概率论与数理统计

崔海英　等编

科学出版社
www.sciencep.com

概率论与数理统计

崔海英 等　编

科学出版社
北京

内 容 简 介

本书是按照我国现行的工科大学本科“概率论与数理统计”课程教学基本要求编写的.全书主要内容包括:随机事件与概率,随机变量及其分布,随机变量的数字特征,数理统计的基本概念,统计推断,方差分析与回归分析,贝叶斯公式与条件分布,大数定律与中心极限定理及其应用,随机过程,SPSS的简单应用.书中有些内容加了“※”号,选用本书时可以根据教学的需要和学时安排略去不讲.

本书适用于高等学校工科类、经济管理类及医学类的本科各专业.

图书在版编目(CIP)数据

概率论与数理统计/崔海英等编.—北京:科学出版社,2010
ISBN 978-7-03-028138-8

Ⅰ.①概… Ⅱ.①崔… Ⅲ.①概率论-高等学校-教材②数理统计-高等学校-教材 Ⅳ.①O21

中国版本图书馆CIP数据核字(2010)第120987号

责任编辑:张中兴/责任校对:刘小梅
责任印制:张克忠/封面设计:耕者设计工作室

科学出版社出版
北京东黄城根北街16号
邮政编码:100717
http://www.sciencep.com

北京市文林印务有限公司印刷

科学出版社发行 各地新华书店经销

*

2010年6月第 一 版 开本:B5(720×1000)
2010年6月第一次印刷 印张:16 1/4
印数:1—3 000 字数:360 000

定价:29.00元

(如有印装质量问题,我社负责调换)

序　言

我国的高等教育近年来有了很大变化与发展，在大学生人数不断增加的同时，新办专业也在不断增加，大学教育已经从精英教育走向了大众化、普及化教育. 面对新的教育形势、不同专业的培养目标及不同学校的生源特点，编写符合各种需要及要求的教材是形势发展的必然要求.

我校教学一线的教师们，经过多年的教学实践及长期思索，为非重点院校工科类、经管类的本科生及专升本的学生，编写了这本《概率论与数理统计》教材.

本教材的编写充分考虑到非重点院校的一般工科类、经管类的本科生及专升本学生的特点，调整了教学内容，讲述了易于学生掌握的概率统计中最基本的概念、理论和最常用的基本方法，达到让学生了解基本思想、掌握基本运算和基本方法的目的，同时突出了概率论与数理统计的应用性，举例贴近生活，源于实际.

在不降低对概率统计课程的基本要求的基础上，本教材兼顾传统理论与时代发展. 例如：

(1) 舍去了一般教材中关于古典概率模型比较详细的讨论，重点介绍了在统计部分涉及的相关概率论的概念；

(2) 通过实际的例子引入概念，对某些定理的证明进行了简化处理；

(3) 根据目前高中教材内容的变化及生源特点，增加了准备知识和衔接内容；

(4) 融入案例分析与统计软件(SPSS 软件)介绍，旨在提高学生解决实际问题的能力.

相信本教材的出版将有益于高等教育事业的发展，并希望作者在今后的教学中不断完善教材内容. 同时欢迎更多适合不同水平学生、不同专业需求的数学教材问世.

2010 年 4 月 25 日

前　言

目前在普通高校的工科类、经济管理类专业都普遍开设有“概率统计”、“数理统计”等必修课，有的甚至还开设有“实用统计方法”等选修课，以适应社会发展对统计分析任务日益增多的需求. 现在，“概率论与数理统计”是高等学校工科类、经济管理类及医学类的本科各专业的一门重要的基础理论课.

