中国科学: 数学 2020 年 第 50 卷 第 5 期: 711～728
SCIENTIA SINICA Mathematica
论 文

基于混合正态的新型多元 Laplace 分布

献给方开泰教授 80 华诞

张弛[1], 邓文礼[2], 李涛[3], 孙源[4], 田国梁[4*]

1. 深圳大学经济学院, 深圳 518060;
2. 香港恒生大学决策科学学院, 香港;
3. 南方科技大学数学系, 深圳 518055;
4. 南方科技大学统计与数据科学系, 深圳 518055
E-mail: chizhang@szu.edu.cn, mltang@hsu.edu.hk, lit6@sustech.edu.cn, 11849457@mail.sustech.edu.cn, tiangl@sustech.edu.cn

收稿日期: 2019-05-21; 接受日期: 2019-09-03; 网络出版日期: 2020-05-08; * 通信作者
国家自然科学基金 (批准号: 11801380 和 11771199) 和深圳大学新引进教师科研启动项目 (批准号: 2019069) 资助项目

摘要 本文从正态方差混合模型出发, 提出一种新的多元 Laplace 分布—II 型多元 Laplace 分布. 区别于经典的多元 Laplace 分布, 新型分布中随机向量的各分量相对应的混合变量可以具有不同的值, 并且各分量之间的关系仅与正态随机向量的相关结构有关. 当正态分布的协方差矩阵是对角矩阵时, 新型分布包含了多个独立的一元 Laplace 分布的乘积. 本文利用容易处理的随机表示得到了新型分布的概率密度函数和其他统计性质, 并通过条件期望最大化 (expectation/conditional maximization, ECM) 算法得到参数的极大似然估计, 此外, 也考虑了 Bayes 方法. 本文通过仿真模拟实验, 对估计方法的表现进行了评价. 实例分析结果表明本文提出的新型多元 Laplace 分布相较于经典的多元 Laplace 分布更加灵活.

关键词 ECM 算法 多元 Laplace 分布 随机表示 一元 Laplace 分布

MSC (2010) 主题分类 62F10, 62F15, 62H12

1 引言

在过去几十年中, 构造比 Gauss 分布更灵活且重尾的分布族成为了很多学者的研究内容, 主要是因为这样的分布族可以解决带有异常值的连续数据的分析问题. 文献 [1] 提出了一元对称的 Gauss 尺度混合模型 (或正态方差混合模型), 这一模型包括了学生 t 分布、Laplace 分布 (或双指数分布)、逻辑分布 [2]、稳定分布族和指数幂分布族 [3,4]. 这些分布已经被广泛地应用于稳健统计模型中 [5], 旨在提供不受异常值或其他偏离正态假设影响的统计方法.

英文引用格式: Zhang C, Tang M L, Li T, et al. A new multivariate Laplace distribution based on the mixture of normal distributions (in Chinese). Sci Sin Math, 2020, 50: 711–728, doi: 10.1360/N012019-00141

© 2020《中国科学》杂志社

在现实生活中, 语音和图像数据通常表现出数据重尾的特性, 在零值或均值处达到峰值 [6]. 这种观察值可以通过稀疏分布来建模, 例如, 人们发现图像经小波变换处理后的系数具有稀疏分布, 并且这一模型在编码和去噪中已经有了广泛的研究 [7,8]. 在机器学习领域中, 也经常应用稀疏分布, 如独立分量的分析问题 [9]. Laplace 分布就是这样一种稀疏分布. 经典 Laplace 分布是一种对称的概率分布, 设连续型随机变量 X 服从一元 Laplace 分布, 则其概率密度函数是 $f_X(x) = (2\sigma)^{-1}\exp(-|x-\mu|/\sigma)$, $x \in \mathbb{R} \hat{=} (-\infty, \infty)$, 记为 $X \sim \text{Laplace}(\mu, \sigma^2)$, 其中 $\mu \in \mathbb{R}$ 是位置/均值参数, $\sigma\ (>0)$ 是尺度参数. 文献 [1] 指出一元 Laplace 随机变量有如下随机表示:

$$X = \mu + \sqrt{2U}Z, \tag{1.1}$$

其中 $U \sim \text{Exp}(1)$, $Z \sim N(0, \sigma^2)$, 且 U 和 Z 相互独立. 由 (1.1) 易得 $E(X) = \mu$, $\text{Var}(X) = 2\sigma^2$. 本文定义服从 Gamma 分布 $\text{Gamma}(\alpha, \beta)$ 的随机变量 Y 的密度函数为 $[\beta^\alpha/\Gamma(\alpha)]y^{\alpha-1}\exp(-\beta y)$, 其中 $y > 0$. 当 $\alpha = 1$ 时, $\text{Gamma}(1, \beta)$ 为指数分布, 记为 $\text{Exp}(\beta)$.

文献 [10–12] 已经从很多不同的扩展角度研究了多元 Laplace 分布, 并由文献 [13] 进行了综述. 文献 [14,15] 考虑了服从一元 Laplace 分布的单变量构成的向量. 此外, 文献 [16,17] 分别研究了有关多元 Laplace 分布和非对称 Laplace 分布的模型推广. 基于 (1.1), 文献 [18] 通过以下随机表示将一元 Laplace 分布推广至多元 Laplace 分布:

$$\mathbf{x} = \begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_d \end{pmatrix} + \sqrt{V}\begin{pmatrix} Y_1 \\ \vdots \\ Y_d \end{pmatrix} = \boldsymbol{\mu} + \sqrt{V}\mathbf{y},$$

其中 $V \sim \text{Exp}(1/\lambda)$, $\mathbf{y} \sim N_d(\mathbf{0}, \boldsymbol{\Omega})$, $\det(\boldsymbol{\Omega}) = 1$ 且 V 和 $\mathbf{y}$ 相互独立. 根据指数分布的放缩性质, 向量 $\mathbf{x}$ 可被等价地表示为

$$\mathbf{x} = \boldsymbol{\mu} + \sqrt{2U}\left(\sqrt{\frac{\lambda}{2}}\,\mathbf{y}\right) = \boldsymbol{\mu} + \sqrt{2U}\mathbf{z}, \tag{1.2}$$

记为 $\mathbf{x} \sim \text{Laplace}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, 其中 $U \sim \text{Exp}(1)$, $\mathbf{z} = (Z_1, \ldots, Z_d)^\top \sim N_d(\mathbf{0}, \boldsymbol{\Sigma})$, $\boldsymbol{\Sigma} = (\sigma_{ij})$, 且 U 和 $\mathbf{z}$ 相互独立. 由 (1.2) 易得

$$E(\mathbf{x}) = \boldsymbol{\mu} \quad \text{和} \quad \text{Var}(\mathbf{x}) = 2\boldsymbol{\Sigma}. \tag{1.3}$$

随机向量 $\mathbf{x}$ 的密度函数详见附录 A.1, 参数极大似然估计的计算过程详见附录 A.2. 称 (1.2) 中定义的随机向量 $\mathbf{x}$ 服从于经典的 I 型多元 Laplace 分布. 由 (1.2) 中定义的随机表示, 可以得到 $X_i = \mu_i + \sqrt{2U}Z_i$, 或

$$X_i \sim \text{Laplace}(\mu_i, \sigma_{ii}). \tag{1.4}$$

(1.2) 中所有的分量 $\{X_i\}_{i=1}^d$ 对于混合变量具有相同的值, 即 $U \sim \text{Exp}(1)$. 换而言之, 每个边际分布只能有相同的重尾. 此外, $\mathbf{x}$ 中任意一对分量 X_i 和 X_j $(i \neq j)$ 之间的相关性取决于公共因子 U 和 $\mathbf{z}$ 中的相应分量. 也就是即使 Z_i 和 Z_j 相互独立, X_i 和 X_j 也总是相关的. 因此, 经典 I 型多元 Laplace 分布不包含 d 个独立的一元 Laplace 分布的乘积这一特殊情形, 使得该分布族的模型灵活性较差.

本文希望通过用独立同分布的标准指数型随机变量替换公共因子 U, 提出一个新的多元 Laplace 分布. 因此, $\mathbf{x}$ 中的分量仍然是相关的, 且 $\mathbf{x}$ 中的相关性仅依赖于 $\mathbf{z}$ 的相关结构, 这样, 当 $\mathbf{z}$ 的协方差矩阵为对角矩阵时, 新的多元 Laplace 分布将退化为多个独立的一元 Laplace 分布的乘积.

本文的其余部分安排如下. 第 2 节通过随机表示给出 II 型多元 Laplace 分布, 并研究其分布特性. 第 3 节将利用 ECM 算法求解参数的极大似然估计, 同时给出自助抽样法的置信区间及分量间独立性的假设检验. 第 4 节考虑 Bayes 方法. 第 5 节通过模拟实验来评价所提出新分布和估计方法的表现. 第 6 节给出两个数据集的实例分析. 第 7 节对本文的内容和方法进行总结, 并对未来可能的研究问题进行讨论与展望.

2 II 型多元 Laplace 分布

令 $\{U_i\}_{i=1}^d \overset{\text{i.i.d.}}{\sim} \mathrm{Exp}(1)$, $\mathbf{z}=(Z_1,\ldots,Z_d)^\top \sim N_d(\mathbf{0},\boldsymbol{\Sigma})$, 且 $\mathbf{u}=(U_1,\ldots,U_d)^\top$ 与 $\mathbf{z}$ 相互独立. 定义

$$\mathbf{x}=\begin{pmatrix}X_1\\ \vdots\\ X_d\end{pmatrix}=\begin{pmatrix}\mu_1\\ \vdots\\ \mu_d\end{pmatrix}+\sqrt{2}\begin{pmatrix}\sqrt{U_1} & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \sqrt{U_d}\end{pmatrix}\begin{pmatrix}Z_1\\ \vdots\\ Z_d\end{pmatrix}=\boldsymbol{\mu}+\sqrt{2}\mathbf{U}^{1/2}\mathbf{z}, \tag{2.1}$$

其中 $\mathbf{U}^{1/2}=\mathrm{diag}(\sqrt{U_1},\ldots,\sqrt{U_d})$, 则称 $\mathbf{x}$ 服从 II 型多元 Laplace 分布, 记为 $\mathbf{x}\sim \mathrm{Laplace}_d^{(\mathrm{II})}(\boldsymbol{\mu},\boldsymbol{\Sigma})$, $\boldsymbol{\mu}$ 是位置/均值向量参数, $\boldsymbol{\Sigma}=(\sigma_{ij})$ 是尺度参数矩阵.

