中国科学: 数学 2020 年 第 50 卷 第 5 期: 571～584
SCIENTIA SINICA Mathematica
综 述

多元统计分析及其应用

献给方开泰教授 80 华诞

李刚[1], 梁家卷[2*], 潘建新[3], 彭小令[4], 田国梁[5]

1. Department of Biostatistics, Jonathan and Karin Fielding of Public Health, University of California at Los Angeles, Los Angeles, CA 90095-1772, USA;
2. College of Business, University of New Haven, West Haven, CT 06516, USA;
3. Department of Mathematics, The University of Manchester, Manchester M13 9PL, UK;
4. 北京师范大学 - 香港浸会大学联合国际学院理工科技学部, 珠海 519087;
5. 南方科技大学理学院统计与数据科学系, 深圳 518055
E-mail: vli@ucla.edu, jliang@newhaven.edu, jianxin.pan@manchester.ac.uk, xlpeng@uic.edu.hk, tiangl@sustech.edu.cn

收稿日期: 2020-03-09; 接受日期: 2020-03-10; 网络出版日期: 2020-05-07; * 通信作者

摘要 自 20 世纪 50 年代以来, 多元统计的理论、方法及其应用受到了越来越广泛的关注. 国内多元统计方向的研究始于 20 世纪 30 年代末至 40 年代初许宝騄在西南联合大学时期. 现代大数据分析的需要使得古典多元统计方法不能完全有效地解决当前的实际问题. 古典多元统计理论从 20 世纪 70 年代以来已经得到了快速发展, 本文旨在对国内学者在推广古典多元统计理论及其应用方面的工作进行概述, 主要包括: 多元统计分析和广义多元统计、一般对称多元分布、增长曲线模型及其他方向. 广义多元统计是正态假设下的传统统计方法论的推广. 其目的是将传统的统计方法论, 如参数估计、假设检验和统计模型等, 推广到更大的多元分布族. 这个分布族称为椭球等高分布族. 一般对称多元分布构成一个更大的多元统计分布族. 这个分布族包含了椭球等高分布族作为其特例. 增长曲线模型包含了一类统计方法, 它允许考虑个体内部及个体之间随着时间变化时的相关关系. 异常观察点及影响观察点的辨别是增长曲线模型研究的一个重要方向.

关键词 copula l_1- 模对称分布 球对称分布 随机表示 椭球等高分布 占有问题 增长曲线模型 左球矩阵分布

MSC (2010) 主题分类 62H10, 62H99

1 引言

多元统计是研究客观事物中多个变量 (或多个因素) 之间相互依赖的规律性的一个统计学分支. 如果研究个体有多个观察数据, 每个观察数据点可看作高维空间中的一个点. 众多的数据点组成高维

英文引用格式: Li G, Liang J J, Pan J X, et al. Multivariate statistics and its applications (in Chinese). Sci Sin Math, 2020, 50: 571–584, doi: 10.1360/SSM-2020-0071

空间中的一个点集. 分析这些点集的统计方法就是多元统计的中心内容. 古典统计学的基本假设就是数据分布的正态性. 虽然早在 19 世纪就出现了处理二维 (二元) 正态分布的一些零星方法, 但系统地处理多元正态总体的统计分析问题主要起始于 20 世纪 30 年代. 英国统计学家 Fisher 通常被公认为数理统计方法论的鼻祖 (参见文献 [1,2]). 中国的数理统计学家许宝騄 (1910–1970) 也是早期多元统计发展的奠基人之一 (参见文献 [3]). 许宝騄培养出了不少在近代中国统计研究与教学方面有着广泛影响力的学者, 他们对中国现代统计事业的推动与发展做出了不可磨灭的贡献. 20 世纪 50 年代后期至 80 年代, 由于此期间概率统计师资匮乏, 教师和科研工作者主要集中在教学方面. 当时的研究课题主要集中在统计在工业和经济等领域的应用, 以及为统计教育提供急需的教科书及相关读物, 例如文献 [4–41].

中国实行改革开放政策至今 40 余年, 中国的统计在教学、科研与应用方面已经有了翻天覆地的变化. 从陈希孺 (1934–2005) 20 世纪 80 年代初期培养出第一批数理统计方向的博士研究生 (白志东、苏淳和赵林城) 开始, 中国的统计科学发展成果已经受到国际瞩目. 中国统计学家们的研究成果辐射到了多个方向. 20 世纪 80 年代至 90 年代当中国统计教学和科研都处于复苏和初期发展阶段时, 已经涌现出一批杰出的统计教育与科研工作者. 例如, 早在 20 世纪 70 年代, 峁诗松就投身数理统计的理论研究和实践探索, 率先将数理统计引入质量管理. 作为我国数理统计专业的开拓者之一, 当年正是在茆诗松等一批数理统计学者的带动下, 我国首次在复旦大学、南开大学和华东师范大学设立了数理统计专业. 他 (或作为第一作者) 撰写的十多本统计学专著 [42–44], 很多成为高校数理统计专业的教材. 20 世纪 80 年代以来, 陈希孺、白志东和赵林城等在参数与非参数统计推断理论、大样本分析和线性模型等领域取得了丰硕成果, 参见代表性专著 [45–51]. 李国英等开创了投影寻踪技术在高维数据分析中的应用并取得了可喜成绩, 参见代表性文献 [52,53].

2 多元统计分析及广义多元统计

本节简要介绍国内统计学者在多元统计的发展及其推广应用方面做出的开创性贡献.

2.1 椭球等高分布理论的发展

随着 20 世纪末统计科学和计算机科学的迅速发展, 正态假设下的古典统计分析已经不能完全满足高维数据分析的需要. 统计学家们早已注意到具有 “宽尾部分布” 数据的现象和实际问题. 正态统计方法对这些数据分析的结果引起了怀疑. 现代计算机技术使得正态假设之外的数据分析成为可能. 推广正态统计方法到更广泛的一类统计分布, 即椭球等高分布族 (elliptically contoured distribution, ECD)[54–56].

随机表示方法在研究 ECD 理论中有着重要作用. 例如, p- 维正态分布 $N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$ 有随机表示

$$\boldsymbol{x}\stackrel{\mathrm{d}}{=}\boldsymbol{\mu}+\boldsymbol{A}\boldsymbol{y}, \tag{2.1}$$

这里 $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}=\boldsymbol{\Sigma}$, $\boldsymbol{y}$ 服从标准正态 $N_p(\boldsymbol{0},\boldsymbol{I})$ ($\boldsymbol{I}$ 表示单位矩阵), $\stackrel{\mathrm{d}}{=}$ 表示方程两边具有相同的分布. (2.1) 称为多元正态向量的随机表示. 可以注意这样一个事实, 对任意的 $p\times p$ 常数正交矩阵 $\boldsymbol{\Gamma}$, $\boldsymbol{\Gamma}\boldsymbol{y}\stackrel{\mathrm{d}}{=}\boldsymbol{y}$ 对服从正态的 $\boldsymbol{y}\sim N_p(\boldsymbol{0},\boldsymbol{I})$ 总是成立的. $\boldsymbol{y}$ 具有分布的旋转不变性或称为球对称性. 通过定义满足下面条件的一族随机向量, 球对称性思想可以推广到一般情形:

$$\boldsymbol{S}_p(\phi)=\{\boldsymbol{x}:\boldsymbol{\Gamma}\boldsymbol{x}\stackrel{\mathrm{d}}{=}\boldsymbol{x}\ \text{对任意}\ p\times p\ \text{常数正交矩阵}\ \boldsymbol{\Gamma}\}, \tag{2.2}$$

