

广义 Pareto 分布的广义有偏 概率加权矩估计方法*

赵旭 程维虎 李婧兰

(北京工业大学应用数理学院, 北京 100124)

(E-mail: barbiezx@sina.com)

摘要 广义 Pareto 分布 (GPD) 是统计分析中一个极为重要的分布, 被广泛应用于金融、保险、水文及气象等领域. 传统的参数估计方法如极大似然估计、矩估计及概率加权矩估计方法等已被广泛应用, 但使用中存在一定的局限性. 虽然提出很多改进方法如广义概率加权矩估计、L 矩和 LH 矩法等, 但都是研究完全样本的估计问题, 而在水文及气象等应用领域常出现截尾样本. 本文基于概率加权矩理论, 利用截尾样本对三参数 GPD 提出一种应用范围广且简单易行的参数估计方法, 可有效减弱异常值的影响. 首先求解出具有较高精度的形状参数的参数估计, 其次得出位置参数及尺度参数的参数估计. 通过 Monte Carlo 模拟说明该方法估计精度较高.

关键词 广义 Pareto 分布; 广义有偏概率加权矩估计; 次序统计量; 截尾样本;
概率加权矩估计

MR(2000) 主题分类 62F12

中图分类 O212.1

1 简介

广义 Pareto 分布 (Generalized Pareto Distribution, 简称 GPD) 是经典 Pareto 分布 (Pareto Distribution, 简称 PD) 的一种推广, 由 Pickands 于 1975 年首次引入^[1], 随后很多学者做了进一步的研究, 在诸多领域发挥重要作用. 在环境科学领域, 如对海浪、风力、气温、降雨等自然现象的研究很多都采用了 GPD. 在金融股票领域, 股票交易量、股票收益率均服从 GPD. 在金融市场的风险管理和经营领域, 投资者经常面临风险度量问题, 而 Value at Risk (简称 VaR) 是最重要并被广泛接受的风险度量工具之一. 计

本文 2010 年 10 月 17 日收到, 2011 年 8 月 1 日收到修改稿.

* 北京市自然科学基金 (KM201010005006) 资助项目.

算 VaR 的前提是估计 GPD 的参数, 采用适当的统计方法估计参数才能得到比较准确的 VaR 值, 从而进行风险评估. 寻找 GPD 参数估计简单可行并且具有较高精度的方法就显得尤为重要.

目前, 国内外学者对 GPD 参数估计做了许多研究. 传统参数估计方法如极大似然估计 (Maximum Likelihood Estimation, 简称 MLE) 方法^[2]、矩法 (Method of Moments, 简称 MOM)^[3]、概率加权矩 (Probability Weighted Moments, 简称 PWM) 方法^[4] 已经得到广泛应用. 虽然这些估计方法各有优点, 但也存在一定的局限性. MLE 依赖于大样本性质, 算法不收敛时可能得不到极大似然估计. MOM 和 PWM 存在时则要求形状参数满足一定条件. 作为 PWM 的线性组合, Hosking^[5] 提出了 L 矩 (L-moments) 方法, 广泛应用于水文气象等领域. 随后, Wang^[6] 进一步研究 LH 矩 (Higher Order L-moments), 可看作是 L 矩的广义形式. 另外, 作为 PWM 的推广形式, Haktanir^[7] 提出了 self-determined 概率加权矩 (Self-Determined Probability Weighted Moments, 简称 SD-PWM). 随后, Whalen^[8] 重新对 SD-PWM 方法进行研究, 并对算法进行改进, 使其更加便于应用. Rasmussen^[9] 研究广义概率加权矩 (Generalized Probability Weighted Moments, 简称 GPWM) 方法, 放宽了估计存在时形状参数的限制范围.

然而, 以上研究都是针对完全样本的, 而在水文及气象等应用领域常出现截尾样本, 截尾样本的估计方法与完全样本不同. Wang^[10,11] 对传统的 PWM 进行改进, 提出有偏概率加权矩 (Partial Probability Weighted Moments, 简称 PPWM). Bhattarai^[12] 应用洪水截尾样本考虑有偏 L 矩 (Partial L-moments), 但这些方法都是利用截尾样本研究广义极值分布的参数估计. 另外, 传统的 PWM 也不具备稳健性^[13], 在极值研究领域, 稳健性是很重要的一个问题. 实际应用中, 极大或极小的观测值通常是不精确地测量或截尾样本所造成, 这些观测值则会影响估计的稳健性. 本文的目的是利用截尾样本, 提出一种可有效减弱异常值影响的 GPD 参数估计方法. 我们考虑广义有偏概率加权矩法 (Generalized Partial Probability Weighted Moments, 简称 GPPWM), 该有偏方法对于可以很好拟合尾部数据的 GPD 而言, 应用方便且具有较高的估计精度.

