

普通高等院校大学数学系列教材

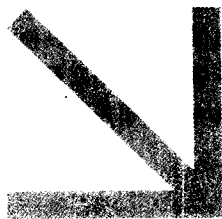
# 概率论与数理统计

龚光鲁 编著

清华大学出版社



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



普通高等院校大学数学系列教材

# 概率论与数理统计

龚光鲁 编著

清华大学出版社  
北京

## 内 容 简 介

本书是为大学层面的工科、经管学科等学生编写的入门教科书。先修知识只要求微积分和线性代数。本书以较小的篇幅阐述概率与统计的思维方法以及最基本的内容、概念与方法。写作时我们力求与国际接轨。内容包括：古典模型与概率空间；随机变量及随机数的生成；极限定理介绍；数据与统计，重点强调描述性统计；点估计与区间估计；假设检验，强调用  $p$  值作为否定零假设的依据，并介绍了用 Microsoft Excel 作统计计算；变量间的统计关系与回归模型；方差分析介绍；非参数方法及数据间的 Spearman 秩相关系数的介绍，正态化相关系数；再抽样法与刀切法介绍等。

版权所有，翻印必究。举报电话：010-62782989 13501256678 13801310933

本书封面贴有清华大学出版社防伪标签，无标签者不得销售。

本书防伪标签采用特殊防伪技术，用户可通过在图案表面涂抹清水，图案消失，水干后图案复现；或将表面膜揭下，放在白纸上用彩笔涂抹，图案在白纸上再现的方法识别真伪。

### 图书在版编目(CIP)数据

概率论与数理统计/龚光鲁编著. —北京：清华大学出版社，2006.5

(普通高等院校大学数学系列教材)

ISBN 7-302-12723-9

I. 概… II. 龚… III. ①概率论—高等学校—教材 ②数理统计—高等学校—教材  
IV. O21

中国版本图书馆 CIP 数据核字(2006)第 023579 号

出 版 者：清华大学出版社 地 址：北京清华大学学研大厦  
http://www.tup.com.cn 邮 编：100084  
社 总 机：010-62770175 客 户 服 务：010-62776969

组稿编辑：佟丽霞

文稿编辑：王海燕

印 刷 者：北京市清华园胶印厂

装 订 者：北京市密云县京文制本装订厂

发 行 者：新华书店总店北京发行所

开 本：170×230 印张：22 字数：403 千字

版 次：2006 年 5 月第 1 版 2006 年 5 月第 1 次印刷

书 号：ISBN 7-302-12723-9/O·522

印 数：1~4000

定 价：28.50 元

普通高等院校大学数学系列教材

编委会名单

主

任

萧树铁

编

委

王萼芳

龚光鲁

扈志明

计东海

陈冬

# 序

大学数学系列课程“微积分”、“线性代数”和“概率论与数理统计”是大学理工、管理等各专业的一个重要基础课程。随着我国经济的高速发展,高等教育的日益普及,需要培养出大批应用型工程技术人员。同重点大学相比,以培养应用型人才为主的普通高校在教学目标、教学内容、教学方式等方面都有很大的不同。而这类普通高校学生规模更大,但师资力量和教学条件却相对较弱。因此,编写高质量的面向此类高校的教材,对于促进教学改革,提高我国高等教育的教学质量,更具迫切性和非常重要的现实意义。

为此,我们组织清华大学、北京大学、哈尔滨理工大学、北京联合大学等高校的老师,编写了这套面向普通高校的“普通高等院校大学数学系列教材”,包括“微积分”、“线性代数”、“概率论与数理统计”及与每门课程主教材配套的教师用书(习题详细解答)、电子教案和学习指导。本套教材的作者均长期从事大学数学的教学工作,学术水平高,教学经验丰富,并编写出版过相关的教材,对大学数学系列课程的教学内容和课程体系改革有深入的研究。同时,来自于普通高校教师的参与使本套教材更有针对性,更符合当前这类高校培养目标的要求和基础数学教学的实际情况。

本套教材编写的主要原则是:强调各门课程整体的理念、基本方法和适当的应用。由于这三门课都属于基础课程,所以对其内容的改革应当慎重。这套教材内容涵盖了教育部发布的“工科类本科数学基础课程教学基本要求”。在取材方面,不是简单地对内容进行增删,而是在努力深入的基础上尽量做到“浅出”。

本套教材的全部讲授时间大约为 260 学时,其中微积分

## 序

140 学时、线性代数、概率论与数理统计各为 60 学时. 教师还可以根据本校的实际情况对课时作一定的增减, 重要的是: 每门课都应配置适当学时(例如, 总学时的  $1/3$  左右)的习题课.