这里 $\phi(\cdot)$ 表示 $\boldsymbol{x}$ 的特征函数, $\boldsymbol{S}_p(\phi)$ 称为球对称分布族或简称为球分布族. 很明显, $\boldsymbol{S}_p(\phi)$ 包含了多元标准正态分布 $N_p(\mathbf{0},\boldsymbol{I})$ 及一些常见的多元分布, 例如, 具有零均值和单位离差矩阵 (dispersion matrix) 的多元 t- 分布. 已经证明, $\boldsymbol{x}\sim\boldsymbol{S}_p(\phi)$ 当且仅当

$$\boldsymbol{x}\overset{\mathrm{d}}{=}R\boldsymbol{U}^{(p)}, \tag{2.3}$$

这里随机向量 $\boldsymbol{U}^{(p)}$ 服从 $\mathbb{R}^p$ (p 维实空间) 中单位球面上的均匀分布, 即 $\|\boldsymbol{U}^{(p)}\|=1$ ($\|\cdot\|$ 表示通常的 Euclid 距离); $R>0$ 是一个独立于 $\boldsymbol{U}^{(p)}$ 的随机变量. 方程 (2.3) 称为球分布向量的随机表示. 对任意一个非退化的随机变量, $\boldsymbol{x}\sim\boldsymbol{S}_p(\phi)$ 满足 $P(\boldsymbol{x}=\mathbf{0})=0$, 已经证明

$$\boldsymbol{x}\overset{\mathrm{d}}{=}\|\boldsymbol{x}\|\cdot\frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}, \tag{2.4}$$

这里 $\|\boldsymbol{x}\|$ 和 $\boldsymbol{x}/\|\boldsymbol{x}\|$ 是相互独立的, 并且 $\boldsymbol{x}/\|\boldsymbol{x}\|\overset{\mathrm{d}}{=}\boldsymbol{U}^{(p)}$. (2.1) 是标准正态分布 $N_p(\mathbf{0},\boldsymbol{I})$ 的一个线性变换.

设 $\boldsymbol{x}\sim\boldsymbol{S}_p(\phi)$, 一般来说, 随机向量 $\boldsymbol{x}$ 的联合密度不一定存在. 但是, 如果 $\boldsymbol{x}$ 的密度存在, 那么它一定具有 $g(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x})$ 这样的形式, 这里 $g(\cdot)$ 是某一个非负的一元函数. 在这种情况下, 一定有

$$\int g(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x})d\boldsymbol{x}=\frac{\pi^{p/2}}{\Gamma(p/2)}\int_0^\infty y^{p/2-1}g(y)dy=1.$$

因此, 一个非负函数 $g(\cdot)$ 能够用来定义某个球对称分布的密度函数 $cg(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x})$ 当且仅当

$$\int_0^\infty y^{p/2-1}g(y)dy<\infty.$$

在这种情况下, 我们应当用这样的记号 $\boldsymbol{x}\sim\boldsymbol{S}_p(g)$ 而不用原来的记号 $\boldsymbol{x}\sim\boldsymbol{S}_p(\phi)$, 并且将 $g(\cdot)$ 称为球对称分布的密度生成元 (density generator). 有如下的重要结果:

如果 (2.3) 成立, 那么 $\boldsymbol{x}$ 拥有密度生成元 $g(\cdot)$ 的充分必要条件为 R 具有密度函数 $f(\cdot)$, 且 $g(\cdot)$ 与 $f(\cdot)$ 的关系如下:

$$f(r)=\frac{2\pi^{p/2}}{\Gamma(p/2)}r^{p-1}g(r^2).$$

通过不同的线性变换而得到一般的多元正态分布族. 这种思想用于球分布族 $\boldsymbol{S}_p(\phi)$, 可以给出更大的一个分布族

$$\mathrm{ECD}_p(\boldsymbol{\mu},\boldsymbol{\Sigma},\phi)=\{\boldsymbol{x};\boldsymbol{x}\overset{\mathrm{d}}{=}\boldsymbol{\mu}+\boldsymbol{A}\boldsymbol{y},\boldsymbol{y}\in\boldsymbol{S}_p(\phi),\boldsymbol{\mu}\in\mathbb{R}^p,\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}=\boldsymbol{\Sigma}\}. \tag{2.5}$$

$\mathrm{ECD}_p(\boldsymbol{\mu},\boldsymbol{\Sigma},\phi)$ 称为椭球等高分布族 (ECD), 或简称椭球分布族. (2.1) 限定 $\boldsymbol{y}\sim\boldsymbol{S}_p(\phi)$ 称为 ECD 的随机表示. 可以想象, ECD 将有类似于多元正态 $N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$ 的诸多性质. 例如, 如果 $\boldsymbol{x}\sim\mathrm{ECD}_p(\boldsymbol{\mu},\boldsymbol{\Sigma},\phi)$ 具有密度函数, 则它一定具有如下形式:

$$f(\boldsymbol{x})=c\boldsymbol{\Sigma}^{-\frac{1}{2}}g[(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})], \tag{2.6}$$

这里 $g(\cdot)>0$ 是一个标度函数, $c>0$ 是一个规范常数. 例如, 如果 $\boldsymbol{x}\sim N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$, 那么

$$g(x)=\exp\left(-\frac{x}{2}\right).$$

假设 $\boldsymbol{x} \sim \mathrm{ECD}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$, 将 $\boldsymbol{x}$、$\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$ 剖分如下:

$$\boldsymbol{x} = \begin{pmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix},$$

这里 $\boldsymbol{x}_1$: $m \times 1$, $\boldsymbol{\mu}_1$: $m \times 1$ 且 $\boldsymbol{\Sigma}_{11}$: $m \times m$, 则

$$\boldsymbol{x}_1 \sim \mathrm{ECD}_m(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, \phi), \quad \boldsymbol{x}_2 \sim \mathrm{ECD}_{p-m}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22}, \phi)$$

且相应的条件分布也属于椭球分布族. 这些性质类似于多元正态分布.

用于方程 (2.1)–(2.5) 的随机表示方法在研究 ECD 理论中发挥着重要作用. 方开泰及其合作者们已经发展了 ECD 均值 $\boldsymbol{\mu}$ 和协方差矩阵 $\boldsymbol{\Sigma}$ 的统计推断理论, 详情参见文献 [57–72]. 他们同时发展了一些球分布 (ECD 的子类) 的拟合优度检验方法, 详情参见文献 [73–78]. 一些检验球对称性和椭球对称性的方法收集在文献 [79, 80].

2.2 球矩阵分布理论与应用

中国学者在多元统计分析及广义多元统计方面的贡献, 包括文章、专著和教科书等, 已经被一些国际学者引用于统计方法论及数据分析方面 (参见文献 [81–84]). 这些学者应用文献 [71] 中的球矩阵分布理论发展了一类多元正态及其相关的线性模型的精确参数检验. 这些检验可以有效应用于高维小样本情形, 即使样本量小于维数的情形也一样有效. Läuter 及他的合作者们发展了一类适用于高维小样本情形的多元正态均值比较问题的精确检验. 这些检验把传统的 Hotelling T^2- 检验推广到多重均值比较 (多元方差分析) 和多元回归模型中回归系数的一般线性检验. 这些检验被证明特别适用于高维小样本情形. 即使样本量小于维数的情形仍然有可观的功效 (参见文献 [85]).

