大學數學叢書⑧
主編／項武義・莫宗堅・康明昌

黎子良・著

統計推論與決策

黎子良著

統　計

推　論　與　決　策

大學數學叢書⑧

主　編

項武義（美國柏克萊加州大學教授）

莫宗堅（美國普渡大學教授）

康明昌（國立臺灣大學教授）

声

对书中信

和国政府关于

容词句一律不

北京国防

82.05.1458

定價：新臺幣 280 元

著　者 黎　子　良
發行人 劉　國　瑞

出版者 聯經出版事業公司
臺北市 忠孝東路 四段 555 號
電　話：7627429・3620137
郵撥電話：6 4 1 8 6 6 2
郵政劃撥帳戶第 0100559-3 號
印刷者：永　裕　印　刷　廠

行政院新聞局出版事業登記證局版臺業字第0130號

ISBN 957-08-0881-0（精裝）

31005-08

大學數學叢書序

聯經出版公司大學數學叢書創刊，約我寫幾句話，很高興有機會談一談我對中國數學的展望。

中國數學有長久和光榮的歷史；從“九章算經”到劉徽、祖冲之的割圓勾股，從“大衍求一”到天元四元，都是人民智慧的結晶。唐代算學設於學官，長期訓練學生。唐朝是中國歷史上最偉大的一個朝代。當時的中國是全世界最強盛的國家，其設施令人景仰。

但是中國數學沒有產生像“歐幾里得”的“幾何原本”這樣偉大的著作。歐幾里得根據一組公理，用邏輯推理導出平面和空間的基本性質。希臘的科學和近代的科學是人類文化史的奇蹟。中國以往的政治社會和文化背景，不能產生這樣的科學!

十六、七世紀後世界數學突飛猛進。基本的現象是數學的範圍擴大了，而這種新數學符合於其他科學的發展，獲得應用。因爲擴大範圍，就需抽象化。例如，從一本書、兩本書到1，2，3，從1 2 3到x y z，都是抽象化。數學的基礎是純粹數學。純粹數學引進基本的概念，將繁化簡，導入進一步的發展。中國數學不如西洋的一個主要原因是過分注意應用。這種短淺的目光，現在還很普遍。

中國的新式大學大多有數學系。到了三十年代有些可喜的發展：南開大學的姜立夫先生首先注意到圖書設備的重要並培養學生；清華大學從其雄厚的財力，在熊慶來、孫光遠、楊武之先生的領導下，培養一批優秀的青年，包括華羅庚、許寶騄、陳省身等；浙江大學在陳建功、蘇步青先生的領導下，訓練學生從事研究工作；其他許多學校也都有蓬勃的朝氣。西南聯合大學，集北大、清華、南開三校的精英，抗戰八年，人才甚盛。

第二次世界大戰後的一個現象是大批中國學生去美國受教育而發展成名。舉國際數學會議爲例：該會每四年舉行一次。有菲爾滋獎，獎勵突出的研究工作。也有約十六個的綜合演講，結論各方面的最近進展，邀講者必爲各方面研究的領導人物。中國數學家曾得菲爾滋獎，也曾受邀作綜合演講前後已有五次。此外作專題演講者甚多。中國數學已經成熟了！

中國現在有大批的年輕的數學工作者，在博士階段前後。人數空前，優秀者甚多。中國將成爲數學大國，是擋也擋不住的。

我的朋友，法國大數學家魏耳（André Weil）喜歡說，二十年後數學家需要學中文。此說已有端倪，大家已經知道：中國人的姓不能確定這個人，此張不是彼張，此王不是彼王。數學家已開始學中國人名了。

