□ 应用统计学丛书

Randomized Estimation and VDR Test

随机估计及 VDR 检验

杨振海

高等教育出版社
HIGHER EDUCATION PRESS

□ 应用统计学丛书

Randomized Estimation and VDR Test

随机估计及 VDR 检验

杨振海

SUIJI GUJI JI VDR JIANYAN

高等教育出版社·北京
HIGHER EDUCATION PRESS BEIJING

图书在版编目（CIP）数据

随机估计及 VDR 检验 / 杨振海著. -- 北京: 高等教育出版社，2014. 1

ISBN 978-7-04-038672-1

Ⅰ. ①随… Ⅱ. ①杨… Ⅲ. ①概率密度函数 Ⅳ. ① O211.3

中国版本图书馆 CIP 数据核字（2013）第 256401 号

策划编辑 王丽萍 责任编辑 李华英 封面设计 姜 磊 版式设计 童 丹
责任校对 胡晓琪 责任印制 尤 静

出版发行	高等教育出版社	咨询电话	400-810-0598
社 址	北京市西城区德外大街4号	网 址	http://www.hep.edu.cn
邮政编码	100120		http://www.hep.com.cn
印 刷	北京四季青印刷厂	网上订购	http://www.landraco.com
开 本	787mm×1092mm 1/16		http://www.landraco.com.cn
印 张	15	版 次	2014 年 1 月第 1 版
字 数	300 千字	印 次	2014 年 1 月第 1 次印刷
购书热线	010-58581118	定 价	59.00 元

本书如有缺页、倒页、脱页等质量问题，请到所购图书销售部门联系调换
版权所有 侵权必究
物 料 号 38672-00

随机估计及 VDR 检验

前言

这本小书包括了作者 10 余年来关于 VDR (Vertical Density Representation) 研究的一些成果及其应用, 主要是 VDR 在假设检验中的应用. 2001 年作者去香港浸会大学访问方开泰教授, 在做有约束的混料设计研究中涉及 VDR. 1991 年 Troutt 首先提出 VDR, 给出了最基本结果. 众所周知, 一维随机变量 X 的分布函数 $F(\cdot)$ 连续时, $F(X)$ 的分布是 $[0,1]$ 上的均匀分布. 若它有连续概率密度函数 $f(\cdot)$, 那么 $f(X)$ 的分布是什么? 若随机向量 $\mathbf{X}$ 有连续概率密度函数 $f(\cdot)$, 那么 $f(\mathbf{X})$ 的分布是什么? 这个问题就不容易回答了. 1991 年 Troutt 回答了该问题, 将此称为 VDR. 详见第四章. 作者和方开泰教授等人将 Troutt 的结果称为 I 型 VDR. 我们提出了 II 型 VDR. 应用 II 型 VDR 很容易导出 Troutt 的结果. 作者又和彭运佳教授等人合作, 应用 II 型 VDR 给出了给定 $f(\mathbf{X}) = v$ 的条件下 $\mathbf{X}$ 的条件密度函数, 完善了 I 型 VDR. VDR 的直接应用是生成非均匀随机数, 我们得到了生成非均匀随机数算法, 只要知道密度函数就可生成它的随机数, 生成速度是可控制的精确算法. 另一应用是构造多元分布密度函数, 作者又和学生戴家佳等人提出中心相似分布, 给出了参数的矩估计. 中心相似分布是构造本书讨论的线性变换分布族的基础.

作者一直苦于找不到 VDR 在统计中的应用. 2010 年下半年在回北京的飞机上作者和何书元教授讨论有关问题, 想到 VDR 可以应用到假设检验, 何书元教授给予积极支持. 作者发现 VDR 检验是通用方法, 可以应用到很多问题, 可以应用到一维参数检验, 也可以应用到多维参数检验. VDR 检验是基于随机估计的概率密度函数构造的. 随机估计源于 Fisher 的信仰推断, 是经典统计理论体系下的概念. Singh, Xie 和 Strawderman (2007) 讨论了一维参数的随机估计. 在给定参数空间上的概率分布条件下, 由 VDR 检验确定的参数置信域具有最小 Lebesgue 测度. 将其应用于一维参数的检验和置信区间, 或得到经典已知结果, 或给出改进结果. 如正态总体期望参数的 t 检验, 可由 VDR 检验导出. 可以得到方差的置信区间最短的两端概率的分配原则. 应用到多元正态总体均值向量参数检验得到大家熟知的 Hotelling 检验.