一个 $n \times p$ 随机矩阵 $\boldsymbol{X}$ 称为具有左球矩阵分布, 记为

$$\boldsymbol{X} \sim LS_{n\times p}(\phi),$$

如果对任意的 $n \times n$ 常数正交矩阵 $\boldsymbol{\Gamma}$, 有

$$\boldsymbol{\Gamma X} \stackrel{\mathrm{d}}{=} \boldsymbol{X}. \tag{2.7}$$

已经证明, $\boldsymbol{X} \sim LS_{n\times p}(\phi)$ 当且仅当 $\boldsymbol{X}$ 有随机表示

$$\boldsymbol{X} \stackrel{\mathrm{d}}{=} \boldsymbol{UV}, \tag{2.8}$$

这里 $\boldsymbol{U}$ $(n \times p)$ 独立于 $\boldsymbol{V}$ $(p \times p)$, 且 $\boldsymbol{U} \sim \boldsymbol{U}^{(n\times p)}$, 即 Stielfel 流型

$$\boldsymbol{Q}(n,p) = \{\boldsymbol{H}_{n\times p} : \boldsymbol{H}^{\mathrm{T}}\boldsymbol{H} = \boldsymbol{I}_p\} \tag{2.9}$$

上的均匀分布. 如果 $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^{\mathrm{T}}$ ($\boldsymbol{X}$: n 行 p 列) 是一个独立同分布 $N_p(\boldsymbol{0}, \boldsymbol{\Sigma})$ 的随机向量组成的一个观察矩阵, 则可证明

$$\boldsymbol{X} \sim LS_{n\times p}(\phi),$$

且 $\boldsymbol{X}$ 具有随机表示 (2.8). 对任意一个随机矩阵 $\boldsymbol{D}_{p\times q}$ $(q \leqslant p)$, 假定 $\boldsymbol{D}$ 是 $\boldsymbol{X}$ 的一个二次型函数

$$\boldsymbol{D} = f(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}),$$

可以证明 $\boldsymbol{XD} \sim LS_{n\times p}(\phi)$, 所以, $\boldsymbol{XD}$ 也有类似于 (2.8) 的随机表示, 即

$$\boldsymbol{XD} \stackrel{\mathrm{d}}{=} \boldsymbol{UA},$$

且 $\boldsymbol{U} \sim \boldsymbol{U}^{(n\times q)}$ 独立于 $\boldsymbol{A}_{q\times q}$. 作为这个随机表示的结果, 任意一个仿射不变的统计量 $T(\boldsymbol{XD})$, 一定满足

$$T(\boldsymbol{XD}) \stackrel{\mathrm{d}}{=} T(\boldsymbol{U}),$$

即其分布由 $\boldsymbol{U}$ 唯一决定, 而不依赖于如何选择二次函数 $\boldsymbol{D} = f(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})$, 人们可以选择 $q \leqslant p$ 作为降维, 并满足

$$\boldsymbol{U} \sim \boldsymbol{U}^{(n\times q)}.$$

例如, 令

$$q = \min(n, p) - 1,$$

这将使得统计量在高维小样本下和样本量小于维数 ($n \leqslant p$) 的情形仍然适用. 这就是 Läuter 及他的合作者们的精确检验方法的主要思想.

通过应用文献 [71] 中的主要理论及 Läuter 和他的合作者们发展他们的精确检验方法的途径 (参见文献 [71,81,83]), 中国学者发展了一类正态拟合优度的非参数检验. 这些检验特别适用于高维小样本情形, 即使样本量小于维数 ($n \leqslant p$) 仍然有效. 他们的研究成果也包含了带有置信区域的检验正态分布的图方法和一类球性检验 (参见文献 [77,78,86–88]). 由中国学者主导的这类方法后来有了进一步的发展, 参见文献 [89–96].

2.3 广义多元分析的国外平行研究及进一步发展

国内学者在 20 世纪 80 年代和 90 年代初期参与并开创了一些广义多元分析研究方向, 参见文献 [65,69–72,97–104]. 20 世纪 90 年代之后, 国际学者在广义多元分析研究的相似领域也取得了可喜成就. 例如, 广义多元分析基础和统计推断及其应用的平行研究可以参见文献 [105–114].

3 一般对称多元分布及相关分布

在椭球等高分布族之外, 有几类常见的对称多元分布. 一般多元连续分布的系统性概述可追溯到 20 世纪 70 年代初期 (参见文献 [115]). 中国学者在构造新的连续多元对称分布方面的研究始于 20 世纪 80 年代 (参见文献 [98,116]).

3.1 从球分布到 l_1- 模对称分布

一般的连续多元对称分布是通过一个在广义单位球面上的均匀分布的非负随机组合来实现的. 通过改变定义广义单位球的距离度量, 我们可以构造不同的连续多元对称分布族. 通过对一般的连续多元对称分布的随机表示进行线性变换, 我们可以得到更广泛的连续多元对称分布. 例如, ECD 是通过对球对称分布的随机表示进行线性变换而得到的. 文献 [117] 提出了一类多元指数分布. 在此基础上, 文献 [118] 构造了基于指数分布的多元分布族. 我们采用文献 [117] 中的记号, 定义如下分布族:

$$F_n = \{L(\boldsymbol{z}) : \boldsymbol{z} \stackrel{\mathrm{d}}{=} R\boldsymbol{u}, R \geqslant 0 \text{ 独立于 } \boldsymbol{u}\}, \tag{3.1}$$

这里 $\boldsymbol{u}=(U_1,\dots,U_n)^{\mathrm{T}}$ 是服从 l_1- 模单位球面 (第一象限) 上

$$\boldsymbol{S}_+^1=\left\{\boldsymbol{z}=(z_1,\dots,z_n)^{\mathrm{T}}: z_i\geqslant 0\ (i=1,\dots,n), \|\boldsymbol{z}\|_1=\sum_{i=1}^n z_i=1\right\} \tag{3.2}$$

的均匀分布, 其中 $\|\boldsymbol{z}\|_1$ 是 $\boldsymbol{z}$ 的 l_1- 模. 文献 [118] 证明了对任意 $\boldsymbol{z}=(Z_1,\dots,Z_n)^{\mathrm{T}}\in F_n$, 它的生存函数

$$P(Z_1>z_1,\dots,Z_n>z_n) \tag{3.3}$$

只依赖于 l_1- 模 $\|\boldsymbol{z}\|_1$. 由此可以构造一个新的分布族:

$$T_n=\{L(\boldsymbol{z}):\boldsymbol{z}=(Z_1,\dots,Z_n)^{\mathrm{T}}\in\mathbb{R}_+^n,\ \text{满足}\ P(Z_1>z_1,\dots,Z_n>z_n)=h(\|\boldsymbol{z}\|_1)\}, \tag{3.4}$$

这里

$$\mathbb{R}_+^n=\{\boldsymbol{z}=(z_1,\dots,z_n)^{\mathrm{T}}: z_i\geqslant 0\ (i=1,\dots,n)\}.$$

文献 [118] 证明了 T_n 包含了一个对称多元分布的子族:

$$D_{n,\infty}=\{L(\boldsymbol{z}):\boldsymbol{z}\overset{\mathrm{d}}{=}R\boldsymbol{x}, R\geqslant 0\ \text{独立于}\ \boldsymbol{x}=(X_1,\dots,X_n)^{\mathrm{T}},\ X_i\overset{\text{i.i.d.}}{\sim}\mathrm{Exp}(\lambda)\}, \tag{3.5}$$

这里 $\mathrm{Exp}(\lambda)$ 表示参数为 $\lambda>0$ 的指数分布. $D_{n,\infty}$ 实际上是混合指数分布族. 文献 [118] 同时证明了三个分布族有趣的关系:

$$D_{n,\infty}\subset T_n\subset F_n, \tag{3.6}$$

这意味着 F_n 是最大的分布族, 它包含了 T_n 作为其子族, T_n 包含了 $D_{n,\infty}$ 作为其子族; 并得到了 $\boldsymbol{z}=(Z_1,\dots,Z_n)^{\mathrm{T}}\in F_n$ 的生存函数具有如下形式:

$$P(Z_1>a_1,\dots,Z_n>a_n)=\int_{\|\boldsymbol{a}\|_1}^{+\infty}\left(1-\frac{\|\boldsymbol{a}\|_1}{r}\right)^{n-1}dG(r), \tag{3.7}$$

这里 $G(r)$ 代表随机表示 (3.1) 中 R 的分布函数,

$$\boldsymbol{a}=(a_1,\dots,a_n)^{\mathrm{T}}\in\mathbb{R}_+^n.$$

如果 $\boldsymbol{z}=(Z_1,\dots,Z_n)^{\mathrm{T}}\in F_n$ 具有密度函数, 则它一定具有形式 $f(\|\boldsymbol{z}\|_1)$ ($\boldsymbol{z}\in\mathbb{R}_+^n$), 其只依赖于 $\boldsymbol{z}$ 的 l_1- 模. 文献 [119] 取得了多元 l_1- 模对称分布的顺序统计量的联合分布. 文献 [120] 提出了矩阵指数分布. 文献 [121] 研究了多元 l_1- 模对称分布的位置参数与刻度参数的统计推断问题.

