

$$\begin{aligned}p(E_1+E_2)&=\sum_{j\text{ 在 }E_1\text{ 或 }E_2\text{ 中}} p_j=\sum_{j\text{ 在 }E_1\text{ 中}} p_j+\sum_{j\text{ 在 }E_2\text{ 中}} p_j\\&=p(E_1)+p(E_2).\end{aligned}$$

这正是(1.9)所断定的.

现在看概率论中另一个重要的概念, 即两个试验的独立性. 作两个试验, 例如掷一粒骰子与洗一付牌并发最上面的一张牌. 我们的常识以及对于规律的一切了解告诉我们, 所谓两个试验是完全相互独立的, 是说一个试验的结果不可能影响另一个试验的结果, 而且两个试验的结果也不受某种公共原因的影响. 现在我们用概率论的语言来确切地叙述独立性的一个重要结论.

给定任意两个试验, 我们可把它们结合成单个的联合试验, 即同时作这两个试验. 设 E 是一个试验结构中的事件, F 是另一个试验结构中的事件. 两个事件 E 与 F 都发生的联合事件记为 EF. 譬如, E 是掷出偶数点这个事件而 F 是抽得一张黑桃, EF 是掷出偶数点并抽得一张黑桃这一事件. 我们断言: 如两个试验是独立的, 那末联合事件 EF 的概率是事件 E 与 F 各别的概率之积:

(1.10) $$p(EF)=p(E)p(F)$$

我们把这个关系叫做独立试验的乘法法则.

现在来阐明如何推导乘法法则. 想象重复联合试验无穷多次, 我们考察这个序列中前 N 个试验. 在前 N 次中, 计算 E 出现的次

数，F 出现的次数以及 EF 出现的次数．我们用 $S(E)$，$S(F)$ 以及 $S(EF)$ 表示这些数．根据一事件的概率的定义(1.5)

$$p(E)=\lim_{N\to\infty}\frac{S(E)}{N}, \tag{1.11$_E$}$$

$$p(F)=\lim_{N\to\infty}\frac{S(F)}{N}, \tag{1.11$_F$}$$

$$p(EF)=\lim_{N\to\infty}\frac{S(EF)}{N}. \tag{1.12}$$

设我们从联合试验序列中挑出 E 出现的试验所成的子序列，在这个子序列中 F 出现的频率是 $S(EF)/S(E)$．如果 E 与 F 两个事件真正是独立的，F 在这个子序列中出现的频率和 F 在原始的序列中出现的频率应是一样的．因此，

$$\lim_{N\to\infty}\frac{S(EF)}{S(E)}=\lim_{N\to\infty}\frac{S(F)}{N}=p(F). \tag{1.13}$$

现在我们把数列 $S(EF)/N$ 写成积的形式：

$$\frac{S(EF)}{N}=\frac{S(EF)}{S(E)}\cdot\frac{S(E)}{N}.$$

按照在§1.4中所讲的收敛数列的运算，两个数列之积的极限等于它们的极限之积：

$$\lim_{N\to\infty}\frac{S(EF)}{N}=\lim_{N\to\infty}\frac{S(EF)}{S(E)}\cdot\lim_{N\to\infty}\frac{S(E)}{N}. \tag{1.14}$$

按照(1.12)，左端的值是 $p(EF)$，而按照(1.13)与 $(1.11)_E$ 右端的乘积是 $p(F)p(E)$，因此由(1.14)便得到乘法法则(1.10)．

