

二元帕累托分布的识别性及参数估计

李国安, 李卫华

(宁波大学 理学院, 浙江 宁波 315211)

摘要: 讨论了二元帕累托分布的识别性及参数估计, 当只有最小值的分布已知时, 那么只有一个参数可识别, 当可识最小值的分布已知时, 那么所有参数皆可识别, 并由此得到了所有参数的最大似然估计.

关键词: 二元帕累托分布; 识别性; 最大似然估计

中图分类号: O212.4

文献标志码: A

文章编号: 1001-5132 (2015) 02-0048-04

1994 年 Veenus 等^[1]给出了我们要讨论的二元帕累托分布的形式及其一个特征, 这个二元帕累托分布是一个混合型分布, 这种分布形式的构造源于文献[2]. 1967 年 Marshall 和 Olkin^[2]首先利用这种构造提出了一类多元指数分布, 所以现有一些文献把这种构造称为 Marshall-Olkin 型. 文献[3]给出了多元指数分布的参数估计, 文献[4]把这种构造称为 mixed. 文献[5]研究了一类混合分布的识别性. 在二元及多元分布理论中涉及混合二字的还有指 2 个及以上分布的混合(mixture), 很容易产生混淆, 这是 2 种完全不同的分布形式. 自文献[6]讨论了二元帕累托分布的参数估计以来, 文献[7-15]讨论了二元帕累托分布的特征及参数估计以及其他特性和应用; 文献[16]研究了多元帕累托分布, 说明了二元及多元帕累托分布在保险精算中的广泛应用; 在已知二元及多元正态分布参数估计的前提下, 文献[17]对二元正态分布的参数识别进行了研究. 可想而知, 研究二元帕累托分布的识别性及参数估计也是必要的, 其必要性在于二元帕累托分布的识别性研究是进行多元帕累托分布的识别性研究及其对应参数估计的前提. 已知 (X_1, X_2) 服从二元帕累托分布, 记作 $(X_1, X_2) \sim P(\lambda_1, \lambda_2, \lambda_0, \beta)$, 记 $F(x, y)$ 为 (X_1, X_2) 的联合分布函数, 设 $Z = \min(X_1, X_2)$, Z 的密度函数和分布函数分别记为 $g(z), G(z)$, 记 I 为一个三值随机变量, 定义 $I = 1, 2, 3$ 分别对应于 $X_1 < X_2, X_1 > X_2, X_1 = X_2$ 时,

(Z, I) 的分布密度记为 $p_i f_i(\bullet), i = 1, 2, 3$. 所谓识别性问题, 是指已知 Z 的分布或 (Z, I) 的分布, 是否唯一确定 (X_1, X_2) 的分布函数中 $F(x, y)$ 的参数问题, 即若已知 $(X_1, X_2) \sim P(\lambda_1, \lambda_2, \lambda_0, \beta)$, $(X'_1, X'_2) \sim P(\lambda'_1, \lambda'_2, \lambda'_0, \beta')$, 这里 $\lambda_i, \lambda'_i, i = 0, 1, 2, \beta, \beta'$ 同为参数, 设 $Z' = \min(X'_1, X'_2)$, 定义随机变量 $I' = 1, 2, 3$ 分别对应 $X'_1 < X'_2, X'_1 > X'_2, X'_1 = X'_2$ 时, Z' 的密度函数和分布函数分别记为 $g'(z), G'(z)$, (Z', I') 的分布密度记为 $p'_i f'_i(\bullet), i = 1, 2, 3$. 若已知 $G(z) = G'(z)$, 则必有 $\lambda_i = \lambda'_i, i = 0, 1, 2, \beta = \beta'$, 4 个等式中至少有一个成立, 称在此情形下等式成立的对应参数是可识别的, 等式不成立的对应参数是不可识别的; 否则, 称在此情形下参数是不可识别的. 类似若已知 $p_i f_i(\bullet) = p'_i f'_i(\bullet), i = 1, 2, 3$, 则必有 $\lambda_i = \lambda'_i, i = 0, 1, 2, \beta = \beta'$, 4 个等式中至少有一个成立, 称在此情形下等式成立的对应参数是可识别的, 等式不成立的对应参数是不可识别的; 否则, 称在此情形下参数是不可识别的.

