研究生教学用书

教育部学位管理与研究生教育司推荐

高等数理统计

(第二版)

Advanced Mathematical Statistics

(Second Edition)

茆诗松　王静龙　濮晓龙　编著

高等教育出版社

研究生教学用书

教育部学位管理与研究生教育司推荐

高等数理统计

(第二版)

Advanced Mathematical Statistics

(Second Edition)

茆诗松　王静龙　濮晓龙　编著

高等教育出版社

内容提要

本书是“教育部推荐研究生教学用书”之一. 全书共分 6 章：基本概念、点估计、假设检验、区间估计、统计决策理论与 Bayes 分析、统计计算方法. 书中含有丰富的例子，着力说明统计思想和统计应用. 书中还配置了足够的习题，可使读者得到各种基本训练. 读完本书即可进入数理统计各分支的学习与研究.

本书可作为数学专业、统计专业研究生的教学用书和统计工作者的参考书.

图书在版编目(CIP)数据

高等数理统计 / 茆诗松，王静龙，濮晓龙编著. —2 版. —北京：高等教育出版社，2006.5

ISBN 7-04-019321-3

Ⅰ. 高… Ⅱ. ①茆… ②王… ③濮… Ⅲ. 数理统计-高等学校-教材 Ⅳ. O212

中国版本图书馆 CIP 数据核字（2006）第 021958 号

策划编辑 杨 波 **责任编辑** 杨 波 **封面设计** 李卫青 **责任绘图** 杜晓丹
版式设计 张 岚 **责任校对** 康晓燕 **责任印制** 尤 静

出版发行	高等教育出版社	购书热线	010-58581118
社 址	北京市西城区德外大街 4 号	免费咨询	800-810-0598
邮政编码	100011	网 址	http://www.hep.edu.cn
总 机	010-58581000		http://www.hep.com.cn
		网上订购	http://www.landraco.com
经 销	蓝色畅想图书发行有限公司		http://www.landraco.com.cn
印 刷	北京东光印刷厂	畅想教育	http://www.widedu.com
		版 次	1998 年 7 月第 1 版
			2006 年 5 月第 2 版
开 本	787×960 1/16		
印 张	30	印 次	2006 年 5 月第 1 次印刷
字 数	560 000	定 价	40.70 元

物料号 19321-00

再版前言

本书出版已有六年多，不少教师选用此书作为概率论与数理统计专业研究生教材．他们在指出不足之处的同时，还给了我们不少鼓励．特别，上海理工大学叶慈南教授把教学中发现的本书不足之处一一向我们指正．在此向各位同仁一并表示我们的谢意．此外，我们还要特别感谢高等教育出版社这次给我们一个再版的机会．

这次趁再版之时，我们修正了原书中所有已发现的不足之处．在保持原书结构的同时，适当增加一些内容，供教师选用．所增内容有五处，它们是：矩的近似（§1.4.5），稳健估计（§2.8），秩检验（§3.8），Bayes 假设检验（§5.5.6）和 Bayes 统计计算软件——Winbugs 简介（§6.4.6）．希望这些内容对扩大研究生的统计基础知识能起到积极的作用．

不当之处敬请读者指正．

茆诗松　王静龙　濮晓龙
2004 年 12 月于华东师范大学统计系

前　　言

本书是为统计学专业及相关专业的学生和统计工作者编写的教科书.阅读此书需要有高等数学基础和概率论与数理统计基础知识.读完本书即可进入数理统计各分支的学习和研究.基于这样的要求,我们在本书中着力于数理统计的基本概念、基本方法和基本理论,充分反映数理统计的现代发展,力求做到理论与实际的结合,为读者进入理论研究领域和实际应用领域打下扎实的基础.

全书共分六章,依次为基本概念,点估计,假设检验,区间估计,统计决策理论与 Bayes 分析,统计计算方法.前五章的前身是一份讲义,曾在华东师范大学统计系研究生高等数理统计课程上使用了十多年.虽经多次修改,总感不足.这次趁出版之际,对前五章作了较大的修改,充实了一些新的内容,另外在叙述上也作了不少改进,使内容有点新意,也更易理解.书中丰富的例子着力说明统计思想和统计应用领域,配置的习题足够让读者得到各种基本训练,掌握本书内容.完成这些习题就能品尝到统计学特有的味道.

