高等职业教育数学系列教材

GAODENG ZHIYE JIAOYU SHUXUE XILIE JIAOCAI

主 编 郭建英

编著者 郭建英 彭明珠 林海平

概率统计

北京大学出版社

PEKING UNIVERSITY PRESS

高等职业教育数学系列教材

概　率　统　计

主　编　郭建英

编著者　郭建英　彭明珠　林海平

北京大学出版社

PEKING UNIVERSITY PRESS

图书在版编目(CIP)数据

概率统计/郭建英主编. —北京：北京大学出版社，2005. 7
(高等职业教育数学系列教材)
ISBN 7-301-09084-6

Ⅰ. 概… Ⅱ. 郭… Ⅲ. ①概率论-高等学校：技术学校-教材 ②数理统计-高等学校：技术学校-教材 Ⅳ. O21

中国版本图书馆 CIP 数据核字(2005)第 047996 号

书　　　名：概率统计
著作责任者：郭建英　主编
责 任 编 辑：刘　勇
标 准 书 号：ISBN 7-301-09084-6/O·0647
出 版 发 行：北京大学出版社
地　　　址：北京市海淀区成府路 205 号　100871
网　　　址：http://cbs.pku.edu.cn　电子信箱：zpup@pup.pku.edu.cn
电　　　话：邮购部 62752015　发行部 62750672　理科编辑部 62752021
排　版　者：北京高新特打字服务社　82350640
印　刷　者：北京大学印刷厂
经　销　者：新华书店
787mm×1092mm　16 开本　13 印张　321 千字
2005 年 7 月第 1 版　　2005 年 7 月第 1 次印刷
印　　　数：0001—4000 册
定　　　价：21.00 元

内 容 简 介

本书是《高等职业教育数学系列教材》之一的工科类、经济类、管理类“概率统计”课程的教材.本书按照教育部制定的高职“概率统计课程教学基本要求”编写，反映了当前高等职业教育培养高素质实用型人材数学课程设置的教学理念.全书共分八章，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.每章配备了A类习题与B类习题，书末附有答案或提示，以供读者参考.为了满足教学的需要，本书还编写了四套模拟试题，其中前两套供专科层次的学生使用，后两套供本科层次的学生使用，读者可用模拟试题自测学习效果.

本书突出体现了作者在教学第一线积累的丰富教学经验，注重对学生基础知识的传授和基本能力的培养.本书针对高职院校开设“概率统计”课程的实际，并考虑到多层次学生的学习目标，由浅入深、循序渐进、形象直观地讲述概率统计的基本概念、基本理论及方法，强调典型实例的应用价值，以培养学生分析问题和解决问题的综合素质.

本书可作为高等职业教育工科类、经济管理类大学生“概率统计”课程的教材或教学参考书，也可供成人教育相关专业的学生学习参考.

高等职业教育系列教材编写委员会

序　言

为了适应我国高等职业教育迅速发展的需要，适应高等职业教育多层次办学的需要，我们编委会应北京大学出版社之邀规划、编写了《高等职业教育系列教材》.

我国高等职业教育兴起于20世纪90年代中期，至今已得到迅速发展，受到人们的广泛关注.为了培养出具有一定科学素质和职业技能的优秀人才，无论是在专科教育还是在本科教育方面，我们都一直在进行艰辛的探索.高等职业教育的教材担负着教改的重任，在教学实践中，直接关系教学质量，在引导教学教法、理论联系实际、指导实践等方面具有重要作用，为此我们始终将教材建设作为教学工作的重要组成部分.

从高等职业教育培养技术型应用人才这一目标来看，高等职业教育的基础课教材应当体现积极的创造性思维的训练，以提高学生的科学素质和工作能力；内容不仅要体现该知识系统的精华，而且应具有系统的伸缩性和可选性，以适应不同层次教学的实际需要；教学内容与课后的训练应具有方便学生的自修性，以发挥学生作为学习主体的积极作用.专业课教材的内容应当具有工作实践的应用性，体现实际工作的规律性，理论印证性的推导内容在不影响今后实践需要的情况下，应代之以翻阅技术资料、查阅工程手册的实际应用能力的培养，只有这样，高等职业教育教材才能走出传统本科教材和专科压缩本科教材的编写模式.

