# 高等数理统计学

GAO DENG SHU LI TONG JI XUE

陈希孺 著



中国科学技术大学出版社

# 高等数理统计学

陈希孺　著

中国科学技术大学出版社

**图书在版编目(CIP)数据**

高等数理统计学/陈希孺著.—合肥:中国科学技术大学出版社,1999.2

ISBN 7-312-00896-8

Ⅰ.高… Ⅱ.陈… Ⅲ.数理统计 Ⅳ.0212

中国版本图书馆CIP数据核字(1999)第02204号

中国科学技术大学出版社出版发行

(安徽省合肥市金寨路96号,230026)

中国科学技术大学印刷厂印刷

全国新华书店经销

开本:850×1168/32 印张:22.75 字数:600千

1999年2月第1版 1999年2月第1次印刷

印数:1—5100册

ISBN 7-312-00896-8/O·192 定价:25.00元

# 序　言

十余年前，笔者写过一部《数理统计引论》，当时的意图是作为一本专著来写，充作教材的想法倒是第二位的。可因当时百废待兴，数理统计学的教学、参考用书都很缺乏，因此该书出版后颇被充作这一用途。由于该书包含了不少超出基础课范围以外的材料，用作为教本殊有其不便之处。另外，习题也太少了一些，而作者一向主张，在打基础的阶段，应强调多做习题。

由于这些问题的存在，并考虑到随着学习数理统计及相近专业的青年人的队伍愈来愈扩大，这类教材今后的需要还会增加，多年来，笔者就有一个心愿，即按一本基础课教科书这个唯一的目标来重写这本书，并大大扩充其习题部分，其结果就是呈现在读者面前的这部书稿。

本书的定位是"基于测度论的数理统计学基础教科书"。内容除预备知识外，其主体是关于几种基本统计推断形式（点及区间估计，假设检验）的大小样本理论和方法，另有一章讲述线性模型的初步理论。凡是只宜在专门课程中展开讨论的内容，则一律不列入。这些看目录即可了然，故不在此细加说明了。

书中习题及提示占了近半的篇幅，从写作时间言，则占了四分之三以上。总计得题五百，若计小题，则不止千数。其中除少量选摘自有关著作外，大半属作者自创。有时一题之设，累日始成，可以说倾注了不少心力。这样做完全是因为，多做习题，尤其是多做难题，对掌握并熟练数理统计学基本的论证方法和技巧上，有着不可替代的重要性。如果通过一个基础课的学习，只是记住了若干概念，背了几个定理，而未能在这方面有所长进，那就真是"入宝山而空返"了。技巧的熟练固非一日之功，但取法乎上，仅得乎中，必须

在开始学基础课时就设定一个高目标。日后进入研究工作，克服难点的能力如何，相当一部分就取决于在这上面修为的深浅了。同时，经验表明，在打基础的阶段因忽视习题而导致素质上的缺陷，在日后不易弥补，或事倍功半。

笔者在学生时代及其后的几年中，对做习题未给予足够重视。当时误认为做题费时间，不增长新知识，不如多读些书，占得实地，以后试做研究工作，就日渐感到其不良后果，表现在碰到问题办法少，容易钻死胡同，克服难点的能力弱，以致对自己缺乏信心，对许多方法，都似雾里看花，似曾识面，而不能切实掌握和灵活运用。有如十八般兵器，样样都见过，但拿到手里，就使不动或很笨拙。欲以此克敌制胜，自难有成，以后稍明白了这一点，做了些亡羊之补，终究晚了一些，所谓"困而学之，又其次也"。"熟能生巧"，前人的经验不诬。而要达到"熟"，舍大量做题，无它捷径可循。几十年来，审了大量的杂志稿件，每见某些工作，由于未经深思，为一个并不难克服之点加上了若干不必须的繁复条件，从而使整个工作流于肤浅。这根子，大略也在于早先在习题上下的工夫不够，以致难于产生别出心裁的想法。

