数理统计基础

杨振海 杨明哲 徐福荣 等编

北京工业大学出版社

数理统计基础

杨振海　杨明哲　徐福荣 等编

北京工业大学出版社

内 容 提 要

本书介绍了概率基础；数理统计的基本概念与抽样分布；参数估计；假设检验；回归分析；方差分析；非参数统计方法；拟合优度检验等数理统计的基本概念、理论和方法，可为进一步学习数理统计理论打下基础。本书可作大专院校数理统计课程的教材或教学参考书。

数 理 统 计 基 础

杨振海 杨明哲 徐福荣 等编

*

北京工业大学出版社出版发行

各 地 新 华 书 店 经 销

北京通县燕山印刷厂印刷

*

1991年10月第1版 1991年10月第1次印刷

787×1092毫米32开本 13.25 印张 289 千字

印数：1～2000册

ISBN7-5639-0149-3/G·99

定价：6.90元

序　言

数理统计是有广泛应用的学科. 数理统计的思想、方法已渗透到众多科学领域和技术领域. 近年来许多高等院校开设了数理统计课或开设了数理统计专业. 北京工业大学应用数学系也于86年开设了数理统计专门化方向. 为适应教学的需要，我们编写了“数理统计引论”讲义，经过四年的教学实践，对原讲义做了某些修改补充而成本书.

本书的目的是想用不大的篇幅，介绍数理统计的基本概念、理论和方法，为学习数理统计的重要分枝奠定基础. 本书可做为大专院校数理统计专业的教学参考书. 书中的主要方法，均以实例说明，并力图讲清实际背景.

阅读本书只须有数学分析基础. 主要讲述了数理统计中参数估计，假设检验等最基本的概念、方法和理论，着重介绍基本概念，同时也介绍了一般数理统计书中没有涉及到的经典非参数方法.

由于数理统计的内容十分丰富，基础知识的范围很难划定，我们的选材是否合适，有待同行批评指正；由于我们水平所限，难免有错误和不妥之处，望广大读者对书中错误和缺点不吝指正.

编　者

1990年12月

目　录

此为试读，需要完整PDF请访问：www.ertongbook.com

第一章　概率基础

§1.1　随机事件及其运算

在客观世界中，各种量的变化规律可以分为两大类．一类是确定性规律，一量可准确唯一地被其他量所确定，即按特定的规律变化．例如，自由落体下落的距离是下落时间的二次函数；水在标准大气压下100°C时沸腾；一段导体中的电流强度与加在导体两端的电压成正比等等．这类现象的共同特点是：可由给定的条件准确地预言其结果．另一类是非确定性或偶然规律，即无法根据已知条件准确地预言其结果．例如，我们无法精确地预言明年6月份的降雨量．但也并非无规律可循，通过对往年降雨量记录及其他气象资料的分析，可预言降雨量的范围．概率及数理统计就是研究这类偶然规律、探索这类规律方法的科学．

无法预言其结果的现象叫偶然现象．如掷一枚硬币，在掷之前我们无法预言是出现正面还是反面．但大量重复地掷，就会发现出现正面的次数和出现反面的次数大致相等，即在大量重复中表现出其规律性．

在相同条件下，可重复进行，不能预言其结果，但知其所有可能结果的试验叫随机试验．如掷骰子，摸扑克牌等都是随机试验．在随机试验中，每一个可能出现的结果叫基本事件，通常用ω表示．全体基本事件组成的集合叫样本空间，记做Ω．Ω的子集叫做事件．常用A、B、C、…表示．事件一般由若干基本事件组成．事件A发生是指随机试验的结果$\omega\in A$．

例1.1.1 掷一均匀的骰子，可能得到的点数有六种：1、2、3、4、5、6. 故共有六个基本事件. 样本空间$\Omega=\{1, 2, 3, 4, 5, 6\}$. 事件“得到的点数为偶数”，由三个基本事件2、4、6组成. 若记该事件为A，则$A=\{2, 4, 6\}$. 同样，Ω的子集$B=\{1, 3, 5\}$表示的事件是“得到的点数为奇数”. 而$C=\{1, 2, 3\}$代表事件“得到的点数不超过3”.

空集Φ是Ω的子集，因此它也是事件，叫不可能事件. 直观意义为不可能发生的事件. Ω也是事件，叫必然事件. 直观意义为必然发生的事件.

定义1.1.1 若事件A发生，B一定发生，则称事件B包含事件A或事件A含于B. 记做$A\subset B$.

事件B包含事件A就是凡是属于事件A的基本事件必属于事件B.

定义1.1.2 A、B是两个事件，事件“A、B两事件至少有一个发生”称为A与B的和事件. 记做$A\cup B$或$A+B$.

A、B的和事件是由A中的基本事件与B中的基本事件一起构成的事件.