文献 [122] 研究了具有旋转对称性的一类分布的大样本性质. 文献 [123] 提出了一类广义对称 Dirichlet 分布. 文献 [124] 得到了一个多元 l_1- 模对称分布的刻画性质. 文献 [125] 构造了一类多元分布, 包含了多元 logistic 分布. 文献 [126] 概述了当时众多已知的对称多元及相关分布. 文献 [126] 提出了一类 v- 球分布族的统计推断和建模应用. 文献 [127] 提出了一类广义 Liouville 分布. 定义方程 (3.1) 中的分布族 F_n 的思想由文献 [128] 推广到 l_p- 模对称分布, 也由文献 [129] 推广到 L_p- 模对称多元分布. 这些文献中, l_p- 模是对非负向量定义的, 而 L_p- 模是对所有实向量定义的.

3.2 其他相关的多元分布

中国学者在对称多元及相关分布的研究持续于 20 世纪 90 年代及之后若干年. 例如, 文献 [130] 引进了 L_1- 模对称分布到 L_1- 模统计推断分析. 这里 L_1- 模区别于前面所用的记号 l_1- 模: l_1- 模是对非负向量定义的, 而 L_1- 模是对所有实向量定义的. 文献 [131] 利用一类对称多元分布的垂直密度表示法提出了生成这些分布的随机数的方法. 文献 [132] 提出了一个推广复数正态分布的新途径. 文献 [133] 提出了一个检验高维对称分布的新方法. 文献 [134] 构造了一类具有非椭球等高线的二维分布族. 文献 [135] 提出了一个使用垂直密度表示法生成多元分布的新途径. 文献 [136] 提出了使用 copula 方法构造已知边际分布的 meta 型椭球分布. 他们的 copula 方法已被众多的国际学者在多个领域中引用. 在众多的构造多元分布的方法中, copula 方法成为构造已知边际分布的多元分布的最常用方法 (参见文献 [137]). 文献 [136] 构造 meta 型椭球分布的思想是建立在 ECD 某些已知性质的基础上的. 如果

$$\boldsymbol{z}=(Z_1,\ldots,Z_n)^{\mathrm{T}}\sim \mathrm{ECD}_n(\mathbf{0},\boldsymbol{R},g)$$

具有方程 (2.6) 给出的概率密度函数和相关系数矩阵 $\boldsymbol{R}$, 那么 $\boldsymbol{z}$ 的每个分量 Z_i $(i=1,\ldots,n)$ 具有边际密度函数

$$q_g(z)=\frac{\pi^{(n-1)/2}}{\Gamma((n-1)/2)}\int_{z^2}^{+\infty}(y-z^2)^{(n-1)/2}g(y)dy \tag{3.8}$$

和分布函数

$$Q_g(z)=\frac{1}{2}+\frac{\pi^{(n-1)/2}}{\Gamma((n-1)/2)}\int_0^z\int_{u^2}^{+\infty}(y-u^2)^{(n-1)/2}g(y)dydu. \tag{3.9}$$

令 $\boldsymbol{x}=(X_1,\ldots,X_n)^{\mathrm{T}}$ 为随机向量, 其每个分量 X_i 具有密度函数 $f_i(x_i)$ 和分布函数 $F_i(x_i)$.

令随机向量

$$\boldsymbol{z}=(Z_1,\ldots,Z_n)^{\mathrm{T}}\sim \mathrm{ECD}_n(\mathbf{0},\boldsymbol{R},g),$$

其中

$$Z_i=Q_g^{-1}(F_i(X_i)),\quad i=1,\ldots,n, \tag{3.10}$$

这里 $Q_g^{-1}(\cdot)$ 是 (3.9) 定义的 $Q_g(\cdot)$ 的逆函数. 文献 [136] 导出了 $\boldsymbol{x}=(X_1,\ldots,X_n)^{\mathrm{T}}$ 的联合密度为

$$h(x_1,\ldots,x_n)=\phi(Q_g^{-1}(F_1(x_1)),\ldots,Q_g^{-1}(F_n(x_n)))\prod_{i=1}^n f_i(x_i), \tag{3.11}$$

这里 ϕ 称为 n- 变量密度权函数,

$$\phi(z_1,\ldots,z_n)=\frac{|\boldsymbol{R}|^{-\frac{1}{2}}g(\boldsymbol{z}^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{z})}{\prod_{i=1}^n q_g(z_i)}. \tag{3.12}$$

如果 $\boldsymbol{x}=(X_1,\ldots,X_n)^{\mathrm{T}}$ 具有 (3.11) 给出的密度函数, 则 $\boldsymbol{X}$ 称为 meta 型椭球分布, 记为

$$\boldsymbol{X}\sim \mathrm{ME}_n(\mathbf{0},\boldsymbol{R},g;\,F_1,\ldots,F_n). \tag{3.13}$$

分布族 (3.13) 包含了众多的多元分布, 例如, $\mathrm{ECD}_n(\mathbf{0},\boldsymbol{R},g)$ 和多种通过选择不同的边际分布 $F_i(x_i)$ 而得到的非对称多元分布. 文献 [136] 取得了一些二维情形有趣的 meta 型 ECD. 一般地, 分布族 (3.13) 是一个极大的分布族. 其任意一个成员的概率密度函数难以获得. 如今 copula 方法已经被广泛应用

于多个领域 (参见文献 [138–140]). 应用文献 [104] 的球分布理论, 文献 [96] 提出了一类检验 L_p- 模对称多元分布的拟合优度的均匀性检验.

文献 [104] 所讨论的对称多元分布的方法给构造新的多元统计分布提供了某种启发性. 例如, 多元偏正态分布 (multivariate skew-normal distribution) 族的构造及其应用可参见文献 [141–153].

4 方向数据、占有问题、增长曲线模型与其他研究方向

改革开放初期, 为了满足中国工业领域对应用统计的需求, 中国学者进行了一系列的应用项目研究, 如聚类分析、占有问题、数理统计与标准化、质量控制和多元观察的图分析等. 这些成果集中于文献 [11, 14–22, 24, 25, 29, 154–157].

4.1 方向数据与占有问题

方向数据产生于多个领域, 如地球科学、天体物理和医学等领域. 文献 [158] 对方向数据给出了一个综合概述. 假设 $\boldsymbol{x} = (x_1, \dots, x_p)^{\mathrm{T}}$ 是球面 $S_p = \{\boldsymbol{x} \in \mathbb{R}^p : \|\boldsymbol{x}\| = 1\}$ 上的一个方向 ($\mathbb{R}^p$ 表示通常的 p- 维 Euclid 空间, $\|\cdot\|$ 表示通常的距离). 方向数据分析的一些主要问题包括球面 S_p 上两个方向 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 上数据的相关性分析, 及 $\boldsymbol{y}$ 对 $\boldsymbol{x}$ 的回归等问题. 方向数据分析在 20 世纪 80 年代后期对中国统计学者是一个全新的领域. 中国学者在这个领域取得了一定的研究成果 (参见文献 [159–166]). 除了集中于统计理论及应用的研究之外, 占有问题是概率论中关于如何合理分配一组球到一组格子里的方法问题. 占有问题源自概率论. 一些关于资源的最优分配的实际问题可以归结为一种占有问题的解. 例如, 酒店空间的占有问题、房间出租、办公室配置, 或将一组人员分配到一个不分割的空间等问题, 都可以被描述为一类占有问题. 有限资源的最优分配问题可以简化成占有问题的求解, 参见文献 [154, 155, 157, 167].