本书的出版是在上海市学位委员会"上海研究生专项经费"资助下实现的,在他们大力支持和倡导下,我们充满信心地完成这本书的充实、完善和改写工作.在此对上海市学位委员会表示衷心的感谢.另外对我校研究生院培养处徐钧涛副教授,高等教育出版社张小萍和翁咏梅二位女士表示衷心的感谢,没有他(她)们的帮助与关心,此书不可能很快出版.

本书的编写和修改得到我系广大师生的帮助,特别是梁小筠教授和程依明副教授.另外还有尤进红、刘忠和何基报三位博士生为本书部分章节的打印、修改做了很多工作,在此一并表示衷心感谢.

本书由茆诗松主编.第一、五章由茆诗松执笔,第三、四章由王静龙执笔,第二、六章由濮晓龙执笔,最后由茆诗松统稿.由于编者水平有限,错缪之处在所难免,恳请国内同行及广大读者批评指正.

茆诗松　王静龙　濮晓龙

1998 年 4 月于华东师范大学统计系

郑重声明

目　录

第一章 基 本 概 念

§1.1 统计结构

§1.1.1 统计结构

概率论和数理统计都是研究随机现象统计规律性的数学学科，它们之间联系密切，但也有根本差别：在概率论中研究的出发点是一个概率空间$(\mathscr{X},\mathscr{B},P)$，即已知一个样本空间$\mathscr{X}$，$\mathscr{X}$中某些子集组成的$\sigma$代数$\mathscr{B}$和在可测空间$(\mathscr{X},\mathscr{B})$上定义的一个概率分布$P$，然后研究这个概率空间的各种性质；而在数理统计中研究的是一组受到随机性干扰的数据，再加上人们对这组数据的一些认识（即各种假设），就形成数理统计研究的出发点，然后对所考虑的问题作出统计推断或预测．为说清这个出发点，我们先看一个例子．

例 1.1（测量问题） 一个试验者对未知的物理量μ进行测量，为了对μ作出估计，大家知道，他的测量值x会受到各种随机因素的影响，以至于使x可认为是μ加上随机误差ε后而得到的，即

$$x = \mu + \varepsilon$$

这里的"可加性"是人们对测量数据构成所作的一个假设，经过多次使用经验，说明这个假设是合理的．另外，由于测量误差ε是受到测量仪器、环境温度、光线、视觉、心理等因素的微小变化而引起的综合结果，据中心极限定理，又可认为ε是服从均值为0和方差为σ^2的正态分布$N(0,\sigma^2)$．这是人们对测量数据的另一个假设（认识）．至此，我们对这个测量值问题的认识有如下三点：

1. 测量值x可取任何实数，实数集$\mathbf{R}$组成样本空间；
2. 实数集$\mathbf{R}$上的 Borel 集的全体组成的σ代数$\mathscr{B}_{\mathbf{R}}$；
3. 在可测空间$(\mathbf{R},\mathscr{B}_{\mathbf{R}})$上的一个概率分布族

$$\mathscr{P}_1 = \{N(\mu,\sigma^2):(\mu,\sigma) \in \mathbf{R} \times \mathbf{R}^+\}$$

其中$\mathbf{R}^+$是正实数集．

这样三件东西 $\mathbf{R}$,$\mathscr{B}_{\mathbf{R}}$ 和 $\mathscr{P}_1$ 就是我们研究测量问题的出发点. 假如不仅了解 ε 服从正态分布,而且还知其方差为 σ_0^2(比如知道测量仪器的精度),那么分布族就缩小为

$$\mathscr{P}_2 = \{N(\mu,\sigma_0^2):\mu \in \mathbf{R}\}$$

这时 $\mathbf{R}$,$\mathscr{B}_{\mathbf{R}}$ 和 $\mathscr{P}_2$ 就成为我们研究这个问题的出发点,假如我们对随机误差 ε 了解甚少,讲不出 ε 的分布是什么类型,只知道它是关于 0 对称的连续分布,那么分布族就扩大为

$$\mathscr{P}_3 = \{P:P\text{ 为 }\mathbf{R}\text{ 上关于 }\mu\text{ 对称的分布}\}$$

这时,研究这个问题的出发点就是 $\mathbf{R}$,$\mathscr{B}_{\mathbf{R}}$ 和 $\mathscr{P}_3$.