定义1.1.3 A、B是两个事件，事件“A发生而B不发生”称为A与B的差事件. 记做$A\setminus B$或$A-B$.

A、B的差事件是由属于A但不属于B的基本事件组成的事件.

定义1.1.4 A、B是两个事件，事件“A与B同时发生”，称为A与B的交事件或积事件. 记做$A\cap B$或AB.

A、B的积事件是由那些既属于A又属于B的基本事件组成的事件.

定义1.1.5 若两事件不能同时发生，则称这两个事件互斥或互不相容.

显见，若事件A、B互斥，则$AB=\Phi$. 反之亦然.

定义1.1.6 A为事件，事件"A不发生"称为A的对立事件. 记做$\bar{A}$.

显然$A+\bar{A}=\Omega$, $A\bar{A}=\Phi$

例1.1.2 考虑例1.1.1中的事件A、B、C，依上述各定义有：

$$A+B=\{1,2,3,4,5,6\}=\Omega$$

$$A\setminus C=\{4,6\},\quad B\setminus C=\{2\}$$

$$AC=\{2\},\quad BC=\{1,3\},\quad AB=\Phi$$

$$\bar{A}=B,\quad \bar{B}=A,\quad \bar{C}=\{4,5,6\}$$

事件间的运算服从以下规律：

1. 交换律：$A+B=B+A$, $AB=BA$
2. 分配律：$A(B+C)=AB+AC$,
   $A(B\setminus C)=AB\setminus AC$
3. 结合律：$A(BC)=(AB)C=ABC$,
   $A+(B+C)=(A+B)+C=A+B+C$

总之，集合论中关于集合运算的公式，对事件的运算均成立. 下面的对偶公式是很有用的.

4. $\overline{\left(\bigcup\limits_{i=1}^{n} A_i\right)}=\bigcap\limits_{i=1}^{n}\bar{A}_i$

5. $\overline{\left(\bigcap_{i=1}^{n} A_i\right)}=\bigcup_{i=1}^{n} \bar{A}_i$

上面两个公式的意义是明显的．4的两端表示事件“没有一个A_i发生”，5的两端表示“所有的A_i不能同时发生”．

§1.2 概率及其运算

对于一随机试验，各种事件发生的“可能性”是不相同的．有的大些，有的小些．观察一随机试验，不仅要观察可能发生的所有结果，而且要观察各种结果发生的“可能性”大小，揭示其内在的规律．概率就是用来刻划事件发生的可能性大小的，是事件的客观属性．多次重复一随机试验时，某一事件发生的频率反映该事件发生的可能性大小．

1.2.1 频率与概率

考虑掷一枚均匀硬币的试验．H表示出现正面，T表示出现反面．重复掷币n次，n_H表示出现正面的次数，比值$f_n(H)=n_H/n$叫做n次试验中H发生的频率．它反映了H发生的可能性大小．但是这个频率不是固定的．一般来说，这n次重复掷币中H发生的频率与另一个n次掷币得到的H发生的频率是不相同的；又当重复次数n发生变化时，频率也会有所变化．幸运的是随着n的增加，H发生的频率稳定在1/2．1/2就是H发生的概率．历史上有人做过掷硬币的试验，其

表 1.1 掷硬币试验结果

试验者	n	n_H	$f_n(H)$
浦丰(Buffon)	4040	2048	0.5070
皮尔逊(K·Pearson)	12000	6019	0.5016
皮尔逊(K·Pearson)	24000	12012	0.5005

此为试读,需要完整PDF请访问: www.ertongbook.com

结果见表1.1.

一般地，设事件A在n次试验中发生n_A次，比值$f_n(A)=n_A/n$叫做事件A在这n次试验中出现的频率. 当n增大时，$f_n(A)$稳定在某一数值，这个值就是A发生的概率.

设对应随机试验E的样本空间为Ω，A、B是两个事件，则在n次重复试验中，频率有如下性质：

1. $0\leqslant f_n(A)\leqslant 1$

2. $f_n(\Omega)=1$

3 若$AB=\Phi$，则$f_n(A+B)=f_n(A)+f_n(B)$

由于当n增大时，频率稳定在概率上，故概率有与频率相似的性质. 以$P(A)$记事件A的概率，Ω为样本空间，$A_1, A_2, \cdots, A_n, \cdots$是事件序列，那么，概率有以下基本性质：

1. $0\leqslant P(A)\leqslant 1$

2. $P(\Omega)=1$

3. 对两两互斥的事件序列$A_1, A_2, \cdots, A_n, \cdots$有：

$$P\left(\sum_{i=1}^{n} A_i\right)=\sum_{i=1}^{n} P(A_i) \qquad (1.2.1)$$

$$P\left(\sum_{i=1}^{\infty} A_i\right)=\sum_{i=1}^{\infty} P(A_i) \qquad (1.2.2)$$

式子(1.2.1)称为有限可加性，(1.2.2)称为可列可加性或σ-可加性. 上面三条性质可作为概率的公理系统，即定义在样本空间的子集族上的函数，满足上述三条性质，就视为概率.