2.1 边际分布、矩和联合密度函数

由 (2.1) 中的随机表示, 有 $X_i=\mu_i+\sqrt{2U_i}Z_i$, 表明

$$X_i\sim \mathrm{Laplace}(\mu_i,\sigma_{ii}), \tag{2.2}$$

即边际分布仍属于一元 Laplace 分布族. 对任意满足 $1\leqslant k_1<\cdots<k_r\leqslant d$ 的正整数 $k_1,\ldots,k_r$, 有 $(X_{k_1},\ldots,X_{k_r})^\top\sim \mathrm{Laplace}_r^{(\mathrm{II})}(\boldsymbol{\mu}^*,\boldsymbol{\Sigma}^*)$, 其中 $\boldsymbol{\mu}^*=(\mu_{k_1},\ldots,\mu_{k_r})^\top$, $\boldsymbol{\Sigma}^*=(\sigma_{ij}^*)$ 是 $\boldsymbol{\Sigma}$ 的 $r\times r$ 阶子矩阵, 其元素 $\sigma_{ij}^*=\sigma_{ij}$, $i,j=k_1,\ldots,k_r$. 可计算出

$$E(\mathbf{x})=\boldsymbol{\mu}\quad 和\quad \mathrm{Var}(\mathbf{x})=\frac{\pi}{2}\boldsymbol{\Sigma}+\left(2-\frac{\pi}{2}\right)\mathrm{diag}(\boldsymbol{\Sigma}). \tag{2.3}$$

因此, $\mathbf{x}$ 的方差 - 协方差矩阵与 (1.3) 给出的 I 型多元 Laplace 分布的方差 - 协方差矩阵不同. 任意两分量 X_i 和 X_j $(i\neq j)$ 之间的相关系数可表示为 $\mathrm{Corr}(X_i,X_j)=\frac{\pi\sigma_{ij}}{4\sqrt{\sigma_{ii}\sigma_{jj}}}$, 其取值的正负性取决于参数 σ_{ij} 的符号. 若以矩阵形式表达, 即 $\mathrm{Corr}(\mathbf{x})=(\pi/4)\boldsymbol{R}$, $\boldsymbol{R}$ 是与方差 - 协方差矩阵 $\boldsymbol{\Sigma}$ 对应的相关系数矩阵, 即 II 型多元 Laplace 分布的分量之间的相关性是 I 型的 $\pi/4$ 倍. 此外, 由 (2.1) 中的随机表示可知, 由于 U_i 和 U_j 是独立的, 因此, X_i 和 X_j 之间的相关性仅来自于 Z_i 和 Z_j 之间的相关性. 当 $\boldsymbol{\Sigma}$ 是对角矩阵时, II 型多元 Laplace 分布就变为 d 个独立的一元 Laplace 分布的乘积.

对于 (2.1) 中的随机表示, 有如下等价的混合表示:

$$U_1,\ldots,U_d\overset{\text{i.i.d.}}{\sim}\mathrm{Exp}(1)\quad 和\quad \mathbf{x}\mid(\mathbf{u}=\boldsymbol{u})\sim N_d(\boldsymbol{\mu},2\boldsymbol{U}^{1/2}\boldsymbol{\Sigma}\boldsymbol{U}^{1/2}), \tag{2.4}$$

其中 $\boldsymbol{u}=(u_1,\ldots,u_d)^\top$ 是 $\mathbf{u}=(U_1,\ldots,U_d)^\top$ 的实现值, $\boldsymbol{U}^{1/2}=\mathrm{diag}(\sqrt{u_1},\ldots,\sqrt{u_d})=\mathrm{diag}(\boldsymbol{u}^{1/2})$ 是 $\mathbf{U}^{1/2}=\mathrm{diag}(\mathbf{u}^{1/2})$ 的实现值. 则可计算得到 $\mathbf{x}$ 的联合密度函数

$$
\begin{aligned}
f_{\mathbf{x}}(\boldsymbol{x}) &= \int_{\mathbb{R}_+^d} f_{\mathbf{x}\mid\mathbf{u}}(\boldsymbol{x}\mid\boldsymbol{u}) f_{\mathbf{u}}(\boldsymbol{u})d\boldsymbol{u} \\
&= (4\pi)^{-\frac{d}{2}}|\boldsymbol{\Sigma}|^{-\frac{1}{2}}\int_{\mathbb{R}_+^d} u_\bullet^{-\frac{1}{2}}\exp\left[-\frac{(\boldsymbol{u}^{-\frac{1}{2}})^\top(\boldsymbol{X}^*\boldsymbol{\Sigma}^{-1}\boldsymbol{X}^*)\boldsymbol{u}^{-\frac{1}{2}}}{4}-u_+\right]d\boldsymbol{u},
\end{aligned}
\tag{2.5}
$$

其中 $u_\bullet=\prod_{i=1}^d u_i$, $\boldsymbol{X}^*=\mathrm{diag}(\boldsymbol{x}-\boldsymbol{\mu})$, $u_+=\sum_{i=1}^d u_i$.

2.2 I 型和 II 型多元 Laplace 密度函数比较

对于相同的参数组 $(\boldsymbol{\mu},\boldsymbol{\Sigma})$, 从 (1.4) 和 (2.2) 可以看出, I 型和 II 型多元 Laplace 分布的每个分量的边际分布是相同的. 但是, 从 (A.1) 和 (2.5) 看出它们联合密度函数的表达式显然是不同的. 图 1 中

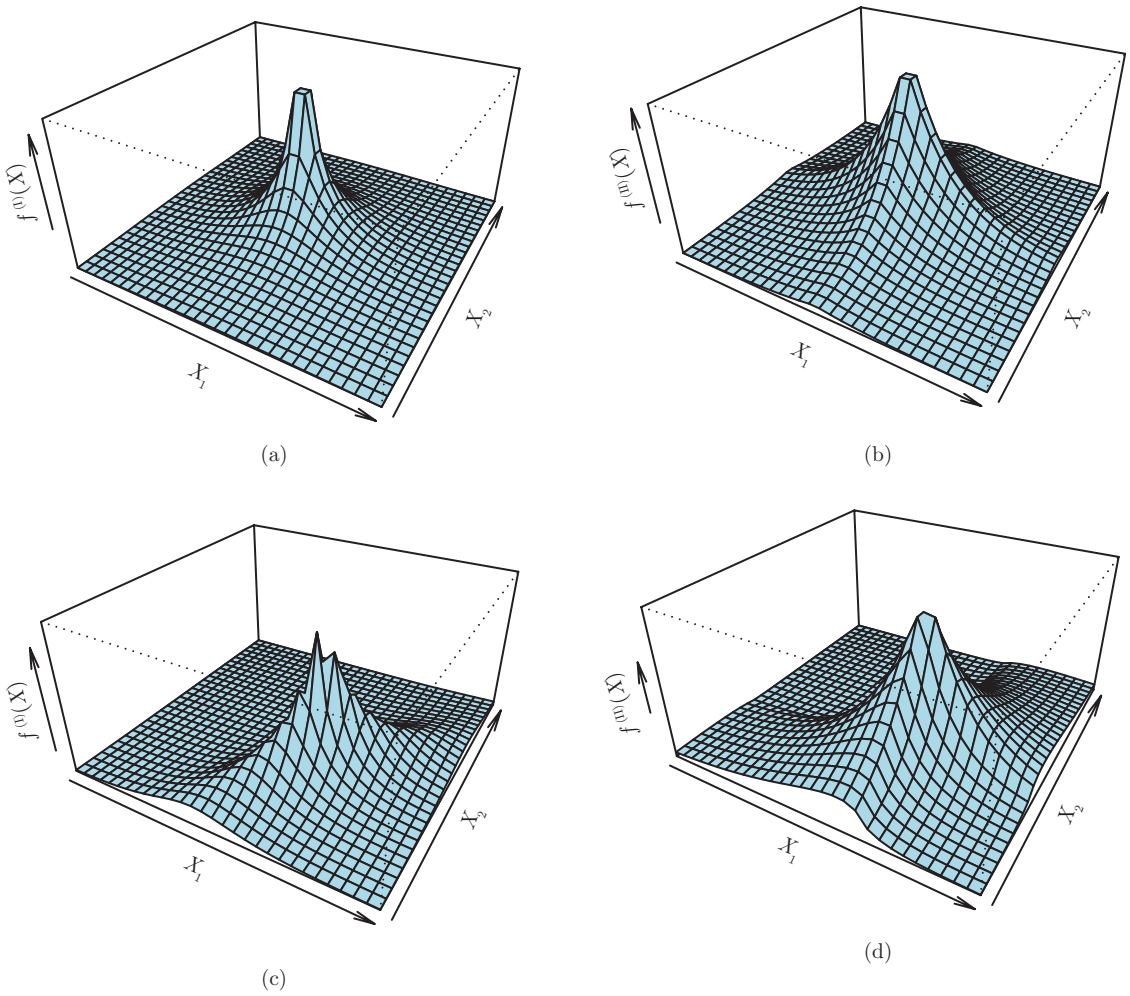

图 1 I 型多元 Laplace 分布和 II 型多元 Laplace 分布. (a) I 型多元 Laplace 分布 : $\boldsymbol{\mu}=(\mathbf{0},\mathbf{0})^\top$, $\boldsymbol{\Sigma}=\mathbf{diag}(\mathbf{1}_2)$; (b) II 型多元 Laplace 分布 : $\boldsymbol{\mu}=(\mathbf{0},\mathbf{0})^\top$, $\boldsymbol{\Sigma}=\mathbf{diag}(\mathbf{1}_2)$; (c) I 型多元 Laplace 分布 : $\boldsymbol{\mu}=(\mathbf{1},-\mathbf{1})^\top$, $\boldsymbol{\Sigma}=[\mathbf{2},\mathbf{2};\mathbf{2},\mathbf{3}]$; (d) II 型多元 Laplace 分布 : $\boldsymbol{\mu}=(\mathbf{1},-\mathbf{1})^\top$, $\boldsymbol{\Sigma}=[\mathbf{2},\mathbf{2};\mathbf{2},\mathbf{3}]$

表 1 I 型和 II 型二元 Laplace 分布的密度函数的最大值和最小值

	$\boldsymbol{\mu}=(0,0)^\top$, $\boldsymbol{\Sigma}=\mathrm{diag}(\mathbf{1}_2)$		$\boldsymbol{\mu}=(1,-1)^\top$, $\boldsymbol{\Sigma}=[2,2;2,3]$	
	$f^{(\mathrm{I})}$	$f^{(\mathrm{II})}$	$f^{(\mathrm{I})}$	$f^{(\mathrm{II})}$
最大值	0.326956	0.203276	0.291899	0.148126
最小值	0.001355	0.000620	0.000010	0.000001

$f^{(\mathrm{I})}$: 由 (A.1) 定义的 I 型多元 Laplace 分布的密度函数; $f^{(\mathrm{II})}$: 由 (2.5) 定义的 II 型多元 Laplace 分布的密度函数.