以本书的习题量，要求学员在课程时间范围内做完，恐不现实。但作者本意并非把这一组题全作为课内习题，而是把它作为"打基础"这个工作的一环，一两年、两三年完成都可以，有空就做一点。根据题的难易将其分为三类：加"＊"号的难度较大，加"°"的相对容易，教师可考虑作为课外作业，不加任何记号的，其难度介乎二者之间。对自学者、已经研究生毕业的青年教师和研究者，可利用这组题测试一下自己解题的能力如何。可能会有一种意见，认为这组题过于偏难。作为课程作业，这的确如此。但笔者觉得，从"打基础"，从锻炼技巧和提高能力这一目标看，非做难题不行，这道理正如训练运动员要加大运动量，做高难动作，不然，在训练的过程中舒服了，就别指望出好成绩。何况，对一个有志于在将来搞基础研究的人，日后在研究工作中将碰到的难点，比起这些习题，

又要高出若干个数量级。如果在现在面对这种习题尚有畏难，那又怎能指望在日后研究工作中能具备克服更大困难的能力和信心？

各题都有详细提示，大多数较难的题都给出了完整解答。这是因为，鉴于这些题的难度，需要有一个解答文本在，以作为依据。对读者而言，笔者切望这部分是备而不用、备而少用。如碰到一个题一时做不出来，宁肯暂时搁一搁，也不要轻易翻看解答。譬如登山，经过艰苦努力上了峰顶，自有其乐趣和成就感。反之，如在未尽全力之前就任人抬上去，则不惟无益，适足以挫折信心。

以上就习题一事唠叨了半天，读者也许烦了，就此打住。千言万语，归结到一点：希望大家多做题，做难题。“千里之行，始于足下”，就从今日开始吧！

本书的出版，得到中国科学技术大学出版社的大力支持。赵林城、方兆本和缪柏其等诸位教授给了很多鼓励和帮助。特别是吴耀华博士，在百忙中拨冗对稿件作了校阅，花费了不少的心力和时间。对以上机构和同志，作者谨借此机会表示衷心的感谢。书中不妥以至谬误之处，在所难免，尚祈广大读者和同行专家不吝指教。

陈希孺

1996年除夕于北京

# 目　录

# 第一章　预备知识

本章的内容有两个方面：一是构成一个统计问题的诸要素，一是**统计量**. 前者描述了一个统计问题的提法中所涉及的种种概念，而统计问题的解决，则总要通过统计量的形式去表达. 故此二者可以说是数理统计基础中的基础.

## 1.1　样本空间与样本分布族

数理统计学的任务，统而言之，是如何获得**样本**（观察或试验结果）和利用样本，以对事物的某些未知方面进行分析、推断以至作出一定的决策. 对“如何收集样本”这个问题的研究，是数理统计学中两个重要分支学科——**试验设计**和**抽样调查**的任务，本书将不涉及这一方面. 就是说，本书只讨论在已有样本的情况下怎样去进行统计分析的问题.

从数学上说，**样本空间**就是一个可测空间$(\mathscr{X},\mathscr{B})$，其中集合$\mathscr{X}$包含了随机元$X$的一切可能取值，$\mathscr{B}$是$\mathscr{X}$的某些子集构成的$\sigma$域. 依$X$的概率分布而从$\mathscr{X}$中随机地抽出的一个元素$x$，就叫做样本. 在许多常见的统计问题中，$X$是一个有限维（$k$维）随机向量，这时总是取$k$维欧氏空间$\mathbf{R}^k$或$\mathbf{R}^k$之一适当的 Borel 子集作为$\mathscr{X}$，而取$\mathscr{X}$的一切 Borel 子集作为$\mathscr{B}$. 这样的样本空间称为**欧氏样本空间**. 有了这个约定，我们就不必在每个场合下对样本空间进行仔细的界定了.

