

二元二点分布的识别性及参数估计

李国安, 李晶晶

(宁波大学 理学院, 浙江 宁波 315211)

摘要: 讨论 1 个二元二点分布的识别性及参数估计, 在可识最小值分布已知时, 若其中 1 个参数已知, 则所有其他参数皆可识, 由此得到了基于随机样本的所有其他参数的最大似然估计.

关键词: 二元二点分布; 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 (2016) 02-0081-04

笔者在满意度市场调查研究中发现了二元二点分布, 在前人研究中也曾发现有二元二点分布的情况^[1-3]. 其中, 文献[1]研究的是一个具自然形式的二元二点分布, 而在本文中, 笔者试图引入文献[1]中具自然形式的二元二点分布的特殊情形, 并依据文献[2]构造二元分布的思路引入二元二点分布. 二元二点分布是所有二元分布中最为简单的分布, 从教学角度来看, 它是二元分布中分布函数、特征函数、数字特征乃至参数估计的一种最简单载体; 从研究角度来看, 参数估计是求解模型的核心问题之一, 而二点分布对应的模型是绩效模型, 即业绩是否达到预期之评价模型. 在文献[4-13]中, 都在已知二元及多元正态分布参数估计的情形下, 研究了二元及多元正态分布参数的识别性. 众所周知, 分布参数的识别性说明分布参数的可估计性, 因此, 正常次序是研究分布参数识别性在先, 研究分布参数估计在后. 本文采用从参数的识别性进而求参数的最大似然估计方式, 讨论二元二点分布的识别性及参数估计.

1 二元二点分布的识别性

二元二点分布定义如下.

定义 1 称 (X, Y) 服从二元二点分布, 是指其存在相互独立的 3 个随机变量 U_1, U_2, U_3 , 其中 $U_1 \sim B(1, p_1)$, $U_2 \sim B(1, p_2)$, $U_3 \sim B(1, p_3)$, 使得 $X = \min(U_1, U_3)$, $Y = \max(U_2, U_3)$, $0 < p_1 < 1$, $0 < p_2 < 1$,

$0 < p_3 < 1$, 记作 $(X, Y) \sim BBD(1, p_1, p_2, p_3)$.

记 $Z = \min(X, Y)$, 定义 $I = 1, 2, 3$ 分别对应于 $X < Y, Y < X, X = Y$.

引理 1 若 $(X, Y) \sim BBD(1, p_1, p_2, p_3)$, 则 Z 的分布律为 $Z \sim B(1, p_1 p_2)$.

证明 $P(Z=1) = P(X=1, Y=1) = P(\min(U_1, U_3)=1, \max(U_2, U_3)=1) = P(U_1=1, U_2=1) = p_1 p_2$.

引理 2 若 $(X, Y) \sim BBD(1, p_1, p_2, p_3)$, 则 (Z, I) 的联合分布律为:

$$P(Z=i, I=j) = \begin{cases} (1-p_1 p_3) p_2 + (1-p_1)(1-p_2) p_3, & i=0, j=1, \\ 0, & i=0, j=2, \\ (1-p_2)(1-p_3), & i=0, j=3, \\ 0, & i=1, j=1, \\ 0, & i=1, j=2, \\ p_1 p_3, & i=1, j=3. \end{cases}$$

证明 $P(Z=0, I=1) = P(X=0, Y=1) = P(\min(U_1, U_3)=0, \max(U_2, U_3)=1) = P(U_1 U_3=0, U_2=1) + P(U_3=0, U_2=0, U_3=1) = (1-p_1 p_3) p_2 + (1-p_1)(1-p_2) p_3$. 同理, $P(Z=0, I=3) = P(X=0, Y=0) = P(\min(U_1, U_3)=0, \max(U_2, U_3)=0) = P(U_2=0, U_3=0) = (1-p_2)(1-p_3)$, $P(Z=1, I=3) = P(X=1, Y=1) = P(\min(U_1, U_3)=1, \max(U_2, U_3)=1) = P(U_1=1, U_3=1) = p_1 p_3$.

定理 1 若 $(X, Y) \sim BBD(1, p_1, p_2, p_3)$, $(X', Y') \sim BBD(1, p'_1, p'_2, p'_3)$, 且已知 Z 与 Z' 同分布, 则所有参数皆不可识别.

证明 由 $p_1 p_2 = p'_1 p'_2$ 可知 p_1, p_2, p_3 皆不可识别.

定理 2 若 $(X, Y) \sim BBD(1, p_1, p_2, p_3), (X', Y') \sim BBD(1, p'_1, p'_2, p'_3)$, 且已知 (Z, I) 与 (Z', I') 同分布, 其中 1 个参数已知, 则其他参数皆可识别.

