

周富臣
易英

王生辉
周鹏翔

编著

常用 数理统计方法 及应用实例



中国计量出版社

CHINA METROLOGY PUBLISHING HOUSE



常用数理统计方法 及应用实例

周富臣 王生辉 编著
易 英 周鹏翔

中国计量出版社

图书在版编目 (CIP) 数据

常用数理统计方法及应用实例/周富臣, 王生辉, 易英, 周鹏翔编著. —北京: 中国计量出版社, 2006. 11

ISBN 7-5026-2518-6

I. 常… II. ①周… ②王… ③易… ④周… III. 数理统计—基本知识 IV. O212

中国版本图书馆 CIP 数据核字 (2006) 第 120332 号

内 容 提 要

本书是从事统计检验和统计质量控制人员的工具书, 也是本科生学习概率统计的理想辅导读物。内容包括: 随机变量及分布, 随机变量的数字特征, 数理统计的基本知识, 参数估计, 假设检验, 回归分析及方差分析。

本书叙述详细, 深入浅出, 便于自学, 具有较高的实用价值和可操作性。可供工矿企业、研究院所从事统计检验和统计质量控制的工程技术人员和管理人员使用, 也可供高等院校师生参考。

中国计量出版社出版

北京和平里西街甲 2 号

邮政编码 100013

电话 (010) 64275360

<http://www.zgjl.com.cn>

北京市密东印刷有限公司印刷

新华书店北京发行所发行

版权所有 不得翻印

*

787 mm×1092 mm 16 开本 印张 24 字数 466 千字

2006 年 11 月第 1 版 2006 年 11 月第 1 次印刷

*

印数 1—2 000 定价: 49.00 元

前 言

为了满足广大从事统计检验和统计质量控制的工程技术人员和管理人员的工作需要,给自学概率统计的人员提供一本理想的参考读物,笔者自甘寂寞,乐于与公式、曲线、数字打交道,在概率与统计的天地里跋涉不息,广收博采,披阅五载,增删多次,悉心编写了这本内容丰富、通俗易懂、信息量大、便于自学的读物。

概率与统计是数学的一个重要分支,它的内在逻辑别具一格,在诸多领域应用广泛。在信息化时代的今天,无论对自然科学还是社会科学,概率统计都显得尤为重要。气象与自然灾害的预报、天体未来位置的预测、一些社会问题的预测与控制、抽样调查、抽样检验、优秀的实验方案和生产方法的确定以及各种社会保险等,均离不开建立在概率论基础上的现代统计方法。

本书取材广泛,博采众长,内容翔实,深入浅出,说理清楚,剖析透彻,紧密联系生产和社会实践,具有较高的实用性和可操作性。

本书推崇经典,倡导实用,注重对理论、原理、公式、结论在实践中的应用及在什么条件下应用。对基本概念、基本理论、基本方法的叙述注意其实际意义的解释说明,并力求清晰、准确。对某些知识点,从不同角度给以阐述,并计算了大量数据列入表格,绘成曲线,使读者加深记忆,并有更深刻、更全面的理解。

本书紧密联系生产和社会生活,展现给读者许多典型实例。内容覆盖工业、农业、科研、商贸、医学、气象、保险、经济及日常生活等领域,其范围之广,数量之多,基本满足诸多学科的需要,为读者提供了广阔的借鉴和应用空间。

本书介绍了六种不同形式的正态分布表的使用及其关系,还介绍了三种不同形式的二项分布表及三种不同形式的泊松分布表的使用。这在其他著作中是很难看到的。 χ^2 分布表、 t 分布表和 F 分布表是收集多种文献上的资料汇集而成的,足以满足一般工作学习上的需要。

在假设检验中,对“同一问题取不同假设将得出不同结论”这一问题提出

了探讨性的看法。在参数估计中，给出了标准差 σ 的估计值 $\hat{\sigma}$ 的标准差 $\sigma_{\hat{\sigma}}$ 的计算公式。这些粗浅之见实为引玉之砖，恭请专家读者不吝赐教。

本书的编写，融入了作者三十多年从事计量检测和质量管理的实践经验，收录了所发表论文中的部分实例。本书由四人合编，王生辉编写第三章，易英编写第四章，周鹏翔编写第五章，周富臣编写第一、二、六、七、八章并通纂全书。

