现代数学基础丛书 154

现代非参数统计

薛留根 著

科学出版社

现代数学基础丛书 154

现代非参数统计

薛留根 著

科学出版社
北京

内 容 简 介

本书阐述现代非参数统计的方法和理论. 本书在取材上侧重内容的科学性和应用性, 体现学术思想; 在写作上注重阐述方法论, 个别章节安排模拟计算和实例分析; 在结构上每章内容自成体系, 方便读者阅读. 本书的内容不仅为从事该领域的科研人员提供了尽可能多的资料, 也为实际应用者提供了一些数据分析的方法, 同时也为想全面了解现代非参数统计的读者提供参考读物.

本书可以作为高等院校统计学及其相关专业的学生的教学用书, 对高等院校和科研机构的研究人员、工程技术人员和研究生有参考价值.

图书在版编目(CIP)数据

现代非参数统计/薛留根著. —北京：科学出版社, 2015.1
(现代数学基础丛书; 154)
ISBN 978-7-03-042400-6
Ⅰ. ①现… Ⅱ. ①薛… Ⅲ. ①非参数统计 Ⅳ. ①O212.7
中国版本图书馆 CIP 数据核字(2014) 第 259077 号

责任编辑: 陈玉琢 / 责任校对: 鲁 素
责任印制: 赵德静 / 封面设计: 陈 敬

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

双青印刷厂 印刷
科学出版社发行 各地新华书店经销
*
2015 年 1 月第 一 版 开本: 720 × 1000 1/16
2015 年 1 月第一次印刷 印张: 25 1/2
字数: 491 000

定价: 128.00 元
(如有印装质量问题, 我社负责调换)

《现代数学基础丛书》序

对于数学研究与培养青年数学人才而言，书籍与期刊起着特殊重要的作用．许多成就卓越的数学家在青年时代都曾钻研或参考过一些优秀书籍，从中汲取营养，获得教益．

20 世纪 70 年代后期，我国的数学研究与数学书刊的出版由于“文化大革命”的浩劫已经破坏与中断了 10 余年，而在这期间国际上数学研究却在迅猛地发展着．1978 年以后，我国青年学子重新获得了学习、钻研与深造的机会．当时他们的参考书籍大多还是 50 年代甚至更早期的著述．据此，科学出版社陆续推出了多套数学丛书，其中《纯粹数学与应用数学专著》丛书与《现代数学基础丛书》更为突出，前者出版约 40 卷，后者则逾 80 卷．它们质量甚高，影响颇大，对我国数学研究、交流与人才培养发挥了显著效用．

《现代数学基础丛书》的宗旨是面向大学数学专业的高年级学生、研究生以及青年学者，针对一些重要的数学领域与研究方向，作较系统的介绍．既注意该领域的基础知识，又反映其新发展，力求深入浅出，简明扼要，注重创新．

近年来，数学在各门科学、高新技术、经济、管理等方面取得了更加广泛与深入的应用，还形成了一些交叉学科．我们希望这套丛书的内容由基础数学拓展到应用数学、计算数学以及数学交叉学科的各个领域．

这套丛书得到了许多数学家长期的大力支持，编辑人员也为其付出了艰辛的劳动．它获得了广大读者的喜爱．我们诚挚地希望大家更加关心与支持它的发展，使它越办越好，为我国数学研究与教育水平的进一步提高做出贡献．

杨　乐

2003 年 8 月

前　言

非参数统计的一个显著特点是它的使用面广，因为它讨论的模型中分布族没有通过有限个实参数去刻画，模型使用的范围更大. 因此，它作为统计学的一个重要分支，在经济、金融、生物、医学等领域有着广泛的应用. 非参数统计的另一个特点是大样本方法占重要位置. 可以说，绝大多数常用的非参数统计方法都是基于有关统计量的某种渐近性质. 因此，某些定理的论证很烦琐，读起来往往感到困难. 目前，非参数统计的方法和理论已渗透到许多学科，在实际中有越来越多的应用，深受人们重视. 人们需要有一本系统介绍现代非参数统计的专著，这就萌发了作者撰写本书的想法. 本书阐述了现代非参数统计的方法和理论，并对一些重要定理给出了证明. 书中丰富的内容一方面为从事该领域的研究人员提供了完善的参考资料，另一方面为实际应用工作者提供了现代统计方法. 此外，本书也为想全面了解现代非参数统计的读者提供了有价值的参考，为进一步进行非参数统计方面的研究奠定基础.

