高等学校规划教材

选煤工业数理统计方法及应用

边炳鑫 主编

煤炭工业出版社

高等学校规划教材

选煤工业数理统计方法及应用

主　编　边炳鑫
副主编　郭　德　丛桂芝　沈笑君

煤炭工业出版社

内 容 提 要

本书是按选矿专业本科教学大纲编写而成，系统地介绍了选煤生产、数据分析和试验设计中常用的数理统计的基本概念、原理和方法。全书分十章，包括概率基础、随机变量分布与特征参数、抽样分布、参数估计、统计检验、质量管理图、方差分析、正交试验设计、相关与回归分析、取样与检验。

本书基本概念叙述清晰，循序渐进，内容较为全面，应用性强，各章配有应用实例和习题，便于教学和自学。

本书作为高等院校矿物加工工程专业教材，也可供煤炭加工利用、选矿和有关领域的科技人员参考。

图书在版编目（CIP）数据

选煤工业数理统计方法及应用/边炳鑫主编. 一北京：煤炭工业出版社，1998

高等学校规划教材

ISBN 7－5020－1607－4

Ⅰ. 选… Ⅱ. 边… Ⅲ. 选煤－数理统计－高等学校－教材

Ⅳ. TD94

中国版本图书馆 CIP 数据核字（98）第 13686 号

高等学校规划教材

选煤工业数理统计方法及应用

主　编　边炳鑫

副主编　郭　德　丛桂芝　沈笑君

责任编辑：袁　筠

*

煤炭工业出版社　出版

（北京朝阳区霞光里 8 号　100016）

煤炭工业出版社印刷厂　印刷

新华书店北京发行所　发行

*

开本 787×1092mm $^{1}/_{16}$　印张 $14^{1}/_{4}$

字数 336 千字　印数 1—1,000

1998 年 10 月第 1 版　1998 年 10 月第 1 次印刷

书号 4376　定价 35.00 元

前　　言

当今，数理统计在生产管理和科学试验中的应用越来越广泛深入，它的重要性和实用性愈来愈被人们所认识。随着我国选煤事业的迅速发展，数理统计方法在选煤厂生产管理中的应用也得到了蓬勃的发展，所以煤炭系统的选矿专业都把《选煤工业数理统计方法及应用》课程列入教学计划。

本教材是在原有教学讲义的基础上，参考有关书籍，针对选矿专业本科教学大纲要求编写的。

编者力图把概念讲得清晰，循序渐进。为帮助学生理解概念，常用选煤生产实例加以说明。在讲解统计原理时，尽量多作直观解释，删去较长的数学证明。

本书包括的数理统计内容较为全面。统计推断是数理统计的核心，本书讲解统计推断时采用先讲参数点估计和区间估计，再讲假设检验的方式。这种方式的优点是两部分内容比较均衡。另外，为使读者便于阅读，第二章中扼要介绍了概率论的基本知识。

本书应用性较强。为了培养学生运用统计方法解决选煤生产实际问题的能力，书中配有大量选煤生产实例与习题，书后附有习题答案。希望把本书写成一本既便于教学也便于自学的教材。

本书可作为高等院校选矿专业教材，也可供煤炭加工利用、冶金和有关专业教学参考书。要求读者具备工科高等数学和概率论基本知识。

参加编写的有黑龙江矿业学院边炳鑫、沈笑君、宋志伟、李哲华北矿业高等专科学校郭德和煤炭科学总院唐山分院丛桂芝。具体编写分工：第一、四章边炳鑫，第二、七章李哲、宋志伟，第三章宋志伟，第六、十章郭德、丛桂芝，第八章沈笑君、郭德，第五、九章宋志伟、边炳鑫、丛桂芝。全书由边炳鑫整理、统稿。

本书承许占贤教授审阅，提出了许多宝贵意见，有关同志也予以支持和帮助，在此一并表示衷心感谢。

由于我们水平所限，书中不足和遗漏之处在所难免，恳请读者指正。

编　者

1998年6月

目　　录

第一章 绪　　论

一、数理统计的基本任务

自然界和社会上发生的现象是多种多样的。有一类现象，在一定条件下必然发生（或必然不发生），例如，向上抛一石子必然下落；同性电荷必不相互吸引等等。这类现象称为**确定性现象**。过去我们学过的微积分和线性代数等就是研究这类现象的数学工具。然而在自然界和社会中还存在着另一类现象，例如，在相同条件下抛同一枚硬币，其结果可能正面朝上，也可能是反面朝上，并且不论怎样控制抛掷条件，在每次抛掷之前无法肯定抛掷的结果。按相同的采样、制样及化验规程操作，对一批精煤产品的质量进行多次重复的测试，但每次所测得结果总是参差不齐。再如一个选煤厂，在原煤性质、工艺设备和操作条件相同的情况下，所选出精煤灰分，虽然总的来说比较稳定，但每一时刻所选出的精煤灰分难免会有波动。这类现象的共同特点是：在基本条件不变的情况下，所进行的一系列试验或重复观测会得到不同的结果，在每次试验或观测之前无法预知确切的结果，即呈现出不确定性。

