概率论与数理统计

主编 刘浩瀚

高等教育出版社

概率论与数理统计

Gailülun yu Shuli Tongji

主　编　刘浩瀚
参　编　陈少云　黄非难　余家树　张松林　李　伟

高等教育出版社·北京

内容简介

本书是作者根据教学实践编著而成，重点培养学生对概率统计知识的实际应用能力。本书内容包括随机变量及其分布、多维随机变量及其分布、随机变量的数字特征、大数定律与中心极限定理、数理统计的基础知识、参数估计、假设检验、方差分析与回归分析。本书合理选择部分例题并附上 Excel 2000 软件求解方法，在章节之后附有相当数量的习题，同时在书末附有常用分布、记号及数字特征表及常用的临界值表，供读者查阅。

本书以提高读者解题能力与解决实际问题能力为出发点，从实例引入抽象的基本概念，又从抽象的数学定理结合 Excel 2000 软件求解，将复杂的概率统计计算问题简单化，有助于读者较快地掌握概率统计的相关知识并求出准确的结果。本书可作为高职高专工科类各专业通用教材，也可作为概率统计应用人员的工具性参考书。

图书在版编目(CIP)数据

概率论与数理统计／刘浩瀚主编. - -北京：高等教育出版社，2015.5（2018.7重印）
ISBN 978-7-04-042400-3

Ⅰ. ①概… Ⅱ. ①刘… Ⅲ. ①概率论-高等职业教育-教材②数理统计-高等职业教育-教材 Ⅳ. ①O21

中国版本图书馆 CIP 数据核字(2015)第 067557 号

策划编辑 崔梅萍　　责任编辑 崔梅萍　　封面设计 王　洋　　版式设计 童　丹
插图绘制 郝　林　　责任校对 殷　然　　责任印制 赵义民

出版发行	高等教育出版社	咨询电话	400-810-0598
社　　址	北京市西城区德外大街 4 号	网　　址	http://www.hep.edu.cn
邮政编码	100120		http://www.hep.com.cn
印　　刷	中国农业出版社印刷厂	网上订购	http://www.landraco.com
开　　本	787mm×1092mm　1/16		http://www.landraco.com.cn
印　　张	10	版　　次	2015 年 5 月第 1 版
字　　数	240 千字	印　　次	2018 年 7 月第 5 次印刷
购书热线	010-58581118	定　　价	17.30 元

物 料 号　42400-00

前　言

概率统计的理论和方法及其应用几乎遍及科学技术领域、工农业生产和国民经济的各个部门。由于它的应用广泛,大多数高职高专院校都把它作为一门公共基础课。根据高职"培育具备足够的基础知识、较强的技术应用能力的实用性人才"这一培养目标的基本要求,遵循"以应用为目的,以必需、够用为度"的教学原则,同时为了适应市场经济对人才的需求,培养学生运用概率统计方法分析解决实际问题的能力和创造性思维能力,我们结合自己及相关专家、学者多年的数学实践经验,以高中基本的概率统计知识为基础,结合传统概率统计理论、方法内容,选择合理的实际应用背景知识,重点强化 Excel 在概率统计求解中的应用,我们编写了本教材。

本书有以下特色:

1. 注重与高中内容和已学知识的衔接。学生在高中阶段学过简单的概率统计的知识和概念,本书对学过的内容进行归纳总结,在已学知识的基础上,适当加深对基本概念、定理的阐述,同时拓广大学所必需的概率统计知识。

2. 注重与软件的结合,切实加强应用。专业的统计软件很多,如 SPSS,SAS 等,我们选择了常用的 Excel 来处理统计中的计算问题,使学生从繁杂的计算中解放出来,更多精力放在问题的处理、方法的选择上,提高学生的学习兴趣和解决实际问题的能力。

3. 切实加强应用。本书在内容讲述上不过分强调理论的推导,而是突出所学知识在实际中的应用。为此,我们精选统计例题和习题,文中例题涉及工业、农业、医学、经济等各个方面,使学生能切实感受到学习这门课的必要性,使学生学有所获。

本书共分八章,内容包括:随机事件的概率、随机变量及其概率分布、随机变量的数字特征、大数定律与中心极限定理、数理统计的基本概念、参数估计、假设检验、方差分析与回归分析等。本书由陈少云(第一—四章),刘浩瀚(第五—八章)编写初稿,各章节的 Excel 软件求解程序由刘浩瀚编写,并由刘浩瀚统稿、定稿。在编写过程中,得到了四川建筑职业技术学院数学教研室教师黄非难、余家树、张松林、李伟的支持和帮助,编者谨致谢意。

限于编者的水平和精力,本书难免存在不足之处,衷心欢迎读者批评指正。

编　者

2015 年 1 月

目 录

预备知识

一、计数原理

1. 排列数公式

$A_n^m = n(n-1)(n-2)\cdots(n-m+1)$，其中 $n, m \in \mathbf{N}^*$ 并且 $m \leqslant n$.