本书编写中，参阅了有关文献，引用了有关资料和例证，在此向原作者表示衷心的感谢。

由于作者水平所限，本书可能存在某些讹谬和不妥、疏漏和笔误，恳祈读者指正。

编著者谨识

2006年7月

目 录

第一章 随机变量及其分布	(1)
第一节 随机事件及概率	(1)
第二节 随机变量	(10)
第三节 几种离散型随机变量的分布	(11)
第四节 连续随机变量的分布函数和密度函数	(41)
第二章 几个重要的概率分布	(47)
第一节 正态分布	(47)
第二节 正态分布应用初步	(61)
第三节 χ^2 分布	(67)
第四节 t 分布	(75)
第五节 F 分布	(83)
第六节 几个分布之间的关系	(89)
第三章 随机变量的数字特征	(94)
第一节 数学期望	(94)
第二节 方差	(100)
第四章 数理统计的基本知识	(108)
第一节 总体与样本	(108)
第二节 统计量及其分布	(110)
第三节 有关样本均值的一些计算	(118)
第五章 参数估计	(122)
第一节 概述	(122)
第二节 期望和方差的点估计	(124)
第三节 正态总体标准差 σ 的估计	(133)
第四节 估计量的评价标准	(144)
第五节 正态总体参数的区间估计	(148)
第六章 假设检验	(173)
第一节 概述	(173)
第二节 一个正态总体的假设检验	(179)
第三节 两个正态总体的假设检验	(201)
第七章 回归分析	(222)
第一节 概述	(222)

第二节	相关图	(224)
第三节	一元线性回归	(228)
第四节	一元线性回归的预测与控制	(252)
第五节	用同一批数据作出两条回归直线的讨论	(258)
第六节	一元非线性回归	(262)
第七节	多元线性回归	(284)
第八节	回归方程的稳定性及两条回归直线的比较	(305)
第八章	方差分析	(311)
第一节	方差分析的意义	(311)
第二节	方差分析的基本思考方法	(311)
第三节	一个因素的方差分析	(313)
第四节	两个因素的方差分析	(329)
附录	常用数理统计表	(342)
附表 1	二项分布表 $P(\xi \leq x) = \sum_{i=0}^x C_n^i p^i q^{n-i} \quad x=0, 1, 2, \dots, n-1$	(342)
附表 2	泊松分布数值表 $P(\xi \leq n) = \sum_{d=0}^n \frac{\lambda^d}{d!} e^{-\lambda}$	(349)
附表 3	阶乘和阶乘的对数表	(354)
附表 4	正态分布的密度函数表 $\varphi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$	(356)
附表 5	正态分布表 $\Phi(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^u e^{-\frac{x^2}{2}} dx$	(357)
附表 6	正态分布的双侧分位数 (u_α) 表 $\alpha = 1 - \frac{1}{\sqrt{2\pi}} \int_{-u_\alpha}^{u_\alpha} e^{-\frac{u^2}{2}} du$	(359)
附表 7	χ^2 分布的上侧分位数 (χ_α^2) 表 $P(\chi_f^2 > \chi_\alpha^2) = \alpha$	(360)
附表 8	t 分布的双侧分位数 (t_α) 表 $P(t > t_\alpha) = \alpha$	(361)
附表 9	t 分布上侧分位数表 $P(t \geq t_\alpha) = \alpha$	(363)
附表 10	F 检验的临界值 (F_α) 表 $P(F > F_\alpha) = \alpha$	(364)
附表 11	符号检验表 $P(S \leq S_\alpha) = \alpha$	(372)
附表 12	检验相关系数 $\rho=0$ 的临界值 (r_α) 表 $P(r > r_\alpha) = \alpha$	(373)
附表 13	多元回归相关系数表	(374)
参考文献		(375)

第一章 随机变量及其分布

第一节 随机事件及概率

一、三类事件

概率论与数理统计是研究随机事件的数量规律的科学。所谓事件就是试验或观测的一种结果。事件可分为三类，即必然事件、不可能事件和随机事件。

在一定条件下必然发生的事件称为“必然事件”。例如，“在标准大气压下，纯水加热到 100°C 一定会沸腾”；“气体在一定条件下满足状态方程 $pV=nRT$ ”等都是必然事件。反之，在一定条件下，必然不发生的事件，称为“不可能事件”。例如，“在室温下，生铁会熔化”；“太阳西出东落”等都是不可能事件。通常用 U 表示必然事件，用 V 表示不可能事件。

在一定条件下，可能发生，也可能不发生的事件称为“随机事件”。简称“事件”，用 ξ 表示。例如，投掷一枚均匀的硬币，“出现正面”及“出现反面”都是随机事件；在工艺正常条件下，加工一批图样规定 $\phi 10 \pm_{0.01}^{0.03} \text{ mm}$ 的轴，加工后“零件尺寸恰好在 $10.01 \text{ mm} \sim 10.02 \text{ mm}$ 之间”也是一个随机事件。

二、随机事件

随机事件在一定条件下，可能出现，也可能不出现。例如，抛掷硬币，每次出现正面或反面纯粹是偶然的。但偶然性和必然性并不是永远互相排斥的两个范畴。虽然随机事件出现之前，不能确切地指出它是出现还是不出现。但是，实践经验证明，各种随机事件发生的可能性是有大小之分的。某随机事件发生可能性的大小，是这个事件的固有属性，它不依赖于人们的主观意识。

