DOI:10.19789/j.1004-9398.2016.06.005

变形广义极值分布的参数估计及实例分析

樊利利

(吕梁学院数学系,山西 吕梁 033000)

摘 要

本文通过对广义极值(GEV)分布运用二次秩变形映射导出变形的广义极值(TGEV)分布的概率分布函数和概率密度函数,并对TGEV分布的参数进行了MLE和矩估计.用R语言中的optim函数对TGEV分布的参数做了MLE估计,并对此估计方法的偏差和均方误差做了模拟,说明对于TGEV分布来说,MLE估计方法效果很好.最后对变形广义极值分布进行实例分析,给出了TGEV分布拟合的结果,得到分位数及其置信区间的估计,并讨论了所得结果的实际意义.

关键词: 二次秩变形映射,变形广义极值分布,参数估计,非参数估计.
中图分类号: O211.4

0 引 言

极值统计是分析和预测小概率事件风险的有力工具,主要研究很少发生,然而一旦发生却产生极大影响的随机时间的统计规律性.这些随机事件(又称极端事件)既包括百年不遇的洪水,飓风和地震等的自然灾害,又包括金融领域中股市价格的暴涨暴跌等经济现象.由于极端时间的发生往往会给我们的生活带来重大的影响,近年来引起了很多学者的关注.目前极值统计已经广泛应用到水文、气象、地震等许多领域,而且在金融、保险等经济领域的风险评估中体现了良好的优越性.

在统计分析中所用的程序的质量很大程度上取决于假设的概率模型或者是概率分布.正因为如此,一直以来,人们将相当大的努力花费在了研究和发展一大类的标准概率分布上,设法将其应用于广泛现实数据的统计模型.然而,在现实世界中,仍然存在许多重要的问题,比如真实的数据不遵循任何古典或标准的概率模型,能实际去应用对称概率模型来模拟的数据很少.因此,对于每一个现实世界的真实数据,常用的正态概率模型不是一个有用的概率密度函数模型.很显然,对于一个服从非对称分布的现实数据来说,正态概率模型就变得不可用.反映非对称变化程度的非对称模型是Modeler工具包中的一个必要组成部分.在文献[1]中,根据Genton先生所介绍,关于非正态的概率分布的研究可以追溯到十九世纪.同样,在文献[2]中也可以看到Edgeworth对非对称频率数据拟合的非对称分布的问题已经进行了研究.而且一些关于峰度和偏度的统计模型最近也出现在文章中.其中,O'Hgana提出的反正态的概率模型被Azzalini进行了广泛的研究并延伸到其他对称模型,而且已在若干领域应用.对此有广泛应用的书目在文献[3]中已提供.

Azzalini's approach(文献[3]):设随机变量X服从一个反对称分布,如果$f(x) = 2g(x)G(\beta x)$,其中$g(\cdot)$和$G(\cdot)$分别是一个对称分布的概率密度函数和累积分布函数,$\beta = (-\infty, \infty)$是skewness参数.

在文献[4]中Aryal等应用斜拉普拉斯概率分布来研究货币汇率数据.如果(基本)的概率分布是非对称的,我们就不能应用Azzalini的方法来生成一个灵活的概率分布族.目前研究的目的就是研究可以从非对称分布(特别是极值分布)得到的一个概率分布,可用于建模和分析现实世界的数据.

1 变形映射

设F_1、F_2为来自同一样本空间的两个分布的累积分布函数,在文献[5]中给出了一般秩变形的定

收稿日期:2015-12-23

(C)1994-2019 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

义如下:

$$G_{R_{12}}(u) = F_2(F_1^{-1}(u)),$$
$$G_{R_{21}}(u) = F_1(F_2^{-1}(u)).$$

注意到其中的逆累积分布函数也叫做分位数函数,定义式为 $F^{-1}(y) = \inf_{x\in R}\{F(x) > y, y \in [0,1]\}$. 则可以看出 $G_{R_{12}}(u)$ 和 $G_{R_{21}}(u)$ 都是由单位区间 $I[0,1]$ 映射到单位区间 $I[0,1]$ 的. 且满足 $G_{R_{ij}}(0) = 0, G_{R_{ij}}(1) = 1$.

