统计检验的理论与方法

史宁中 主编

概率统计系列 1

研究生教学丛书

史宁中 /著

科学出版社
www.sciencep.com

中国科学院科学出版基金资助出版

概率统计系列研究生教学丛书 1

统计检验的理论与方法

史宁中 著

科 学 出 版 社

北 京

内 容 简 介

本书以测度论为基础，阐述假设检验的理论、思想和方法. 通过引入统计空间的概念，借助若干典型案例和数据(如美国法律判刑是否存在种族歧视、两种 Zucker 鼠的进食行为以及中国 1978～2002 年人均收入和消费支出等)，引出统计问题，激发统计思想，探索统计方法. 在内容方面，注意论述深度的把握和学科的发展，兼顾基础性与前沿性.

本书可用作统计专业高年级本科生和研究生的教材，也适合作为广大理工类科研人员的参考书.

图书在版编目(CIP)数据

统计检验的理论与方法/史宁中著. —北京：科学出版社，2008

(概率统计系列研究生教学丛书；1)

ISBN 978-7-03-020393-9

I. 统… II. 史… III. 数理统计-应用-质量管理 IV. F273.2

中国版本图书馆 CIP 数据核字(2008) 第 031062 号

责任编辑：陈玉琢 房 阳 / 责任校对：陈玉凤

责任印制：赵德静 / 封面设计：王 浩

科学出版社出版

北京东黄城根北街 16 号

邮政编码：100717

http://www.sciencep.com

铭浩彩色印装有限公司印刷

科学出版社发行 各地新华书店经销

*

2008 年 6 月第 一 版 开本：B5(720 × 1000)

2008 年 6 月第一次印刷 印张：15 1/2

印数：1—3 000 字数：292 000

定价：48.00 元

(如有印装质量问题，我社负责调换〈长虹〉)

《概率统计系列研究生教学丛书》序

概率论与数理统计是一门研究随机现象规律性的数学学科. 它一方面有自己独特的概念和方法, 形成了结构宏大的理论; 另一方面, 它与其他数学分支又有紧密的联系, 它是近代数学的重要组成部分. 在培养高素质科学技术人才中具有其独特的、不可替代的重要作用. 它不单是一种知识、方法或工具, 更在于它可以有效地培养和训练学生的随机思维模式、培养学生的一种素养.

大体上说, 概率论是统计学的理论和方法的依据, 而统计学可视为概率论的一种应用. 统计方法的应用促进了科学技术的进步; 反过来科技的进步推动了统计学突飞猛进的发展. 统计学的一些新方法应运而生, 比如 EM 法、GEE 方法、MCMC 方法、经验似然、贝叶斯网络、大维数据分析等. 而计算机技术和信息技术的飞速发展为数据分析的复杂化和多样化提供了强有力的平台, 过去许多不敢想像的方法成为可能, 如 Data Mining, Bootstrap 和 Jack-knife 等方法. 英国统计学家哈斯利特说:“统计方法的应用是这样普遍, 在我们的生活和习惯中, 统计的影响是这样巨大, 以致统计的重要性无论怎样强调也不过分.”

为了适应国内概率统计教学的现状以及社会对人才培养的需求, 并拓宽统计学应用的领域, 在科学出版社的大力支持下, 我们组织了一批专家编写了该系列适用于概率统计专业高年级本科生、研究生以及有关教师的教材 (教学参考书). 该丛书力求提高理论水平、突出前沿思想、侧重实际应用和学科渗透, 其中凝聚了该系列丛书作者的多年教学和科研经验.

我们衷心希望该系列丛书的出版能为我国高等院校教学改革作出贡献, 更希望能促进统计学在诸多领域的广泛应用.

史宁中

2008 年 5 月

于东北师范大学

前 言

假设检验是数理统计中一个非常重要的研究领域,是现代统计专业以及相关专业本科生和研究生的必修课程. 可惜的是,我国一直没有比较系统地阐述这方面内容的教科书. 自 1989 年回国以来,我曾给研究生开过几轮假设检验的课程,都是临时拼凑的讲义,深感需要一本相对成体系的教科书. 因此,在几轮课程的基础上,又花费了三年多的时间写出了本书,希望能够为学习、研究和应用假设检验的学生、教师或研究工作者提供一些帮助.