随机事件既有偶然性，也有规律性。它的规律性可以通过大量重复试验观测到。著名的统计学家蒲丰、皮尔逊等曾做过大量投掷硬币的试验，其结果见表 1-1。

表 1-1 投掷硬币的试验结果

实验者	投掷次数	出现正面的次数	出现正面的频率
Demorgan	2048	1061	0.5181
Buffon	4040	2048	0.5069
Pearson	12000	6019	0.5016
Pearson	24000	12012	0.5005

由此可以看出，出现正面的频率总在 0.5 附近波动，投掷次数越多，频率越接近 0.5。

这个 0.5 就反映出现正面的可能性的大小, 即投掷均匀硬币这个随机事件的统计规律。

每个随机事件出现的频率都有一个常数与之对应, 即频率具有稳定性。将试验次数无限增大时, 所逐渐稳定的那个常数就定义为该随机事件的概率。概率统计是研究客观世界中随机事件统计规律的一门学科, 它揭示偶然性与必然性之间的联系, 指明在什么条件下, 偶然性向必然性转化。概率则是随机事件出现的可能性大小的度量。

显然, 必然事件 U 的概率为 1, 即

$$P(U)=1$$

不可能事件 V 的概率为 0, 即

$$P(V)=0$$

随机事件 ξ 的概率总在 $0 \sim 1$ 之间, 即

$$0 \leq P(\xi) \leq 1$$

虽然必然事件和不可能事件并非随机事件, 但在实际问题中, 把必然事件和不可能事件看成是随机事件的特殊情况往往是方便的, 有时并不将它们严格区分。

三、小概率事件

当某一事件的概率非常小, 与“0”非常接近时, 称之为“小概率事件”。小概率事件不是不可能事件, 但是它在大量次数的试验中出现的频率很小, 也可以讲在一次试验中出现的频率很小。因此, 通常认为在一次试验中, 小概率事件几乎不会出现。

通常把概率值为 0.01, 0.05 的事件称为小概率事件。

四、随机事件的几个常见关系

1. 事件的包含与相等

若事件 A 发生必然导致事件 B 发生, 则称事件 B 包含事件 A , 记作 $B \supset A$ 或 $A \subset B$ 。

例 1-1 抛掷骰子, 令 A = “出现 2”, B = “出现偶数”, 则 $B \supset A$ 。

例 1-2 掷两枚匀称硬币, 令 A = “正好一个正面向上”, B = “至少一个正面向上”, 则 $A \subset B$ 。

若事件 A 包含事件 B , 同时事件 B 也包含事件 A , 则称事件 A 与 B 相等 (或称等价), 记作 $A=B$ 。例 1-1 中, 令 A = “出现 2, 4, 6”, B = “出现偶数”, 则 $A=B$ 。

2. 事件的和 (也称和事件)

事件 A 与事件 B 至少有一个发生的事件称为事件 A 与 B 的和, 记作 $A+B$ 或 $A \cup B$ 。“ $A+B$ ”发生意味着: 或事件 A 发生, 或事件 B 发生, 或事件 A 和事件 B 都发生。事件 A 与 B 的和是一事件。

例 1-3 令 A = “随机点落入图 1-1 (a) 中的小圆内”, B = “随机点落入图 1-1 (b) 的大圆内”, 则事件 $A+B$ = “随机点落入任一圆内”, 如图 1-1 (c) 所示。

类似地, 随机事件 $A_1, A_2, \dots, A_n$ 的和是一事件, 它表示事件 $A_1, A_2, \dots, A_n$ 中至少有一事件发生。记作 $A_1+A_2+\dots+A_n$, 或 $A_1 \cup A_2 \cup \dots \cup A_n = \bigcup_{k=1}^n A_k$ 。

3. 事件的积 (也称积事件)

随机事件 A 与 B 的积是一事件, 它表示事件 A 与 B 都发生。记作 $A \cdot B$ 或 AB 或 $A \cap B$

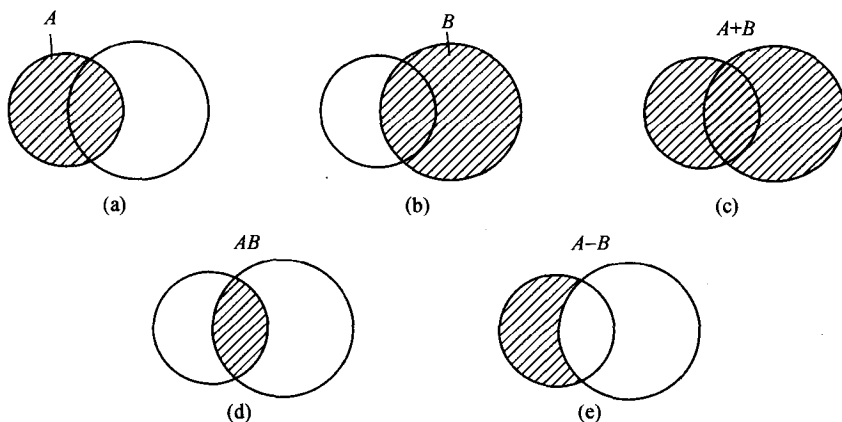


图 1-1 几个事件的示意

B。在例 1-3 中，事件 AB 就是随机点落入两圆的公共部分内，如图 1-1 (d) 所示。

类似地，随机事件 $A_1, A_2, \dots, A_n$ 的积是一事件，它表示事件 $A_1, A_2, \dots, A_n$ 都发生。记作 $A_1 A_2 \dots A_n$ ，或 $A_1 \cap A_2 \cap \dots \cap A_n = \bigcap_{k=1}^n A_k$ 。