假设一种试验有 m 个可能的结果，另一种试验有 n 个可能的结果．它们的概率分别记为 $p_1,\cdots,p_m$ 与 $q_1,\cdots,q_n$，联合试验便有 mn 个可能的结果，即，所有的结果对 (j,k)．如果两种试验是独立的，那末乘法法则(1.10)告诉我们联合试验的结果 (j,k) 具有概率

$$p_jq_k. \tag{1.15}$$

这个公式在概率论中起很重要的作用，现在来说明它的用处．

设我们现在要讨论的两个试验都是掷一枚骰子，那末联合试验便是掷一对骰子．每个试验都有6个可能的结果，而其概率均为1/6．根据公式(1.10)，有36个联合结果，每个均具概率1/36．现问：掷得7点这个事件的概率是多少？显然有六种方法可掷得7点：(1,6)，(2,5)，(3,4)，(4,3)，(5,2)，(6,1)．根据公式(1.8)，掷得7点的概率是构成这个事件的六个结果的概率之和．这个和数是

$$6\times\frac{1}{36}=\frac{1}{6}.$$

同样我们能计算掷得2与12之间任一点数的概率．希望读者完成算出掷得2，3，…，12各点的概率．其结果如下表：

(1.16)

掷得的点　数	2	3	4	5	6	7	8	9	10	11	12
概　率	1/36	1/18	1/12	1/9	5/36	1/6	5/36	1/9	1/12	1/18	1/36

现在我们来看概率论的另一个重要概念，试验的**数值结果**．在用来测量单个物理量的物理试验中，数值结果就是对问题中的量所测得的值．对掷一对骰子这个简单例子来说，数值结果可以是每个骰子面上的点数之和．对发一手桥牌的试验来说，数值结果可以是这手桥牌的点数*．一般地，**一种试验的数值结果的意义是对每个可能的结果给一个实数** x_j，$j=1,2,\cdots,n$．如上面那些例子所表明的那样，到底给些什么数值那是试验者的事，并且要由他希望作这个试验所起的作用以及他作这试验的有关理论所决定．

如果试验是非确定性的，在试验的任何**单独一次**的实施中其结果是不可预测的．然而想象该试验实施了无穷多次，并把它们的结果按出现的顺序排下来：

* 一手桥牌13张的“点”数，是根据一定计算法则算出的点数，它是衡量这手牌优劣的一种标准．——译者注

$$j_1, j_2, \cdots, j_N, \cdots.$$

每个 j_N 是 1 与 n 之间的一个整数. 也把数值结果排起来:

(1.17) $$a_1, a_2, \cdots, a_N, \cdots.$$

第N个数值结果 a_N 当然是

(1.18) $$a_N = x_{j_N}.$$

在这一节中我们考虑的试验，虽然单独一个是不可预测的但平均起来是可预测的. 现在我们证明在这种情况下，数值结果所成的数列(1.17)有一个平均值$\bar{x}$，它定义为当N趋于无穷时头N个数值结果的算术平均数的极限:

(1.19) $$\bar{x} = \lim_{N\to\infty} \frac{a_1 + a_2 + \cdots + a_N}{N}.$$

再者，我们将证明平均数值结果 $\bar{x}$ 和数值结果 x_j 以及它们的概率 p_j 的关系是由下列简单的公式所给出:

(1.20) $$\bar{x} = p_1 x_1 + \cdots + p_n x_n.$$

$\bar{x}$ 称为数 $x_1, x_2, \cdots, x_n$ 的加权平均数，$p_1, p_2, \cdots, p_n$ 称为权.

公式 (1.20) 容易证明. 用 S_j 表示在头N次中第 j 个结果出现的次数. 因而根据公式(1.18)在 a_1, a_2, $\cdots$, a_N 中有 S_1 个等于 x_1, S_2 个等于 x_2, $\cdots$, S_n 个等于 x_n. 于是

$$a_1 + a_2 + \cdots + a_N = S_1 x_1 + S_2 x_2 + \cdots + S_n x_n.$$

用N去除得到

(1.21) $$\frac{a_1 + \cdots + a_N}{N} = \frac{S_1}{N} x_1 + \cdots + \frac{S_n}{N} x_n,$$

再应用 §1.4 所讲的关于收敛数列和的法则: 收敛数列之和的极限等于各数列之极限的和. 对数列

$$\frac{S_1}{N} x_1, \quad \frac{S_2}{N} x_2, \quad \cdots, \quad \frac{S_n}{N} x_n$$

的和应用上述法则，由(1.1)知这些数列的极限是 $p_1 x_1, p_2 x_2, \cdots, p_n x_n$，因此它们的和的极限是 $p_1 x_1 + \cdots + p_n x_n$. 由于(1.21)我们得

到(1.20).

