# 随机数据处理方法

（修订本）

常兆光　王清河
宋岱才　何苏阳　编著

石油大学出版社

# 随机数据处理方法

（修订本）

常兆光　王清河
宋岱才　何[redacted]　编著

石油大学出版社

## 内 容 简 介

本书系统介绍处理随机数据的统计方法和统计方法的应用。内容包括：概率论、数理统计初步、回归分析、方差分析、正交试验设计、判别分析、聚类分析、主成分分析、因子分析、P.P. 判别、残差图分析等。对非线性回归、稳健统计也作了简单介绍。书中收录了许多实际例子。各章有例题及习题，书末有附录。

本书可作为应用统计工作者和工程技术人员的参考书，也可作为高等院校非数学专业本科生、研究生教材。

**随机数据处理方法**

（修订本）

常兆光　王清河
宋岱才　何苏阳　编著

石油大学出版社出版发行
（山东省东营市）
新华书店经销

山东电子工业印刷厂印刷
*
开本 850×1168　1/32　11.375 印张　296 千字
1997 年 8 月第 2 版　　1997 年 8 月第 2 次印刷
印数 3001－6500 册

ISBN 7-5636-0997-0/O・53
定价：11.80 元

# 前　言

概率统计是数学的一个分支，其方法已经广泛应用于许多领域，如地质、石油、气象、水文等。特别是近十几年来，由于电子计算机的普及，它的发展更加活跃和深入，其方法已普遍受到人们的重视。

概率统计方面的书已有很多，多数是将一元统计与多元统计分开。有的偏重于理论，有的偏重于方法，而以应用为主的很少介绍近代统计方法。为了弥补这种不足，我们编写了本书，既顾及理论，也强调方法。我们采取的原则是：(1) 基本概念和基本方法详细介绍，对工程中应用较广泛的方法着重从应用的角度介绍其应用背景、统计思想、应用条件及具体实现步骤；(2) 强调工程应用，尽可能介绍现场应用统计方法的例子，提高读者用统计方法处理实际问题的能力；(3) 在介绍基本统计方法的同时，尽可能介绍近年发展起来的一些新的统计方法，如投影寻踪(P. P. 方法)、回归诊断、非线性回归、稳健统计等。(4) 对古典内容作适当压缩，使读者在有限的篇幅内能了解随机数据处理方法的概貌。

王才经教授审阅了本书的全稿，并提出了许多宝贵意见。本书的出版还得到了石油大学出版社的大力支持，我们在此一并表示衷心的感谢。

限于编者水平，书中不妥之处还请专家、同仁批评指正。

编　者

1992 年 9 月 22 日

# 再版前言

第一版出版后，曾收到、听到许多读者的热情意见和建议，现依据1995年国家教委高教司重新修订的《概率论与数理统计课程教学基本要求》，结合我们教学工作的体会，对本书作了如下修改：(1) 对若干内容及表达方式进行了修改，并将原第一章作了适当补充，分为四章；(2) 对印刷错误作了校正；(3) 增加了几何概型、中心极限定理、单边检验、二正态总体未知参数的区间估计等内容；(4) 对原有例题作了适当调整补充，并增加了部分习题，使之题型更全面、更有代表性，更注意统计方法的应用性。

本书再版的修订过程中，得到了王才经教授、李元教授、施宝正教授、张广禄教授的大力支持和帮助；李元教授审阅了全书，并提出若干修改建议。在此一并表示感谢。竭诚地希望读者继续对本书提出批评建议。

编　者

1997年4月

此为试读，需要完整PDF请访问：www.ertongbook.com

# 目　录

# 第一章　随机事件与概率

在自然界和社会中，人们观察到的现象大体上可分为两类。一类是在一定条件下必然发生的现象，如上抛的石子必然下落等等。这类现象称为必然现象。另一类是在一定条件下可能发生，也可能不发生的现象。如抛一枚硬币可能正面（数字）朝上，也可能反面（国徽）朝上；掷一颗骰子可能出现的点数为 1，2，3，…6。这类现象称为随机现象。随机现象看起来杂乱无章，似乎毫无规律可言，但是人们经过大量观察和研究发现，随机现象的发生是具有某种规律性的。概率论与数理统计就是研究和揭示随机现象统计规律性的一门数学学科。