4. 事件的差（差事件）

事件 A 发生而事件 B 不发生的事件称为事件 A 与事件 B 的差事件，记作 $A-B$ 。

在例 1-3 中，事件 $A-B$ 就是图 1-1 (e) 所示的阴影部分。

5. 互不相容事件（互斥事件）

若事件 A 与事件 B 不能同时发生，即 $AB=V$ ，则称事件 A 与事件 B 是互不相容事件，简称事件 A 与事件 B 互不相容。

在例 1-1 中，令 A = “出现偶数”， B = “出现奇数”，则 $AB=V$ ，即 A 与 B 互不相容。

对于 n 个事件 $A_1, A_2, \dots, A_n$ ，若它们两两互不相容，则称 $A_1, A_2, \dots, A_n$ 是互不相容的。

6. 对立事件

“事件 A 发生”与“事件 A 不发生”称互为对立事件。事件 A 的对立事件记作 $\bar{A}$ 。不难看出， A 与 $\bar{A}$ 也是互不相容的。应指出，对立事件一定是互不相容事件，但互不相容事件不一定是对立事件。在例 1-1 中，令 A = “出现偶数”， B = “出现奇数”，则 $B=\bar{A}$ ，同理 $A=\bar{B}$ 。

由对立事件的定义和差事件的定义知

$$A-B=A\bar{B}$$

这从图 1-1 (e) 也可看出，该图表示了以下三个事件：

$$A-B \quad A\bar{B} \quad A-A\bar{B}$$

对立事件 A 和 $\bar{A}$ 满足： $A\bar{A}=V$ （不可能事件）， $A+\bar{A}=U$ （必然事件）。显然， A 与 $\bar{A}$ 互不相容。

7. 互不相容事件完备组（完备群）

若事件 $A_1, A_2, \dots, A_n$ 互不相容，又在一次试验中，事件 $A_1, A_2, \dots, A_n$ 必发生其

中之一，即

$$A_1 + A_2 + \cdots + A_n = U$$

又

$$A_i A_j = V \quad (i \neq j)$$

则称 $A_1, A_2, \dots, A_n$ 为互不相容事件完备组。

显然， A 与 $\bar{A}$ 构成互不相容事件完备组。

8. 独立事件

若两事件 A 与 B ，其中任一事件发生与否都不影响另一事件发生的可能性，则 A 与 B 相互独立。例如，有放回抽样的每次抽样都是独立的。

五、概率加法公式

1. 互不相容事件

若事件 A, B 互不相容，则

$$P(A+B) = P(A) + P(B) \quad (1-1)$$

式 (1-1) 表达了概率的最重要的特性：可加性。

由于 $A + \bar{A} = U$ ， A 与 $\bar{A}$ 互不相容，由式 (1-1) 可知

$$P(A) + P(\bar{A}) = P(A + \bar{A}) = P(U) = 1 \quad (1-2)$$

从而得

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

或

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

将式 (1-1) 推广，设 n 个事件 $A_1, A_2, \dots, A_n$ 互不相容，则

$$P(A_1 + A_2 + \cdots + A_n) = P(A_1) + P(A_2) + \cdots + P(A_n) \quad (1-3)$$

若 $A_1, A_2, \dots, A_n$ 构成互不相容事件完备组，即

$$A_1 + A_2 + \cdots + A_n = U$$

则

$$P(A_1 + A_2 + \cdots + A_n) = P(U) = 1$$

即

$$P(A_1 + A_2 + \cdots + A_n) = P(A_1) + P(A_2) + \cdots + P(A_n) = 1$$

2. 任意事件

对任意两事件 A, B ，有

$$P(A+B) = P(A) + P(B) - P(AB) \quad (1-4)$$

下面介绍用加法公式 (1-1) 来证明公式 (1-4)：

由于 $A+B = A + B\bar{A}$ 且 A 和 $B\bar{A}$ 互不相容，由式 (1-1) 知

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

又 $B = BA + B\bar{A}$ [参看图 1-1 (b)] 且 BA 和 $B\bar{A}$ 也互斥，则

$$P(B) = P(BA + B\bar{A}) = P(BA) + P(B\bar{A})$$

故有

$$P(B\bar{A}) = P(B) - P(BA) \quad ②$$

将式②代入式①即得式 (1-4)。

在式 (1-4) 中，当事件 A, B 互斥时，则 $AB = V$ ，故 $P(AB) = 0$ ，从而式 (1-4) 变为式 (1-1)，即式 (1-1) 是式 (1-4) 的特殊情况。

