

四参数二元 McKay 型伽马分布的几个性质

李国安¹, 李建峰¹, 李穆真²

(1. 宁波大学 理学院, 浙江 宁波 315211; 2. 美国密苏里哥伦比亚大学 统计系, 密苏里 哥伦比亚 65201)

摘要: 讨论了四参数二元 McKay 型伽马分布的几个性质. 通过变量替换技术, 导出了四参数二元 McKay 型伽马分布的特征及模拟性质; 通过已知可识最小值及差值的分布密度, 所有参数皆可识别, 得到识别性质; 通过计算 X, Y 之间的相关系数, 证明了相关系数落在 $(0, 1)$ 内, 得到相关性质, 证明了 X 的尾部关于 Y 的尾部渐近不相关以及 Y 的尾部关于 X 的尾部渐近正相关, 得到渐近性质; 最后, 得到了四参数二元 McKay 型伽马分布参数的矩估计及最大似然估计. 从应用层面考虑, 采用二步拟合, 先导出了位置参数的最大似然估计, 进而导出尺度参数和形态参数的矩估计, 得到估计性质. 利用已有文献中的实例数据, 给出了基于四参数二元 McKay 型伽马分布的地震灾害评估参数的估计值.

关键词: 四参数二元 McKay 型伽马分布; 相关性质; 渐近性质; 二步拟合

中图分类号: O211.4

文献标志码: A

文章编号: 1001-5132 (2018) 01-0062-07

Pasari 和 Dikshit 在文献[1-2]中用一元三参数一般指数分布和三参数 Weibull 分布拟合地震数据, 在文献[3]中用二参数伽马分布(含尺度参数和形态参数, 无位置参数)拟合地震数据, 从而实现了地震发生的概率预测, 但不能实现不同级别地震发生的概率预测. 李国安等^[4]也用一元三参数一般指数分布拟合了地震数据. 从充分利用统计信息的角度看, 此类地震发生的概率预测模型有改进的余地. Hristopulos 等^[5]虽然提到了地震强度统计数据, 但仍是用一元生存分布拟合地震时间发生间隔数据, 本质上是用一元生存分布建立预测模型. Bacchi 等^[6]用二元指数分布拟合了强度和极值雨量的下雨持续时间. Dargahi-Noubary 等^[7]用二元指数分布建立地震灾害风险评估模型, 2 个随机变量分别表示地震灾害发生强度及由地震导致的地震波的持续时间. 因此, 为推广一元二参数伽马分布的地震概率预测模型, 可从二元伽马分布中去选择符合条件的二元伽马分布, 表示地震时间发生间隔长度的随机变量的边缘分布是一元三参数伽马分布(含尺度参数、形态参数以及位置参数).

本文筛选出四参数二元 McKay 型伽马分布. McKay^[8]提出了二元 McKay 型伽马分布; Gupta 等^[9]讨论了服从二元 McKay 型伽马分布的 2 个随机变量的和、积及商的分布; NADARAJAH 等^[10]基于特征性质介绍了二类二元伽马分布, 并导出其密度函数的不同表示; 李国安等^[11]讨论了二元 McKay 型伽马分布的一个特征及其参数估计; Yue 等^[12]对二元伽马分布在水文学中的应用作了综述; Basu 等^[13]给出了二元正态分布的识别性; Domma^[14]研究了二元 Burr III 型分布的几个性质; 李国安^[15]研究了二元 Marshall-Olkin 型指数分布的独立性与不相关性; Lin 等^[16]研究了二元 Marshall-Olkin 型指数分布的相关性质; 李国安等^[17]研究了二元 Lawrance-Lewis 型指数分布的识别性及渐近不独立性; Kotz 等^[18]给出了二元 Arnold-Strauss 型指数分布的特征; Yu^[19]给出了二元 Arnold-Strauss 型指数分布的模拟; Nadarajah 等^[20]讨论了基于二元指数分布的可靠性模型; Gupta^[21]讨论了基于二元 Arnold-Strauss 型指数分布的可靠性模型; Vinh 等^[22]给出了相依样本下基于二元 Burr III 型分布的基尼系数的最大似然

估计; 董永权等^[23]给出了尾部相关性的定义; 依据 Marshall 等^[24]添加一个参数创造新分布的方式以及为达到更好的拟合效果对三参数二元 McKay 型伽马分布添加位置参数.