此为试读, 需要完整PDF请访问: www.ertongbook.com

应用到协方差阵的检验, 得到基于 Wishart 分布的精确的 VDR 检验. 实际上是将一维正态总体结果推广到多维正态总体. VDR 检验可以应用于很广的线性变换分布族, 它是正态分布族的推广. 多元正态分布族是标准正态分布经线性变换得到的分布族. 将标准正态分布换成已知的中心相似分布, 得到异于正态分布族的分布族, 其分布参数就是线性变换的参数. VDR 检验可以实现这些参数的检验. 展示了非正态多元统计分析前景和实际应用前景. 很有可能一类实际问题可以由一特定中心相似分布形成的线性变换分布族描述. 将多元正态分布密度函数中分量平方和换成绝对值的正数幂和就得到一类中心相似分布, 为广义球生成的分布. 将 VDR 检验应用于回归分析的参数检验, 当误差项是正态分布时得到经典结果, 当误差项是已知的刻度参数分布族时, 给出参数的精确检验. 非正态误差回归模型是有实际应用价值的. 也期望应用到非正态误差自回归模型.

特别感谢韦博成教授、林金官教授和濮晓龙教授, 作者多次去东南大学、华东师范大学进行交流, 受益匪浅, 尤其是韦博成教授提出了许多宝贵意见, 本书许多章节是按他的意见修改的. 感谢赵林城教授和胡太忠教授, 提供了 Xie Mingge 在中国科学技术大学关于 CD 的演讲稿. 感谢张忠占教授和程维虎教授对本书写作的支持.

作者于北京

2013 年 9 月 14 日

符号说明

基本原则: 白体表示标量, 黑体表示向量; 小写表示常量, 同名大写表示随机变量. 如 s_n^2 是样本方差, 是常数; S_n^2 是随机样本方差, 是随机变量. a 表示常数, $\mathbf{a}$ 表示常数向量.

$\mathbf{1}_p$ 是各分量均为 1 的 p 维列向量, $\bar{\mathbf{1}} = \dfrac{1}{\sqrt{p}}\mathbf{1}$; $\mathbf{0}_p$ 是各分量均为 0 的 p 维列向量.

I_p 是 $p \times p$ 单位阵.

由定义在欧氏空间 $\mathfrak{R}^p$ 上的非负函数 f 定义以下集合:

$$D_{[f]} = \{\mathbf{x}_{p+1} = (\mathbf{x}', x_{p+1})' : 0 < x_{p+1} \leqslant f(\mathbf{x}), \mathbf{x} \in \mathfrak{R}^p,\ x_{p+1} \in \mathfrak{R}\} \subset \mathfrak{R}^{p+1};$$

$$D_{[f]}(v) = \{\mathbf{x} : f(\mathbf{x}) \geqslant v, \mathbf{x} \in \mathfrak{R}^p\} \subset \mathfrak{R}^p, v > 0;$$

$$D_{[f]}(0+) = \{\mathbf{x} : f(\mathbf{x}) > 0, \mathbf{x} \in \mathfrak{R}^p\} = \bigcup_{n=1}^{\infty} D_{[f]}\left(\frac{1}{n}\right) = \bigcup_{v>0} D_{[f]}(v).$$

$D_{[f]}(0+)$ 是非负函数 f 的支撑. $D_{[f]}$ 也可以表示为

$$D_{[f]} = \{\mathbf{x}_{p+1} = (\mathbf{x}', x_{p+1})' : 0 < x_{p+1} \leqslant f(\mathbf{x}), \mathbf{x} \in D_{[f]}(0+),\ x_{p+1} \in \mathfrak{R}\}.$$

“$\stackrel{d}{=}$” 表示两端随机向量有相同的分布函数.