随机元$X$有一定的概率分布$F$. 对数理统计学的问题，$F$总是未知的，或至少是部分未知的. 可以把这个意思说成：$F$属于某个**分布族**$\mathscr{F}$，$\mathscr{F}$的具体含义要在特定的统计问题中作确切的界

定. 有时,$\mathscr{F}$ 表述了我们对某一特定问题在理论或经验上的了解,有时则仅仅是一种假定,实际上往往是二者兼而有之. 例如,依中心极限定理及经验的积累,有一定理由认为 $F$ 应近似于正态,但未必是严格正态,于是取 $\mathscr{F}$ 为正态分布族就也包含有数学假定的成分.

在数理统计学上把分布 $F$ 称为**样本分布**,而 $\mathscr{F}$ 则称为**样本分布族**,也有称之为**总体分布**和**总体分布族**的. 这二者其实是一个问题的两面:在统计上把随机元 $X$ 称为总体,$F$ 是 $X$ 的分布,自可称为总体分布. 另一方面,样本 $x$ 又是"依 $X$ 的分布(即 $F$)从 $\mathscr{X}$ 中随机抽取"所得,故 $x$ 也有分布 $F$,称之为样本分布亦在理中. 但是,在一个重要情况下这两个概念的通常理解有些差别,后面将要说明.

我们来解释一下概率分布的**负荷集**或称**支撑**(support)的概念. 设 $F$ 是 $\mathbf{R}^k$ 上的一个概率分布,$a\in\mathbf{R}^k$. 以 $a$ 为中心 $r$ 为半径的球体记为 $S(a,r)$. 若对任何 $r>0$,$S(a,r)$为 $F$-正测度集,则称 $a$ 为 $F$ 的一个负荷点或支撑点. 一切这样的 $a$ 构成的集 $A$,就称为 $F$ 的负荷集或支撑. 容易证明 $A$ 为闭集,故 $F(A)$有意义. 不难证明:$F(A)=1$(习题 1),就是说,$F$ 的全部概率都在集 $A$ 上. 这解释了"负荷"或"支撑"这种称谓的由来. 更一般一些,也可以把任一满足如下两条件的 Borel 集 $B$ 称为 $F$ 的支撑:$B\subset A$,$F(B)=1$. 给出这个灵活性有其道理,见下. 关于负荷的概念也显然可推广到 $F$ 为一般距离空间上的非负测度.

设有一个定义在 $\mathbf{R}^k$ 上的概率分布族 $\mathscr{F}$. 对 $F\in\mathscr{F}$,以 $A_F$ 记 $F$ 的支撑,则集 $A=\bigcup_{F\in\mathscr{F}}A_F$ 称为 $\mathscr{F}$ 的支撑. 若存在一个 Borel 集 $A$,它是每一个 $F\in\mathscr{F}$ 的支撑,则称 $\mathscr{F}$ 有**公共支撑**. 有公共支撑的一个最重要的情况如下:设分布族 $\mathscr{F}$ 定义于欧氏可测空间 $(\mathscr{X},\mathscr{B})$,设 $\mu$ 为此空间上的一个 $\sigma$ 有限的测度(对"测度",如无相反申明,概指非负测度),满足条件

$$\mu(C)=0\Rightarrow F(C)=0,\quad 对任何\ F\in\mathscr{F}.$$

则称 $\mathscr{F}$ 受控于 $\mu$,可记为 $\mathscr{F}\ll\mu$. 依 Radon-Nikodym 定理,对每个 $F\in\mathscr{F}$ 相应地有一个定义于 $\mathscr{X}$ 的非负 Borel 可测函数 $f$,使 $\mathrm{d}F=f\mathrm{d}\mu$. 在统计上称 $f$ 为分布 $F$ 对测度 $\mu$ 的**密度**. 记

$$A(f)=\{x:x\in\mathscr{X},f(x)>0\}.$$

易证(习题 2):若 $A(f)=A$ 与 $F\in\mathscr{F}$ 无关,则 $\mathscr{F}$ 有公共支撑,它是 $A$ 的某一子集.