教材归根结底是为学生服务的，是为学生今后从事工作打基础的，因此教材内容还需要体现该学科或该专业的科学性和先进性，以适应未来工作的实际需要；内容安排上必须循序渐进，由浅入深，把握好学生知识水平的可接受性；在陈述上必须通俗易懂，简练明了，注重化抽象为具体再由具体到抽象的过程，这样才能确保学生在学习中真正掌握知识.

编委会组织编写的教材力图体现上述编写原则，集优秀教师的教学经验认真编好每一部教材，为高等职业教育教改做出自己应有的贡献是我们的宗旨.

编委会

2004年7月于北京

前　言

本书是高等职业院校“概率统计”课程的教材.本书按照教育部制定的相关教学大纲,以高等职业教育的培养目标为宗旨、以提高学生科学素质和综合能力为指导来编写.编写中既考虑知识结构的完善性又注重其实用性,反映了当前高等职业教育数学课程设置的教学理念.本书共分八章,前四章是概率论部分,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;后四章是数理统计部分,内容包括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.

概率统计是借助数量研究随机现象规律性的数学学科,它是工科、经济管理类学生重要的基础课程.随着时代的发展及知识的更新,概率统计广泛应用到自然科学与社会科学的各个领域,并取得了重要成果.在编写中,我们紧扣教学大纲,针对工科及经济类等学科的实际要求,结合教学经验,并考虑到多层次学生的学习目标,采用了由浅入深、循序渐进、融会贯通的教学原则与直观形象的教学方法,注重概率统计的基本概念、基本理论、基本方法的阐述;同时又强调培养学生基本运算能力和分析问题、解决问题的综合素质.教材配备了大量的几何图形与实用型案例,以达到丰富课堂教学以及便于学生自学的目的.

本书加 * 号的内容供任课教师根据学时和教学实际酌情选用.为了帮助多层次学生更好地掌握本课程的教学内容,本教材每章配备了 A 类习题与 B 类习题,其中带有 * 号的题目是最新《全国硕士研究生入学统一考试・数学考试大纲》中提供的习题,以供学生掌握基本练习题及具有一定难度的提高练习题.此外,我们还配备了模拟试题一、二(专科适用)和模拟试题三、四(本科适用),以供读者复习,巩固所学知识.书后附有习题答案与提示,可供教师与学生参考.

本书承蒙赵佳因副教授审稿,特此表示感谢.限于编者的水平和时间仓促,书中的不当之处,恳请读者批评指正.

我们谨将本书奉献给广大读者,希望它能成为读者学习概率统计的良师益友.本书的编写与出版得到了北京大学出版社的大力支持与帮助,在此表示衷心的感谢.

编　者

2005 年 1 月

目　录

预备知识

在概率论与数理统计中，常常用到排列与组合的知识，更重要的是要用到排列与组合的思维方法，为此我们介绍排列组合的有关内容，把它们分三部分作为预备知识介绍给读者.

一、基本原理

乘法原理 完成一件事情，需要分 m 个步骤进行，其中第一个步骤有 n_1 种方法，第二个步骤有 n_2 种方法，……，第 m 个步骤有 n_m 种方法，则完成这件事情的方法共有

$$N = n_1 \times n_2 \times \cdots \times n_m = \prod_{i=1}^{m} n_i$$

种方法.

例 1 从北京经过天津到上海，若从北京到天津的路线有 A_1 和 A_2 两条，从天津到上海的路线有 B_1, B_2, B_3 三条，问从北京经天津到达上海的路线共有多少种?