下面给出二次秩变形映射(Quadratic Rank Transmutation Map,简记为 QRTM)的定义: $G_{R_{12}}(u) = u + \lambda u(1-u)$, $|\lambda| \leqslant 1$, 其分布函数

$$F_2(x) = (1+\lambda)F_1(x) - \lambda(F_1(x))^2. \tag{1.2.1}$$

对其求导得

$$f_2(x) = f_1(x)[(1+\lambda) - 2\lambda F_1(x)], \tag{1.2.2}$$

其中 $f_1(x)$, $f_2(x)$ 分别为 $F_1(x)$, $F_2(x)$ 的概率密度函数,关于二次秩变形映射的更多信息请看文献[5],在此就不多介绍了.

下面的定理 1 建立了 Azzalini's formulation 和 QRTM formulation 之间的联系,其中随机变量服从均匀分布. 下面先给出 Azzalini's formulation:

Azzalini's approach(文献[3]): 若随机变量 X 服从一个反对称分布,当且仅当其密度函数为 $f(x) = 2g(x)G(\beta x)$ 的形式,其中 $g(\cdot)$ 和 $G(\cdot)$ 分别是一个对称分布的概率密度函数和累积分布函数,$\beta = (-\infty, \infty)$ 是 skewness 参数.

定理 1: 设 X 来自均匀分布 $(-\theta, \theta)$ 的随机变量,其概率密度函数为 $f(x)$,累积分布函数为 $F(x)$,则 $F^{-1}(x)$ 关于 0 对称. 那么对 $\forall \theta > 0, \beta \in R$, 有

$$g(x|\beta) = 2f(x)F(\beta x) = \frac{\max\{\min(\beta x, \theta), -\theta\} + \theta}{2\theta^2}, -\theta < x < \theta,$$

是一个概率密度函数. 如果在定义的变形均匀分布中 λ 被看作是形状参数的话,则 $|\beta| \leqslant 1$ 时有 $\lambda = -\beta$.

注意到 Azzalini's formulation 只适用于对称分布,而 QRTM formulation 则适用于任何分布(对称和非对称分布),另外,在 Azzalini's formulation 中形状参数 $\beta \in R$, 而在 QRTM formulation 中形状参数 $\lambda \in [-1,1]$, 且两个方程都是在形状参数取 0 值时,其分布等于基分布.

2　变形广义极值分布

设随机变量 X 服从广义极值分布,则其累积分布函数为

$$F(x) = \exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\},$$

其中 $1+\xi\left(\frac{x-\mu}{\sigma}\right) > 0$, $-\infty < \mu < \infty$ 是位置参数, $\sigma > 0$ 是尺度参数, $-\infty < \xi < \infty$ 是形状参数. 相应的概率密度函数为

$$f(x) = \frac{1}{\sigma}\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}-1} \cdot \exp\left\{\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\}.$$

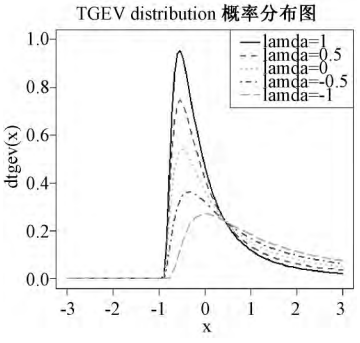
注意到形状参数 ξ 决定了分布的尾部形状,当 $\xi > 0$ 时,分布的尾部较长; 当 $\xi = 0$ 时,分布的尾部呈指数状; 当 $\xi < 0$ 时,分布具有有限的上端点.

$\xi = 0$ 是 GEV 分布的特殊情况,这时的分布称作 Gumbel 分布且此时 $x \in R$,同时也注意到 $\xi > 0$ 和 $\xi < 0$ 时 GEV 分布分别趋向于 Frechet 分布和 Weibull 分布.

应用 QRTM 中的(1.2.1),(1.2.2)两式可得到变形的广义极值分布(简记为 TGEV)的累积分布函数和概率密度函数

$$F(x) = (1+\lambda)\exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\} - \lambda\exp\left\{-2\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\} \tag{1.3.1}$$

$$f(x) = \frac{1}{\sigma}\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}-1} \cdot \exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\} \times \left[(1+\lambda) - 2\lambda\exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\}\right]. \tag{1.3.2}$$



上图是 TGEV 分布的概率密度图,其中 $\mu = 0$, $\sigma = 1$, $\xi = 1$ 和 $\lambda = -1, -0.5, 0, 0.5, 1$.