本文讨论二元帕累托分布的参数识别性问题, 分别讨论了 Z 分布已知、 (Z, I) 的分布已知时二元帕累托分布的参数识别性; 讨论二元帕累托分布的参数估计, 分别讨论了 Z 分布已知、 (Z, I) 的分布已知时二元帕累托分布的参数估计.

1 二元帕累托分布的识别性

文献[1]给出了如下二元帕累托分布:

定义 1 设 (X_1, X_2) 为二维随机变量, 有如下的尾概率:

$$\bar{F}(x_1, x_2) = \left(\frac{x_1}{\beta}\right)^{-\lambda_1} \left(\frac{x_2}{\beta}\right)^{-\lambda_2} [\max(\frac{x_1}{\beta}, \frac{x_2}{\beta})]^{-\lambda_0},$$

$x_1 \geq \beta, x_2 \geq \beta$, 则称 (X_1, X_2) 服从二元帕累托分布, 记作 $(X_1, X_2) \sim P(\lambda_1, \lambda_2, \lambda_0, \beta)$, 这里 $\lambda_i > 0, i=1, 2, \lambda_0 \geq 0, \beta$ 为任意常数. 记 $\lambda = \sum_{i=0}^2 \lambda_i$.

引理 1 若 $(X_1, X_2) \sim P(\lambda_1, \lambda_2, \lambda_0, \beta)$, 则 Z 的分布密度为:

$$g(z) = \lambda \beta^\lambda z^{-\lambda-1}, z \geq \beta.$$

证明 由

$$P(Z > z) = P(\min(X_1, X_2) > z) =$$

$$P(X_1 > z, X_2 > z) = \bar{F}(z, z) =$$

$$\left(\frac{z}{\beta}\right)^{-\lambda_1} \left(\frac{z}{\beta}\right)^{-\lambda_2} [\max(\frac{z}{\beta}, \frac{z}{\beta})]^{-\lambda_0} = \left(\frac{z}{\beta}\right)^{-\lambda},$$

得

$$g(z) = \lambda \beta^\lambda z^{-\lambda-1}, z \geq \beta.$$

引理 2 若 $(X_1, X_2) \sim P(\lambda_1, \lambda_2, \lambda_0, \beta)$, 则 (Z, I) 的分布密度为:

$$p_i f_i(z) = \begin{cases} \lambda_1 \beta^{\lambda_1} z^{-\lambda_1-1}, i=1, z \geq \beta, \\ \lambda_2 \beta^{\lambda_2} z^{-\lambda_2-1}, i=2, z \geq \beta, \\ \lambda_0 \beta^{\lambda_0} z^{-\lambda_0-1}, i=3, z \geq \beta. \end{cases}$$

证明 二元帕累托分布为混合型分布, 现对其3部分 $x_1 < x_2 \geq \beta, x_1 > x_2 \geq \beta, x_1 = x_2 \geq \beta$ 分别计算对应的密度. 由

$$\bar{F}_1(x_1, x_2) = P(X_1 > x_1, X_2 > x_2) =$$

$$\left(\frac{x_1}{\beta}\right)^{-\lambda_1} \left(\frac{x_2}{\beta}\right)^{-\lambda_2-\lambda_0}.$$

$$f_1(x_1, x_2) = \frac{\partial^2 F_1(x_1, x_2)}{\partial x_1 \partial x_2} =$$

$$\lambda_1 (\lambda_2 + \lambda_0) \beta^{\lambda_1} x_1^{-\lambda_1-1} x_2^{-\lambda_2-\lambda_0-1}, x_2 > x_1 \geq \beta,$$

同理

$$f_2(x_1, x_2) = \frac{\partial^2 F_2(x_1, x_2)}{\partial x_1 \partial x_2} =$$

$$\lambda_2 (\lambda_1 + \lambda_0) \beta^{\lambda_2} x_2^{-\lambda_2-1} x_1^{-\lambda_1-\lambda_0-1},$$

$$x_1 < x_2 \geq \beta,$$

$$P(Z > z, I = 1) = P(\min(X_1, X_2) > z, I = 1) =$$

$$P(X_1 > z, X_2 > X_1) =$$

$$\int_z^{+\infty} dx_1 \int_{x_1}^{+\infty} f_1(x_1, x_2) dx_2 =$$

$$\int_z^{+\infty} dx_1 \int_{x_1}^{+\infty} \lambda_1 (\lambda_2 + \lambda_0) \beta^{\lambda_1} x_1^{-\lambda_1-1} x_2^{-\lambda_2-\lambda_0-1} dx_2 =$$

$$\int_z^{+\infty} \lambda_1 \beta^{\lambda_1} x_1^{-\lambda_1-1} dx_1,$$

所以

$$p_1 f_1(z) = \lambda_1 \beta^{\lambda_1} z^{-\lambda_1-1}, i=1, z \geq \beta,$$