2. 全排列

$A_n^n = n!$，另外规定 $0! = 1$.

3. 组合数公式

$$C_n^m = \frac{A_n^m}{A_m^m} = n(n-1)(n-2)\cdots\frac{(n-m+1)}{m!} \text{或} C_n^m = \frac{n!}{m!\ (n-m)!},$$

$C_n^m = C_n^{n-m}$,

$C_{n+1}^m = C_n^m + C_n^{m-1}$.

4. 二项式定理

$(a+b)^n = C_n^0 a^n + C_n^1 a^{n-1} b + \cdots + C_n^k a^{n-k} b^k + \cdots + C_n^n b^n$，其中 $k \in \{0, 1, 2, \cdots, n\}$.

二、随机变量及其分布

1. 离散型随机变量及其分布列

随着试验结果变化而变化的变量称为随机变量(random variable).常用字母 $X, Y, \xi, \eta, \cdots$ 表示.

所有取值可以一一列出的随机变量，称为离散型随机变量(discrete random variable).

若离散型随机变量 X 可能取的不同值为 $x_1, x_2, \cdots x_n$，X 取每个值 $x_i (i=1,2,\cdots,n)$ 的概率 $P\{X=x_i\} = p_i$，以表 0-1 形式表示如下：

表 0-1

X	x_1	x_2	…	x_i	…	x_n
P	p_1	p_2	…	p_i	…	p_n

表 0-1 称为离散型随机变量 X 的概率分布列(probability distribution series)，简称为 X 的分布列(distribution series).

离散型随机变量有以下性质：

(1) $p_i \geqslant 0, i=1,2,\cdots,n$;

(2) $\sum_{i=1}^{n} p_i = 1$.

随机变量的分布列以表 0-2 形式表示如下

表 0-2

X	0	1
P	$1-p$	p

的分布列称为两点分布列.

如果随机变量 X 的分布列为两点分布列,就称 X 服从两点分布(two-point distribution),而称 $p=P\{x=1\}$ 为成功概率.

分布列以表 0-3 形式表示如下:

表 0-3

X	0	1	…	m
P	$C_M^0C_{N-M}^{n-0}/C_N^n$	$C_M^1C_{N-M}^{n-1}/C_N^n$	…	$C_M^mC_{N-M}^{n-m}/C_N^n$

为超几何分布列.

如果随机变量 X 的分布列为超几何分布列,则称随机变量 X 服从超几何分布(hyper geometric distribution).

2. 条件概率

一般地,设 A,B 为两个事件,且 $P(A)>0$,称 $P(B|A)=P(AB)|P(A)$ 为在事件 A 发生的条件下,事件 B 发生的条件概率(conditional probability).

条件概率基本性质:$0\leqslant P(B|A)\leqslant 1$;

如果事件 B,C 为互斥事件,则 $P(B\cup C|A)=P(B|A)+P(C|A)$.

设 A,B 为两个事件,如果 $P(AB)=P(A)P(B)$,则称事件 A 与事件 B 相互独立.

3. 二项分布

一般地,在相同条件下重复做 n 次试验称为 n 次独立重复试验(independent and repeated trials).在 n 次独立重复试验中,设事件 A 发生的次数为 X,在每次试验中事件 A 发生的概率为 p,那么在 n 次独立重复试验中,事件 A 恰好发生 k 次的概率为

$$P\{X=k\}=C_n^kp^k(1-p)^{n-k},k=0,1,2,\cdots,n.$$

此时称随机变量 X 服从二项分布.

三、离散型随机变量的均值与方差

1. 离散型随机变量的均值

离散型随机变量 X 的分布如表 0-1 所示,则称 $E(X)=x_1p_1+x_2p_2+\cdots+x_np_n$ 为随机变量 X 的均值(mean)或数学期望(mathematical expectation),它反映了随机变量取值的平均水平.

若 X 为随机变量,a,b 为常数,则 $E(aX+b)=aE(X)+b$.

若 X 服从两点分布,则 $E(X)=p$;

若 $X\sim B(n,p)$,则 $E(X)=np$.

2. 离散型随机变量的方差

设离散型随机变量 X 分布列为表 0-1,则$(x_i-E(X))^2$ 描述了 $x_i(i=1,2,\cdots,n)$ 相对于均值

$E(X)$的偏离程度，称 $D(X)=\sum_{i=1}^{n}(x_i-E(X))^2p_i$ 为随机变量 X 的方差(variance)，其算术平方根 $\sqrt{D(X)}$ 为随机变量的标准差(standard deviation).

