成　平　陈希孺　陈桂景　吴传义

参数估计

上海科学技术出版社

参 数 估 计

成 平　陈希孺
陈桂景　吴传义

上海科学技术出版社

参 数 估 计

成　平　陈希孺

陈桂景　吴传义

上海科学技术出版社出版

（上海瑞金二路 450 号）

新华书店上海发行所发行　上海东方印刷厂印刷

开本850×1168　1/32　印张 15　字数 396,000

1985 年 4 月第 1 版　1985 年 5 月第 1 次印刷

印数：1－4,000

统一书号：13119·1225　　定价：3.65 元

内 容 简 介

参数估计是数理统计学的一个重要分支，近几十年来发展很快．本书对参数估计的基本理论给予严格而比较系统的论述，内容包括基本概念、无偏估计、C-R 型不等式、Bayes 估计、大样本估计理论、极小化极大估计、同变估计和估计的容许性等．本书可作为数理统计专业中本课程的教材和参考书．供研究生、青年研究工作者、教师以及高年级学生阅读。

序　言

参数估计是一种基本的统计推断形式，也是数理统计学的一个重要分支．本世纪二十年代，英国著名统计学家 R. A. Fisher 奠定了参数估计的理论基础；战后近四十年来，这个分支有了很大的发展．这一方面是由于 A. Wald 的统计决策理论的影响，给参数估计理论提出了许多新的研究课题；另一方面是由于极限理论的发展，为估计的大样本理论的深入发展准备了条件．另外，Bayes 学派在战后的崛起，也是一个重要的因素．时至今日，参数估计仍不失为数理统计学中的一个活跃分支．

本书的目的，是对这个分支的基本理论给予严格而比较系统的论述．读者对象，主要是数理统计学中本分支和相近分支的研究生、青年研究工作者、教师以及本专业高年级学生．本书属于专著性质，有些方面的论述达到或接近当前研究的前沿．但是，由于这个分支内容很庞杂，因篇幅所限，有些性质过于专门或特殊的材料，就只好舍弃了，例如稳健估计的理论、自适应估计的理论以及有关容许性问题的大量深入的研究成果等．

本书第一、二章由陈桂景执笔，陈希孺作了一些修改；第三、五、六章由成平执笔，其中，第三章的写作得到李国英的协助；第七、八章由成平和吴传义共同执笔，陈桂景协助审订；第四章由陈希孺执笔．另外，陶波和吴启光给作者提供了不少帮助．上海科学技术出版社徐福生同志参与了最后定稿的工作，提供了一些有益的意见．作者谨在此表示衷心的感谢．

作　者

符 号 说 明 表

$\mathscr{X}$ 或 $(\mathscr{X}, \mathscr{B}_{\mathscr{X}})$ 样本空间

$\mathscr{P}$ 或 $\{P_\theta, \theta\in\Theta\}$ 分布族

Θ 参数空间

$A, (A, \mathscr{B}_A)$ 行动空间

$\mathscr{B}_{\mathscr{X}}$ 空间 $\mathscr{X}$ 中的 σ- 域

$(R^n, \mathscr{B}^n)$ n 维欧氏 Borel 可测空间

$L(\theta, a)$ 损失函数

$\delta(x)$ 非随机化判决函数

$\delta(D|x)$ 随机化判决函数

$R(\theta, \delta)$ 风险函数

$B_H(\delta)$ Bayes 风险

$H(d\theta)$ 先验分布

$H(d\theta|x)$ 后验分布

$\mathscr{H}$ 先验分布族

$R_H(\delta, x)$ 后验风险

$B(n, p)$ 二项分布

$\mathscr{P}(\lambda)$ Poisson 分布

$N(\mu, \sigma^2)$ 均值为 μ 方差为 σ^2 的正态分布

$R(0, \theta)$ 在 $[0, \theta]$ 上的均匀分布

$G(\alpha, p)$ Gamma 分布

$B(p, q)$ β- 分布

$N(\mu, A)$ 期望为向量 μ 协方差阵为 A 的多元正态分布

$\mathscr{L}(X)$ 随机变量 X 的分布

c. f. 特征函数

$M(t)$ 母函数

$I(\theta)$ 信息函数(矩阵)