经典的假设检验是基于大样本的,如 Pearson 的卡方检验和 Fisher 的似然比检验,其判定准则依赖于 p 值. 现在通行的假设检验理论是基于 Neyman 与 Pearson 在 20 世纪 30 年代的思考,特别要归功于 Lehmann(1959, 1986) 的整理和发展. 后者的判定准则依赖于优化:控制检验统计量的第一类错误而使第二类错误尽可能地小. 几十年的教学和研究实践证明,后者在理论上是完美的,但是并没有更多地提供构建新的检验统计量的方法. 事实上,对于检验问题,还是要根据实际问题的背景,借助良好的统计直观给出合适的统计量,在此基础上再进一步分析检验统计量的性质. 因此,需要把上述两种方法有机地结合起来,这一点是本书所遵循的基本思路. 另外,当今假设检验的研究,大多数是针对参数模型的,这也是本书写作的重点.

全书分三个部分共计 5 章. 第一部分即第 1 章,讨论统计空间的一些基本性质. 虽然“统计空间”的名字是本书所给出的,但其概念却是长期通用的,即在可测空间加上概率测度族. 这样,统计空间与通常的概率空间就有了本质的不同:对于一个给定的可测集合,通常的概率空间关心的是度量“随机变量属于这个集合的可能性的大小”;统计空间关心的则是用哪一个概率测度来度量“随机变量属于这个集合的可能性的大小”更合适. 这一章讨论的统计空间的性质都是在这个基本想法下展开的. 这一章还总结了估计的一些方法和性质,因为作者认为检验在本质上是依赖于估计的.

第二部分包括第 2 章和第 3 章,讨论检验统计量的基本概念. 第 2 章侧重于构造检验统计量的基本思路和方法,包括参数和非参数的方法. 第 3 章讨论如何判断一个检验统计量的好坏, Neyman–Pearson 基本引理也被看作一种判断的方法. 除此之外,还有功效函数的比较、稳健性的比较等. 另外,基于检验的区间估计是非常重要的,第 3 章的最后部分讨论了检验接受域与区间估计的关系以及如何构造区间估计.

第三部分包括第 4 章和第 5 章,讨论参数模型的检验. 目前检验理论与方法的

研究大多数都是针对参数模型的. 事实上, 新的检验思想和检验方法也往往产生于此, 其中包括对于模型选择问题的研究. 这一部分结合真实的统计数据讨论了方差分析与 AIC 准则、对数线性模型与列联表、回归分析与 Logistic 模型、时间序列等模型的检验及模型选择问题. 特别是在方差分析的部分, 引入趋势性检验的内容, 这不仅仅是因为作者比较熟悉这方面的内容, 事实上, 随着对于生物信息学研究的逐渐深入, 趋势性的检验变得越来越重要.

本书的撰写过程得到郭建华博士、张宝学博士、高巍博士、陶剑博士、郑术蓉博士和胡果荣博士的帮助, 在此表示感谢.

史宁中

2007 年 8 月

目　录

第 1 章　统 计 空 间

数理统计研究随机现象的统计规律, 本质上, 是通过对随机变量 X 的一些观测值来进行研究的. 有时候, 也称这些观测值为样本, 通常用小写字母 x 来表示观测值的具体取值. 用 $\mathscr{X}$ 表示随机变量 X 的取值空间, 用 $\mathscr{A}$ 表示由 $\mathscr{X}$ 中的某些子集组成的 σ 代数. 可测空间 $(\mathscr{X},\mathscr{A})$ 是我们研究问题的出发点, 称为样本空间. 针对研究问题的需要, 通常有三种方式在可测空间 $(\mathscr{X},\mathscr{A})$ 上定义测度: $(\mathscr{X},\mathscr{A},\nu)$, 其中 ν 是一个 σ 有限测度, 称为**测度空间**; $(\mathscr{X},\mathscr{A},P)$, 其中 P 是一个概率测度, 称为**概率空间**; $(\mathscr{X},\mathscr{A},\mathscr{P})$, 其中 $\mathscr{P}$ 是一个概率测度族或者概率分布族, 为方便起见, 我们称之为**统计空间**. 在很多情况下, 概率分布族仅仅依赖概率分布中的参数:

$$\mathscr{P}=\{P_\theta:\theta\in\Theta\},$$

其中 Θ 为参数空间. 例如, 方差已知的正态分布族仅仅依赖均值参数, 这时参数 θ 是一个实数; 而当方差也未知时, 分布则依赖均值和方差两个参数, 这时候参数 θ 是一个二维向量. 统计检验的主要工作, 就是要通过样本来推断分布的参数是否等于某一个给定的参数 $H:\theta=\theta_0$, 其中 $\theta_0\in\Theta$; 或者是否属于某一个给定的参数空间的子集 $H:\theta\in\Theta_0$, 其中 $\Theta_0\subset\Theta$. 在讨论这个问题之前, 这一章要回顾一些以后要涉及的基本概念. 首先分析统计空间的一些性质, 然后讨论条件概率和充分统计量以及指数分布族的一些特征, 最后, 回顾估计的基本方法. 已经了解这些内容的读者可以越过本章.