(C)1994-2019 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

3 随机数的生成和参数估计

3.1 随机数的生成

设随机变量 X 服从变形广义极值分布,其累积分布函数在前面已给出,即(1.2.1)式,在这里我们使用逆变换的方法可以生成变形广义极值分布的随机数如下:

$$F(x) = (1+\lambda)\exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\} - \lambda\exp\left\{-2\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\} = u,\ u \sim U(0,1),$$

则 $\xi \neq 0, \lambda \neq 0$ 时,是 TGEV 分布,

$$x = \mu + \frac{\sigma}{\xi}\left[\left(-\ln\frac{1+\lambda-\sqrt{(1+\lambda)^2-4\lambda u}}{2\lambda}\right)^{-\xi} - 1\right]. \quad (1.4.1)$$

$\xi = 0, \lambda \neq 0$ 时,是 TGumbel 分布,

$$x = \mu - \sigma\ln\left[-\ln\left(\frac{1+\lambda-\sqrt{(1+\lambda)^2-4\lambda u}}{2\lambda}\right)\right], \quad (1.4.2)$$

$\xi \neq 0, \lambda = 0$ 时,是 GEV 分布,

$$x = \mu + \frac{\sigma}{\xi}\left[(-\ln u)^{-\xi} - 1\right],$$

$\xi = 0, \lambda = 0$ 时,是 Gumbel 分布,

$$x = \mu - \sigma\ln\left[-\ln u\right].$$

当参数 $\mu, \sigma, \xi, \lambda$ 都已知的时候,上面(1.4.1)式就可以生成 TGEV 分布的随机数.

3.2 极大似然估计

极大似然估计方法是统计中最重要的方法之一,其基本思想就不在这赘述了.

设 $X_1, X_2, \cdots, X_n$ 为来自 TGEV 分布的随机样本,下面给出随机样本 $(X_1, X_2, \cdots, X_n)$ 的似然函数:

$$L = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n}\left\{\frac{1}{\sigma}\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}-1} \cdot \exp\left\{-\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\} \times \left[(1+\lambda) - 2\lambda\exp\left\{-\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\}\right]\right\}.$$

其对数似然函数为

$$\ln L = \sum_{i=1}^{n}\ln f(x_i) = -n\ln\sigma - \left(\frac{1}{\xi}+1\right)\sum_{i=1}^{n}\ln\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right] - \sum_{i=1}^{n}\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}} + \sum_{i=1}^{n}\ln\left[(1+\lambda) - 2\lambda\exp\left\{-\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\}\right].$$

然后将其分别对 $\mu, \sigma, \xi, \lambda$ 求偏导,分别令 $\frac{\partial\ln L}{\partial\mu} = 0$, $\frac{\partial\ln L}{\partial\sigma} = 0, \frac{\partial\ln L}{\partial\xi} = 0, \frac{\partial\ln L}{\partial\lambda} = 0$ 得到含四个参数的四个方程,解此方程组便可得到 $\hat{\mu}, \hat{\sigma}, \hat{\xi}, \hat{\lambda}$. 此方程组解起来比较复杂,我们可以采用数值方法求解,常用的有 Newton Raphson's 方法,也可以用 R 语言中的 optim 函数(文献[6])来解.

3.3 矩估计

用矩法求估计是由 K. Pearson 在 20 世纪初提出,其基本思想是用样本矩及其函数估计相应的总体矩及其函数,具体做法步骤就不详细介绍了.

设 $X_1, X_2, \cdots, X_n$ 为来自 TGEV 分布的随机样本,下面给出随机样本 $(X_1, X_2, \cdots, X_n)$ 的 n 阶矩:

$$E[X^n] = \int\frac{x^n}{\sigma}\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}-1} \cdot \exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\} \times \left[1+\lambda-2\lambda\exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\}\right]dx$$

其中 $1+\xi\left(\frac{x-\mu}{\sigma}\right) > 0$, $-\infty < \mu < \infty$, $\sigma > 0$, $-\infty < \xi < \infty$, $|\lambda| \leqslant 1$.