我们来看几个简单例子. 设 $\mathscr{F}$ 为**一维指数分布族**

$$\mathrm{d}F_\theta(x)=\theta\mathrm{e}^{-\theta x}I(x>0)\mathrm{d}x,\quad \theta>0, \tag{1.1}$$

这里 $I$ 是**指示函数**. 对每个 $\theta>0$,任何 $x\geqslant 0$ 都是 $F_\theta$ 的负荷点,而 $x<0$ 则不是. 故 $F_\theta$ 的支撑是 $A=\{x:x\geqslant 0\}$,它与 $\theta$ 无关,因而是公共支撑. 按前述对支撑更一般的定义,也可以取 $\widetilde{A}=\{x:x>0\}$ 作为支撑. 后者有一个好处,即突出了相应变量只取正值的性质.

另一个例子是**离散分布**. 设 $F$ 是 $\mathbf{R}^k$ 上之一离散分布,则通常总是把集合 $A=\{a:a\in\mathbf{R}^k,F(\{a\})>0\}$ 作为 $F$ 的支撑,这也要按前述更一般的定义才行. 例如,若把 $\mathbf{R}^k$ 的一切有理点排列为 $a_1$, $a_2,\cdots$,而定义分布 $F$ 为:$F(\{a_r\})=2^{-r},r=1,2,\cdots$. 则易见 $\mathbf{R}^k$ 中任一点都是 $F$ 的负荷点,而按起初的定义 $F$ 的支撑应是 $\mathbf{R}^k$,而不是 $A=\{a_1,a_2,\cdots\}$,这就不自然了. 但按后一定义,仍可取 $F$ 的支撑为 $A=\{a_1,a_2,\cdots\}$,这就是我们要把支撑的定义略为推广的理由之一. 另一个理由是:若不这样做,则一些实质上有共同支撑的分布族,严格说来却是没有共同支撑.

现在让我们再回到样本空间与样本分布族的问题. 有一个特例值得讨论一下. 把随机元 $X$ 记为$(X_1,\cdots,X_n)$,相应地把样本 $x$ 记为$(x_1,\cdots,x_n)$. 假定 $X_1,\cdots,X_n$ 为 iid.(独立同分布). 取一个与 $X_1$ 同分布的随机变量 $\widetilde{X}$. 则我们可以把 $x_1,\cdots,x_n$ 视为在完全同等的条件下对 $\widetilde{X}$ 所作的 $n$ 次独立观察值. 若 $\widetilde{X}$ 有分布 $\widetilde{F}$,则 $X$ 有分布 $F=\widetilde{F}^n=\widetilde{F}\times\cdots\times\widetilde{F}$,完全由 $\widetilde{F}$ 所决定. 故在这个结构中,可以取消 $X$ 和 $F$ 的地位,只要 $\widetilde{X}$ 和 $\widetilde{F}$ 就行. 这时,通常把 $\widetilde{X}$ 的分布 $\widetilde{F}$ 的族称为总体分布族,它决定了样本 $x$ 的分布的族——样本分

布族. 故在这个特例下，总体分布与样本分布有明确不同的含义. 相应地，通常就把 $x_1,\cdots,x_n$ 称为“**从总体（或变量）$\widetilde{X}$ 中抽取的** iid.（**或随机**）**样本**”. $n$ 常称为 $x=(x_1,\cdots,x_n)$ 的**样本量**或**样本大小**，以后我们只用前一名称.

样本空间连同赋予其上的样本分布族，构成一个统计问题的基本要素，它的确定或指定，给予了问题一个确定的**统计模型**，也可称**概率模型**. 这样的规定派生出一个事实，即现代统计分析的理论、方法，乃至统计分析所得结果的解释，莫不仰赖于概率论. 这对于现代统计学中的两大流派——**频率学派**（或**古典学派**）及 **Bayes 学派**都一样成立，只是对样本分布的意义的理解上有所不同而已. 说到这个样本分布族 $\mathscr{F}$（对 idd. 样本的情况也可以取总体分布族代替之），按其结构的复杂性可以划分为以下几类：