## §1　随机试验与随机事件

### 一、随机试验

在工农业生产和现实生活中，我们遇到过各种各样的试验。在这里，我们把试验作为一个含义广泛的术语，它包括各种各样的科学试验，甚至对某一事物的某一特征的观察。例如：

$E_1$：抛一枚硬币，观察正（$H$）、反（$T$）面出现的情况；

$E_2$：将一枚硬币抛二次，观察正、反面出现的情况；

$E_3$：将一枚硬币抛二次，观察正面出现的次数；

$E_4$：掷一颗骰子，观察出现的点数；

$E_5$：一射手进行射击，直到击中目标为止，观察射击次数；

$E_6$：在一批灯泡中任意抽取一只，测其寿命。

以上几个试验具有以下三个共同的特点：

(1) 可以在相同条件下重复进行。

(2) 每次试验的可能结果不止一个，但能事先明确试验的所有可能结果。

(3) 进行一次试验之前，不能确定哪一个结果会出现。

我们把具有上述三个特点的试验称为随机试验，以后简称为试验，记为 $E$。

## 二、随机事件与样本空间

我们将试验 $E$ 的所有可能出现的结果组成的集合称为 $E$ 的样本空间，记为 $\Omega$。$\Omega$ 中的每个元素称为样本点。例如上述试验 $E_i$ $(i=1,2,3,4,5,6)$的样本空间 $\Omega_i$ 分别为：

$\Omega_1=\{H,T\}$；

$\Omega_2=\{HH,TT,HT,TH\}$；

$\Omega_3=\{0,1,2\}$；

$\Omega_4=\{1,2,3,4,5,6\}$；

$\Omega_5=\{1,2,3,4,\cdots\}$；

$\Omega_6=\{t\mid t\geqslant 0\}$。

需要指出的是：样本空间中的元素是由试验目的所确定的。例如在 $E_2$ 和 $E_3$ 中同是将一枚硬币连抛二次，由于试验目的不同，其样本空间也不一样。

样本空间包含了试验 $E$ 的所有可能结果，每一个可能的结果称为随机事件，简称为事件。通常用大写字母 $A$、$B$、$C$、…等表示。只包含一个样本点的事件称为基本事件。例如掷骰子试验中，每一个可能出现的点数都是基本事件。而由两个或两个以上的基本事件(样本点)组成的事件称为复合事件。例如掷骰子试验中，$A_1=\{1,3,5\}$、$A_2=\{2,4,6\}$、$A_3=\{4,5,6\}$都是复合事件。

在每次试验中，一定发生的事件叫做必然事件，而一定不发生的事件叫做不可能事件，记作$\varnothing$。无论是必然事件、随机事件还是不可能事件，都是相对于“一定条件”而言的。条件发生变化，事件

性质也随之变化。例如，抛掷两颗骰子，“出现的点数之和为 5 点”，这是一个随机事件，若同时掷 6 颗骰子“出现的点数为 5 点”，则是不可能事件了。为了研究问题方便，通常把必然事件与不可能事件看成是特殊的随机事件。

由以上讨论知，事件是样本空间 $\Omega$ 的一个集合，因此可把集合运算推广到事件之间。

设 $E$ 的样本空间为 $\Omega$；$A$、$A_i$、$B_i(i=1,2,\cdots)$ 为事件。

(1) 包含：若 $A\subset B$，则称事件 $B$ 包含事件 $A$，它表示 $A$ 发生必然导致 $B$ 发生。如图 1-1 所示。

对任何一个事件 $A$，都有 $\varnothing\subset A\subset\Omega$。

(2) 相等：如果 $A\subset B$，且 $B\subset A$，则称事件 $A$ 与 $B$ 相等，记为 $A=B$。

(3) 和事件：事件 $A\cup B$ 称为事件 $A$ 与 $B$ 的和事件，它表示事件 $A$ 和 $B$ 至少有一个发生，如图 1-2 所示。

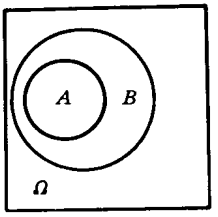


图 1-1

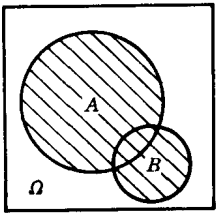


图 1-2

事件 $\bigcup_{i=1}^{n}=A_1\cup A_2\cup\cdots\cup A_n$ 称为事件 $A_1,A_2,\cdots,A_n$ 的和事件，它表示事件 $A_1,A_2,\cdots,A_n$ 中至少有一个事件发生。