同理

$$p_2 f_2(z) = \lambda_2 \beta^{\lambda_2} z^{-\lambda_2-1}, i=2, z \geq \beta,$$

$$P(Z > z) = P(Z > z, I = 1) +$$

$$P(Z > z, I = 2) + P(Z > z, I = 3),$$

两边关于 z 求导得:

$$\lambda \beta^\lambda z^{-\lambda-1} = \lambda_1 \beta^{\lambda_1} z^{-\lambda_1-1} +$$

$$\lambda_2 \beta^{\lambda_2} z^{-\lambda_2-1} + p_3 f_3(z), z \geq \beta,$$

由此可得, 混合部分的密度为:

$$p_3 f_3(z) = \lambda_0 \beta^{\lambda_0} z^{-\lambda_0-1}, i=3, z \geq \beta.$$

定理 1 若 $(X_1, X_2) \sim P(\lambda_1, \lambda_2, \lambda_0, \beta)$, $(X'_1, X'_2) \sim P(\lambda'_1, \lambda'_2, \lambda'_0, \beta')$, 若已知 Z 与 Z' 同分布, 则只有参数 β 可识别, 其余参数皆不可识别.

证明 由 $g(z) = g'(z)$, 得:

$$\lambda \beta^\lambda z^{-\lambda-1} = \lambda' \beta'^{\lambda'} z^{-\lambda'-1}, \forall z \geq \max(\beta, \beta'),$$

首先得 $\lambda = \lambda'$, 由此得到 $\beta = \beta'$; 也可以直接由 $z \geq \beta, z \geq \beta'$ 得 $\beta = \beta'$.

注记 1 3 个参数的和是可识别的.

定理 2 若 $(X_1, X_2) \sim P(\lambda_1, \lambda_2, \lambda_0, \beta)$, $(X'_1, X'_2) \sim P(\lambda'_1, \lambda'_2, \lambda'_0, \beta')$, 若已知 (Z, I) 与 (Z', I') 同分布, 则所有参数皆可识别.

证明 由 $p_i f_i(z) = p'_i f'_i(z)$, 得:

$$\lambda_i \beta^{\lambda_i} z^{-\lambda_i-1} = \lambda'_i \beta'^{\lambda'_i} z^{-\lambda'_i-1}, i=0, 1, 2, \text{得:}$$

$$\lambda = \lambda', \lambda_i = \lambda'_i, i=0, 1, 2, \beta = \beta'.$$

2 二元帕累托分布的参数估计

文献[6]研究了二元帕累托分布的参数估计, 这里我们从定理 1 和定理 2 出发, 直接分别获得了参数 β , 以及所有参数的最大似然估计.

定理 3 设 (X_1, X_2) 是总体, $(X_{11}, X_{21}), \dots, (X_{1n}, X_{2n})$ 是来自总体 (X_1, X_2) 的容量为 n 的样本, 记

$$Z = \min(X_1, X_2), Z_i = \min(X_{1i}, X_{2i}),$$

$$i=1, 2, \dots, n,$$

若 Z 具有以下的分布密度:

$$g(z) = \lambda \beta^\lambda z^{-\lambda-1}, z \geq \beta,$$

$Z_1, \dots, Z_n$ 是来自总体 Z 的容量为 n 的样本, 则参数 β 的最大似然估计为:

$$\hat{\beta} = \min_i(Z_i).$$

证明 略.

定理 4 设 (X_1, X_2) 是总体, $(X_{11}, X_{21}), \dots, (X_{1n}, X_{2n})$ 是来自总体 (X_1, X_2) 的容量为 n 的样本, 记 $Z = \min(X_1, X_2)$, 定义随机变量 $I = 1, 2, 3$ 分别对应 $X_1 < X_2, X_1 > X_2, X_1 = X_2$ 时, 记 $Z_i = \min(X_{1i}, X_{2i})$. 定义随机变量 $I_i = 1, 2, 3$ 分别对应 $X_{1i} < X_{2i}, X_{1i} > X_{2i}, X_{1i} = X_{2i}$ 时, $i = 1, \dots, n$.