显示了在两种不同的参数配置下 I 型和 II 型二元 Laplace 分布的密度函数图像. 表 1 总结了这两个分布的密度函数在每种情形下的最大值和最小值.

3 II 型多元 Laplace 分布的似然统计推断

假设 $\mathbf{x}_1,\ldots,\mathbf{x}_n$ 为来自 II 型多元 Laplace 分布 $\mathrm{Laplace}_d^{(\mathrm{II})}(\boldsymbol{\mu},\boldsymbol{\Sigma})$ 的一组样本量为 n 的随机样本, 其中 $\mathbf{x}_j=(X_{1j},\ldots,X_{dj})^\top$, $j=1,\ldots,n$. 令 $\boldsymbol{x}_j=(x_{1j},\ldots,x_{dj})^\top$ 表示随机向量 $\mathbf{x}_j$ 的实现值, 则观测数据为 $Y_{\mathrm{obs}}=\{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n\}$.

3.1 矩估计法

基于 (2.3) 给出的期望和方差 - 协方差表达式, 利用矩估计法估计参数 $\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$, 得到以下矩估计表达式:

$$\hat{\boldsymbol{\mu}}^{\mathrm{M}}=\frac{1}{n}\sum_{j=1}^{n}\boldsymbol{x}_j \quad \text{和} \quad \hat{\boldsymbol{\Sigma}}^{\mathrm{M}}=(\hat{\sigma}_{ii'}^{\mathrm{M}}), \tag{3.1}$$

其中

$$\hat{\sigma}_{ii}^{\mathrm{M}}=\sum_{j=1}^{n}\frac{(x_{ij}-\bar{x}_i)^2}{2(n-1)}, \quad i=1,\ldots,d;$$
$$\hat{\sigma}_{ii'}^{\mathrm{M}}=2\sum_{j=1}^{n}\frac{(x_{ij}-\bar{x}_i)(x_{kj}-\bar{x}_k)}{(n-1)\pi}, \quad i,i'=1,\ldots,d,$$

且 $i\neq i'$, $\bar{x}_i=\sum_{j=1}^{n}x_{ij}/n$.

3.2 极大似然估计的 ECM 算法

基于 (2.4) 中的混合表示, 对固定的 j $(j=1,\ldots,n)$, 引入一组独立同分布于 Exp(1) 的潜在变量 $U_{1j},\ldots,U_{dj}$, 则有

$$\mathbf{x}_j\mid(\mathbf{u}_j=\boldsymbol{u}_j)\sim N_d(\boldsymbol{\mu},2\boldsymbol{U}_j^{1/2}\boldsymbol{\Sigma}\boldsymbol{U}_j^{1/2}),$$

其中 $\boldsymbol{u}_j=(u_{1j},\ldots,u_{dj})^\top$ 是 $\mathbf{u}_j=(U_{1j},\ldots,U_{dj})^\top$ 的实现值,

$$\boldsymbol{U}_j^{1/2}=\mathrm{diag}(\sqrt{u_{1j}},\ldots,\sqrt{u_{dj}})=\mathrm{diag}(\boldsymbol{u}_j^{1/2})$$

是 $\mathbf{U}_j^{1/2}=\mathrm{diag}(\mathbf{u}_j^{1/2})$ 的实现值. 则缺失数据为 $Y_{\mathrm{mis}}=\{\boldsymbol{u}_j\}_{j=1}^n$, 完全数据为 $Y_{\mathrm{com}}=\{Y_{\mathrm{obs}},Y_{\mathrm{mis}}\}$. 因此, $(\boldsymbol{\mu},\boldsymbol{\Sigma})$ 的完全数据似然函数为

$$\begin{aligned}L(\boldsymbol{\mu},\boldsymbol{\Sigma}\mid Y_{\mathrm{com}})&=\prod_{j=1}^n\left\{(4\pi)^{-\frac{d}{2}}|\boldsymbol{\Sigma}|^{-\frac{1}{2}}\left(\prod_{i=1}^d u_{ij}\right)^{-\frac{1}{2}}\exp\left[-\frac{1}{4}(\boldsymbol{x}_j-\boldsymbol{\mu})^{\top}\boldsymbol{U}_j^{-\frac{1}{2}}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_j^{-\frac{1}{2}}(\boldsymbol{x}_j-\boldsymbol{\mu})-\sum_{i=1}^d u_{ij}\right]\right\}\\&\propto|\boldsymbol{\Sigma}|^{-\frac{n}{2}}\exp\left[-\frac{1}{4}\sum_{j=1}^n(\boldsymbol{x}_j-\boldsymbol{\mu})^{\top}\boldsymbol{U}_j^{-\frac{1}{2}}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_j^{-\frac{1}{2}}(\boldsymbol{x}_j-\boldsymbol{\mu})\right].\end{aligned}$$

由上式易知, $(\boldsymbol{\mu},\boldsymbol{\Sigma})$ 的完全数据对数似然函数为

$$\ell(\boldsymbol{\mu},\boldsymbol{\Sigma}\mid Y_{\mathrm{com}})=c-\frac{n}{2}\log|\boldsymbol{\Sigma}|-\frac{1}{4}\,\mathrm{tr}\left\{\boldsymbol{\Sigma}^{-1}\sum_{j=1}^n[\boldsymbol{U}_j^{-\frac{1}{2}}(\boldsymbol{x}_j-\boldsymbol{\mu})][\boldsymbol{U}_j^{-\frac{1}{2}}(\boldsymbol{x}_j-\boldsymbol{\mu})]^{\top}\right\},$$

其中 c 为与参数 $(\boldsymbol{\mu},\boldsymbol{\Sigma})$ 无关的常数. 我们采用 ECM 算法 [19] 来导出 $(\boldsymbol{\mu},\boldsymbol{\Sigma})$ 的极大似然估计. ECM 算法的条件最大化 (conditional maximization, CM) 步为计算 $\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$ 的完全数据条件极大似然估计, 分别得到

$$\hat{\boldsymbol{\mu}}=\left(\sum_{j=1}^n\boldsymbol{U}_j^{-\frac{1}{2}}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_j^{-\frac{1}{2}}\right)^{-1}\left(\sum_{j=1}^n\boldsymbol{U}_j^{-\frac{1}{2}}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_j^{-\frac{1}{2}}\boldsymbol{x}_j\right),\tag{3.2}$$

$$\hat{\boldsymbol{\Sigma}}=\frac{1}{2n}\sum_{j=1}^n\boldsymbol{U}_j^{-\frac{1}{2}}(\boldsymbol{x}_j-\boldsymbol{\mu})(\boldsymbol{x}_j-\boldsymbol{\mu})^{\top}\boldsymbol{U}_j^{-\frac{1}{2}}.\tag{3.3}$$

期望 (expectation, E) 步将 (3.2) 和 (3.3) 中的 $u_{ij}^{-\frac{1}{2}}u_{kj}^{-\frac{1}{2}}$ 由其条件期望 $E(U_{ij}^{-\frac{1}{2}}U_{kj}^{-\frac{1}{2}}\mid Y_{\mathrm{obs}},\boldsymbol{\mu},\boldsymbol{\Sigma})$ 进行替代, 其中 $i,k=1,\ldots,d$, $j=1,\ldots,n$. 由 (2.5) 可计算 $\mathbf{u}_j\mid\mathbf{x}_j$ 的条件分布为

$$f_{\mathbf{u}_j\mid\mathbf{x}_j}(\boldsymbol{u}_j\mid\boldsymbol{x}_j)=\frac{1}{c_j}\left(\prod_{i=1}^d u_{ij}\right)^{-\frac{1}{2}}\exp\left[-\frac{1}{4}(\boldsymbol{u}_j^{-\frac{1}{2}})^{\top}(\boldsymbol{X}_j^*\boldsymbol{\Sigma}^{-1}\boldsymbol{X}_j^*)\boldsymbol{u}_j^{-\frac{1}{2}}-\sum_{i=1}^d u_{ij}\right],\tag{3.4}$$

其中 $\boldsymbol{X}_j^*=\mathrm{diag}(\boldsymbol{x}_j-\boldsymbol{\mu})$, c_j 是归一化常数. 由于无法给出上述条件分布的显式表达, 我们可以通过数值方法近似地计算 (3.5) 中的归一化常数 c_j 和 (3.6) 中的条件期望, 可借助 R 软件包 cubature 中用于超立方体上的多元积分计算的函数 adaptIntegrate. 因此, c_j 的数值可通过以下积分计算:

$$c_j=\int_{\mathbb{R}_+^d}\left(\prod_{i=1}^d u_{ij}\right)^{-\frac{1}{2}}\exp\left[-\frac{1}{4}(\boldsymbol{u}_j^{-\frac{1}{2}})^{\top}(\boldsymbol{X}_j^*\boldsymbol{\Sigma}^{-1}\boldsymbol{X}_j^*)\boldsymbol{u}_j^{-\frac{1}{2}}-\sum_{i=1}^d u_{ij}\right]d\boldsymbol{u}_j,\tag{3.5}$$

$j=1,\ldots,n$, 从而可计算相应的条件期望

$$\begin{aligned}&E(U_{ij}^{-\frac{1}{2}}U_{kj}^{-\frac{1}{2}}\mid\boldsymbol{x}_j,\boldsymbol{\mu},\boldsymbol{\Sigma})\\&=\frac{1}{c_j}\int_{\mathbb{R}_+^d}u_{ij}^{-\frac{1}{2}}u_{kj}^{-\frac{1}{2}}\left(\prod_{i=1}^d u_{ij}\right)^{-\frac{1}{2}}\exp\left[-\frac{1}{4}(\boldsymbol{u}_j^{-\frac{1}{2}})^{\top}(\boldsymbol{X}_j^*\boldsymbol{\Sigma}^{-1}\boldsymbol{X}_j^*)\boldsymbol{u}_j^{-\frac{1}{2}}-\sum_{i=1}^d u_{ij}\right]d\boldsymbol{u}_j,\end{aligned}\tag{3.6}$$

$i,k=1,\ldots,d$ 且 $j=1,\ldots,n$.