令 $y = \left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}$,则 $x = \frac{\sigma}{\xi}(y^{-\xi}-1)+\mu$, $dx = -\sigma y^{-\xi-1}dy$. 由于 $1+\xi\left(\frac{x-\mu}{\sigma}\right) > 0$,当 $\xi > 0$ 时, $-\frac{\sigma}{\xi}+\mu < x < \infty$,由 $x \to -\frac{\sigma}{\xi}+\mu$ 可得 $y \to \infty$,而 $x \to \infty$ 时 $y \to 0$;当 $\xi < 0$ 时, $-\infty < x < -\frac{\sigma}{\xi}+\mu$,由 $x \to -\frac{\sigma}{\xi}+\mu$ 可得 $y \to 0$,而 $x \to -\infty$ 时 $y \to \infty$. 故 $\xi \neq 0$ 时,对 y 积分,积分下限是 ∞,积分上限是 0. 因此,就有

(C)1994-2019 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

$$E[X^n] = \int_0^{\infty}\left[\mu+\frac{\sigma}{\xi}(y^{-\xi}-1)\right]^n \times \exp(-y) \times [1+\lambda-2\lambda\exp(-y)]\,dy$$

$$= \int_0^{\infty}\left[\left(\mu-\frac{\sigma}{\xi}\right)+\frac{\sigma}{\xi}y^{-\xi}\right]^n \times \exp(-y) \times [1+\lambda-2\lambda\exp(-y)]\,dy$$

$$= \sum_{k=0}^{n}\binom{n}{k}\left(\mu-\frac{\sigma}{\xi}\right)^{n-k}\left(\frac{\sigma}{\xi}\right)^k\left\{\int_0^{\infty}(1+\lambda)(y^{-\xi})^k \cdot \exp(-y)\,dy-\int_0^{\infty}2\lambda(y^{-\xi})^k\exp(-2y)\,dy\right\}$$

$$= \sum_{k=0}^{n}\binom{n}{k}\left(\mu-\frac{\sigma}{\xi}\right)^{n-k}\left(\frac{\sigma}{\xi}\right)^k\Gamma(1-\xi k)\cdot(1+\lambda-2^{\xi k}\lambda),$$

其中 $\int_0^{\infty}x^{v-1}\exp(-\mu x)(\ln x)^n dx = \frac{\partial^n}{\partial v^n}\{\mu^{-v}\Gamma(v)\}$, $n=0,1,2,3\cdots$, 此式来自文献[7].

故得到随机变量 X 的 n 阶矩 $E(X^n) = \sum_{k=0}^{n}\binom{n}{k}\left(\mu-\frac{\sigma}{\xi}\right)^{n-k}\left(\frac{\sigma}{\xi}\right)^k\Gamma(1-\xi k)(1+\lambda-2^{\xi k}\lambda)$. 由此可求出 X 的一阶矩、二阶矩、三阶矩、四阶矩, 根据矩方法的思想, 我们分别用样本矩代替中心矩, 根据方程组便可得到参数的矩估计 $\hat{\mu},\hat{\sigma},\hat{\xi},\hat{\lambda}$. 由于此方程组解起来比较复杂, 在这就不写出各参数估计的具体表达形式.

4　变形广义极值分布的参数估计的模拟结果

前面我们用两种方法对 TGEV 的参数进行了估计, 下面用 R 语言中的 optim 函数对 TGEV 分布的参数进行最大似然估计.

首先, 对参数估计进行模拟, 假设 $\mu=0.5$, $\sigma=0.5$, $\xi=0.5$, $\lambda=-0.1$, 则 $x=0.5+\left[\left(-\log\frac{0.9-\sqrt{0.9^2+0.4\mu}}{-0.2}\right)^{-0.5}-1\right]$, 由逆变换法生成样本 $(X_1,X_2,\cdots,X_n)$, 对样本容量 $n=50$, 100, 500, 1000, 用 R 语言编程可得 $\hat{\mu},\hat{\sigma},\hat{\xi},\hat{\lambda}$, 且通常我们通过计算参数估计的偏差(Bias)和均方误差(MSE)来评价估计得有效性. 具体模拟结果如表 1.

