数学研究生暑期教学中心讲座

统计推断与决策

黎子良

南开大学出版社

统计 推断与决策

美国哥伦比亚大学
数理统计教授
黎子良讲授

王公恕（南开大学） 孙嘉阳（北京大学）整理

南开大学出版社

1987年

内 容 简 介

本书系统而又扼要地阐述了统计决策与统计推断的基本理论，介绍了它们的历史和在当代的新进展，注重阐明统计理论的思想以及概念间的内在联系。

本书内容翔实，深入浅出，主要包括统计决策理论、假设检验和估计引论，书末开列了大量参考文献以备读者进一步学习和研究之用。

本书读者对象为高等院校学习数理统计的研究生和高年级学生、理论和应用统计工作者，部分内容对数理统计的初学者也很有裨益。

统 计 推断与决策

黎 子 良

南开大学出版社出版
（天津八里台南开大学校内）
新华书店天津发行所发行
河北省河北邮电印刷厂印刷

1988年1月第1版　　1988年1月第1次印刷
开本：850×1168 1/32　印张：4.5　插页2
字数：97千　　印数：1—6000
ISBN7 310-00039-0/O·8　　定价：1.25 元

编 者 的 话

受原教育部委托，南开大学于1985年6月10日至7月19日在南开数学研究所举办了全国第二期“数学研究生暑期教学中心”，该中心由世界著名数学家、南开数学研究所所长陈省身教授主持，聘请了六位外国专家为来自全国的近200名数学研究生和青年教师主讲六门课程（每门课约32学时）和一个讲座，其中的《统计——推断与决策》一课由著名的数理统计专家、美国哥伦比亚大学数理统计教授黎子良（*Lai,T.L*）讲授（黎子良现任美国斯坦福大学统计系教授），这本书就是根据他的讲稿和课堂笔记整理而成的，草稿经过本人审阅、编排和修改。

数理统计的任务是研究如何用有效的方式收集、整理和分析受到随机性影响的数据，以对所考察的问题作出推断、预测，直至为采取决策和行动提供依据和建议。因此，无论在自然科学领域，社会、经济领域，还是在工业、农业、医药卫生等技术领域，数理统计都有着广泛的应用。在这方面，甚至难以开列一张较详细的清单，可以预见，在我国，随着教育水平的提高和计算机的日趋普及，数理统计的理论研究和应用的开发都必将以更快的步伐向深度和广度进军，我国理、工、农、医、财经、管理等各类院校，包括成人教育学校、电视大学都普遍开设了数理统计方面的课程，虽然国内已经出版了不少这方面的教材和专著，但以较小的篇幅提纲挈领而又系统地介绍数理统计基本理论、思想和方法的书并不多，为适应多层次的读者的需要，我们编辑出版了这本书。

本书共分四章。第一章绪论简要介绍了统计学发展的历史、

它所研究的一些课题，以及概率论方面的预备知识；第二章讲统计决策理论；第三章是假设检验；第四章是估计理论。全书的论述系统、简明、扼要，特别注重阐发统计理论中的思想和概念之间的联系，具有内容翔实、深入浅出的特点，书中介绍了数理统计学若干领域在当代的新进展和新成果，其中包括不少作者本人的工作。书末开列了大量参考文献，以备读者进一步学习和研究之用。

本书供高等院校学习数理统计的研究生和高年级学生，理论及应用统计工作者使用，书中关于统计理论的背景、发展历史和统计思想的论述对于初学者也很有裨益。

编　者

1987.5.

目　录

第一章　绪论……………………………………………（1）

§1　历史背景……………………………………………（1）

§2　统计决策论和推断理论的一些课题…………（2）

§3　分布、模型、噪声数据……………………（4）

§4　概率………………………………………………（5）

第二章　统计决策理论……………………………（10）

§1　统计决策问题的一般形式…………………（10）

§2　Bayes 法则和后验分布……………………（13）

§3　极小极大（Minimax）法则　………………（21）

§4　完全类定理………………………………………（28）

§5　可容许与不可容许的广义Bayes估计………（35）

§6　复合决策理论和经验Bayes方法……………（42）

§7　序贯决策问题……………………………………（50）

§8　不变决策法则……………………………………（63）

第三章　假设检验……………………………………（69）

§1　数据与模型的差异和显著性检验……………（69）

§2　Neyman-Pearson 理论、单调似然比和决策论的应用……………………………………（72）

§3　不变检验…………………………………………（78）

§4　一般线性假设……………………………………（83）

§5　非参数秩检验和渐近相对效率………………（85）

§6　序贯检验…………………………………………（88）

第四章　估计引论……………………………………………………(96)
§ 1　无偏估计、U统计量和相合性………………………(96)
§ 2　矩法和最小二乘法………………………………………(105)
§ 3　渐近效率和极大似然方法……………………………(109)
§ 4　稳健（Robust）估计论 ………………………………(119)
§ 5　序贯估计和递推估计……………………………………(123)
参考文献……………………………………………………………(135)