因此, 采用 ECM 算法来计算 $(\boldsymbol{\mu},\boldsymbol{\Sigma})$ 的极大似然估计就是联合 (3.2) 和 (3.3) 中的 CM 步和 (3.6) 中的 E 步. 令 $\boldsymbol{\theta}^{(t+1)}$ 表示 $\hat{\boldsymbol{\theta}}$ 的第 $t+1$ 次迭代结果, 则设定 ECM 算法的停止法则为

$$\|\boldsymbol{\theta}^{(t+1)}-\boldsymbol{\theta}^{(t)}\|\leqslant\delta,$$

其中 δ 为预先设定的精度.

由以上 ECM 算法的具体步骤可以看出参数的完全数据极大似然估计均有显式表达, 难点在于计算潜在变量及其函数的条件期望, 所涉及的多元积分问题可借助现有的技术方法解决. 第 5 和 6 节的模拟实验及实例分析的结果也表明所提出的 ECM 算法在实际中可以较好地被运用.

3.3 自助抽样法的置信区间

自助抽样 (bootstrap) 方法可用来构造关于参数 $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 的置信区间. 参数 $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 的极大似然估计 $(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$ 可由 (3.2)、(3.3) 和 (3.6) 定义的 ECM 算法得到. 令 $\vartheta = h(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 表示 $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 的任一函数, 则 ϑ 的极大似然估计为 $\hat{\vartheta} = h(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$. 基于所得到的极大似然估计 $(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$, 结合 (2.1) 所定义的随机表示可以独立地从分布 $\mathrm{Laplace}_d^{(\mathrm{II})}(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$ 中产生一组自助抽样样本 $Y_{\mathrm{obs}}^* = \{\boldsymbol{x}_j^*\}_{j=1}^n$. 依据 (3.2)、(3.3) 和 (3.6) 定义的 ECM 算法计算极大似然估计 $(\hat{\boldsymbol{\mu}}^*, \hat{\boldsymbol{\Sigma}}^*)$ 从而得到 $\hat{\vartheta}^*$. 将此过程重复 G 次, 得到 G 个自助抽样参数估计 $\{\hat{\vartheta}_1^*, \ldots, \hat{\vartheta}_G^*\}$. 因此, $\hat{\vartheta}$ 的标准误差 $\mathrm{se}(\hat{\vartheta})$ 可由 G 个自助抽样的样本标准差来估计, 即

$$\widehat{\mathrm{se}}(\hat{\vartheta}) = \left\{ \frac{1}{G-1} \sum_{g=1}^{G} \left[\hat{\vartheta}_g^* - \frac{\hat{\vartheta}_1^* + \cdots + \hat{\vartheta}_G^*}{G} \right]^2 \right\}^{1/2}. \tag{3.7}$$

如果 $\{\hat{\vartheta}_1^*, \ldots, \hat{\vartheta}_G^*\}$ 近似正态分布, 则 ϑ 的一个置信水平为 $(1-\alpha)100\%$ 的自助抽样置信区间为

$$[\hat{\vartheta} - z_{\alpha/2} \cdot \widehat{\mathrm{se}}(\hat{\vartheta}), \hat{\vartheta} + z_{\alpha/2} \cdot \widehat{\mathrm{se}}(\hat{\vartheta})], \tag{3.8}$$

其中 $z_{\alpha/2}$ 表示标准正态分布的上 $\alpha/2$ 分位点. 如果 $\{\hat{\vartheta}_1^*, \ldots, \hat{\vartheta}_G^*\}$ 并非近似正态分布, 则 ϑ 的一个置信水平为 $(1-\alpha)100\%$ 的自助抽样置信区间为

$$[\hat{\vartheta}_{\mathrm{L}}, \hat{\vartheta}_{\mathrm{U}}], \tag{3.9}$$

其中 $\hat{\vartheta}_{\mathrm{L}}$ 和 $\hat{\vartheta}_{\mathrm{U}}$ 分别为 $\{\hat{\vartheta}_1^*, \ldots, \hat{\vartheta}_G^*\}$ 的 $\alpha/2$ 和 $1-\alpha/2$ 分位点.

(3.8) 和 (3.9) 分别基于参数自助抽样估计的正态性和非正态性假设构建了两种置信区间. 已有的一些正态性检验统计量如 Anderson-Darling 检验、Cramer-Von Mises 检验和 Kolmogorov-Smirnov-Lilliefors 检验等可用于检验正态假设的有效性. 考虑到在实际情形中, 样本本身的抽样分布一般偏离正态分布较远且样本量有限, 因此在多数情形下假定参数的自助抽样估计的样本服从正态分布并不是一个十分合理的假设. 在实际应用中, 采用 (3.9) 定义的自助抽样置信区间更加可靠.

3.4 分量的独立性检验

由 (2.1) 定义的随机表示易知 $U_1, \ldots, U_d$ 是相互独立的. 如果所有的 $\sigma_{ij} = 0$ $(i \neq j)$, 则所有的分量 $X_1, \ldots, X_d$ 是相互独立的. 现要进行如下检验:

$$H_0\colon\ \sigma_{ij} = 0,\ \text{其中}\ i, j = 1, \ldots, d,\ i \neq j \quad \text{vs.} \quad H_1\colon\ H_0\ \text{非真}. \tag{3.10}$$

在原假设 H_0 下, 似然比检验 (likelihood ratio test, LRT) 统计量为

$$T = -2\{\ell(\hat{\boldsymbol{\mu}}_0, \hat{\boldsymbol{\Sigma}}_0 \mid Y_{\mathrm{obs}}) - \ell(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}} \mid Y_{\mathrm{obs}})\} \mathrel{\dot{\sim}} \chi^2\left(\frac{d(d-1)}{2}\right), \tag{3.11}$$

其中 $(\hat{\boldsymbol{\mu}}_0, \hat{\boldsymbol{\Sigma}}_0)$ 表示在原假设下参数 $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 带约束的极大似然估计, $(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$ 表示全空间下的极大似然估计, 则 $(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$ 为每一个分量的参数 (μ_i, σ_{ii}) 在边际分布中的极大似然估计的集合. 在原假设成立时, 检验统计量 T 近似地服从于自由度为 $d(d-1)/2$ 的卡方分布, 且相应的 p 值可由下式计算:

$$p = \begin{cases} \Pr(T > t \mid H_0), & \text{当 } d \leqslant 2 \text{ 时}, \\ 2\min\{\Pr(T > t \mid H_0), \Pr(T \leqslant t \mid H_0)\}, & \text{当 } d \geqslant 3 \text{ 时}. \end{cases} \tag{3.12}$$

对于给定的显著性水平 α, 当 $p \leqslant \alpha$ 时拒绝原假设.

4 Bayes 方法

4.1 后验模型的 ECM 算法

4.1.1 无信息先验分布

如果实验前我们对参数的了解很少, 那么就需要考虑以下无信息 (或弥散) 先验分布:

$$\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{d+1}{2}}. \tag{4.1}$$

特别地, 当 $d = 1$ 时, (4.1) 右边变成 $1/\sigma^2$. 通过引入与第 3.2 小节相同的潜在变量 U_{ij}, 结合 (3.2) 和 (4.1), 可得到完全数据的后验分布:

$$p(\boldsymbol{\mu}, \boldsymbol{\Sigma} \mid Y_{\text{com}}) \propto |\boldsymbol{\Sigma}|^{-\frac{n+d+1}{2}} \exp\left[-\frac{1}{4}\sum_{j=1}^{n}(\boldsymbol{x}_j - \boldsymbol{\mu})^\top \boldsymbol{U}_j^{-\frac{1}{2}} \boldsymbol{\Sigma}^{-1} \boldsymbol{U}_j^{-\frac{1}{2}} (\boldsymbol{x}_j - \boldsymbol{\mu})\right].$$

CM 步是计算参数 $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 的完全数据的后验众数:

$$\tilde{\boldsymbol{\mu}} = \boldsymbol{A}_1^{-1}\boldsymbol{b}_1 \quad \text{和} \quad \tilde{\boldsymbol{\Sigma}} = \frac{\boldsymbol{C}_1}{n+d+1}, \tag{4.2}$$

其中

$$\begin{aligned} \boldsymbol{A}_1 &= \frac{1}{2}\sum_{j=1}^{n} \boldsymbol{U}_j^{-\frac{1}{2}} \boldsymbol{\Sigma}^{-1} \boldsymbol{U}_j^{-\frac{1}{2}}, \\ \boldsymbol{b}_1 &= \frac{1}{2}\sum_{j=1}^{n} \boldsymbol{U}_j^{-\frac{1}{2}} \boldsymbol{\Sigma}^{-1} \boldsymbol{U}_j^{-\frac{1}{2}} \boldsymbol{x}_j, \\ \boldsymbol{C}_1 &= \frac{1}{2}\sum_{j=1}^{n} \boldsymbol{U}_j^{-\frac{1}{2}} (\boldsymbol{x}_j - \boldsymbol{\mu})(\boldsymbol{x}_j - \boldsymbol{\mu})^\top \boldsymbol{U}_j^{-\frac{1}{2}}. \end{aligned}$$

E 步是将 (4.2) 中的 $u_{ij}^{-\frac{1}{2}} u_{kj}^{-\frac{1}{2}}$ 替换为各自相应的条件期望值.