科学研究发现造成结果不确定性的原因是由于试验结果除了受人们可以控制的基本条件影响外，还受许多没有被人们认识和掌握、或者即使认识到也无法控制，或者对试验结果影响太小没有必要加以控制的因素影响，例如，周围环境、气象条件、操作人员的情绪波动等；而且这些因素的影响往往是随机的、波动的，具有一种偶然性。这些因素称之为**随机因素**。

随机因素对结果的影响虽然具有偶然性，但在大量重复试验或观测下其结果却呈现出某种规律性，例如，多次重复抛一枚硬币得到正面朝上次数大致是半数；选煤厂在正常生产情况下选出的精煤灰分虽然会有波动，但它却是有规律地在某一区间内围绕着一个平均值来回波动。这种在大量重复试验和观测中所呈现出的固有规律性，就称为**统计规律性**。

一类现象，在个别试验中呈现出不确定性；在大量重复试验中，又具有统计规律性，称为**随机现象**，随机现象的每一种结果称为**随机事件**，它的取值称为**随机变量**。概率论和数理统计就是研究随机现象的统计规律性的数学学科。

从概率论的基本概念与方法的学习讨论中，我们知道在概率论的许多问题中，通常随机变量的概率分布是已知的，或假设为已知，从而一切计算与推理是在已知概率分布的基础上得出来的。因此概率论是对随机现象统计规律性演绎的研究。而数理统计则先从所要研究的对象整体中抽取一部分进行试验或观测，得到有关信息，即取得数据，再对整体的分布（或参数）作出推理。因此数理统计是对随机现象统计规律性归纳的研究。数理统计与概率论两者在研究方法上存在着明显的差异。然而两者都是研究随机现象统计规律性的一门数学学科，它们之间又相互渗透，紧密联系。

由于试验或观测是随机现象，那么有限次数的试验或观测对整体不可能作出正确的判断，总是有一定程度的不正确性。而这种不正确性用概率的大小来描述是最合适的。概率大，推断就比较可靠，概率小，推断的可靠性就差。**数理统计**就是以概率论为理论基础，根

据试验或观测所取得的有限信息，对整体作出推断，且这样的每一个推断必须伴有一定的概率以表明推断的可靠程度。这种伴有一定概率的推断称为**统计推断**。综上所述，**数理统计的基本任务**是：以概率论的理论为基础，以有效的方式收集、整理和分析受到随机因素影响的数据，以对所研究的问题作出描述、估计、预测直至为准备作出的决策和所要采取的行动提供依据或建议。

数理统计作为数据的艺术和科学，在现代科学研究和生产管理中已被广泛应用。随着选煤工业的迅速发展，数理统计方法在选煤工业中的应用也得到了蓬勃的发展。例如，选煤生产中产品质量的检测和控制、工艺流程改进前后分选效果的鉴定、精煤产品质量稳定性的统计分析等等。

二、数理统计的基本内容

数理统计研究的内容随着科学技术和生产实际的不断发展而逐步扩大，但概括起来可分为两大类：第一，试验设计和研究。研究如何对随机现象进行观测、试验，以取得有代表性的局部观测值，即研究以有效的方式收集数据及描述这类数据，这一部分内容称为**描述统计学**。它满足于对整体的调查和描述。第二，统计推断。研究如何对这些已得的观测值进行整理、分析，并作出决策的方法，以推断整体的规律性，这一部分内容称为**推断统计学**。下面用选煤生产实例具体说明。

