应用数学丛书

统计推断与微分几何

韦博成

清华大学出版社

清华大学
应用数学丛书
第 6 卷

统计推断与微分几何

韦博成

国家自然科学基金资助项目

清华大学出版社

内 容 简 介

本书系统介绍近年来利用微分几何研究统计问题的主要成果．力求揭示微分几何概念与数理统计基本概念之间的内在联系和解决问题的思想方法．内容包括参数分布族和指数族的几何，曲指数族的几何及其统计分析，Edgeworth 展开式的几何理论及其对估计理论的应用等．附录中简要介绍了本书所用的微分几何知识．

本书对象为数理统计工作者，高等学校数学系教师、研究生和高年级大学生等．

统计推断与微分几何

韦博成

责任编辑：潘真微

清华大学出版社出版

北京　清华园

中国科学院印刷厂印刷

新华书店北京发行所发行

☆

开本：850×1168 1/32　印张：5.75　字数：148 千字

1988 年 8 月第 1 版　1988 年 8 月第 1 次印刷

印数：0001～6000　定价：精 3.70 元　平 1.20 元

精 ISBN 7-302-00258-4/O·49(课)

平 ISBN 7-302-00259-2/O·50(课)

关于《应用数学丛书》

为了满足广大科技人员、高等院校教师、研究生进一步学习应用数学的需要，我们编辑出版本丛书. 丛书内容将包括应用数学的各个方面、有关的边缘科学以及应用数学的方法等. 限于我们的水平和经验，丛书中难免有不少错误和不足之处，诚恳希望广大读者批评指正.

清华大学《应用数学丛书》编辑委员会

1983.4

序 言

近年来，微分几何方法已经广泛应用于数理科学的许多分支，成果累累. 统计学中采用微分几何方法起步较晚，首先由美国统计学家 Efron 在他 1975 年的著名论文中提出（见[11]）. 此后，微分几何方法逐步进入统计学领域. 由于许多统计学家的共同努力，特别是日本统计学家甘利俊一（即 Amari）的杰出工作（见[1]—[5]），使得统计学中的微分几何方法得到学术界的广泛注意，成为统计学的一个令人瞩目的新分支. 这个分支尚有待进一步探讨和研究. 目前，国内外介绍这一分支的论著尚少. 本书试图用比较短的篇幅系统介绍其基础内容，希望引起读者的兴趣.

本书共分四章. 第一章系统阐述微分几何中的基本概念，诸如流形、切空间、度量、联络、曲率等概念与统计学中的分布族、似然函数、Fisher 信息、最大似然估计量等基本概念之间的内在联系. 并介绍统计学中最常见的指数族的几何结构. 第二、第三章介绍曲指数族的几何及统计理论. 重点讨论与估计有关的渐近理论. 诸如一阶和高阶渐近有效性、渐近充分性、信息和信息损失、估计量的条件分布等问题. 第四章介绍关于 α-联络的几何. 这是一个研究统计问题很有前途的几何方法.

本书试图反映近年来统计学家利用微分几何研究统计问题的思想方法和主要成果. 但是，对读者的基础知识要求并不高，只要了解大学本科数理统计课程的内容，就能阅读本书. 为了便于读者阅读，并增强本书的自封性，书末还有二个附录. 附录A 介绍本书用到的微分几何知识，使统计工作者不必查阅有关的几何书籍

就能阅读本书. 附录 B 简要介绍了大学数理统计教材中较少见的 Edgeworth 展开式.

本书未能介绍某些重要的内容. 例如,假设检验的几何理论、时间序列的几何理论、非线性回归模型的几何理论等等. 如果加入这些内容,篇幅可能要增加一倍以上,因此只好割爱舍去. 有兴趣的读者可参阅 [3], [4], [15], [16],另外,作者已写出"非线性回归模型"一书初稿,打算作为本书的姐妹篇另行出版.

本书初稿曾在南京、武汉、贵州等地的统计讲习班、研究生班进行讲授. 通过这些学术活动对初稿进行了多次修改. 不少同志提出了许多宝贵意见,并对本书的出版表示热情的关注. 清华大学肖树铁教授、武汉大学张尧庭教授、江西师范大学倪国熙教授以及南京工学院高金衡教授、陶永德教授对本书的出版一直表示极大的支持与关心,在此一并表示衷心的感谢.