第一章 绪 论

§1 历史背景

早在罗马帝国时期就已经有了统计观念，那时人们用它数人口。在大约二百年前，随着概率论的发展及其在天文学等领域中的应用，统计学的一些基本理论也应运而生。早期作出重要贡献的有 Bayes (1763)、Laplace (1773, 1812)、Gauss (1816)、Fourier(1826)、Gavarret(1840)、Lexis(1875, 1877)等人。上世纪末，在F.Galton和Karl Pearson的带领下，统计学在英国有一段集中的发展时期。本世纪二十年代，R.A.Fisher 的工作奠定了今天统计推断理论的基础。他研究的课题主要是从生物学（如优生学、遗传学及其在农业上的应用等等）中提出来的。后来J.Neyman和E.S.Pearson建立了一整套关于假设检验的理论(1928, 1933, 1936)，他们的工作是统计决策论的先导，这套假设检验理论加上四十年代初由 Von Neumann和Morgenstern所发展的博奕论促使Wald 在四十年代末期提出了统计决策论。在第二次世界大战期间，Wald为了国防上的需要（如怎样使得军火的抽样检验更快等等）还发明了序贯分析方法，而序贯分析与统计决策论一起又引起了动态规划（dynamic programming）在统计学和运筹学中的发展，到五十年代初期已经有了一套基本的统计决策方法和统计推断理论，在过去的三十年中，这套方法和理论更是取得了不少重要的进展。

当今的八十年代常被称为信息时代，我们已经有了高效能地

收集资料（信息）的技术，因此如何处理与综合所得到的大量资料就成了十分重要的问题。收集资料，对于在科技和商业中的实际应用来说，还只是第一步，根本的问题还在于如何利用这些资料去作出正确的决策。所以，在应用上以至理论上，统计决策论和推断方法都面临着新的挑战。统计学中的很多方法已经成为许多其它应用学科（如自动控制、运筹学、通讯工程、人工智能等等）的基本的数学工具。

§2 统计决策论和推断理论的一些课题

统计决策论的出发点是信息（information），终点是决策（decision）。其基本问题是怎样有效地利用信息去得到最优（或最低限度，合理）的决策，在一些带有序贯反馈的问题中，一个阶段的决策行动还会影响到下一阶段的信息内容。

我们用下面的框图来表达统计决策论的基本要素：

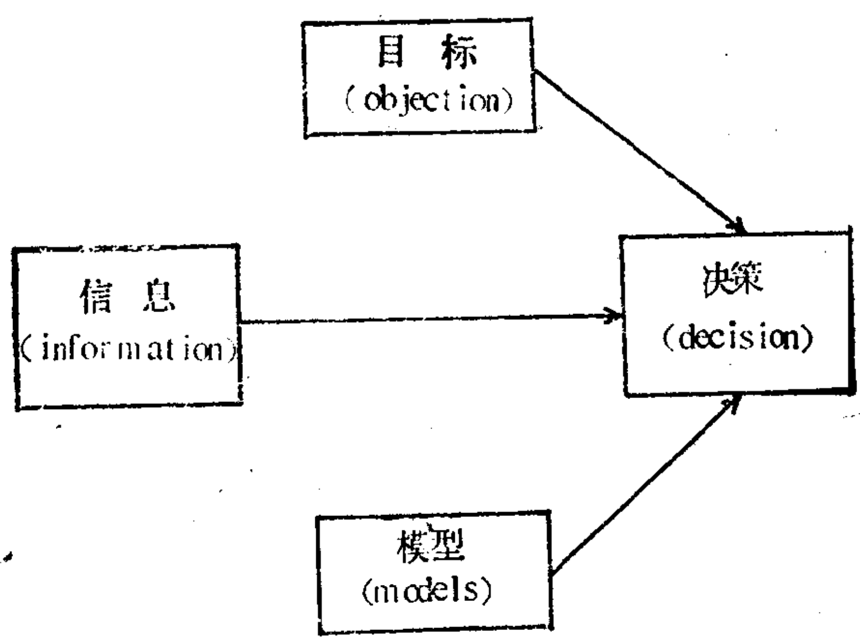

图 1.1

在图1.1中，除了“信息”和“决策”两个方框以外，还有“目标”和“模型”两个方框，“目标”表示决策的最终目的。在统计学中，人们常用损失函数（loss function）来描述决策问题的目标，也就是说，用损失函数作为“目标”的数学表示（见第二章），在其它的科技应用问题中，人们还常常用效用函数(utility function）、目标函数（objective function）、收益函数(gain function）等等作为“目标”的数学表示。