清华大学出版社对本套教材的编写和出版给与了各方面的支持, 佟丽霞编辑为本书做了大量的组织和文字工作.

尽管作者都有良好的愿望和多年的教学经验, 但由于这一工作的难度较大, 时间又比较仓促, 各方面的问题肯定不少. 欢迎广大师生和各方人士提出宝贵意见, 以便进一步修改.

萧树铁

2006 年 4 月

# 前言

随着我国在科学、工程、管理、经济与人文等诸多领域的研究与应用的迅猛发展,概率统计的思维和方法,已经成为一个通向实用的平台.为了适应这个潮流,我们撰写了这本教材,希望它能成为广大朋友学习的益友.

本书是为大学层面的工科、经管学科等的学生编写的入门教科书.先修知识只要求微积分和线性代数.我们以较小的篇幅,阐述概率与统计的思维方法以及最基本的内容、概念与方法.写作时我们力求与国际接轨.

本书带有\*号的章节与例子是更为扩展的内容,初学时可以忽略.

本书有如下特点:

1. 强调统计作为建模工具的作用.概率论更多的是提供模型,而统计是用数据资料建模,在拟合模型的基础上作统计分析.
2. 对描述性统计给予充分的重视.
3. 强调用样本  $p$ -值的大小作为拒绝零假设根据的优越性.
4. 介绍随机数的生成和随机模拟的作用.在没有确切数学表示的随机现象中,随机模拟能为研究标的随机变量及随机环境提供作数值模拟的“实验”手段.
5. 介绍如何使用 Microsoft Office 中的通用程序 Excel 作统计计算.
6. 给出术语的目录与详细索引,使读者非常方便地查阅.
7. 给出术语的英文名称,便于与外文教材及文献接轨.

## 前 言

8. 引理、命题、定理、推论采用同一标识序号,查阅十分方便.
  9. 介绍了再抽样法与刀切法.
  10. 介绍了数据间的 Spearman 秩相关系数,正态化相关系数.
- 本书也可以供有关人员参考使用.

作 者

2006 年 2 月



# 常用符号

符 号	含 义
$a^+ = \begin{cases} a & (a > 0) \\ 0 & (a \leq 0) \end{cases}$	实数 $a$ 的正部
$A^c = \Omega \setminus A$	$A$ 的对立事件
$A \cup B$	事件“ $A$ 或 $B$ ”
$A \cap B$	事件“ $A$ 且 $B$ ”
$B(n, p)$	二项分布
$\text{cov}(X, Y), \sigma_{XY}$	随机变量 $X, Y$ 的协方差
$(CV)_X$	随机变量 $X$ 的变异系数
$\gamma(X)$	随机变量 $X$ 的偏度系数
$EX, \mu_X$	随机变量 $X$ 的(数学)期望, 均值
$E(Y X=x)$	条件 $X=x$ 下的条件期望
$E(Y X)$	作为随机变量的条件期望: $E(Y X) \equiv \varphi(X)$ , 其中 $\varphi(x) \equiv E(Y X=x)$
$\exp_\lambda$	均值为 $\frac{1}{\lambda}$ 的指数分布
$\mathcal{F}$	事件体, $\sigma$ -代数
$f_X(x); F_X(x)$	随机变量 $X$ 的概率函数, (边缘)分布; 分布函数
$f_{X,Y}(x, y), F_{X,Y}(x, y)$	随机向量 $(X, Y)$ 的联合分布, 联合分布函数
$f_{Y X}(y x)$	$X=x$ 条件下, 随机变量 $Y$ 的条件分布
$F(n, m), F_p(n, m)$	自由度为 $(n, m)$ 的 $F$ 分布, 及其 $1-p$ 分位点(或尾概率为 $p$ 的点)
$\Phi(x) \left( = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{u^2}{2}} du \right)$	标准正态随机变量的分布函数
$G(z)$	非负随机变量的母函数
$\Gamma(\alpha)$	gamma 函数
$H_0, H_1$	零假设, 对立假设