本书取材是根据作者三十多年来对该领域的研究成果及所积累的资料撰写而成，其中相当一部分内容是最新成果，反映了本学科的现代面貌. 同时，考虑到尽可能扩大读者面，本书在取材上作了精心安排，内容由浅入深，既有广度，又有深度. 考虑到非参数统计的特点，本书在写作上就如下三个方面做了努力：一是侧重介绍现代非参数统计的方法和理论，定理的证明尽可能简化；二是注重介绍现代非参数统计的最新成果；三是有选择地安排了一些模拟计算的内容，并安排了若干应用案例. 本书在语言叙述上尽量通俗易懂，便于读者阅读. 本书包含了现代非参数统计模型的一些内容，这样可以使读者读完本书就能够了解到非参数统计的研究前沿，为他们进入理论和应用研究打下良好基础；同时使他们掌握现代非参数统计的处理技术，并将这些技术应用到解决实际问题之中.

本书的出版得到了科学出版社陈玉琢编辑的鼓励和帮助；得到了国家自然科学基金 (11171012, 11331011)、高等学校博士学科点专项科研基金 (20121103110004)、北京市自然科学基金 (1142003)、北京市自然科学基金与北京市科学技术研究院联合资助项目 (L140003) 的资助；同时也得到了冯三营、田瑞琴、李万斌等同志的帮助，作者谨在此一并表示感谢.

薛留根

2014 年 6 月

符　号　表

iid	独立同分布
$\mathbf{R}^d$	d 维 Euclidean 空间
$\\|\cdot\\|$	Euclidean 模
A^{T}	向量或矩阵 A 的转置
$\mathrm{tr}(A)$	矩阵 A 的迹
$I(B)$	集合 B 的示性函数
$A \otimes B$	矩阵 A 与 B 的 Kronecker 乘积
$A^{\otimes 2}$	AA^{T}
$\mathrm{diag}\{a_1, \cdots, a_n\}$	由元素 $a_1, \cdots, a_n$ 组成的对角阵
$\xrightarrow{D}$	依分布收敛
$\xrightarrow{P}$	依概率收敛
a.s.(a.e.)	几乎处处
$\triangleq$	“定义为”或“记为”
$y = O(1)$	y 是有界变量
$y = o(1)$	y 是无穷小量
$\xi_n = o_p(\eta_n)$	对任一 $\varepsilon > 0$, 有 $P\{\\|\xi_n\\| \geqslant \varepsilon\\|\eta_n\\|\} \to 0$
$\xi_n = o_P(1)$	ξ_n 依概率收敛到 0
$\xi_n = O_P(1)$	ξ_n 依概率有界
$f'(t)$	函数 $f(t)$ 的一阶导数
$N(\mu, \Sigma)$	均值为 μ, 协方差阵为 Σ 的正态分布
χ_p^2	自由度为 p 的 χ^2 分布
$\chi_p^2(1-\alpha)$	自由度为 p 的 χ^2 分布的 $1-\alpha$ 分位数
$z_{1-\alpha/2}$	标准正态分布的 $1-\alpha/2$ 分位数
x_+	变量 x 的正部
$h_n \searrow 0$	h_n 单调下降趋于 0
$C_n \nearrow \infty$	C_n 单调上升趋于无穷大
$\log^+ x$	对数函数 $\log x$ 的正部
$\lfloor x \rfloor$	不大于 x 的最大整数
$\lceil x \rceil$	不小于 x 的最小整数

目　录

第1章　预 备 知 识

在本章中, 我们对本书用到的一些知识和已知结果给以阐述, 其中包括一些重要的概率不等式和主要的随机极限定理. 由于这些内容涉及面广, 难以在此全面地叙述. 我们的主要目的是将一些常用的结果不加证明地汇集起来, 以便于查阅. 了解这些结果的确切意义而不必涉及其证明细节, 就可以读懂本书的有关内容.