表 1　参数估计结果以及估计的偏差与均方误差

	Size	μ	σ	ξ	λ
MLE 估计	100	0.514 452 0	0.514 959 4	0.590 577 5	0.345 974 2
	500	0.475 815 8	0.455 790 2	0.520 890 4	−0.115 591 2
	1 000	0.414 979 3	0.435 969 3	0.526 825 5	−0.331 860 0
Bias	50	0.085 968 96	0.106 104 0	0.074 408 38	0.181 783 8
	100	0.084 460 1	0.105 452 9	0.067 165 47	0.199 541 7
	500	0.090 028 2	0.112 933 8	0.060 148 44	0.214 917 0
	1 000	0.101 742 8	0.181 737 8	0.056 880 63	0.205 062 8
MSE	50	0.007 390 663	0.011 258 05	0.005 536 606	0.033 045 34
	100	0.007 133 51	0.0111 203 1	0.004 511 2	0.039 816 89
	500	0.008 105 075	0.012 754 05	0.003 617 834	0.046 189 33
	1 000	0.010 351 6	0.033 028 63	0.003 235 406	0.042 050 76

根据数据结果, MLE 对 TGEV 参数估计的偏差和均方误差都很小, 是一种有效的估计方法, 而且对小样本也具有稳定性.

5　用变形广义极值分布进行实例分析

通过分析实际例子, 给出 TGEV 分布拟合的结果, 得到分位数及其各参数置信区间的估计, 并讨论所得结果的实际意义.

例: Pirie 港海平面年最大值.

图 1 是 1923—1987 年位于澳大利亚南部的 Pirie 港海平面年最大值(文献[8]). 假设澳大利亚南部的 Pirie 港每年的最大海平面高度值服从 TGEV 分布, 则用 R 语言编程可得到参数的极大似然估计为

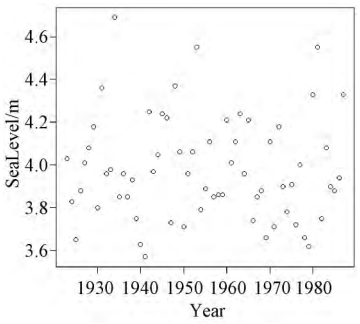


图 1　澳大利亚南部的 Pirie 港每年的最大海平面高度

$$(\hat{\mu},\hat{\sigma},\hat{\xi},\hat{\lambda}) = (3.865\ 259\ 32, 0.194\ 950\ 85, -0.042\ 739\ 73, -0.076\ 563\ 78),$$

(C)1994-2019 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

相应的对数似然函数为 4.342 217. 参数估计量的近似协方差矩阵为

$$V=\begin{bmatrix} 0.014\,607\,125 & 0.004\,506\,861 & -0.011\,872\,968 & 0.111\,908\,78 \\ 0.004\,506\,861 & 0.001\,758\,509 & -0.004\,214\,131 & 0.034\,937\,49 \\ -0.011\,872\,968 & -0.004\,214\,131 & 0.017\,989\,267 & -0.087\,614\,70 \\ 0.111\,908\,782 & 0.034\,937\,486 & -0.087\,614\,696 & 0.904\,570\,77 \end{bmatrix}$$

矩阵 V 对角元素分别是估计量 $\hat{\mu},\hat{\sigma},\hat{\xi},\hat{\lambda}$ 的方差,开方后得到 $\hat{\mu},\hat{\sigma},\hat{\xi},\hat{\lambda}$ 的标准误差,分别为 0.120 859 94,0.041 934 58,0.134 124 07,0.951 089 26. 由估计值和标准误差可知 μ,σ,ξ,λ 置信度近似为 95% 的置信区间分别为 [3.628 374,4.102 145], [0.112 759 1, 0.277 142 6], [-0.305 622 9, 0.220 143 4], [-1.940 699, 1.787 571]. 虽然 ξ 的极大似然估计是负值,但 ξ 的置信区间却有一部分大于 0,所以认为这些数据服从有界分布的证据不足. 由轮廓似然函数可以得到更高精度的置信区间. 图 2 给出了 ξ 的轮廓对数似然函数曲线,由此得到 ξ 的 95% 的置信区间 [-0.271 8,0.42 3] 与前面得到的置信区间稍有差别.