例 1－1　某选煤厂月生产精煤 120 批，每批 900t。为了掌握精煤灰分情况，质量检查员每批抽取一个试样。于是提出下列问题：

(1) 如何从每批 900t 精煤产品中抽取部分有代表性的试样？取得试样后，如何进行制样和化验分析，以便获得样本数据？

(2) 如何用 120 个样本数据来估计全月精煤灰分均值和这 120 批精煤灰分偏离均值的离散程度？

(3) 若本厂规定精煤灰分为 10.00%，如何利用 120 个样本数据来判断本月精煤灰分均值与规定精煤灰分有无差异？

(4) 抽样获得的 120 个灰分值有大有小，如果本月生产采取不同的操作制度，那么灰分呈现的差异是由于操作制度不同造成的，还是仅仅由随机因素产生的呢？

(5) 如果精煤灰分与原煤灰分有关，那么如何利用抽取的 120 个样本数据与对应的 120 组原煤灰分值，来描述表示本月精煤灰分与原煤灰分之间的关系？

问题 (1) 实际就是如何制定具有一定代表性而又比较省工的采样、制样、化验分析方案以获得样本数据。在数理统计中解决这类问题的方法称为**抽样检验**。

问题 (2) 实质是要从 120 个灰分数据出发估计本月精煤灰分分布的某些数字特征，这里要估计数学期望与方差。数理统计中解决这类问题的方法称为**参数估计**。

问题 (3) 是要求根据抽样获得的数据去检查精煤灰分分布的某项数字特征与规定标准的差异，这里检验数学期望。数理统计中解决这类问题的方法是先作一个假设（例如假设与规定标准无差异），然后利用概率反证法检验这一假设是否成立，这种方法称为**统计假设检验**，简称**统计检验**。

问题 (4) 是要分析造成数据误差的原因，当有多个因素起作用时，还要分析哪些因素起主要作用，这种分析法称为**方差分析**。

问题 (5) 是要根据观察数据研究变量间的关系，这里是研究精煤灰分与原煤灰分两个

变量的关系。有时还要研究多个变量间的关系。这种研究方法称为**相关与回归分析**。

以上列举的抽样检验、参数估计、统计检验、方差分析、相关与回归分析都是数理统计中研究的基本内容，这些内容和质量管理图、正交试验设计将在以后各章分别加以研究讨论。此外还有时间序列分析、可靠性理论、统计决策理论等也是数理统计研究的重要内容。由于受篇幅、学时等限制，本教材就不一一讨论研究了。

三、统计术语

1. 母体与样本

由于数理统计方法主要是通过样本来了解母体，因此，正确理解母体与样本的基本概念，是学习数理统计的基础。

1）母体

在数理统计中，常把研究对象的全体称为**母体**或**总体**，而把组成母体的每个单元称为个体。例如，我们要研究一个工厂的产品质量时，该厂所产每件产品的质量就是一个个体，全部产品的质量就是母体；当研究的是本月精煤按批的平均质量时，则本月所产各批精煤平均质量的全体就是母体（例 1－1 中的 120 批精煤的质量就是一个母体），而其中每一批精煤的平均质量则为个体（例 1－1 中的每批 900t 精煤的平均质量就是一个个体）。

实际上，我们所关心的不是母体中个体的一切方面，而是它的某一数量特征（即数量指标，例如精煤灰分）及其在母体中的分布情况（例如灰分在 9.50%到 10.50%之间的精煤在 120 批中所占比例），就某一数量特征（例如精煤灰分）X 而言，每一个个体所取的值不一定相同，以例 1－1 为例，有的一批是 10.00%，有的一批是 10.50%，又因为对一个母体而言，个体的取值是按一定规律分布的，例如这 120 批精煤中各种灰分的精煤所占的比例是确定的，是客观存在的，所以，任取一批精煤，其灰分 X 究竟取什么值是有一定概率分布的。正因为如此，对母体来说，这一数量特征 X，就是一个随机变量，母体就是这个随机变量 X 取值的全体，其中每一个个体都是一个实数。如果要研究的指标不止一个（例如对于精煤产品质量，我们要研究灰分、水分、硫分等等），那么可分为几个母体来研究。如果表征母体的随机变量 X 的分布函数为 $F(x)$，那么，我们就称母体 X（或母体）的分布为 $F(x)$。

由于任何一个母体，都可以用一个相应的随机变量来描述，即母体与随机变量之间有着自然对应联系，我们就可以在研究数理统计问题中方便地引用概率论中的结论，而对统计问题进行深入的研究讨论。

