

一类指数混合分布族的结果

卫 飏

(南京政治学院 基础部 江苏 南京 210003)

摘要:在统计分析中,多元 t 分布是一种重要的分布以及指数分布族也是一种重要的分布族。模仿从多元正态分布得到多元 t 分布的过程,把指数分布族 p_θ 与gamma分布进行混合,得到一种新的分布族,称为指数混合分布族 $\tilde{p}_\theta$,并讨论了参数函数在指数混合分布族 $\tilde{p}_\theta$ 下的无偏性以及讨论指数分布族 p_θ 和指数混合分布族 $\tilde{p}_\theta$ 有相同的完全统计量和充分统计量。

关键词:多元 t 分布;指数混合分布族;充分统计量

中图分类号: O211.4

文献标识码: A

文章编号: 1671-5322(2013)02-0031-04

由于一般情况下,多元正态分布都属于指数分布族。因此,本文将从多元正态分布得到多元 t 分布的过程推广到一般指数分布族中,定义了一类新的分布族,称为指数混合分布族。并研究了该类分布族的数学及统计性质。下面首先给出多元 t 分布和指数族分布的定义。

定义 1^[1] 设 $t_m(X\beta, V, q)$ 表示密度函数为

$$t_m(y; X\beta, V, q) =$$

$$c_m(q, V) [(q-2) + (y - X\beta)' V^{-1} (y - X\beta)]^{-\frac{q+m}{2}},$$

$\beta \in R^p, V \in v$ 的多元 t 分布,其中 $c_m(q, V) = [\pi^{\frac{m}{2}} \Gamma(\frac{q}{2}) |V|^{\frac{1}{2}}]^{-1} (q-2)^{\frac{m}{2}} \Gamma(\frac{q+m}{2})$,则显然多元 t 分布 $t_m(X\beta, V, q)$ 是正态分布 $N_m(X\beta, V, q)$ 与参数为 $(\frac{q}{2}, \frac{1}{2})$ 的gamma分布的混合,即

$$t_m(y; X\beta, V, q) =$$

$$\int_0^{+\infty} f_m(y; X\beta, \frac{q-2}{z} V) g(z; q) dz =$$

$$\int_0^{+\infty} (2\pi)^{\frac{m}{2}} \frac{q-2}{z} |V|^{-\frac{1}{2}}$$

$$\exp\left\{-\frac{z}{2(q-2)}(y - X\beta)' V^{-1} (y - X\beta)\right\} \cdot$$

$$g(z; q) dz$$

其中 $f_m(y; X\beta, V) = (2\pi)^{-\frac{m}{2}} |V|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(y -$

$X\beta)' V^{-1} (y - X\beta)]$ 表示正态分布的密度函数,

$g(z, q) = [2^{\frac{q}{2}} \Gamma(\frac{q}{2})]^{-1} z^{\frac{q}{2}-1} e^{-\frac{z}{2}}, (z > 0)$ 表示参数

为 $(\frac{q}{2}, \frac{1}{2})$ 的gamma分布密度函数。

设随机变量 X 的样本空间为 (χ, β_x) , μ 为 (χ, β_x) 上一个 σ -有限测度。 $\{P_\theta, \theta \in \Theta\}$ 是一个概率族, θ 为参数, Θ 为参数空间,它关于 $\mu(x)$ 的密度函数族为 $\{f(x, \theta): \theta \in \Theta\}$,其中 $f(x, \theta) = \frac{dP_\theta}{d\mu(x)}$ 。

定义 2^[2] 在上述记号下,若存在自然数 k ,定义于 χ 上的有限可测函数 $T_1(x), \dots, T_k(x)$,以及定义于 Θ 上的有限函数 $c(\theta)$,致

$$p_\theta = \{f(x; \theta) d\mu(x) =$$

$$c(\theta) \exp\left(\sum_{i=1}^k \theta_i T_i(x)\right) d\mu(x): \theta \in \Theta\}$$

对任何 $\theta \in \Theta$,则称分布族 $\{p_\theta: \theta \in \Theta\}$ 为一个指数分布族,其中 $\theta = (\theta_1, \theta_2, \dots, \theta_k)$ 是未知参数, $\theta \in \Theta$ 。

1 指数混合分布族的概念

本章假定上述指数族的自然参数空间 Θ^* 满足条件:对任意的 $\theta \in \Theta^*$ 及任意 $c > 0, c\theta \in \Theta^*$ 。