1.1　概率不等式

概率不等式在统计学中起着非常重要的作用, 它是研究统计大样本理论的极其得力的工具. 不夸张地说, 如果没有概率不等式, 我们所研究的某些大样本问题就很难有新进展. 本节我们仅就一些常用的不等式加以介绍, 而其证明不是主要的. 读者要想深入了解这些结果, 可以在有关文献中查找.

1.1.1　概率的指数型不等式

我们首先给出 Bernstein 不等式, 其证明可参看文献 Hoeffding(1963) 或林正炎和白志东 (2006).

定理 1.1.1 (Bernstein 不等式)　设随机变量 $X_1,\cdots,X_n$ 相互独立, 均值为 0, 且存在有限常数 b, 使得 $P\{|X_i|\leqslant b\}=1$, $i=1,\cdots,n$, 则对任何 $\varepsilon>0$, 有

$$P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}X_i\right|\geqslant\varepsilon\right\}\leqslant 2\exp\left\{-\frac{n\varepsilon^2}{2\sigma^2+b\varepsilon}\right\},$$

其中 $\sigma^2=\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\mathrm{var}(X_i)$.

由定理 1.1.1 立刻可得到下面的推论.

推论 1.1.1　设随机变量 Y 服从二项分布 $B(n,p)$, 则对任何 $\varepsilon>0$, 有

$$P\left\{\left|\frac{Y}{n}-p\right|\geqslant\varepsilon\right\}\leqslant 2\exp\left\{-\frac{n\varepsilon^2}{2p+\varepsilon}\right\}.$$

下面介绍由 Talagrand(1994, 1996) 给出的经验过程的指数型不等式, 该不等式是由 Vapnik-Červonenkis(VC) 类函数族所标识的经验过程的矩不等式来完成. 设 $(S,\mathcal{S})$ 是一个可测空间, $\mathcal{F}$ 为其上的一致有界的可测函数的全体. 如果类 $\mathcal{F}$ 是可分

的, 且存在正数 A 和 v, 使得对 $(S,\mathcal{S})$ 上的每个概率测度 P 和每个 $0<\tau<1$,

$$N(\mathcal{F}, L_2(P), \tau\|F\|_{L_2(P)}) \leqslant \left(\frac{A}{\tau}\right)^v,$$

其中 $\|F\|_{L_2(P)} = \sup\limits_{x\in L_2(P)} |F(x)|$, $N(T,d,\tau)$ 表示距离空间 (T,d) 的 τ 覆盖数, 即需要覆盖 T 的球的最小数目, 其中球以 T 为中心且半径不大于 τ, 则称 $\mathcal{F}$ 为一个有界的可测 VC 函数类. 在上述不等式中, d 是 $L_2(P)$ 距离. 我们将把 A 和 τ 看作不等式对所有 P 都成立的类 $\mathcal{F}$ 的 VC 特征集, 并设 $A\geqslant 3\sqrt{\mathrm{e}}$, $v\geqslant 1$. 记 $\|\varPsi\|_{\mathcal{F}} = \sup\limits_{f\in\mathcal{F}} |\varPsi(f)|$. 设 P 为 $(S,\mathcal{S})$ 上的任何概率测度, 又设 $\xi_i : S^N \mapsto S$, $i\in N$ 是坐标函数. 那么有下列结果, 它的证明可参看 Talagrand(1994, 1996).

定理 1.1.2　设 $\mathcal{F}$ 是可测且一致有界的 VC 函数族, 又设 σ^2 和 U 是任何正数, 使得 $\sigma^2 \geqslant \sup\limits_{f\in\mathcal{F}} \mathrm{var}_P(f)$, $U\geqslant \sup\limits_{f\in\mathcal{F}}\sup\limits_{x}|f(x)|$, $0<\sigma<U$. 如果

$$0<\sigma<U/2, \quad \sqrt{n}\sigma \geqslant U\sqrt{\log(U/\sigma)},$$

则存在仅依赖于 A 和 τ 的正的常数 L 和 C_1, 使得对任意 $\lambda\geqslant C_1$ 和满足

$$C_1\sqrt{n}\sigma\sqrt{\log(U/\sigma)} \leqslant \varepsilon \leqslant \lambda n\sigma^2/U$$