1. **参数族**. 当 $\mathscr{F}$ 中的分布的一般数学形式已知，但包含若干个（通常为数甚少）未知**参数**时，称 $\mathscr{F}$ 为参数族. 形式上可以这样说：存在某一个维数 $r$ 的欧氏空间 $\mathbf{R}^r$ 的一 Borel 子集 $\Theta$，使 $\mathscr{F}$ 中的分布 $F$ 可以与 $\Theta$ 中之点 $\theta$ 建立一一对应的关系. 这样，可以把 $\mathscr{F}$ 中的分布 $F$ 记为 $F_\theta$，$\theta$ 称为参数，而 $\Theta$ 称为**参数空间**. 为方便计，也可以把任何包含 $\Theta$ 的 Borel 集叫做参数空间. 通常总要设法取最小的 $r$，这时 $r$ 可以称为所论统计模型的维数，故有一维统计问题（也称单参数统计问题）和多维统计问题（多参数统计问题）之别，这个维数当然不能与分布 $F$ 本身的维数混为一谈.

很多在统计理论或应用中常见的参数分布族，读者当然都能耳熟能详，诸如**二项**、**Poisson**、**正态**、**指数**和 **Weibull** 分布族之类，在此也暂不一一列举了.

2. **非参数族**. 当 $\mathscr{F}$ 中的分布不能通过有限个未知参数去刻画时，$\mathscr{F}$ 称为非参数（分布）族. 例如，$\mathscr{F}$ 是一切一维对称分布；是一切其期望（或某阶矩）存在有限的一维分布；是一切处处连续的一维分布等等，都是非参数族的典型例子. 在这里，我们仍可以形式地把 $\mathscr{F}$ 写成 $F_\theta:\theta\in\Theta$，并称 $\Theta$ 为“参数空间”而 $\theta$ 为“参数”，

只不过 $\theta$ 就是 $F$ 本身，而 $\Theta$ 就是 $\mathscr{F}$ 而已. 这种形式的记法使我们在表述一个问题时不必一定要分别参数和非参数两种情况去陈述.

如果一个统计模型中的样本分布族 $\mathscr{F}$ 是(非)参数性的，则该问题称为(**非**)**参数统计问题**. 研究非参数统计问题的统计学分支称为**非参数统计**，相应地当然也可以定义**参数统计**. 习惯上并不把参数统计作为统计的一个分支，狭义地说，也可以把 $\mathscr{F}$ 为正态分布族(或加上另一些常见的重要的参数族)时的统计问题定义为参数统计的内容.

3. **半参数族或部分参数族**. 这个名词产生较晚，约在 80 年代. 本来，据上述定义，任一指定的分布族 $\mathscr{F}$，必能归入"参数"或"非参数"两类之一，因而没有所谓"半参数族"存在的余地. 但是，如果不是只着眼于样本(或总体)的分布本身而是把注意力放在其某一重要特征上，则情况有所不同. 下面就一个近年来研究很多的模型来说明这一点.

考虑一个以 $p$ 维变量($p\geqslant 1$)$X$ 和一维变量 $T$ 为自变量，一维变量 $Y$ 为因变量的**均值回归模型**. 设回归函数，即给定 $X=x$ 和 $T=t$ 时 $Y$ 的条件期望值，有形式

$$\mathrm{E}(Y|X=x,T=t)=\alpha+x'\beta+g(t), \tag{1.2}$$

其中 $\alpha$ 和 $\beta$ 分别为一维及 $p$ 维未知向量，$g$ 为定义在某区间上的，其形式未知的函数. 可以要求 $g$ 满足某些一般性条件，如连续，或二阶导数存在有界之类. 对这种模型，如按其分布去划分，应归入非参数一类，因($X,T,Y$)的分布族并不能通过有限个参数去刻划之. 但是，由于在回归分析中我们主要重视的是作为该分布的一项特征——回归函数(1.2)，而(1.2)可分为两部分：$\alpha+x'\beta$ 和 $g(t)$，前者是参数性的而后者是非参数性的，据这一点，把上述统计模型称为半参数(或部分参数)的. 将这个例子的精神加以类推，可以举出其他一些半参数的例子. 但是，"半参数"的概念难于用完全精确的语言加以界定，说到底，也没有这个必要. 因为，把一

个统计模型归入何类，对其处理方法并没有多大的影响.