下面把指数分布族 p_θ 与gamma分布的混合,可以得到一种新的分布族——指数混合分布族。

收稿日期:2013-04-08

作者简介:卫飏(1977-),男,江苏无锡人,讲师,硕士,主要研究方向为概率统计。

定义 3 若存在自然数 k , 定义于 χ 上的有限可测函数 $T_1(x), \dots, T_k(x)$, 以及定义于 Θ 上的有限函数 $c(\theta)$, 对一切 $\theta \in \Theta$, 由积分

$$\int_0^{+\infty} f(x; \frac{z}{q-2}\theta) g(z; q) dz = \int_0^{+\infty} c(\theta) \exp\{\frac{z}{q-2} \cdot \sum_{i=1}^k \theta_i T_i(x)\} g(z; q) dz = h(x; \theta, q)$$

可以得到一个新的分布族, 记作 $\tilde{p}_\theta$, 即

$$\tilde{p}_\theta = \{h(x; \theta, q) d\mu(x) : \theta \in \Theta\}$$

$\tilde{p}_\theta$ 是指数分布族 p_θ 与 gamma 分布的混合, 则称之为指数混合分布族, 简称指数混合族。其中记

$g(z; q) = [2^{\frac{q}{2}} \Gamma(\frac{q}{2})]^{-1} z^{\frac{q}{2}-1} e^{-\frac{z}{2}}$ 是参数为 $(\frac{q}{2}, \frac{1}{2})$ 的 gamma 分布密度函数, 其中参数 $q > 0$ 已知或未知, 且当 q 未知时, 设 $q \in Q, Q \subset R$ 。

2 指数混合族 $\tilde{p}_\theta$ 的性质

定理 1 设 θ 满足: 对任意的 $c > 0$ 及 $\theta \in \Theta$, $c\theta \in \Theta$, 并设参数函数 $S(\theta)$ 满足条件: 对任意 $c > 0, S(c\theta) = \frac{1}{c}S(\theta)$ 或 $S(c\theta) = S(\theta)$, 则 $\alpha(X)$ 在指数混合族 p_θ 下是 $S(\theta)$ 的无偏估计的充分必要条件是 $\alpha(X)$ 在指数族 p_θ 下是 $S(\theta)$ 的无偏估计。

证明 充分性 设 $\alpha(X)$ 在指数族 p_θ 下是 $S(\theta)$ 的无偏估计, 即

$$E_{p_\theta}(\alpha(X)) = \int_\chi \alpha(x) f(x; \theta) d\mu(x) = S(\theta)$$

对一切 $\theta \in \Theta$, 则对任意 $\theta \in \Theta$, 任意 $z > 0$ 及 $q \in Q$, 有

$$\int_\chi \alpha(x) f(x; \frac{z}{q-2}\theta) d\mu(x) = \int_\chi \alpha(x) f(x; \frac{z}{q-2}\theta) d\mu(x) = S(\frac{z}{q-2}\theta) = \frac{q-2}{z} S(\theta) \text{ (或 } S(\theta))$$

从而有

$$\begin{aligned} E_{\tilde{p}_\theta}(\alpha(X)) &= \int_\chi \alpha(X) h(x; \theta, q) d\mu(x) = \\ &= \int_\chi \left[\int_0^{+\infty} \alpha(x) f(x; \frac{z}{q-2}\theta) g(z; q) dz \right] d\mu(x) = \\ &= \int_0^{+\infty} S(\frac{z}{q-2}\theta) g(z; q) dz, \text{ 对一切 } \theta \in \Theta, \text{ 且 } q \in Q \end{aligned} \quad (1)$$

在 $S(\frac{z}{q-2}\theta) = \frac{q-2}{z} S(\theta)$ 或 $S(\frac{z}{q-2}\theta) = S(\theta)$ 两种情况下, 显然都有

$$\int_0^{+\infty} S(\frac{z}{q-2}\theta) g(z; q) dz = S(\theta),$$

对一切 $\theta \in \Theta$, 且 $q \in Q$

从而(1)式变为

$$E_{\tilde{p}_\theta}(\alpha(X)) = S(\theta), \text{ 对一切 } \theta \in \Theta, \text{ 且 } q \in Q$$

这说明, 在指数混合族 $\tilde{p}_\theta$ 下, $\alpha(X)$ 是 $S(\theta)$ 的无偏估计。

必要性 设在指数混合族 $\tilde{p}_\theta$ 下, $\alpha(X)$ 是 $S(\theta)$ 的无偏估计, 即对一切 $\theta \in \Theta, q \in Q$, 有

$$E_{\tilde{p}_\theta}(\alpha(X)) = \int_\chi \alpha(X) h(x; \theta, q) d\mu(x) = S(\theta),$$