为了实现第 4.2 小节将介绍的数据增强 (data augmentation, DA) 算法, 我们需要用到 $\boldsymbol{\mu} \mid (Y_{\text{com}}, \boldsymbol{\Sigma})$ 和 $\boldsymbol{\Sigma} \mid (Y_{\text{com}}, \boldsymbol{\mu})$ 的条件后验分布, 分别为

$$\boldsymbol{\mu} \mid (Y_{\text{com}}, \boldsymbol{\Sigma}) \sim N_d(\boldsymbol{A}_1^{-1}\boldsymbol{b}_1, \boldsymbol{A}_1^{-1}) \quad \text{和} \quad \boldsymbol{\Sigma} \mid (Y_{\text{com}}, \boldsymbol{\mu}) \sim \text{IWishart}(\boldsymbol{C}_1, n). \tag{4.3}$$

4.1.2 共轭先验分布

当均值和协方差矩阵未知时, 多元正态分布的共轭先验分布是正态逆 Wishart (normal inverse Wishart, NIW) 分布 [20]. 考虑以下共轭先验:

$$\boldsymbol{\mu} \mid \boldsymbol{\Sigma} \sim N_d\left(\boldsymbol{\mu}_0, \frac{\boldsymbol{\Sigma}}{\kappa_0}\right) \quad \text{和} \quad \boldsymbol{\Sigma} \sim \mathrm{IWishart}_d(\boldsymbol{\Lambda}_0, \nu_0),$$

其中 $\nu_0 > d-1$, 也就是 $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 服从 NIW 先验分布, 记为 $\mathrm{NIW}(\mu_0, \kappa_0, \boldsymbol{\Lambda}_0, \nu_0)$. 我们可以得到完全数据的后验分布:

$$\begin{aligned} p(\boldsymbol{\mu}, \boldsymbol{\Sigma} \mid Y_{\mathrm{com}}) &\propto L(\boldsymbol{\mu}, \boldsymbol{\Sigma} \mid Y_{\mathrm{com}}) \cdot \pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \\ &\propto |\boldsymbol{\Sigma}|^{-\frac{n+\nu_0+d+2}{2}} \exp\Bigg[-\frac{1}{4}\sum_{j=1}^{n}(\boldsymbol{x}_j-\boldsymbol{\mu})^{\top}\boldsymbol{U}_j^{-\frac{1}{2}}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_j^{-\frac{1}{2}}(\boldsymbol{x}_j-\boldsymbol{\mu}) \\ &\quad -\frac{1}{2}\kappa_0(\boldsymbol{\mu}-\boldsymbol{\mu}_0)^{\top}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}_0) - \frac{1}{2}\operatorname{tr}(\boldsymbol{\Lambda}_0\boldsymbol{\Sigma}^{-1})\Bigg]. \end{aligned}$$

CM 步是计算参数 $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 的完全数据的后验众数:

$$\tilde{\boldsymbol{\mu}} = \boldsymbol{A}_2^{-1}\boldsymbol{b}_2 \quad \text{和} \quad \tilde{\boldsymbol{\Sigma}} = \frac{\boldsymbol{C}_2}{n+\nu_0+d+2}, \tag{4.4}$$

其中

$$\begin{aligned} \boldsymbol{A}_2 &= \frac{1}{2}\sum_{j=1}^{n}\boldsymbol{U}_j^{-\frac{1}{2}}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_j^{-\frac{1}{2}} + \kappa_0\boldsymbol{\Sigma}^{-1}, \\ \boldsymbol{b}_2 &= \frac{1}{2}\sum_{j=1}^{n}\boldsymbol{U}_j^{-\frac{1}{2}}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_j^{-\frac{1}{2}}\boldsymbol{x}_j + \kappa_0\boldsymbol{\Sigma}^{-1}\mu_0, \\ \boldsymbol{C}_2 &= \frac{1}{2}\sum_{j=1}^{n}\boldsymbol{U}_j^{-\frac{1}{2}}(\boldsymbol{x}_j-\boldsymbol{\mu})(\boldsymbol{x}_j-\boldsymbol{\mu})^{\top}\boldsymbol{U}_j^{-\frac{1}{2}} + \kappa_0(\boldsymbol{\mu}-\boldsymbol{\mu}_0)(\boldsymbol{\mu}-\boldsymbol{\mu}_0)^{\top} + \boldsymbol{\Lambda}_0. \end{aligned}$$

E 步是将 (4.4) 中的 $u_{ij}^{-\frac{1}{2}}u_{kj}^{-\frac{1}{2}}$ 替换为各自相应的条件期望值.

为了实现第 4.2 小节将介绍的 DA 算法, 我们需要用到 $\boldsymbol{\mu} \mid (Y_{\mathrm{com}}, \boldsymbol{\Sigma})$ 和 $\boldsymbol{\Sigma} \mid (Y_{\mathrm{com}}, \boldsymbol{\mu})$ 的条件后验分布, 分别为

$$\boldsymbol{\mu} \mid (Y_{\mathrm{com}}, \boldsymbol{\Sigma}) \sim N_d(\boldsymbol{A}_2^{-1}\boldsymbol{b}_2, \boldsymbol{A}_2^{-1}) \quad \text{和} \quad \boldsymbol{\Sigma} \mid (Y_{\mathrm{com}}, \boldsymbol{\mu}) \sim \mathrm{IWishart}_d(\boldsymbol{C}_2, n+\nu_0+1). \tag{4.5}$$

4.2 生成后验样本的内嵌 AR 算法的 DA 算法

为了对参数 $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 进行完整的 Bayes 推断, 我们需要使用 DA 算法 [21] 来生成服从后验分布 $p(\boldsymbol{\mu}, \boldsymbol{\Sigma} \mid Y_{\mathrm{obs}})$ 的数据. 输入 (imputation, I) 步是从 (3.4) 中的条件分布 $f_{\mathbf{u}_j \mid \mathbf{x}_j}(\boldsymbol{u}_j \mid \boldsymbol{x}_j)$ 中抽取变量 $\{\mathbf{u}_j\}_{j=1}^n$, 记其实现值为 $\{\boldsymbol{u}_j\}_{j=1}^n$. 由于函数没有显式表达, 所以使用接受 - 拒绝 (acceptance-rejection, AR) 算法 [22] 来实现这一步. 对于每个未归一化的目标密度函数 $f_{\mathbf{u}_j \mid \mathbf{x}_j}(\boldsymbol{u}_j \mid \boldsymbol{x}_j)$, 选择如下套密度函数 (envelope density function):

$$g_{\mathbf{u}_j}(\boldsymbol{u}_j) = \prod_{i=1}^{d}\mathrm{Gamma}(u_{ij} \mid 0.5, 1),$$

其中 $\mathrm{Gamma}(u \mid \alpha, \beta)$ 表示 $\mathrm{Gamma}(\alpha, \beta)$ 的密度函数. 由于 $\boldsymbol{\Sigma}$ 作为多元正态分布的协方差矩阵, 是一个非负定矩阵, 所以, $\boldsymbol{X}_j^* \boldsymbol{\Sigma}^{-1} \boldsymbol{X}_j^*$ 也是非负定的, 我们有 $(\boldsymbol{u}_j^{-\frac{1}{2}})^\top (\boldsymbol{X}_j^* \boldsymbol{\Sigma}^{-1} \boldsymbol{X}_j^*) \boldsymbol{u}_j^{-\frac{1}{2}} \geqslant 0$ 成立, 则

$$f_{\mathbf{u}_j \mid \mathbf{x}_j}(\boldsymbol{u}_j \mid \boldsymbol{x}_j) \leqslant \prod_{i=1}^{d} u_{ij}^{-\frac{1}{2}} \mathrm{e}^{-u_{ij}} = (\sqrt{\pi})^d \cdot g(\boldsymbol{u}_j).$$

套常数 (envelope constant) $c = (\sqrt{\pi})^d > 1$. 通过 AR 算法生成 $\{\mathbf{u}_j = \boldsymbol{u}_j\}_{j=1}^n$ 的过程如下.

步骤 1 首先, 从均匀分布 $U(0,1)$ 中抽取变量 $V_j = v_j$, 并独立地从分布 $\mathrm{Gamma}(0.5, 1)$ 中生成数据 $Y_{ij} = y_{ij}$, $i = 1, \ldots, d$. 设 $\mathbf{y}_j = (Y_{1j}, \ldots, Y_{dj})^\top$ 且 $\boldsymbol{y}_j = (y_{1j}, \ldots, y_{dj})^\top$.

步骤 2 如果满足

$$v_j \leqslant \frac{f_{\mathbf{u}_j \mid \mathbf{x}_j}(\boldsymbol{y}_j \mid \boldsymbol{x}_j)}{(\sqrt{\pi})^d \prod_{i=1}^{d} \mathrm{Gamma}(y_{ij} \mid 0.5, 1)},$$

则将 $\boldsymbol{y}_j$ 赋值给 $\mathbf{u}_j$; 否则回到步骤 1.

重复上述步骤从 $j = 1$ 到 $j = n$. 后验 (posterior, P) 步是通过 Gibbs 采样方法, 在给定 $(Y_{\mathrm{obs}}, \boldsymbol{u}_1, \ldots, \boldsymbol{u}_n)$ 时, 生成数据 $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, 即在给定 $(Y_{\mathrm{com}}, \boldsymbol{\mu})$ 时生成 $\boldsymbol{\Sigma}$, 然后根据 (4.3) 或 (4.5), 在给定 $(Y_{\mathrm{com}}, \boldsymbol{\Sigma})$ 时更新 $\boldsymbol{\mu}$.

5 模拟研究

本节通过仿真模拟来评估所提出的矩估计和极大似然估计方法的性能. 首先, 设定 $d = 2$ 和 $d = 3$, 分别配置不同的 $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 组合, 并选定样本量为 $n = 20$ 和 $n = 50$. 在给定的 $(n, d, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ 下, 基于 (2.1) 中的随机表示分别产生 100 组来自于 $\mathrm{Laplace}_d^{(\mathrm{II})}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 的随机样本. 对于每一组随机样本, 利用 (3.1)–(3.3) 和 (3.6) 定义的 ECM 算法, 可以计算参数 $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 的矩估计值和极大似然估计值, 结果如表 2 和 3 所示.