由于本人水平有限,书中肯定有不足和欠妥之处,恳请读者给予批评指正.

作　者

丛 书 目 录

目　录

第一章　参数分布族的几何与指数族

在参数统计中，对于一个含有有限个参数的统计模型，人们总是联系其参数空间来研究有关的统计问题．从几何观点来看，若把统计模型中每一个分布看成一个点，它与参数空间中的点是一一对应的．所有这些点组成的集合在通常条件下可以视为一个流形．这个流形的几何结构显然与相应分布的性质有关．在一定条件下，可以根据统计模型的分布在流形上建立微分结构，从而通过微分流形来研究分布的统计性质．这个想法是很自然的，早在1945年 Rao 就提出来了，并建议以 Fisher 信息阵定义流形上的 Riemann 度量（见[19]）．但由于种种原因，直到1975年 Efron 引入统计曲率的概念（见[11]）．特别是1982年甘利俊一（即 Amari）定义了一个单参数族的联络以后，才使微分流形的理论与方法逐步进入统计领域（见[2]）．

本章第一节将引入统计模型的流形概念，并定义相应的几何要素：切空间、度量、联络及其他有关的量，同时讨论这些几何量与统计概念之间的联系．第二节则专门讨论统计中最常见的指数分布族的几何结构．关于统计推断问题，将在以后各章逐步展开讨论．

§1.1　参数分布族的几何

设 $\mathscr{X}$ 为欧氏空间，$\mathscr{B}$ 为其 Borel 子集构成的 σ-域．考虑样本空间 $(\mathscr{X}, \mathscr{B})$ 上的一族随机变量 X 的分布族 $\{P_\theta,\ \theta\in\Theta\}$，其

中 P_θ 为定义在 $\mathscr{B}$ 上的概率测度，$\theta=(\theta^1,\cdots,\theta^n)'$，$\Theta$ 为欧氏空间 R^n 中的一个开集．设对任何 $\theta\in\Theta$，测度 P_θ 对某个 σ-有穷测度 μ 绝对连续：$P_\theta\ll\mu$，因而 P_θ 关于 μ 存在密度函数 $dP_\theta(x)/d\mu=p(x,\theta)$．本章及以后各章，如无特别声明均假设 $p(x,\theta)$ 为满足以下正则条件的正则密度族：

(1) $\theta_1\neq\theta_2$ 时，$p(x,\theta_1)$ 与 $p(x,\theta_2)$ 表示不同的密度函数．并假定对一切 $p(x,\theta)$，$\theta\in\Theta$ 存在有共同的支撑集合，使得在支撑集合上 $p(x,\theta)>0$．

(2) 定义函数 $l(x,\theta)\triangleq\log p(x,\theta)$(*)，假定 $l(x,\theta)$ 在 Θ 上关于 θ 存在所需要的各阶连续导数．令

$$\partial_i\triangleq\frac{\partial}{\partial\theta^i},\quad \partial_i l=\frac{\partial}{\partial\theta^i}l(x,\theta);\quad i=1,\cdots,n.$$

假设对于固定的 θ，$\partial_i l$ $(i=1,\cdots,n)$ 作为 x 的函数为线性无关．

(3) 对于任意我们所考虑的 μ 可积函数 $f(x,\theta)$，偏导数 ∂_i 与积分可以交换次序，即

$$\partial_i\left[\int f(x,\theta)d\mu(x)\right]=\int\partial_i[f(x,\theta)]d\mu(x).$$

(4) $\partial_i l(X,\theta)$ 作为随机变量 X 的函数(**)存在所需要的各阶矩．

下面我们来着手建立统计模型的流形．

在函数空间中定义集合

$$S=\{p(x,\theta)|\theta\in\Theta\},\tag{1.1}$$

其中 Θ 为欧氏空间 R^n 中的开集．这个开集本身可以看成一个拓扑．根据正则密度的条件(1)，S 和 Θ 之间存在一一对应关系．这种一一对应关系使得在 S 上诱导出与 Θ 同胚的拓扑结构．今考虑

(*) 符号 $\triangleq$ 表示定义或规定．

(**) 今后我们对随机变量 $l(X,\theta)$ 与实函数 $l(x,\theta)$ 并不严格加以区别，从上下文可以明确其意义．$p(x,\theta)$ 等亦然．

S 上任一点 $P=p(x,\theta)$ 和 P 的一个邻域 U，定义 U 到 R^n 的映射 $\phi:\phi(P)=\theta$. 显然，所有这样的邻域 U 是 S 的一个覆盖，ϕ 是一个微分同胚映射，这就在 S 上建立了一个微分结构，因而 S 形成一个微分流形，称为统计模型的流形，或简称为统计流形. θ 称为统计流形的自然坐标，当然我们还可以选取其他参数得到其他坐标.