证明 由

$$\begin{cases} (1-p_1 p_3) p_2 + (1-p_1)(1-p_2) p_3 = (1-p'_1 p'_3) \cdot \\ p'_2 + (1-p'_1)(1-p'_2) p'_3, \\ (1-p_2)(1-p_3) = (1-p'_2)(1-p'_3), \\ p_1 p_3 = p'_1 p'_3, \end{cases}$$

得:

$$\begin{aligned} p_1 &= \frac{p'_1 p'_3}{p_3}, \\ p_2 &= \frac{1 - (1-p'_2)(1-p'_3)}{1-p_3}, \end{aligned}$$

代入后, 可得:

$$\begin{aligned} (1-p'_1 p'_3) - \frac{(1-p'_2)(1-p'_3)}{(1-p_3)} + \frac{p_3(1-p'_2)(1-p'_3)}{(1-p_3)} = \\ (1-p'_1 p'_3) p'_2 + (1-p'_1)(1-p'_2) p', \end{aligned}$$

恒等式为:

$$\begin{aligned} (1-p'_1 p'_3) - (1-p'_2)(1-p'_3) = \\ (1-p'_1 p'_3) p'_2 + (1-p'_1)(1-p'_2) p'. \end{aligned}$$

记 $p_1 p_3 = u, (1-p_2)(1-p_3) = v$, 得:

$$p_1 = u / p_3, \quad p_2 = 1 - p_3 v / (p_3 - u).$$

同样可得:

$$\begin{aligned} p'_1 &= \frac{u}{p'_3}, \\ p'_2 &= 1 - \frac{p'_3 v}{p'_3 - u}, \end{aligned}$$

若其中 1 个参数已知, 则其他参数皆可识别.

2 二元二点分布的参数估计

从定理 2 出发, 若其中 1 个参数已知, 可得其他参数的最大似然估计.

定理 3 设 $(X, Y) \sim BBD(1, p_1, p_2, p_3)$ 是总体, $(X_1, Y_1), \dots, (X_n, Y_n)$ 是来自总体 (X, Y) 的容量为 n 的样本, 记 $Z = \min(X, Y)$, 定义随机变量 $I = 1, 2, 3$ 分别对应于 $X < Y, Y < X, X = Y$; 记 $Z_i = \min(X_i, Y_i)$, 定义随机变量 $I_i = 1, 2, 3$ 分别对应于 $X_i < Y_i, Y_i < X_i, X_i = Y_i$ 时, $i = 1, \dots, n$, 若 (Z, I) 具有以下的联合分布:

$$P(Z=i, I=j) = \begin{cases} (1-p_1 p_3) p_2 + (1-p_1)(1-p_2) p_3, & i=0, j=1, \\ 0, & i=0, j=2, \\ (1-p_2)(1-p_3), & i=0, j=3, \\ 0, & i=1, j=1, \\ 0, & i=1, j=2, \\ p_1 p_3, & i=1, j=3, \end{cases}$$

$(Z_1, I_1), \dots, (Z_n, I_n)$ 是来自总体 (Z, I) 的容量为 n 的样本, 若其中 1 个参数已知, 则其他参数的最大似然估计可通过似然方程直接求解可得, 其中,

$$\begin{aligned} w_{1i}^0 &= \begin{cases} 1, & Z_i=0, I_i=1, \\ 0, & \text{其他}, \end{cases} \quad w_{1i}^1 = \begin{cases} 1, & Z_i=1, I_i=1, \\ 0, & \text{其他}, \end{cases} \\ w_{2i}^0 &= \begin{cases} 1, & Z_i=0, I_i=2, \\ 0, & \text{其他}, \end{cases} \quad w_{2i}^1 = \begin{cases} 1, & Z_i=1, I_i=2, \\ 0, & \text{其他}, \end{cases} \\ w_{3i}^0 &= \begin{cases} 1, & Z_i=0, I_i=3, \\ 0, & \text{其他}, \end{cases} \quad w_{3i}^1 = \begin{cases} 1, & Z_i=1, I_i=3, \\ 0, & \text{其他}, \end{cases} \\ i &= 1, \dots, n. \end{aligned}$$

证明 假设 3 个参数均未知, 似然函数为:

$$\begin{aligned} L = & ((1-p_1 p_3) p_2 + (1-p_1)(1-p_2) p_3)^{\sum_{i=1}^n w_{1i}^0} \cdot \\ & ((1-p_2)(1-p_3))^{\sum_{i=1}^n w_{3i}^0} (p_1 p_3)^{\sum_{i=1}^n w_{3i}^1}, \end{aligned}$$