现给出平均数值结果的公式(1.20)的一个例子. 作掷一对骰子的试验，按掷得 2, 3, …, 12 点进行分类，我们取这些数作为试验的数值结果，每个结果的概率在表(1.16)中给出. 利用这个表和公式(1.20)，我们得到掷一对骰子的平均数值结果为下列的值:

$$\bar{x}=\frac{1}{36}2+\frac{1}{18}3+\frac{1}{12}4+\frac{1}{9}5+\frac{5}{36}6+\frac{1}{6}7+$$
$$+\frac{5}{36}8+\frac{1}{9}9+\frac{1}{12}10+\frac{1}{18}11+\frac{1}{36}12=7.$$

习　　题

1.1. 设E是由一试验的某一组结果所构成的事件，从我们所关心的事件来看称这些结果是有利的. 所有不利的结果的全体，即不属于E的结果的全体，称为E的对立事件，它常记作$\bar{E}$. 证明

$$p(E)+p(\bar{E})=1.$$

1.2. 根据公式(1.15)，两个独立试验的联合所得的结果(j, k)的概率是

$$p_j q_k.$$

证明所有这些概率之和为 1.

1.3. 如每当E发生时F也发生，就说事件E是包含在事件F之中. 表达这个关系的另一种说法是：组成E的那些结果是组成F的结果的一个子集. “事件E包含在事件F之中”这句话用记号$E\subset F$来表达. 例如，抽得一张黑桃的事件E是包含在抽得一张黑牌这个事件F之中的. 证明，如$E\subset F$，则

$$p(E)\leqslant p(F). \tag{1.22}$$

1.4. 设$E_1, E_2, E_3, \cdots, E_m$是$m$个事件，它们是互斥的，意即没有一个结果能属于一个以上的事件. 用E表示事件E_i之和:

$$E=E_1+\cdots+E_m.$$

证明加法法则成立:

$$p(E)=p(E_1)+\cdots+p(E_m).$$

1.5. 利用一个伪随机数产生器去模拟掷一对骰子所得的结果数列. 看看头一千次投掷的平均接近 7 的程度如何.

[提示：大多数计算机都有生成伪随机数的功能．设 p 是一个在区间

$$0\leqslant p\leqslant 1$$

上均匀分布的随机数．把这个区间分成 6 等分．规定与第一个区间中的数相应取值 $x_1=1$，与第二区间中的数相应取值 $x_2=2$，…，等等，而与第 6 个区间的数相应取值 $x_6=6$，生成一对随机数 (p_j, p_k) 并规定这次实验的结果为和 x_j+x_k．]

6.2 信息论或感兴趣的事有多有趣

有些信息是不引人入胜的，有些信息是有趣的，这是人类普遍的经验．人咬狗是一奇闻，狗咬人便不是．在这一节里，我们讲一种方法对信息的价值给予一种量的量度．

在这个讨论中，“信息”的意义是得知其出现为随机的某一事件 E 已经发生．§6.1 中定义过，一个事件是一试验的一些可能结果所成的一个集合．在大量的试验中事件 E 出现的频率就是它的概率 $p(E)$．在这个理论中，我们假定知道从一个事件已经发生所得的信息只依赖于该事件的概率 p．于是我们用 $f(p)$ 表示所得的信息，换句话说，我们可把 $f(p)$ 想象为由事件出现所产生的意外成分的一种量度．

这个函数 f 有一些什么性质呢？我们说必须有下列三个性质：

(i) 当 p 递减时 $f(p)$ 递增，

(2.1) (ii) $f(1)=0$，

(iii) 当 p 趋于 0 时 $f(p)$ 趋于 ∞。

性质 (i) 表达这样的事实：较小可能事件的出现比较大可能事件的出现是更为意外，因而会带来较多的信息．性质(ii) 说明一个几乎必然事件的出现没有给出新的信息．性质(iii)说明一稀有事件的出现最有意义而且提供许多新的信息．