$\xrightarrow{P}$ 依概率收敛

$\xrightarrow{r}$ 正则收敛

$\longrightarrow$a. s. 以概率1收敛

$\mu \ll \nu$ (σ-有限)测度μ关于(σ-有限)测度ν绝对连续

$\{\mu\} \ll \{\nu\}$ 测度族$\{\mu\}$被测度族$\{\nu\}$所控

$\{\mu\} = \{\nu\}$ 测度族$\{\mu\}$与测度族$\{\nu\}$等价

$dP/d\mu$ 概率测度P关于测度μ的Radom-Nikodgm导数

$P^T(\cdot)$ 统计量T导出的概率测度

$E[Y|\mathscr{T}]$ 在σ-域$\mathscr{T}$下Y的条件期望

$E[Y|X]$, $E[Y|x]$ 在X给定下Y的条件期望

$P(A|T)$, $P(A|t)$, $P(A|\mathscr{T})$ 条件概率

$P(A, t)$ (正则)条件概率分布

a. s. μ 关于测度μ几乎处处

$\mathbb{1}_n$ 元素全为1的n维列向量

A', A^τ 向量或矩阵A的转置

$I_A(\cdot)$ 集(事件)A的示性函数

$\hat{\theta}$ 参数θ的估计量

$\hat{g}$ 参数函数g的估计量

UMVUE 方差一致最小无偏估计

LSE 最小二乘估计

BLUE 一致最优线性无偏估计

MLE 极大似然估计

BANE 最好渐近正态估计

GBDF 广义Bayes判函数

CGBDF 比较广义Bayes判决函数

NGBDF 正规广义Bayes判决函数

iid. 独立同分布

$X_{(1)}, X_{(2)}, \cdots, X_{(n)}$ 样本$X_1, \cdots, X_n$的次序统计量

$\bar{X}$ 样本$X_1, \cdots, X_n$的均值

$\forall$ 任意

$\mathscr{G}$ 变换群

$\|a\|$ 向量a的欧氏模

$\delta_1 \prec \delta_2$ 判决函数 δ_1 优于 δ_2

$\mathscr{D}$ 判决函数类

$\mathscr{D}_T$ 仅依赖于统计量 T 的判决函数类

$\mathrm{Var}(X)$ 随机变量 X 的方差

$\mathrm{VAR}(X)$ 随机向量 X 的协方差阵

$\mathrm{cov}(X, Y)$ X 与 Y 的协方差

A^- 矩阵 A 的广义逆

$R(A)$, $\mathrm{Rank}(A)$ 矩阵 A 的秩

$\mathrm{tr}(A)$ 方阵 A 的迹

$B(\varphi{:}s)$ 函数 φ 的双边 Laplace 变换

$\mathscr{F}(\varphi{:}s)$ 函数 φ 的 Fourier 变换

$M(\varphi{:}s)$ 函数 φ 的 Mellin 变换

目　录

第一章 预备知识

§1 统计判决的基本概念

四十年代末期，A. Wald 提出了一种观点，把统计推断问题看成是人和自然的一种“博奕”. 这就是对以后的统计发展起了相当影响的统计判决理论. 从这个理论看来，本书的主题——参数估计，不过是一种具有某些特点的统计判决问题. 事实证明，这种观点对参数估计的发展，确实起了一定的推动作用. 不过，其作用主要在于它开拓了问题的面，而不在于提供了多少解决问题的工具.

为了本书今后的应用，在本节中我们对统计判决理论的若干基本概念作一个简要的介绍.

一、统计判决问题的三要素

为了估计一个未知参数，给出一定的估计量；为了检验某个假设，使用一定的检验函数，等等，它们都是相应的统计问题的解. 一般地说，一个统计问题的解是所谓“统计判决函数”. 为了定义统计判决函数这一重要概念，需要对构成一个统计判决问题的基本要素给以清楚的描述. 这些要素是：样本空间和分布族、行动空间以及损失函数. 以下逐点介绍.

1. 样本空间和分布族　样本是统计推断所依据的原始资料. 从理论上说，样本无非就是具有一定分布(至少是部分未知)的随机变量. 在以后，我们常常采用如下的配套记法：若以 X 记样本，则以 x 记 X 的具体观察值(在意义不致含混时，有时样本和样本值的记号不加区分)，而以 $\mathscr{X}$ 记 X 的一切可能值构成的集. 为在

数学上处理方便起见，有时 $\mathscr{X}$ 还包含一些 X 的值以外的点．例如，在一定的问题中，X 是只取非负值的随机变量，但为方便计，可以把 $\mathscr{X}$ 取为 R^1(此处及以后，总以 R^n 记 n 维欧氏空间)．