若 (Z, I) 具有以下的分布密度:

$$p_i f_i(z) = \begin{cases} \lambda_1 \beta^\lambda z^{-\lambda-1}, & i = 1, z \geq \beta, \\ \lambda_2 \beta^\lambda z^{-\lambda-1}, & i = 2, z \geq \beta, \\ \lambda_0 \beta^\lambda z^{-\lambda-1}, & i = 3, z \geq \beta, \end{cases}$$

$(Z_1, I_1), \dots, (Z_n, I_n)$ 是来自总体 (Z, I) 的容量为 n 的样本, 记

$$w_{1i} = \begin{cases} 1, & I_i = 1, \\ 0, & I_i \neq 1, \end{cases} \quad i = 1, \dots, n,$$

$$w_{2i} = \begin{cases} 1, & I_i = 2, \\ 0, & I_i \neq 2, \end{cases} \quad i = 1, \dots, n,$$

$$w_{3i} = \begin{cases} 1, & I_i = 3, \\ 0, & I_i \neq 3, \end{cases} \quad i = 1, \dots, n,$$

则参数 β 的最大似然估计为 $\hat{\beta} = \min_i(Z_i)$.

参数 λ_j 的最大似然估计为:

$$\hat{\lambda}_j = \frac{\sum_{i=1}^n w_{ji}}{\sum_{i=1}^n \ln z_i - \ln \hat{\beta}}, \quad j = 1, 2, 3.$$

证明 似然函数为:

$$L(\lambda_1, \lambda_2, \lambda_3, \beta) = (\lambda_1 \beta^\lambda (\prod_i z_i)^{-\lambda-1})^{\sum_{i=1}^n w_{1i}} \cdot$$

$$(\lambda_2 \beta^\lambda (\prod_i z_i)^{-\lambda-1})^{\sum_{i=1}^n w_{2i}} \cdot$$

$$(\lambda_0 \beta^\lambda (\prod_i z_i)^{-\lambda-1})^{\sum_{i=1}^n w_{3i}} =$$

$$(\lambda_1)^{\sum_{i=1}^n w_{1i}} (\lambda_2)^{\sum_{i=1}^n w_{2i}} (\lambda_0)^{\sum_{i=1}^n w_{3i}}.$$

$$\beta^\lambda (\prod_i z_i)^{-\lambda-1}, \quad z_i \geq \beta,$$

则参数 β 的最大似然估计为 $\hat{\beta} = \min_i(Z_i)$, 并有似然方程:

$$\begin{cases} \frac{\partial \ln L}{\partial \lambda_1} = \sum_{i=1}^n w_{1i} / \lambda_1 + \ln \hat{\beta} - \sum_{i=1}^n \ln z_i = 0, \\ \frac{\partial \ln L}{\partial \lambda_2} = \sum_{i=1}^n w_{2i} / \lambda_2 + \ln \hat{\beta} - \sum_{i=1}^n \ln z_i = 0, \\ \frac{\partial \ln L}{\partial \lambda_0} = \sum_{i=1}^n w_{3i} / \lambda_0 + \ln \hat{\beta} - \sum_{i=1}^n \ln z_i = 0, \end{cases}$$

得:

$$\begin{cases} \hat{\lambda}_1 = \sum_{i=1}^n w_{1i} / (\sum_{i=1}^n \ln z_i - \ln \hat{\beta}), \\ \hat{\lambda}_2 = \sum_{i=1}^n w_{2i} / (\sum_{i=1}^n \ln z_i - \ln \hat{\beta}), \\ \hat{\lambda}_0 = \sum_{i=1}^n w_{3i} / (\sum_{i=1}^n \ln z_i - \ln \hat{\beta}). \end{cases}$$

注记 2 二元帕累托分布随机变量的计算机生成过程如下:

先直接生成 3 个相互独立的一元帕累托分布随机变量

$$U_1 \sim \bar{F}_{U_1}(u_1) = \left(\frac{u_1}{\beta}\right)^{-\lambda_1},$$

$$U_2 \sim \bar{F}_{U_2}(u_2) = \left(\frac{u_2}{\beta}\right)^{-\lambda_2},$$

$$U_3 \sim \bar{F}_{U_3}(u_3) = \left(\frac{u_3}{\beta}\right)^{-\lambda_0},$$

并置 $X_1 = \min(U_1, U_3)$, $X_2 = \min(U_1, U_3)$, 则

$$(X_1, X_2) \sim P(\lambda_1, \lambda_2, \lambda_0, \beta),$$

这里模拟略去.