概率统计是一门实用性较强的学科，但实际情况是课时受限，因此我们编写这本教材的主要思路如下：首先根据生源特点及目前高中教材内容的变化，注重衔接，增加了准备知识的章节. 其次对内容有所取舍，突出重点. 本书的特点是概率论和数理统计的内容各占一半，相对于传统的概率论与数理统计教材，概率部分略有减弱，统计部分有所加强，因此在概率论部分关于古典概型只是一般讨论，关于概率论的大部分知识是点到为止，但是却加强了对于非常重要的正态分布的学习. 编写中注重概念和理论的直观解释，尽量避免纯数学化的论证，同时保持内容的完整性和严谨性；注重概率统计方法在各个领域的应用，侧重对概率统计方法的介绍，培养学生对基本概念的准确理解及对常用方法的熟练掌握；介绍统计理论时，更偏重于方法和统计背景的介绍，同时以 SPSS 软件为平台，介绍各种统计方法的应用实现，旨在强调理论与应用的有机结合. 希望通过本书的学习，使学生掌握概率论与数理统计的基本概念、基本理论及相应的处理随机事件的基本思想和方法，培养学生运用概率统计方法分析和解决实际问题的能力.

本书中的例题有相当一部分来自于实际问题，以实际问题为背景，举例贴近生活，消除学生对数学的陌生感、抽象感，增强学好数学、用好数学的信心，并且对一些例子的计算步骤和结论有较详细的讨论. 希望通过实例的讨论，使读者对于应用中遇到的问题会进行分析，并能够得出相应的结果.

本书是按照我国现行的工科大学本科“概率论与数理统计”课程教学基本要求编写的. 全书主要内容包括：随机事件与概率，随机变量及其分布，随机变量的数字特征，数理统计的基本概念，统计推断，方差分析与回归分析，贝叶斯公式与条件分布，大数定律与中心极限定理及其应用，随机过程，SPSS 的简单应用. 书中有些内容加了“※”号，选用本书时可以根据教学的需要和学时安排略去不讲.

本书由崔海英任主编，杨静任副主编．参加本书编写的老师有：陈冬、尚学海、崔海英、冯远福、杨静、崔菊莲．全书由陈冬、崔海英、杨静统稿、校订．

北京工业大学高旅端教授详细审阅了本书初稿，并提出了宝贵意见与建议．北京联合大学任开隆教授也对本书提出了建设性意见．在本书编写和出版过程中得到了科学出版社的大力支持，同时北京联合大学教务处、基础部也给予了很多指导与关怀，在此一并表示衷心的感谢．

由于编写时间仓促，编者水平有限，书中难免存在不妥之处，敬请读者批评指正．

编 者

2010 年 3 月

目　　录

第0章 准备知识

本章叙述一些准备知识，其中包括概率统计一些常用公式和简化计算方法，介绍了均值、方差、中位数等基本概念，为今后各章的学习提供一个更直观的背景. 同时还介绍了概率统计的应用实例，以了解如何运用概率统计的思想方法.

0.1 数学符号“$\sum$”和“$\prod$”

0.1.1 求和号 $\sum$

$\sum$ 是一个符号，表示若干个数连加. n 个数 $a_1,a_2,\cdots,a_n$ 相加，可用 $\sum\limits_{i=1}^{n}a_i$ 表示，即

$$\sum_{i=1}^{n}a_i=a_1+a_2+\cdots+a_n,$$

式中 i 称为流动下标(均为整数)，a_i 称为通项. 很明显

$$\sum_{i=1}^{n}a_i=a_1+a_2+\cdots+a_n=\sum_{d=1}^{n}a_d.$$

可见流动下标用什么符号表示对求和没有影响，这一点我们以后要常常用到. 其次，只有通项 a_i 有了明确的定义时，求和才有意义. 例如 $a_i=x^i$，则有 $\sum\limits_{i=1}^{k}a_i=x+x^2+\cdots+x^k$.

下面给出求和号今后常用的几个运算性质.

性质 0.1.1 $\sum\limits_{i=1}^{n}c=nc$(特别有 $\sum\limits_{i=1}^{n}1=n$).

证明 按定义，通项 $a_i=c$，因此

$$\sum_{i=1}^{n}c=\underbrace{c+c+\cdots+c}_{n\text{个}}=nc.$$

性质 0.1.2 $\sum\limits_{i=1}^{n}(a_i+b_i)=\sum\limits_{i=1}^{n}a_i+\sum\limits_{i=1}^{n}b_i$.