符 号	含 义
$I_A$	集合 $A$ 的示性函数
$I_n(\theta) = (I_n(\theta)_{ij})$	Fisher 信息量矩阵
$\kappa(X)$	随机变量 $X$ 的峰度系数
$M(t), M_X(t) (= Ee^{tX})$	随机变量 $X$ 的矩母函数
$\text{mul}(n; p_1, \dots, p_k)$	多项分布
$\mu_k = E(X - EX)^k$	$k$ 阶中心矩
$\mu'_k = EX^k$	$k$ 阶原点矩
$N(\mu, \sigma^2)$	正态分布
$(X, Y) \sim N(\mu_1, \mu_2, \rho, \sigma_1^2, \sigma_2^2)$	二维正态分布
$o(l)$	无穷小量
$o(h)$	比 $h$ 高阶的无穷小量
$O(l)$	有界量
$O(h)$	与 $h$ 同阶的变量
$P_B(A), P(A B)$	已知事件 $B$ 发生的条件下, 事件 $A$ 发生的条件概率
Poisson <sub><math>\lambda</math></sub>	Poisson 分布
$\pi(\theta)$	先验分布密度, 或先验概率函数
$\pi(\theta   x_1, \dots, x_n)$ ,	后验分布密度, 或后验概率函数
$\pi_{\theta   x_1, \dots, x_n}(\theta   x_1, \dots, x_n)$	
$r_{X,Y}$	随机变量 $X, Y$ 的相关系数
$\Sigma$	(协)方差矩阵
$s_x^2 = \frac{(x_1 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}$	数据 $x_1, \dots, x_n$ 的方差(离差)
$S_x^2 = \frac{(X_1 - \bar{X})^2 + \dots + (X_n - \bar{X})^2}{n}$ ,	
$S_0^2 = \frac{(X_1 - \mu_0)^2 + \dots + (X_n - \mu_0)^2}{n}$	
$s_x^2 = \frac{(x_1 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n-1}$	数据 $x_1, \dots, x_n$ 的修正了的无偏方差
$S^2 = \frac{(X_1 - \bar{X})^2 + \dots + (X_n - \bar{X})^2}{n-1}$	
s. d.	标准差
s. e.	标准误差
$t(n), t_p(n)$	自由度为 $n$ 的 $t$ 分布, 及其 $1-p$ 分位点(或尾概率为 $p$ 的点)
$U[a, b]$	区间 $[a, b]$ 上的均匀分布

续表

符 号	含 义
$\text{var}(X), \sigma_X^2$	随机变量 $X$ 的方差
$\text{var}(Y X=x)$	随机变量 $Y$ 在条件 $X=x$ 下的条件方差
$\text{var}(Y X)$	作为随机变量的条件方差: $(\text{var}(Y X) \equiv \psi(X), \text{其中}$ $\psi(x) \equiv \text{var}(Y X=x))$
$\hat{y} = \hat{a} + \hat{b}x$	回归直线
$z_p$	标准正态分布的 $1-p$ 分位点(或尾概率为 $p$ 的点)
$\omega$	基本事件, 样本点
$\Omega$	样本空间, 全集, 必然事件
$(\Omega, \mathcal{F}, P)$	概率空间
$X(\omega), X$	随机变量
$x^{(1)}, \dots, x^{(n)}$	数据 $x_1, \dots, x_n$ 的顺序(从小到大的)数据
$x_1^*, \dots, x_n^*$	数据 $x_1, \dots, x_n$ 的标准化
$\hat{\theta}$	未知参数 $\theta$ 的样本估计(例如: $\widehat{\text{var}}(X)$ 是 $\text{var}(X)$ 的样本估计)
$X_n \xrightarrow{P} X$	随机变量列的依概率收敛
$X_n \approx N(0, 1)$	$n$ 大时, 随机变量 $X_n$ 的分布近似地为 $N(0, 1)$
$\emptyset$	空集, 不可能事件
$\chi^2(n), \chi_p^2(n)$	自由度为 $n$ 的 chi 平方分布, 及其 $1-p$ 分位点(或尾概率为 $p$ 的点)