的 ε, 有

$$P\left\{\left\|\sum_{i=1}^n [f(\xi_i) - Ef(\xi_1)]\right\|_{\mathcal{F}} \geqslant \varepsilon\right\} \leqslant L\exp\left\{-\frac{\log(1+\lambda/(4L))}{L\lambda}\frac{\varepsilon^2}{n\sigma^2}\right\}. \tag{1.1.1}$$

特别地, 如果取

$$\varepsilon = C_2\sqrt{n}\sigma\sqrt{\log(U/\sigma)}, \quad C_2>C_1,$$

则

$$\begin{aligned} &P\left\{\left\|\sum_{i=1}^n [f(\xi_i) - Ef(\xi_1)]\right\|_{\mathcal{F}} \geqslant C_2\sqrt{n}\sigma\sqrt{\log(U/\sigma)}\right\} \\ &\leqslant L\exp\left\{-\frac{C_2\log(1+C_2/(4L))}{L}\log(U/\sigma)\right\}. \end{aligned} \tag{1.1.2}$$

事实上, 定理 1.1.2 在更弱的条件下成立: 对 $(S,\mathcal{S})$ 上的每个概率测度 P 和每个 $0<\tau<1$,

$$N(\mathcal{F}, L_2(P), \tau\|F\|_\infty) \leqslant \left(\frac{A}{\tau}\right)^v.$$

Chen 和 Zhao(1987) 给出了下列重要的不等式.

定理 1.1.3 设随机变量 $X_1, \cdots, X_n$ 具有共同的连续分布函数 $F(x)$, 其经验分布函数为 $F_n(x)$. $F(x)$ 和 $F_n(x)$ 对应的概率测度仍分别记为 F 和 F_n. 用 $\mathcal{A}$ 表示直线上某些区间构成的集合. 如果对某个 $b > 0$,

$$\sup_{A \in \mathcal{A}} F(A) \leqslant b \leqslant 1,$$

则存在正的常数 $c_i (i = 0, 1, \cdots, 4)$, 使得对任何 $\varepsilon > 0$, 当 $n/\log n > c_0/\varepsilon$ 时, 有

$$P\left\{\sup_{A \in \mathcal{A}} |F_n(A) - F(A)| \geqslant \varepsilon\right\} \leqslant c_1 \left(\frac{\sqrt{b}}{\varepsilon\sqrt{n}} + \frac{1}{b}\right) \exp\left\{-\frac{c_2 n \varepsilon^2}{b}\right\} + c_3 \exp\{-c_4 n \varepsilon\}.$$

Devroye 和 Wanger(1980) 在一维情形下给出了与定理 1.1.3 类似的结果. 将他们的结果推广到 d 维的情形, 有下述结果.

定理 1.1.4 设 $F_n(x)$ 和 $F(x)$ 分别为 d 维随机变量 $X_1, \cdots, X_n$ 的经验分布和理论分布, 其对应的概率测度仍记为 F_n 和 F. 用 $\mathcal{V}_a$ 表示 $\mathbf{R}^d$ 中半径为 $a > 0$ 的球所构成的集合. 如果对某个 $b > 0$,

$$\sup_{V \in \mathcal{V}_a} F(V) \leqslant b < \frac{1}{4},$$

则对任意的 $\varepsilon > 0$, 当 $n \geqslant \{1/b, 8b/\varepsilon^2\}$ 时, 有

$$P\left\{\sup_{V \in \mathcal{V}_a} |F_n(V) - F(V)| \geqslant \varepsilon\right\} \leqslant 4(2n)^{2d} \exp\left\{-\frac{n\varepsilon^2}{64b + 4\varepsilon}\right\} + 8n \exp\left\{-\frac{nb}{10}\right\}.$$

1.1.2 随机变量的概率不等式

定理 1.1.5 (Montgomery-Smith 最大值不等式) 利用定理 1.1.2 的记号. 对任意的 $\varepsilon > 0$,

$$P\left\{\max_{1 \leqslant k \leqslant n} \left\|\sum_{i=1}^{k} [f(\xi_i) - Ef(\xi_1)]\right\|_{\mathcal{F}} \geqslant \varepsilon\right\} \leqslant 9P\left\{\left\|\sum_{i=1}^{n} [f(\xi_i) - Ef(\xi_1)]\right\|_{\mathcal{F}} \geqslant \frac{\varepsilon}{30}\right\}.$$

本定理的证明可参看 Montgomery-Smith(1993).