本文讨论了四参数二元 McKay 型伽马分布的几个性质. 通过变量替换技术, 导出四参数二元 McKay 型伽马分布的特征及模拟性质; 通过在已知可识最小值及差值的分布密度时所有参数皆可识别, 得到其识别性质; 通过计算 X, Y 之间的相关系数, 证明相关系数落在 $(0, 1)$ 内, 得到其相关性; 证明 X 的尾部关于 Y 的尾部渐近不相关, 以及 Y 的尾部关于 X 的尾部渐近正相关, 得到其渐近性质; 最后讨论四参数二元 McKay 型伽马分布的参数估计, 对于出现位置参数或截断参数情形下的参数估计, 参考文献[25-26]中参数估计方法, 采用二步估计, 先导出位置参数的最大似然估计, 进而导出尺度参数和形态参数的矩估计, 得到估计性质. 利用文献[1-2]中的实例数据, 给出了基于四参数二元 McKay 型伽马分布的地震灾害评估参数的估计值.

1 四参数二元 McKay 型伽马分布的模拟性质

定义 1 称 (X, Y) 服从二元 McKay 型伽马分布, 其联合密度函数为:

$$f(x, y) = \frac{a^{p+q}}{\Gamma(p)\Gamma(q)} x^{p-1} (y-x)^{q-1} \exp(-ay),$$

$$y > x > 0, a > 0, p > 0, q > 0, \text{ 记作}^{[3-4]}$$

$$(X, Y) \sim BMGD(a, p, q).$$

定义 2 称 (X, Y) 服从四参数二元 McKay 型伽马分布, 其联合密度函数为:

$$f(x, y) = \frac{a^{p+q}}{\Gamma(p)\Gamma(q)} (x-\beta)^{p-1} (y-x)^{q-1} \exp[-a(y-\beta)],$$

$$y > x \geq \beta, a > 0, p > 0, q > 0, \text{ 记作}$$

$$(X, Y) \sim BMGD(a, p, q, \beta).$$

引理 1 若 $(X, Y) \sim BMGD(a, p, q, \beta)$, 则 $X \sim GD(a, p, \beta)$, $Y \sim GD(a, p+q, \beta)$. 这里 $X \sim GD(a, p, \beta)$ 表示 X 服从参数为 a, p, β 的伽马分布, 其中 β 是位置参数.

证明

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy =$$

$$\int_x^{\infty} \frac{a^{p+q}}{\Gamma(p)\Gamma(q)} (x-\beta)^{p-1} (y-x)^{q-1} \exp[-a(y-\beta)] dy =$$

$$\frac{a^{p+q}}{\Gamma(p)\Gamma(q)} (x-\beta)^{p-1} \int_0^{\infty} y^{q-1} \exp(-ay) dy =$$

$$\frac{a^p}{\Gamma(p)} (x-\beta)^{p-1} \exp[-a(x-\beta)], x \geq \beta, \text{ 即}$$

$$X \sim GD(a, p, \beta).$$

同理可证 $Y \sim GD(a, p+q, \beta)$.

定理 1 $(X, Y) \sim BMGD(a, p, q, \beta)$ 当且仅当 $W = Y - X$ 与 $V = X$ 相互独立, 且 $W \sim GD(a, q)$, $V \sim GD(a, p, \beta)$.

证明 直接由 2 个联合密度函数相互转换可得, 先得到服从 $GD(a, p, \beta)$ 的伪随机数 v , 再得到服从 $GD(a, q)$ 的伪随机数 w , 记 $x = v$, $y = v + w$, 则 (x, y) 为来自 $BMGD(a, p, q, \beta)$ 的伪随机数.

2 四参数二元 McKay 型伽马分布的识别性质

记 $Z = \min(X, Y)$, 记 Z 的分布密度为 $f_Z(z)$, 定义随机变量 $I = 1, 2$ 分别对应 $X < Y$, $X \geq Y$ 时, 记 $p_i = P(I = i)$, $i = 1, 2$, 记 $f_i(z)$ 表示给定 $I = i$ 时 Z 的条件密度函数; 记 $p_i f_i(\cdot)$, $i = 1, 2$ 表示 (Z, I) 的联合密度; 记 $p_i f_i(z) f_{|Y-X|}^{(i)}(u)$ 表示 $(Z, I, |Y-X|)$ 的联合密度.

引理 2 若 $(X, Y) \sim BMGD(a, p, q, \beta)$, 则 Z 的分布密度为:

$$f_Z(z) = \frac{a^p}{\Gamma(p)} (z-\beta)^{p-1} \exp[-a(z-\beta)], z > 0.$$

证明 直接计算可得, 这里略去.