1.1　统计空间的基本性质

上述三种度量空间, 在形式上看是一致的, 但其内涵却有着本质的不同: 对于 $A\in\mathscr{A}$, 测度空间的测度记为 $\nu(A)$, 表示集合 A 的大小; 概率空间的测度记为 $P(X\in A)$, 表示随机变量 X 取值于集合 A 的可能性的大小; 统计空间的测度记为 $P_\theta(X\in A)$, 是寻求用哪一个参数 θ 来度量这个可能性的大小最为合适. 关于 " 合适", 可以根据研究问题的不同给出不同的判断标准. 我们先讨论统计空间的一般性质.

1.1.1　统计空间的测度

令 $\mathscr{P}$ 和 ν 分别为定义在样本空间 $(\mathscr{X},\mathscr{A})$ 上的概率分布族和 σ 有限测度. 如

果

$$A \in \mathscr{A}, \nu(A) = 0 \quad \Rightarrow \quad P(A) = 0, \forall P \in \mathscr{P}, \tag{1.1.1}$$

则称 $\mathscr{P}$ 关于 ν 是绝对连续的, 记为 $\mathscr{P} \ll \nu$. 如果概率分布族仅仅依赖概率分布中的参数, 则 (1.1.1) 式又可以写为

$$A \in \mathscr{A}, \nu(A) = 0 \quad \Rightarrow \quad P_\theta(A) = 0, \forall \theta \in \Theta. \tag{1.1.2}$$

以后的讨论, 我们将把注意力集中在由 (1.1.2) 式表示的定义上. 由 Radon–Nikodym 定理 (Halmos, 1957), 如果 $\mathscr{P} \ll \nu$, 则对 $\forall \theta \in \Theta$, 存在 $\mathscr{A}$ 可测函数 f_θ, 对 $\forall A \in \mathscr{A}$, 有

$$P_\theta(A) = \int_A f_\theta(x) \mathrm{d}\nu(x). \tag{1.1.3}$$

并且在测度 ν 下, f_θ 是几乎处处唯一存在的. 这时称 f_θ 为 P_θ 的密度函数, 记为

$$\frac{\mathrm{d}P_\theta}{\mathrm{d}\nu} = f_\theta \quad 或 \quad \mathrm{d}P_\theta = f_\theta \mathrm{d}\nu.$$

关于几乎处处唯一性可以有下面两种等价的理解. 如果还存在 f_θ^*, 使得 $\mathrm{d}P_\theta = f_\theta^* \mathrm{d}\nu$, 则有

$$\nu(x; f_\theta(x) \neq f_\theta^*(x)) = 0,$$

或对 $\forall A \in \mathscr{A}$, 有

$$\int_A f_\theta(x) \mathrm{d}\nu(x) = \int_A f_\theta^*(x) \mathrm{d}\nu(x).$$

因此, 几乎处处不仅仅与测度 ν 有关, 也与 σ 代数 $\mathscr{A}$ 有关.

定理 1.1.1　令 $\mathscr{P}$ 和 ν 分别为定义在样本空间 $(\mathscr{X}, \mathscr{A})$ 上的概率分布族和 σ 有限测度. 令 g 为 $\mathscr{A}$ 可测函数. 如果 $\mathscr{P} \ll \nu$, 则对 $\forall \theta \in \Theta$, 有

$$\int_{\mathscr{X}} g(x) \mathrm{d}P_\theta(x) = \int_{\mathscr{X}} g(x) \frac{\mathrm{d}P_\theta}{\mathrm{d}\nu} \mathrm{d}\nu(x) = \int_{\mathscr{X}} g(x) f_\theta(x) \mathrm{d}\nu(x). \tag{1.1.4}$$

证明　由 (1.1.3) 式, 当 $g(x) = I_A(x), A \in \mathscr{A}$ 时, (1.1.4) 式显然成立. 当 g 为非负简单函数时, 即存在 $A_i \in \mathscr{A}, A_i \bigcap A_j = \phi (i \neq j)$ 和 $a_i > 0 (i = 1, \cdots, n)$ 使得

$$g(x) = a_1 I_{A_1}(x) + \cdots + a_n I_{A_n}(x),$$

则 (1.1.4) 式成立, 即

$$\sum_{i=1}^n a_i P_\theta(A_i) = \sum_{i=1}^n a_i \int_{\mathscr{X}} I_{A_i}(x) f_\theta(x) \mathrm{d}\nu(x).$$