解 从北京到上海分两个步骤进行，第一个步骤从北京到天津的路线有 2 条，所以从北京到天津的方法有 2 种；第二个步骤从天津到上海的方法有 3 种. 因为从北京到上海是分两个步骤进行的，所以总的方法数为每一步骤的方法数相乘，即从北京经天津到达上海共有

$$N = 2 \times 3 = 6$$

种方法，可给出下面的路线图：

$$\text{北京}\begin{cases}\text{第一种路线}(A_1)\text{ 至天津}\begin{cases}\text{第一种路线}(B_1)\text{ 至上海}\\ \text{第二种路线}(B_2)\text{ 至上海}\\ \text{第三种路线}(B_3)\text{ 至上海}\end{cases}\\ \text{第二种路线}(A_2)\text{ 至天津}\begin{cases}\text{第一种路线}(B_1)\text{ 至上海}\\ \text{第二种路线}(B_2)\text{ 至上海}\\ \text{第三种路线}(B_3)\text{ 至上海}\end{cases}\end{cases}$$

加法原理 完成一件事情，共有 m 类方式，其中第一类方式中有 n_1 种方法，第二类方式中有 n_2 种方法，……，第 m 类方式中有 n_m 种方法，并且任何一类方式中的任何一种方法都能完成这件事情，则完成这件事情共有

$$N = n_1 + n_2 + \cdots + n_m = \sum_{i=1}^{m} n_i$$

种方法.

例 2 从北京到上海，有三类交通工具，其中第一类是火车，有 A_1, A_2, A_3 三班次火车；第二类是飞机，有 B_1, B_2 两班次飞机；第三类是汽车，有 C_1, C_2, C_3, C_4 四班次汽车. 问从北京到达上海的方法一共有多少种?

解 从北京到达上海的方式有三类：第一类乘火车，第二类乘飞机，第三类乘汽车. 每一类的任何一种方法都能到达上海，完成从北京到上海这件事情，其中第一类乘火车的方法有 3 种，第二类乘飞机的方法有 2 种，第三类乘汽车的方法有 4 种，则从北京到上海共有

$$N = 3 + 2 + 4 = 9$$

种方法,可给出下面的路线图:

$$\text{北京}\begin{cases}\text{火车}\begin{cases}A_1(\text{班次})\text{到上海}\\ A_2(\text{班次})\text{到上海}\\ A_3(\text{班次})\text{到上海}\end{cases}\\ \text{飞机}\begin{cases}B_1(\text{班次})\text{到上海}\\ B_2(\text{班次})\text{到上海}\end{cases}\\ \text{汽车}\begin{cases}C_1(\text{班次})\text{到上海}\\ C_2(\text{班次})\text{到上海}\\ C_3(\text{班次})\text{到上海}\\ C_4(\text{班次})\text{到上海}\end{cases}\end{cases}$$

在应用基本原理时,必须**注意**加法原理和乘法原理的**根本区别**.若完成一件事情有多类方式,其中任何一类方式的任何一种方法都可以完成这件事情,这时用加法原理;若完成一件事情需要分多个步骤依次完成,这时用乘法原理.

二、排列

1. 不同元素不重复的排列

定义 1 从 n 个**不同**的元素中,任取其中 m $(m\leqslant n)$个元素排成有顺序的一列,所有这样排列的方法数称为**排列数**,简称**排列**,记作 P_n^m.

如何计算 P_n^m? 抽取这 m 个元素要分 m 个步骤进行,并且每一步只抽取其中一个元素,在抽取时请读者注意两点:第一,已给的 n 个元素都不同,故每一种元素只有一个;第二,每一个元素抽取后都放在一排的某个位置上,抽取后不再放回.因此,抽取第一个元素时有 n 个不同元素,故有 n 种方法;第二个元素在剩下的$(n-1)$个元素中抽取,故有$(n-1)$种方法;抽取第三个元素时只剩下$(n-2)$个元素,故只有$(n-2)$种方法……抽取第 m 个元素时前面已抽取$(m-1)$个元素,故只剩下$[n-(m-1)]=(n-m+1)$个元素,故只有$(n-m+1)$种方法,由于是分 m 个步骤完成的,根据乘法原理:所有排列(数)为

$$\mathrm{P}_n^m = n(n-1)(n-2)\cdots(n-m+1).$$

若 $m<n$ 时,称排列为**选排列**;

若 $m=n$ 时,称排列为**全排列**,记为 P_n^n,且 $\mathrm{P}_n^n=n(n-1)(n-2)\cdots 2\cdot 1=n!$.