证明

$$\begin{aligned}\sum_{i=1}^{n}(a_i+b_i)&=(a_1+b_1)+(a_2+b_2)+\cdots+(a_n+b_n)\\&=(a_1+a_2+\cdots+a_n)+(b_1+b_2+\cdots+b_n)\\&=\sum_{i=1}^{n}a_i+\sum_{i=1}^{n}b_i.\end{aligned}$$

性质 0.1.3 $\sum\limits_{i=1}^{n}ca_i=c\sum\limits_{i=1}^{n}a_i$.

证明

$$\begin{aligned}\sum_{i=1}^{n}ca_i&=ca_i+ca_2+\cdots+ca_n\\&=c(a_1+a_2+\cdots+a_n)\\&=c\sum_{i=1}^{n}a_i.\end{aligned}$$

例 0.1.1(等差数列求和) 设 $a_i=a+id(i=0,1,2,\cdots,n)$,求 $\sum\limits_{i=0}^{n}a_i$ 的值.

解

$$\begin{aligned}\sum_{i=0}^{n}i&=0+1+2+\cdots+n\\&=\frac{1}{2}(1+2+\cdots+n+1+2\cdots+n)\\&=\frac{1}{2}[(1+n)+(2+n-1)+\cdots+(n+1)]\\&=\frac{1}{2}n(n+1),\end{aligned}$$

即

$$\sum_{i=0}^{n}i=\frac{1}{2}n(n+1).$$

所以

$$\begin{aligned}\sum_{i=0}^{n}(a+id)&=\sum_{i=0}^{n}a+\sum_{i=0}^{n}id=\sum_{i=0}^{n}a+d\sum_{i=0}^{n}i=(n+1)a+d\sum_{i=0}^{n}i\\&=(n+1)a+d\cdot\frac{1}{2}n(n+1)\\&=(n+1)\cdot\frac{1}{2}[a+(a+nd)]\\&=\frac{1}{2}\cdot\text{项数}\cdot(\text{首项}+\text{末项}).\end{aligned}$$

例 0.1.2(等比数列求和) 设 $a_i=x^i, x\neq 1, i=0,1,2,\cdots,n$,求 $\sum\limits_{i=0}^{n} a_i$ 的值.

解 记 $s_n \triangleq \sum\limits_{i=0}^{n} a_i, n=0,1,2,\cdots$,于是

$s_0 = a_0 = x^0 = 1$,

$s_1 = a_0 + a_1 = x^0 + x = 1 + x = \dfrac{1-x^2}{1-x}, \quad (x \neq 1)$,

$s_2 = a_0 + a_1 + a_2 = x^0 + x + x^2 = 1 + x + x^2 = \dfrac{1-x^3}{1-x}, \quad (x \neq 1)$,

容易看出

$$\begin{aligned}&(1-x)(1+x+x^2+\cdots+x^n)\\=&1+x+x^2+\cdots+x^n-x-x^2-\cdots-x^n-x^{n+1}\\=&1-x^{n+1}.\end{aligned}$$

因此有

$$1+x+x^2+\cdots+x^n = \frac{1-x^{n+1}}{1-x}, \quad (x\neq 1),$$

即

$$\sum_{i=0}^{n} a_i = \frac{1-x^{n+1}}{1-x}, \quad (x\neq 1),$$

由此可知

$$\sum_{i=0}^{n} ax^i = \frac{a(1-x^{n+1})}{1-x}, \quad (x\neq 1).$$

例 0.1.3(双重连加号)

$$\begin{aligned}\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij} &= \sum_{i=1}^{m}(a_{i1}+a_{i2}+\cdots a_{in})\\&=(a_{11}+a_{12}+\cdots+a_{1n})+(a_{21}+a_{22}+\cdots+a_{2n})+\\&\quad\cdots+(a_{m1}+a_{m2}+\cdots+a_{mn})\end{aligned}$$