函数自变量列表中 “;” 右端为分布参数, “;” 仅在分位数、密度函数、分布函数和枢轴量自变量列表中出现. 如参数 $\boldsymbol{\eta}$ 的随机估计 W 的概率密度函数 $f_{\boldsymbol{\eta}}(\cdot; \bar{x}, s_n^2)$, $\bar{x}, s_n^2$ 是随机估计的分布参数.

向量相除: $\dfrac{\mathbf{a}}{\mathbf{b}} = \dfrac{(a_1, \cdots, a_p)'}{(b_1, \cdots, b_p)'} = \left(\dfrac{a_1}{b_1}, \cdots, \dfrac{a_p}{b_p}\right)'$, 即对应分量相除.

向量相乘: $\mathbf{ab} = (a_1, \cdots, a_p)'(b_1, \cdots, b_p)' = (a_1 b_1, \cdots, a_p b_p)'$. 向量相乘、向量相除运算级别最高, 系列运算时最先作向量相乘和向量相除运算.

$$\sqrt{\mathbf{a}} = \sqrt{(a_1, \cdots, a_p)'} = (\sqrt{a_1}, \cdots, \sqrt{a_p})';$$

$$\mathbf{a}^b = ((a_1, \cdots, a_p)')^b = (a_1^b, \cdots, a_p^b)'.$$

A_p 是 A 的前 p 个分量, A 可以是向量或向量等式. 若 A 是矩阵, A_p 是 A 的前 p 行.

记

$$\begin{aligned}\mathbf{L}(\mathbf{x}) &= \mathbf{L}_{[f]}(\mathbf{x}) = (L_1(\mathbf{x}), \cdots, L_p(\mathbf{x}))' \\ &= -\frac{1}{f(\mathbf{x})}\frac{\partial}{\partial \mathbf{x}}f(\mathbf{x}) = -\left(\frac{1}{f(\mathbf{x})}\frac{\partial}{\partial x_1}f(\mathbf{x}), \cdots, \frac{1}{f(\mathbf{x})}\frac{\partial}{\partial x_p}f(\mathbf{x})\right)',\end{aligned}$$

当不引起误解时省略下标记为 $\mathbf{L}(\cdot)$, 如一维密度函数记作 $L_{[f]}(\cdot)$, 高维密度函数记作 $\mathbf{L}(\cdot)$.

$\mathbf{W}_{\cdot}$ 是下标参数的随机估计, 下标表示 "某参数的", 如 W_μ 是参数 μ 的随机估计, $\mathbf{W}_{\boldsymbol{\theta}}$ 是参数 $\boldsymbol{\theta}$ 的随机估计. 当不引起混淆时省略下标, 用 $\mathbf{W}$ 表示参数的随机估计.

$Q_Z(\gamma;\mathcal{X})$ 表示检验变量 Z 的 γ 分位点, $\mathcal{X}$ 是样本, 是随机估计的分布参数. $Q_V(\gamma)$ 表示 $V = f_h(\mathbf{T})$ 的 γ 分位点, $f_h(\cdot)$ 是枢轴量的分布密度函数, $\mathbf{T}$ 是随机向量, 它的密度函数是 $f_h(\cdot)$. $Q_F(\alpha)$ 是一元分布函数 $F(\cdot)$ 的分位点.

$pchi(\cdot, k)$ 是自由度为 k 的 χ^2 分布的分布函数, 其密度函数是 $dchi(\cdot, k)$; $pt(\cdot, k)$ 是自由度为 k 的 t 分布的分布函数, 其密度函数是 $dt(\cdot, k)$. 其他分布有类似表示.