注记 3 由二元帕累托分布参数的最大似然估计表达式可自然地写出对应的多元帕累托分布参数的最大似然估计表达式.

注记 4 由

$$EZ = \frac{\lambda \beta}{-\lambda + 1}, \quad EZ^2 = \frac{\beta^2 \lambda}{-\lambda + 2}, \quad EW_1 = \frac{\lambda_1}{\lambda},$$

$$EW_2 = \frac{\lambda_2}{\lambda}, \quad EW_3 = \frac{\lambda_0}{\lambda},$$

可得二元帕累托分布参数的矩估计表达式, 不同于对应参数的最大似然估计表达式.

结论 仅是 Z 的分布已知时, 只有参数 β 可识别, 即只有参数 β 可估计, 当 (Z, I) 的分布已知

时, 所有参数皆可识别, 即所有参数皆可估计.

参考文献:

- [1] Veenus P, Nair K R M. Characterization of a bivariate Pareto distribution[J]. J Ind Statist Assoc, 1994, 32:15-20.
- [2] Marshall A W, Olkin I. A multivariate exponential distribution[J]. Journal of the American Statistical Association, 1967, 62(1):30-44.
- [3] Arnold B C. Parameter estimation for a multivariate exponential distribution[J]. Journal of the American Statistical Association, 1968, 63:848-852.
- [4] Johnson N L, Kotz S. Distribution in statistics: Continuous multivariate distributions[M]. NewYork: Wiley, 1972:266.
- [5] 李国安. 混合分布的识别性及其应用[J]. 宁波大学学报: 理工版, 1993, 6(2):28-38.
- [6] Jeevanand E S, Veenus P. Parameter estimation for a bivariate Pareto distribution[J]. Statistical Papers, 1996, 37:153-164.
- [7] Gupta R C. Reliability studies of bivariate distributions with Pareto conditionals[J]. Journal of Multivariate Analysis, 2001, 76(2):214-225.
- [8] Xekalaki E, Dimaki C. Characterizations of bivariate Pareto and Yule Distributions[J]. Communications in Statistics Theory and Methods, 2004, 33(12):3033-3042.
- [9] Nadarajah S, Kotz S. Performance measures for some bivariate Pareto distributions[J]. International Journal of General Systems, 2006, 35(4):387-393.
- [10] Ali M M, Nadarajah S. A truncated bivariate generalized Pareto distribution[J]. Computer Communications, 2007, 30:1926-1930.
- [11] Gupta A K, Nadarajah S. Information matrices for some bivariate Pareto distribution[J]. Applied Mathematics and Computation, 2007, 184(2):1069-1079.
- [12] Chacko M, Thomas P Y. Estimation of a parameter of bivariate Pareto distribution by ranked set sampling[J]. Journal of Applied Statistics, 2007, 34(4):703-714.
- [13] Michel R. Estimation of the angular density in bivariate generalized Pareto models[J]. Statistics: Theoretical and Applied Statistics, 2009, 43(2):187-202.
- [14] Nadarajah S. A bivariate Pareto model for drought[J]. Stochastic Environmental Research and Risk Assessment, 2009, 23(6):811-822.
- [15] Sankaran P G, Kundu D. A bivariate Pareto model[J]. Statistics: Theoretical and Applied Statistics, 2014, 48(2): 241-255.
- [16] Asimit A V, Furman E, Vernic R. On a multivariate Pareto distribution[J]. Insurance: Mathematics and Economics, 2010, 46:308-316.
- [17] Basu A P, Ghosh J K. Identifiability of the multinormal and other distributions under competing risks model[J]. Journal of Multivariate Analysis, 1978, 8(3):413-429.

Identifiability and Parameter Estimation of a Bivariate Pareto Distribution

LI Guo-an, LI Wei-hua

(Faculty of Science, Ningbo University, Ningbo 315211, China)

Abstract: This paper shows that provided any two random variable observe a bivariate pareto distribution, and the distribution of the minimum is known, then only one of parameters can be identified; while the distribution of the identified minimum is known, then all of parameters can be identified. As a result, the maximum likelihood estimator of all of parameters can hence be derived.

Key words: bivariate Pareto distribution; identifiability; maximum likelihood estimator

(责任编辑 史小丽)