类似地，事件 $\bigcup_{i=1}^{\infty}A_i=A_1\cup A_2\cup\cdots\cup A_n\cup\cdots$ 称为事件 $A_1,A_2,\cdots,A_n,\cdots$ 的和事件，表示事件 $A_1,A_2,\cdots$ 中至少有一个发生。

(4) 积事件：事件 $A\cap B$(或表示成 $AB$)称为事件 $A$ 与 $B$ 的积事件，它表示事件 $A$ 与 $B$ 同时发生，如图 1-3 所示。



此为试读，需要完整PDF请访问：www.ertongbook.com

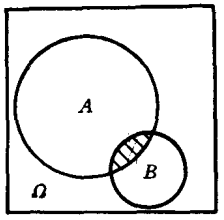


图 1-3

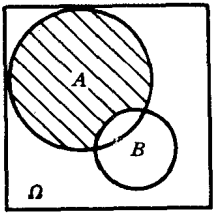


图 1-4

事件 $\bigcap_{i=1}^{n} A_i = A_1 \cap A_2 \cap \cdots \cap A_n$ 称为事件 $A_1, A_2, \cdots, A_n$ 的积事件，它表示事件 $A_1, A_2, \cdots, A_n$ 同时发生。

类似地，事件 $\bigcap_{i=1}^{\infty} A_i = A_1 \cap A_2 \cap \cdots$ 称为事件 $A_1, A_2, \cdots, A_n, \cdots$ 的积事件，它表示 $A_1, A_2, \cdots, A_n, \cdots$ 同时发生。

(5) 差事件：事件 $A-B$ 称为事件 $A$ 与 $B$ 的差事件，它表示事件 $A$ 发生而事件 $B$ 不发生，如图 1-4 所示。

(6) 互不相容事件：若 $A \cap B = \varnothing$，则称事件 $A$ 与 $B$ 互不相容。它表示事件 $A$ 与 $B$ 不能同时发生，如图 1-5 所示。基本事件是互不相容的。

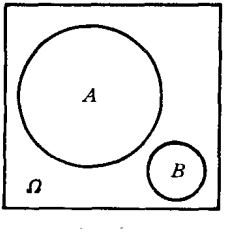


图 1-5

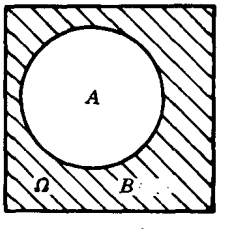


图 1-6

类似地，若 $\forall\ i \neq j, A_i A_j = \varnothing\ (i, j = 1, 2, \cdots)$ 则称事件 $A_1, A_2, A_3, \cdots, A_n, \cdots$ 是两两互不相容的。

(7) 对立事件：若事件 $AB=\varnothing$，且 $A\cup B=\Omega$，则称事件 $A$ 和 $B$ 互逆，又称事件 $A$ 是 $B$ 的对立事件(或事件 $B$ 是 $A$ 的对立事件)，记作 $A=\overline{B}$(或 $B=\overline{A}$)，如图 1-6。

需要注意，事件 $A$ 与 $B$ 互逆则一定互不相容，但互不相容不一定互逆。

根据以上所述，不难验证下列定律成立：

设 $A$、$B$、$C$ 为三个事件，则有：

交换律　$A\cup B=B\cup A$，$A\cap B=B\cap A$

结合律　$A\cup(B\cup C)=(A\cup B)\cup C$

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

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

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

德·莫根定律(对偶公式)

$$\begin{cases}\overline{A\cup B}=\overline{A}\cap\overline{B}, & \overline{A\cap B}=\overline{A}\cup\overline{B}\\ \overline{\bigcup\limits_{i=1}^{n}A_i}=\bigcap\limits_{i=1}^{n}\overline{A}_i, & \overline{\bigcap\limits_{i=1}^{n}A_i}=\bigcup\limits_{i=1}^{n}\overline{A_i}\end{cases}$$