目录

此为试读,需要完整PDF请访问: www.ertongbook.com

第一章

引言

本章简要介绍本书的基本内容. 主要是 Vertical Density Representation (VDR) 基础理论及其在构造多元概率密度函数的应用, 枢轴量、随机估计和 VDR 检验的基本思想.

1.1 VDR 理论和构造多元概率密度函数

1.1.1 什么是 VDR

1991 年 Troutt 首先提出 VDR 概念, 并给出了最基本结果. 众所周知, 当一维随机变量 X 的分布函数 $F(\cdot)$ 连续时, $F(X)$ 的分布是 $[0,1]$ 上的均匀分布. 若它有连续概率密度函数 $f(\cdot)$, 那么 $f(X)$ 的分布是什么? 若 p 维随机向量 $\mathbf{X}$ 有连续概率密度函数 $f(\cdot)$, 那么 $f(\mathbf{X})$ 的分布是什么? 初次遇到这个问题往往不知如何解决. 1991 年 Troutt 回答了该问题, 将此称为 VDR. 设 $\mathbf{X}$ 是 $\Re^p$ 上的随机向量, 其概率密度函数是 $f(\cdot)$. 若随机变量 $V = f(\mathbf{X})$ 有概率密度函数 $g(\cdot)$, Troutt 证明了

$$g(v) = -v\frac{dL_p(D_{[f]}(v))}{dv},$$

$$D_{[f]}(v) = \{\mathbf{x} : f(\mathbf{x}) \geqslant v, \quad \mathbf{x} \in \Re^p\},$$

其中 $L_p(\cdot)$ 是 $\Re^p$ 上的 Lebesgue 测度. 上式成立的条件是 $L_p(D_{[f]}(v))$ 在 $\Re^+$ 上连续可微. $D_{[f]}(v)$ 有清晰的几何意义. 概率密度函数 $f(\cdot)$ 的图像是空间 $\Re^{p+1}$ 上的 p 维曲面 $\bar{D}_{[f]} = \{(\mathbf{x}', x_{p+1})' : x_{p+1} = f(\mathbf{x}), \mathbf{x} \in \Re^p\}$. 超平面 $= \{(\mathbf{x}', x_{p+1})' : x_{p+1} = h\}$ 记作 $L(h)$. $D_{[f]}$ 是 $\bar{D}_{[f]}$ 和超平面 $L(0)$ 围成的区域. $D_{[f]}(v)$ 是 $L(v) \bigcap D_{[f]}$ 在 $\Re^p$ 上的投影. 详见第四章. 方开泰和杨振海等 (2001) 将 Troutt 的结果称为 I 型 VDR, 并提

出了基于几何概率的 II 型 VDR. 在 $\mathfrak{R}^{p+1}$ 中集合 $D_{[f]}$ 表示为

$$D_{[f]} = \{(\mathbf{x}', x_{p+1})' : 0 < x_{p+1} \leqslant f(\mathbf{x}), x_{p+1} \in \mathfrak{R}, \mathbf{x} \in \mathfrak{R}^p\},$$

$$L_{p+1}(D_{[f]}) = \int_{\mathfrak{R}^p} f(\mathbf{x}) d\mathbf{x} = 1.$$

II 型 VDR 的基本内容是:

随机向量 $(\mathbf{X}', X_{p+1})'$ 在 $D_{[f]}$ 上均匀分布的充要条件是 X_{p+1} 的概率密度函数是

$$f_{p+1}(v) = \begin{cases} L_p(D_{[f]}(v)), & 0 < v \leqslant f_0 = \sup\limits_{\mathbf{x} \in \mathfrak{R}^p} f(\mathbf{x}), \\ 0, & \text{其他}, \end{cases}$$

及给定 $X_{p+1} = v$ 的条件下 $\mathbf{X}$ 的条件分布是 $D_{[f]}(v)$ 上的均匀分布, 等价于 $\mathbf{X}$ 的概率密度函数是 $f(\cdot)$, 给定 $\mathbf{X} = \mathbf{x}$ 的条件下 X_{p+1} 的条件分布是 $(0, f(\mathbf{x})]$ 上的均匀分布.