顺便指出,也可按 $A+B=A\bar{B}+\bar{A}B+AB$ [参看图 1-1 (c)] 来证式 (1-4)。

对任意三事件 A, B, C , 有

$$P(A+B+C)=P(A)+P(B)+P(C)-P(AB)-P(BC)-P(AC)+P(ABC) \quad (1-5)$$

例 1-4 箱中有球 30 个, 其中白球 5 个, 黄球 3 个, 红球 8 个, 黑球 10 个, 蓝球 4 个。求取得一个亮球 (白或黄球) 的概率, 取得一个暗球的概率。

解 设 A = “取得白球”, B = “取得黄球”, C = “取得亮球”, D = “取得红球”, E = “取得黑球”, F = “取得蓝球”, H = “取得暗球”。显然 $C=A+B$, $H=D+E+F$ 。由于 A, B, D, E, F 均为两两互不相容事件, 故

$$P(C)=P(A+B)=P(A)+P(B)=\frac{5}{30}+\frac{3}{30}=\frac{8}{30}$$

$$P(H)=P(D+E+F)=P(D)+P(E)+P(F)=\frac{8}{30}+\frac{10}{30}+\frac{4}{30}=\frac{22}{30}$$

由此可知:

$$P(A)+P(B)+P(D)+P(E)+P(F)=\frac{5}{30}+\frac{3}{30}+\frac{8}{30}+\frac{10}{30}+\frac{4}{30}=1$$

这是由于事件 A, B, D, E, F 是互斥的, 在一次抽取中其中之一必然发生, 它们构成互斥事件完备组, 即构成互斥事件完备组的这些事件的概率的和等于 1。若 $A_1, A_2, \dots, A_n$ 构成互斥事件完备组, 则有

$$P(A_1)+P(A_2)+\dots+P(A_n)=1 \quad (1-6)$$

由此可知:

$$P_{\text{亮}}+P_{\text{暗}}=P(C)+P(H)=\frac{8}{30}+\frac{22}{30}=1$$

任取一球时, 非暗球则亮球, 非亮球则暗球, 这是仅有两个互不相容事件构成的完备组。这两个事件是对立事件。对立事件的概率的和等于 1。式 (1-2) 表达的就是这个结论。

例 1-5 箱中有 100 个产品, 其中有 3 个次品, 现从中任取 5 个。问①恰有 1 个次品的概率; ②至少有 1 个次品的概率。

解 ① $P=\frac{C_3^1 \cdot C_{97}^4}{C_{100}^5}=0.138064$

② “至少有 1 个次品” 是 “恰有 1 个次品 4 个正品”、“恰有 2 个次品 3 个正品”、“恰有 3 个次品 2 个正品” 这 3 个互斥事件的和。令 A = “至少有 1 个次品”, A_i = “恰有 i 个次品, $5-i$ 个正品” ($i=1, 2, 3$), 则 $A=A_1+A_2+A_3$, 又 A_1, A_2, A_3 互不相容, 由概率加法公式得

$$P(A)=P(A_1+A_2+A_3)=P(A_1)+P(A_2)+P(A_3)=\frac{C_3^1 \cdot C_{97}^4}{C_{100}^5}+\frac{C_3^2 \cdot C_{97}^3}{C_{100}^5}+\frac{C_3^3 \cdot C_{97}^2}{C_{100}^5}=0.144$$

或按对立事件解。“至少有 1 个次品” 的对立事件是 “全是正品”。而 “全是正品” 的概率为

$$P(\bar{A})=\frac{C_{97}^5}{C_{100}^5}=0.855998$$

故

$$P(A)=1-P(\bar{A})=1-0.855998=0.144002 \approx 0.144$$

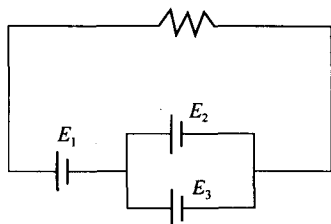


图 1-2 例 1-6 示意图

例 1-6 电路由电池 E_1 和两个并联的电池 E_2 及 E_3 串联而成 (图 1-2)。设电池 E_1, E_2, E_3 损坏的概率分别是 0.3, 0.2, 0.2, 求电路断电的概率。

解 设 $A = “E_1 损坏”$, $B = “E_2 与 E_3 同时损坏”$, $C = “电路断电”$, 则 $C = A + B$, 而 $P(A) = 0.3$, $P(B) = 0.2 \times 0.2 = 0.04$, 故电路断电的概率为

$$\begin{aligned} P(C) &= P(A+B) = P(A) + P(B) - P(AB) \\ &= 0.3 + 0.04 - 0.3 \times 0.04 \\ &= 0.328 \end{aligned}$$