2）样本

在生产和科学研究中，尽管我们研究的对象是母体，但是母体通常很大，即使不大，有时也不可能对其中的每个个体都进行考察，因为有些检验具有破坏性，例如产品的寿命、精煤的灰分等等。所以就必须对母体进行抽样观测，根据抽样观测所得的结果来推断母体的性质。这种从母体中抽取若干个体来观测母体某种数量指标的取值过程，称为**抽样**（又称**取样**、**采样**）。从母体中随机抽取的一部分个体称为**样本**；样本中所包含的个体数目 n 称为**样本容量**或**样本大小**。

我们抽取样本的目的是为了对母体的性质进行各种分析推测，要确保分析推测的可靠性，样本必须有代表性，必须是从母体中随机抽取的。如果母体中样本元素的每种可能出现的组合能有同样的机会被抽到，我们称抽样是随机的；抽取的样本称为**随机样本**。随机

样本必须具备独立性（即每个观察结果既不影响其它观察结果，也不受其它观察结果的影响）和代表性〔即要求样本每个分量 X_i（$i=1, 2, \cdots, n$）必须与总体 X 具有相同的分布 $F(x)$〕。

综上所述，从数学角度而言，所谓母体就是指一个随机变量 X，所谓样本就是 n 个相互独立且与母体 X 有相同概率分布的随机变量 X_i（$i=1, 2\cdots, n$）所组成的 n 维随机变量。我们每一次具体的抽样，所得的数据是这 n 维随机变量的值（样本观测值），用 $x_1, x_2, \cdots, x_n$ 来表示。容易看出，样本具有两重性，它本身是随机变量，但一经抽取便是一组确定的具体值。

2. 统计量

样本是母体的代表及反映，但在抽取样本之后，并不直接利用样本的几个观测值进行推断，而需要对这些值进行加工，提炼。把样本中所包含的有关我们关心的事物的信息集中起来，这便是针对不同的问题构造样本的某种函数，这种样本函数在数理统计中称为**统计量**。

当我们获得母体 X 的一个样本（$X_1, X_2, \cdots, X_n$）时，为了推断母体的性质，往往从某些数字特征去推断母体的相应数字特征。样本的数字特征常用的有两种：一种是表现观测值的位置特征，如样本均值（$\overline{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$），另一种是表示观测值的离散特征，如样本方差（$s^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2$）。

容易看出，样本数字特征是样本（$X_1, X_2, \cdots, X_n$）的函数。为了充分利用样本来推断母体，在数理统计中除了用样本数字特征外，还要用到样本的其它函数。为此，我们引入下列定义。

定义 设 $X_1, X_2, \cdots, X_n$ 为母体 X 的一个样本，$g(X_1, X_2, \cdots, X_n)$ 为一个连续函数。如果 g 中不包含任何未知参数，则称 $g(X_1, X_2, \cdots, X_n)$ **为一个统计量**。

如果 $x_1, x_2, \cdots, x_n$ 是样本 $X_1, X_2, \cdots, X_n$ 的观测值，则 $g(x_1, x_2, \cdots, x_n)$ **称为统计量的值**，简称**统计值**。

例如，$X_1, X_2, \cdots, X_n$ 是从正态母体 $N(\mu, \sigma^2)$ 中抽取的一个样本，其中 μ 为已知，σ^2 为未知，则 $\sum_{i=1}^{n}(X_i-\mu)^2$ 是统计量，但$\left(\sum_{i=1}^{n}X_i\right)/\sigma$ 不是统计量。

从统计量定义还可看到，由于样本（$X_1, X_2, \cdots, X_n$）是随机变量，所以作为样本函数的统计量 $g(X_1, X_2, \cdots, X_n)$ 也是随机变量，它应有确定的概率分布，并且统计量也具有两重性。如果母体的分布函数为已知，则统计量的分布总是可以求出的。

第二章　概　率　基　础

第一节　事件的基本性质

一、样本空间与事件

1. 样本空间

对于随机事件，尽管在每次试验之前不能预知试验结果，但是试验的所有结果所组成的集合是已知的。将随机试验 E 的所有可能组成的集合称为 **E 的样本空间**。样本空间的元素，即 E 的每个结果，叫做**样本点**。