均匀分布于 $D_{[f]}$ 上的随机向量 $(\mathbf{X}', X_{p+1})'$ 的密度函数记作 $f_{D_{[f]}}(\cdot)$, 则 II 型 VDR 的基本内容表达为

$$f_{D_{[f]}}(\mathbf{x}_p, x_{p+1}) = f_{p+1}(x_{p+1}) f_{\vec{p}}(\mathbf{x}_p | x_{p+1}) = f(\mathbf{x}_p) f_{p+1}(x_{p+1} | \mathbf{x}_p),$$

其中 $f_{p+1}(\cdot)$ 是 X_{p+1} 的边缘密度函数, $f_{p+1}(x_{p+1}|\mathbf{x}_p)$ 是给定条件 $\mathbf{X}_p = \mathbf{x}_p$ 下 X_{p+1} 的条件密度函数, $f_{\vec{p}}(\mathbf{x}_p|x_{p+1})$ 是给定条件 $X_{p+1} = x_{p+1}$ 下 $\mathbf{X}_p$ 的条件密度函数. 计算得

$$\begin{aligned} f_{\vec{p}}(\mathbf{x}|X_{p+1} = x) &= \begin{cases} \dfrac{1}{L_p(D_{[f]}(x))}, & \text{若 } f(\mathbf{x}) \geqslant x, \\ 0, & \text{若 } f(\mathbf{x}) < x \end{cases} \\ &= \frac{I_{D_{[f]}}(\mathbf{x}_p)}{L_p(D_{[f]}(x))}. \end{aligned}$$

应用 II 型 VDR 可简洁地导出 Troutt 的结果. 事实上,

$$\begin{aligned} &\{(\mathbf{x}', x_{p+1})' : f(\mathbf{x}) \leqslant v, (\mathbf{x}', x_{p+1})' \in D_{[f]}\} \\ &= \{(\mathbf{x}', x_{p+1})' : x_{p+1} \leqslant v, (\mathbf{x}', x_{p+1})' \in D_{[f]}\} \setminus \{D_{[f]}(v) \times (0, v]\}, \end{aligned}$$

于是

$$\begin{aligned} P(f(\mathbf{X}) \leqslant v) &= L_{p+1}(\{(\mathbf{x}', x_{p+1})' : f(\mathbf{x}) \leqslant v, (\mathbf{x}', x_{p+1})' \in D_{[f]}\}) \\ &= L_{p+1}(\{(\mathbf{x}', x_{p+1})' : x_{p+1} \leqslant v, (\mathbf{x}', x_{p+1})' \in D_{[f]}\}) - L_{p+1}(D_{[f]} \times (0, v]) \\ &= \int_0^v L_p(D_{[f]}(u)) du - v L_p(D_{[f]}(v)). \end{aligned}$$

两端对 v 求导

$$\begin{aligned} g(v) &= \frac{d}{dv} P(f(\mathbf{X}) \leqslant v) \\ &= L_p(D_{[f]}(v)) - L_p(D_{[f]}(v)) - v \frac{dL_p(D_{[f]}(v))}{dv} \\ &= -v \frac{dL_p(D_{[f]}(v))}{dv}, \end{aligned}$$

恰是 Troutt 的结果. 但是 Troutt 没有给出在给定 $f(\mathbf{X})=v$ 的条件下 $\mathbf{X}$ 的条件密度函数. 彭运佳和杨振海等 (2001) 应用 II 型 VDR 给出了给定 $f(\mathbf{X})=v$ 的条件下 $\mathbf{X}$ 的条件密度函数 $f(\mathbf{x}|v)$, 完成了 I 型 VDR:

$$f(\mathbf{x})=\int_0^{f_0} f(\mathbf{x}|v)g(v)dv, \quad f_0=\sup_{\mathbf{x}} f(\mathbf{x}).$$

VDR 的直接应用是生成非均匀随机数, 彭运佳和杨振海等 (2002) 得到了生成非均匀随机数的通用算法, 只要知道密度函数就可生成它的随机数, 是生成速度可控制的精确算法.