定义 1.1 设$(\mathscr{X},\mathscr{B})$为可测空间,$\mathscr{P}$为其上的一个概率分布族,则称三元组$(\mathscr{X},\mathscr{B},\mathscr{P})$为统计结构,或称为统计模型. 假如分布族 $\mathscr{P}$仅依赖于某个参数(或参数向量)θ,即

$$\mathscr{P} = \{P_\theta:\theta \in \Theta\}$$

其中 Θ 为参数空间,则称此结构为参数(统计)结构,或称为参数(统计)模型,否则称为非参数(统计)结构或非参数(统计)模型.

在例 1.1 中,$(\mathbf{R},\mathscr{B}_{\mathbf{R}},\mathscr{P}_1)$,$(\mathbf{R},\mathscr{B}_{\mathbf{R}},\mathscr{P}_2)$,$(\mathbf{R},\mathscr{B}_{\mathbf{R}},\mathscr{P}_3)$是三个不同的统计结构. 它们之间的差别可能反映实际背景的差别,也可能反映人们对实际情况认识上的差别,因此这三个结构可能都是合理的,至于选用哪一个结构,这已不是一个理论问题,而是一个实践性很强的问题,人们常凭借经验积累、专业知识和抽象概括等来确定统计结构.

在例 1.1 中,$(\mathbf{R},\mathscr{B}_{\mathbf{R}},\mathscr{P}_1)$和$(\mathbf{R},\mathscr{B}_{\mathbf{R}},\mathscr{P}_2)$是参数结构,而$(\mathbf{R},\mathscr{B}_{\mathbf{R}},\mathscr{P}_3)$是非参数结构. 人们往往希望从参数结构出发来研究问题,因为参数结构含有较多的信息,由此出发,可以获得精度较高的参数估计,但这样做要意识到是有风险的,因为当参数结构不真时,那推断结果可能离实际更远了. 若选用非参数结构,所冒风险就要小得多,因为非参数结构所含的信息较少,适应面广,但精度一般不会很高. 以后会看到,在这两类结构下所用的统计推断方法有很大差别,以至于在今天已形成统计中的参数方法与非参数方法两类.

§1.1.2 乘积结构与重复抽样结构

由简单的统计结构可以派生出一些比较复杂的统计结构.

定义 1.2 设$(\mathscr{X},\mathscr{B},\mathscr{P})$和$(\mathscr{X}',\mathscr{B}',\mathscr{P}')$是两个统计结构,则称$(\mathscr{X}\times\mathscr{X}',\mathscr{B}\otimes\mathscr{B}',\mathscr{P}\otimes\mathscr{P}')$为两者的乘积结构,并记为$(\mathscr{X},\mathscr{B},\mathscr{P})\otimes(\mathscr{X}',\mathscr{B}',\mathscr{P}')$,其中

$$\mathscr{P}\otimes\mathscr{P}' = \{P\otimes P':P\in\mathscr{P},P'\in\mathscr{P}'\}$$

类似地可以给出多于两个统计结构的乘积结构. 特别,n 个相同统计结构($\mathscr{X}$, $\mathscr{B}$,$\mathscr{P}$)的乘积结构称为重复抽样结构,记为($\mathscr{X}$,$\mathscr{B}$,$\mathscr{P}$)^n 或($\mathscr{X}^n$,$\mathscr{B}^n$,$\mathscr{P}^n$).

乘积结构在实际中相当于独立观察系统,重复抽样结构相当于对一个总体进行有限次独立抽样结果的描述. 今后,"从一个总体(或分布)抽取一个样本"与"从一个统计结构抽取一个样本"这两种说法是表示同一个意思.

例 1.2 在方差相等的两个正态均值的比较的问题中,若对于第一个正态总体 $N(\mu_1,\sigma^2)$获得 n_1 个观察值,对于第二个正态总体 $N(\mu_2,\sigma^2)$获得 n_2 个观察值,那么研究这个问题所涉及的统计结构是一个乘积结构.

$$(\mathbf{R}^{n_1},\mathscr{B}_{\mathbf{R}}^{n_1},\mathscr{P}_1^{n_1}) \otimes (\mathbf{R}^{n_2},\mathscr{B}_{\mathbf{R}}^{n_2},\mathscr{P}_2^{n_2}) = (\mathbf{R}^{n_1+n_2},\mathscr{B}_{\mathbf{R}}^{n_1+n_2},\mathscr{P}_1^{n_1} \otimes \mathscr{P}_2^{n_2})$$