在本节的末尾我们要介绍一种在理论和应用上都很重要的样本或总体分布族——**指数型分布族**. 它包含了统计上最常见的一些分布族.

设有欧氏可测空间$(\mathscr{X},\mathscr{B})$及其上的$\sigma$有限测度$\mu$. 设$h\geqslant 0$和$T_1,\cdots,T_k$都是定义于$\mathscr{X}$的 Borel 可测函数. 设$\Theta$为一集合，$C>0$和$Q_1,\cdots,Q_k$都是定义在$\Theta$上的取有限实值的函数，满足条件

$$0<1/C(\theta)=\int_{\mathscr{X}}\exp\Big(\sum_{i=1}^{k}Q_i(\theta)T_i(x)\Big)h(x)\mathrm{d}\mu(x)<\infty,\ \theta\in\Theta, \tag{1.3}$$

令

$$\mathrm{d}F_\theta(x)=C(\theta)\exp\Big(\sum_{i=1}^{k}Q_i(\theta)T_i(x)\Big)h(x)\mathrm{d}\mu(x), \tag{1.4}$$

则称$\{F_\theta,\theta\in\Theta\}$为一指数型分布族.

作为例子，容易看出统计上常用的几个分布族都是指数型，例如：

**正态分布族**：$X=(X_1,\cdots,X_n)$，$X_1,\cdots,X_n$为 iid.，公共分布为$N(a,\sigma^2)$，参数空间$\Theta=\{\theta=(a,\sigma^2):-\infty<a<\infty,0<\sigma^2<\infty\}$为开半平面. 取$\mu$为$\mathbf{R}^n$中的$L$测度，易把$X$的密度写成(1.4)的形式，其中$k=2$，而

$$\begin{gathered}h(x)=1,\quad T_1(x)=\bar{x}=\sum_{i=1}^{n}x_i/n,\quad T_2(x)=-\sum_{i=1}^{n}x_i^2,\\ C(\theta)=(2\pi\sigma^2)^{-n/2}\exp(-na^2/2\sigma^2),\\ Q_1(\theta)=na/\sigma^2,\ Q_2(\theta)=1/2\sigma^2.\end{gathered} \tag{1.5}$$

**二项分布族**：：$X\sim B(n,\theta)$，$0<\theta<1$. 取$\mu$为集$\{0,1,\cdots,n\}$上的计数测度，易将$X$对$\mu$的密度写成(1.4)的形状，其中$k=1$，而

$$h(x)=\binom{n}{x},\quad T_1(x)=x,\quad C(\theta)=(1-\theta)^n,$$

$$C_1(\theta) = \log(\theta/(1-\theta)). \tag{1.6}$$

此表达方式当 $\theta=0$ 及 $\theta=1$ 时失效，而在二项分布原来的形式中，这两值都是允许的.

**Poisson 分布族**：$P_\theta(X=i)=\mathrm{e}^{-\theta}\theta^i/i!, i=0,1,2,\cdots,\theta>0$. 即 $\mu$ 为集$\{0,1,2,\cdots\}$上的计数测度，易将 $X$ 对 $\mu$ 的密度写成(1.4)的形状，其中 $k=1$，而

$$h(x) = 1/x!,\ T_1(x) = x\ ,\ C(\theta) = \mathrm{e}^{-\theta},\ Q_1(\theta) = \log\theta. \tag{1.7}$$

此表达方式当 $\theta=0$ 时失效，而在 Poisson 分布原来的形式中，$\theta$ 可以取 0 为值.