若 X 服从两点分布，则 $D(X)=p(1-p)$；

若 $X\sim B(n,P)$，则 $D(X)=np(1-p)$；

$D(aX+b)=a^2D(X)$.

四、正态分布

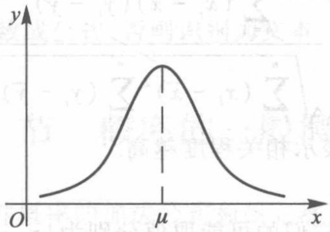

图 0-1　正态分布曲线

图 0-1 所示曲线为函数

$$\varphi_{\mu,\sigma}(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}},x\in(-\infty,+\infty)$$

的图像，其中 $\mu,\sigma(\sigma>0)$ 为参数.称 $\varphi_{\mu,\sigma}(x)$ 的图像为正态分布的密度曲线，称为正态曲线.

一般地，如果对于任何实数 $a<b$，随机变量 X 满足：

$$P\{a<X\leqslant b\}=\int_a^b\varphi_{\mu,\sigma}(x)\mathrm{d}x,$$

则称 X 的分布为正态分布(normal distribution)，正态分布由参数 μ,σ 确定，通常记为 $N(\mu,\sigma^2)$.

五、统计案例

1. 回归分析

回归分析(regression analysis)是对具有相关关系的两个变量进行统计分析的一种常用方法.在研究两个变量间的关系时，首先要根据散点图来粗略判断它们是否线性相关，是否可以用线性回归模型来拟合数据.

对于一组具有线性相关关系的数据

$$(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n),$$

线性回归模型为

$$\begin{cases}y=bx+a+e,\\E(e)=0,D(e)=\sigma^2.\end{cases}$$

回归方程 $y=bx+a$ 的截距和斜率的最小二乘估计公式分别为：

$$\hat{a}=\bar{y}-\hat{b}\,\bar{x},$$

$$\hat{b}=\frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sum_{i=1}^{n}(x_i-\overline{x})^2},$$

其中$\overline{x}=\frac{1}{n}\sum_{i=1}^{n}x_i,\overline{y}=\frac{1}{n}\sum_{i=1}^{n}y_i$.

可以用相关系数 r 来衡量两个变量之间的线性相关关系：

$$r=\frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\overline{x})^2\sum_{i=1}^{n}(y_i-\overline{y})^2}},$$

其中 $r>0$ 表示正相关，r 越接近 1 表示相关程度越高.

2. 独立性检验

假设有两个分类变量 X 和 Y，它们的可能取值分别为$\{x_1,x_2\}$和$\{y_1,y_2\}$，其样本频数列联表（称为 2×2 列联表）为：

2×2 列联表

	y_1	y_2	总　计
x_1	a	b	$a+b$
x_2	c	d	$c+d$
总计	$a+c$	$b+d$	$a+b+c+d$

若要推断的论述为

$$H_1:X\text{与}Y\text{有关系},$$

可以利用独立性检验来考察两个分类变量是否有关系，并且能较精确地给出这种判断的可靠程度. 具体做法是：

① 根据实际问题需要的可信程度确定临界值 k_0；

② 利用公式，由观测数据计算得到随机变量 $K^2=\frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}(n=a+b+c+d)$ 的观测值 k；

③ 如果 $k>k_0$，就以$(1-P\{K^2\geqslant k_0\})\times100\%$的把握认为“$X$ 与 Y 有关系”；否则就说样本观测数据没有提供“X 与 Y 有关系”的充分证据.

第一章　随机变量及其分布

本章在引入随机变量后,把对随机事件及其概率的研究转化为对随机变量及其取值的概率规律性(即分布)的研究,并重点讨论几个有代表性的常用分布及随机变量函数的分布.本书中引入 Excel 2000 进行求解.Excel 具有非常强大的函数功能.Excel 2000 中的每一个单元格都是一个功能强大的计算器,它可以提供 12 类共 400 多条函数.利用这些函数和自己构造的公式可以高效地完成概率统计中的计算问题.在 Excel 中使用函数,必须先输入'='或'+','-',凡以'='或'+'开始的数据,Excel 即识别为函数或公式,否则识别为文本.

第一节　概率的一般概念

本节复习概率的概念、互斥事件概率的加法公式和独立事件概率的乘法公式.

一、概率的统计定义

在一个随机试验中,随机事件是否发生是很重要的,但更重要的是随机事件发生的可能性大小,它是事件本身的客观属性,是可以度量的.于是,把刻画事件 A 在试验中发生的可能性大小的量 p 称为事件 A 的概率,记作 $P(A)=p$.