下面我们来推导函数 f 的一个关键的性质．设两个事件 E 与 F 是独立的，因为这样的事件是完全无关连的，所以，知道它们两

者都出现并不比知道每个分别出现会给出更多的信息，即，由知道两事件都发生所得的信息是分别知道两个事件发生所得的信息之和．分别用 p 与 q 表示事件 E 与 F 的概率．按照乘法法则(1.10)联合事件 EF 的概率是乘积 pq．于是上面所讲的那个规律可用下列方程来表达：

$$f(pq)=f(p)+f(q). \tag{2.2}$$

在§5.2中我们已看到对数函数满足上列函数方程，反过来，满足函数方程(2.2)的函数只是 $\log p$ 的一个常数倍，k 倍．因此我们得到结论

$$f(p)=k\log p. \tag{2.3}$$

这个常数的值是什么呢？按照(2.1)的性质(i)，$f(p)$ 随 p 的递减而递增．因为 $\log p$ 是随 p 的递增而递增，我们得出这个常数必定是负的结论．它的大小怎样呢？不首先采用信息的任一种单位的话，是没有办法确定它的大小的．为了方便起见，我们选取这个常数为 -1，因而

$$f(p)=-\log p. \tag{2.4}$$

关于(2.1)的性质(ii)和性质(iii)怎样呢？我们断言由(2.4)所定义的 f 是满足这两个性质的．我们把对它们的验证留给读者作为练习．

现在考虑有 n 个分别具有概率 $p_1, p_2, \cdots, p_n$ 的可能结果的一个试验．如果在单独一次试验中第 j 个结果发生，按照公式(2.4)，我们所得信息的量是 $-\log p_j$．现在我们问：如果我们重复做多次试验，那末平均的信息量是什么？这个问题的答案包含在§6.1中关于一系列试验的平均数值结果的公式(1.20)之中．按照那个公式，如果第 j 个数值结果是 x_j，那末平均数值结果是 $p_1x_1+\cdots+p_nx_n$．在现在的情形，我们取在第 j 个结果中所得的信息作为数值结果：

$$x_j = -\log p_j.$$

所以平均的信息量 I 是

(2.5) $$I = -(p_1\log p_1 + p_2\log p_2 + \cdots + p_n\log p_n).$$

为了指明 I 依赖于那些概率，我们将写成

(2.6) $$I = I(p_1, \cdots, p_n).$$

公式(2.5)是克劳德·谢农*(Claude Shannon)所得到的.

我们考虑最简单的情形：只有两种可能的结果，分别具有概率 p_1 与 p_2，因为按照(1.3)所有的概率之和是 1, $p_1 + p_2 = 1$，所以我们可以把 p_2 表达成

$$p_2 = 1 - p_1.$$

利用这个关系并把足码去掉，我们可把关于信息量的公式 (2.5) 写成

(2.7) $$\begin{aligned} I &= -(p_1\log p_1 + p_2\log p_2) \\ &= -p\log p + (p-1)\log(1-p). \end{aligned}$$

I 是怎样地依赖于 p 呢？ 为了研究 I 如何随 p 改变，我们利用微积分的方法：对于 p 求 I 的导数得

$$\begin{aligned} \frac{dI}{dp} &= -\log p - 1 + \log(1-p) + 1 \\ &= -\log p + \log(1-p). \end{aligned}$$

利用对数函数的函数方程，我们可以改写这个结果为

(2.8) $$\frac{dI}{dp} = \log\left(\frac{1-p}{p}\right).$$

我们在 § 5.2 中已看到，对于 $x>1$, $\log x$ 是正的；对于 $x<1$, $\log x$ 是负的. 显然有

---

* 谢农的信息理论的一本好的入门书是：CE. 谢农和 W. 魏弗尔(Weaver)的小书《通讯的数学理论》. Urbana: University of Illinois Press, 1949.