4.2 增长曲线模型与其他研究方向

增长曲线模型是 20 世纪 80 年代中后期中国学者继承发展的一个研究方向. 关于增长曲线模型方法的一般概述可以参见文献 [168]. 异常观察点判别、有影响的观察点识别和协方差结构的研究是增长曲线模型理论的一些重要课题. 增长曲线模型的一般表达模式可以写为 (参见文献 [169])

$$\boldsymbol{Y}_{p\times n} = \boldsymbol{X}_{p\times m}\boldsymbol{B}_{m\times r}\boldsymbol{Z}_{r\times n} + \boldsymbol{E}_{p\times n}, \tag{4.1}$$

这里 $\boldsymbol{X}$ 和 $\boldsymbol{Z}$ 是已知的设计矩阵. 它们的秩分别为 $m < p$ 和 $r < n$, 回归系数矩阵 $\boldsymbol{B}$ 为未知. 进一步假设误差系数矩阵 $\boldsymbol{E}$ 的所有列相互独立且具有零均值与未知协方差矩阵 $\boldsymbol{\Sigma} > 0$ 的 p- 维正态分布. 增长曲线模型 (4.1) 的表达模式也可以写为矩阵正态分布 $\boldsymbol{Y} \sim N_{p\times n}(\boldsymbol{XBZ}, \boldsymbol{\Sigma} \otimes \boldsymbol{I}_n)$ ($\otimes$ 表示 Kronecker 乘积). 由矩阵正态分布的假设可以容易地得到回归系数矩阵 $\boldsymbol{B}$ 和未知协方差矩阵 $\boldsymbol{\Sigma}$ 的极大似然估计. 文献 [169] 利用均值平移回归模型提出了一个多个异常点同时识别的新方法. 文献 [170] 通过利用经验影响函数的方法研究了一个观察子集对增长曲线模型拟合优度的影响. 文献 [171] 采用 Bayes 局部影响方法对 Rao 的简单协方差结构下的增长曲线模型提出了一个模型诊断方法. 文献 [172] 研究了一类在抽象置换框架下没有固定协方差结构的增长曲线模型的局部影响判别. 文献 [173] 讨论了增长曲线模型协方差矩阵的后验分布. 文献 [174] 把文献 [171] 中 Rao 的简单协方差结构情形推广到没有

固定协方差结构情形. 文献 [175] 把投影寻踪技术应用于多元数据分析中的多个异常点判别. 增长曲线模型的近期研究成果可以参见专著 [176].

进入 20 世纪 90 年代以后, 中国统计学家与数学家合作, 将纯粹数学中的数论方法应用于多元统计, 发展了一类建立在数论基础上的多元统计方法, 详见专著 [177] 和综述文献 [178]. 由中国学者开创的数论方法与多元统计相结合途径已经广泛应用于与多元统计相关的数量分析问题, 例如, 文献 [179] 发展了一类多元统计中的全局优化算法, 文献 [180] 概述了伪 Monte Carlo 方法和多元统计在计量经济中的应用. 进入 2000 年之后, 中国的多元统计及其应用在老一辈统计学家已经开创的众多领域继续得到发扬光大. 谨以此文纪念那些为中国统计事业的发展历经辛苦、孜孜不倦和永不停息的科研工作者.

致谢　作者衷心感谢方开泰教授引领我们进入近代多元统计分析领域并悉心指导我们走上了学术之路. 我们从与方教授结下师生之缘起几十年来一直受益良多. 祝愿方教授: 人生八十再展旌旗, 看诺奖得主九十七. 本文作者按照拼音字母序排列.

参考文献

1 Fisher R A, Owen A R G. An appreciation of the life and work of Sir Ronald Aylmer Fisher. Statistician, 1962, 12: 313–319
2 Efron B, Cox D R, Kass R, et al. R. A. Fisher in the 21st century (Invited paper presented at the 1996 R. A. Fisher Lecture). Statist Sci, 1998, 13: 95–122
3 Chen D, Olkin I. Pao-Lu Hsu (Xu, Bao-lu): The grandparent of probability and statistics in China. Statist Sci, 2012, 27: 434–445
4 方开泰, 董泽清, 韩继业. 平稳无后效流的结构. 应用数学和计算数学学报, 1965, 2: 84–90
5 中国科学院数学研究所统计组. 介绍几种简易的数理统计方法. 数学的实践与认识, 1972, 5: 58–66
6 中国科学院数学研究所统计组. 常用数理统计方法. 北京: 科学出版社, 1973
7 中国科学院数学研究所统计组. 正交试验设计. 有色金属, 1974, (8): 39–56
8 方开泰, 刘璋温. 极差在方差分析中的应用. 数学的实践与认识, 1976, (1): 37–51
9 方开泰, 刘璋温. 如何使用概率纸 (I). 数学的实践与认识, 1976, (3): 49–55
10 方开泰, 刘璋温. 如何使用概率纸 (II). 数学的实践与认识, 1976, (4): 55–61
11 轻工业部二局服装标准组, 中国科学院数学研究所统计组. 运用 “条件分布的理论” 制定全国服装统一号型标准. 标准化通讯, 1977, (4): 9–19
12 中国科学院数学研究所统计组. 方差分析. 北京: 科学出版社, 1977
13 孙尚拱, 方开泰. 多元分析的附加信息检验法. 应用数学学报, 1977, 3: 81–91
14 方开泰. 聚类分析 (I). 数学的实践与认识, 1978, (4): 66–80
15 方开泰. 聚类分析 (II). 数学的实践与认识, 1978, (4): 54–62
16 方开泰. 数理统计和标准化 (II). 标准化通讯, 1978, (2): 13–22
17 方开泰. 数理统计和标准化 (IV). 标准化通讯, 1978, (3): 13–20
18 方开泰. 数理统计和标准化 (V). 标准化通讯, 1978, (4): 12–22
19 方开泰. 数理统计和标准化 (VI). 标准化通讯, 1978, (4): 23–30
20 方开泰. 质量控制. 标准化通讯, 1979, (2): 19–30
21 方开泰. 在质量控制中工序能力指数 C_p. 标准化通讯, 1980, (1): 10–17
22 方开泰. 聚类分析及其应用. 见: 数学地质特刊 (I). 北京: 地质出版社, 1980, 12–23
23 刘璋温, 戴树森, 方开泰. 概率纸浅说. 北京: 科学出版社, 1980
24 方开泰. 多变量样本的图分析法 (一). 数学的实践与认识, 1981, (3): 63–71
25 方开泰. 多变量样本的图分析法 (二). 数学的实践与认识, 1981, (4): 42–48
26 方开泰. 方差分析 (第二版). 北京: 科学出版社, 1981
27 马毅林, 方开泰, 吴传义. 数理统计与标准化. 北京: 技术标准出版社, 1981
28 方开泰, 潘恩沛. 聚类分析. 北京: 地质出版社, 1982
29 孙尚拱, 方开泰. 多元分析的附加信息检验法. 应用数学学报, 1977, 3: 81–91
30 张尧庭, 方开泰. 多元统计分析引论. 北京: 科学出版社, 1982
31 方开泰, 许建伦. 统计分布. 北京: 科学出版社, 1987
32 方开泰, 全辉, 陈庆云. 实用回归分析. 北京: 科学出版社, 1988