在一个随机试验下,怎样确定事件 A 的概率呢?为此,先介绍事件 A 的频率的概念.

定义 1(事件的频率)　若事件 A 在 N 次重复试验中出现 M 次,则称 M 为事件 A 在这 N 次试验中出现的频数,称 M/N 为事件 A 在这 N 次试验中出现的频率,记作 $f_N(A)$.于是,

$$f_N(A)=\frac{M}{N}.$$

实践表明,当试验次数 N 逐渐增大时,频率 $f_N(A)$ 虽然不尽相同,但却稳定在某常数 p 附近.例如,掷一枚质地均匀的硬币,随着抛掷次数的不断增加,正面朝上(设为事件 A)的频率 $f_N(A)$ 就在 0.5 附近摆动,抛掷次数越多,摆动范围越小,越接近 0.5.这一点已由表 1-1 中诸试验者的试验所验证.

表 1-1　抛掷硬币试验中正面朝上的频率

试验者	抛掷硬币的次数	正面朝上的频数	正面朝上的频率
德摩根	2 048	1 061	0.518 1
蒲丰	4 040	2 048	0.506 9
皮尔逊	12 000	6 019	0.501 6
皮尔逊	24 000	12 012	0.500 5

表 1-1 的数据充分揭示了随机事件的一个极其重要的特性:频率的稳定性.它表明常数 p 是事件本身客观存在的一种固有属性,反映了事件本身所蕴含的规律性.因此,常数 p 可以对事件 A 发生的可能性大小进行度量.

定义 2(概率的统计定义)　事件 A 出现的频率 $f_N(A)$ 随着试验次数 N 的增大而在某个常数 p 附近摆动,则称 p 为事件 A 的概率,记作 $P(A)$,即

$$P(A)=p.$$

概率的统计定义指出,任一事件 A 发生的概率是客观存在的.在实际问题中,往往不知 $P(A)$ 为何值,这时可取一定数量试验下事件 A 的频率 $f_N(A)$ 作为 $P(A)$ 的近似值.这正是统计定义的优点所在.

二、概率的性质

由概率和频率的关系,易见概率具有下列性质:

性质 1　对任意事件 A 有 $0\leqslant P(A)\leqslant 1$;

性质 2　必然事件 Ω 的概率为 1,即 $P(\Omega)=1$;

性质 3　不可能事件 $\varnothing$ 的概率为 0,即 $P(\varnothing)=0$.

三、概率的古典定义

由概率的统计定义只能大致估计事件 A 的概率 $P(A)$,但是在某些特殊的随机试验中,我们可以先行计算事件的概率,这种试验就是下述的古典概型.

如果随机试验只有有限个基本事件(有限性),每个基本事件发生的可能性相同(等可能性),则称这种随机试验为古典概型.

定义 3(概率的古典定义)　对于给定的古典概型,若样本空间中的基本事件总数为 n,事件 A 包含了 m 个基本事件,则事件 A 的概率为

$$P(A)=\frac{m}{n}.$$

例 1(古典定义求概率)　有 10 件产品,其中 2 件次品,从中任取 3 件,求(1)3 件产品全是正品的概率;(2)3 件产品中恰有 1 件次品的概率;(3)3 件产品中至少有 1 件次品的概率.

解　设 A =“全是正品”,B =“恰有 1 件次品”,C =“至少有 1 件次品”.从 10 件中取出 3 件,共有 C_{10}^3 种取法,即基本事件总数 $n=\mathrm{C}_{10}^3$.

(1) 取出 3 件全是正品有 C_8^3 种取法,故所求概率为

$$P(A)=\frac{\mathrm{C}_8^3}{\mathrm{C}_{10}^3}=\frac{7}{15};$$

(2) 取出 3 件中恰有 1 件次品有 $\mathrm{C}_2^1\mathrm{C}_8^2$ 种取法,故所求概率为

$$P(B)=\frac{\mathrm{C}_2^1\mathrm{C}_8^2}{\mathrm{C}_{10}^3}=\frac{7}{15};$$

(3) 取出 3 件中至少有 1 件次品有 $\mathrm{C}_{10}^3-\mathrm{C}_8^3$ 种取法,故所求概率为

$$P(C)=\frac{\mathrm{C}_{10}^3-\mathrm{C}_8^3}{\mathrm{C}_{10}^3}=\frac{8}{15}.$$

四、互斥事件概率的加法公式

若事件 A 与 B 在一次试验中不可能同时发生,即满足 $AB=\varnothing$,则称事件 A 与 B 是互斥(互

不相容)事件.如上例中的 A 与 B 是互斥事件,A 与 C 也是互斥事件.