文章第 2 节介绍 GPD, 第 3 节给出广义有偏概率加权矩估计的表达式. 第 4, 5 节进行 Monte Carlo 模拟及给出总结.

2 统计建模

设 X 为随机变量, 三参数 GPD 定义如下:

$$F(x|\gamma, \mu, \sigma) = \begin{cases} 1 - \left(1 + \gamma \frac{x - \mu}{\sigma}\right)^{-1/\gamma}, & \gamma \neq 0, \\ 1 - \exp\left(-\frac{x - \mu}{\sigma}\right), & \gamma = 0, \end{cases} \quad (1)$$

这里 $x > \mu$, $\sigma > 0$ 且 $1 + \gamma \frac{x - \mu}{\sigma} > 0$. 当 $\gamma \geq 0$ 时, $\mu \leq x < \infty$; 当 $\gamma < 0$ 时, $\mu < x < \mu - \sigma/\gamma$, 其中 μ 为位置参数, σ 为尺度参数, γ 为形状参数. 当 $\gamma > 0$ 时, GPD 是厚尾的, 通常, 称 $1/\gamma$ 为尾指数, GPD 的尾部形状随形状参数 γ 的变化而变化. 特别地, 当 $\gamma > 0$ 时, $F(x|\gamma, \mu, \sigma)$ 是重新参数化的普通 Pareto 分布, 其形状参数

为 $1/\gamma$, 即 GPD 是 PD 的广义形式; 当 $\gamma = 0$ 时, 对应的是指数分布; $\gamma = -1$, $\mu = 0$ 时, 对应 $(0, \sigma)$ 上的均匀分布.

其分布函数对应的分位数为:

$$x(F) = \begin{cases} \mu - \frac{\sigma}{\gamma}[1 - (1 - F)^{-\gamma}], & \gamma \neq 0, \\ \mu + \sigma[-\ln(1 - F)], & \gamma = 0, \end{cases} \quad (2)$$

3 广义有偏概率加权矩估计

首先, PWM^[4] 定义如下:

$$M_{p,r,s} = \int_0^1 [x(F)]^p F^r (1 - F)^s dF,$$

其中, X 是服从分布函数为 F 的随机变量, p, r 和 s 是给定的正整数.

设 X 是服从 GPD 的随机变量, 令 $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$ 为 X 的次序统计量, 仿照 PWM 定义方法, GPPWM 表示为

$$\begin{aligned} M'_{1,0,s} &= \int_{F_0}^{1-F_1} x(F) \cdot (1 - F)^s dF = \sigma \int_{F_0}^{1-F_1} \frac{x(F) - \mu}{\sigma} (1 - F)^s dF + \mu \int_{F_0}^{1-F_1} (1 - F)^s dF \\ &= -\frac{\sigma}{\gamma} \int_{F_0}^{1-F_1} [1 - (1 - F)^{-\gamma}] (1 - F)^s dF + \mu \int_{F_0}^{1-F_1} (1 - F)^s dF \\ &= -\frac{\sigma}{\gamma} \int_{F_0}^{1-F_1} (1 - F)^s dF + \frac{\sigma}{\gamma} \int_{F_0}^{1-F_1} (1 - F)^{s-\gamma} dF + \mu \int_{F_0}^{1-F_1} (1 - F)^s dF \\ &= (\mu - \frac{\sigma}{\gamma}) \int_{F_0}^{1-F_1} (1 - F)^s dF + \frac{\sigma}{\gamma} \int_{F_0}^{1-F_1} (1 - F)^{s-\gamma} dF \\ &= \mu \frac{(1 - F_0)^{s+1} - F_1^{s+1}}{s+1} + \frac{\sigma}{\gamma} \left[\frac{(1 - F_0)^{s-\gamma+1} - F_1^{s-\gamma+1}}{s-\gamma+1} - \frac{(1 - F_0)^{s+1} - F_1^{s+1}}{s+1} \right], \end{aligned} \quad (3)$$

其中 $F_0 = F(u_0)$, $F_1 = F(u_1)$, u_0, u_1 为给定值, u_0 称为门限值, $x(F)$ 见公式 (2). 当 $s > -1$ 及 $\gamma < s + 1$ 时该矩存在. 若 s 为正整数且 $F_0 = 0$, $F_1 = 1$, 则 GPPWM 退化为 PWM, 若 $s > -1$, $F_0 = 0$, $F_1 = 1$, 则 GPPWM 退化为 GPWM, 即 GPPWM 可看作 PWM 和 GPWM 的广义形式. 类似于 PWM, GPPWM 的样本矩 a'_s 表示为^[14]:

$$a'_s = \frac{1}{n} \sum_{i=1}^n X'_{i:n} (1 - p_{i:n})^s, \quad (4)$$

其中, n 为样本容量, $p_{i:n} = i/(n+1)$ 且

$$X'_{i:n} = \begin{cases} 0, & X_{i:n} \leq u_0 \text{ 或 } X_{i:n} \geq u_1, \\ X_{i:n}, & u_0 < X_{i:n} < u_1, \end{cases}$$

其中 $X_{i:n}$ 为第 i 个次序统计量, F_0 可由 n_0/n 估计, n_0 是样本中观测值小于等于 u_0 的个数, F_1 可由 n_1/n 估计, n_1 是样本中观测值大于等于 u_1 的个数.