如果 g 为非负可测函数, 则存在非负简单函数列 g_n, 满足

$$0 \leqslant g_1(x) \leqslant g_2(x) \leqslant \cdots, \quad g_n(x) \to g(x).$$

由单调收敛定理

$$\begin{aligned}\int_{\mathscr{X}} g(x)\mathrm{d}P_\theta(x) &= \lim_{n\to\infty}\int_{\mathscr{X}} g_n(x)\mathrm{d}P_\theta(x) \\ &= \lim_{n\to\infty}\int_{\mathscr{X}} g_n(x) f_\theta(x)\mathrm{d}\nu(x) \\ &= \int_{\mathscr{X}} g(x) f_\theta(x)\mathrm{d}\nu(x).\end{aligned}$$

当 g 为一般可测函数时, 可以把 g 分为正部和负部. 从而证明了定理. □

上面定理的证明手法通常称为**I 方法**. 这种方法说明在一个统计空间里, 对可测集合成立的命题往往对可测函数也成立. 通常称 (1.1.4) 式为当参数为 θ 时 $g(x)$ 的**均值**, 表示为

$$E_\theta g(X) = \int_{\mathscr{X}} g(x)\mathrm{d}P_\theta(x) = \int_{\mathscr{X}} g(x) f_\theta(x)\mathrm{d}\nu(x). \tag{1.1.5}$$

当 (1.1.5) 式有限时, 称

$$V_\theta g(X) = E_\theta(g(X) - E_\theta g(X))^2 \tag{1.1.6}$$

是参数为 θ 时 $g(x)$ 的**方差**. 特别当 $g(x) = x$ 时, 称 $E_\theta X$ 和 $V_\theta X$ 分别是参数为 θ 时随机变量 X 的均值和方差. 令 $h(x)$ 也为可测函数, 当 $E_\theta h(X)$ 有限时, 称

$$CV_\theta(g(X), h(X)) = E_\theta(g(X) - E_\theta g(X))(h(X) - E_\theta h(X)) \tag{1.1.7}$$

是参数为 θ 时 $g(X)$ 和 $h(X)$ 的**协方差**.

可以看到, 在统计空间中, 测度 ν 起着关键的作用. 通常我们只考虑两种形式的 ν, 一种是计数测度, 一种是 Lebesgue 测度.

令 $\mathbf{Z}_+$ 表示由零和正整数组成的集合, $\mathscr{A}_{\mathbf{Z}_+}$ 表示由 $\mathbf{Z}_+$ 的某些子集构成的 σ 代数, 则计数测度 ν 依赖于定义在 $(\mathbf{Z}_+, \mathscr{A}_{\mathbf{Z}_+})$ 上的示性函数, 即对 $A \in \mathscr{A}_{\mathbf{Z}_+}$,

$$\nu(A) = \sum_{x\in A} I_A(x),$$

这表示了 A 中的元素的个数. 令 $\mathscr{P}$ 为定义在 $(\mathbf{Z}_+, \mathscr{A}_{\mathbf{Z}_+})$ 上的概率分布族. 如果 $\mathscr{P} \ll \nu$, 其中 ν 为计数测度, 则称 $\mathscr{P}$ 为**离散型概率分布**.

例 1.1.1　二项分布 $Bi(n,\theta)$, 其中 $\theta\in(0,1)$. 令 $A=\{x;x\leqslant n,x\in\mathbf{Z}_+\}$, 则二项分布密度函数可以表示为

$$P_\theta(X=x)=\begin{pmatrix} n \\ x \end{pmatrix}\theta^x(1-\theta)^{n-x}I_A(x).$$

可以得到分布的均值和方差分别为

$$E_\theta X=n\theta,\quad V_\theta X=n\theta(1-\theta).$$

例 1.1.2　Poisson 分布 $P(\theta)$, 其中 $\theta>0$. 密度函数为

$$P_\theta(X=x)=\mathrm{e}^{-\theta}\frac{\theta^x}{x!}I_{\mathbf{Z}_+}(x).$$

分布的均值和方差分别为

$$E_\theta X=\theta,\quad V_\theta X=\theta.$$

例 1.1.3　多项分布 $M(n;\boldsymbol{\theta})$, 其中 $\boldsymbol{\theta}=(p_1,\cdots,p_k),p_i>0,\sum\limits_{i=1}^k p_i=1$. 令 $A=\{\boldsymbol{x}=(x_1,\cdots,x_k):x_i\in\mathbf{Z}_+,\sum\limits_{i=1}^k x_i=n\}$, 则密度函数为

$$P_\theta(X_1=x_1,\cdots,X_k=x_k)=\frac{n!}{x_1!\cdots x_k!}p_1^{x_1}\cdots p_k^{x_k}I_A(\boldsymbol{x}).$$