引理 3 若 $(X, Y) \sim BMGD(a, p, q, \beta)$, 则 (Z, I) 的分布密度为:

$$p_i f_i(z) = \begin{cases} \frac{a^p}{\Gamma(p)} (z-\beta)^{p-1} \exp[-a(z-\beta)], \\ i = 1, z \geq \beta, \\ 0, i = 2, z \geq \beta. \end{cases}$$

证明 直接计算可得, 这里略去.

引理 4 若 $(X, Y) \sim BMGD(a, p, q, \beta)$, 则 $(Z, I, |Y-X|)$ 的联合密度为:

$$p_i f_i(z) f_{|Y-X|}^{(i)}(u) = \begin{cases} \frac{a^{p+q}}{\Gamma(p)\Gamma(q)} (z-\beta)^{p-1} u^{q-1} \exp(-a(z+u-\beta)), \\ i=1, z \geq \beta, u > 0, \\ 0, i=2, z \geq \beta, u > 0. \end{cases}$$

证明 直接计算可得, 这里略去.

定理 2 若 $(X, Y) \sim BMGD(a, p, q, \beta)$, $(X', Y') \sim BMGD(a', p', q', \beta')$, 若已知 Z 与 Z' 同分布, 则参数 a, p, β 是可识别的.

证明 由 $f_Z(z) = f_{Z'}(z)$ 得:

$$\frac{a^p}{\Gamma(p)} (z-\beta)^{p-1} \exp[-a(z-\beta)] = \frac{a'^{p'}}{\Gamma(p')} (z-\beta')^{p'-1} \exp[-a'(z-\beta')],$$

由此得 $a = a'$, $p = p'$, $\beta = \beta'$.

定理 3 若 $(X, Y) \sim BMGD(a, p, q, \beta)$, $(X', Y') \sim BMGD(a', p', q', \beta')$, 若已知 (Z, I) 与 (Z', I') 同分布, 则参数 a, p, β 是可识别的.

证明 同定理 2 证明, 这里略去.

定理 4 若 $(X, Y) \sim BMGD(a, p, q, \beta)$, $(X', Y') \sim BMGD(a', p', q', \beta')$, 若已知 $(Z, I, |Y-X|)$ 与 $(Z', I', |Y'-X'|)$ 同分布, 则参数 a, p, q, β 是可识别的.

证明 由 $p_i f_i(z) f_{|Y-X|}^{(i)}(u) = p'_i f'_i(z) f_{|Y'-X'|}^{(i)}(u)$, 得:

$$a = a', \quad p = p', \quad q = q', \quad \beta = \beta'.$$

3 四参数二元 McKay 型伽马分布的相关性质

定理 5 若 $(X, Y) \sim BMGD(a, p, q, \beta)$, 则

$$\rho_{(X,Y)} = \sqrt{1/(1+q/p)}, \text{ 即 } 0 < \rho_{(X,Y)} < 1.$$

证明 由 $Cov(X, Y) = Cov(X - \beta, Y - \beta)$,

$$DX = D(X - \beta), \quad DY = D(Y - \beta),$$

不妨设 $(X, Y) \sim BMGD(a, p, q)$,

$$EX = \int_0^\infty x \frac{a^p}{\Gamma(p)} x^{p-1} \exp(-ax) dx =$$

$$\int_0^\infty \frac{1}{a\Gamma(p)} t^p \exp(-t) dt = \frac{\Gamma(p+1)}{a\Gamma(p)} = \frac{p}{a},$$

$$EX^2 = \int_0^\infty x^2 \frac{a^p}{\Gamma(p)} x^{p-1} \exp(-ax) dx =$$

$$\int_0^\infty \frac{1}{a^2\Gamma(p)} t^{p+1} \exp(-t) dt =$$

$$\frac{\Gamma(p+2)}{a^2\Gamma(p)} = \frac{(p+1)p}{a^2},$$

$$EY = \int_0^\infty y \frac{a^{p+q}}{\Gamma(p+q)} y^{p+q-1} \exp(-ay) dy =$$

$$\int_0^\infty \frac{1}{a\Gamma(p+q)} t^{p+q} \exp(-t) dt =$$

$$\frac{\Gamma(p+q+1)}{a\Gamma(p+q)} = \frac{p+q}{a},$$

$$EY^2 = \int_0^\infty y^2 \frac{a^{p+q}}{\Gamma(p+q)} y^{p+q-1} \exp(-ay) dy =$$