中國數學正在經歷史上未有的變化。如果大家努力，二十年後或者每一數學家的書架上都將有幾本這個叢書。

陳省身

1987年12月於加州

序

在1985年夏天，我在南開大學「數學研究生暑期教學中心」主講統計推論與決策論，本書的初稿是根據我的講稿和課堂筆記編成，這門課程主要是給數學大學部的高年級學生和剛剛進入研究所的學生關於統計推論和決策論一個簡明扼要的介紹，讓他們可以在短短一個月每週八個小時的課程內，能夠由淺入深，從這方面的基本概念和方法到一些最近研究的領域和成果，有一些初步認識，以便他們能在感到有興趣的課題上作進一步的學習和研究。

我在此特別感謝南開大學的王公恕教授和密西根大學的孫嘉陽教授幫我整理課堂筆記和編寫這本書的初稿，中央大學的張憶壽和熊昭教授以及他們的研究生劉正夫，黃連成，陳麗如幫我修正初稿和謄繕複稿，普渡大學的莫宗堅教授和臺灣大學康明昌教授的鼓勵和寶貴意見。

黎子良

謹識於史丹福大學統計系

1991年12月10日

目次

第一章　緒論

§1　歷史背景

早在羅馬帝國時期就已經有了統計觀念，那時人們用它數人口。在大約二百年前，隨著機率論的發展及其在天文學等領域中的應用，統計學的一些基本理論也應運而生。早期作出重要貢獻的有 Bayes (1763)、Laplace (1773, 1812)、Gauss (1816)、Fourier (1826)、Gavarret (1840)、Lexis (1875, 1877) 等人。上世紀末，在 F. Galton 和 Karl Pearson 的帶領下，統計學在英國有一段集中的發展時期。本世紀二十年代，R. A. Fisher 的工作奠定了今天統計推斷理論的基礎。他研究的課題主要是從生物學（如優生學、遺傳學及其在農業上的應用等等）中提出來的。後來 J. Neyman 和 E. S. Pearson 建立了一整套關於假設檢驗的理論 (1928, 1933, 1936)，他們的工作是統計決策論的先導，這套假設檢驗理論加上四十年代初由 Von Neumann 和 Morgenstern 所發展的博奕論促使 Wald 在四十年代末期提出了統計決策論。在第二次世界大戰期間，Wald 爲了國防上的需要（如怎樣使得軍火的抽樣檢驗更快等等）還發明了序貫分析方法，而序貫分析與統計決策論一起又引起了動態規劃 (dynamic programming) 在統計學和運籌學中的發展，到五十年代初期已經有了一套基本的統計決策方法和統計推斷理論，在過去的三十年中，這套方法和理論更是取得了不少重要的進展。

在當今大家所稱的信息時代裏，我們已經有了高效能地收集資料（信息）的技術，因此如何處理與綜合所得到的大量資料就成了十分重要的問題。收集資料，對於在科技和商業中的實際應用來說，還只是第一步，根本的問題還在於如何利用這些資料去作出正確的決策。所以，在應用上以至理論上，統計決策論和推論方法都面臨著新的挑戰。統計學中的很多方法已經成爲許多其它應用學科（如自動控制、運籌學、通訊工程、人工智能等等）的基本的數學工具。

§2　統計決策論和推斷理論的一些課題

統計決策論的出發點是信息(information)，終點是決策 (decision)。其基本問題是怎樣有效地利用信息去得到最優（或最低限度，合理）的決策，在一些帶有序貫反饋的問題中，一個階段的決策行動還會影響到下一階段的信息內容。

我們用下面的框圖來表達統計決策論的基本要素:

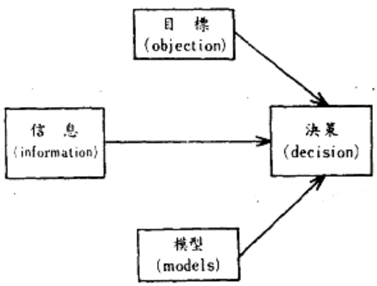

圖 1.1

在圖 1.1 中，除了“信息”和“決策”兩個方框以外，還有“目標”和“模型”兩個方框，“目標”表示決策的最終目的。在統計學中，人們常用損失函數 (loss function) 來描述決策問題的目標，也就是說，用損失函數作爲“目標”的數學表示（見第二章），在其它的科技應用問題