要定义 X 的概率分布，需要给出一个由 $\mathscr{X}$ 的某些子集所构成的 σ-域．我们常把这个 σ-域相应地记为 $\mathscr{B}_x$．那么，可测空间 $(\mathscr{X}, \mathscr{B}_x)$ 就称为(所讨论问题的)样本空间．在 σ-域 $\mathscr{B}_x$ 的意义明确时，我们也称 $\mathscr{X}$ 为样本空间．在绝大多数统计问题中，$\mathscr{X}$ 为有限维欧氏空间 R^n 或其一非空 Borel 子集，在这种情况下，$\mathscr{B}_x$ 总取为 $\mathscr{X}$ 的一切 Borel 子集构成的 σ-域，并称这样的样本空间 $(\mathscr{X}, \mathscr{B}_x)$ 为欧氏的．我们也常说：变量 X 的样本空间为 $(\mathscr{X}, \mathscr{B}_x)$．

如上所述，样本 X 有一定的概率分布，这个分布至少是部分未知的．观察 X 的目的，在于希望从样本观察值 x 得到有关 X 的分布的信息．这样，我们只能假定 X 的分布属于一定的分布族 $\mathscr{P}$．在一定的统计问题里，$\mathscr{P}$ 假定是已知的．自然要提出一个问题：当我们面临一个统计问题时，怎样给出 $\mathscr{P}$？对这个在应用上十分重要但在很大程度上不能由数学来回答的问题，我们只能一般地说：它有时可根据所论问题的专业理论知识来定，有时依赖于以往积累的经验资料，而有时不过是一种纯粹假设，其中考虑在数学上处理简单往往占重要地位．例如，经常采用“正态模型”．在各种问题里，以上三个方面都可能是一种重要的因素．

为了标明分布族 $\mathscr{P}$ 中某一确定的分布，可以对 $\mathscr{P}$ 中的每一分布赋予一个标号，例如 θ．这样，也可以把 $\mathscr{P}$ 写为 $\{P_\theta, \theta\in\Theta\}$，其中 Θ 为一切标号集，当 θ 跑遍 Θ 时，P_θ 跑遍 $\mathscr{P}$．习惯上称 θ 为分布参数．以后，我们总假定在参数值不同时，所对应的分布也不相同．Θ 作为参数 θ 的全部可能值的集，习惯上称为参数空间．通常，Θ 可以理解为这样的一种集，它包含参数 θ 的所有可能值，但不必与后者重合．本书讨论的绝大多数情况是：Θ 为有限维欧氏空间或其一非空 Borel 子集．有时，有必要在 Θ 中引进一个 σ-域 $\mathscr{B}_\Theta$．当 Θ 为欧氏空间或其一 Borel 子集时，$\mathscr{B}_\Theta$ 总取为由 Θ 的一切 Borel 子集构成的 σ-域，即参数(可测)空间 $(\Theta, \mathscr{B}_\Theta)$ 为欧

氏的. 最常见的情况是: 分布族 $\mathscr{P}=\{P_\theta,\ \theta\in\Theta\}$ 满足如下条件: 存在定义于 $\mathscr{B}_x$ 上的 σ-有限测度 μ, 使对一切的 $\theta\in\Theta$ 有 $\mathscr{P}_\theta\ll\mu$ (这时也记为 $\mathscr{P}\ll\mu$). 于是可把 $\mathscr{P}$ 写成 $\mathscr{P}=\{p_\theta d\mu,\ \theta\in\Theta\}$ 的形式, 其中 p_θ 为 P_θ 对 μ 的 Radom-Nikodym 导数: $p_\theta(x)=dP_\theta(x)/d\mu(x)$. 以后, 我们常见的情况是, μ 为 Lebesgue 测度, 或为某个可列集上的计数测度, 即对任一集 A, $\mu(A)=A$ 中的点数.

把分布族 $\mathscr{P}=\{P_\theta,\ \theta\in\Theta\}$ 与样本空间 $(\mathscr{X},\ \mathscr{B}_x)$ 配在一起, 构成一个三元组 $(\mathscr{X},\ \mathscr{B}_x,\ \mathscr{P})=(\mathscr{X},\ \mathscr{B}_x,\ P_\theta,\ \theta\in\Theta)$, 称其为(所讨论的问题的)概率空间. 有时也把它称作样本空间.