这里, 事件 A 与 B 不是互斥事件而是任意事件, 因 E_1, E_2, E_3 的损坏是随机的。

六、条件概率和概率乘法公式

1. 条件概率

若 A, B 是条件 S 下的两个随机事件, $P(A) \neq 0$, 则称在 A 发生的前提下 B 发生的概率为条件概率, 记作 $P(B|A)$ 。

例 1-7 有五个球 (三个新, 两个旧), 每次取出一个, 无放回地取两次。求①第一次取到新球的概率; ②第二次取到新球的概率; ③在第一次取得新球的条件下第二次取到新球的概率。

解 设 $A = “第一次取到新球”$, $B = “第二次取到新球”$, $B|A = “第一次取到新球后第二次取到新球”$, 则

$$\textcircled{1} P(A) = \frac{3}{5}$$

②事件 B 没有对第一次取到什么球作限制或假设, 因此下面分两种情况进行讨论。

情况一: 第一次取到的是新球, 第二次也取到新球。

显然, 第一次取到新球的概率为 $\frac{3}{5}$, 在此条件下, 第二次取到新球的概率为 $\frac{2}{4}$, 这 $\frac{2}{4}$ 是在 “ $\frac{3}{5}$ ” 这个附加条件下的 $\frac{2}{4}$, 所以第二次取到新球的概率为 $\frac{3}{5} \times \frac{2}{4} = \frac{6}{20}$ 。

情况二: 第一次取到的是旧球, 第二次才取到新球。

易知, 第一次取到旧球的概率为 $\frac{2}{5}$, 在这之后, 第二次取到新球的概率为 $\frac{3}{4}$, 这 $\frac{3}{4}$ 是在 “ $\frac{2}{5}$ ” 这个条件下的 $\frac{3}{4}$, 所以第二次取到新球的概率为 $\frac{2}{5} \times \frac{3}{4} = \frac{6}{20}$ 。

故第二次取到新球的概率为

$$P(B) = \frac{6}{20} + \frac{6}{20} = \frac{3}{5}$$

由此可见, 第一次取到新球与第二次取到新球的概率是相等的。这就是抓阄这个公认为公平方法的道理所在。可以断言, 在第一次、第二次取球后, 第三次取得新球的概率也为 $\frac{3}{5}$ 。

③第一次取到新球后, 剩下两个新球、两个旧球共四个, 故第二次取到新球的概率为

$$P(B|A) = \frac{2}{4}$$

2. 乘法公式

(1) 乘法公式 1

条件概率 $P(B|A)$ 与事件的原概率的关系 [要求 $P(A) \neq 0$] 为

$$P(B|A) = \frac{P(AB)}{P(A)} \quad (1-7)$$

式 (1-7) 是从大量社会实践中总结出来的普遍规律, 不是用纯数学方法推导出来的。但在古典概型的情况下可用数学方法给出证明 (证略)。

由式 (1-7) 可得

$$P(AB) = P(A)P(B|A)$$

同理可得

$$P(A|B) = \frac{P(AB)}{P(B)} \quad (1-8)$$

$$P(AB) = P(B)P(A|B)$$

故

$$P(AB) = P(A)P(B|A) = P(B)P(A|B) \quad (1-9)$$

式 (1-9) 可叙述为二事件的积的概率等于其中一事件的概率与另一事件在另一事件已发生的条件下的条件概率的乘积。

将乘法公式推广到有限个随机事件的情况, 有

①若 $P(AB) \neq 0$, 则

$$P(ABC) = P(A)P(B|A)P(C|AB) \quad (1-10)$$

②对任意 n 个事件 $A_1, A_2, \dots, A_n$, 若 $P(A_1 A_2 \dots A_{n-1}) > 0$, 则

$$P(A_1 A_2 \dots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 A_2) \dots P(A_n|A_1 A_2 \dots A_{n-1}) \quad (1-11)$$

即有限个事件的积的概率等于这些事件的概率的乘积, 其中每一事件的概率是在它前面的一切事件都已发生的条件下的概率。

例 1-8 有五个球 (三个新球、两个旧球), 作无放回抽取, 每次取一件, 连取两次, 求取出的两个均为新球的概率。

解 令 $A =$ “第一次取得新球”, $B =$ “第二次取得新球”, $C =$ “取出的两个均为新球”, 则 $C = AB$, 故

$$P(C) = P(AB) = P(A)P(B|A) = \frac{3}{5} \times \frac{2}{4} = \frac{6}{20} = \frac{3}{10}$$

或用古典概型方法:

$$P(C) = \frac{C_3^2}{C_5^2} = \frac{3}{10}$$

(2) 乘法公式 2

所谓事件 A, B 相互独立, 就是一事件的发生并不影响另一事件发生的概率。

若事件 B 的发生不影响事件 A 的概率, 即

$$P(A|B) = P(A) \quad (1-12)$$

则称事件 A 对事件 B 是独立的。

若事件 A 对事件 B 是独立的, 则事件 B 对事件 A 也是独立的。因为如果

$$P(A|B)=P(A)$$

则由式 (1-9) 可得

$$P(B|A)=P(B) \quad (1-13)$$

由此可见, 随机事件的独立性是一种相互对称的性质。故又可将独立性定义为: 若二事件中任何一事件的发生不影响另一事件的概率, 则称它们是独立的。若二事件 A 与 B 独立, 则下列各对事件: A 与 $\bar{B}$, $\bar{A}$ 与 B , $\bar{A}$ 与 $\bar{B}$ 都是独立的。