# 目 录

概  
率  
论  
与  
数  
理  
统  
计

## 第 1 章 古典模型与概率空间 ..... 1

### 1.1 概率的古典模型与对概率认识的经验概括 ..... 1

#### 1.1.1 引言 ..... 1

#### 1.1.2 概率的古典模型 ——等可能性分析 ..... 2

#### 1.1.3 事件的运算及古典模型中概率的 加法法则 ..... 6

#### 1.1.4 古典模型的条件概率 ..... 9

#### 1.1.5 频率与其加法法则 ..... 10

### 1.2 概率的公理模型——概率空间 ..... 11

#### 1.2.1 引言 ..... 11

#### 1.2.2 概率的公理模型 ..... 12

#### 1.2.3 概率的公理模型的直接推论 ——概率的性质及计算 ..... 14

#### 1.2.4 公理化概率空间中的条件概率 ..... 19

#### 1.2.5 乘法公式 ..... 20

#### 1.2.6 全概率公式 ..... 21

#### 1.2.7 Bayes 公式(逆概率公式) ..... 26

#### 1.2.8 随机事件的独立性 ..... 29

### 习题 1 ..... 34

## 第 2 章 离散随机变量 ..... 42

### 2.1 Bernoulli 随机变量及其分布 ..... 42

#### 2.1.1 Bernoulli 随机试验列与描述它的 随机变量 ..... 42

2.1.2	成功率很小时二项概率的近似计算、Poisson 近似	45
2.2	离散随机变量与其分布函数	46
2.2.1	离散随机变量与 Poisson 随机变量	46
2.2.2	离散随机向量	54
2.2.3	边缘分布	56
2.3	离散随机变量的条件分布、独立性	58
2.3.1	离散随机变量的条件分布	58
2.3.2	离散随机变量的独立性	60
2.4	离散随机变量的数字表征	63
2.4.1	离散随机变量的数学期望(均值)	63
2.4.2	一些常见的离散分布的期望	65
2.4.3	离散随机变量 $Y$ 在条件 $X=x$ 下的期望与 $Y$ 关于 $X$ 的条件期望	68
2.4.4	离散随机变量的方差和变异系数	70
2.4.5	一些常见的离散分布的方差和变异系数	73
2.4.6	离散随机变量的条件分布的方差	76
2.5	两个随机变量的协方差与相关系数	76
2.5.1	两个随机变量的协方差与相关系数的定义及其统计含义	76
2.5.2	协方差的计算与性质	80
	习题 2	82
<b>第 3 章</b>	<b>连续型随机变量</b>	<b>87</b>
3.1	连续型随机变量	87
3.1.1	连续型随机变量的定义	87
3.1.2	概率密度函数的一些性质	89
3.2	连续型随机向量与边缘分布及条件分布	92
3.2.1	联合密度与边缘分布	92
3.2.2	联合分布函数与边缘分布函数	93
3.2.3	连续型随机变量的数字表征	96
3.3	常见的连续型随机变量的分布及其数字表征	98

3.3.1	均匀随机变量、均匀分布与随机数 .....	98
3.3.2	指数随机变量与指数分布 .....	101
3.3.3	正态随机变量与正态分布 .....	104
3.3.4	二维正态随机向量与二维正态分布 .....	110
3.4	连续型随机变量的条件密度、混合分布与条件期望 .....	113
3.4.1	连续型随机变量的条件概率与条件密度 .....	113
* 3.4.2	混合分布 .....	116
* 3.4.3	连续型条件期望与全期望公式 .....	118
习题 3	.....	119
<b>第 4 章</b>	<b>随机变量的函数与矩母函数 .....</b>	<b>128</b>
4.1	随机变量的函数 .....	128
4.1.1	计算随机变量函数的分布的一般原则 .....	128
4.1.2	矩母函数 .....	130
4.1.3	对数正态分布 .....	132
4.1.4	两个随机变量和的分布 .....	133
4.2	gamma 族分布 .....	135
* 4.2.1	gamma 分布 .....	135
4.2.2	$\chi^2$ 分布 .....	138
4.3	随机变量的多个函数的联合分布 .....	139
4.3.1	一般公式 .....	139
4.3.2	商的分布 .....	142
4.4	随机数与随机模拟 .....	144
4.4.1	生成随机数的逆函数方法 .....	144
* 4.4.2	生成随机数的 von Neumann 取舍原则 .....	148
习题 4	.....	151
<b>第 5 章</b>	<b>极限定理介绍 .....</b>	<b>154</b>
5.1	Chebyshev 不等式与大数定律 .....	154
5.1.1	Chebyshev 不等式 .....	154
5.1.2	经典大数定律 .....	155