并有似然方程如下:

$$\begin{cases} \frac{\partial \ln L}{\partial p_1} = \frac{\sum_{i=1}^n w_{3i}^1}{p_1} - \frac{p_3 \sum_{i=1}^n w_{1i}^0}{(1-p_1 p_3) p_2 + (1-p_1)(1-p_2) p_3} = 0, \\ \frac{\partial \ln L}{\partial p_2} = -\frac{\sum_{i=1}^n w_{3i}^0}{1-p_2} + \frac{(1-p_3) \sum_{i=1}^n w_{1i}^0}{(1-p_1 p_3) p_2 + (1-p_1)(1-p_2) p_3} = 0, \\ \frac{\partial \ln L}{\partial p_3} = \frac{\sum_{i=1}^n w_{3i}^1}{p_3} - \frac{\sum_{i=1}^n w_{3i}^0}{1-p_3} + \frac{(1-p_1-p_2) \sum_{i=1}^n w_{1i}^0}{(1-p_1 p_3) p_2 + (1-p_1)(1-p_2) p_3} = 0, \end{cases}$$

得:

$$\frac{(\sum_1^n w_{3i}^1) / p_1}{(\sum_1^n w_{3i}^0) / (1-p_2)} = \frac{p_3}{1-p_3},$$

$$p_3 = \frac{(1-p_2) \sum_1^n w_{3i}^1}{p_1 \sum_1^n w_{3i}^0 + (1-p_2) \sum_1^n w_{3i}^1}.$$

记 $\frac{p_1}{1-p_2} = a$, 则,

$$p_1 = a(1-p_2),$$

$$p_3 = \frac{\sum_1^n w_{3i}^1}{a \sum_1^n w_{3i}^0 + \sum_1^n w_{3i}^1},$$

代入后可得:

$$p_2 = (a \sum_1^n w_{3i}^0 \sum_1^n (w_{li}^0) \sum_1^n w_{3i}^1 (1-a)) /$$

$$((\sum_1^n w_{3i}^1 (a \sum_1^n w_{3i}^0 + \sum_1^n w_{3i}^1) \sum_1^n (w_{3i}^0 + w_{3i}^1) -$$

$$(a \sum_1^n w_{3i}^0 + \sum_1^n w_{3i}^1 \sum_1^n w_{3i}^1)^2 +$$

$$a(1-a) \sum_1^n w_{3i}^0 \sum_1^n w_{li}^0 \sum_1^n (w_{3i}^1)),$$

$$p_1 = (\sum_1^n (w_{3i}^1) (a \sum_1^n w_{3i}^0 + \sum_1^n w_{3i}^1) \sum_1^n (w_{3i}^0 + w_{3i}^1) -$$

$$(a \sum_1^n w_{3i}^0 + \sum_1^n w_{3i}^1 \sum_1^n w_{3i}^1)^2) /$$

$$(\sum_1^n w_{3i}^1 (a \sum_1^n w_{3i}^0 + \sum_1^n w_{3i}^1) \sum_1^n (w_{3i}^0 + w_{3i}^1) -$$

$$(a \sum_1^n w_{3i}^0 + \sum_1^n w_{3i}^1 \sum_1^n w_{3i}^1)^2 +$$

$$a(1-a) \sum_1^n w_{3i}^0 \sum_1^n w_{li}^0 \sum_1^n (w_{3i}^1)),$$

因此,

$$\left\{ \begin{aligned} p_1 &= (\sum_1^n w_{3i}^1 (a \sum_1^n w_{3i}^0 + \sum_1^n w_{3i}^1) \sum_1^n (w_{3i}^0 + w_{3i}^1) - \\ & (a \sum_1^n w_{3i}^0 + \sum_1^n w_{3i}^1 \sum_1^n w_{3i}^1)^2) / (\sum_1^n w_{3i}^1 \cdot \\ & (a \sum_1^n w_{3i}^0 + \sum_1^n w_{3i}^1) \sum_1^n (w_{3i}^0 + w_{3i}^1) - (a \sum_1^n w_{3i}^0 + \\ & \sum_1^n w_{3i}^1 \sum_1^n w_{3i}^1)^2 + a(1-a) \sum_1^n w_{3i}^0 \sum_1^n w_{li}^0 \sum_1^n (w_{3i}^1)), \end{aligned} \right.$$

$$\left\{ \begin{aligned} p_2 &= (a \sum_1^n w_{3i}^0 \sum_1^n w_{li}^0 \sum_1^n w_{3i}^1 (1-a)) / ((\sum_1^n w_{3i}^1 \cdot \\ & (a \sum_1^n w_{3i}^0 + \sum_1^n w_{3i}^1) \sum_1^n (w_{3i}^0 + w_{3i}^1) - (a \sum_1^n w_{3i}^0 + \\ & \sum_1^n w_{3i}^1 \sum_1^n w_{3i}^1)^2 + a(1-a) \sum_1^n w_{3i}^0 \sum_1^n w_{li}^0 \sum_1^n (w_{3i}^1)), \\ p_3 &= \sum_1^n w_{3i}^1 / (a \sum_1^n w_{3i}^0 + \sum_1^n w_{3i}^1), \end{aligned} \right.$$