例 1.1 一元正态分布密度族. 设一维密度函数为

$$p(x,\mu,\sigma^2)=\frac{1}{\sqrt{2\pi}\,\sigma}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\},$$

记为 $N(\mu,\sigma^2)$. 这时参数是二维的：$\theta=(\theta^1,\theta^2)$，$\theta^1=\mu,\theta^2=\sigma$. 参数空间很自然的定义为半平面

$$\Theta=\{(\theta^1,\theta^2)|-\infty<\theta^1<+\infty,\theta^2>0\}.$$

因此，$S=\{p(x,\theta)|\theta\in\Theta\}$ 形成一个二维微分流形.

注记 对于统计流形，不能简单地把它等同于参数空间 Θ. 例如，设 $p_1(x,\theta)$ 对应于 $N(0,\theta^2)$，$p_2(x,\theta)=\theta e^{-\theta x}$（当 $x\geqslant 0$）. 二者的参数空间 Θ 都是 $(0,+\infty)$. 显然 $S_1=\{p_1(x,\theta),\theta>0\}$ 不同于 $S_2=\{p_2(x,\theta),\theta>0\}$. 另外，不同的分布将在流形上建立起不同的度量联络等几何结构.

下面开始讨论统计流形 (1.1) 的基本几何量及相应的统计意义.

（一）切空间及其统计表示

流形 S 上一个定点 P 处的切空间 T_P 是一个 n 维向量空间，其实质就是 S 在 P 点附近的线性近似. 在自然坐标 $\theta=(\theta^i)$ 下，切空间的自然基就是 $\partial_i=\partial/\partial\theta^i\ (i=1,\cdots,n)$，任一切向量 $A\in T_P$ 可表示为 $A=A^i\partial_i$，$A^i\ (i=1,\cdots,n)$ 为切向量 A 的坐标. 这里我们使用了 Einstein 求和约定，即凡是上下指标相同表示求和，即

$$A^i\partial_i=\sum_{i=1}^{n}A^i\partial_i.$$

这时切空间可表示为

$$T_P = \{A \mid A = A^i\partial_i\}.$$

P 点某邻域向量场的集合记为 $T(S)$，不致混淆时也简记为 T.

上述切空间对于讨论统计问题并无直接帮助. 因为无论是 ∂_i 或 $A = A^i\partial_i$ 作为微分算子它们与统计概念相距甚远. 为了使切空间能密切地联系到统计问题，我们设法使它与某种类型的随机变量建立同构对应. 这种同构对应称为切空间的统计表示. 最直接的统计表示为下列同构映射:

$$A \to Al,\ A \in T_P,\ l = l(x, \theta).$$

基向量 ∂_i 对应于 $\partial_i l$. 因此在自然坐标下，正则条件 (2) 确保这个映射是同构的. 这时切空间 T_P 同构于下列空间:

$$T_P^{(1)} = \{A(x) \mid A(x) = A^i\partial_i l(x, \theta)\}.$$

T_P 中的切向量 (微分算子) $A = A^i\partial_i$ 对应于 $T_P^{(1)}$ 中的随机变量 $A(x) = A^i\partial_i l(x, \theta)$. 从代数观点来看，可以认为 T_P 与 $T_P^{(1)}$ 是等同的. T_P 是切空间微分算子形式的表示，$T_P^{(1)}$ 是同一切空间随机变量形式的表示. 前者便于运算，后者则使我们有可能把流形观点应用于某些统计问题. $T_P^{(1)}$ 称为切空间的 1-表示，这是所谓 α-表示的特殊情形，可参看第四章.

切空间的 1-表示有一个很好的特性，我们把它写成一个引理

引理 1.1 对任一向量 $A(x) \in T_P^{(1)}$，我们有

$$E_\theta[A(x)] = 0. \tag{1.2}$$

其中 E_θ 表示随机变量关于 $p(x, \theta)$ 的数学期望.