33 方开泰, 项可风, 刘光仪. 测试方法的精密度. 北京: 中国标准出版社, 1988
34 方开泰. 实用多元统计分析. 上海: 华东师范大学出版社, 1989
35 陈希孺. 数理统计引论. 北京: 科学出版社, 1981
36 倪国熙, 陈希孺. 线性统计与线性代数参考材料. 合肥: 安徽大学出版社, 1981
37 陈希孺, 王松桂. 近代实用回归分析. 南宁: 广西人民出版社, 1984
38 陈希孺, 苏淳. 统计学漫话. 北京: 科学出版社, 1987
39 陈希孺, 王松桂. 近代回归分析原理方法及应用. 合肥: 安徽教育出版社, 1987
40 陈希孺, 倪国熙. 数理统计学教程. 上海: 上海科学技术出版社, 1988
41 陈希孺. 统计学概貌. 北京: 科学技术文献出版社, 1989
42 茆诗松. 贝叶斯统计. 北京: 中国统计出版社, 1999
43 茆诗松. 概率论与数理统计教程. 北京: 高等教育出版社, 2004
44 茆诗松, 周纪芗, 陈颖. 试验设计 (第二版). 中国统计出版社, 2012
45 陈希孺. 线性模型参数的估计理论专著第 14 号. 北京: 科学出版社, 1985
46 陈希孺, 陈桂景, 吴启光, 等. 线性模型参数的估计理论. 北京: 科学出版社, 1985
47 中国科学技术大学概率统计课题组. 非参数统计. 上海: 上海科学技术出版社, 1989
48 陈希孺, 赵林城. 线性模型中的 M 方法. 上海: 上海科学技术出版社, 1996
49 陈希孺. 广义线性模型的拟似然法. 合肥: 中国科学技术大学出版社, 2011
50 陈希孺, 方兆本. 陈希孺文集非参数统计. 合肥: 中国科学技术大学出版社, 2012
51 Bai Z, Silverstei J W. Spectral Analysis of Large Dimensional Random Matrices. New York: Springer, 2006
52 Li G, Chen Z. Projection-pursuit approach to robust dispersion matrices and principal components: Primary theory and Monte Carlo. J Amer Statist Assoc, 1985, 80: 759–766
53 Li G, Cheng P. Some recent developments in projection pursuit in China. Statist Sinica, 1993, 3: 351–51
54 Cambanis S, Huang S, Simons G. On the theory of elliptically contoured distributions. J Multivariate Anal, 1981, 11: 368–385
55 Dawid A P. Spherical matrix distributions and a multivariate model. J R Stat Soc Ser B Stat Methodol, 1977, 39: 254–261
56 Dawid A P. Extendibility of spherical matrix distributions. J Multivariate Anal, 1978, 8: 559–566
57 张尧庭, 方开泰, 陈汉峰. 矩阵椭球等高分布族. 数学物理学报, 1985, 5: 341–353
58 方开泰, 许建伦. 椭球等高分布参数的似然比检验. 经济数学, 1985, 2: 1–9
59 方开泰. 广义多元分析简介—椭球等高分布族理论. 数学进展, 1987, 16: 1–15
60 方开泰, 范剑青, 许建伦. 随机投影阵的二次型分布及其应用. 应用概率统计, 1987, 3: 289–297
61 Fang K T, Xu J L. The Mills' ratio of multivariate normal distributions and spherical distributions. Acta Math Sin Engl Ser, 1987, 30: 211–220
62 全辉, 方开泰. 广义多元分布参数检验的无偏性. 应用数学学报, 1987, 10: 215–234
63 张洪青, 方开泰. 有关左右球对称矩阵分布的几个性质. 应用概率统计, 1987, 3: 97–105
64 方开泰, 许建伦, 滕成业. 一类矩阵球对称分布参数的似然比检验. 东北数学, 1987, 4: 241–252
65 Fan J Q, Fang K T. Inadmissibility of the usual estimator for location parameters of spherically symmetric distributions. Chinese Sci Bull, 1989, 34: 533–537
66 Fang K T, Liang J J. Inequalities for the partial sums of elliptical order statistics related to genetic selection. Canad J Statist, 1989, 17: 439–446
67 Quan H, Fang K T, Teng C Y. The applications of information function for spherical distributions. Northeast Math J, 1989, 5: 27–32
68 Xu J L, Fang K T. Expected values of zonal polynomials of spherical matrix distributions. Acta Math Appl Sin Engl Ser, 1989, 5: 6–14
69 Anderson T W, Fang K T. On the theory of multivariate elliptically contoured distributions and their applications. In: Statistical Inference in Elliptically Contoured and Related Distributions. New York: Allerton Press, 1990, 1–23
70 Anderson T W, Fang K T. Inference in multivariate elliptically contoured distributions based on maximum likelihood. In: Statistical Inference in Elliptically Contoured and Related Distributions. New York: Allerton Press, 1990, 201–216
71 Fang K T, Zhang Y T. Generalized Multivariate Analysis. Beijing-Berlin: Science Press-Springer-Verlag, 1990
72 Anderson T W, Fang K T. Theory and Applications elliptically contoured and related distributions. In: The Development of Statistics: Recent Contributions from China. London: Longman, 1992, 41–62
73 Fang K T, Zhu L X, Bentler P M. A necessary test of goodness of fit for sphericity. J Multivariate Anal, 1993, 45: 34–55
74 Zhu L X, Fang K T, Zhang J T. A projection NT-type test for spherical symmetry of a multivariate distribution. In: Tiit E M, Kollo T, Niemi H, eds. Multivariate Statistics and Matrices in Statistics. New Trends in Probability and