中，人們還常常用效用函數 (utility function)、目標函數 (objective function)、收益函數 (gain function)等等作爲“目標”的數學表示。

我們將在本章§3中討論“模型”方框。

關於“信息”，有這樣一些統計學的以及與計算機科學有關的課題：試驗設計 (experimental design)、抽樣方法 (sampling methods)、信號處理 (signal processing)、數據分析 (data analysis)、統計計算與作圖 (statistical computing and graphics)、數據管理(database management)。

關於“決策”的一些經典課題有：

1.推論：參數推論、非參數推論、假設檢驗、點估計和區間估計。

2.模型的擬合與診斷 (diagnostics)：回歸模型（線性、非線性、非參數等）、多元分析（因子分析、方差分析、聚類分析等）、時間序列模型、系統辨識、機率分布擬合數據與分布密度的估計、回歸診斷 (regression diagnostics) 等等。

3.預測與控到：時間序列模型的預測、過程控制、隨機控制、質量控制、抽樣調查、可靠性理論等等。

非常規的信息，由於其性質不同於通常的信息，故往往導致新的決策方法。例如，序貫信息在工程和醫藥實驗中的應用導致了序貫分析的研究；最近，如何利用缺失數據 (censored data，例如在醫藥實驗和可靠性檢驗等問題中) 進行推論成了統計學的熱門課題(Kaplan 和 Meier, 1958; Cox, 1972, 1975; Kalbfleisch 和 Prentice, 1980; Lawless, 1982)。

§3 分布、模型、噪聲數據

圖1.1 所表達的實際上還包括統計學以外的很多類其它學科的問題，統計學的特殊之處在於它還有一個分布問題。在統計學中，“信息”中有一個基本概念——“分布”，這可以歸結爲以下幾個方面：

1.總體與樣本：在古典統計中就有了抽樣的概念，即有一個我們所看不見的總體，我們不能獲得這個總體的全部信息，我們所能做的，只是從

這個總體中抽取樣本， 也就是說， 我們能够觀察到的 只是樣本而不是總體。

2.帶有誤差的測量（噪聲數據）：在實際問題中，如在工程問題中，我們看到的都是受了干擾的信號（如信號之間的干擾，外界因素的干擾等等），所以我們得到的數據是所謂噪聲數據，而要得到樣本的分布，就還需要測量干擾（誤差）。當干擾太大時，我們所能作出的最好的推論也不一定好，當信噪比比較大時，情況可能好一些，問題在於要評價我們的決策推論的優劣，即要判定樣本的變異性。

3.統計量的抽樣分布。

4.Bayes 推論中的後驗分布。

5.預測與控制問題中的條件分布。

圖1.1中的“模型”方框可以分成以下幾方面討論：

1.物理模型 (physical models) 從科學原理 中所得出的模型， 如用數學方程式表示的物理定律等等。

2.經驗模型 (empirical models) 這時沒 有什麼 可資借鑒 的基本原理，全憑我們的觀察，從所看到的現象中總結出模型，如經驗曲線、經驗公式等等。

3.隨機模型 (stochastic models) 例如，根據物理定律， 一個物體的位置 x 與它運動的速度 $\dot{x}$、運動的時間 t 以及對它施加的控制 u 有關，即有