* 5.1.3 依概率收敛的性质 .....	157
5.2 中心极限定理 .....	159
5.2.1 极限定理的矩母函数方法 .....	159
5.2.2 独立同分布随机变量列的中心极限定理 .....	159
习题 5 .....	164
<b>第 6 章 数据与统计、正态抽样分布 .....</b>	<b>167</b>
6.1 总体与样本、随机样本 .....	167
6.1.1 总体 .....	167
6.1.2 样本、随机样本 .....	168
6.1.3 描述性统计与推断性统计 .....	170
6.2 随机数据与数据的描述性统计 .....	171
6.2.1 类别数据与定量数据 .....	171
6.2.2 类别数据的图形表示法 .....	171
6.2.3 定量数据的图形表示法 .....	172
6.2.4 一个应用——数据是否来自正态的一个经验做法 .....	176
6.3 定量数据的数字特征 .....	177
6.3.1 数据的均值、方差、标准差、偏度与峰度 .....	177
6.3.2 二维数据的数字特征 .....	178
6.4 抽样分布 .....	179
6.4.1 $Z$ 分布 .....	179
6.4.2 $\chi^2$ 分布 .....	180
6.4.3 $t$ 分布 .....	181
6.4.4 $F$ 分布 .....	183
习题 6 .....	186
<b>第 7 章 点估计方法 .....</b>	<b>188</b>
7.1 矩估计方法与百分位数匹配方法 .....	189
7.1.1 矩估计原则 .....	189
* 7.1.2 百分位数匹配方法 .....	191
7.2 极大似然估计方法 .....	192

7.2.1	似然函数与极大似然估计 .....	192
*7.2.2	不同分布的数据、删失数据、不独立数据情形的 极大似然估计 .....	194
7.3	估计优良的一些标准 .....	198
7.3.1	估计的偏差与无偏性 .....	198
7.3.2	估计的相合性要求 .....	200
7.3.3	估计量的均方误差、方差和方差的估计 .....	201
*7.3.4	Fisher 信息量与无偏估计的方差的 Cramer-Rao 下界 .....	203
*7.4	Bayes 估计 .....	205
7.4.1	Bayes 方法 .....	205
7.4.2	共轭先验分布 .....	207
习题 7	.....	211
<b>第 8 章</b>	<b>参数的区间估计 .....</b>	<b>218</b>
8.1	大样本情形的置信区间 .....	218
8.1.1	大样本情形参数区间估计的一般原则 .....	218
8.1.2	一个总体的情形 .....	219
8.1.3	两个总体的情形 .....	222
8.2	小样本情形正态总体的参数的置信区间 .....	223
8.2.1	一个正态总体的情形 .....	223
8.2.2	两个正态总体的情形 .....	224
8.3	均值的单侧置信区间、随机变量的风险值与容忍限 .....	226
8.3.1	均值的单侧置信区间 .....	226
*8.3.2	随机变量的风险值 .....	226
*8.3.3	正态数据的容忍限 .....	227
8.4	利用区间估计作统计判断 .....	228
习题 8	.....	229
<b>第 9 章</b>	<b>假设检验 .....</b>	<b>230</b>
9.1	假设检验的基本概念 .....	230



9.1.1	零假设与统计否定法 .....	231
9.1.2	两个类别变量间是否显著的不独立 ——两种因素的无关性检验 .....	233
9.1.3	区间估计与假设检验的关系 .....	239
9.2	大样本情形总体参数的假设检验问题 .....	239
9.2.1	大样本情形总体参数的 $Z$ 检验的一般原则 .....	239
9.2.2	一个总体大样本情形的 $Z$ 检验 .....	240
9.2.3	两个总体大样本情形的 $Z$ 检验 .....	241
9.3	正态总体小样本情形的假设检验 .....	242
9.3.1	一个正态总体的情形 .....	242
9.3.2	两个正态总体的情形 .....	245
9.3.3	均值的单侧检验 .....	248
9.4	假设检验的拒绝域与两种可能的错误 .....	249
9.4.1	检验的两种可能的错误 .....	249
*9.4.2	假设检验问题的法庭、病检与医疗事故责任 检验类比 .....	251
9.5	$\chi^2$ 检验用于列联表和拟合度 .....	253
9.5.1	关系的显著性检验 .....	253
9.5.2	类别变量的拟合度 (goodness) 检验与分布函数的 拟合 .....	256
习题 9	.....	259
<b>第 10 章</b>	<b>变量间的统计关系与回归模型</b> .....	<b>263</b>
10.1	一元线性回归模型 .....	264
10.1.1	正态随机误差模型 .....	264
10.1.2	正态误差模型的参数估计 .....	266
10.1.3	非正态误差的一元线性回归模型及其模型 的参数估计 .....	268
10.1.4	一元线性回归模型参数的假设检验和置信区间 ..	271
*10.1.5	一元线性回归模型的方差分析 .....	273
10.1.6	一元线性回归模型的预测 .....	275