我们将在本章§3中讨论“模型”方框。

关于“信息”，有这样一些统计学的以及与计算机科学有关的课题：试验设计(experimental design)、抽样方法(sampling methods）、信号处理（signal processing）、数据分析(data analysis）、统计计算与作图（statistic computing and graphics）、数据管理（data base management）。

关于“决策”的一些经典课题有：

1. 推断：参数推断、非参数推断、假设检验、点估计和区间估计。

2. 模型的拟合与诊断(diagnostics)：回归模型（线性、非线性、非参数等)、多元分析(因子分析、方差分析、聚类分析等)、时间序列模型、系统辨识、概率分布拟合数据与分布密度的估计、回归诊断（Regression Diagnostics）等等。

3. 预测与控制：时间序列模型的预测、过程控制、随机控制、质量控制、抽样调查、可靠性理论等等。

非常规的信息，由于其性质不同于通常的信息，故往往导致新的决策方法。例如，序贯信息在工程和医药实验中的应用导致了序贯分析的研究；最近，如何利用缺失数据(censored data，例如在医药实验和可靠性检验等问题中）进行推断成了统计学的热门课题（Kaplan和Meier，1958；Cox，1972，1976；Kalbsfleisch和Prenctice，1980；Lawless，1982）。

§3 分布、模型、噪声数据

图1.1所表达的实际上还包括统计学以外的很多类其它学科的问题，统计学的特殊之处在于它还有一个分布问题。在统计学中，“信息”中有一个基本概念——“分布”，这可以归结为以下几个方面：

1. 总体与样本：在古典统计中就有了抽样的概念，即有一个我们所看不见的总体，我们不能获得这个总体的全部信息，我们所能做的，只是从这个总体中抽取样本，也就是说，我们能够观察到的只是样本而不是总体。

2. 带有误差的测量(噪声数据)：在实际问题中，如在工程问题中，我们看到的都是受了干扰的信号（如信号之间的干扰，外界因素的干扰等等），所以我们得到的数据是所谓噪声数据，而要得到样本的分布，就还需要测量干扰(误差)。当干扰太大时，我们所能作出的最好的推断也不一定好，当信噪比比较大时，情况可能好一些，问题在于要评价我们的决策推断的优劣，即要判定样本的变异性。

3. 统计量的抽样分布。

4. Bayes推断中的后验分布。

5. 预测与控制问题中的条件分布。

图1.1中的“模型”方框可以分成以下几方面讨论：

1. 物理模型（physical models）从科学原理中所得出的模型，如用数学方程式表示的物理定律等等。

2. 经验模型（empirical models）这时没有什么可资借鉴的基本原理，全凭我们的观察，从所看到的现象中总结出模型，如经验曲线、经验公式等等。

3. 随机模型(stochastic models)例如，根据物理定律，

一个物体的位置 x 与它运动的速度 $\dot{x}$、运动的时间 t 以及对它施加的控制 u 有关，即有

$$x=f(\dot{x},\ t,\ u)$$

但实际上还有许许多多不确定的因素影响着物体的位置，所以我们实际观察到的物体的位置与随机干扰 ε 有关，即

$$x=f(\dot{x},\ t,\ u)+\varepsilon$$

这就是一个随机模型。

4. 模型类（class of possible models）常有这样的情况：很多模型在一定的意义下都能与观察值相拟合。究竟从其中选取哪个模型呢？这就要根据我们的最终目的或者模型的简洁（parsimony）程度来决定了。

§4 概　　率

怎样描述一个分布呢？如何用数学表达它呢？最自然的就是用概率。下面，我们先复习一下有关的概率知识。

1. 随机变量、概率分布。

前面，我们曾谈到带误差的测量，由于测量值是有分布的，所以可以把它看成一个随机变量，而把观测到的数据看作该随机变量所取的值。从这个意义上说，概率分布与随机变量是分不开的，测量时，我们就是在观察分布，观测到的数据越多，对分布的估计就越精确，为形象起见，人们常常对数据进行整理，然后把它们作成直方图，从这种图中，可以看出分布的概貌。