**例** 1-1　设 $A$、$B$、$C$ 是三个事件，用 $A$、$B$、$C$ 的运算关系表示下列事件：

(1) $A$ 发生而 $B$ 与 $C$ 都不发生；

(2) $A$、$B$、$C$ 恰有一个发生；

(3) $A$、$B$、$C$ 至少有一个发生。

**解**　以上三个事件依次可表示为：

(1) $A\cap\overline{B}\cap\overline{C}$ 或 $A-B-C$；

(2) $(A\cap\overline{B}\cap\overline{C})\cup(\overline{A}\cap B\cap\overline{C})\cup(\overline{A}\cap\overline{B}\cap C)$；

(3) $A\cup B\cup C$。

**例** 1-2　在随机试验 $E_2$ 的样本空间 $\Omega_2$ 中，$A_1=\{HT,HH\}$ 表示第一次出现正面，$A_2=\{HH,TT\}$ 表示两次出现同一面，$A_3=\{HT,TH\}$ 表示只出现一次正面。则有：

$$A_1\cup A_2=\{HH,HT,TT\}$$

$$A_1 \cap A_2 = \{HH\}, \qquad A_1 - A_2 = \{HT\}$$

$A_2 \cup A_3 = \Omega_2$，且 $A_2A_3 = \varnothing$，因此，$A_2$ 与 $A_3$ 是互不相容的，并且是互逆的。

## §2 频率与概率

随机事件的发生是偶然的，在许多情况下，我们想知道的往往是随机事件发生的可能性大小。如建造水坝，为了确定坝高，需要知道建造水坝地段每年最大洪水达到某高度的可能性大小。在概率论中，将描述随机事件 $A$ 发生的可能性大小的数记为 $P(A)$，称为随机事件 $A$ 的概率。那么如何来确定事件的概率呢？一种方法是通过反复做试验来确定，为此先来讨论频率的概念。

### 一、频率

在掷硬币试验中，将一枚硬币在相同条件下连掷 $n$ 次，其中正面 $H$ 出现的次数记为 $n(H)$，则 $n(H)/n$ 在一定程度上能反映出现正面 $H$ 的可能性大小，将其记为 $f_n(H)=n(H)/n$。

表 1-1 列出了抛一枚均匀硬币次数 $n=5,50,500$，且均做 10 遍的 $n(H)$ 与 $f_n(H)$；历史上有人不厌其烦地进行了成千上万次试验，其结果见表 1-2。

表 1-1

| 实验序号 | $n=5$ | | $n=50$ | | $n=500$ | |
|---|---|---|---|---|---|---|
| | $n(H)$ | $f_n(H)$ | $n(H)$ | $f_n(H)$ | $n(H)$ | $f_n(H)$ |
| 1 | 2 | 0.4 | 22 | 0.44 | 251 | 0.502 |
| 2 | 3 | 0.6 | 25 | 0.50 | 249 | 0.498 |
| 3 | 1 | 0.2 | 21 | 0.42 | 256 | 0.512 |
| 4 | 5 | 1.0 | 25 | 0.50 | 253 | 0.506 |

续表 1-1

| 实验序号 | $n=5$ | | $n=50$ | | $n=500$ | |
|---|---|---|---|---|---|---|
| | $n(H)$ | $f_n(H)$ | $n(H)$ | $f_n(H)$ | $n(H)$ | $f_n(H)$ |
| 5 | 1 | 0.2 | 24 | 0.48 | 251 | 0.502 |
| 6 | 2 | 0.4 | 21 | 0.42 | 246 | 0.492 |
| 7 | 4 | 0.8 | 18 | 0.36 | 244 | 0.488 |
| 8 | 2 | 0.4 | 24 | 0.48 | 258 | 0.516 |
| 9 | 3 | 0.6 | 27 | 0.54 | 262 | 0.524 |
| 10 | 3 | 0.6 | 31 | 0.62 | 247 | 0.494 |

表 1-2

| 实验者 | $n$ | $n(H)$ | $f_n(H)$ |
|---|---|---|---|
| 蒲丰 | 4 040 | 2 048 | 0.507 0 |
| K·皮尔逊 | 12 000 | 6 019 | 0.501 5 |
| K·皮尔逊 | 24 000 | 12 012 | 0.500 5 |

由表 1-1 可见，即使试验的次数相同，$f_n(H)$也不尽相同，但总在$\frac{1}{2}$附近波动；将表 1-1 和表 1-2 结合起来看，当 $n$ 较小时，$f_n(H)$在$\frac{1}{2}$附近摆动幅度较大；随着 $n$ 的增大，摆动的幅度减小，逐渐稳定于$\frac{1}{2}$，这个值称为 $f_n(H)$的稳定值。通常把这个稳定值称为出现正面 $H$ 的概率，而 $f_n(H)$称为 $n$ 次试验中正面 $H$ 发生的频率。一般定义如下：