定理 1(两个互斥事件概率的加法公式)　两个互斥事件 A 与 B 的和(或并)的概率等于这两个事件的概率的和,即

$$P(A+B)=P(A)+P(B).$$

在上例中,设 $D=$"取出 3 件中次品数不超过 1 件",则 $D=A+B$,因此事件 D 发生的概率为

$$P(D)=P(A+B)=P(A)+P(B)=\frac{7}{15}+\frac{7}{15}=\frac{14}{15}.$$

推论 1(多个互斥事件概率的加法公式)　若事件 $A_1,A_2,\cdots,A_n$ 两两互斥,则它们的和的概率等于它们概率的和,即

$$P(A_1+A_2+\cdots+A_n)=P(A_1)+P(A_2)+\cdots+P(A_n).$$

如果事件 A 与 B 在一次试验中必有一个发生,且仅有一个发生,即满足 $A+B=\Omega$, $AB=\varnothing$,则称事件 A 与 B 互为逆事件(对立事件),记作 $B=\overline{A}$, $A=\overline{B}$. 上例中"至少 1 件次品"与"全是正品"就互为逆事件,即 $C=\overline{A}$.

推论 2(逆事件的概率公式)　互逆事件的概率和等于 1,即

$$P(A)+P(\overline{A})=1.$$

例 2(互斥事件概率的加法公式求概率)　一盒电子元件共有 50 个,其中有 5 个次品,从这盒电子元件中任取 3 个,求其中有次品的概率.

解法一　设 $A=$"取出的 3 个元件中有次品",$A_i=$"取出的 3 个元件中恰有 i 个次品"($i=1,2,3$),显然,事件 A_1,A_2,A_3 两两互斥,且 $A=A_1+A_2+A_3$. 由于

$$P(A_1)=\frac{C_5^1C_{45}^2}{C_{50}^3}\approx 0.252\ 6,P(A_2)=\frac{C_5^2C_{45}^1}{C_{50}^3}\approx 0.023\ 0,P(A_3)=\frac{C_5^3}{C_{50}^3}\approx 0.000\ 5,$$

故所求概率为

$$P(A)=P(A_1+A_2+A_3)=P(A_1)+P(A_2)+P(A_3)\approx 0.276\ 1.$$

解法二　设 $A=$"取出的 3 个元件中有次品",则 $\overline{A}=$"取出的 3 个元件全是正品",由于

$$P(\overline{A})=\frac{C_{45}^3}{C_{50}^3}\approx 0.724\ 0,$$

故所求概率为

$$P(A)=1-P(\overline{A})\approx 0.276\ 0.$$

注:两解法最后结果第四位不同是由舍入误差引起的.

五、独立事件概率的乘法公式

在现实中,某些事件的发生不相互影响.例如,抛掷两枚硬币,观察出现正反面的情况,令 A 表示"第一枚硬币出现正面",B 表示"第二枚硬币出现反面",很明显事件 A 与 B 之间没有必然的联系,其中任一个事件发生与否,都不影响另一个事件发生的概率,我们称事件 A 与 B 相互独立,简称 A 与 B 独立.

下面给出事件相互独立的严格定义.

定义 4(两事件独立)　若对于事件 A 与 B，有 $P(AB)=P(A)P(B)$，则称事件 A 与 B 相互独立，简称 A 与 B 独立.

可以证明，若 A 与 B 相互独立，则 A 与 $\overline{B}$，$\overline{A}$ 与 B，$\overline{A}$ 与 $\overline{B}$ 都相互独立.一般地，A 与 B，A 与 $\overline{B}$，$\overline{A}$ 与 B，$\overline{A}$ 与 $\overline{B}$ 只要有一组相互独立，另外三组也各自相互独立.

定义 5(多个事件独立)　设有 n 个事件 $A_1,A_2,\cdots,A_n$ ($n \geqslant 3$)，若对于其中任意 k 个事件 $A_{i_1},A_{i_2},\cdots,A_{i_k}$ ($2 \leqslant k \leqslant n$) 有

$$P(A_{i_1}A_{i_2}\cdots A_{i_k})=P(A_{i_1})P(A_{i_2})\cdots P(A_{i_k}),$$

则称这 n 个事件 $A_1,A_2,\cdots,A_n$ 是相互独立的.

由该定义可知，若 $A_1,A_2,\cdots,A_n$ 是相互独立的，则从中任选 k ($1 \leqslant k \leqslant n$) 个仍相互独立.还可以证明，任选 k 个事件，并把其中一些事件换成其对立事件，这样得到的新事件组仍相互独立.