下面是几个典型样本空间的表示方法：

S_1：$\{H, T\}$

S_2：$\{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$

S_3：$\{1, 2, 3, 4, 5, 6, 7\}$

S_4：$\{0, 1, 2, 3, \cdots\}$

S_5：$\{t \mid t \geqslant 0\}$

S_6：$\{(x, y) \mid T_0 \leqslant x \leqslant y \leqslant T_1\}$

要注意的是：样本空间的元素是由试验的目的决定的，试验目的不一样，样本空间也不一样。

2. 事件

事件是概率研究的对象。事件可分为必然事件，不可能事件、随机事件三类。

二、事件间的关系及运算

事件是一个集合，所以事件间的关系与事件的运算应按集合论中的集合间的关系和运算来处理。

设试验 E 的样本空间为 S，而 A，B，A_k（$k=1, 2, \cdots$）是 S 的子集。

1. 事件的包含与相等

若事件 A 发生必有事件 B 发生，则称**事件 B 包含事件 A**，记作 $A \subset B$。

若 $A \subset B$ 且 $B \subset A$，则称**事件 A 与事件 B 相等**，记作 $A=B$。

2. 和事件

事件 A 与事件 B 至少发生一个所构成的事件，称为**事件 A 与事件 B 的和事件**，记作 $A \cup B$。

类似地，称 $\bigcup_{k=1}^{n} A_k$ 为 n 个事件 A_1，A_2，$\cdots$，A_n 的和事件；称 $\bigcup_{k=1}^{n} A_k$ 为可列事件 A_1，$A_2\cdots$，A_n 的和事件。

3. 积事件

事件 A 与事件 B 都发生所构成的事件，称为**事件 A 与事件 B 的积事件**，记作 $A \cap B$ 或 AB。

类似地，称$\bigcap\limits_{k=1}^{n}A_k$为$n$个事件$A_1$，$A_2$，…，$A_n$的积事件；称$\bigcap\limits_{k=1}^{n}A_k$为可列事件$A_1$，$A_2$，…，$A_n$的积事件。

4. 差事件

事件A发生而事件B不发生所构成的事件，称为**事件A与事件B的差事件**，记作$A-B$。

5. 互不相容

事件A与事件B不可能同时发生，即$A\cap B=\phi$，则称**事件A与事件B互不相容**。

6. 对立事件

若$A\cup B=S$且$A\cap B=\phi$，则称**事件A与事件B互为逆事件**，也可称**事件A与事件B互为对立事件**。"A不发生"的事件为A的对立事件，记作$\overline{A}$。

7. 相互独立事件

若事件A的发生对事件B能否发生没有影响，或事件B发生的概率与事件A无关，则称**事件B对于事件A是独立的**。事件间的独立性是相互的。

8. 完备事件组

若有n个事件A_i ($i=1$，2，…，n)，在每次试验中必然发生，且仅能发生其中之一，即这n个事件满足两两互不相容且它们的和为必然事件，则称这**n个事件是一个"完备事件组"**。例如，互逆事件就是完备事件组。在概率的古典定义中，基本事件的全体也是一个完备事件组。

第二节 随机事件的概率

随机事件在一次试验中可能会发生，也可能不会发生。在实际试验中是不可能对某一随机事件能否发生事先作出准确的判断，但可以对这一事件在限定的试验条件下发生的可能性的大小作出估计。一个随机事件发生可能性大小，就是在一定条件下该事件发生的概率。

我们规定：必然事件的概率为1，不可能事件的概率为0。

随机事件发生的概率在0～1之间，并且任何事件的概率既不可能大于1，也不可能为负值，这就是概率的基本属性。

一、概率的定义

概率的定义有许多种，这里只是简单介绍概率的统计定义和古典概率定义。

1. 概率的统计定义

随机事件在一次试验中可能会发生，也可能不会发生，试验结果是很难预测的。只有对其进行大量的重复试验或观察，才有可能发现其中的规律。

例2－1 "抛硬币"试验。将一枚硬币抛掷5次、50次、500次，各做10遍，得到的数据如表2－1所示（其中n_H表示发生的频数，f_n（H）表示H发生的频率）。

同样的试验在历史上曾有人作过，得到如下结果：

实验者	抛的次数	"抛出正面"次数	频 率
Buffon	4040	2048	0.5069
K. Pearson	12000	6019	0.5016
K. Pearson	24000	12012	0.5005

表 2—1

实验序号	n=5		n=50		n=500	
	n_H	$f_r(H)$	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
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

从这里可以得出：①对于同样的 n，所得的 $f_n(H)$ 并不相同，频率具有随机波动性；②当 n 较小时，$f_n(H)$ 随机波动的幅度比较大，但随着 n 的增大，$f_n(H)$ 趋于稳定，$f_n(H)$ 总是在 0.5 左右摆动，并逐渐稳定于 0.5。即随着试验次数的增加，频率与概率接近的可能性也越大。频率和概率间的这种关系，在概率论中称为“随机事件出现的频率依概率收敛与它的概率”。概率的统计定义就是以这种客观规律为基础建立起来的。