若二事件 A 与 B 独立, 则乘法公式 (1-9) 变为

$$P(AB)=P(A)P(B) \quad (1-14)$$

即二独立事件的积的概率等于这二事件的概率的乘积。

对于有限个独立事件, 则乘法公式 (1-11) 变为

$$P(A_1 A_2 \cdots A_n)=P(A_1)P(A_2) \cdots P(A_n) \quad (1-15)$$

即有限个独立事件的积的概率等于这些事件的概率的乘积。

例 1-9 有五个球 (三个新球, 两个旧球), 从中每次取一个, 有放回地连续取两次, 求两次取得都是新球的概率。

解 此例与例 1-8 的区别为: 这里是有放回抽样。故两次抽样是独立的, 则

$$P(A_1 A_2)=P(A_1)P(A_2)=\frac{3}{5} \times \frac{3}{5}=\frac{9}{25}$$

例 1-10 一工人看管三台车床, 在 1 h 内车床不需要工人照管的概率: 第一台为 0.9, 第二台为 0.8, 第三台为 0.7。求在 1 h 内三台车床至少有一台需要工人照管的概率。

解法一 令 A = “甲车床需照管”, B = “乙车床需照管”, C = “丙车床需照管”, 则 $P(A)=0.1$, $P(B)=0.2$, $P(C)=0.3$ 。 D = “至少有一台需照管”, 则 $D=A+B+C$, 故

$$P(D)=P(A+B+C)$$

$$=P(A)+P(B)+P(C)-P(AB)-P(BC)-P(AC)+P(ABC)$$

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

$$=0.1+0.2+0.3-0.1 \times 0.2-0.2 \times 0.3-0.1 \times 0.3+0.1 \times 0.2 \times 0.3$$

$$=0.496$$

这里之所以用任意事件的加法公式是由于 A, B, C 并不是互斥事件; 又由于 A, B, C 是独立的, 故用独立事件的乘法公式。

解法二 “至少有一台需照管”的对立事件是“三台都不需要照管”。令“三台都不需照管”= $\bar{A}\bar{B}\bar{C}$, 因 $\bar{A}, \bar{B}, \bar{C}$ 相互独立, 故

$$P(\bar{A}\bar{B}\bar{C})=P(\bar{A})P(\bar{B})P(\bar{C})=0.9 \times 0.8 \times 0.7=0.504$$

所求概率为

$$P=1-P(\bar{A}\bar{B}\bar{C})=1-0.504=0.496$$

例 1-11 条件同例 1-10, 求在 1 h 内三台车床最多有一台需要照管的概率。

解 “三台中最多有一台需要照管”是“三台都不需要照管”和“三台中恰有一台需要照管”这两个互斥事件的和。

令 $A =$ “甲车床不需照管”, $B =$ “乙车床不需照管”, $C =$ “丙车床不需照管”, 且 $P(A)=0.9$, $P(B)=0.8$, $P(C)=0.7$ 。

令 $ABC =$ “三台车床都不需要照管”, 因各车床独立, 故

$$P(ABC) = P(A)P(B)P(C) = 0.9 \times 0.8 \times 0.7 = 0.504$$

令 “三台中恰有一台需照管” $= \bar{A}BC + A\bar{B}C + AB\bar{C}$, 则

$$\begin{aligned} P(\bar{A}BC + A\bar{B}C + AB\bar{C}) &= P(\bar{A}BC) + P(A\bar{B}C) + P(AB\bar{C}) \\ &= P(\bar{A})P(B)P(C) + P(A)P(\bar{B})P(C) + P(A)P(B)P(\bar{C}) \\ &= 0.1 \times 0.8 \times 0.7 + 0.9 \times 0.2 \times 0.7 + 0.9 \times 0.8 \times 0.3 \\ &= 0.398 \end{aligned}$$

$$\begin{aligned} \text{故 } P(\text{三台中最多有一台需要照管}) &= P(ABC) + P(\bar{A}BC + A\bar{B}C + AB\bar{C}) \\ &= 0.504 + 0.398 \\ &= 0.902 \end{aligned}$$

这里, $\bar{A}BC$, $A\bar{B}C$, $AB\bar{C}$ 三事件互斥, A , B , C 相互独立。

七、贝努里概型 (独立试验序列概型)

1. 贝努里试验

在许多实际问题中, 对试验感兴趣的是试验中某事件 A 是否发生。例如, 在产品抽样检查中关心的是抽到废品, 还是抽到正品; 在掷硬币时关心的是出现正面还是反面……这种每次试验只有两个可能结果的试验称为贝努里 (Jacob Bernoulli, 瑞士, 1654 ~ 1705) 试验。