2. 不同元素可重复的排列

前面我们介绍了不同元素不重复的排列,现在介绍不同元素可重复的排列.若有 n 种元素,每一种元素可以重复抽取,则从这 n 种元素中抽取其中 m 个元素排成一排的方法总数为 n^m.

这是因为每一步骤抽取一个元素的方法都是 n 种,故 m 个步骤的方法总数为

$$\underbrace{n\times n\times\cdots\times n}_{m\text{个}} = n^m.$$

例 3 目前北京市区的普通电话号码为八位,试分析以 6232 为前四位的电话号码共有多少个?

解 由于电话号码前四位为 6232,只需确定后四位的数字,每种数码可以重复使用,因此

可以排成的后四位数字的方法数共有 $10^4=10000$(种).

例 4 由 0,1,2,3,4,5,6,7,8,9 这 10 个数码,任取其中 4 个数码构成四位偶数,可以有多少种构成方法?

解 由于 4 位数可以重复,且第一位数不能为 0,所以有 9 种取法;第二与第三位数无特殊要求,有 10 种取法;第四位数只能取 0,2,4,6,8,所以有 5 种取法. 故所有的构成方法数共有 $9\times10\times10\times5=4500$(种).

例 5 在北京、上海、广州这三座城市间需印多少种不同的飞机票?

解 每一张飞机票上有两座城市,顺序不同,机票也不同,而且每一张机票上两座城市不能相同,即属于不重复选排列,故有 $\mathrm{P}_3^2=3\times2=6$(种).

例 6 某邮政局设有 4 个邮箱,现将 3 封信逐一投放邮箱,问共有多少种投法?

解 将 3 封信逐一投放邮箱,必须依次经过三个步骤:第一个步骤是将第一封信投入 4 个邮箱的其中 1 个,有 4 种方法;第二个步骤是将第二封信投入 4 个邮箱的其中的一个邮箱,有 4 种方法;同理,第三个步骤,也有 4 种方法. 根据乘法原理,共有 $4^3=64$(种)投放方法.

三、组合

定义 2 从 n 个不同的元素中,任取其中 $m(m\leqslant n)$ 个元素组成一组,所有的方法数称**组合数**,简称**组合**,记作 C_n^m.

下面我们推导组合数 C_n^m 的计算公式.

上面我们已给出排列数 P_n^m 的计算公式

$$\mathrm{P}_n^m=n(n-1)\cdots(n-m+1).$$

实际上,从 n 个不同元素中,任取其中 m 个元素排列成有顺序的一排的方法数 P_n^m,我们也可这样分析:

第一步:先取其中 m 个元素组成与顺序无关的一组,它的方法数为 C_n^m.

第二步:在每一个由 m 个元素组成的一组再进行全排列,方法数为 $m!$.

根据乘法原理有 $\mathrm{P}_n^m=m!\mathrm{C}_n^m$,由此得

$$\mathrm{C}_n^m=\frac{\mathrm{P}_n^m}{m!}=\frac{n(n-1)\cdots(n-m+1)}{1\cdot2\cdot\cdots\cdot m}.$$

上面的公式就是**组合数的计算公式**,同时规定:$\mathrm{C}_n^0=\mathrm{C}_n^n=1$.

注意:排列与组合的区别在于前者与顺序有关,后者与顺序无关.

性质 1 $\mathrm{C}_n^m=\mathrm{C}_n^{n-m}$ (当 $m>n/2$ 时利用组合性质计算组合数要方便).

性质 2 $\mathrm{C}_n^m\overset{①}{=\!=}\mathrm{C}_{n-1}^m+\mathrm{C}_{n-1}^{m-1}$.

性质 3 $\sum\limits_{k=0}^{n}\mathrm{C}_n^k=2^n$.

例 7 根据组合数计算:

(1) C_5^2; (2) C_7^3; (3) C_{100}^{97}; (4) C_{1000}^{998}.

① 从 n 个不同元素中取出 m 个元素的组合,对于某个元素 a 来说,可分为两类:一类是含 a 的组合,共有 C_{n-1}^{m-1}种;另一类是不含 a 的组合,共有 C_{n-1}^m种. 因此有 $\mathrm{C}_n^m=\mathrm{C}_{n-1}^{m-1}+\mathrm{C}_{n-1}^m$.