表 2 和 3 分别给出不同的 $(n, d, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ 组合下参数的矩估计、极大似然估计、均方误差和每组样本利用 ECM 算法所需收敛次数的平均值, 其中 $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 的矩估计和极大似然估计分别为 100 组样本得到的估计的平均值, 均方误差由 100 组样本得到的估计的方差和偏差的平方之和计算得到. 从估计结果可以看出, 参数 $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 的极大似然估计值比矩估计值更接近真实值, 矩估计法得到的均方误差总是

表 2 II 型二元 Laplace 分布参数极大似然估计与矩估计的比较

样本量	估计方法	参数					迭代次数
		$\mu_1 = 0$	$\mu_2 = 0$	$\sigma_{11} = 0.2$	$\sigma_{12} = 0.2$	$\sigma_{22} = 0.3$	
$n = 20$	极大似然估计	0.0013	−0.0024	0.2033	0.1936	0.2873	15.13
		(0.0047)	(0.0041)	(0.0083)	(0.0091)	(0.0195)	
	矩估计	−0.0398	0.0014	0.2082	0.2110	0.3011	
		(0.0226)	(0.0301)	(0.0109)	(0.0096)	(0.0188)	
$n = 50$	极大似然估计	0.0012	−0.0073	0.1960	0.1946	0.2928	13.33
		(0.0020)	(0.0023)	(0.0029)	(0.0029)	(0.0062)	
	矩估计	−0.0048	−0.0061	0.1935	0.1943	0.2990	
		(0.0085)	(0.0120)	(0.0041)	(0.0036)	(0.0077)	

注: 极大似然估计的精度为 10^{-3}.

表 3 II 型三元 Laplace 分布参数极大似然估计与矩估计的比较

样本量	估计方法	参数									迭代次数
		$\mu_1=-1$	$\mu_2=0$	$\mu_3=1$	$\sigma_{11}=1$	$\sigma_{12}=0.5$	$\sigma_{13}=-0.5$	$\sigma_{22}=1$	$\sigma_{23}=-0.5$	$\sigma_{33}=1$	
$n=20$	极大似然估计	−0.9955	−0.0329	1.0092	0.9660	0.4864	−0.4850	0.9908	−0.4883	0.9650	9.07
		(0.0279)	(0.0268)	(0.0253)	(0.2339)	(0.1050)	(0.1086)	(0.2440)	(0.0978)	(0.1762)	
	矩估计	−0.9965	−0.0126	1.0294	1.0381	0.5590	−0.5501	1.0497	−0.5911	1.0426	
		(0.1152)	(0.1277)	(0.1037)	(0.3149)	(0.1407)	(0.1348)	(0.3476)	(0.1469)	(0.2235)	
$n=50$	极大似然估计	−1.0268	0.0082	0.9909	1.0269	0.4996	−0.5179	1.0531	−0.5015	1.0026	7.90
		(0.0108)	0.0138	(0.0076)	(0.0800)	(0.0340)	(0.0301)	(0.0848)	(0.0306)	(0.0727)	
	矩估计	−0.9853	−0.0219	1.0073	1.0462	0.5061	−0.5337	1.0332	−0.5160	1.0420	
		(0.0409)	(0.0372)	(0.0346)	(0.1004)	(0.0564)	(0.0465)	(0.1157)	(0.0492)	(0.1077)	

注: 极大似然估计的精度为 10^{-2}.

大于极大似然估计法得到的均方误差, 且两种方法的估计值都随着样本量的增加越来越接近真实值, 均方误差随着样本量的增加而减小.

6 实例分析

6.1 雅典道路交通事故数据分析

6.1.1 数据描述和边际分析

文献 [23] 提供了 1987 至 1991 年雅典 24 条中心道路的事故数量. 数据仅包含造成人员伤亡的事故, 并且每条道路是在一定的长度上进行统计. 我们通过将事故数除以相应的道路长度, 将数据转化为每公里道路上的平均事故数, 从而假设数据是连续的. 我们从中选择 1988 和 1991 年的数据进行分析, 设 X_1 和 X_2 分别表示 1988 和 1991 年的平均事故数, 记为 $\mathbf{x}=(X_1,X_2)^\top$. 首先采用一元正态分布、t 分布和 Laplace 分布分别对 X_1 和 X_2 的边际分布进行拟合. 基于对数似然 (log-likelihood) 准则、Akaike 信息准则 (Akaike information criterion, AIC) 和 Bayes 信息准则 (Bayesian information criterion, BIC) 得到的结果列在表 4 中. 由表中数据可知在 AIC 和 BIC 下, Laplace 分布是最适合的 X_1 和 X_2 的边际分布, 即

$$X_1\sim \mathrm{Laplace}(\hat{\mu}_1=12.695,\hat{\sigma}_1^2=7.539^2)\quad 和\quad X_2\sim \mathrm{Laplace}(\hat{\mu}_2=5.414,\hat{\sigma}_2^2=2.715^2).$$

同时根据描述性统计量

$$\bar{\mathbf{x}}=(14.314,6.055)^\top,\quad \mathrm{Cov}(\mathbf{x})=\begin{pmatrix}123.983 & 19.199\\ 19.199 & 17.827\end{pmatrix},\quad \mathrm{Corr}(X_1,X_2)=0.408,$$

考虑使用 II 型二元 Laplace 分布来拟合该数据.

6.1.2 统计分析

假设 $\mathbf{x}_j=(X_{1j},X_{2j})^\top$ $(j=1,\ldots,n)$ 是来自 $\mathrm{Laplace}_2^{(\mathrm{II})}(\boldsymbol{\mu},\boldsymbol{\Sigma})$ 的独立同分布的随机样本, $\boldsymbol{x}_j=$

表 4　雅典道路交通事故数据边际分布比较

变量	分布	对数似然准则	AIC	BIC
X_1	正态分布	−91.386	186.771	189.127
	t 分布	−88.757	183.514	187.048
	Laplace 分布	−89.118	182.236	184.592
X_2	正态分布	−68.112	140.225	142.581
	t 分布	−64.470	134.940	138.474
	Laplace 分布	−64.609	133.217	135.573

$(x_{1j}, x_{2j})^\top$ 表示 $\mathbf{x}_j$ 的观测值, 其中 $n = 24$. 由 (3.1) 可得参数的矩估计如下:

$$\hat{\boldsymbol{\mu}}^{\mathrm{M}} = (14.314, 6.055)^\top, \quad \hat{\boldsymbol{\Sigma}}^{\mathrm{M}} = \begin{pmatrix} 61.992 & 12.222 \\ 12.222 & 8.914 \end{pmatrix}.$$

为了计算参数的极大似然估计, 选取上述矩估计值作为参数的初始值. 根据 ECM 算法 (3.2)、(3.3) 和 (3.6), 我们得到精度为 $\delta = 10^{-3}$ 的 $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 的极大似然估计值, 如表 5 第 2 列所示. 利用第 3.3 小节中介绍的自助抽样法产生 $G = 1{,}000$ 组自助抽样样本后, 表 5 的第 3–5 列分别展示了参数 $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 的自助抽样标准差及由 (3.8) 和 (3.9) 定义的两个置信水平为 95% 的自助抽样置信区间.

我们要检测 X_1 与 X_2 之间的独立性, 即检验原假设 H_0: $\sigma_{12} = 0$ 与备择假设 H_1: $\sigma_{12} \neq 0$. 由 (3.11), 可计算得似然比统计量值为 $t = 11.581$, 由 (3.12) 可知相应的 p 值为 0.001. 因此, 对于给定的显著性水平 0.05, 拒绝原假设.

此外, 采用 $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 的共轭先验分布分别为 $\boldsymbol{\mu} \mid \boldsymbol{\Sigma} \sim N_2(\mathbf{0}, \boldsymbol{\Sigma}/1)$ 和 $\boldsymbol{\Sigma} \sim \mathrm{IWishart}_2(\mathbf{I}_2, 2)$. ECM 算法 (4.4) 和 (3.6) 收敛到后验众数, 如表 6 的第 2 列所示. 为了计算 $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 的 Bayes 可信区间 (credible interval), 我们使用内嵌了 AR 算法的 DA 算法 (详情见第 4.2 小节) 生成 $L = 6{,}000$ 组后验样本. 丢弃前半部分样本后, 可以计算得到表 6 最后 3 列展示的参数后验均值、后验标准差和 95% Bayes 可信区间.

6.1.3　模型比较

由 (1.2) 定义的 I 型多元 Laplace 分布不适用于这个数据集. 因此, 在多元正态分布 $N_2(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$、多元 t 分布 $t_2(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \nu)$ 和 II 型多元 Laplace 分布 $\mathrm{Laplace}_2^{(\mathrm{II})}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 之间进行模型选择. 可分别计算在 3 个分布模型下对应参数的极大似然估计值. 表 7 列出 3 种分布的结果比较, 其中当 $\nu > 2$ 时, 二元 t 分布的协方差矩阵为 $[\nu/(\nu-2)]\boldsymbol{\Sigma}_2$. 二元 II 型 Laplace 分布的协方差矩阵由 (2.3) 给出. 我们可以看到在 AIC 和 BIC 下, II 型多元 Laplace 分布表现更优.

表 5　雅典道路交通事故数据参数的极大似然估计和自助抽样置信区间

参数	极大似然估计	标准差 [B]	95% 自助抽样置信区间 [†]	95% 自助抽样置信区间 [‡]
μ_1	11.124	0.288	[10.557, 11.685]	[10.491, 11.725]
μ_2	5.0000	0.105	[4.7950, 5.2060]	[4.7650, 5.2390]
σ_{11}	57.923	12.11	[34.536, 82.016]	[36.850, 84.156]
σ_{12}	14.728	4.083	[6.7340, 22.740]	[7.6240, 23.306]
σ_{22}	7.2440	1.542	[4.2160, 10.261]	[4.4590, 10.647]

注: 标准差 [B] 由 (3.7) 定义, 95% 自助抽样置信区间 [†] 由 (3.8) 定义, 95% 自助抽样置信区间 [‡] 由 (3.9) 定义.