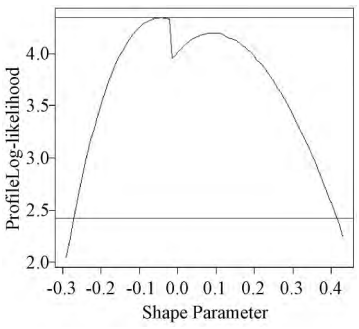


图 2 Pirie 港海平面形状参数 ξ 的轮廓似然值

将参数估计值代入式(2.1.1)以及应用 delta 方法,可得重现水平的估计值和置信区间. 例如: 估计 10 年期的重现水平,就是 $u=1-\frac{1}{10}=0.9$ 分位数,得到 $\hat{x}_{0.9}=4.296\,022$, $\mathrm{Var}(\hat{x}_{0.9})=\nabla x_u{}^T V \nabla x_u = 0.003\,061\,22$,其中 ∇x_u 为 $\left(\frac{\partial x_u}{\partial\mu},\frac{\partial x_u}{\partial\sigma},\frac{\partial x_u}{\partial\xi},\frac{\partial x_u}{\partial\lambda}\right)$ 在 $(\hat{\mu},\hat{\sigma},\hat{\xi},\hat{\lambda})$ 处的值. 且

$$\frac{\partial x_u}{\partial\mu}=1,$$

$$\frac{\partial x_u}{\partial\sigma}=\frac{1}{\xi}\left[\left(-\ln\frac{1+\lambda-\sqrt{(1+\lambda)^2-4\lambda u}}{2\lambda}\right)^{-\xi}-1\right],$$

$$\frac{\partial x_u}{\partial\xi}=-\frac{\sigma}{\xi^2}\left[\left(-\ln\frac{1+\lambda-\sqrt{(1+\lambda)^2-4\lambda u}}{2\lambda}\right)^{-\xi}-1\right]-\frac{\sigma}{\xi}\left(-\ln\frac{1+\lambda-\sqrt{(1+\lambda)^2-4\lambda u}}{2\lambda}\right)^{-\xi}\cdot\ln\left(-\ln\frac{1+\lambda-\sqrt{(1+\lambda)^2-4\lambda u}}{2\lambda}\right),$$

$$\frac{\partial x_u}{\partial\lambda}=-\frac{\sigma}{\lambda}\left(-\ln\frac{1+\lambda-\sqrt{(1+\lambda)^2-4\lambda u}}{2\lambda}\right)^{-\xi-1}\cdot\left[\frac{1-\frac{(1+\lambda)-2u}{\sqrt{(1+\lambda)^2-4\lambda u}}+\lambda\left(1-\frac{(1+\lambda)-2u}{\sqrt{(1+\lambda)^2-4\lambda u}}\right)}{1+\lambda-\sqrt{(1+\lambda)^2-4\lambda u}}\right].$$

得到 $\left(\frac{\partial x_u}{\partial\mu},\frac{\partial x_u}{\partial\sigma},\frac{\partial x_u}{\partial\xi},\frac{\partial x_u}{\partial\lambda}\right)$ 在 $(\hat{\mu},\hat{\sigma},\hat{\xi},\hat{\lambda})$ 处的值为 (1,2.209 594,0.491 636 3,-0.158 297 8). 因此,95% 的置信区间为 $\hat{x}_{0.9}\pm1.96\times\sqrt{Var(\hat{x}_{0.9})}=4.296\,022\pm1.96\times\sqrt{0.003\,061\,22}=$ [4.187 579, 4.404 465]. 相应地,100 年重现水平的估计值为 $\hat{x}_{0.99}=4.691\,155$, $\mathrm{Var}(\hat{x}_{0.99})=\nabla x_u{}^T V\nabla x_u=0.027\,302\,06$,其中 ∇x_u 为 $\left(\frac{\partial x_u}{\partial\mu},\frac{\partial x_u}{\partial\sigma},\frac{\partial x_u}{\partial\xi},\frac{\partial x_u}{\partial\lambda}\right)$ 在 $(\hat{\mu},\hat{\sigma},\hat{\xi},\hat{\lambda})$ 处的值为 (1,4.236 429,1.865 751,-0.147 803 2). 因此,95% 的置信区间为 $\hat{x}_{0.99}\pm1.96\times\sqrt{\mathrm{Var}(\hat{x}_{0.99})}=4.691\,155\pm1.96\times\sqrt{0.027\,302\,06}=$ [4.367 298,5.015 012]. 对比实际数据,在 65 年的实际观测中,有 4 次超过了区间的下界 4.367 298,这表明 100 年重现水平不会那样小.