$$x=f(\dot{x}, t, u)。$$

但實際上還有許許多多不確定的因素影響著物體的位置，所以我們實際觀察到的物體的位置與隨機干擾 ε 有關，即

$$x=f(\dot{x}, t, u)+\varepsilon。$$

這就是一個隨機模型。

4.模型類 (class of possible models) 常有這樣的情況： 很多模型在

一定的意義下都能與觀察值相擬合。究竟從其中選取哪個模型呢？這就要根據我們的最終目的或者模型的簡潔（parsimony）程度來決定了。

§4　機率

怎樣描述一個分布呢？如何用數學表達它呢？最自然的就是用機率。下面，我們先複習一下有關的機率知識。

1.隨機變量、機率分布。

前面，我們曾談到帶誤差的測量，由於測量值是有分布的，所以可以把它看成一個隨機變量，而把觀測到的數據看作該隨機變量所取的值。從這個意義上說，機率分布與隨機變量是分不開的，測量時，我們就是在觀察分布，觀測到的數據越多，對分布的估計就越精確，爲形象起見，人們常常對數據進行整理，然後把它們作成直方圖，從這種圖中，可以看其分布的概貌。

分布還有一些數字特徵，如均值、方差、中位數等等。

2.古典極限定理。

我認爲，在所有統計方法中，從哲學的觀點講，最終的基本點就是怎樣平衡噪聲或樣本波動的問題。也就是說，怎樣從雜亂無章的數據中剔除噪聲的影響或樣本波動的問題，大數定律很好地說明了這一基本點。

(1)大數定律　設隨機變量 $X_1, X_2, \cdots$ 爲i. i. d(independent, identically distributed)，$EX_1=\mu<\infty$，則有

弱大數定律：

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

即對任何 $\varepsilon>0$，有

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

強大數定律:

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

即

$$P\{\lim_{n\to\infty}\overline{X}_n=\mu\}=1\text{。}$$

對經驗分布函數有：對任何固定的x，有

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

其中 I_A 表示A的示性函數，$F(x)$ 爲 X_1 的分布函數。

還有更強的結果（Glivenko-Cantelli 定理）：

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

對於在抽象空間（如巴拿赫空間）上取値的隨機變量也有相應的大數定律。此外，若Φ是連續函數，則有

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

對於平穩序列有遍歷性定理。

⑵中心極限定理　設隨機變量 $X_1, X_2, \cdots$ i. i. d，$EX_1=\mu$，$\text{Var}X_1=\sigma^2$，則對充分大的n，有

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

即

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

這裏$\mathscr{D}$表示依分布收斂，所以上式意爲：對任何$a\leqslant b$，

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

在統計中，還常用到下述關於 F_n 的中心極限定理：設 $F(x)$ 是連續的分布函數，$F_n(x)$ 爲相應的經驗分布函數，則

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

(3)重對數定理　設隨機變量 $X_1, X_2, \cdots$ i. i. d, $EX_1=\mu$，$\mathrm{Var}X_1=\sigma^2$，則

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

其中$S_n=\sum_{i=1}^{n} X_i$。

關於 F_n 的重對數定理，請參看 Finkelstein (1971)。

對於取值於巴拿赫空間的隨機變量也有相應的結果。有興趣的讀者請參看 Kuelbs (1977) 的文章。

3.機率測度族（參看 Lehmann (1959) 書中的第二章）。

我們常用 $(\mathscr{X}, \mathscr{F}, P)$ 表示一個機率空間，其中$\mathscr{X}$是樣本空間，$\mathscr{F}$是$\mathscr{X}$上的 σ- 代數，P是可測空間$(\mathscr{X}, \mathscr{F})$上的機率測度，與機率論不同，在統計問題中，我們不知道具體的P，知道的是一個機率測度族 $(P_\theta, \theta \in \Theta)$，其中$\Theta$叫做參數空間。我們只知道樣本$X$是$(\mathscr{X}, \mathscr{F})$上的隨機變量，其分布屬於分布族$(P_\theta, \theta \in \Theta)$，而不知道$\theta$的具體值。通常，統計問題正是要根據樣本$X$，對未知參數$\theta$作爲某種推論。

例 1.1　設 $X_1, \cdots, X_n$ 爲 i. i. d 隨機變量，$X_1 \sim N(\theta, 1)$，其中θ未知，則$\Theta=(-\infty, +\infty)$，$P_\theta$爲$n$個$N(\theta, 1)$的乘積測度。