X_i 的均值和方差分别为

$$E_\theta X_i=np_i,\quad V_\theta X_i=np_i(1-p_i).$$

当 $i\neq j$ 时, X_i 和 X_j 协方差为

$$CV_\theta(X_i,X_j)=-np_ip_j.$$

令 $\mathbf{R}$ 表示实数空间, $\mathscr{B}$ 为由 $\mathbf{R}$ 中的某些子集构成的 Borel 代数. 则 Lebesgue 测度 ν 依赖于定义在 $(\mathbf{R},\mathscr{B})$ 上的区间长度, 即对半开区间 $(a,b]\in\mathscr{B}$,

$$\nu((a,b])=b-a.$$

令 $\mathscr{P}$ 为定义在 $(\mathbf{R},\mathscr{B})$ 上的概率分布族. 如果 $\mathscr{P}\ll\nu$, 其中 ν 为 Lebesgue 测度, 则称 $\mathscr{P}$ 为**连续型概率分布**.

例 1.1.4　均匀分布 $U(a,b)$, 其中 $a<b$. 密度函数为

$$f_\theta(x)=\begin{cases}\dfrac{1}{b-a}, & x\in[a,b],\\ 0, & \text{其他},\end{cases}$$

其中 $\theta=(a,b)$. X 的均值和方差分别为

$$E_\theta X=\frac{1}{2}(a+b),\quad V_\theta X=\frac{1}{12}(b-a)^2.$$

例 1.1.5 正态分布 $N(\mu,\sigma^2)$, 其中 $\sigma^2>0$. 密度函数为

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

其中 $\theta=(\mu,\sigma^2)$. X 的均值和方差分别为

$$E_\theta X=\mu,\quad V_\theta X=\sigma^2.$$

特别是当 $\mu=0,\sigma=1$ 时, 称之为标准正态分布, 密度函数为

$$f_\theta(x)=\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}x^2\right\}.$$

例 1.1.6 指数分布 $E(\mu,\sigma)$, 其中 $\sigma>0$. 密度函数为

$$f_\theta(x)=\begin{cases}\dfrac{1}{\sigma}\exp\left\{-\dfrac{1}{\sigma}(x-\mu)\right\}, & x\geqslant\mu,\\ 0, & \text{其他},\end{cases}$$

其中 $\theta=(\mu,\sigma)$. X 的均值和方差分别为

$$E_\theta X=\mu+\sigma,\quad V_\theta X=\sigma^2.$$

例 1.1.7 Cauchy 分布 $C(\mu,\sigma)$, 其中 $\sigma>0$. 密度函数为

$$f_\theta(x)=\frac{1}{\pi}\frac{\sigma}{\sigma^2+(x-\mu)^2},$$

其中 $\theta=(\mu,\sigma)$. Cauchy 分布的均值和方差均不存在.

例 1.1.8 χ^2 分布 $\chi^2(n)$ 与非心 χ^2 分布 $\chi^2(n,\alpha^2)$. 令 $X_1,\cdots,X_n$ 为相互独立的随机变量, $X_i\sim N(\mu_i,1),i=1,\cdots,n$. 令 $\alpha^2=\sum\limits_{i=1}^n\mu_i^2,X=\sum\limits_{i=1}^n X_i^2$, 则 X 的密度函数为

$$f_\theta(x)=\begin{cases}\displaystyle\sum_{s=0}^{\infty}\mathrm{e}^{-\frac{\alpha^2}{2}}\frac{\left(\frac{\alpha^2}{2}\right)^s}{s!}\frac{1}{2^{\frac{n}{2}+s}\Gamma\left(\frac{n}{2}+s\right)}x^{\frac{n}{2}+s-1}\mathrm{e}^{-\frac{x}{2}}, & x>0,\\ 0, & \text{其他},\end{cases}$$

其中 $\theta=(n,\alpha^2)$. 特别是当 $\alpha=0$ 时, 称之为自由度是 n 的 χ^2 分布, 密度函数为

$$f_\theta(x)=\begin{cases}\dfrac{1}{2^{\frac{n}{2}}\Gamma\left(\dfrac{n}{2}\right)}x^{\frac{n}{2}-1}\mathrm{e}^{-\frac{x}{2}}, & x>0,\\ 0, & \text{其他}.\end{cases}$$