Statistics, vol. 3. London: VSP International Science Publisher, 1995, 109–122
75 Zhu L X, Fang K T, Bhatti M I, et al. Testing sphericity of a high-dimensional distribution based on bootstrap approximation. Pakistan J Statist, 1995, 11: 49–65
76 Li R Z, Fang K T, Zhu L X. Some Q-Q probability plots to test spherical and elliptical symmetry. J Comput Graph Statist, 1997, 6: 435–450
77 Liang J, Fang K T. Some applications of Läuter's technique in tests for spherical symmetry. Biom J, 2000, 42: 923–936
78 Liang J, Fang K T, Hickernell F J. Some necessary uniform tests for spherical symmetry. Ann Inst Statist Math, 2008, 60: 679–696
79 Fang K T, Liang J. Tests of spherical and elliptical symmetry. In: Johnson N L, Kotz S, eds. Encyclopedia of Statistical Sciences, Update, vol. 3. New York: John Wiley & Sons, 1999, 686–691
80 Fang K T, Liang J. Spherical and elliptical symmetry, tests of. In: Johnson N L, Kotz S, eds. Encyclopedia of Statistical Sciences, 2nd ed. New York: John Wiley & Sons, 2006, 7924–7930
81 Läuter J. Exact t and F tests for analyzing studies with multiple endpoints. Biometrics, 1996, 52: 964–970
82 Läuter J, Glimm E, Kropf S. New multivariate tests for data with an inherent structure. Biom J, 1996, 38: 5–23
83 Läuter J, Glimm E, Kropf S. Multivariate tests based on left-spherically distributed linear scores. Ann Statist, 1998, 26: 1972–1988
84 Glimm E, Läuter J. On the admissibility of stable spherical multivariate tests. J Multivariate Anal, 2003, 86: 254–265
85 Kropf S, Läuter J, Kose D, et al. Comparison of exact parametric tests for high-dimensional data. Comput Statist Data Anal, 2009, 53: 776–787
86 Fang K T, Li R Z, Liang J J. A multivariate version of Ghosh's T_3-plot to detect non-multinormality. Comput Statist Data Anal, 1998, 28: 371–386
87 Liang J, Li R, Fang H, et al. Testing multinormality based on low-dimensional projection. J Statist Plann Inference, 2000, 86: 129–141
88 Fang K T, Liang J, Hickernell F J, et al. A stabilized uniform Q-Q plot to detect non-multinormality. In: Hsiung A C, Ying Z, Zhang C H, eds. Random Walk, Sequential Analysis and Related Topics. Singapore: World Scientific, 2007, 254–268
89 Liang J, Ng K W. A multivariate normal plot to detect nonnormality. J Comput Graph Statist, 2009, 18: 52–72
90 Liang J, Tang M L. Generalized F-tests for the multivariate normal mean. Comput Statist Data Anal, 2009, 53: 1177–1190
91 Liang J, Tang M L, Chan P S. A generalized Shapiro-Wilk statistic for testing high-dimensional normality. Comput Statist Data Anal, 2009, 53: 3883–3891
92 Ai M, Liang J, Tang M L. Generalized T_3-plot for testing high-dimensional normality. Front Math China, 2016, 11: 1363–1378
93 Liang J. Exact F-tests for a class of elliptically contoured distributions. J Adv Statist, 2016, 1: 212–217
94 Liang J. A generalized F-test for the mean of a class of elliptically contoured distributions. J Adv Statist, 2017, 2: 10–15
95 Liang J, Tang M L, Zhao X. Testing high-dimensional normality based on classical skewness and Kurtosis with a possible small sample size. Comm Statist Theory Methods, 2019, 48: 5719–5732
96 Liang J, Ng K W, Tian G. A class of uniform tests for goodness-of-fit of the multivariate L_p-norm spherical distributions and the l_p-norm symmetric distributions. Ann Inst Statist Math, 2019, 71: 137–162
97 Anderson T W, Fang K T. Distributions of quadratic forms and Cochran's theorem for elliptically contoured distributions and their applications. Technical Report: 53. ONR Contract N00014-75-C 0442. California: Stanford University, 1982
98 Fang K T, Chen H F. Relationships among classes of spherical matrix distributions. Acta Math Appl Sin, 1984, 1: 138–148
99 范剑青, 方开泰. 椭球等高分布位移参数的 minimax 估计与 Stein 两阶段估计. 应用概率统计, 1985, 1: 103–114
100 范剑青, 方开泰. 椭球等高分布样品均值和样品回归参数的不允许性. 东北数学, 1985, 1: 68–81
101 范剑青, 方开泰. 球对称分布位移参数的常用估计量的不容许性. 科学通报, 1987, 27: 1361–1364
102 Anderson T W, Hsu H, Fang K T. Maximum-likelihood estimates and likelihood-ratio criteria for multivariate elliptically contoured distributions. Canad J Statist, 1986, 14: 55–59
103 Anderson T W, Fang K T. On the theory of multivariate elliptically contoured distributions. Sankhya, 1987, 49: 305–315
104 Fang K T, Kotz S, Ng K W. Symmetric Multivariate and Related Distributions. London-New York: Chapman & Hall, 1990

105 Gupta A K, Varga T. Elliptically Contoured Models in Statistics. Dordrecht: Kluwer Academic Publishers, 1993
106 Gupta A K, Varga T. A new class of matrix variate elliptically contoured distributions. J Statist Soc, 1994, 3: 255–270
107 Gupta A K, Varga T. Moments and other expected values for matrix variate elliptically contoured. Distrib Statist, 1994, 54: 361–373
108 Gupta A K, Varga T. Some inference problems for matrix variate elliptically contoured distributions. Statistics, 1995, 26: 219–229
109 Gupta A K, Varga T. Characterization of matrix variate elliptically contoured distributions. In: Johnson N L, Balakrishnan N, eds. Advances in the Theory and Practice of Statistics: A Volume in Honor of Samuel Kotz. New York: John Wiley & Sons, 1997, 455–467
110 Gupta A K. Multivariate elliptically contoured and generalized normal models. Random Oper Stoch Equ, 1998, 6: 281–290
111 Gupta A K, Nagar D K. Matrix Variate Distributions. Boca Raton: Chapman & Hall/CRC, 1999
112 Arashi M, Tabatabaey S M M. A note on classical Stein-type estimators in elliptically contoured models. J Statist Plann Inference, 2010, 140: 1206–1213
113 Coelho C A, Marques F J. Near-exact distributions for the likelihood ratio test statistic to test equality of several variance-covariance matrices in elliptically contoured distributions. Comput Statist, 2012, 27: 627–659
114 Gupta, A K, Varga T, Bodnar T. Elliptically Contoured Models in Statistics and Portfolio Theory. New York: Springer, 2013
115 Johnson N L, Kotz S. Distributions in Statistics: Continuous Multivariate Distributions. New York: John Wiley & Sons, 1972
116 方开泰, 吴月华. 二次型分布与 COCHRAN 定理. 经济数学, 1984, 1: 29–48
117 Fang K T, Fang B Q. A new family of multivariate exponential distributions. Kexue Tongbao, 1986, 31: 1510–1511
118 Fang K T, Fang B Q. Some families of mutivariate symmetric distributions related to exponential distribution. J Multivariate Anal, 1988, 24: 109–122
119 方碧琪, 方开泰. 一个和指数分布有关的多元对称分布族的次序统计量及其应用. 应用概率统计, 1988, 4: 44–52
120 方开泰, 方碧琪. 矩阵负指数分布族. 东北数学, 1988, 4: 16–28
121 Fang B, Fang K T. Maximum likelihood estimates and likelihood ratio criteria for location and scale parameters of the multivariatel l_1-norm symmetric distributions. Acta Math Appl Sin, 1988, 4: 13–22
122 Fang K T, Fan J Q. Large sample properties for distributions with rotational symmetries. Northeast Math J, 1988, 4: 379–388
123 Fang K T, Fang B Q. A class of generalized symmetric Dirichlet distributions. Acta Math Appl Sin Engl Ser, 1988, 4: 316–322
124 Fang B Q, Fang K T. A characterization of multivariate l_1-norm symmetric distributions. Statist Probab Lett, 1989, 7: 297–299
125 Fang K T, Xu J L. A class of multivariate distributions including the multivariate logistic. J Math Research Exposition, 1989, 9: 91–100
126 Fernández C, Osiewalski J, Steel M F J. Modeling and inference with v-spherical distributions. J Amer Statist Assoc, 1995, 90: 1331–1340
127 Gupta A K, Song D. Generalized liouville distribution. Comput Math Appl, 1996, 32: 103–109
128 Yue X, Ma C. Multivariate l_p-norm symmetric distributions. Statist Probab Lett, 1995, 24: 281–288
129 Song D, Gupta A K. L_p-norm uniform distribution. Proc Amer Math Soc, 1997, 125: 595–602
130 Fang K T, Kotz S, Ng K W. On the L_1-norm distributions. In: L_1-Statistical Analysis and Related Methods. Amsterdam: Elsevier Science Publishers-North Holland, 1992, 401–413
131 Kotz S, Fang K T, Liang J J. On multivariate vertical density representation and its application to random number generation. Statistics, 1997, 30: 163–180
132 Rosen V D, Fang H B, Fang K T. An extension of the complex normal distribution. In: Johnson N L, Balakrishnan N, eds. Advances in the Theory and Practice of Statistics: A Volume in Honor of Samuel Kotz. New York: John Wiley & Sons, 1997, 415–427
133 Zhu L X, Fang K T, Li R Z. A new approach for testing symmetry of a high-dimensional distribution. Bull Hong Kong Math Soc, 1997, 1: 35–46
134 Fang K T, Fang H B, Rosen D. A family of bivariate distributions with non-elliptical contours. Comm Statist Theory Methods, 2000, 29: 1885–1898
135 Fang K T, Yang Z, Kotz S. Generation of multivariate distributions by vertical density representation. Statistics, 2001, 35: 281–293