当然, 有些试验的结果不止两个。例如, 产品寿命是不小于 0 的任一数值, 可以根据需要可以把寿命大于 $k(h)$ 的当作合格品, 其余作为次品。这样, 这类问题也可归结为贝努里试验。

显然, 贝努里试验的两个结果是相互对立的事件 A 及事件 $\bar{A}$ 。对立事件的概率和等于 1, 即 $P(A) + P(\bar{A}) = 1$, 若

$$P(A) = p \quad P(\bar{A}) = 1 - p = q \quad (0 < p < 1)$$

显然, $p + q = 1$ 。

2. n 重贝努里试验 (独立试验序列)

在确定的条件下, 重复进行 n 次独立的贝努里试验, 称为 n 重贝努里试验。这样的一系列试验也称作 “独立试验序列”。这里的 “重复” 是指在每次试验中事件 A 的概率 $P(A)$ 保持不变, 从而事件 $\bar{A}$ 出现的概率也保持不变。所谓 “独立” 是指假设每次试验的结果与其他各次试验的结果无关, 即事件 A 的概率 $P(A)$ 在整个系列试验中保持不变。由于 n 次独立重复试验即独立试验序列是贝努里首先研究的, 故称 n 重独立重复试验为贝努里概型。

3. 贝努里概型的概率计算

贝努里概型是概率论中一个非常重要的概率模型。它是最早研究也是得到最多研究的模型之一, 在理论上具有重要意义, 在实践中有十分广泛的应用。例如, 在工业产品的质量检查中, 在群体遗传学中它都占有重要地位。

贝努里概率模型是 “在同样条件下进行重复试验或观察” 的一种数学模型。对这种模

型,关心的是在 n 次独立重复试验中,事件 A 恰好发生 i 次 ($0 \leq i \leq n$) 的概率 $P_n(i)$, $i = 0, 1, 2, \dots, n$ 。若在独立试验序列中事件 A 的概率为 p ($0 < p < 1$),则在 n 次试验中事件 A 恰好发生 i 次的概率为

$$P_n(i) = C_n^i p^i q^{n-i} = \frac{n!}{i! (n-i)!} p^i q^{n-i} \quad (i=0,1,2,\dots,n) \quad (1-16)$$

其中, $p+q=1$ 。

这就是后面要介绍的二项分布求概率公式。二项分布用于确定 n 重贝努里试验中事件 A 出现 i 次的概率。

若只进行一次贝努里试验,则或是事件 A 出现,或是事件 $\bar{A}$ 出现,其概率为 $P(A)=p$, $P(\bar{A})=1-p=q$ 。显然 $p \geq 0$, $q \geq 0$, 且 $p+q=1$ 。这是最简单的情况,称为贝努里分布即二点分布。

贝努里概型是一种十分重要的概率模型。贝努里分布、二项分布、几何分布、帕斯卡分布都属于贝努里概型中的分布。掷硬币、射击、有放回抽样(进行 n 次独立重复抽样)等都是贝努里概型。

第二节 随机变量

随机变量就是在随机试验的结果中能取得不同数值的量。例如,电话员在 1 min 内接到的呼唤次数;扔一枚骰子出现的点数;在一定的生产条件下,机床加工出来的产品尺寸与规定尺寸的偏差;射击时击中点与目标中心的偏差等。随机变量具有以下两个特点:

①取值的不确定性。由于随机事件在一次试验中可能出现,也可能不出现,具有一定的随机性,因此随机变量的取值也是随机的,它究竟取什么值,要随试验结果而定。

②随机变量的取值遵从一定的分布。虽然随机变量的取值是不确定的,但由于随机变量出现的可能性的规律,因此随机变量的取值也是有规律的,它遵从一定的分布。

按照随机变量可能取值的特点,可以把它们分为两个基本类型,即离散型随机变量和非离散型随机变量。

离散型随机变量仅可能取有限个或无穷可数多个的数值,即所有可能的取值能按照一定顺序排列起来,表示为数列 $x_1, x_2, \dots, x_n, \dots$ 。例如,电话员在 1 h 内接到的呼唤次数的可能值为 $0, 1, 2, \dots, n, \dots$ 。

若随机变量的所有可能取值不能按照一定的顺序排列起来,则称为非离散型随机变量。在范围很广的非离散型随机变量中,最重要、最常见的是连续型随机变量。

连续型随机变量可以取得某一区间内的任何数值。例如,机床加工的零件尺寸与规定尺寸的偏差;炮弹着地点与目标的距离;人的身高、体重等,均属于连续型随机变量。

常见的离散型随机变量的分布有:二点分布(也称 0-1 分布)、二项分布、泊松分布和超几何分布等。

常见的连续型随机变量的分布有:均匀分布、指数分布、正态分布、 χ^2 分布、 t 分布和 F 分布等。