例 1.1.9　t 分布 $t(n)$ 与非心 t 分布 $t(n,\alpha)$. 令 Y 与 Z 是相互独立的随机变量, 其中 $Y\sim N(\alpha,1), Z\sim\chi^2(n)$. 令 $X=Y/\sqrt{Z}$, 则 X 服从自由度为 n 的非心 t 分布, 密度函数为

$$f_\theta(x)=\sum_{s=0}^{\infty}\frac{\mathrm{e}^{-\frac{\alpha^2}{2}}}{s!}\frac{\Gamma\left(\dfrac{n+s+1}{2}\right)}{\sqrt{\pi}\Gamma\left(\dfrac{n}{2}\right)}\frac{(\sqrt{2}\alpha x)^s}{(1+x^2)^{\frac{1}{2}(n+s+1)}},$$

其中 $\theta=(n,\alpha)$. 特别是当 $\alpha=0$ 时, X 服从自由度为 n 的 t 分布, 密度函数为

$$f_\theta(x)=\frac{\Gamma\left(\dfrac{n+1}{2}\right)}{\sqrt{\pi}\Gamma\left(\dfrac{n}{2}\right)}(1+x^2)^{-\frac{n+1}{2}}.$$

例 1.1.10　F 分布 $F(m,n)$ 与非心 F 分布 $F(m,n,\alpha^2)$. 令 Y 与 Z 是相互独立的随机变量, 其中 $Y\sim\chi^2(m,\alpha^2), Z\sim\chi^2(n)$. 令 $X=Y/Z$, 则 X 服从自由度为 m 和 n 的非心 F 分布, 密度函数为

$$f_\theta(x)=\begin{cases}\displaystyle\sum_{s=0}^{\infty}\mathrm{e}^{-\frac{\alpha^2}{2}}\frac{\left(\dfrac{\alpha^2}{2}\right)^s}{s!}\frac{\Gamma\left(\dfrac{m+n}{2}+s\right)}{\Gamma\left(\dfrac{m}{2}+s\right)\Gamma\left(\dfrac{n}{2}\right)}\frac{x^{\frac{m}{2}+s-1}}{(1+x)^{\frac{m+n}{2}+s}}, & x>0,\\ 0, & \text{其他},\end{cases}$$

其中 $\theta=(m,n,\alpha^2)$. 特别是当 $\alpha=0$ 时, X 服从自由度为 m 和 n 的 F 分布, 分布函数为

$$f_\theta(x)=\begin{cases}\dfrac{\Gamma\left(\dfrac{m+n}{2}\right)}{\Gamma\left(\dfrac{m}{2}\right)\Gamma\left(\dfrac{n}{2}\right)}\dfrac{x^{\frac{m}{2}-1}}{(1+x)^{\frac{m+n}{2}}}, & x>0,\\ 0, & \text{其他}.\end{cases}$$

例 1.1.11　多维正态分布 $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$, 其中 $\boldsymbol{\mu}$ 是 n 维向量, $\boldsymbol{\Sigma}$ 为 $n\times n$ 的正定阵. 密度函数为

$$f_\theta(\boldsymbol{x})=\left(\frac{1}{2\pi}\right)^{\frac{n}{2}}\frac{1}{\sqrt{\det\boldsymbol{\Sigma}}}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\},$$

其中 $\theta=(\boldsymbol{\mu},\boldsymbol{\Sigma}), \boldsymbol{x}=(x_1,\cdots,x_n)', E_\theta\boldsymbol{X}=\boldsymbol{\mu}, \boldsymbol{\Sigma}=(\sigma_{ij}), \sigma_{ij}=E_\theta(X_i-\mu_i)(X_j-\mu_j)$.

例 1.1.12 Beta 分布 $Be(a,b)$, 其中 $a>0,b>0$. 密度函数为

$$f_\theta(x)=\begin{cases}\dfrac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1}, & 0\leqslant x\leqslant 1,\\ 0, & \text{其他},\end{cases}$$

X 的均值和方差分别为

$$E_\theta X=\frac{a}{a+b},\quad V_\theta X=\frac{ab}{(a+b+1)(a+b)^2}.$$

在上面的讨论中可以看到, 在统计空间中, 概率分布族关于一个 σ 有限测度的绝对连续性是很重要的, 它构建了依赖于参数的概率计算的基础. 那么, 对于一个给定的概率分布族, 如何判断是否存在一个 σ 有限测度, 使得这个概率分布族关于这个 σ 有限测度绝对连续呢? 这要涉及概率分布族的可分等概念.

1.1.2 概率分布族的对等与可分

讨论两个测度之间的关系. 令 ν 和 λ 是定义在样本空间 $(\mathscr{X},\mathscr{A})$ 上的两个 σ 有限测度, 如果 $\nu\ll\lambda$ 并且 $\lambda\ll\nu$, 则称 ν 和 λ 是**相互绝对连续**的. 在下面的定理中可以看到, 通常要求一个测度是 σ 有限的必要性.