0.1.2 连乘号 $\prod$

$\prod$ 是一个符号,表示若干个数连乘. n 个数 $a_1, a_2, \cdots, a_n$ 相乘,可用 $\prod\limits_{i=1}^{n} a_i$ 表示,即

$$\sum_{i=1}^{n} a_i = a_1 \cdot a_2 \cdot \cdots \cdot a_n,$$

式中 i 称为流动下标(均为整数),a_i 称为通项. 很明显

$$\prod_{i=1}^{n} a_i = a_1 \cdot a_2 \cdot \cdots \cdot a_n = \prod_{k=1}^{n} a_k.$$

下面给出求连乘号今后常用的几个运算性质.

性质 0.1.4 $\prod\limits_{k=1}^{n} k = 1 \cdot 2 \cdot 3 \cdots \cdot n = n!$.（读作 n 的阶乘）

性质 0.1.5 $\prod\limits_{i=1}^{n} c = c^n$.

证明 按定义，通项 $a_i = c$，因此

$$\prod_{i=1}^{n} c = \underbrace{c \cdot c \cdot \cdots \cdot c}_{n\text{个}} = c^n.$$

性质 0.1.6 $\prod\limits_{i=1}^{n} ca_i = c^n \prod\limits_{i=1}^{n} a_i$

证明

$$\begin{aligned}\prod_{i=1}^{n} ca_i &= (ca_1) \cdot (ca_2) \cdot \cdots \cdot (ca_n) \\ &= c^n \prod_{i=1}^{n} a_i\end{aligned}$$

0.2 概率统计中若干常用基本概念

0.2.1 均值

定义 0.2.1 给定一组数据 $x_1, x_2, \cdots, x_n$. 我们称

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i$$

为该组数据的均值.

均值就是通常所说的算数平均数，或简称为平均数. 像“平均身高”、“平均成绩”、“平均产值”等在日常生活中经常使用，是人们所熟悉的概念. 由于均值反映一组数的总的情况，所以它具有代表性，且 $\bar{x}$ 具有如下性质：

性质 0.2.1 $\sum\limits_{i=1}^{n} (x_i - \bar{x}) = 0$.

证明

$$\begin{aligned}\sum_{i=1}^{n} (x_i - \bar{x}) &= \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \bar{x} = \sum_{i=1}^{n} x_i - n \cdot \bar{x} \\ &= \sum_{i=1}^{n} x_i - n \cdot \frac{1}{n}\sum_{i=1}^{n} x_i = 0.\end{aligned}$$

性质 0.2.2 $\sum_{i=1}^{n}(x_i-c)^2 \geqslant \sum_{i=1}^{n}(x_i-\bar{x})^2$，等号当且仅当 $c=\bar{x}$ 时成立.

证明

$$\begin{aligned}\sum_{i=1}^{n}(x_i-c)^2 &= \sum_{i=1}^{n}[(x_i-\bar{x})+(\bar{x}-c)]^2 \\ &= \sum_{i=1}^{n}(x_i-\bar{x})^2+2\sum_{i=1}^{n}(x_i-\bar{x})(\bar{x}-c)+n(\bar{x}-c)^2 \\ &= \sum_{i=1}^{n}(x_i-\bar{x})^2+n(\bar{x}-c)^2,\end{aligned}$$

因为$(\bar{x}-c)^2\geqslant 0$，$n>0$，所以 $n(\bar{x}-c)^2\geqslant 0$，且 $\sum_{i=1}^{n}(x_i-\bar{x})^2$ 与 c 无关，所以$\sum_{i=1}^{n}(x_i-c)^2 \geqslant \sum_{i=1}^{n}(x_i-\bar{x})^2$ 恒成立，且等号仅在 $c=\bar{x}$ 时成立.

0.2.2 加权平均数

定义 0.2.2 给定一组数据 $x_1,x_2,\cdots,x_n$，又给定了一组正数 $p_1,p_2,\cdots,p_n$，且 $\sum_{i=1}^{n}p_i=1$，则称

$$\bar{x}=x_1p_1+x_2p_2+\cdots+x_np_n=\sum_{i=1}^{n}x_ip_i$$

为 $x_1,x_2,\cdots,x_n$ 的加权平均数，p_i 称为 x_i 的权.