概率的统计定义：进行大量重复试验时，事件 A 发生的频率具有稳定性。频率的稳定值 p 称为**事件 A 的概率**，记作 $p=p(A)$。

2. 古典概率定义

随机试验有两个特点：

(1) 随机试验只可能出现 n 种（n 有限）基本结果，而全体基本结果构成两两不相容事件完备组。

(2) 在一次随机试验中各个基本结果发生可能性相等。

设试验的样本空间为 $S=\{e_1, e_2, \cdots, e_n\}$. 由于在试验中每个基本事件发生的可能性相等，即：

$$P(e_1)=P(e_2)=\cdots=P(e_n)$$

又由于基本事件是两两不相容的，所以：

$$1=P(S)=P(\{e_1\}\cup\{e_2\}\cup\cdots\cup\{e_n\})=P(\{e_1\})+P(\{e_2\})+\cdots+P(\{e_n\})=nP(\{e_i\})$$

$$P(\{e_i\})=\frac{1}{n}\quad i=1, 2, \cdots, n$$

如果事件 A 包含 k 个基本事件，

$A=\{e_{i_1}\}\cup\{e_{i_2}\}\cup\cdots\cup\{e_{i_k}\}$ $(1\leqslant i_1<i_2<\cdots<i_k\leqslant n)$ 则有：

$$P(A)=\frac{k}{n} \tag{2—1}$$

$P(A)$ 就是**事件 A 的古典概率**，也称为**等可能概率**。

二、概率的性质

由概率的定义，可以推得概率的一些重要性质。

(1) $P(\phi)=0$

(2) 若 A_1，A_2，…，A_n 是两两互不相容的事件，则有

$$P=(A_1\cup A_2\cup\cdots\cup A_n)=P(A_1)+P(A_2)+\cdots+P(A_n) \tag{2-2}$$

(2—2) 式称为**概率的有限可加性**。

(3) 设 A，B 两个事件，若 $A\subset B$，则有

$$P(B-A)=P(B)-P(A) \tag{2-3}$$

$$P(B)\geqslant P(A) \tag{2-4}$$

(4) 对于任一事件 A，有

$$P(A)\leqslant 1 \tag{2-5}$$

(5) 对于任一事件 A，有

$$P(\overline{A})=1-P(A) \tag{2-6}$$

(6) 对于任一两事件 A、B，有

$$P(A\cup B)=P(A)+P(B)-P(AB) \tag{2-7}$$

第三节　条件概率与独立事件

一、条件概率的定义

条件概率是概率论中的一个重要而实用的概念，所考虑的是事件 A 已发生的条件下事件 B 发生的概率。

定义　设 A，B 是两个事件，若 $P(A)>0$，称：

$$P(B|A)=\frac{P(AB)}{P(A)} \tag{2-8}$$

为**事件 A 发生的条件下事件 B 发生的条件概率**，记作 $P(B|A)$。

例 2—2　一个家庭中有两个小孩，已知其中有一个是女孩，问另一个小孩也是女孩的概率为多大？(假定一个小孩是男还是女是等可能的)

解　据题意事件的样本空间为：

$S=\{$(男，男)，(男，女)，(女，男)，(女，女)$\}$

$A=\{$已知有一个是女孩$\}=\{$(男，女)，(女，男)，(女，女)$\}$

$B=\{$另一个也是女孩$\}=\{$(女，女)$\}$

因此，所求的概率为：

$$P(B|A)=\frac{P(AB)}{P(A)}=\frac{1/4}{3/4}=1/3$$

例 2—3　有外形相同的球分装三个盒子，每盒 10 个，其中，第一个盒子中 7 个球标有字母 A，3 个球标有字母 B；第二个盒子中有红球和白球各 5 个；第三个盒子中则有红球 8 个，白球 2 个，试验按如下规则进行：先在第一个盒子中任取一个球，若取得标有字母 A 的球，则在第二个盒子中任取一个球；若第一次取得标有字母 B 的球，则在第三个盒子中任取一个球、如果第二次取出的是红球，则称试验为成功。求试验成功的概率。