定理1.2.1 设Ω为样本空间，A、B是两个事件，P是定义在Ω上的概率函数，则：

(a) $P(A)=1-P(\bar{A})$，$P(\Phi)=0$

(b) $P(A+B)=P(A)+P(B)-P(AB)$ (1.2.3)

证明： (a) $\because A+\bar{A}=\Omega$，且$A\bar{A}=\Phi$，由有限可加性

$$1=P(\Omega)=P(A+\bar{A})=P(A)+P(\bar{A})$$

$$\therefore P(A)=1-P(\bar{A})$$

$$P(\Phi)=1-P(\bar{\Phi})=1-P(\Omega)=1-1=0$$

(b) $\because A+B=A+(B-A)$，$B=AB+(B-A)$

$$\therefore P(A+B)=P(A)+P(B-A),$$

$$P(B)=P(AB)+P(B-A)$$

由上两式易得

$$P(A+B)=P(A)+P(B)-P(AB)$$

1.2.2 古典概型

前面叙述过的掷硬币和掷骰子的随机试验有以下特点：

1. 样本空间只有有限个点．即随机试验的基本事件数目有限；

2. 每个基本事件出现的可能性相同．

一般讲，若样本空间有限，且各基本事件有相同的概率，就称这种模型为古典概型．对古典概型，$\Omega=\{\omega_1,\omega_2,\cdots,\omega_n\}$，$P(\{\omega_i\})=1/n(i=1,2,\cdots,n)$．

对任意事件A，以n_A记A中所含基本事件的数目．$n_A=\#\{\omega_i:\ \omega_i\in A\}$（$\#A$表示A中元素的个数）．则$P(A)=n_A/n$．有时，称含于A中的基本事件为A的有利基本事件，$n_A$称为有利基本事件数．

上述模型是概率发展初期研究的对象，因此称为古典概率模型，简称古典概型．若随机试验有象掷硬币等的均匀性或几何上的对称性，或者无理由认为某些基本事件出现的可能性偏大（或偏小）时，可用古典概型计算概率．

例1.2.1 同时掷两枚不同的硬币，观察正反面出现的情

况．写出样本空间Ω．设事件A为"恰有一个正面出现"，事件B为"至少有一个正面出现"，求$P(A)$、$P(B)$．

解：以H表示出现正面，T表示出现反面．则

$$\Omega=\{(H,H),\ (H,T),\ (T,H),\ (T,T)\}$$

易见　$n=4$，$n_A=2$，$n_B=3$

$\therefore$　$P(A)=n_A/n=2/4=1/2$　　$P(B)=n_B/n=3/4$

例1.2.2　同时掷m个硬币，试求至少有一个出现正面的概率是多少？

解：令$A=\{m$个硬币中至少有一个出现正面$\}$

则$\bar{A}=\{m$个硬币全是反面$\}$

易见$n=2^m$，$P(\bar{A})=1/2^m$

故　$P(A)=1-P(\bar{A})=1-1/2^m$

本例的解法是计算概率常用的方法之一．

例1.2.3　（超几何分布）今有N件产品，其中有M件次品，从中随机地抽取n件，问其中恰有m件次品的概率是多少？

解：基本事件总数：从N件产品中抽取n件共有$\binom{N}{n}$种取法．

有利基本事件数：n件产品中有m件次品，这m件次品只能从M件次品中抽取，共有$\binom{M}{m}$种取法；而$n-m$件正品取自$N-M$件正品，共有取法$\binom{N-M}{n-m}$种．故有利基本事件数为$\binom{M}{m}\binom{N-M}{n-m}$，于是

$$P(n\text{件产品中有}m\text{件次品})=\binom{M}{m}\binom{N-M}{n-m}\Big/\binom{N}{n} \tag{1.2.4}$$

$m=0,1,2,\cdots,\min(M,n)$

(1.2.4)称为超几何分布．记做$H(m,N,n)$

例1.2.4 有m个球，每个球落入$N(N\geqslant m)$个盒子中任一个的可能性相等．假设每个盒子盛球的个数无限．求事件$A=\{$某预先指定的m个盒子中各落入一个球$\}$的概率．

解： 由于每个球可落入N个盒子中的任一个，故m个球共有N^m种落法；而m个球落入指定m个盒子且每个盒子一个球的落法共有$m!$种．于是$P(A)=m!/N^m$．