$$\int_0^\infty \frac{1}{a^2\Gamma(p+q)} t^{p+q+1} \exp(-t) dt =$$

$$\frac{\Gamma(p+q+2)}{a^2\Gamma(p+q)} = \frac{(p+q+1)(p+q)}{a^2},$$

$$EXY = \int_0^\infty \int_x^\infty \frac{a^{p+q} xy}{\Gamma(p)\Gamma(q)} x^{p-1} (y-x)^{q-1} \exp(-ay) dy dx =$$

$$\int_0^\infty dx \int_0^\infty \frac{a^{p+q} x(x+t)}{\Gamma(p)\Gamma(q)} x^{p-1} t^{q-1} \exp[-a(x+t)] dt \cdot$$

$$\int_0^\infty \frac{a^p x^2}{\Gamma(p)} x^{p-1} \exp(-ax) dx + \int_0^\infty \frac{a^{p-1} qx}{\Gamma(p)} x^{p-1} \cdot$$

$$\exp(-ax) dx = \frac{\Gamma(p+2)}{a^2\Gamma(p)} + \frac{q\Gamma(p+1)}{a^2\Gamma(p)} = \frac{p(p+q+1)}{a^2},$$

$$Cov(X, Y) = \frac{p(p+q+1)}{a^2} - \frac{p}{a} \frac{p+q}{a} = \frac{p}{a^2},$$

$$DX = \frac{(p+1)p}{a^2} - \frac{p^2}{a^2} = \frac{p}{a^2},$$

$$DY = \frac{(p+q+1)(p+q)}{a^2} - \frac{(p+q)^2}{a^2} = \frac{(p+q)}{a^2},$$

$$\rho_{(X,Y)} = \frac{p/a^2}{\sqrt{(p+q)/a^2} \sqrt{p/a^2}} = \sqrt{\frac{1}{1+q/p}}.$$

4 四参数二元 McKay 型伽马分布的渐近性质

二元分布的 2 个随机变量尾部相关性的定义参见文献[17], 其值为 $\eta_{Y|X} = \lim_{u \rightarrow \infty} P(Y > u | X > u)$, $\eta_{X|Y} = \lim_{u \rightarrow \infty} P(X > u | Y > u)$. 若 $0 < \eta_{Y|X} \leq 1$, 则称 Y 的尾部关于 X 的尾部渐近相关; 若 $\eta_{Y|X} = 0$, 则称 Y 的尾部关于 X 的尾部渐近不相关或渐近独立.

引理 5 若 $(X, Y) \sim BMGD(a, p, q, \beta)$, 则

$$P(X > u, Y > u) = \int_u^\infty dy \int_u^y \frac{a^{p+q}}{\Gamma(p)\Gamma(q)} \cdot \\ (x - \beta)^{p-1} (y - x)^{q-1} \exp[-a(y - \beta)] dx.$$

证明 直接计算可得, 这里略去.

定理 6 若 $(X, Y) \sim BMGD(a, p, q, \beta)$, 则 X 的尾部关于 Y 的尾部渐近不相关, 且 Y 的尾部关于 X 的尾部渐近正相关.

证明

$$\eta_{X|Y} = \lim_{u \rightarrow \infty} P(X > u | Y > u) = \\ \lim_{u \rightarrow \infty} P([X - \beta] > u | [Y - \beta] > u) = \\ \lim_{u \rightarrow \infty} \frac{\int_u^\infty dy \int_u^y \frac{a^{p+q} x^{p-1} (y-x)^{q-1}}{\Gamma(p)\Gamma(q)} \exp(-ay) dx}{\int_u^\infty \frac{y^{p+q-1} a^{p+q} \exp(-ay)}{\Gamma(p+q)} dy} = \\ \lim_{u \rightarrow \infty} \frac{\int_u^\infty \frac{1}{\Gamma(p)\Gamma(q)} x^{q-1} \exp(-ax) dx}{u^q / \Gamma(p+q)} = \\ \lim_{u \rightarrow \infty} \frac{\Gamma(p+q)}{q\Gamma(p)\Gamma(q)} \exp(-au) = 0,$$

X 的尾部关于 Y 的尾部渐近不相关. 同理

$$\eta_{Y|X} = \lim_{u \rightarrow \infty} P(Y > u | X > u) = \\ \lim_{u \rightarrow \infty} \frac{\frac{a^p u^{p-1} \exp(-au)}{\Gamma(p)\Gamma(q)} \int_0^\infty a^q (t)^{q-1} e^{-at} dt}{u^{p-1} a^p \exp(-au) / \Gamma(p)} = \\ \lim_{u \rightarrow \infty} \frac{a^p u^{p-1} \exp(-au) / \Gamma(p)}{u^{p-1} a^p \exp(-au) / \Gamma(p)} = 1,$$

Y 的尾部关于 X 的尾部渐近正相关.