解　设　A=从第一个盒子中取出标有字母 A 的球；

$P(A)=0.7$

B=从第一个盒子中取出标有字母 B 的球：

$P(B)=0.3$

C=第二个盒子中取出的是红球：

D=第二个盒子中取出的是白球：

$$P(C/A)=0.5 \quad P(D|A)=0.5$$

$$P(C|B)=0.8 \quad P(D|B)=0.2$$

$$\begin{aligned} P(R) &= P(C|A)\times P(A)+P(C|B)\times P(B) \\ &= 0.5\times 0.7+0.8\times 0.3 \\ &= 0.59 \end{aligned}$$

二、概率乘法定理

由条件概率的定义既可推得概率的乘法定理。

概率乘法定理 若 $P(A)>0$ 则有：

$$P(AB)=P(B|A)P(A) \tag{2-9}$$

一般，设 A_1，A_2，…，A_n 为 n 个事件，$n\geqslant 2$，且 $P(A_1, A_2, \cdots, A_{n-1})>0$，则有：

$$P(A_1A_2\cdots A_n)=P(A_n|A_1A_2\cdots A_{n-1})P(A_{n-1}|A_1A_2\cdots A_{n-2})\cdots P(A_2|A_1)P(A_1) \tag{2-10}$$

例 2－4　设一个袋中有两个篮球与三个排球，从袋中取出两个球，问两个都是篮球的概率是多少？

解　设：

A=第一个球是篮球

B=第二个球是篮球

由于

$$P(A)=2/5 \quad P(B|A)=1/4$$

所以取出两个都是篮球的概率是 $P(AB)=P(B|A)P(A)=0.4\times 0.25=0.1$

概率乘法定理应用到选煤具体实践中具有重要的意义。例如，在 n 批精煤中灰分合格的有 m 批，水分合格的有 k 批，灰分和水分都合格的有 r 批。很明显：$r\leqslant m$，$r\leqslant k$。设事件 A=抽测结果灰分合格；B=抽测结果水分合格。那么，抽测结果灰分和水分都合格的概率是多少？按照集合论的原理，抽测结果灰分和水分都合格是 A 事件与 B 事件的交集，即 AB。

根据式（2－9）可得：

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

其中，$P(A)=\dfrac{m}{n}$　$P(B|A)=\dfrac{r}{m}$

所以，$P(AB)=\dfrac{m}{n}\cdot\dfrac{r}{m}$

三、全概率公式和贝叶斯（Bayes）公式

1. 全概率公式

定理　设试验 E 的样本空间为 S，A 为 E 的事件，B_1，B_2，…，B_n 为 S 的一个划分，且 $P(B_i)>0$ $(i=1, 2, \cdots, n)$，则：

$$P(A)=P(A|B_1)P(B_1)+P(A|B_2)P(B_2)+\cdots+P(A|B_n)P(B_n) \tag{2-11}$$

式（2－11）称为**全概率公式**。

在实际工作中 $P(A)$ 是很难直接求得的，但较容易找到 S 的一个的划分 B_1，B_2，…，B_n，且 $P(B_i)$ 和 $P(A|B_i)$ 为已知或容易求得，那么就可以按式（2－11）求得 $P(A)$。

例 2－5　已知某选煤厂的跳汰精煤快灰合格的概率为 0.90，浮选精煤快灰合格的概率为 0.95。当这两种精煤快灰都合格时，销售精煤灰分合格；假定当跳汰精煤灰分合格而浮选精煤灰分不合格时，销售精煤灰分合格的可能性有 70%；当跳汰精煤灰分不合格而浮选精煤灰分合格时，销售精煤灰分合格的可能性只有 20%。问该选煤厂销售精煤灰分合格的概率是多少？

解　将销售精煤的质量这个随机现象划分成四种情况，即：

B_1：跳汰与浮选精煤快灰都合格；

B_2：跳汰精煤灰分合格，浮选精煤灰分不合格；

B_3：跳汰精煤灰分不合格，浮选精煤灰分合格；

B_4：跳汰和浮选精煤灰分都不合格。

显然，B_1，B_2，B_3，B_4 是一个完备事件组，各事件出现的概率分别为：

$$P(B_1)=0.90\times0.95=0.855$$
$$P(B_2)=0.90\times0.05=0.045$$
$$P(B_3)=0.10\times0.95=0.095$$
$$P(B_4)=0.10\times0.05=0.005$$