由轮廓似然估计可得到更精确的重现水平估计,图 3 和图 4 表示 10 年与 100 年期重现水平的轮廓似然函数曲线. 由文献[5]中定理 3.5 以及轮廓似然图像可知 $x_{0.9}$ 和 $x_{0.99}$ 的置信区间约为 [4.20, 4.45] 和 [4.51,5.35]. 前者与 delta 方法得到的区间相似,后者则相差较大,其原因是轮廓似然函数的不对称性随着重现期的增加而赠加,也就是说不对称的轮廓似然曲线表明重现水平取大值时的不定性大于取小值时的不定性,重现期越长,数据提供的信息就越少,出现不对称性也是意料之中的. 也再次说明了 delta 方法给出的置信区间并不是保守的,当需要考虑不定性的更精确度量时,应该采用轮廓似然函数方法.

(C)1994-2019 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

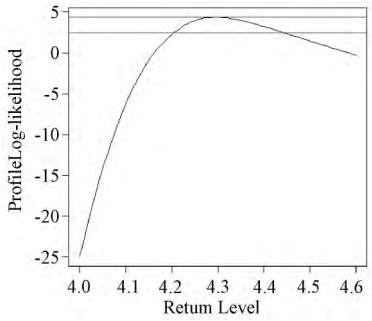


图 3　Pirie 港海平面 10 年期重现水平轮廓似然函数

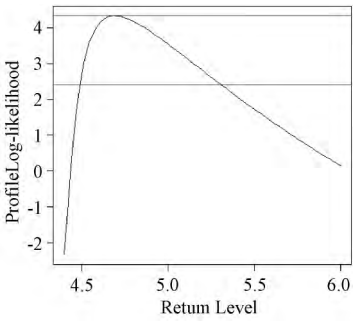


图 4　Pirie 港海平面 100 年期重现水平轮廓似然函数

图 5 是用 TGEV 模型拟合 Pirie 港海平面数据的诊断图. 由 P-P 图和 Q-Q 图都不能否定所拟合的模型,因为所有点都几乎在一条直线上. 由于 ξ 的估计值是负的,因此相应分布的支撑就有有限上界,故重现水平曲线应渐近地趋于某个有限值; 另一方面由于 ξ 的估计值接近于零,估计得重现水平曲线还应近似为线性. 所以重现水平曲线也给出了对 TGEV 模型满意的表示. 最后密度曲线的估计与直方图也是吻合的. 因此,四个诊断图都支持拟合的 TGEV 模型.

6　结论与展望

在本研究中,我们给出了变形广义极值分布的概率分布分析框架,并且建立了基于均匀分布的

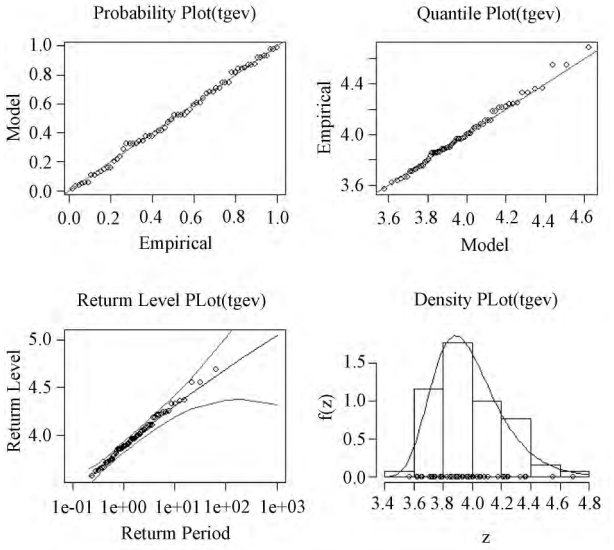


图 5　Pirie 港海平面数据 TGEV 拟合诊断图

Azzalini's formulation 和 QRTM formulation 之间的关系,之后得到了基本的分析表达式,同时还给出了最大似然估计和矩估计方法对参数估计表达式. 并用 R 语言中的 optim 函数对 TGEV 分布的参数做了 MLE 估计并对此估计方法的偏差和均方误差做了模拟,说明对于 TGEV 分布来说,MLE 估计方法效果很好. 为了说明变形广义极值分布的实用性和有效性,我们采用实际的数据对变形广义极值分布进行实例分析,给出了 TGEV 分布拟合的结果,得到分位数及其置信区间的估计,并讨论了所得结果的实际意义.