下面考察几个简例:

例 1.1 为估计一物体的重量 b, 对该物体作了 n 次独立称量, 结果记为 $X_1,\ \cdots,\ X_n$. 假定每次称量的误差服从正态分布 $N(0,\ \sigma^2)$, $\sigma^2>0$ 是未知的. 对本问题, 样本为 $X=(X_1,\ \cdots,\ X_n)$, 样本空间为 $(R^n,\ \mathscr{B}^n)$ ($\mathscr{B}^n$ 为 R^n 的一切 Borel 子集构成的 σ-域), 参数空间 Θ 为 $\{\theta:\theta=(b,\ \sigma),\ -\infty<b<\infty,\ \sigma>0\}$, 分布族 $\mathscr{P}=\{P_\theta,\ \theta\in\Theta\}$, 其中

$$dP_\theta(x)=(\sqrt{2\pi}\,\sigma)^{-n}\exp\left\{-\frac{1}{2\sigma^2}\sum_1^n(x_i-b)^2\right\}dx_1\cdots dx_n.$$

例 1.2 为估计一批产品(有 N 个)的废品率 θ, 从其中随机抽出 n 个, 以 X 记其中的废品个数. 当 $\frac{n}{N}\approx0$ 或抽样是有放回时, 可以认为 X 服从二项分布 $B(n,\ \theta)$. 就本例而言, 样本空间自然地取为 $\mathscr{X}=\{0,\ 1,\ \cdots,\ n\}$, 参数空间为 $\Theta=\{\theta:0\leqslant\theta\leqslant1\}$, 分布族可形式地表为

$$dP_\theta(x)=\binom{n}{x}\theta^x(1-\theta)^{n-x}d\mu(x),\ 0\leqslant\theta\leqslant1,$$

其中 μ 为 $\mathscr{X}$ 上的计数测度.

例 1.3 设在例 1.1 中, 假定各次试验的误差均值为 0 且独立同分布, 其它一无所知. 则样本空间仍与例 1.1 一样, 而分布族 $\mathscr{P}$ 则由这样的一些分布 P 构成: $P=F\times F\times\cdots\times F$ (n 重直积), F 是任一有均值的一维分布. 在此, 可以取 F 作为分布 P 的“标号”

即参数(而上面的 P 写为 P_F). 从而有

$$\Theta=\left\{F: F\text{ 为一维分布}, \int_{-\infty}^{\infty}|x|dF(x)<\infty\right\}.$$

本例与前两例不同处在于：不仅参数空间 Θ 不是欧氏的，而且分布族 $\mathscr{P}$ 中任一分布 P_F 对"参数值"F 的依赖关系也无法用简单的解析形式表达出来. 一般地，称前两例（以及与之类似的情况）中的分布族或者说所涉及的统计问题为*参数性的*,而例 1.3 则为*非参数性的*. 但是,这种区别只是一种比较笼统的概念,不可能也无必要划出一条泾渭分明的界限来.

2. 行动空间(也称为判决空间) 对一个统计问题的回答,依问题的性质而异，它可以是一个数(在点估计中)、一个区间(在区间估计中),或者一个决定——接受或拒绝(在假设检验中). 而统计判决问题是一种抽象化了的统计问题,关于它的回答,也就是关于形形色色的具体统计问题的回答的抽象——"判决". 因为一定的判决(如接受或否定某一假设)常导致采取一定的行动，所以也常把关于统计判决问题的回答称为一个"*行动*". 对一个给定的统计判决问题，采取何种行动，这与问题的性质、所获样本以及所采用的"优良性准则"(见后面)有关. 但是,对一定的问题,所可能采取的全部行动，是事先就可能知道的. 这个由全部可能行动构成的集 A 就称为*行动空间*. 有时，需要将 A 适当扩大一些，在不少情况下，有必要引进一个由 A 的某些子集所构成的 σ-域 $\mathscr{B}_A$，以形成*可测空间*$(A, \mathscr{B}_A)$. 当 A 为有限维欧氏空间或其一 Borel 子集(这是最常见的情况)时，总把 $\mathscr{B}_A$ 取为由 A 的一切 Borel 子集构成的 σ-域.

例 1.4 在例 1.1 中,若问题是作 b 的点估计，则行动空间可取为 $A=R^1$；若要作 b 的区间估计，则可取 $A=\{a: a=[a_1, a_2], a_1\leqslant a_2\}$. 若问题是同时估计 b 和 σ，则行动空间可取为 $A=\{a=(a_1, a_2): -\infty<a_1<\infty, a_2>0\}$.

例 1.5 设有 c 个正态总体 $N(a_i, \sigma^2)(i=1, \cdots, c)$. 要求根据对各总体的一些观察结果，对这 c 个总体的均值 $a_1, a_2, \cdots, a_c$

从小到大排定一个次序. 在此, 每个"行动"都是 1, 2, $\cdots$, c 的一个置换 $(i_1, \cdots, i_c)$, 它表示我们所作的判决是: a_{i_1} 最小, a_{i_2} 次之, $\cdots$, a_{i_c} 最大. 行动空间 A 由一切这样的置换所构成, 它一共包含 $c!$ 个点.