参数空间 $\Theta$ 中每一元 $\theta$ 都满足(1.3)，但 $\Theta$ 不一定包含了全部满足(1.3)的 $\theta$. 全部这样的 $\theta$ 构成一集 $\Theta_1$，它是该指数族参数空间可能最大的扩充，称为该指数型分布族的**自然参数空间**. 以上所举诸例中的参数空间都是自然参数空间.

若按 $\mathrm{d}\nu=h\mathrm{d}\mu$ 引进测度 $\nu$，则(1.4)右边的 $h(x)\mathrm{d}\mu(x)$用$\mathrm{d}\nu(x)$取代，因此不失普遍性不妨设 $h\equiv1$. 保留这个 $h$ 的意义在于：在应用上常见的指数型分布族或是(对 $L$ 测度)绝对连续，或是离散的. 这时可取 $\mathrm{d}\mu=\mathrm{d}x$ 或 $\mu$ 为某可列集上的计数测度，这种取法就不可能排除函数 $h$. 在有些理论性的讨论中 $\mu$ 的具体形式不重要，这时取 $h\equiv1$ 是方便的. (1.4)的另一个有用的简化如下：取新参数 $\theta^*=(\theta_1^*,\cdots,\theta_k^*)$，其中 $\theta_i^*=Q_i(\theta)$，可以把(1.4)写为

$$\mathrm{d}F_{\theta^*}(x) = C^*(\theta^*)\exp\Big(\sum_{i=1}^k\theta_i^*T_i(x)\Big)h(x)\mathrm{d}\mu(x),$$

把 $\theta^*$ 和 $C^*$ 仍改回到 $\theta$ 和 $C$，得

$$\mathrm{d}F_\theta(x) = C(\theta)\exp\Big(\sum_{i=1}^k\theta_iT_i(x)\Big)h(x)\mathrm{d}\mu(x). \tag{1.8}$$

就是说，不失普遍性可设(1.4)中的 $Q_i(\theta)$就是 $\theta$ 的第 $i$ 分量 $\theta_i$. 当然，(1.4)和(1.8)中的 $\theta$ 的含义不同. 常把(1.8)称为指数型族的**自然形式**，通常在有关这个分布族的理论研究中，总是取这个自然

形式作为出发点. 这个形式,除了较简单之外,还有一个优点,即(1.8)式中的"自然参数"$\theta$是取实值,而在原形式(1.4)中则不必如此. 因此,自然形式下指数型族的自然参数空间$\Theta$是$\mathbf{R}^k$之一子集. 很容易证明:$\Theta$是一个凸集. 我们把这个事实的简单证明留给读者. 因此,若$\theta$为一维,则自然参数空间必是一个区间,这区间可以是有限或无限,开、闭或半开半闭都有可能(习题 32).

由于$\exp\left(\sum_{i=1}^{k}\theta_i T_i(x)\right) > 0$而$h$与$\theta$无关,按前面提到的判别法,指数型分布族必有共同支撑. 这是指数型族的一个很好的性质. 由此可知,均匀分布族$\{R(0,\theta):\theta>0\}$,以及一般的截尾型分布族,都不可能是指数型的. 另一方面,有共同支撑的分布族也可以不是指数型. Cauchy 分布$\{(1+(x-\theta)^2)^{-1}\pi^{-1}\mathrm{d}x, -\infty<\theta<\infty\}$就是一例(习题 39).

指数型族的另一个优点是其良好的分析性质,见于下述定理:

**定理 1.1** 设$a=(a_1,\cdots,a_k)$是指数型族(1.8)的参数空间$\Theta$的一个内点,则函数

$$H(\theta) \equiv (C(\theta))^{-1} = \int_{\mathscr{X}} \exp\left(\sum_{i=1}^{k}\theta_i T_i(x)\right)\mathrm{d}\nu(x), \quad \mathrm{d}\nu = h\mathrm{d}\mu \tag{1.9}$$

在$a$点连续,其在$a$点的任意阶偏导数皆存在且可以在积分号下求导.