定理 1.1.6(Lévy 型不等式) 设 $X_1, \cdots, X_n$ 为独立随机变量, 满足 $E(X_i) = 0$, $E(X_i^2) < \infty$, $i = 1, \cdots, n$. 记 $S_n = \sum_{i=1}^{n} X_i$, $b_n = \sum_{i=1}^{n} E(X_i^2)$, 则对任何 $x \in \mathbf{R}$,

$$P\left\{\max_{1 \leqslant k \leqslant n} S_k \geqslant x\right\} \leqslant 2P\left\{S_n \geqslant x - \sqrt{2b_n}\right\}.$$

本定理的证明可参看林正炎和白志东 (2006).

上面两个不等式经常被用来证明强大数定律.

1.1.3　用随机变量的矩估计概率的界

定理 1.1.7　设 X 为随机变量, $g(x) > 0$ 为偶函数且在 $[0,\infty)$ 非降. 如果 $E[g(X)] < \infty$, 则对任意的 $\varepsilon > 0$,

$$\frac{E[g(X)] - g(\varepsilon)}{\text{a.s.}\sup g(X)} \leqslant P\{|X| \geqslant \varepsilon\} \leqslant \frac{E[g(X)]}{g(\varepsilon)},$$

其中 $\text{a.s.}\sup g(X) = \inf\{t : P(g(X) \geqslant t) = 0\}$.

本定理的证明可参看林正炎和白志东 (2006).

上面不等式右端是著名的 Chebyshev-Markov 不等式的一般形式, 取不同的 $g(x)$ 可以得到不同的具体不等式.

(a) Chebyshev 不等式: 对任意的 $\varepsilon > 0$,

$$P\{|X - E(X)| \geqslant \varepsilon\} \leqslant \frac{\text{var}(X)}{\varepsilon^2}.$$

(b) Markov 不等式: 对任意的 $r > 0$ 和 $\varepsilon > 0$,

$$P\{|X| \geqslant \varepsilon\} \leqslant \frac{E(|X|^r)}{\varepsilon^r}.$$

定理 1.1.8 (Kolmogorov 不等式)　设 $X_1, \cdots, X_n$ 为独立随机变量, 满足 $E(X_i) = 0$, $E(X_i^2) < \infty$, $i = 1, \cdots, n$. 记 $S_n = \sum_{i=1}^{n} X_i$, 则对任意的 $\varepsilon > 0$,

$$P\left\{\max_{1\leqslant k\leqslant n} |S_k| \geqslant \varepsilon\right\} \leqslant \frac{\text{var}(S_n)}{\varepsilon^2}.$$

此外, 如果存在常数 $c > 0$ 使得 $|X_i| \leqslant c$, $1 \leqslant j \leqslant n$, 则

$$P\left\{\max_{1\leqslant k\leqslant n} |S_k| \geqslant \varepsilon\right\} \geqslant 1 - \frac{(\varepsilon + c)^2}{\text{var}(S_n)}.$$

本定理的证明可参看林正炎和白志东 (2006).

1.1.4　随机变量之和 (积) 的矩不等式

下面的不等式是由随机变量的数学期望 $E(\cdot)$ 给出的矩不等式. 如果将 $E(\cdot)$ 替换为条件数学期望 $E(\cdot|\mathcal{A})$, 相应的条件数学期望也成立. 我们仅列出一些常用的矩不等式, 其证明可以参看林正炎和白志东 (2006).

(1) Minkowski 不等式. 设 $X_1, \cdots, X_n$ 为任意随机变量, 而 $p \geqslant 1$, 则

$$\left[E\left|\sum_{i=1}^{n} X_i\right|^p\right]^{1/p} \leqslant \sum_{i=1}^{n} [E(|X_i|^p)]^{1/p}.$$