本文只是对参数做了 MLE 估计并对此估计方法的偏差和均方误差做了计算机随机模拟,也可以对参数做矩估计、PWM 估计和 Pickans 估计,然后对不同估计方法的偏差和均方误差做模拟,进行比较,这个想法我实践过,但是最终因为程序出了问题没找到原因,无法得到模拟结果而放弃,希望之后的学者们能在这方面多做考虑,使之完善.

参　考　文　献

[1] Genton M G. Skew-Elliptical Distributions and their Applications-A Journey Beyond Normality [J]. Champman Hall/CRC, 2004.

[2] Edgeworth F Y. XLII. The law of error and the elimination of chance [J]. Philosophical Magazine, 1886,21(131):308-324.

[3] Azzalini A. References on the skew-normal distribution and related ones [OL]. (2008-07-22). http://tango.stat.unipd.it/SN/list-publ.pdf.

[4] Aryal G, Tsokos C P. An application of skew Laplace distribution in finance [J]. Journal of Statistical Theory and

(C)1994-2019 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

Applications,2007,6:45 -60.
[5] Shaw W T, Buckley Ian R C. The alchemy of probability distributions: beyond Gram-Charlier expansions, and a skew-kurtotic-normal distrition from a rank transmutation map [OL]. (2009-01-05). http://arxiv.org/pdf/0901.0434v1.pdf.
[6] 汤银才. R 语言与统计分析 [M]. 北京: 高等教育出版社,2008.
[7] Coles S G. An introduction to statistical modelling of extreme values [M]. London: Springer,2001.
[8] 茆诗松,王静龙,濮晓龙. 高等数理统计 [M]. 北京: 高等教育出版社,2006.
[9] 史道济. 实用极值统计分析 [M]. 天津: 天津科学技术出版社,2006.
[10] Stuart Coles. 极值统计建模导论 [M]. 北京: 世界图书出版公司北京公司,2008.

Parameter Estimation and Case Analysis of Generalized Extreme Value Distribution

Fan Lili

(Department of Mathematics, Lyuliang University, Lyuliang Shanxi 033000)

Abstract

Based on the generalized extreme value (GEV) distribution, using the quadratic Rank Transmutation Map, we derive the probability distribution function and probability density function of the transmuted generalized extreme value (TGEV) distribution, and provide MLE and the moment estimator from TGEV distribution. The optim function of R language is used to do MLE of TGEV distribution and to simulate for bias and mean square error of MLE methods. The MLE method is better for the TGEV distribution. Finally, the transmuted generalized extreme valus distribution is applied in a case study, giving the TGEV distribution fitting results, obtaining estimation of quantile and confidence intervals, and discussing the practical significance of the results.

Key words: Quadratic rank transmutation map, transmuted generalized extreme value distribution, parameter estimation, nonparametric estimation.

(上接第 28 页)

Theoretical Calculations on Oscillator Strengths of $1s^2nd-1s^2n'f$ ($n\leqslant 9$) Dipole Transitions for Ni^{25+} Ion

Wang Yi Yao Simeng Yu Shuang Zhang Jingyi Hu Muhong Liu Lijuan

(School of Physics and Electronic Technology, Liaoning Normal University, Dalian Liaoning 116029)

Abstract

The wave function of $1s^2nl$ ($n\leqslant 9$, $l=d,f$) states for highly charged Ni^{25+} ion is constructed with multi-configuration interaction method. The oscillator strengths in length, velocity and acceleration expressions for $1s^2nd-1s^2n'f$ ($n,n'\leqslant 9$) dipole transitions are calculated, the agreement among them is satisfying. Combining the oscillator strength with the single channel quantum defect theory, the discreet oscillator strengths and oscillator strength density corresponding to the bounding-free transition from a certain initial state to all the final states are obtained.

Key words: highly ionized state, wavefunction, oscillator strengths, quantum defect theory.

(C)1994-2019 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net