定理 2(独立事件概率的乘法公式)　若事件 $A_1,A_2,\cdots,A_n$ 相互独立，则它们的积的概率等于它们的概率的积，即

$$P(A_1A_2\cdots A_n)=P(A_1)P(A_2)\cdots P(A_n).$$

需要指出的是，在实际应用中，事件的独立性往往不是根据定义来判断，而是根据实际意义来判断，即通常把不存在明显影响的若干事件看做是相互独立的.例如，电路中的各元件正常工作、放回抽样的各次抽取结果都是常见的独立性模型.

例 3(独立事件概率乘法公式求概率)　甲、乙两人独立射击同一目标，已知甲击中目标的概率为 0.7，乙击中目标的概率为 0.8.求(1)两人都击中目标的概率；(2)目标被击中的概率.

解　设 A =“甲击中目标”，B =“乙击中目标”，则

$$P(A)=0.7,P(B)=0.8,$$

显然，事件 A 与 B 相互独立.于是有

(1)两人都击中目标的概率

$$P(AB)=P(A)P(B)=0.7\times 0.8=0.56;$$

(2)目标被击中的概率

$$P(A+B)=P(A)+P(B)-P(AB)=0.7+0.8-0.56=0.94.$$

对于 n 个独立事件 $A_1,A_2,\cdots,A_n$，其和事件的概率可以通过下式来计算：

$$P(A_1+A_2+\cdots+A_n)=1-P(\overline{A_1}\;\overline{A_2}\cdots\overline{A_n})=1-P(\overline{A_1})P(\overline{A_2})\cdots P(\overline{A_n}),$$

这便是独立事件概率的加法公式.

例 4(相互独立事件的和的概率)　若每个人血清中含有肝炎病毒的概率为 0.004，混合 100 个人的血清，求此血清中含有肝炎病毒的概率.

解　设 A_i =“第 i 个人的血清含有肝炎病毒”，$i=1,2,\cdots,100$，显然，它们相互独立.所求概率为

$$P(A_1+A_2+\cdots+A_{100})=1-P(\overline{A_1})P(\overline{A_2})\cdots P(\overline{A_{100}})=1-0.996^{100}\approx 0.330\,2,$$

虽然每个人有病毒的概率很小，但是混合后则有很大的概率，在实际工作中，这类效应值得重视.

课堂练习

1. 连续抛掷一枚质量均匀的硬币 3 次，求只有 1 次正面朝上的概率.

2. 口袋中有10个球,分别标有号码1到10,现从中任取3只,记下取出的号码,求最小号码为5的概率.

3. 袋中有10个球,8个白球2个红球.

(1) 不放回地每次从中任取一个,连取3次,求第三次取到红球的概率;

(2) 每次从中任取一个,有放回地连取3次,求取到3个红球的概率.

4. 500个人中,至少有一个人的生日是在7月1日的概率为多少?(1年按365天计算.)

5. 已知A,B为两个相互独立的事件,$P(A) = 0.2, P(B) = 0.4$,则$P(A + B) =$ ________.

6. 三人独立地破译一个密码,他们能单独译出的概率分别为$\frac{1}{5},\frac{1}{3},\frac{1}{4}$,求该密码被破译的概率.

习题1-1

1. 罐中有12粒围棋子,其中8粒白子4粒黑子,从中任取3粒,求

(1) 取到的都是白子的概率;

(2) 取到2粒白子1粒黑子的概率;

(3) 至少取到1粒黑子的概率;

(4) 取到的3粒棋子颜色相同的概率.

2. 袋中有7个球,其中5个白球2个红球,不放回地取球2次,求

(1) 两次都取到红球的概率;

(2) 第一次取得白球,第二次取得红球的概率;

(3) 两次取得的球中一个白球一个红球的概率;

(4) 取得的两个球颜色相同的概率.

3. 将3名学生的学生证混放在一起,现将其随意地发给这3名学生,求至少有一名学生拿到自己的学生证的概率.

4. 某专业研究生复试时,有3张考签,3个考生应试,一个人抽一张看后立刻放回,再让另一个人抽,如此3个人各抽一次,求抽签结束后,至少有一张考签没有被抽到的概率.

5. 有甲、乙两批种子,发芽率分别为0.8和0.7,在两批种子中各取一粒,求

(1) 两粒种子都发芽的概率;

(2) 至少有一粒种子发芽的概率;

(3) 恰好有一粒种子发芽的概率.

第二节　离散型随机变量及其分布列

本节将介绍随机变量的定义和类型,并给出离散型随机变量的定义和性质,讨论几种有代表性的离散型随机变量及其分布.随机变量的引入,能让我们充分认识随机现象的统计规律性,使我们能运用高等数学的方法来研究随机现象,实现把随机事件及其概率的研究转化为对随机变量及其取值的概率规律性(即分布)的讨论.