136 Fang H B, Fang K T, Kotz S. The meta-elliptical distributions with given marginals. J Multivariate Anal, 2002, 82: 1–16
137 Kotz S, Seeger J P. A new approach to dependence in multivariate distributions. In: Dall'Aglio G, Kotz S, Salinetti G, eds. Advance in Probability Distributions with Given Marginals. Dordrecht: Kluwer Academic, 1991, 13–50
138 Cherubini U, Luciano E, Vecchiato W. Copulas Methods in Finance. New York: John Wiley & Sons, 2004
139 Nelsen R B. An Introduction to Copulas. New York: Springer, 2006
140 Fan Y, Patton A J. Copulas in econometrics. Annu Rev Econ, 2014, 6: 179–200
141 Azzalini A. The multivariate skew-normal distribution. Biometrika, 1996, 83: 715–726
142 Azzalini A, Capitanio A. Statistical applications of the multivariate skew normal distribution. J R Stat Soc Ser B Stat Methodol, 1999, 61: 579–602
143 Azzalini A. The skew-normal distribution and related multivariate families. Scand J Statist, 2005, 32: 159–188
144 Vernic R. Multivariate skew-normal distributions with applications in insurance. Insurance Math Econom, 2006, 38: 413–426
145 Counsell N, Cortina-Borja M, Lehtonen A, et al. Modelling psychiatric measures using skew-normal distributions. Eur Psychiatry, 2011, 26: 112–114
146 Branco M D, Dey D K. A general class of multivariate skew-elliptical distributions. J Multivariate Anal, 2001, 79: 99–113
147 Azzalini A, Capitanio A. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. J R Stat Soc Ser B Stat Methodol, 2003, 65: 367–389
148 Genton M G. Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality. Boca Raton: Chapman & Hall/CRC, 2004
149 Ma Y, Genton M G. Flexible class of skew-symmetric distributions. Scand J Statist, 2004, 31: 459–468
150 Arellano-Valle R B, Genton M G. On fundamental skew distributions. J Multivariate Anal, 2005, 96: 93–116
151 Genton M G, Loperfido N M R. Generalized skew-elliptical distributions and their quadratic forms. Ann Inst Statist Math, 2005, 57: 389–401
152 Nadarajah S, Kotz S. Skewed distributions generated by the Cauchy kernel. Braz J Probab Stat, 2005, 19: 39–51
153 Pourmousa R, Mashinchi M. An interpretation of skew-elliptical distributions in terms of fuzzy events. Sci Iran, 2012, 19: 1870–1875
154 Fang K T. A restricted occupancy problem. J Appl Probab, 1982, 19: 707–711
155 Fang K T. A restricted occupancy problem and its central limit theorem. Kexue Tongbao, 1982, 27: 572–573
156 方开泰, 马逢时. 聚类分析中的分解法及其应用. 应用数学学报, 1982, 5: 339–534
157 Fang K T, Niedzwiecki D. A unified approach to distributions in restricted occupancy problem. In: Contributions to Statistics, Essays in Honour of Professor Norman Lloyd Johnson. Amsterdam: North-Holland, 1983, 147–158
158 Mardia K V. Directional data analysis: An overview. J Appl Stat, 1988, 15: 115–122
159 方开泰, 范剑青, 全辉, 等. 方向数据的统计分析 (I). 数理统计与管理, 1989, 8(1): 59–61
160 方开泰, 范剑青, 全辉, 等. 方向数据的统计分析 (II). 数理统计与管理, 1989, 8(2): 58–65
161 方开泰, 范剑青, 全辉, 等. 方向数据的统计分析 (III). 数理统计与管理, 1989, 8(3): 57–64
162 方开泰, 范剑青, 全辉, 等. 方向数据的统计分析 (IV). 数理统计与管理, 1989, 8(4): 56–65
163 方开泰, 范剑青, 全辉, 等. 方向数据的统计分析 (V). 数理统计与管理, 1989, 8(5): 56–64
164 方开泰, 范剑青, 全辉, 等. 方向数据的统计分析 (VI). 数理统计与管理, 1989, 8(6): 57–64
165 方开泰, 范剑青, 全辉, 等. 方向数据的统计分析 (VII). 数理统计与管理, 1990, 9(1): 56–64
166 方开泰, 范剑青, 全辉, 等. 方向数据的统计分析 (VIII). 数理统计与管理, 1990, 9(2): 59–65
167 Fang K T. Occupancy problems. In: Kotz S, Johnson N L, eds. Encyclopedia of Statistical Sciences 6. New York: John Wiley & Sons, 1985, 402–406
168 Von Rosen D. The growth curve model: A review. Comm Statist Theory Methods, 1991, 20: 2791–2822
169 Pan J X, Fang K T. Multiple outlier detection in growth curve model with unstructured covariance matrix. Ann Inst Statist Math, 1995, 47: 137–153
170 Pan J X, Fang K T. Influential observation in the growth curve model with unstructured covariance matrix. Comput Statist Data Anal, 1996, 22: 71–87
171 Pan J X, Fang K T, Liski E P. Bayesian local influence for the growth curve model with Rao's simple covariance structure. J Multivariate Anal, 1996, 58: 55–81
172 Pan J X, Fang K T, von Rosen D. Local influence assessment in the growth curve model with unstructured covariance. J Statist Plann Inference, 1997, 62: 263–278
173 Pan J X, Fang K T, von Rosen D. On the posterior distribution of the covariance matrix of the growth curve model. Statist Probab Lett, 1998, 38: 33–39

174 Pan J X, Fang K T, von Rosen D. Bayesian local influence in growth curve model with unstructured covariance. Biom J, 1999, 41: 641–658
175 Pan J X, Fung W K, Fang K T. Multiple outlier detection in multivariate data using projection pursuit techniques. J Statist Plann Inference, 2000, 83: 153–167
176 Pan J X, Fang K T. Growth Curve Models and Statistical Diagnostics. New York: Springer, 2002
177 Fang K T, Wang Y. Number-Theoretic Methods in Statistics. London: Chapman & Hall, 1994
178 Fang K T, Wang Y, Bentler P M. Some applications of number-theoretic methods in statistics. Statist Sci, 1994, 9: 416–428
179 Fang K T, Hickernell F J, Winker P. Some global optimization algorithms in statistics. In: Lecture Notes in Operations Research. Singapore: World Scientific, 1996, 14–24
180 Winker P, Fang K T. Randomness and quasi-Monte Carlo approaches: Some remarks on fundamentals and applications in statistics and econometrics. Jahrb National Statist, 1999, 218: 215–228

Multivariate statistics and its applications

Gang Li, Jiajuan Liang, Jianxin Pan, Xiaoling Peng & Guoliang Tian

Abstract Multivariate statistics and its applications have received more and more attention since 1950s. Multivariate statistical research in China was initiated by Pao-Lu Hsu during the end of 1930s and the beginning of 1940s in the so-called "Southwest United University". Modern big data analysis makes classical multivariate statistical theory unable to accurately and effectively solve practical problems. The theory of generalized multivariate statistics has been developing very fast since 1970s. This paper aims to introduce the summary contributions that Chinese scholars have made in the development of generalized multivariate statistics and its applications in several aspects: (1) multivariate statistics and generalized multivariate statistics; (2) general symmetric multivariate distributions; and (3) growth curve modeling and miscellaneous directions. Generalized multivariate statistics is an extension to the traditional statistics under the normal assumption. It aims to generalize the traditional statistical methodologies like parametric estimation, hypothesis testing, and modeling to a much wider family of multivariate distributions that are called elliptically contoured distributions (ECDs). General symmetric multivariate distributions form an even wider class of multivariate probability distributions that includes the ECDs as its special case. Growth curve modeling (GCM) includes statistical methods that allow for consideration of inter-individual variability in intra-individual patterns of chance over time. Outlier detection and identification of influential observations are important topics in the area of GCM.

Keywords copula, l_1-norm symmetric distributions, spherically symmetric distribution, stochastic representation, elliptically contoured distribution, occupancy problem, growth curve model, left-spherical matrix distribution

MSC(2010) 62H10, 62H99

doi: 10.1360/SSM-2020-0071