分布还有一些数字特征，如均值、方差、中位数等等。

2. 古典极限定理。

我认为，在所有统计方法中，从哲学的观点讲，最终的基本点就是怎样平衡噪声或样本波动的问题。也就是说，怎样从杂乱无章的数据中剔除噪声的影响或样本波动的问题，大数定律很好

地说明了这一基本点。

（1）大数定律 设随机变量X_1，X_2，…i.i.d(independent identically distribution)，$EX_1=\mu<\infty$，则有

弱大数定律：

$$\overline{X}_n \triangleq \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} EX_1=\mu, \qquad (n\to\infty)$$

即对任何$\varepsilon>0$，有

$$P\{|\overline{X}_n-\mu|\leqslant\varepsilon\}\to 1 \qquad (n\to\infty)$$

强大数定律：

$$\lim_{n\to\infty}\overline{X}_n=\mu \qquad \text{a.s.}$$

即

$$P\left\{\lim_{n\to\infty}\overline{X}_n=\mu\right\}=1$$

对经验分布函数有：对任何固定的x，有

$$F_n(x)\triangleq\frac{1}{n}\sum_{i=1}^{n} I_{\{X_i\leqslant x\}} \xrightarrow{\text{a.s.}} EI_{\{X_1\leqslant x\}}\triangleq F(x) \quad (n\to\infty)$$

其中I_A表示A的示性函数，$F(x)$为X_1的分布函数。

还有更强的结果（Glivenko-Cantelli定理）：

$$P\{\sup_x|F_n(x)-F(x)|\xrightarrow[n\to\infty]{}0\}=1$$

对于在抽象空间（如巴拿赫空间）上取值的随机变量也有相应的大数定律。此外，若Φ是连续函数，则有

$$\Phi(F_n)\xrightarrow{\text{a.s.}}\Phi(F) \quad (n\to\infty)$$

对于平稳序列有遍历性定理。

（2）中心极限定理 设随机变量X_1，X_2，…i.i.d，$EX_1=\mu$，$\mathrm{Var}X_1=\sigma^2$，则对充分大的n，有

$$\overline{X}_n \approx N\left(\mu,\ \frac{\sigma^2}{n}\right),$$

即

$$Z_n \triangleq \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{\mathscr{D}} N(0,1)$$

这里$\mathscr{D}$表示依分布收敛，所以上式意为：对任何 $a \leqslant b$，

$$P\{a \leqslant Z_n \leqslant b\} \longrightarrow P\{a \leqslant N(0,1) \leqslant b\} \qquad (n \to \infty)$$

在统计中，还常用到下述关于F_n的中心极限定理：设 $F(x)$ 是连续的分布函数，$F_n(x)$为相应的经验分布函数，则

$$\sqrt{n}\,[F_n(x) - F(x)] \xrightarrow{\mathscr{D}} [0,1]\text{上的布朗桥}$$

（3）重对数定理　设随机变量 X_1，X_2，…i.i.d，$EX_1=\mu$，$VarX_1=\sigma^2$，则

$$P\left\{\overline{\lim_{n\to\infty}}\left|\frac{S_n - n\mu}{\sqrt{2n\log\log n}}\right| = \sigma\right\} = 1,$$

其中$S_n = \sum\limits_{i=1}^{n} X_i$

关于F_n的重对数定理，请参看 Finkelstein（1971）的书。

对于取值于巴拿赫空间的随机变量也有相应的结果。有兴趣的读者请参看Kuelbs（1977）的文章。

3. 概率测度族（参看Lehmann（1959）书中的第二章）。

我们常用（$\mathscr{X}$，$\mathscr{F}$，P）表示一个概率空间，其中$\mathscr{X}$是样本空间，$\mathscr{F}$是$\mathscr{X}$上的σ-代数，P是可测空间（$\mathscr{X}$，$\mathscr{F}$）上的概率测度，与概率论不同，在统计问题中，我们不知道具体的P，知道的是一个概率测度族（P_θ，$\theta\in\Theta$），其中Θ叫做参数空间。我们只知道样本X是（$\mathscr{X}$，$\mathscr{F}$）上的随机变量，其分布属于分布族（P_θ，$\theta\in\Theta$），而不知道θ的具体值。通常，统计问题正是要根据样本$\mathbf{X}$，对未知参数θ作为某种推断。

例1.1 设X_1, …, X_n为i.i.d随机变量，$X_1 \sim N(\theta,1)$，其中θ未知，则$\Theta=(-\infty, +\infty)$，$P_\theta$为$n$个$N(\theta,1)$的乘积测度。

例1.2 设X_1, …, X_n i.i.d，X_1的分布F未知，则$\Theta=\mathscr{F} \triangleq \{F: F$是分布函数$\}$。