5 四参数二元 McKay 型伽马分布的估计性质

定理 7 设总体 $(X, Y) \sim BMGD(a, p, q, \beta)$, $(X_1, Y_1), \dots, (X_n, Y_n)$ 为来自总体容量为 n 的样本, 记 $W_i = Y_i - X_i, i = 1, \dots, n, n\bar{X} = \sum_{i=1}^n X_i, n\bar{W} = \sum_{i=1}^n W_i$, 则参数 a, p, q, β 的矩估计分别为:

$$\hat{a} = n\bar{W} / (\sum_{i=1}^n W_i^2 - n\bar{W}^2), \\ \hat{p}_{1,2} = (1 + 4\hat{a}\bar{x} \pm \sqrt{1 + 8\hat{a}\bar{x} + 8\hat{a}^2\bar{x}^2}) / 4, \\ \hat{q} = \frac{n\bar{W}^2}{\sum_{i=1}^n W_i^2 - n\bar{W}^2}, \\ \hat{\beta}_{1,2} = \bar{x} - \frac{1 + 4\hat{a}\bar{x} \pm \sqrt{1 + 8\hat{a}\bar{x} + 8\hat{a}^2\bar{x}^2}}{4\hat{a}}.$$

证明 由 $W = Y - X \sim GD(a, q)$ 可得:

$$\hat{a} = \frac{n\bar{W}}{\sum_{i=1}^n W_i^2 - n\bar{W}^2}, \quad \hat{q} = \frac{n\bar{W}^2}{\sum_{i=1}^n W_i^2 - n\bar{W}^2},$$

由 $X \sim GD(a, p, \beta)$ 可得:

$$EX = \frac{p}{a} + \beta, \quad EX^2 = \frac{p(p+1)}{a^2} + \frac{2p\beta}{a} + \beta^2,$$

由 $\beta = \bar{x} - \frac{p}{\hat{a}}$ 代入得 $\frac{2p^2}{\hat{a}^2} - (\frac{1}{\hat{a}^2} + \frac{4\bar{x}}{\hat{a}})p + \bar{x}^2 = 0$, 得:

$$\hat{p}_{1,2} = \frac{1 + 4\hat{a}\bar{x} \pm \sqrt{1 + 8\hat{a}\bar{x} + 8\hat{a}^2\bar{x}^2}}{4}, \\ \hat{\beta}_{1,2} = \bar{x} - \frac{1 + 4\hat{a}\bar{x} \pm \sqrt{1 + 8\hat{a}\bar{x} + 8\hat{a}^2\bar{x}^2}}{4\hat{a}}.$$

定理 8 设总体 $(X, Y) \sim BMGD(a, p, q, \beta)$, $(X_1, Y_1), \dots, (X_n, Y_n)$ 为来自总体容量是 n 的样本, 则参数 β 的最大似然估计为 $\hat{\beta} = \min(x_j)$, 其他参数 a, p, q 的最大似然估计通过迭代求解可得.

证明 由 $W = Y - X \sim GD(a, q)$, 得似然函数:

$$L(a, q) = \frac{a^{nq}}{(\Gamma(q))^n} (\prod_{i=1}^n w_i)^{q-1} \exp(-a \sum_{i=1}^n w_i),$$

$$\ln L(a, q) = nq \ln a - n \ln \Gamma(q) +$$

$$(q-1) \sum_{i=1}^n \ln w_i - a \sum_{i=1}^n w_i,$$

由此得方程组:

$$\begin{cases} \frac{nq}{a} - \sum_{i=1}^n w_i = 0, \\ n \ln a - \frac{n\Gamma'(q)}{\Gamma(q)} + \sum_{i=1}^n \ln w_i = 0, \end{cases} \quad (1)$$

$$\begin{cases} \frac{nq}{a} - \sum_{i=1}^n w_i = 0, \\ n \ln a - \frac{n\Gamma'(q)}{\Gamma(q)} + \sum_{i=1}^n \ln w_i = 0, \end{cases} \quad (2)$$

从式(1)得:

$$a = \frac{nq}{\sum_{i=1}^n w_i},$$

代入式(2)得超级方程:

$$n \ln \frac{nq}{\sum_{i=1}^n w_i} - \frac{n\Gamma'(q)}{\Gamma(q)} + \sum_{i=1}^n \ln w_i = 0,$$

用数值计算可得 $\hat{q}$, 回代得 $\hat{a}$.