**证明** 找$c>0$使$2^k$个点$(d_1,\cdots,d_k)=(a_1\pm c,\cdots,a_k\pm c)$都在$\Theta$内. 若$b=(b_1,\cdots,b_k)$而$|b_i|<c, 1\leqslant i\leqslant k$,则有

$$\begin{aligned}\exp((a_i+b_i)T_i(x)) \leqslant{}& \exp((a_i+c)T_i(x)) \\ &+ \exp((a_i-c)T_i(x)), \quad 1\leqslant i\leqslant k,\end{aligned}$$

于是有

$$\exp\left(\sum_{i=1}^{k}(a_i+b_i)T_i(x)\right) \leqslant \sum{}^{*}\exp\left(\sum_{i=1}^{k}d_i T_i(x)\right) \equiv H(x),$$

此处$\sum{}^{*}$表示对$(d_1,\cdots,d_k)$求和且遍及上述$2^k$个点. 按$c$的取法

知 $\int_{\mathscr{X}} H\mathrm{d}\nu < \infty$，因此由控制收敛定理知当 $b_i \to 0, 1 \leqslant i \leqslant k$ 时，有

$$\int_{\mathscr{X}} \exp\Big(\sum_{i=1}^{k}(a_i + b_i)T_i(x)\Big)\mathrm{d}\nu(x) \to \int_{\mathscr{X}} \exp\Big(\sum_{i=1}^{k} a_i T_i(x)\Big)\mathrm{d}\nu(x),$$

这证明了定理的前一结论.

为证后一结论，我们来讨论 $H$ 对 $\theta_1$ 的一阶偏导数. 这个场合的证法很容易推广到一般的高阶混合偏导数的情形(用归纳法，设 $\partial H(\theta)/\partial\theta_1^{r_1}\cdots\partial\theta_k^{r_k}$ 对 $r \leqslant m$ 时已证，去证明它对 $r = m+1$ 也对). 其次，在考虑对 $\theta_1$ 的偏导数时 $\theta_2, \cdots, \theta_k$ 之值保持不变，故为方便计且不失普遍性，不妨设 $k=1$，并分别以 $\theta$ 和 $a$ 记 $\theta_1$ 及 $a_1$，$T$ 记 $T_1$. 有

$$\begin{aligned}&(H(a+b) - H(a))/b \\ &= \int_{\mathscr{X}} \exp(aT(x))(\exp(bT(x)) - 1)b^{-1}\mathrm{d}\nu(x).\end{aligned}$$

取 $c>0$ 充分小，使 $a \pm 2c \in \Theta$. 设 $0<|b|<c$，有

$$|\exp(bT(x)) - 1|/|b| = |T(x)|\exp(\tilde{b}T(x)),\ |\tilde{b}| \leqslant |b| < c.$$

找常数 $M$，使 $|t| \leqslant M\mathrm{e}^{ct} + M\mathrm{e}^{-ct}$，对任何实数 $t$. 则由此与上式，并利用 $|\tilde{b}|<c$，有

$$\begin{aligned}&|\exp(bT(x)) - 1|/|b| \\ &\leqslant M\exp((\tilde{b}+c)T(x)) + M\exp((\tilde{b}-c)T(x)),\end{aligned}$$

因此有

$$\begin{aligned}&|\exp(aT(x))(\exp(bT(x)) - 1)/b| \\ &\leqslant M\exp((a+\tilde{b}+c)T(x)) + M\exp((a+\tilde{b}-c)T(x)) \\ &\leqslant M\exp((a+2c)T(x)) + 2M\exp(aT(x)) \\ &\quad + M\exp((a-2c)T(x)) \equiv H(x).\end{aligned}$$

由 $a \pm 2c \in \Theta$，知 $\int_{\mathscr{X}} H\mathrm{d}\nu < \infty$. 于是由控制收敛定理，

$$\begin{aligned}&\lim_{b\to 0}(H(a+b) - H(a))/b \\ &= \lim_{b\to 0}\int_{\mathscr{X}} \exp(aT(x))(\exp(bT(x)) - 1)b^{-1}\mathrm{d}\nu(x)\end{aligned}$$