证明 根据正则条件 (3)，我们有

$$\begin{aligned} E_\theta[\partial_i l] &= \int \partial_i l(x, \theta) p(x, \theta) d\mu(x) \\ &= \int \frac{\partial_i p(x, \theta)}{p(x, \theta)} p(x, \theta) d\mu(x) \\ &= \partial_i \int p(x, \theta) d\mu(x) = 0. \end{aligned}$$

由于 $A(x)=A^i\partial_i l$，因此可得 $E_\theta[A(x)]=0$. 证毕.

这个引理指出了切空间 $T_P^{(1)}$ 的一个重要特征，今后经常用到.

注记 切空间的 1-表示可以作如下形象的理解：设

$$S^{(1)}=\{l(x,\theta)\mid\theta\in\Theta\}, \tag{1.3}$$

由于 $l=\log p$ 为单调函数，因此 $S^{(1)}$ 和 S 之间具有基于单调函数的一一对应. $S^{(1)}$ 可以认为是 S 的 1-表示，它可以反映 S 的许多几何特性. 我们形象地来看一看 $S^{(1)}$ 的切空间. 设 c 为 $S^{(1)}$ 上通过某点 P 的一条曲线

$$c=\{l(x,\theta(t))\mid\theta(t)\in\Theta,\ t\in(-1,1)\}.$$

为简单起见，设 P 点的坐标对应于 $t=0$，过 c 上 P 点的切向量可表示为

$$\frac{dl}{dt}=\partial_i l\cdot\dot\theta^i|_{t=0},\quad \dot\theta^i=\frac{d\theta^i}{dt}.$$

这正好是 $\partial_i l$ 的线性组合. 由 c 的任意性，我们可以断定 $\partial_i l(i=1,\cdots,n)$ 恰好是 $S^{(1)}$ 在 P 点处的切空间的一组基(事实上，$\partial_i l$ 为坐标曲线 c_i: $\theta^j(t)=\theta^j(0)+\delta_i^j t$, $j=1,\cdots,n$ 的切向量，其中 $\theta^j(0)$ 为 P 点的坐标). 因此 $S^{(1)}$ 在 P 处的切空间正好是 $T_P^{(1)}$. (1.3) 式经常可以帮助我们理解有关统计流形的问题.

以上讨论都是在自然坐标 $\theta=(\theta^i)$ 中进行的，当然还可以采用其它坐标. 今假定 $\omega=(\omega^\alpha)$, $\alpha=1,\cdots,n$ 为另一组坐标，这时切空间的另一组基及相应切向量可表示为

$$\partial_\alpha\triangleq\frac{\partial}{\partial\omega^\alpha},\quad A=A^\alpha\partial_\alpha,$$

∂_α 和 ∂_i 有如下关系

$$\partial_\alpha=B_\alpha^i\partial_i,\quad B_\alpha^i\triangleq\frac{\partial\theta^i}{\partial\omega^\alpha}. \tag{1.4}$$

通常假定 Jacobi 矩阵 (B_α^i) 是非退化的，因此

$$\partial_i=B_i^\alpha\partial_\alpha,\quad B_i^\alpha\triangleq\frac{\partial\omega^\alpha}{\partial\theta^i}. \tag{1.5}$$

向量 $A=A^\alpha\partial_\alpha=A^i\partial_i$; A^α, A^i 有如下关系

$$A^{\alpha} = B^{\alpha}_{i} A^{i}, \quad A^{i} = B^{i}_{\alpha} A^{\alpha}. \tag{1.6}$$

对于向量的 1-表示有类似的式子，只是注意

$$A(x) = A^{\alpha}\partial_{\alpha} l = A^{i}\partial_{i} l.$$

从统计上来讲，换一组坐标就相当于对统计模型选择另一组刻画模型的参数.

例 1.2 一元正态分布族 $N(\mu, \sigma^2)$ 的切空间(续例 1.1).