设销售精煤灰分合格的事件为 A，由题意可知在上述四种情况下销售精煤灰分合格的概率，即是事件 A 对于事件 B_1，B_2，B_3，B_4 的条件概率分别为：

$$P(A|B_1)=1.00,\quad P(A|B_2)=0.70,$$
$$P(A|B_3)=0.20,\quad P(A|B_4)=0.00,$$

根据全概率公式式（2－11）得：

$$P(A)=\sum_{i=1}^{4}P(A|B_i)P(B_i)=1.00\times0.855+0.70\times0.045+0.20\times0.095$$
$$+0.00\times0.005=0.9055=90.55\%$$

2. 贝叶斯公式

在选煤生产中经常发生这样情况，已知某日的销售精煤的灰分不合格，希望估计和分析产生这种结果的原因，以及时解决问题。这类问题实质上是计算事件 B_1，B_2，B_3，B_4 对销售精煤灰分不合格这一事件（$\overline{A}$）的条件概率 $P(B_i|\overline{A})$。这个问题就要应用贝叶斯公式来解决。

定理　设试验 E 的样本空间为 S，A 为 E 的事件，B_1，B_2，…，B_n 为 S 的一个划分，且 $P(A)>0$，$P(B_i)>0$ $(i=1, 2, \cdots, n)$，则：

$$P(B_i|A)=\frac{P(A|B_i)P(B_i)}{\sum_{j=1}^{n}P(A|B_j)P(B_j)},\quad i=1, 2, \cdots, n. \tag{2-12}$$

例 2－6　在例 2－5 中，若发现某日销售精煤灰分不合格，问题发生在 B_1，B_2，B_3，B_4

中哪种可能性最大？

解 设精煤灰分不合格为C事件，由题意可知在上述四种情况下销售精煤灰分合格的概率，即是事件C对于事件B_1，B_2，B_3，B_4的条件概率分别为：

$$P(C)=1-P(A)=1-0.9055=0.0945$$

按式（2—12）可分别求出：

$$P(B_1|C)=\frac{P(C|B_1)\cdot P(B_1)}{P(C)}=\frac{0.00\times0.855}{0.0945}=0.00$$

$$P(B_2|C)=\frac{P(C|B_2)\cdot P(B_2)}{P(C)}=\frac{0.30\times0.045}{0.0945}=0.1429$$

$$P(B_3|C)=\frac{P(C|B_3)\cdot P(B_3)}{P(C)}=\frac{0.80\times0.095}{0.0945}=0.8042$$

$$P(B_4|C)=\frac{P(C|B_4)\cdot P(B_4)}{P(C)}=\frac{1.00\times0.05}{0.0945}=0.0529$$

由此可见，造成销售精煤灰分不合格的各种原因中，B_3，即“跳汰精煤灰分不合格，浮选精煤灰分合格”的可能性最大。

四、独立事件的概念与定义

定义 设A，B是两事件，如果具有等式：

$$P(AB)=P(A)P(B) \tag{2—13}$$

则称**事件A与B相互独立或独立**。

定理 若$P(A)>0$，$P(B)>0$，则事件A与B相互独立的必要条件是：

$$P(A|B)=P(A)\text{ 或 }P(B|A)=P(B) \tag{2—14}$$

此定理说明：事件A，B独立表示A发生的可能性不依赖B的发生；反之亦然。

定义 若事件A_1，A_2，…，A_n满足：

$P(A_{i_1},A_{i_2},\cdots A_{i_k})=P(A_{i_1})P(A_{i_2})\cdots P(A_{i_k})$，$2\leqslant k\leqslant n$，其中$i_1$，$i_2$，…，$i_k$是1，2，…，$n$中任意$k$个数，则称**事件$A_1$，$A_2$，…，$A_n$相互独立或独立**。

事件A_1，A_2，…，A_n相互独立或独立表示它们的发生互不影响。

第四节 随机变量的分布及数字特征

一、随机变量的分布

1. 随机变量的定义

随机变量是随机现象结果的数量指标。因为在给定的条件下，随机现象在每次的试验中可能会有不同的试验结果，所以，它是一个取值会随机变化的变量。

定义 设E是随机试验，它的样本空间是$S=\{e\}$。如果对于每一个$e\in S$，有一个实数$X(e)$与之对应，这样就得到一个定义在S上的单值实值函数$X=X(e)$，称为**随机变量**。

由于随机变量X的取值随试验的结果而定，而试验的各个结果的出现有一定的概率，因而X的取值也有一定的概率。

随机变量与普通函数的区别在于：随机变量取值的不确定性和取值的概率性；随机变量定义在样本空间上而普通函数定义在实数轴上。

有很多试验的结果其本身就是用数量来描述的，样本点e就是数。即$X(e)=e$。如某