则对任意给定的 $\theta_0 \in \Theta$ 且 $q_0 \in Q$ 和任意的 $c > 0$, 有

$$\begin{aligned} S(c\theta_0) &= \int_0^{+\infty} \left[\int_\chi \alpha(x) f(x; \frac{z}{q_0-2}c\theta_0) d\mu(x) \right] g(z; q_0) dz = \\ &= \int_0^{+\infty} \left[\int_\chi \alpha(x) f(x; \frac{\omega}{q_0-2}\theta_0) d\mu(x) \right] c^{-1} g(\frac{\omega}{c}; q_0) d\omega \\ &\text{从而得到} \\ &\int_0^{+\infty} [\tilde{\alpha}(\omega; \theta_0) - S(\frac{\omega}{q_0-2}\theta_0)] c^{-1} g(\frac{\omega}{c}; q_0) d\omega = 0, \\ &\text{对一切 } c > 0 \end{aligned} \quad (2)$$

其中 $\tilde{\alpha}(\omega; \theta_0) = \int_\chi \alpha(x) f(x; \frac{\omega}{q_0-2}\theta_0) d\mu(x)$ 。

对给定的 $q_0 \in Q$, 利用指数族 $\{c^{-1}g(\frac{\omega}{c}; q_0) : c > 0\}$ 的完全性, 从(2)式就可以得到对一切 $\omega > 0$, 有 $\tilde{\alpha}(\omega; \theta_0) = S(\frac{\omega}{q_0-2}\theta_0)$, 又由指数族的性质, $\tilde{\alpha}(\omega; \theta_0)$ 是 ω 的连续函数。因此, 对一切 $\omega > 0$, 有

$$\begin{aligned} S(\frac{\omega}{q_0-2}\theta_0) &= \tilde{\alpha}(\omega; \theta_0) = \\ &= \int_\chi \alpha(x) f(x; \frac{\omega}{q_0-2}\theta_0) d\mu(x) \end{aligned}$$

由上式, 及 θ_0 和 $\omega > 0$ 的任意性即知, $\alpha(X)$ 在指数族 p_θ 下是 $S(\theta)$ 的无偏估计。

定理 2 设参数空间 Θ 满足条件: 对任意的 $c > 0$, 及 $\theta \in \Theta$, 有 $c\theta \in \Theta$, 则指数分布族 p_θ 与指数混合分布族 $\tilde{p}_\theta$ 有相同的完全统计量。

证明 充分性 设 $T(X)$ 在指数混合族 $\tilde{p}_\theta$ 下是完全的, 下证它在指数分布族 p_θ 下也是完全的。

设 $\alpha(t)$ 是任意满足下述条件的可测函数,

$$E_{p_\theta}(\alpha(T(X))) = \int_\chi \alpha(T(X)) f(x; \theta) d\mu(x) = 0,$$

对一切 $\theta \in \Theta$, (3)

则有

$$E_{\tilde{p}_\theta}(\alpha(T(X))) = \int_X \alpha(T(x)) h(x; \theta) d\mu(x) = \int_0^{+\infty} \left[\int_X \alpha(T(x)) f(x; \frac{z}{q-2}\theta) d\mu(x) \right] g(z; q) dz = 0$$

对一切 $\theta \in \Theta$ 及 $q \in Q$ 成立。

由上式及 $T(X)$ 是在指数混合族 $\tilde{p}_\theta$ 下的完全统计量知, 必有 $\alpha(T(x)) = 0$, 由此及 (3) 式知, $T(X)$ 在指数分布族 p_θ 下是完全的。

必要性 设 $T(X)$ 是在指数分布族 p_θ 下的完全统计量, 下证它在指数混合族 $\tilde{p}_\theta$ 下也是完全的。

设 $\alpha(t)$ 是任一满足下式的可测函数,

$$E_{\tilde{p}_\theta}(\alpha(T(X))) \stackrel{\Delta}{=} \int_X \alpha(T(x)) h(x; \theta, q) d\mu(x) = 0, \text{ 对一切 } \theta \in \Theta \text{ 及 } q \in Q.$$

现取 Θ 定的一个内点 θ_0 , 及 $q_0 \in Q$, 则当 c 充分接近于 1 时, 不妨设 $1-d < c < 1+d$, 有 $c\theta_0 \in \Theta$, 从而当 $c \in (1-d, 1+d)$ 时,

$$\int_0^{+\infty} \left[\int_X \alpha(T(x)) f(x; \frac{z}{q-2}c\theta_0) d\mu(x) \right] g(z; q_0) dz = \int_0^{+\infty} \left[\int_X \alpha(T(x)) f(x; \frac{\omega}{q_0-2}\theta_0) d\mu(x) \right] c^{-1} g(\frac{\omega}{c}; q_0) d\omega = 0$$