由 $X \sim GD(a, p, \beta)$, 得似然函数:

$$L(a, p, \beta) = \frac{a^{np}}{(\Gamma(p))^n} (\prod_{i=1}^n (x_i - \beta))^{p-1} \cdot$$

$$\exp(-a \sum_{i=1}^n (x_i - \beta)) I(\beta \leq \min(x_j)),$$

得:

$$\hat{\beta} = \min(x_j),$$

由此得方程组:

$$\begin{cases} \frac{np}{a} - \sum_{i=1}^n (x_i - \hat{\beta}) = 0, \\ n \ln a - \frac{n\Gamma'(p)}{\Gamma(p)} + \sum_{i=1}^n \ln(x_i - \hat{\beta}) = 0, \end{cases} \quad (3)$$

从式(3)得:

$$\hat{p} = \frac{\hat{a} \sum_{i=1}^n (x_i - \hat{\beta})}{n},$$

再代入式(4)解得:

$$\hat{a} = \exp\left[\frac{\Gamma'(\hat{p})}{\Gamma(\hat{p})} - \frac{\sum_{i=1}^n \ln(x_i - \hat{\beta})}{n}\right],$$

形成迭代求解.

初始值 $\hat{q}_0 = \hat{q}$ 代入, 得 $\hat{a}_0 = n\hat{q}_0 / \sum_{i=1}^n w_i$, 代入得 $\hat{p}_0 = \hat{a}_0 \sum_{i=1}^n (x_i - \hat{\beta}) / n$, 代入得:

$$\tilde{a}_1 = \exp[\Gamma'(\hat{p}_0) / \Gamma(\hat{p}_0) - \sum_{i=1}^n \ln(x_i - \hat{\beta}) / n],$$

回代得 $\hat{q}_1, \hat{a}_1, \hat{p}_1$, 再得 $\tilde{a}_2$, 继续直至符合给定精度.

由文献[26]可知: 矩方法在位置参数的估计中存在缺陷, 四参数二元 McKay 型伽马分布求解参数的最大似然估计的超级方程中出现 $\Gamma'(p)$, 计算量大甚至不容易求解. 另外, 从定理 7 可知, 参数 p, β 有 2 组估计: $\hat{p}_{1,2}$ 和 $\hat{\beta}_{1,2}$. 因此从应用层面考虑, 吸收最大似然估计法的优点, 采用二步拟合, 先导出位置参数的最大似然估计, 进而导出尺度参数和形态参数的矩估计.

定理 9 设总体 $(X, Y) \sim BMGD(a, p, q, \beta)$, $(X_1, Y_1), \dots, (X_n, Y_n)$ 为样本, 则参数 β 的最大似然估计为 $\hat{\beta} = \min_j(x_j)$, 参数 q 的矩估计为:

$$\tilde{q} = \frac{n\bar{W}^2}{\sum_{i=1}^n W_i^2 - n\bar{W}^2},$$

参数 a, p 的二步矩估计分别为:

$$\hat{a} = \frac{\bar{x} - \hat{\beta}}{\frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2}, \quad \hat{p} = \frac{(\bar{x} - \hat{\beta})^2}{\frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2}.$$

证明 由 $X \sim GD(a, p, \beta)$, 得 $\hat{\beta} = \min_j(x_j)$, 假设 $X - \hat{\beta} \sim GD(a, p)$, 则由定理 5 的证明知:

$$E(X - \hat{\beta}) = \frac{p}{a}, \quad E(X - \hat{\beta})^2 = \frac{(p+1)p}{a^2},$$

可得:

$$\frac{E(X - \hat{\beta})^2}{E(X - \hat{\beta})} = \frac{(p+1)}{a}, \text{ 得:}$$

$$a = \frac{E(X - \hat{\beta})}{E(X - \hat{\beta})^2 - (E(X - \hat{\beta}))^2}.$$

a 的二步矩估计为:

$$\hat{a} = \frac{\bar{x} - \hat{\beta}}{\frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2}.$$

p 的二步矩估计为:

$$\hat{p} = \frac{(\bar{x} - \hat{\beta})^2}{\frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2}.$$

由 $W = Y - X \sim GD(a, q)$, 得 q 的矩估计为:

$$\tilde{q} = \frac{n\bar{W}^2}{\sum_{i=1}^n W_i^2 - n\bar{W}^2}.$$

注记 1 由文献[1]可得 X, Y 的观察数据(表 1), 其中 X 表示发生 2 次相邻地震灾害的时间间隔, Y 表示地震级别乘 3.09.