表 6 雅典道路交通事故数据参数的后验估计

参数	后验众数	后验均值	后验标准差	95% Bayes 可信区间
μ_1	11.078	10.463	1.2970	[7.814, 13.04]
μ_2	4.9030	4.5520	0.5520	[3.543, 5.631]
σ_{11}	43.494	46.111	10.485	[27.41, 70.85]
σ_{12}	12.922	12.927	2.8270	[6.864, 17.52]
σ_{22}	6.2200	6.3210	1.2480	[3.956, 8.753]

表 7 雅典道路交通事故数据的模型比较

	多元正态分布	多元 t 分布	II 型多元 Laplace 分布
对数似然	−157.309	−147.227	−147.936
AIC	324.618	306.454	305.872
BIC	330.508	313.522	311.762
均值向量 ($\boldsymbol{\mu}$)	$(14.314,\ 6.055)^\top$	$(11.431,\ 5.365)^\top$	$(11.124,\ 5.000)^\top$
协方差矩阵	$\begin{pmatrix} 118.817 & 18.399 \\ 18.399 & 17.084 \end{pmatrix}$	$\begin{pmatrix} 458.065 & 141.990 \\ 141.990 & 90.167 \end{pmatrix}$	$\begin{pmatrix} 115.845 & 23.119 \\ 23.119 & 14.489 \end{pmatrix}$
相关系数	0.408	0.699	0.564

6.2 犯罪数据分析

6.2.1 数据描述

文献 [23] 报告了 1997 年从希腊国家统计局获取的 5 种不同类型犯罪的数据集. 我们选取其中 3 种犯罪的数据, 分别是强奸 (X_1)、走私文物 (X_2) 和偷渡 (X_3). 这组数据包含了希腊 50 个地方行政区域的犯罪情形, 人口计数以百万为单位. 将罪案数量除以相应的人口, 我们可以得到每百万人口的平均犯罪数. 设 $\mathbf{x} = (X_1, X_2, X_3)^\top$, 则样本均值和样本相关系数矩阵分别为

$$\bar{\mathbf{x}} = (22.425, 11.335, 16.195)^\top$$

和

$$\mathrm{Corr}(\mathbf{x}) = \begin{pmatrix} 1.000 & 0.106 & 0.268 \\ 0.106 & 1.000 & 0.215 \\ 0.268 & 0.215 & 1.000 \end{pmatrix}.$$

为了检验这 3 种不同违法活动之间的关系, 考虑用 II 型三元 Laplace 分布来拟合数据集.

6.2.2 统计分析

假设 $\mathbf{x}_j = (X_{1j}, X_{2j}, X_{3j})^\top$ $(j = 1, \ldots, n)$ 为来自 $\mathrm{Laplace}_3^{(\mathrm{II})}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 的独立同分布的随机样本, 其中 $n = 50$. 我们选用相应的矩估计值作为参数 $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 的初始值, 参数 $\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$ 的极大似然估计值经过 80 次迭代后, 以精度 $\delta = 0.01$ 收敛至

$$\hat{\boldsymbol{\mu}} = (7.680, -0.008, -0.040)^\top$$

表 8　II 型三元 Laplace 分布拟合犯罪数据下的参数的极大似然估计与自助抽样置信区间

参数	均值	标准差 B	95% 自助抽样置信区间 †	95% 自助抽样置信区间 ‡
μ_1	7.644	0.480	[6.7040, 8.585]	[6.7110, 8.647]
μ_2	−0.008	0.317	[−0.630, 0.615]	[−0.710, 0.651]
μ_3	−0.037	0.464	[−0.947, 0.872]	[−1.109, 0.926]
σ_{11}	368.487	52.008	[266.552, 470.422]	[275.957, 483.671]
σ_{12}	96.733	28.504	[40.864, 152.601]	[43.163, 157.664]
σ_{13}	124.951	40.165	[46.228, 203.674]	[53.899, 210.158]
σ_{22}	132.526	19.778	[93.761, 171.292]	[97.806, 175.976]
σ_{23}	82.786	24.485	[34.795, 130.776]	[40.115, 139.730]
σ_{33}	262.865	38.378	[187.645, 338.086]	[196.336, 350.365]

注: 标准差 B 由 (3.7) 定义, 95% 自助抽样置信区间 † 由 (3.8) 定义, 95% 自助抽样置信区间 ‡ 由 (3.9) 定义.

和

$$\hat{\boldsymbol{\Sigma}} = \begin{pmatrix} 369.071 & 97.724 & 124.147 \\ 97.724 & 131.087 & 81.377 \\ 124.147 & 81.377 & 262.831 \end{pmatrix}.$$

在产生 $G = 1{,}000$ 组自助抽样样本后, 自助抽样样本的标准差和参数的两个置信水平为 95% 的自助抽样置信区间如表 8 所示.

我们要检验 $\mathbf{x}$ 中分量的独立性, 即检验原假设 H_0: $\sigma_{12} = \sigma_{13} = \sigma_{23} = 0$ 与备择假设 H_1: H_0 非真. 在原假设 H_0 下, 参数 $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 有约束的极大似然估计即是每个一元的边际估计值. 基于 (3.11) 和 (3.12), 似然比检验统计量的值为 $t = 14.212$, 对应的 p 值为

$$p = 2\min\{\Pr(\chi^2(3) > t), \Pr(\chi^2(3) \leqslant t)\} = 0.005 \leqslant 0.05.$$

因此, 对于给定的显著性水平 0.05, 拒绝原假设 H_0.

6.2.3　模型比较

如果我们使用 (1.2) 中定义的 I 型多元 Laplace 分布来拟合这组数据, 并采用似然比检验统计量来检验 $\boldsymbol{\Sigma}$ 的相关结构, 需计算有约束和无约束情形下参数的极大似然估计值 (见附录 A.2), 可得到检验统计量值为 $t' = 1.3741$, 相应的 p 值为

$$p' = 2\min\{\Pr(\chi^2(3) > t'), \Pr(\chi^2(3) \leqslant t')\} = 0.577 \geqslant 0.05.$$

因此, 我们不能拒绝原假设, 即 $\boldsymbol{\Sigma}$ 中所有的非对角元素都为 0. 注意到当 $\boldsymbol{\Sigma}$ 为对角矩阵时, 我们有这样的结论: 向量 $\mathbf{z}$ 中的所有分量是相互独立的. 但是由于公共因子 U 的作用, $\mathbf{x}$ 中的各分量仍然是相关的. 基于上述结果, 我们可以利用 AIC 和 BIC 比较 I 型多元 Laplace 分布与 II 型多元 Laplace 分布的性能. 从表 9 中可以看出, 在 AIC 和 BIC 下, 犯罪数据使用 II 型三元 Laplace 分布拟合的结果优于使用 I 型三元 Laplace 分布. 这个例子说明了我们所提出的 II 型多元 Laplace 分布与经典多元 Laplace 分布是不同的. II 型三元 Laplace 分布变量间的关联只与 $\boldsymbol{\Sigma}$ 相关, 而对于 I 型三元 Laplace 分布的变量, 相关性由 $\boldsymbol{\Sigma}$ 和 U 共同决定. 在本例中, 由于 $\boldsymbol{\Sigma}$ 是一个对角矩阵, 所以相

表 9 犯罪数据的模型比较

	I 型三元 Laplace 分布	II 型三元 Laplace 分布
对数似然	−663.991	−652.566
AIC	1339.982	1323.133
BIC	1351.454	1340.341

关性仅与 U 有关, 并且从这个例子可以看出 II 型多元 Laplace 分布的性能超过了经典多元 Laplace 分布.

7 结论与展望

本文提出了一种不同于经典的 Laplace 分布的 II 型多元 Laplace 分布. 虽然 II 型多元 Laplace 分布的密度函数没有显式表达式, 但是我们利用随机表示推导出了许多有用的分布性质, 并构建了基于似然估计和 Bayes 方法的统计推断. 此外, 可进一步探讨该分布其他的一些性质. 例如, 根据 (2.1) 和 Laplace 分布的性质, II 型多元 Laplace 分布不具有线性变换封闭性. 由于 $\mathbf{x}$ 中分量的相关性仅取决于 $\mathbf{z}$ 的相关结构, 因此, 当 $\boldsymbol{\Sigma}$ 是分块矩阵时, 相应的分量之间也相互独立, 均服从 II 型多元 Laplace 分布. 由 (2.3) 可知分布的均值向量为 $\boldsymbol{\mu}$, 若进行均值检验, 则所提出的 ECM 算法适用于计算原假设和备择假设下的参数估计, 因此似然比检验可作为均值检验的考虑方法之一.

鉴于 Laplace 分布在稀疏数据分析中的作用, 下一步工作考虑引入协变量的 II 型多元 Laplace 回归模型. 此外, 可以将 (2.1) 中的随机表示扩展到

$$\mathbf{x} = \boldsymbol{\mu} + \mathbf{U}^*\boldsymbol{\mu}^* + \sqrt{2}\mathbf{U}^{1/2}\mathbf{z}, \tag{7.1}$$

其中 $\mathbf{U}^* = \mathrm{diag}(U_1^*, \ldots, U_d^*)$, 称 $\mathbf{x}$ 服从 II 型多元非对称 Laplace 分布, 记为

$$\mathbf{x} \sim \mathrm{ALD}_d^{(\mathrm{II})}(\boldsymbol{\mu}, \boldsymbol{\Sigma}).$$

当 $\mathbf{U}^* = \mathbf{U}$ 时, $\mathbf{x}$ 服从 I 型多元非对称 Laplace 分布, 记为

$$\mathbf{x} \sim \mathrm{ALD}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}).$$

参考文献

1 Andrews D F, Mallows C L. Scale mixtures of normal distributions. J R Stat Soc Ser B Stat Methodol, 1974, 36: 99–102

2 Stefanski L A. A normal scale mixture representation of the logistic distribution. Statist Probab Lett, 1991, 11: 69–70

3 Box G E P, Tiao G C. Bayesian Inference in Statistical Analysis. Reading: Addison-Wesley, 1973

4 West M. On scale mixtures of normal distributions. Biometrika, 1987, 74: 646–648

5 Huber P J. Robust Statistics, 2nd ed. New York: John Wiley & Sons, 2009

6 Martin R, Breithaupt C. Speech enhancement in the DFT domain using Laplacian speech priors. In: Proceedings of International Workshop on Acoustic Echo and Noise Control. Kyoto, 2003, 87–90

7 Portilla J, Strela V, Wainwright M J, et al. Image denoising using scale mixtures of Gaussians in the wavelet domain. IEEE Trans Image Process, 2003, 12: 1338–1351

8 Solbo S, Eltoft T. Homomorphic wavelet-based statistical despeckling of SAR images. IEEE Trans Geosci Remote Sens, 2004, 42: 711–721