其中

$$\mathscr{P}_1 = \{N(\mu_1,\sigma^2):(\mu_1,\sigma) \in \mathbf{R} \times \mathbf{R}^+\}$$

$$\mathscr{P}_2 = \{N(\mu_2,\sigma^2):(\mu_2,\sigma) \in \mathbf{R} \times \mathbf{R}^+\}$$

($\mathbf{R}^{n_1}$,$\mathscr{B}_{\mathbf{R}}^{n_1}$,$\mathscr{P}_1^{n_1}$)是第一个正态总体的重复抽样结构,而($\mathbf{R}^{n_2}$,$\mathscr{B}_{\mathbf{R}}^{n_2}$,$\mathscr{P}_1^{n_2}$)是第二个正态总体的重复抽样结构.

从定义 1.2 可以看出,从统计结构($\mathscr{X}$,$\mathscr{B}$,$\mathscr{P}$)抽取容量为 n 的样本和从重复抽样结构($\mathscr{X}^n$,$\mathscr{B}^n$,$\mathscr{P}^n$)抽取容量为 1 的样本所给出的信息是一样的.

定义 1.3 设($\mathscr{X}$,$\mathscr{B}$,$\mathscr{P}$)^n 为重复抽样结构,对每个样本$(X_1,\cdots,X_n) \in \mathscr{X}^n$,由

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{\{X_i \leqslant x\}}, \qquad \forall x \in \mathbf{R}$$

所确定的($\mathscr{X}$,$\mathscr{B}$)上的分布称为样本分布或经验分布. 对每一个样本观察值来说,$F_n(x)$是一个分布函数,称为样本分布函数或经验分布函数. 对每个固定的 $x \in \mathbf{R}$,$F_n(x)$又是样本 $X_1,\cdots,X_n$ 的一个函数,故 $F_n(x)$又是一个随机变量.

由于可把诸示性函数 $I_{\{X_i \leqslant x\}}, i=1,\cdots,n$ 看作是独立同分布,仅取 0 或 1 的随机变量,故有

$$\mathrm{E}F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}I_{\{X_i \leqslant x\}} = \frac{1}{n}\sum_{i=1}^{n} \mathrm{P}(X_i \leqslant x) = F(x)$$

$$\begin{aligned} \mathrm{Var}[F_n(x)] &= \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}I_{\{X_i \leqslant x\}} \\ &= \frac{1}{n}F(x)[1 - F(x)] \leqslant \frac{1}{4n} \leqslant \frac{1}{4} \end{aligned}$$

其中 $F(x)$ 为分布族 $\mathscr{P}$ 中某个 P 的分布函数，常称 $F(x)$ 为某总体的分布函数. 由大数定律可知，对任意 $\varepsilon>0$，总有

$$P(\,|F_n(x)-F(x)|\geqslant\varepsilon)\to 0\quad(n\to\infty)\tag{1.1}$$

这表明，只要 n 愈来愈大，样本的经验分布函数 $F_n(x)$ 可以愈来愈接近总体分布函数 $F(x)$，因此可以用 $F_n(x)$ 的各阶矩（如样本均值，样本方差，样本相关系数，样本协方差阵等）研究统计结构 $(\mathscr{X},\mathscr{B},\mathscr{P})$ 的某些特征. 这一想法在统计中是经常被使用的.

关于经验分布函数 $F_n(x)$，还有比(1.1)更强的结论，那就是如下的格里汶科定理.

定理 1.1（格里汶科） 对任意给定的正整数 n，设 $x_1,\cdots,x_n$ 是取自总体分布函数 $F(x)$ 的一个样本观察值，$F_n(x)$ 为其经验分布函数，记

$$D_n=\sup_{-\infty<x<\infty}|F_n(x)-F(x)|$$

则有

$$P(\lim_{n\to\infty}D_n=0)=1$$

这个定理的证明在很多教科书[5]上都可找到，这里不再叙述了. 容易看出，上述 D_n 可用来衡量 $F_n(x)$ 与 $F(x)$ 之间在所有的 x 值上最大差异程度. 该定理表明：在 n 无限大时，对于所有的 x 值，$F_n(x)$ 与 $F(x)$ 之差的绝对值是一致地愈来愈小，这个事件发生的概率为 1.