即 3 个参数不可能同时可估计. 显然, 不妨设参数 p_3 已知, 可得似然方程:

$$\left\{ \begin{aligned} \frac{\partial \ln L}{\partial p_1} &= \frac{\sum_1^n w_{3i}^1}{p_1} - \\ & \frac{p_3 \sum_1^n w_{li}^0}{(1-p_1 p_3) p_2 + (1-p_1)(1-p_2) p_3} = 0, \\ \frac{\partial \ln L}{\partial p_2} &= -\frac{\sum_1^n w_{3i}^0}{1-p_2} + \\ & \frac{(1-p_3) \sum_1^n w_{li}^0}{(1-p_1 p_3) p_2 + (1-p_1)(1-p_2) p_3} = 0, \end{aligned} \right.$$

及

$$p_1 = \frac{(1-p_2)(1-p_3) \sum_1^n w_{3i}^1}{p_3 \sum_1^n w_{3i}^0}.$$

代入后可得:

$$(1-p_2)(1-p_3) \sum_1^n w_{li}^0 =$$

$$p_2 \sum_1^n w_{3i}^0 - (1-p_2)(1-p_3) \sum_1^n w_{3i}^1 +$$

$$(1-p_2) p_3 \sum_1^n w_{3i}^0.$$

所以, 由此可得:

$$\hat{p}_2 = 1 - \frac{\sum_1^n w_{3i}^0}{n(1-p_3)}, \hat{p}_1 = \frac{\sum_1^n w_{3i}^1}{np_3}.$$

3 结论

仅是 Z 的分布已知时, 所有参数皆不可识别; 当 (Z, I) 的分布已知时, 若其中 1 个参数已知, 则

其他参数皆可识, 即其他参数皆可估计.

参考文献:

- [1] MARSHALL A W, OLKIN I. A family of bivariate distributions generated by the bivariate bernoulli distribution[J]. J Amer Statist Assoc, 1985, 390(80):332-338.
- [2] OLUYEDE B O. A family of bivariate binomial distributions generated by extreme bernoulli distributions[J]. Communications in Statistics Theory and Methods, 1994, 23(5):1531-1547.
- [3] ISLAM M A, ALZAID A A, CHOWDHURY R I, et al. A generalized bivariate bernoulli model with covariate dependence[J]. Journal of Applied Statistics, 2013, 40(5): 1064-1075.
- [4] 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.
- [5] GILLILAND D C, HANNAN J. Identification of the ordered bivariate normal distribution by minimum variate [J]. J Am Stat Assoc, 1980, 75:651-654.
- [6] MUKHERJEA A, STEPHENS R. The problem of identification of parameters by the distribution of the maximum random: Solution for the trivariate normal case[J]. Journal of Multivariate Analysis, 1990, 34(1):95-115.
- [7] MUKHERJEA A, STEPHENS R. Identification of parameters by the distribution of the maximum random variable: The general multivariate normal case[J]. Probability Theory and Related Fields, 1990, 34(1):289-296.
- [8] GONG J, MUKHERJEA A. Solution of a problem on the identification of parameters by the distribution of the maximum random variable: a multivariate normal case[J]. J Theor Probab, 1991, 4(4):783-790.
- [9] MUKHERJEA A, STEPHENS R. Identification of the parameters of a trivariate normal vector by the distribution of the minimum[J]. Journal of Statistical Planning and Inference, 1999, 78(1):23-37.
- [10] DAI M, MUKHERJEA A. Identification of the parameters of a multivariate normal vector by the distribution of the maximum[J]. J Theor Probab, 2001, 14:267-298.
- [11] DAVIS J, MUKHERJEA A. Identification of parameters by the distribution of the minimum: The tri-variate normal case with negative correlations[J]. Journal of Multivariate Analysis, 2007, 98(6):1141-1159.
- [12] BI L, MUKHERJEA A. Identification of parameters and the distribution of the minimum of the tri-variate normal [J]. Statistics & Probability Letters, 2010, 80(23):1819-1826.
- [13] MUKHERJEA A, ELNAGGAR M. Solution of the problem of the identified minimum for the tri-variate normal[J]. Proc Indian Acad Sci (Math Sci), 2012, 122(4): 645-660.

Identifiability and Parameter Estimation of a Bivariate Bernoulli Distribution

LI Guo-an, LI Jing-jing

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

Abstract: This paper proves that, given the distribution of identified minimum on the condition that two random variables have a bivariate bernoulli distribution, and that one of the parameters is known, then the other parameters can be identified. Based on this proof, the maximum likelihood estimator of the other parameters is derived.

Key words: a bivariate bernoulli distribution; identifiability; maximum likelihood estimator

(责任编辑 章践立)