$D_{[f]}$ 可以更准确地定义, 使上述结论适用性更广泛. 令

$$D_{[f]}(0+)=\bigcup_{v>0} D_{[f]}(v)=\{\mathbf{x}: f(\mathbf{x})>0\},$$

$$D_{[f]}=\{(\mathbf{x}_p, x_{p+1})': \mathbf{x}\in D_{[f]}(0+), 0<x_{p+1}\leqslant f(x)\}.$$

$D_{[f]}(0+)$ 是 $f(\cdot)$ 的支撑, VDR 理论是就 $L_p(D_{[f]}(0+))>0$ 的情形论述的. 但是当 $L_p(D_{[f]}(0+))=0$ 时 VDR 理论仍然成立. 如果 $L_p(D_{[f]}(0+))=0, L_{p-1}(D_{[f]}(0+))>0$, 在前面论述中将 $L_p(\cdot)$ 换成 $L_{p-1}(\cdot)$ 即可. Dirichlet 分布就是这种情形. 详见第四章.

1.1.2 多元概率密度函数的结构

II 型 VDR 的重要作用在于给出了构造多元概率密度函数的方法. 若 $f(\cdot)$ 是 $\Re^p$ 上的密度函数, 则由 $f(\cdot)$ 定义了集族 $\mathfrak{P}_{[f]}=\{D_{[f]}(v), 0<v\leqslant f_0\}$, $(0, f_0]$ 是集族的参数空间. $L_p(D_{[f]}(v))$ 恰是集族参数空间上的概率密度函数. II 型 VDR 断言

$$\mathbf{X}\sim f(\mathbf{x}), \mathbf{x}\in\Re^p, \quad f(\mathbf{x})\geqslant 0, \quad \int_{\Re^p} f(\mathbf{x})d\mathbf{x}=1,$$

等价于

$$\mathbf{X}\sim U(D_{[f]}(X_{p+1})), \quad X_{p+1}\sim L_p(D_{[f]}(v)), \quad 0<v\leqslant f_0, \tag{1.1}$$

这里 $U(A)$ 表示集合 A 上的均匀分布. 陈述为任何有连续密度函数的随机向量可表示为随机集族上的均匀分布, 也是生成随机向量的方法. 首先生成有密度 $L_p(D_{[f]}(v))$, $0<v\leqslant f_0$ 的随机数 X_{p+1}, 再生成随机向量 $\mathbf{U}\sim U(D_{[f]}(X_{p+1}))$, 输出 $\mathbf{X}=\mathbf{U}$, 则 $\mathbf{X}\sim f(\cdot)$. 这个过程分为两步:

第一步由 $f(\cdot)$ 生成集族 $\mathfrak{P}_{[f]}$ 和其参数空间 $(0, f_0]$ 上的概率密度函数 $L_p(D_{[f]}(\cdot))$;

第二步由集族 $\mathfrak{P}_{[f]}$ 和其参数空间 $(0, f_0]$ 上的概率密度函数 $L_p(D_{[f]}(\cdot))$ 生成随机向量 $\mathbf{X}$. 显然这步是生成随机向量的方法, 集族和其参数空间上的概率密度函数可任意给定.

现将集族 $\mathfrak{P}_{[f]}$ 的参数空间标准化. 令

$$R=-\ln\left(\frac{X_{p+1}}{f_0}\right), \quad 0\leqslant R<\infty,$$

则

$$R \sim g(r) = L_p(D_{[f]}(f_0 e^{-r})) f_0 e^{-r}, \quad g(r) > 0, r \in \mathfrak{R}^+, \quad \int_0^{\infty} g(r) dr = 1.$$