**定义** 1-1　设随机事件 $A$ 在 $n$ 次试验中出现的次数为 $n(A)$，则称比值 $n(A)/n$ 为事件 $A$ 在 $n$ 次试验中出现的频率。记为：

$$f_n(A)=\frac{n(A)}{n}$$

其中 $n(A)$ 称为 $A$ 出现的频数，并用 $P(A)$ 来表示 $f_n(A)$ 的稳定值。

由频率的定义，不难验证它有下列性质：

(1) 非负性：$0 \leqslant f_n(A) \leqslant 1$；

(2) 规范性：$f_n(\Omega)=1$；

(3) 可加性：若事件 $A$ 与 $B$ 互不相容，则

$$f_n(A \cup B) = f_n(A) + f_n(B) \tag{1-1}$$

若事件 $A_1, A_2, \cdots, A_n$ 两两互不相容，则

$$f_n(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} f_n(A_i) \tag{1-2}$$

由抛硬币试验可见，用频率来刻画一个事件 $A$ 出现的可能性大小比较直观，但是它有随机波动的缺陷。因此用其稳定值 $P(A)$ 来刻画事件 $A$ 出现的可能性大小是比较恰当的，但在实际中我们不可能对每一个事件都通过做大量的试验来获得 $P(A)$。为此，我们以频率的性质和频率的稳定值 $P(A)$ 为背景，采用抽象化方法给出概率的一般定义。

## 二、概率的公理化定义

**定义** 1-2　设 $E$ 是随机试验，$\Omega$ 是它的样本空间，如果对 $\Omega$ 中每一个事件 $A$ 赋予一个实数，记为 $P(A)$，满足：

(1) 非负性：$0 \leqslant P(A) \leqslant 1$；

(2) 规范性：$P(\Omega)=1$；

(3) 可加性：若 $A_1, A_2 \cdots, A_n, \cdots$ 两两互不相容，有

$$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i) \tag{1-3}$$

则称 $P(A)$ 为事件 $A$ 的概率。

概率 $P(A)$ 性质如下：

(1) $P(\varnothing)=0$；

(2) 若 $A_1, A_2, \cdots, A_n$ 为两两互不相容事件，则

$$P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i) \tag{1-4}$$

(1-4)式称为概率的有限可加性。(1)、(2)的证明请读者自行完成。

(3) $P(A)=1-P(\overline{A})$ (1-5)

**证明** 由于 $A\cup\overline{A}=\Omega$，而 $A\cap\overline{A}=\varnothing$，因此由性质(2)及规范性得：

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

所以

$$P(A) = 1 - P(\overline{A})$$

(4) 若 $A\subset B$，则

$$P(B - A) = P(B) - P(A) \tag{1-6}$$

**证明** 由于 $B=A\cup(B-A)$，且 $A\cap(B-A)=\varnothing$，所以由性质(2)

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

故得

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

由(1-6)知，当 $A\subset B$ 时，有 $P(A)\leqslant P(B)$。

(5) $P(A\cup B)=P(A)+P(B)-P(AB)$ (1-7)

**证明** 由于 $A\cup B=A\cup(B-AB)$，且 $A\cap(B-AB)=\varnothing$，所以由性质(2)、(4)得

$$\begin{aligned} P(A \cup B) &= P(A) + P(B - AB) \\ &= P(A) + P(B) - P(AB) \end{aligned}$$

由性质(5)可推出 $n$ 个事件的和事件概率公式：

$$P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i) - \sum_{1\leqslant i<j\leqslant n} P(A_iA_j) + \cdot$$

$$\sum_{1\leqslant i<j<k\leqslant n} P(A_iA_jA_k) + \cdots + (-1)^{n-1}P(A_1A_2\cdots A_n)$$

特别地，对于事件 $A_1$、$A_2$、$A_3$ 有

$$\begin{aligned} P(A_1\cup A_2\cup A_3) = &P(A_1)+P(A_2)+P(A_3)-P(A_1A_2) \\ &-P(A_1A_3)-P(A_2A_3)+P(A_1A_2A_3) \end{aligned}$$



此为试读，需要完整PDF请访问：www.ertongbook.com