对假设检验问题而言, 一般可能的行动只有两个(接受, 拒绝), 因此行动空间特别简单.

3. 损失函数 面对一个统计判决问题, 采取一定的行动, 就会招致一定的后果(经济的或其它的). 例如, 在例 1.2 中, 问题是要决定是否接受该批产品, 显然, 当废品率 θ 小时, 对接受这批产品的后果是有利的; 否则, 是不利的.

统计判决理论的一个基本假设, 是认为这种后果最终都可以用数量的形式表达出来, 因此, 若从一定的基点起算, 我们就不妨把后果看成是一种"损失", 它总是非负的, 行动愈接近正确, 损失就愈小; 否则, 就愈大. 这就导致"损失函数"这个重要概念. 损失函数 $L(\theta, a)$ 是一个定义在 $\Theta\times A$ 上的非负函数. 当参数真值为 θ 而采取行动 a 时, 所遭受的损失为 $L(\theta, a)$.

例如, 在例 1.1 中, 若要作 b 的点估计, 则一个常用的损失函数是

$$L(\theta, a)=(b-a)^2 \quad (\theta=(b, \sigma)).$$

这就是著名的平方损失函数, 在本书所讨论的点估计理论中占有特殊重要的地位, 许多重要结果都是在这个损失函数下取得的. 更一般地, 可考虑

$$L(\theta, a)=c(\theta)w(b-a),$$

这里 $c(\theta)>0$, 而 $w(x)$ 为 x 的偶函数, 在 $x\geqslant 0$ 处是非降的. $w(x)$ 为凸函数的情况有相当的重要性, 例如绝对值损失函数, 其中 $w(x)=|x|$.

如果问题是要作 b 的区间估计, 则必须考虑到: 当使用区间 $a=[a_1, a_2]$ 时, 所遭受的损失应从两方面来衡量: 一方面是区间 $[a_1, a_2]$ 是否包含 b; 另一方面是区间的长度. 例如, 可取

$$L(\theta, a)=m[1-I_{[a_1, a_2]}(b)]+(a_2-a_1),$$

其中 $m>0$ 为常数. 或

$$L(\theta,\ a)=|a_1-b|+|a_2-b|$$

等等. 在假设检验问题中，由于可能行动只有两个，情况比较简单. 例如，可采用 0-1 损失函数(决定正确时损失为 0，否则为 1)，这相当于通常的 Neyman-Pearson 理论.

函数 L 需满足一定的可测性要求. 一般地说，在行动空间 A 中总是引进了 σ-域 $\mathscr{B}_A$. 这时要求对任意的 $\theta\in\Theta$，$L(\theta,\ a)$ 作为 a 的函数，为 $\mathscr{B}_A$ 可测. 如果在 Θ 中也引进了 σ-域 $\mathscr{B}_\Theta$，则 L 要求对 $\mathscr{B}_\Theta\times\mathscr{B}_A$ 可测.

判决函数理论的一个弱点在于：不仅假定一切行动的后果都可以数量化是不现实的，即使在这种数量化是合理的情况下，也往往缺乏足够的根据去选择一个合适的损失函数. 因此，损失函数的选择不能不带有相当的人为性，其中数学上的简单可行占有重要地位.

二、判决函数及其风险函数

1. 判决函数 沿用前面的记号. 我们曾提到，判决函数是统计判决问题的解. 所谓"解"，就是指明所要采取的"行动"，而这个行动应依据所得到的样本 x. 因此，所谓判决函数，就是指定义于样本空间 $\mathscr{X}$ 上而取值于 A 内的函数. 若选定了判决函数 δ，而得到样本 x，则所采取的行动由 $\delta(x)$ 给出. 对判决函数要求有一定的可测性. 具体地说，任一判决函数 δ 都必须是由样本空间 $(\mathscr{X},\ \mathscr{B}_x)$ 到行动空间 $(A,\ \mathscr{B}_A)$ 的可测变换.

上面所定义的那种判决函数通常叫做"非随机化的". 所谓"非随机"，是指只要样本 x 一经确定，则所要采取的行动 $\delta(x)$ 也唯一地规定，无任何随机性(在得出 x 以后无任何随机性). 更普遍一些，可以考虑下面情况：当得出样本 x 时，并不能唯一地规定所要采取的行动 a，而只能规定一个在 $\mathscr{B}_A$ 上的概率分布 $\delta(\cdot|x)$，其意义是：在得出样本 x 以后，对任何集 $D\in\mathscr{B}_A$，采取属于 D 的行动的概率为 $\delta(D|x)$. 我们可以设想，通过某种随机性的机制来