$\mathfrak{P}_{[f]} = \{A(r) : A(r) = D_{[f]}(f_0 e^{-r}), L(A(r)) > 0, r \in \mathfrak{R}^+\}$ 是以 $\mathfrak{R}^+ = \{r : r > 0\}$ 为参数的集族, $g(\cdot)$ 是 $\mathfrak{R}^+$ 上的概率密度函数, 则

$$\mathbf{X} \sim U(A(R)), \quad R \sim g(\cdot). \tag{1.2}$$

作为构造多元随机向量的方法, 首先构造以 $\mathfrak{R}^+$ 为参数的一集族

$$\mathfrak{P} = \{A(r) : A(r) \subset \mathfrak{R}^p, L_p(A(r)) > 0, \forall r \in \mathfrak{R}^+\},$$

并确定 $\mathfrak{R}^+$ 上的概率密度函数 $g(\cdot)$, 则随机向量

$$\mathbf{Y} = U(A(R)), \quad R \sim g(\cdot)$$

的密度函数是

$$p(\mathbf{y}) = \int_{a(\mathbf{y})} \frac{g(r)}{L_p(A(r))} dr, \tag{1.3}$$

其中

$$a(\mathbf{y}) = \{r : \mathbf{y} \in A(r), r \in \mathfrak{R}^+\}.$$

因为在给定 $R = r$ 的条件下 $\mathbf{Y}$ 的条件概率密度函数是

$$p(\mathbf{y}|R = r) = \begin{cases} \dfrac{1}{L_p(A(r))}, & 若\ \mathbf{y} \in A(r), \\ 0, & 若\ \mathbf{y} \notin A(r). \end{cases}$$

$\mathfrak{P}$ 和 $g(\cdot)$ 可任意选取. 当集族 $\mathfrak{P}$ 满足某些特性时就产生特定分布族.

若集族 $\mathfrak{P}$ 是单调的, 即

$$A(r) \subseteq A(r'), \quad \forall r < r',$$

则 (1.3) 成为

$$\begin{aligned} p(\mathbf{y}) &= \int_{b(\mathbf{y})}^{\infty} \frac{g(r)}{L_p(A(r))} dr, \\ b(\mathbf{y}) &= \inf\{r : \mathbf{y} \in A(r), r \in \mathfrak{R}^+\}. \end{aligned} \tag{1.4}$$

选取特定集族和 $\mathfrak{R}^+$ 上的概率密度函数就得到常见分布.

若集族 $\mathfrak{P}$ 是相似的, 即

$$A(r) = r \cdot A(1) = \{\mathbf{y} : \mathbf{y} = r\mathbf{z}, \mathbf{z} \in A(1)\}, \quad \forall r > 0,$$

则 $\mathbf{Y}$ 的分布是中心相似分布. 杨振海和 Kotz (2003) 基于 VDR 提出了中心相似分布的概念, 杨振海和戴家佳等 (2004) 给出了中心相似分布参数的矩估计. 中心相似分布是多元正态分布的推广, 取作本书讨论的线性变换分布族的基础分布. 如前所述, 用超平面 $L(v)$ 截 p 维密度函数 $f(\cdot)$ 的图像和 $L(0)$ 围成的集合 $D_{[f]}$, 截口在 $\mathfrak{R}^p$ 上的投影就是 $D_{[f]}(v)$. 当 $D_{[f]}(v), v > 0$ 是相似集族, $f(\cdot)$ 是中心相似的. 当 $D_{[f]}(v), v > 0$ 是球体集族, $f(\cdot)$ 是球对称分布密度函数. 再若 $g(\cdot)$ 是自由度为 $p+2$ 的 χ^2 分布的密度函数, $f(\cdot)$ 是 p 维标准正态分布密度函数, 即 p 维标准正态分布向

量 $\mathbf{X}$ 可表示为

$$\mathbf{X} = R\mathbf{V}, \quad R \sim pchi(\cdot, p+2), \quad \mathbf{V} \sim U(S_p^2), \text{ 且 } R \text{ 与 } \mathbf{V} \text{ 相互独立}, \tag{1.5}$$