若取 $\theta = (\theta^1, \theta^2) = (\mu, \sigma)$，则 $\partial_1 = \partial/\partial\mu$，$\partial_2 = \partial/\partial\sigma$，$T_P$ 由这二个算子生成. 再看 $T_P^{(1)}$，

$$l(x, \theta) = -\frac{(x-\mu)^2}{2\sigma^2} - \log(\sqrt{2\pi}\,\sigma),$$

$$\partial_1 l = \frac{x-\mu}{\sigma^2}, \quad \partial_2 l = \frac{(x-\mu)^2}{\sigma^3} - \frac{1}{\sigma}. \tag{1.7}$$

$T_P^{(1)}$ 由这二个随机变量生成，因此由引理 1.1，$T_P^{(1)}$ 可写成

$$T_P^{(1)} = \{A(x) = ax^2 + bx + c \mid E[A(x)] = 0\}.$$

即要求 $c = -E[ax^2 + bx] = -a(\sigma^2 + \mu^2) - b\mu.$

我们也可以取前二阶矩作为模型的参数，即取 ω 如下

$$\omega^1 = E(x) = \mu = \theta^1,$$
$$\omega^2 = E(x^2) = \mu^2 + \sigma^2 = (\theta^1)^2 + (\theta^2)^2.$$

这时坐标变换的 Jacobi 矩阵为

$$(B^{i}_{\alpha}) = \left(\frac{\partial\theta^{i}}{\partial\omega^{\alpha}}\right) = \begin{pmatrix} 1 & 0 \\ -\dfrac{\mu}{\sigma} & \dfrac{1}{2\sigma} \end{pmatrix}.$$

由此可求出二组基之间的关系.

(二) Riemann 度量和 Fisher 信息

流形上的度量张量由切空间的内积所决定. 切空间内积定义的任意性是很大的，但是在统计流形上引入 Riemann 度量必须使之联系到统计概念，以有利于统计问题的讨论. 比较自然的想

法是使度量张量与 Fisher 信息阵相一致(*). 今正式定义如下:

定义 1.1 任给切空间 T_P 的二个向量 A、B，其 $T_P^{(1)}$ 空间的同构向量为 $A(x)$、$B(x)$. 则它们的内积同时定义为

$$\langle A, B\rangle = \langle A(x), B(x)\rangle \triangleq \mathrm{Cov}[A(x), B(x)].$$

设在自然坐标下，$A = A^i\partial_i$，$B = B^j\partial_j$，则

$$\langle A, B\rangle = \langle A(x), B(x)\rangle = E_\theta[(A^i\partial_i l)(B^j\partial_j l)]$$
$$= A^iB^jE_\theta[(\partial_i l)(\partial_j l)].$$

因此在此坐标系中,内积可表示为

$$\langle A, B\rangle = \langle A(x), B(x)\rangle = A^iB^jg_{ij}, \tag{1.8}$$

$$g_{ij}(\theta) \triangleq E_\theta[\partial_i l(x, \theta)\partial_j l(x, \theta)]. \tag{1.9}$$

由 (1.9) 式可知，度量矩阵 (g_{ij}) 就是分布族 $p(x, \theta)$ 关于参数 θ 的 Fisher 信息阵. 由正则条件 (2) 可知它是正定的，其逆阵记为 (g^{ij}). 由 (1.9) 式可知 g_{ij} 是二阶协变张量的分量,通常称为度量张量, g^{ij} 有时也称为度量张量(反变形式).

由正则条件 (3) 可以得到度量张量 (1.9) 式的一个方便的计算公式

$$g_{ij}(\theta) = -E_\theta[\partial_i\partial_j l(x, \theta)]. \tag{1.10}$$

事实上,由等式 $\partial_i\partial_j l = p^{-2}[p(\partial_i\partial_j p) - (\partial_i p)(\partial_j p)]$ 得

$$\partial_i\partial_j l = p^{-1}(\partial_i\partial_j p) - (\partial_i l)(\partial_j l),$$

$$E(\partial_i\partial_j l) = \int(\partial_i\partial_j p)d\mu(x) - E_\theta[(\partial_i l)(\partial_j l)]$$

$$= \partial_i\partial_j\left[\int p(x, \theta)d\mu(x)\right] - g_{ij}(\theta)$$

$$= -g_{ij}(\theta).$$

假如选取另一组坐标 $\omega = (\omega^\alpha)$，则由 (1.4) 式可知

$$g_{\alpha\beta}(\omega) \triangleq \langle\partial_\alpha, \partial_\beta\rangle = E[(\partial_\alpha l)(\partial_\beta l)]$$
$$= B_\alpha^i B_\beta^j g_{ij}. \tag{1.11}$$

(*) 事实上,从某种意义的不变性来讲，Fisher 信息阵是唯一合适的 Riemann 度量. 参见第四章.