例1.3 指数分布族，如果P_θ关于σ有限测度μ绝对连续（即$P_\theta \ll \mu$，对一切$\theta \in \Theta$），则P_θ关于μ的Radon-Nikodym导数存在，记为$p_\theta=\dfrac{dP_\theta}{d\mu}$，我们把$p_\theta$叫做$P_\theta$关于$\mu$的密度函数。

若随机变量X的密度函数形如

$$p_\theta(x)=c(\theta)h(x)\exp\left\{\sum_{i=1}^{k} Q_i(\theta)T_i(x)\right\}, \quad \theta \in \Theta.$$

其中$c(\theta)$，$Q_1(\theta)$，…，$Q_K(\theta)$是Θ上的有限函数，$h(x)$为定义于$\mathscr{X}$上的非负$\mathscr{F}$可测函数，$T_1(x)$，…，$T_K(x)$是$\mathscr{X}$上的$\mathscr{F}$可测函数，则我们把分布族$\{P_\theta;\ \theta\in\Theta\}$叫做$K$维指数分布族。

若定义$\widetilde{\Theta}=\{(Q_1(\theta),\cdots,Q_k(\theta)):\theta\in\Theta\}$，则$\widetilde{\Theta}$中的每一点对应指数分布族中的一个分布，并且，为了方便起见，常把$h(x)$"吸收"到μ中，于是，指数分布族的密度函数就可以写成

$$p_\theta(x)=c(\theta)\exp\left\{\sum_{i=1}^{k} \theta_i T_i(x)\right\},$$

其中$(\theta_1, \cdots, \theta_k)\in\widetilde{\Theta}$，我们把它叫做自然形式的指数族，把$\widetilde{\Theta}$叫做自然参数空间。

将$\int p_\theta(x)d\mu(x)=1$的两边对$\theta_j$求导，便可以得到

$$ET_j(X)=-\frac{\partial}{\partial\theta_j}\log c(\theta),$$

$$\mathrm{Cov}(T_i,T_j)=-\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log c(\theta)$$

还可以把指数分布族写成更简单的形式：

$$p_{\theta}(x)=e^{\theta x-\psi(\theta)} \quad （1 维），$$

$$p_{\underset{\sim}{\theta}}(\underset{\sim}{x})=e^{\underset{\sim}{\theta}'\underset{\sim}{x}-\psi(\underset{\sim}{\theta})}$$

$$=e^{\sum_{i=1}^{k}\theta_i x_i-\psi(\theta_1,\cdots,\theta_k)} \quad （k 维），$$

其中$\psi'(\theta)=E_{\theta}X$（1 维），$\nabla\psi(\underset{\sim}{\theta})=E_{\underset{\sim}{\theta}}\underset{\sim}{X}$（$k$维）。

第二章　统计决策理论

§1 统计决策问题的一般形式

1．统计决策问题的要素

（1）三个可测空间：样本空间$(\mathscr{X}, \mathscr{B}_1)$，其中$\mathscr{X}$是所有可能的数据值的集合；参数空间$(\Theta, \mathscr{B}_2)$，其中$\Theta$是所有可能的模型的集合；行动空间$(\mathscr{A}, \mathscr{B}_3)$，其中$\mathscr{A}$是所有可能的行动(action)的集合，$\mathscr{B}_1$，$\mathscr{B}_2$，$\mathscr{B}_3$分别为$\mathscr{X}$，$\Theta$，$\mathscr{A}$上的$\sigma$-代数。

（2）样本空间$(\mathscr{X}, \mathscr{B}_1)$上的概率测度族$\mathscr{P}=\{P_\theta, \theta\in\Theta\}$，其中$P_\theta$满足：对任意的$B\in\mathscr{B}_1$，$P_\theta(B)$是$\theta$的可测函数。

（3）损失函数L，它是$\Theta\times\mathscr{A}\to(0, +\infty)$的可测(关于$\mathscr{B}_1\times\mathscr{B}_2$)函数，$L(\theta, a)$表示参数真值为$\theta$时，统计学家采取行动$a$所蒙受的损失。

通常，我们用$(\mathscr{X}, \Theta, \mathscr{A}, P_\theta, L)$表示一个统计决策问题。

顺便提一下，在有些领域里，人们说的不是“损失”(loss)，而是“收益”(gain)、“报酬”(reward)或“效用”(utiiity)，当然，人们总是希望采取使收益(＝负损失)达到最大的决策。

2．统计决策问题的解

（1）所谓纯(即非随机化)决策法则就是一个可测函数d：$\mathscr{X}$