定理 1.1.2 令 $(\mathscr{X},\mathscr{A},\nu)$ 是一个测度空间, 如果 ν 是 σ 有限的, 则存在一个概率测度与 ν 是相互绝对连续的.

证明 因为 $\mathscr{A}$ 是一个 σ 代数, ν 是 σ 有限的, 则存在一个集合列 $\{A_i\}$, 使得

$$A_i\in\mathscr{A},\quad \nu(A_i)<\infty,\quad A_1\subset A_2\subset\cdots,\quad \bigcup_{i=1}^{\infty}A_i=\mathscr{X}.$$

不失一般性, 令 $\nu(A_1)>0$. 对 $\forall A\in\mathscr{A}$, 令

$$\lambda(A)=\sum_{i=1}^{\infty}\frac{1}{2^i}\frac{\nu\left(A\bigcap A_i\right)}{\nu(A_i)}.$$

显然 λ 是一个概率测度, 并且有 $\nu\ll\lambda$ 和 $\lambda\ll\nu$. □

令 $\mathscr{P}$ 和 ν 分别为定义在样本空间 $(\mathscr{X},\mathscr{A})$ 上的概率分布族和 σ 有限测度. 如果 $\mathscr{P}\ll\nu$, 并且对 $\forall A\in\mathscr{A}$, 如果 $\forall\theta\in\Theta, P_\theta(A)=0$ 必有 $\nu(A)=0$, 则称 $\mathscr{P}$ 与 ν **是对等的**.

定理 1.1.3 令 $\mathscr{P}$ 为定义在概率空间 $(\mathscr{X},\mathscr{A})$ 上的概率分布族. $\mathscr{P}$ 关于某一个 σ 有限测度绝对连续的充分必要条件是, 存在与 $\mathscr{P}$ 对等的如下形式的概率测度:

$$\lambda=\sum_{i=1}^{\infty}c_iP_{\theta_i}, \tag{1.1.8}$$

其中, $\theta_i \in \Theta, c_i > 0, \sum_{i=1}^{\infty} c_i = 1$.

证明 充分性已由对等的条件给出, 下证必要性.

令 ν 是一个 σ 有限测度, 满足 $\mathscr{P} \ll \nu$. 由定理 1.1.2, 可以考虑 ν 是一个概率测度. 令

$$\mathscr{Q} = \left\{Q : Q = \sum_{i=1}^{\infty} c_i P_{\theta_i}, \theta_i \in \Theta, c_i > 0, \sum_{i=1}^{\infty} c_i = 1\right\},$$

则 $\mathscr{Q}$ 也是定义在 $(\mathscr{X}, \mathscr{A})$ 上的一个概率分布族, 并且有 $\mathscr{Q} \ll \nu$. 令

$$\sup_{Q \in \mathscr{Q}} \nu\left\{x : \frac{\mathrm{d}Q(x)}{\mathrm{d}\nu(x)} > 0\right\} = \alpha.$$

由上确界的定义, 存在 $\mathscr{Q}$ 中的概率分布列 $\{Q_n\}$, 使得

$$\lim_{n \to \infty} \nu\left\{x : \frac{\mathrm{d}Q_n(x)}{\mathrm{d}\nu(x)} > 0\right\} = \alpha. \tag{1.1.9}$$

进一步, 令 $\lambda = \sum_{n=1}^{\infty} \frac{1}{2^n} Q_n$, 则 $\lambda \in \mathscr{Q}$, 并且由 (1.1.9) 式,

$$\nu\left\{x : \frac{\mathrm{d}\lambda(x)}{\mathrm{d}\nu(x)} > 0\right\} = \alpha.$$

现在证明 λ 与 $\mathscr{P}$ 对等. 因为 $\lambda \in \mathscr{Q}$, 则对 $\forall A \in \mathscr{A}$, 如果对 $\forall \theta \in \Theta, P_\theta(A) = 0$, 则必然有 $\lambda(A) = 0$. 下面证明 $\mathscr{P} \ll \lambda$. 用反证法. 如果存在 $A \in \mathscr{A}, \theta \in \Theta$, 使得 $\lambda(A) = 0$ 并且 $P_\theta(A) > 0$. 令 $Q_0 = \frac{1}{2}(\lambda + P_\theta)$. 显然 $Q_0 \in \mathscr{Q}$. 由 $\lambda(A) = 0$ 知, 在 A 上 $\mathrm{d}\lambda(x)/\mathrm{d}\nu(x) = 0$ 几乎处处成立. 因此,