表 1 地震灾害发生时间的间隔长度及震级

序号	X/a	Y
1	21.6	23.2
2	16.5	22.6
3	12.4	26.9
4	11.5	23.2
5	3.4	24.4
6	6.1	23.5
7	5.2	21.9
8	7.4	23.5
9	12.7	22.2
10	2.9	23.2
11	0.9	24.4
12	3.1	26.6
13	1.3	21.6
14	2.3	23.2
15	6.9	23.5
16	15.3	21.6
17	12.2	22.2
18	2.4	21.9
19	4.5	21.9

由表 1 得参数的估计值分别为: $\hat{a} = 0.2065$, $\hat{p} = 1.4291$, $\tilde{q} = 6.382$, $\hat{\beta} = 0.9$.

注记 2 文献[3]用二参数伽马分布拟合地震间隔时间数据, 在已知最小间隔值为 $\hat{\beta} = 0.9$ 情况下, 仍然采用无位置参数的二参数伽马分布, 使得

根据已知样本数据信息得出在 $0 < x < 0.9$ 范围内没有样本数据值, 而经拟合后的二参数伽马分布可在 $0 < x < 0.9$ 内有密度, 即若抽样可能在此范围内抽到样本值, 违背了一致性原则. 从拟合精度看, 由于把部分密度分摊到无样本数据或有可能不存在密度的范围, 降低了拟合精度. 因此, 本文采用三参数伽马分布拟合地震间隔时间数据, 显然其拟合效果优于用二参数伽马分布的拟合效果. 在参数估计上, 设参数 β 的矩估计为 $\tilde{\beta}$, 若 $\tilde{\beta} < 0.9$, 采用其作为 β 的估计, 显然不符合最大似然原则; 若 $\tilde{\beta} > 0.9$, 采用其作为 β 的估计, 显然与已有样本最小值 $\min(x_j) = 0.9$ 矛盾, 且 3 个参数同时求矩估计计算量也很大, 甚至方程组求解有困难. 因此, 对于 β 的参数估计, 只能采用其最大似然估计值作为估计值, 也就是拟合数据过程不得不采用二步拟合. 4 个参数的估计值有 3 个名称, 分别为最大似然估计、矩估计和二步估计.

本文用四参数二元 McKay 型伽马分布拟合地震间隔时间与级别数据, 由二者之间的相关系数的估计值 $\hat{\rho}_{(x,y)} = 0.4277$ 得出二者之间存在一定的相关性, 故利用此相关信息的拟合效果优于用三参数伽马分布的拟合效果(可通过拟合优度检验数据来判定). 通过用四参数二元 McKay 型伽马分布拟合可预报某一段时间内不超过某级别地震发生的概率, 也可预报某一段时间内某级别区间地震发生的概率. 在某种程度上, 地震灾害损失额可看成地震级别的函数, 由此可建立与地震灾害保险精算对应的地震间隔时间与级别的二元生存分布拟合模型.

参考文献:

- [1] Pasari S, Dikshit O. Three-parameter generalized exponential distribution in earthquake recurrence interval estimation[J]. Nat Hazards, 2014, 73:639-656.
- [2] Pasari S, Dikshit O. Impact of three-parameter Weibull models in probabilistic assessment of earthquake hazards[J]. Pure and Applied Geophysics, 2014, 171: 1251-1281.
- [3] Pasari S, Dikshit O. Distribution of earthquake interevent times in northeast India and adjoining regions[J]. Pure and Applied Geophysics, 2015, 172:2533-2544.
- [4] 李国安, 李建峰. 指数分布抽样基本定理及在三参数一般指数分布参数估计中的应用[J]. 数学的实践与认识, 2017, 47(3):165-169.
- [5] Hristopulos D T, Mouslopoulou V. Strength statistics and the distribution of earthquake interevent times[J]. Physica A, 2013, 393:485-495.
- [6] Bacchi B, Bacciu G, Kottegoda N T. Bivariate exponential model applied to intensities and durations of extreme rainfall[J]. Journal of Hydrology, 1994, 155(1/2): 225-236.
- [7] Dargahi-Noubary G R, Razzaghi M. Earth-quake hazard assessment based on bivariate exponential distributions[J]. Reliability Engineering & System Safety, 1994,44(2): 153-166.
- [8] McKay A T. Sampling from batches[J]. Journal of the Royal Statistical Society, 1934(S):207-216.
- [9] Gupta A K, Nadarajah S. Sums, products and ratios for McKay's bivariate gamma distribution[J]. Mathematical and Computer Modelling, 2006, 43:185-193.
- [10] Nadarajah S, Gupta A K. Some bivariate gamma distributions[J]. Applied Mathematics Letters, 2006, 19: 767-774.
- [11] 李国安, 李建峰. 例说基于可识最小值之识别性与参数估计及特征的关系[J]. 大学数学, 2016, 32(4):20-29.
- [12] Yue S, Ouarda T B M J, Bobée B. A review of bivariate gamma distributions for hydrological application[J]. Journal of Hydrology, 2001, 246(1):1-18.
- [13] Basu A P, Ghosh J K. Identifiability of the multinorma and other distributions under competing risks model[J]. Journal of Multivariate Analysis, 1978, 8(3):413-429.
- [14] Domma F. Some properties of the bivariate Burr type III distribution[J]. Journal of Theoretical and Applied Statistics, 2010, 44(2):203-215.
- [15] 李国安. 二元 Marshall-Olkin 型指数分布的独立性和不相关性[J]. 大学数学, 2013, 29(4):91-93.
- [16] Lin G D, Lai C D, Govindaraju K. Correlation structure of the Marshall-Olkin bivariate exponential distribution [J]. Statistical Methodology, 2016, 29:1-9.
- [17] 李国安, 李建峰. 二元 Lawrance-Lewis 型指数分布的识别性及渐近不独立性[J]. 高等数学研究, 2016, 19(6): 44-47.
- [18] Kotz S, Navarro J, Ruiz J M. Characterizations of Arnold and Strauss' and related bivariate exponential models[J]. Journal of Multivariate Analysis, 2007, 98(7):1494-1507.
- [19] Yu Y M. Efficient simulation of a bivariate exponential conditionals distribution[J]. Computational Statistics & Data Analysis, 2008, 52:2273-2276.
- [20] Nadarajah S, Kotz S. Reliability models based on bivariate exponential distributions[J]. Probabilistic

- Engineering Mechanics, 2006, 21:338-351.
- [21] Gupta R C. Reliability studies of bivariate distribution with exponential conditionals[J]. Mathematical and Computer Modelling, 2008, 47:1009-1018.
- [22] Vinh A, Griffiths W E, Chotikapanich D. Bivariate income distributions for assessing inequality and poverty under dependent samples[J]. Economic Modelling, 2010, 27:1473-1483.
- [23] 董永权, 王占民. 关于相关系数 P 的几点注释[J]. 大学数学, 2008, 24(2):182-186.
- [24] Marshall A W, Olkin I. A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families[J]. Biometrika, 1997, 84:641-652.
- [25] Dixit U J, Nasiri P N. Estimation of parameters of a right truncated exponential distribution[J]. Statistical Papers, 2008, 49:225-236.
- [26] 李国安, 黄林涛, 李建峰. 最大似然估计法的探究式教学[J]. 大学数学, 2013, 29(3):144-146.

Some properties of the four-parameter McKay's bivariate gamma distribution

LI Guo-an¹, LI Jian-feng¹, LI Mu-zhen²

(1.Faculty of Science, Ningbo University, Ningbo 315211, China;

2.Department of Statistics, Columbia University of Missouri-Columbia, Columbia 65201, USA)

Abstract: Some properties of the four-parameter McKay's bivariate gamma distribution are considered in this paper. By using the technique of variate substitution, the characterization of the McKay's bivariate gamma distribution are derived, from which an efficient simulation of the McKay's bivariate gamma distribution is obtained. Also if the joint distribution of identified minimum and differential value is known, then all of parameters are identified; and correlation coefficient of X and Y is calculated. It is proven that the value of correlation coefficient is between zero and one; it also shows that the tail of X with respect to the tail of Y is asymptotically independent, while the tail of Y with respect to the tail of X is asymptotically positive related. The moment estimators and maximum likelihood estimators of parameters of the four-parameter McKay's bivariate gamma distribution are obtained. Finally, taking into account the application, fitting data by way of two steps, the moment estimator of one parameter, the maximum likelihood estimator of another parameter, as well as the modified moment estimators of the other two parameter of the four-parameter McKay's bivariate gamma distribution are derived based on integrating the method of moment estimator with the method of maximum likelihood estimator. In the end, the results of fitting to the earthquake data from the reference are provided.

Key words: the four-parameter McKay's bivariate gamma distribution; correlation property; asymptotical property; fitting data by way of two steps

(责任编辑 史小丽)