一、随机变量

在现实中,很多随机试验的结果本身就是用数量表示的.如抛掷一枚骰子,出现的点数用X

表示，当一次试验中出现的点数为 1 时 $X=1$，当出现的点数为 2 时 $X=2$，…，当出现的点数为 6 时 $X=6$，X 就是一个随机变量.又如 110 报警台 24 小时内接到的报警次数 Y，商店某批灯泡的寿命 Z，飞机着陆点 W 等都是随机变量.

有些随机试验的结果本身不是数量，但可以人为地数量化.如抛掷一枚硬币一次，有两种结果：正面朝上或反面朝上，结果不是数量.设 X 表示抛掷一枚硬币一次正面朝上的次数，则 $X=1$ 表示正面朝上，$X=0$ 表示反面朝上. X 随着试验的不同结果而取不同的值，X 为随机变量.又如某足球队参加比赛，这一随机试验的结果有三种：胜、平、负.记 Y 为一场足球比赛的积分数，则 Y 为随机变量，当结果为"胜"时 $Y=3$，结果为"平"时 $Y=1$，结果为"负"时 $Y=0$.

从上面的例子看到，在随机试验中，存在一个变量，依试验的结果不同而取不同的值，但在试验前变量取值又具有随机性，那么就称这个变量为随机变量.下面给出随机变量的严格定义.

定义 1（随机变量） 设随机试验的样本空间为 Ω，如果对于每一个结果（样本点）$\omega\in\Omega$，有一个实数 X 与之对应，则称实值函数 $X=X(\omega)$ 为 Ω 上的随机变量.随机变量通常用 $X, Y, Z, \cdots$ 或 $X_1, X_2, \cdots$ 表示.

这种变量之所以称为随机变量，是因为它的取值随试验的结果而定，而试验结果的出现具有随机性，因而它的取值是随机的.在一次试验之前，我们不能预先确定随机变量取什么值，但由于试验的所有可能结果是预先知道的，故对于每一个随机变量，我们可知道它的取值范围，且可知它取各个值的可能性大小.这一性质显示了随机变量与普通函数有本质的区别.

需要说明的是，对于一个随机试验，可以有多个与之关联的随机变量，而不是仅有一个，随着不同的研究需要定义不同的随机变量.

引入随机变量后，就可以用随机变量描述事件及事件的概率.如从有 2 件次品的 10 件产品中任取 3 件，设 X 为取出 3 件产品中的次品数，则其取值范围是 $X=0,1$ 或 2，$\{X=0\}$ 表示事件"3 件全为正品"，$P\{X=0\}=\dfrac{7}{15}$；$\{X=1\}$ 表示事件"3 件中恰有 1 件次品"，$P\{X=1\}=\dfrac{7}{15}$；$\{X\geqslant 1\}$ 表示事件"3 件中至少 1 件次品"，并且 $\{X\geqslant 1\}=\{X=1\}+\{X=2\}$，$\{X=1\}$ 与 $\{X=2\}$ 互斥，$\{X\geqslant 1\}$ 与 $\{X=0\}$ 互为逆事件，即

$$P\{X\geqslant 1\}=P\{X=1\}+P\{X=2\}=\frac{7}{15}+\frac{1}{15}=\frac{8}{15}$$

或
$$P\{X\geqslant 1\}=1-P\{X=0\}=1-\frac{7}{15}=\frac{8}{15}.$$

用随机变量描述事件，可以让我们不只是孤立地研究一个或几个事件，而是通过随机变量把各个随机事件联系起来，进而去认识随机试验的全貌.

若随机变量可能取值的全体是有限个或无穷可列个实数，那么称其为离散型随机变量.如抛掷骰子出现的点数，110 报警台一段时间内接到的报警次数都是离散型随机变量.若随机变量可能取某一实数区间上的所有值，那么称其为连续型随机变量.如电子元件的使用寿命，在收银台前排队的时间都是连续型随机变量.

二、离散型随机变量的分布列及其性质

定义 2（分布列） 设 X 为离散型随机变量，可能取值为 $x_1, x_2, \cdots, x_k, \cdots$，且

$$P\{X=x_k\}=p_k, k=1,2,\cdots,$$

则称上式为离散型随机变量 X 的分布列.

分布列也可以用表格的形式表示：

X	x_1	x_2	$\cdots$	x_k	$\cdots$
P	p_1	p_2	$\cdots$	p_k	$\cdots$

离散型随机变量的分布列实质上是用以表达“可能取值”及其“对应概率”两要素的方式，全面描述了离散型随机变量取值的概率规律.对于 X 可能取值的正、负并无限制，但通常约定：$x_1<x_2<\cdots<x_k<\cdots$.