其中 $pchi(\cdot, k)$ 是自由度为 k 的 χ^2 分布的分布函数, S_p^2 和 $\bar{S}_p^2$ 分别是 p 维单位球和单位球面. 这和已知结论是一致的. 事实上, 众所周知 p 维标准正态分布向量 $\mathbf{X}$ 通常可表示为

$$\mathbf{X} = R^*\mathbf{U}, \quad R \sim pchi(\cdot, p), \quad \mathbf{U} \sim U(\bar{S}_p^2), \text{ 且 } R^* \text{ 与 } \mathbf{U} \text{ 相互独立}. \tag{1.6}$$

(1.6) 和 (1.5) 两种表示的关系是

$$R^* = R\|\mathbf{V}\|, \quad \mathbf{U} = \frac{\mathbf{V}}{\|\mathbf{V}\|}. \tag{1.7}$$

若 $\mathfrak{P}$ 是相似的, (1.2) 等价于

$$\mathbf{V} = \frac{\mathbf{Y}}{R}\bigg|_{R=r} \sim \frac{U(A(R))}{R}\bigg|_{R=r} = \frac{U(A(r))}{r} = U(A(1)).$$

于是 $\mathbf{Y}$ 可以表示为

$$\mathbf{Y} = R\mathbf{V}, \quad R \sim g(\cdot), \quad \mathbf{V} \sim U(A(1)).$$

由于 $\mathbf{V} = \dfrac{\mathbf{Y}}{R}$ 的分布是 $U(A(1))$, 与 R 无关, 故 $\mathbf{V}$ 与 R 相互独立.

若将球体换成含原点的任意实心有界集合 $D, L_p(D) > 0, \mathbf{0} \in D \subset S_p^2(r)$,

$$\mathbf{Y} = R\mathbf{V}, \quad R \sim g(\cdot), \quad \mathbf{V} \sim U(D), \text{ 且 } R \text{ 与 } \mathbf{U} \text{ 相互独立}, \tag{1.8}$$

其中 D 是实心的, 即 $\mathbf{x} \in D$ 蕴含 $\{r\mathbf{x}, 0 \leqslant r \leqslant 1\} \subset D$. 于是 p 维随机向量 $\mathbf{Y}$ 的概率密度函数是

$$f_Y(\mathbf{y}) = \frac{1}{L_p(D)} \int_{\frac{\|\mathbf{y}\|}{b\left(\frac{\mathbf{y}}{\|\mathbf{y}\|}\right)}}^{\infty} \frac{g(r)}{r^p} dr,$$

其中

$$b(\mathbf{y}) = \sup\{r : r\mathbf{y} \in D\} < \infty, \quad \forall \|\mathbf{y}\| = 1,$$
$$D = \left\{\mathbf{x} : \|\mathbf{x}\| \leqslant b\left(\frac{\mathbf{x}}{\|\mathbf{x}\|}\right)\right\}.$$

见第四章. 用平行于 $\mathfrak{R}^p$ 的超平面截它的密度函数图像, 截口是与 A 相似的集合. $\mathbf{Y}$ 的分布就是中心相似分布. $\mathbf{Y}$ 也可表示为 (1.6) 的形式

$$\mathbf{Y} = R\mathbf{V} = R\|\mathbf{V}\|\frac{\mathbf{V}}{\|\mathbf{V}\|} = R'\mathbf{U}^*,$$

不过 $\mathbf{U}^*$ 是球面上非均匀随机向量.

1.2 枢轴量、置信分布、随机估计和 VDR 检验

用枢轴量构造检验统计量是经典统计中的典型方法. 用经典统计观点分析枢轴量产生置信分布 (CD), 用 Fisher 思想分析枢轴量产生信仰推断, 进而产生随机估计. VDR 基本理论应用于随机估计产生了参数 VDR 检验和参数 VDR 置信集, 这是通用方法, 可用于多种模型.

此为试读,需要完整PDF请访问: www.ertongbook.com