这个例子的结果可解所谓生日问题．若两人同月同日（可不同年）生，则说两人同生日．试求r个人中至少有两人同生日的概率．该事件的对立事件为"r个人的生日均不相同"．由上例，r个人为r个指定的不同生日的概率为$r!/365^r$，又r个不同生日的取法有$\binom{365}{r}$种，故所求的概率为：

$$P=1-\binom{365}{r}r!/365^r$$

$$=1-\frac{1}{365^r}\frac{365!}{(365-r)!}$$

$$=1-\prod_{i=1}^{r-1}\left(1-\frac{i}{365}\right)$$

经计算给出下面的结果：

r	2	4	5	10	15	20
P	0.003	0.016	0.027	0.095	0.253	0.414

r	23	25	30	35	40	50
P	0.507	0.569	0.706	0.813	0.891	0.97

由上表可以看出：当$r=23$时，大约50%的可能性至少

有两人同生日．特别有趣的是美国前36位总统中有三人生于同日，四人死于同日．在50个人中（随机地组合在一起），几乎肯定地说至少两人有相同的生日．这个结论是出乎人们直观预料的．

例1.2.5 有k张电影票，$n(n>k)$个人想要．由抓阄决定谁得票．做n个阄，其中有k个上写“有”字，其余为空白．一个人接一个人地抓取．试问第i个人抓到“有”阄的概率．

解： n个阄n个人一个接一个地抓取，共有$n!$种抓法．这是基本事件总数．

第i个人抓到某特定阄的抓法共有$(n-1)!$种，而他抓到k个“有”阄中任何一个都是抓到“有”，故第i个人抓到“有”阄的抓法共有$k(n-1)!$种，这是有利基本事件数．从而

$$P\{\text{第}i\text{个人抓到“有”阄}\}=k(n-1)!/n!=k/n$$

由此可见，抓阄是公平的，尽管先抓者对后抓者有影响，但有利影响和不利影响相互抵消．这一点在条件概率（§1.3）中看得更清楚．

概率公理系统是概率论中三条基本规律．其它的概率规律均可由此推出．下面的定理称为概率连续性．

定理1.2.2 设$\{A_n\}$是随机事件的递降序列，$A=\bigcap\limits_{i=1}^{\infty} A_i$

则
$$P(A)=\lim_{n\to\infty} P(A_n) \qquad (1.2.5)$$

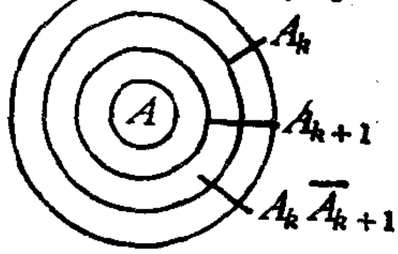

图1.1

证明： 由于$\{A_n\}$是递降序列，即

$A_1\supset A_2\supset\cdots\supset A_n\supset\cdots$

故 对任意的n有$A_n=A+\bigcup_{k=n}^{\infty}A_k\bar{A}_{k+1}$(如图1.1).

由公式(1.2.3)

$$P(A_n)=P(A)+P(\bigcup_{k=n}^{\infty}A_k\bar{A}_{k+1})-P(A(\bigcup_{k=n}^{\infty}A_k\bar{A}_{k+1}))$$

而 $A\left(\bigcup_{k=n}^{\infty}A_k\bar{A}_{k+1}\right)=\bigcup_{k=n}^{\infty}AA_k\bar{A}_{k+1}$, $A\bar{A}_{k+1}=\Phi$

由可列可加性 $P(A\cdot(\bigcup_{k=n}^{\infty}A_k\bar{A}_{k+1}))=0$

故 $$P(A_n)=P(A)+\sum_{k=n}^{\infty}P(A_k\bar{A}_{k+1}) \tag{1.2.6}$$

令 $n=1$, $1\geqslant P(A_1)=P(A)+\sum_{k=1}^{\infty}P(A_k\bar{A}_{k+1})$

故 级数$\sum_{k=1}^{\infty}P(A_k\bar{A}_{k+1})$收敛

对(1.2.6)两边取极限($n\to\infty$),得

$$P(A)=\lim_{n\to\infty}P(A_n)$$

§1.3 条件概率和独立性

考虑例1.2.5中抓阄问题.若已知第一个人抓到“有”阄,那么第二个人抓到“有”阄的概率是$(k-1)/(n-1)$,不再是k/n.这是因为有了附加的信息即条件“第一个人抓到有”.故称这个概率为条件概率.记做$P(A_2|A_1)$ A_i表示“第i个人抓到有”.$|A_1$表示在A_1发生的条件下.显然,$P(A_2|A_1)\neq P(A_2)$.那么$P(A_2|A_1)$与$P(A_1)$、$P(A_2)$有什么关系呢?易见:$P(A_1A_2)=k(k-1)/[n(n-1)]$于是

此为试读,需要完整PDF请访问: www.ertongbook.com