$$\begin{aligned}
\nu\left\{x : \frac{\mathrm{d}Q_0(x)}{\mathrm{d}\nu(x)} > 0\right\} &= \nu\left\{x : \frac{\mathrm{d}\lambda(x)}{\mathrm{d}\nu(x)} > 0 \quad \text{或} \quad \frac{\mathrm{d}P_\theta(x)}{\mathrm{d}\nu(x)} > 0\right\} \\
&\geqslant \nu\left\{x : \frac{\mathrm{d}\lambda(x)}{\mathrm{d}\nu(x)} > 0\right\} + \nu\left\{x \in A : \frac{\mathrm{d}P_\theta(x)}{\mathrm{d}\nu(x)} > 0\right\} \\
&> \nu\left\{x : \frac{\mathrm{d}\lambda(x)}{\mathrm{d}\nu(x)} > 0\right\} \\
&= \alpha.
\end{aligned}$$

这与 α 的定义矛盾. □

上面的定理虽然给出了一个判定条件, 但是条件本身却是很难验证的. 为此, 需要对概率分布族本身进行比较深入的分析. 下面先来回顾一下距离的定义.

令 d 是定义在集合 $\mathscr{X}$ 上的二元非负函数. 如果对 $\forall x, y, z \in \mathscr{X}$, 有

(1) $d(x,x)=0$,

(2) $d(x,y)=d(y,x)$,

(3) $d(x,z)\leqslant d(x,y)+d(y,z)$,

则称 d 为**拟距离**, $\mathscr{X}$ 为**拟距离空间**. 如果 (1) 加上条件 $d(x,y)=0$ 当且仅当 $x=y$, 则称 d 为**距离**, $\mathscr{X}$ 为**距离空间**.

例 1.1.13 令 $(\mathscr{X},\mathscr{A},P)$ 为一个概率空间. 对 $A_1,A_2\in\mathscr{A}$, 令

$$A_1\triangle A_2=(A_1-A_2)\bigcup(A_2-A_1),$$

$$d(A_1,A_2)=P(A_1\triangle A_2),$$

则 σ 代数 $\mathscr{A}$ 对 d 构成一个拟距离空间.

例 1.1.14 令 $(\mathscr{X},\mathscr{A},\mathscr{P})$ 为一个统计空间, 对 $\theta_1,\theta_2\in\Theta$, 令

$$d(\theta_1,\theta_2)=\sup_{A\in\mathscr{A}}|P_{\theta_1}(A)-P_{\theta_2}(A)|,$$

则概率分布族 $\mathscr{P}$ 对 d 构成一个拟距离空间. 如果 θ 与 P_θ 是一一对应的, 则 $\mathscr{P}$ 对 d 构成一个距离空间.

例 1.1.15 令 $\mathscr{P}$ 和 ν 分别为定义在样本空间 $(\mathscr{X},\mathscr{A})$ 上的概率分布族和 σ 有限测度, $\mathscr{P}\ll\nu$. 令 $\mathscr{F}$ 为由所有密度函数构成的集合, 即

$$\mathscr{F}=\left\{f_\theta:f_\theta=\frac{\mathrm{d}P_\theta}{\mathrm{d}\nu},\theta\in\Theta\right\}.$$

对 $f_{\theta_1},f_{\theta_2}\in\mathscr{F}$, 令

$$d^*(f_{\theta_1},f_{\theta_2})=\int\limits_{\mathscr{X}}|f_{\theta_1}(x)-f_{\theta_2}(x)|\mathrm{d}\nu(x),$$

则 $\mathscr{F}$ 对 d^* 构成拟距离空间, 并且几乎处处构成距离空间, 即

$$d^*(f_{\theta_1},f_{\theta_2})=0\quad\Leftrightarrow\quad f_{\theta_1}(x)=f_{\theta_2}(x),\quad \text{a.e.}\nu.$$

可以验证, 如果 θ 与 P_θ 是一一对应的, 则例 1.1.14 和例 1.1.15 定义的距离之间存在一个等价关系, 即对 $\theta_1,\theta_2\in\Theta$, 有

$$\begin{aligned}d(\theta_1,\theta_2)&=\sup_{A\in\mathscr{A}}|P_{\theta_1}(A)-P_{\theta_2}(A)|\\&=\frac{1}{2}\int\limits_{\mathscr{X}}|f_{\theta_1}(x)-f_{\theta_2}(x)|\mathrm{d}\nu(x)\\&=\frac{1}{2}d^*(f_{\theta_1},f_{\theta_2}),\end{aligned}$$