分布列具有下列性质：

(1) $p_k \geqslant 0, k=1,2,\cdots$（非负性）；

(2) $\sum\limits_k p_k=1$（规范性）.

上式中，当 X 取有限个可能值时，$\sum\limits_k p_k$ 表示有限项的和；当 X 取无穷可列个可能值时，$\sum\limits_k p_k$ 表示无穷收敛级数的和.

应当指出的是，只有满足了非负性和规范性的数列 $p_1,p_2,\cdots,p_k,\cdots$ 才能成为随机变量的分布列.

例 1（确定分布列中的参数） 设离散型随机变量的分布列为

X	0	1	2
P	0.2	a	0.5

求常数 a.

解 由分布列的规范性知

$$1=0.2+a+0.5,$$

所以

$$a=0.3.$$

例 2（确定分布列中的参数） 设离散型随机变量的分布列为

$$P\{X=k\}=a\left(\frac{1}{3}\right)^k, \quad k=1,2,\cdots,$$

求常数 a.

解 由分布列的规范性知

$$1=\sum_{k=1}^{\infty}a\left(\frac{1}{3}\right)^k=a\frac{\frac{1}{3}}{1-\frac{1}{3}}=\frac{1}{2}a,$$

所以

$$a=2.$$

例 3（求分布列） 抛掷一枚质量均匀的硬币两次，求正面朝上次数的分布列.

解 设正面朝上的次数为 X，则 X 的可能取值为 0,1 和 2.其相应的概率为

$$P\{X=0\}=0.25, P\{X=1\}=0.5, P\{X=2\}=0.25,$$

故所求分布列为

X	0	1	2
P	0.25	0.5	0.25

例 4(已知分布列求事件的概率)　设离散型随机变量 X 的分布列为

X	-1	0	1	2
P	0.1	0.2	0.3	0.4

求(1) $P\{X<0\}$;(2) $P\{|X|=1\}$;(3) $P\{X\geqslant 0\}$.

解　(1) $P\{X<0\}=P\{X=-1\}=0.1$;

(2) $P\{|X|=1\}=P\{X=-1\}+P\{X=1\}=0.1+0.3=0.4$;

(3) $P\{X\geqslant 0\}=P\{X=0\}+P\{X=1\}+P\{X=2\}=0.2+0.3+0.4=0.9$

或 $P\{X\geqslant 0\}=1-P\{X<0\}=1-0.1=0.9$.

三、几种常见的离散型分布

1. 两点分布

若随机变量 X 只有两个可能取值 0 和 1,且

$$P\{X=1\}=p,P\{X=0\}=q,$$

其中 $0<p<1$, $q=1-p$, 则称 X 服从两点分布(或 0-1 分布),其分布列也可记为

X	0	1
P	q	p

两点分布可简记为 $X\sim B(1,p)$. 凡是只取两种状态或可归结为两种状态的随机试验均可用两点分布来描述.如新生儿是男是女,检查产品质量是否合格,某单位的电力消耗是否超过负荷,单选题的答案是否正确等等.

2. 二项分布

在实践中,经常会遇到这样的一种随机试验:试验可以在相同条件下重复进行 n 次,且每次试验的结果互不影响,即各次试验是独立的;做一次试验只可能出现两种结果 A 和 $\overline{A}$; 每次试验 A 出现的概率都相同,若记 $P(A)=p$ ($0<p<1$),则 $P(\overline{A})=1-p=q$. 具有以上特点的随机试验称为 n 重伯努利试验.如抛掷同一枚硬币 10 次就是 10 重伯努利试验.

在 n 重伯努利试验中,事件 A 发生的次数 X 是个随机变量,其取值为 $0,1,2,\cdots,n$. 事件 A 在某指定的 k 次试验中发生而在其余的 $n-k$ 次试验中不发生的概率为 $p^k q^{n-k}$ (为什么?),而 A 发生的 k 次在总共 n 次试验中出现有 C_n^k 种不同的情况,它们又是两两互斥的,于是事件 A 在 n 重伯努利试验中发生 k 次的概率为 $P\{X=k\}=\mathrm{C}_n^k p^k q^{n-k}$. 称随机变量 X 所服从的分布为二项分布.

若随机变量 X 的分布列为

$$P\{X=k\}=\mathrm{C}_n^k p^k q^{n-k},k=0,1,2,\cdots,n,$$

其中 $0<p<1$, $p+q=1$, 则称随机变量 X 服从参数为 n,p 的二项分布,记作 $X\sim B(n,p)$.