9 Park H J, Lee T W. Modeling nonlinear dependencies in natural images using mixture of Laplacian distribution. In: Advances in Neural Information Processing Systems. Cambridge: MIT Press, 2005, 1041–1048
10 McGraw D K, Wagner J F. Elliptically symmetric distributions. IEEE Trans Inform Theory, 1968, 14: 110–120
11 Pillai R N. Semi-α-Laplace distributions. Comm Statist Theory Methods, 1985, 14: 991–1000
12 Anderson D N. A multivariate Linnik distribution. Statist Probab Lett, 1992, 14: 333–336
13 Kotz S, Johnson N L, Balakrishnan N. Continuous Univariate Distributions, vol. 2, 2nd ed. New York: John Wiley & Sons, 2000
14 Marshall A W, Olkin I. Maximum likelihood characterizations of distributions. Statist Sinica, 1993, 3: 157–171
15 Fernández C, Osiewalski J, Steel M F J. Modeling and inference with v-spherical distributions. J Amer Statist Assoc, 1995, 90: 1331–1340
16 Kozubowski T J, Podgórski K. A multivariate and asymmetric generalization of Laplace distribution. Comput Statist, 2000, 15: 531–540
17 Kotz S, Kozubowski T J, Podgórski K. An asymmetric multivariate Laplace distribution. In: The Laplace Distribution and Generalizations. Boston: Birkhäuser, 2001, 239–272
18 Eltoft T, Kim T, Lee T W. On the multivariate Laplace distribution. IEEE Signal Process Lett, 2006, 13: 300–303
19 Meng X L, Rubin D B. Maximum likelihood estimation via the ECM algorithm: A general framework. Biometrika, 1993, 80: 267–278
20 Murphy K P. Conjugate Bayesian analysis of the Gaussian distribution. Technical Report. Vancouver: University of British Columbia, 2007
21 Tanner M A, Wong W H. The calculation of posterior distributions by data augmentation. J Amer Statist Assoc, 1987, 82: 528–540
22 Von Neumann J. Various techniques used in connection with random digits. Nat Bureau Standards, 1951, 12: 36–38
23 Karlis D. An EM algorithm for multivariate Poisson distribution and related models. J Appl Stat, 2003, 30: 63–77
24 Barndorff-Nielsen O E. Normal inverse Gaussian distributions and stochastic volatility modelling. Scand J Statist, 1997, 24: 1–13

附录 A　I 型多元 Laplace 分布

附录 A.1　密度函数

基于 I 型多元 Laplace 分布随机向量的随机表示 (1.2), 可以得到一个等价的混合表示: $U \sim \mathrm{Exp}(1)$ 和 $\mathbf{x} \mid U \sim N_d(\boldsymbol{\mu}, 2U\boldsymbol{\Sigma})$, 其中 $\mathbf{x} \mid (U=u)$ 的条件密度函数为

$$f_{\mathbf{x}\mid U}(\boldsymbol{x} \mid u) = (4\pi u)^{-d/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\left[-\frac{(\boldsymbol{x}-\boldsymbol{\mu})^{\top}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}{4u}\right].$$

然后可以推导出 $\mathbf{x}$ 的边际分布为

$$\begin{aligned} f_{\mathbf{x}}(\boldsymbol{x}) &= \int_0^{\infty}(4\pi u)^{-\frac{d}{2}}|\boldsymbol{\Sigma}|^{-\frac{1}{2}}\exp\left[-\frac{(\boldsymbol{x}-\boldsymbol{\mu})^{\top}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}{4u}\right]\mathrm{e}^{-u}du \\ &= (4\pi)^{-\frac{d}{2}}|\boldsymbol{\Sigma}|^{-\frac{1}{2}}\frac{\sqrt{2\pi}}{\delta}\mathrm{e}^{-\delta\gamma}\cdot\int_0^{\infty}u^{-d/2+3/2}\frac{\delta}{\sqrt{2\pi}}\mathrm{e}^{\delta\gamma}u^{-3/2}\exp\left[-\frac{1}{2}(\gamma^2u+\delta^2u^{-1})\right]du \\ &= (4\pi)^{-d/2}|\boldsymbol{\Sigma}|^{-1/2}\frac{\sqrt{2\pi}}{\delta}\mathrm{e}^{-\delta\gamma}\mu_{\mathrm{IG}}\left(\frac{3}{2}-\frac{d}{2}\right), \end{aligned}$$

其中 $\gamma^2=2$, $\delta^2=(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})/2 \mathrel{\hat{=}} q(\boldsymbol{x})/2$, 且 $\mu_{\mathrm{IG}}(3/2-d/2)$ 是逆 Gauss 分布[24] 的 $3/2-d/2$ 阶矩

$$\mu_{\mathrm{IG}}(k) = \left(\frac{\delta}{\gamma}\right)^k\frac{K_{-1/2+k}(\delta\gamma)}{K_{-1/2}(\delta\gamma)},$$

其中 $K_m(x)$ 是取值于 x 的第二类修正 Bessel 函数. 最后可以得到

$$f_{\mathbf{x}}(\boldsymbol{x}) = \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}} \frac{K_{d/2-1}(\sqrt{q(\boldsymbol{x})})}{(\sqrt{q(\boldsymbol{x})})^{d/2-1}}. \tag{A.1}$$

附录 A.2 参数的极大似然估计

假设 $\mathbf{x}_j \overset{\text{i.i.d.}}{\sim} \text{Laplace}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $j = 1, \ldots, n$, 其中 $\boldsymbol{x}_j$ 表示 $\mathbf{x}_j$ 的实现值. 对于观测值 $Y_{\text{obs}} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$, 引入了潜在数据 $Y_{\text{mis}} = \{u_1, \ldots, u_n\}$, 则完全数据的似然函数为

$$\begin{aligned} L(\boldsymbol{\mu}, \boldsymbol{\Sigma} \mid Y_{\text{com}}) &= \prod_{j=1}^{n} \left\{ (2\pi)^{-\frac{d}{2}} |2u_j \boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left[-\frac{(\boldsymbol{x}_j - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_j - \boldsymbol{\mu})}{4u_j} \right] \cdot \exp(-u_j) \right\} \\ &\propto |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left[-\sum_{j=1}^{n} \frac{(\boldsymbol{x}_j - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_j - \boldsymbol{\mu})}{4u_j} \right]. \end{aligned}$$

由此, 可以得到 $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 在完全数据下的极大似然估计:

$$\hat{\boldsymbol{\mu}} = \frac{\sum_{j=1}^{n} \boldsymbol{x}_j / u_j}{\sum_{j=1}^{n} 1/u_j}, \quad \hat{\boldsymbol{\Sigma}} = \frac{1}{2n} \sum_{j=1}^{n} \frac{(\boldsymbol{x}_j - \hat{\boldsymbol{\mu}})(\boldsymbol{x}_j - \hat{\boldsymbol{\mu}})^\top}{u_j}.$$

E 步是用 $\{u_j\}_{j=1}^n$ 对应的条件期望替换 $\{u_j\}_{j=1}^n$:

$$E(U_j^{-1} \mid \boldsymbol{x}_j, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{\mu_{\text{IG}}(1/2 - d/2)}{\mu_{\text{IG}}(3/2 - d/2)} = \left(\frac{\sqrt{q(\boldsymbol{x}_j)}}{2} \right)^{-1} \frac{K_{d/2}(\sqrt{q(\boldsymbol{x}_j)})}{K_{d/2-1}(\sqrt{q(\boldsymbol{x}_j)})}, \quad j = 1, \ldots, n,$$

其中 $K_m(x)$ 是取值于 x 的第二类修正 Bessel 函数.

我们检验 $\boldsymbol{\Sigma}$ 中的非对角元素是否全为 0, 也就是检验 $\sigma_{ij} = 0$ 是否对所有满足 $i, j = 1, \ldots, d$ 且 $i \neq j$ 的情形都成立. 在原假设下, 变量之间是相关的, 有约束参数 $\boldsymbol{\mu}$ 和 $\{\sigma_{ii}\}_{i=1}^d$ 的极大似然估计值可以通过期望最大化 (expectation maximization, EM) 算法得到. 最大化 (maximization, M) 步需要计算

$$\hat{\boldsymbol{\mu}}_R = \frac{\sum_{j=1}^{n} \boldsymbol{x}_j / u_j}{\sum_{j=1}^{n} 1/u_j}$$

和

$$\hat{\sigma}_{ii,R} = \frac{1}{2n} \sum_{j=1}^{n} \frac{(x_{ij} - \mu_i)^2}{u_j}.$$

期望 (expectation, E) 步用 $\{u_j\}_{j=1}^n$ 对应的条件期望替换 $\{u_j\}_{j=1}^n$:

$$E_R(U_j^{-1} \mid \boldsymbol{x}_j, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \left(\frac{\sqrt{q_R(\boldsymbol{x}_j)}}{2} \right)^{-1} \frac{K_{d/2}(\sqrt{q_R(\boldsymbol{x}_j)})}{K_{d/2-1}(\sqrt{q_R(\boldsymbol{x}_j)})},$$

其中

$$q_R(\boldsymbol{x}_j) = \sum_{i=1}^{d} \frac{(x_{ij} - \mu_i)^2}{2\sigma_{ii}}.$$

A new multivariate Laplace distribution based on the mixture of normal distributions

Chi Zhang, Man-Lai Tang, Tao Li, Yuan Sun & Guo-Liang Tian

Abstract In this paper, we propose a new multivariate Laplace distribution from normal variance mixture models, called as Type II multivariate Laplace distribution. Unlike the original multivariate Laplace distribution whose all components must have the same value for the mixing variate, the random components in the new distribution could have different value for its own mixing variate and are correlated only through the dependence structure of the normal random vector. Thus, it contains the multiplication of independently identical distribution univariate Laplace distributions as a special case if the normal covariance matrix is diagonal. A tractable stochastic representation is used to derive the probability density function and other statistical properties. The maximum likelihood estimates of parameters via an ECM (expectation/conditional maximization) algorithm and the Bayesian methods are derived. Some simulation studies are conducted to evaluate the performance of the proposed methods. Applications in two real data sets indicate that the Type II multivariate Laplace distribution could have a better performance and is distinct from the original one.

Keywords **ECM algorithm, multivariate Laplace distribution, stochastic representation, univariate Laplace distribution**

MSC(2010) **62F10, 62F15, 62H12**

doi: 10.1360/N012019-00141