解 (1) $C_5^2=\frac{5\times4}{2\times1}=10.$　　(2) $C_7^3=\frac{7\times6\times5}{3\times2\times1}=35.$

(3) $C_{100}^{97}=C_{100}^{3}=\frac{100\times99\times98}{3\times2\times1}=161700.$

(4) $C_{1000}^{998}=C_{1000}^{2}=\frac{1000\times999}{1\times2}=499500.$

例 8 在 6 件不同的产品中,其中有 4 件正品,2 件次品,问:

(1) 任取出其中两件,有多少种不同取法?

(2) 任取出其中两件中恰有一件次品,有多少种不同取法?

解 由于取出的两件产品与顺序无关,故可用组合数计算.

(1) 6 件不同的产品中任取出其中的两件,有 $C_6^2=\frac{6\times5}{1\times2}=15$ 种不同取法.

(2) 6 件不同的产品中,任取出其中两件恰有一件次品,有 $C_4^1\cdot C_2^1=8$ 种不同取法.

第一章　随机事件的概率

概率论是研究大量随机现象统计规律性的一门学科，是统计学的重要基础，它研究的主要对象是事件及其概率、随机变量及其概率分布和数字特征. 在自然科学、经济领域、社会科学、工程和技术以及其他领域，概率论都有广泛的应用.

§1　随 机 事 件

一、确定性现象和随机现象

在自然界和经济活动中的各种现象大致可以分为两种现象，一种现象是事先可以预言的，即在一定条件下必然发生或必然不发生的现象，这种现象称为**确定性现象**. 例如，积压商品一定会增加库存费用. 再如，从一批合格产品中任意抽取一件产品一定是合格品. 另一种现象是事先不能预言的，即在一定的条件下，可能发生也可能不发生的现象，这种现象称为**随机现象**(指带有不确定性或偶然性的现象). 实质上，确定性现象是指在一定的条件下，它出现与否，以及出现时所产生的结果是完全确定的，是可以事先确切预测的. 在自然界中存在大量确定性现象：在标准大气压下，纯水加热到 100℃ 时沸腾，在 0℃ 时会结冰；导体在通电时会发热等. 而随机现象是指在一定的条件下，它出现与否，以及出现时产生的结果是事先不能确定的. 例如：某商店的日销售总额；某年的农产品的产量；在相同的条件下生产的一大批产品的次品率等等. 随机现象有大量与个别之分，在相同条件下可以重复进行观测的随机现象，称为**大量随机现象**；带有偶然性特点的，但原则上不能在相同条件下重复进行观测的随机现象，称为**个别随机现象**，例如：一些带有偶然性特点的具体历史事件，自然灾害等等. 以后，我们研究的随机现象是指**大量随机现象**.

通常，人们不论研究何种现象，都离不开对其进行观察(测)或进行实验. 为简便起见，我们把对某现象或对某事物的某个特征的观察(测)，以及各种各样的科学实验，统称**试验**. 为了研究随机现象，同样需要进行试验. 这类试验的特征是，在一定的条件下，其试验的可能结果不止一个；一次试验中，可能出现某一结果，也可能出现另一个结果，究竟会出现哪一个试验结果，事先无法准确地预言. 对于这类试验，人们实践中发现，就一次试验而言，其试验结果表现出不确定性(偶然性)，似乎难以捉摸，但在大量重复试验下，其试验结果却呈现出某种规律性. 例如，抛掷硬币试验，一次抛掷，哪一面朝上是随机的(或偶然的)，但把同一硬币进行成千上万次抛掷，人们发现，“正面朝上”与“反面朝上”这两个试验结果出现的次数大致各占一半. 又例如，从含有不合格品的一批某种产品中，任意抽取一件进行检验，“抽到合格品”与“抽到不合格品”两个试验结果都有可能发生，该试验具有随机性(或偶然性)，但当重复抽取时，“抽到合格品”的次数与抽取总次数之比却呈现出某种稳定性. 随机现象的这种隐蔽着的内在规律性叫做**统计规律性**. 概率论的任务就是要研究和揭示这种规律性.