由定义不难看出以下三点：

(1) 对于同一组数，给出不同的权(只要满足权均为正数及和为 1 两个条件)，可得出不同的加权平均数，即同一组数的加权平均数不止一个.

(2) 对于一组数，用某一组不全相等的数进行加权平均时，权较大的数据对加权平均的结果影响较大. 例如用不同的仪表测量物体运动的速度，有的仪表精度高，有的就差，把它们所测得的数据等同看待，显然是不妥的，就应对不同精度的仪表所测得的数据赋以不同的权，然后求加权平均数.

(3) 一般的均值(算术平均数)也可看作是一种加权平均数，相应的权 p_i 全一样，都是$\frac{1}{n}$. 也就是说，求均值时，我们是把 n 个数据等同看待的.

0.2.3 中位数

定义 0.2.3 将一组(有限个)数据，按从小到大的次序排好，设为

$$x_1,x_2,\cdots,x_n,\quad x_i\leqslant x_j,\quad (1\leqslant i<j\leqslant n)$$

当 n 为奇数时，我们称 $M=x_{\frac{n+1}{2}}$ 为这一组数据的中位数；

当 n 为偶数时，我们称 $M=\frac{1}{2}(x_{\frac{n}{2}}+x_{\frac{n}{2}+1})$ 为这一组数据的中位数.

在有些实际问题中，常用中位数作一组数据的代表值.

0.2.4 方差与标准差

定义 0.2.4 给定一组数据 $x_1,x_2,\cdots,x_n$，我们称

$$s^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

为这一组数据的方差. 其中 $\bar{x}$ 为 $x_1,x_2,\cdots,x_n$ 的均值；称

$$s=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

为这一组数据的标准差.

很明显，s^2 是各 x_i 与 $\bar{x}$ 的偏差之平方 $(x_1-\bar{x})^2,\cdots,(x_n-\bar{x})^2$ 的均值的 $\frac{n}{n-1}$ 倍，是平方偏差，因此称它为方差. 标准差 s 是方差 s^2 的算术平方根，它与 x_1，$x_2,\cdots,x_n$ 有相同的度量单位. 容易看出，s^2 越大，这组数据就越“分散”，或者说，这组数据的变异性（即互相不同的程度）就越大；s^2 越小，这组数据的变异性就越小，也就更“集中”. 当 $s^2=0$ 时，$x_1=x_2=\cdots=x_n=\bar{x}$，就没有变异.

因此，对一组数据 $x_1,x_2,\cdots,x_n$ 作分析时，$\bar{x}$ 与 s（或 s^2）是最常用的两个量：$\bar{x}$ 是代表性的值；s 或 s^2 是描述数据的变异性的值.

0.2.5 简化计算法

在概率统计中经常会遇到均值、方差的计算，为了使计算简化，下面几个公式是经常会用到的.

命题 0.2.1 若 s^2 是一组数据 $x_1,x_2,\cdots,x_n$ 的方差，则

$$s^2=\frac{1}{n-1}\left(\sum_{i=1}^{n}x_i^2-n\bar{x}^2\right).$$

证明 由定义 0.2.4 知，$s^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$，而

$$\begin{aligned}\sum_{i=1}^{n}(x_i-\bar{x})^2&=\sum_{i=1}^{n}(x_i^2-2x_i\cdot\bar{x}+\bar{x}^2)=\sum_{i=1}^{n}x_i^2-2\bar{x}\cdot\sum_{i=1}^{n}x_i+n\cdot\bar{x}^2\\&=\sum_{i=1}^{n}x_i^2-2\bar{x}\cdot n\bar{x}+n\cdot\bar{x}^2=\sum_{i=1}^{n}x_i^2-n\cdot\bar{x}^2,\end{aligned}$$

所以