例 1.2　設$X_1, \cdots, X_n$ i. i. d, X_1的分布F未知，則$\Theta=\mathscr{F} \triangleq \{F: F$是分布函數$\}$。

例 1.3　指數分布族，如果P_θ關於σ有限測度μ絕對連續(卽$P_\theta \ll \mu$，

對一切 $\theta\in\Theta$)，則 P_θ 關於 μ 的 Radon-Nikodym導數存在，記爲 $p_\theta=\frac{dP_\theta}{d\mu}$，我們把 p_θ 叫做 P_θ 關於 μ 的密度函函數。

若隨機變量X的密度函數形如

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

其中 $c(\theta),Q_1(\theta),\cdots,Q_k(\theta)$ 是 Θ 上的有限函數，$h(x)$ 爲定義於 $\mathscr{X}$ 上的非負$\mathscr{F}$可測函數，$T_1(x),\cdots,T_k(x)$ 是$\mathscr{X}$上的$\mathscr{F}$可測函數，則我們把分布族 $\{P_\theta,\theta\in\Theta\}$ 叫做 k 維指數分布族。

若定義$\widetilde{\Theta}=\{(Q_1(\theta),\cdots,Q_k(\theta))：\theta\in\Theta\}$，則$\widetilde{\Theta}$中的每一點對應指數分布族中的一個分布，並且，爲了方便起見，常把 $h(x)$ “吸收” 到 μ 中，於是，指數分布族的密度函數就可以寫成

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

其中 $(\theta_1,\cdots,\theta_k)\in\widetilde{\Theta}$，我們把它叫做自然形式的指數族，把$\widetilde{\Theta}$ 叫做自然參數空間。

將$\int p_\theta(x)d\mu(x)=1$ 的兩邊對 θ_j 求導，便可以得到

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

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

還可以把指數分布族寫成更簡單的形式：

$$p_\theta(x)=e^{\theta x-\psi(\theta)}\qquad(1\text{ 維}),$$

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

$$=e^{\sum_{i=1}^{k}\theta_ix_i-\psi(\theta_1,\cdots,\theta_k)}\qquad(k\text{ 維})。$$

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

第二章　統計決策理論

§1　統計決策問題的一般形式

1. 統計決策問題的要素

(1)三個可測空間：樣本空間$(\mathscr{X},\mathscr{B}_1)$，其中$\mathscr{X}$是所有可能的數據值的集合；參數空間$(\Theta,\mathscr{B}_2)$，其中$\Theta$是所有可能的模型的集合；行動空間$(\mathscr{A},\mathscr{B}_3)$，其中$\mathscr{A}$是所有可能的行動(action)的集合，$\mathscr{B}_1,\mathscr{B}_2,\mathscr{B}_3$分別爲$\mathscr{X},\Theta,\mathscr{A}$上的$\sigma$-代數。

(2)樣本空間$(\mathscr{X},\mathscr{B}_1)$上的機率測度族$\mathscr{P}=\{P_\theta,\ \theta\in\Theta\}$，其中$P_\theta$滿足：對任意的$B\in\mathscr{B}_1$，$P_\theta(B)$是$\theta$的可測函數。

(3)損失函數L，它是$\Theta\times\mathscr{A}\to(0,+\infty)$的可測(關於$\mathscr{B}_2\times\mathscr{B}_3$)函數，$L(\theta,a)$表示參數眞値爲$\theta$時，統計學家採取行動$a$所蒙受的損失。

通常，我們用$(\mathscr{X},\Theta,\mathscr{A},P_\theta,L)$表示一個統計決策問題。

順便提一下，在有些領域裏，人們說的不是"損失"(loss)而是"收益"(gain)、"報酬"(reward)或"效用"(utility)，當然，人們總是希望採取使收益(=負損失)達到最大的決策。

2. 統計決策問題的解