显然，要获得随机现象的统计规律性，必须在相同的条件下，大量重复地做试验. 在概率统

计中，把这类试验称为**随机试验**，它具有下述三个特性：

(1) 试验可以在相同的条件下重复进行；

(2) 每次试验的可能结果不止一个，而究竟会出现哪一个结果，在试验前不能准确地预言；

(3) 试验所有的可能结果在试验前是明确（已知）的，而每次试验必有其中的一个结果出现，并且也仅有一个结果出现.

随机试验简称为**试验**，并用字母 E 或 E_1, E_2 等表示. 我们就是通过随机试验去研究随机现象的.

例 1 将一枚质地均匀的硬币连掷两次，观察出现正、反面的情况. 这里把硬币连掷两次作为一次试验，就是一个随机试验，记为 E_1. 若用记号 $\omega_{\text{正}}$ 表示出现正面，$\omega_{\text{反}}$ 表示出现反面，则 E_1 共有四种可能的结果：

$$(\omega_{\text{正}},\omega_{\text{正}}),\quad (\omega_{\text{正}},\omega_{\text{反}}),\quad (\omega_{\text{反}},\omega_{\text{正}}),\quad (\omega_{\text{反}},\omega_{\text{反}}).$$

例 2 投掷一枚骰子，观察出现的点数. 这个试验就是随机试验，记为 E_2. 若用 ω_k 表示骰子出现 k 点，则 E_2 共有六种可能的结果：

$$\omega_1,\quad \omega_2,\quad \omega_3,\quad \omega_4,\quad \omega_5,\quad \omega_6.$$

二、样本空间与随机事件

在随机试验中，尽管在每次试验之前不能预知试验的结果，但试验的所有可能结果组成的集合是已知的. 我们将随机试验 E 所有可能出现的结果组成的集合称为 E 的**样本空间**（或称为**基本事件组**），用大写希腊字母 Ω 表示样本空间，试验的每个可能结果称为**样本点**（或**基本事件**），用小写希腊字母 ω 表示（样本点的表示尽量符合原意），即：随机试验的全体样本点（或基本事件）组成的集合就是样本空间.

例 3 若用 E_1 表示连续两次投掷一枚质地均匀的硬币，$\omega_{\text{正}}$ 表示出现正面，$\omega_{\text{反}}$ 表示出现反面，则随机试验 E_1 的样本空间为

$$\Omega = \{(\omega_{\text{正}},\omega_{\text{正}}),(\omega_{\text{正}},\omega_{\text{反}}),(\omega_{\text{反}},\omega_{\text{正}}),(\omega_{\text{反}},\omega_{\text{反}})\}.$$

例 4 若用 E_2 表示投掷一枚骰子，ω_k 表示骰子出现 k $(k=1,2,3,4,5,6)$点的样本点，则随机试验 E_2 的样本空间为

$$\Omega = \{\omega_1,\omega_2,\omega_3,\omega_4,\omega_5,\omega_6\}.$$

例 5 若用 E_3 表示在 10 件产品中有 3 件次品，从中任取 4 件，用 ω_k 表示所取的 4 件产品中恰有 k $(k=0,1,2,3)$件次品的事件，则随机试验 E_3 的样本空间为

$$\Omega = \{\omega_0,\omega_1,\omega_2,\omega_3\}.$$

例 6 若用 E_4 表示某人射击目标，直到击中目标为止，用 ω_k 表示该人射击 k 次，则样本空间为

$$\Omega = \{\omega_1,\omega_2,\cdots,\omega_n,\cdots\}.$$

例 7 若用 E_5 表示从一大批某类电子元件，任意抽取一件，测试其使用寿命. 用 t 表示电子元件使用的时间，则随机试验 E_5 的样本空间为

$$\Omega = \{t \mid t \in [0, +\infty)\}.$$

从上面的例子可以看出，随机试验样本点的总数可以是有限多个，也可以是无限多个.

需要指出的是，试验的目的不同，样本空间也不同. 例如：某人进行两次投篮，(1) 若只观