§1.1.3 可控结构

目前对统计结构研究最多，所获结果也较多的是可控结构，为了叙述可控结构，我们从测度的绝对连续性谈起.

定义 1.4 设 μ 与 v 是可测空间 $(\mathscr{X},\mathscr{B})$ 上的两个 σ 有限测度，假如对 $\mathscr{B}$ 中使 $\mu(N)=0$ 的某个集合 N，也使 $v(N)=0$，则称 v 对 μ 是绝对连续的，或者 v 被 μ 所控制，记为 $v\ll\mu$.

根据 Radon-Nikodym 定理，在 $v\ll\mu$ 条件下，一定存在这样一个 $\mathscr{B}$ 可测函数 $p(x)$（定义在 $\mathscr{X}$ 上），使得

$$v(B)=\int_B p(x)\mu(\mathrm{d}x),\qquad\forall B\in\mathscr{B}$$

此函数 $p(x)$ 称为 v 对 μ 的 Radom-Nikodym 导数，记为

$$p(x)=\frac{\mathrm{d}v}{\mathrm{d}\mu},\quad\text{a.s.}[\mu]$$

并且这个函数$p(x)$在a. s. $[\mu]$意义下是唯一的，这是a. s. $[\mu]$表示"除测度μ为零的集合外都成立"，常称"对μ几乎处处成立".

定义1.5 设$(\mathscr{X},\mathscr{B},\mathscr{P})$为一统计结构，若在可测空间$(\mathscr{X},\mathscr{B})$上存在这样一个$\sigma$有限测度$\mu$，使得$\mathscr{P}$中每一个概率分布$P$对$\mu$都是绝对连续的，即

$$P \ll \mu, \quad \forall P \in \mathscr{P}$$

则称该结构是可控的，相应的Radon-Nikodym导数$\mathrm{d}P/\mathrm{d}\mu$称为概率密度函数.

在数理统计中常用来作控制的σ有限测度是两种：计数测度和Lebesgue测度.

例1.3（计数测度） 设$\mathscr{X}=\mathbf{R}$，$\mathscr{B}$是直线上一切Borel集组成的σ代数，在$(\mathscr{X},\mathscr{B})$上定义如下测度

$$\mu(B) = B\text{中非负整数的个数}, \quad \forall B \in \mathscr{B}$$

容易验证，这样的测度是σ有限测度，并称为计数测度. 它可以控制任一个定义在非负整数集合$\mathbf{N}$（或其子集）上的概率分布族，其Radon-Nikodym导数就是通常的概率分布列. 如对Poisson分布族而言，任一个不含非负整数的Borel集A的计数测度$\mu(A)$为零，而在这样的集合上Poisson概率$P(A)$必为零，而对任一个Borel集B，Poisson概率$P(B)$可表示为

$$P(B) = \int_B \frac{\lambda^x}{x!}\mathrm{e}^{-\lambda}\mu(\mathrm{d}x) = \sum_{x\in B\cap \mathbf{N}} \frac{\lambda^x}{x!}\mathrm{e}^{-\lambda}$$

这里的x只能取非负整数. 所以Poisson分布对计数测度的概率密度函数

$$P_\lambda(x) = \begin{cases} \dfrac{\lambda^x}{x!}\mathrm{e}^{-\lambda}, & x = 0,1,\cdots \\ 0, & \text{其它} \end{cases}$$

类似地可对二项分布族，负二项分布族作出解释. 今后对离散分布所谈论的概率密度函数就是指该分布对计数测度的Radon-Nikodym导数.

例1.4（Lebesgue测度） 设$\mathscr{X}=\mathbf{R}$，$\mathscr{B}$是直线上的一切Borel集组成的σ代数，在$(\mathscr{X},\mathscr{B})$上以区间长度为基础定义的Lebesgue测度

$$\mu(B) = B\text{中不相交区间的长度之和或其极限}, \forall B \in \mathscr{B}$$

容易验证，Lebesgue测度是σ有限测度，它可以控制任一个定义在实数集$\mathbf{R}$上的连续分布，其Radon-Nikodym导数就是通常的概率密度函数$p(x)$，如对正态分布族而言，任一个Borel集B的概率总可表示为

$$P(B) = \int_B \frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-(x-u)^2/2\sigma^2}\mathrm{d}x$$