再由指数族 $\{c^{-1}g(\frac{\omega}{c}; q_0) : c \in (1-d, 1+d)\}$ 的完全性知, 必有

$$\int_X \alpha(T(x)) f(x; \frac{\omega}{q_0-2}\theta_0) d\mu(x) = 0,$$

对一切 $\omega > 0$,

由 $\omega > 0$ 及 q_0 和 θ_0 的任意性, 就得到

$$\int_X \alpha(T(x)) f(x; \theta) d\mu(x) = 0,$$

对一切 $\theta \in \Theta$,

由上式及 $T(X)$ 在指数分布族 p_θ 下的完全性知, 必有 $\alpha(T(x)) = 0$, 因此 $T(X)$ 在指数混合族 $\tilde{p}_\theta$ 下也是完全的。

引理 若 $T(X)$ 在指数分布族 p_θ 下是充分统计量, 则它在指数混合族 $\tilde{p}_\theta$ 下仍然是充分的。

证明 若 $T(X)$ 在指数族 p_θ 下是充分统计量, 则由充分性因子分解定理, 存在非负可测函数 $h(x)$ 及函数 $L(t; \theta)$, 使对任何给定的 $\theta \in \Theta$, $L(t; \theta)$ 是 t 的非负可测函数, 且使得 X 的密度函数为

$$f(x; \theta) = c(\theta) \exp \left\{ \sum_{i=1}^k \theta_i T_i(x) \right\} = L(T(x); \theta) h(x)$$

从而得到 X 在指数混合族 $\tilde{p}_\theta$ 下的密度函数为

$$h(x; \theta, q) = \int_0^{+\infty} f(x; \frac{z}{q-2}\theta) g(z; q) dz = h(x) \int_0^{+\infty} L(T(x); \frac{z}{q-2}\theta) g(z; q) dz \stackrel{\Delta}{=} h(x) L(T(x); \theta, q)$$

对一切 $x \in X$, $\theta \in \Theta$ 及 $q \in Q$, 其中 $\tilde{L}(t; \theta, q) = \int_0^{+\infty} L(t; \frac{z}{q-2}\theta) g(z; q) dz$, 显然 $\tilde{L}(t; \theta, q)$ 仍然是 t 的非负可测函数, 因此, 由因子分解定理知,

$T(X)$ 是指数混合族 $\tilde{p}_\theta$ 下的充分统计量。

由定理 2 和引理 1 立即得到下述重要结果。

定理 3 设参数空间 Θ 有内点集, 则 $(T_1(X), T_2(X), \dots, T_k(X))$ 是指数混合族 $\tilde{p}_\theta$ 下的完全充分统计量。

证明略。

参考文献:

- [1] 杨国庆. 一类线性模型中的参数估计问题[D]. 北京: 中国科学院, 2002.
- [2] 陈希孺. 数理统计引论[M]. 北京: 科学出版社, 1981.

Fingdings of Mixed Exponential Distribution

WEI Biao

(Department of Basic Science, Nanjing Political College, Nanjing Jiangsu 210003, China)

Abstract: In the statistical analysis, multivariate t distribution is an important distribution and exponential distribution family is a family of distributions is important. This paper imitated the multivariate normal distribution to obtain a multivariate distribution and mixed exponential distribution and distribution, a new family of distributions obtained, known as the mixed exponential distribution family. The function of parameters in the mixed exponential distribution family under the unbiased was discussed. This paper pointed out that the exponential distribution family and exponential mixture distribution family had the same complete statistic and the sufficient statistics.

Keywords: multivariate t - distribution; mixed exponential distribution; sufficient statistics

(责任编辑:张英健)

(上接第 12 页)

Product Design Time Forecast by Using Gaussian Margin Kernel Regression

SHANG Zhi-gen

(School of Electrical Engineering, Yancheng Institute of Technology, Yancheng Jiangsu 224051, China)

Abstract: There exist problems of small samples and heteroscedastic noise in design time forecast. To solve them, Gaussian margin kernel regression (GMKR) is proposed. First, the Gaussian distribution over weight vectors for the kernel - based regression is assumed for GMKR, and the optimization objective function of GMKR is designed by considering both the relative entropy and the sum of the natural log of the output probability densities. Then, the optimization problem of GMKR is simplified by assuming the covariance matrix of the Gaussian distribution to be a diagonal matrix, and its relevant optimization problem is solved based on particle swarm optimization algorithm. Finally, the effectiveness of GMKR is verified by our experiment results on the time forecast of plastic injection mold design.

Keywords: Design time; Forecast; Kernel function; Relative entropy; Heteroscedasticity

(责任编辑:张振华)