将 (4) 式代入 (3) 式, 令样本矩等于总体矩, 可得矩估计方程为

$$a'_s = \mu \frac{(1-F_0)^{s+1} - F_1^{s+1}}{s+1} + \frac{\sigma}{\gamma} \left[\frac{(1-F_0)^{s-\gamma+1} - F_1^{s-\gamma+1}}{s-\gamma+1} - \frac{(1-F_0)^{s+1} - F_1^{s+1}}{s+1} \right]. \quad (5)$$

分别取 $s = s_1, s_2, s_3 > -1$.

通过化简可得以下 (6)-(8) 式

$$c_1 = \mu + \frac{\sigma}{\gamma} \cdot c_1(\gamma), \quad (6)$$

$$c_1 - c_2 = \frac{\sigma}{\gamma} [c_1(\gamma) - c_2(\gamma)], \quad (7)$$

$$\frac{c_1 - c_2}{c_2 - c_3} = \frac{c_1(\gamma) - c_2(\gamma)}{c_2(\gamma) - c_3(\gamma)}, \quad (8)$$

其中

$$c_i = \frac{(s_i + 1)a'_{s_i}}{(1-F_0)^{s_i+1} - F_1^{s_i+1}}, \quad c_i(\gamma) = \frac{c_i[(1-F_0)^{s_i+1-\gamma} - F_1^{s_i+1-\gamma}]}{a'_{s_i}(s_i+1-\gamma)} - 1, \quad i = 1, 2, 3.$$

形状参数 γ 的估计由式 (8) 通过数值方法解出. 令 $\hat{\gamma}$ 为 γ 的估计, 则 σ 和 μ 的估计可表示为

$$\hat{\sigma} = \frac{(c_1 - c_2)\hat{\gamma}}{c_1(\hat{\gamma}) - c_2(\hat{\gamma})}, \quad \hat{\mu} = c_1 - \frac{\hat{\sigma}}{\hat{\gamma}}c_1(\hat{\gamma}). \quad (9)$$

4 模拟

参数估计的好坏取决于很多因素, 如: 样本容量, 样本分布的位置、尺度及形状参数, 算法的稳定性以及估计方法的精度等. 估计效果可由偏差 (Bias) 和均方误差 (mean square error, 简称 MSE) 来进行评价. 偏差反映预测值与真值之间的差异程度. 均方误差反映预测值与真值之间的离散程度, 值越小, 说明估计越准确. 偏差和均方误差公式如下:

$$\text{偏差} = E(\theta_{\text{est}} - \theta_{\text{true}}), \quad \text{均方误差} = E[(\theta_{\text{est}} - \theta_{\text{true}})^2], \quad (10)$$

其中 $\theta_{\text{est}}, \theta_{\text{true}}$ 分别为参数的估计值和真值. 本文中 θ 代表 μ, σ 和 γ .

为了评价参数估计的效果, 对于不同样本容量 n, F_0 及 s_1, s_2, s_3 , 我们进行 1000 次 Monte Carlo 模拟. 表 1 给出分别取 n 为 15, 25, 50, 100, $F_1 = 0$ 以及不同的 $F_0, \mu, \sigma, \gamma, s_1, s_2, s_3$ 组合下, 三参数 GPD 的估计偏差及均方误差. 模拟表明对于不同的形状参数 γ, s_1 通常是负的且在 $(-0.3, -0.1)$ 之间. 另外, 通常当形状参数 γ 为正时, 有 $s_2 \rightarrow s_3$, 计算时我们采用 $s_2 = s_3 + 0.01$. 具体结果见表 1.

表 1 三参数 GPD 的估计偏差及均方误差

样本容量	F_0	(s_1, s_2, s_3)	μ			σ			γ		
			真值	偏差	均方误差	真值	偏差	均方误差	真值	偏差	均方误差
15	0	(-0.25, 1.01, 1)	0.15	-0.0798	0.0089	0.4	0.1520	0.0603	0.2	-0.2110	0.1020
		(-0.25, 1.01, 1)	0.15	-0.0183	0.0063	0.4	0.0647	0.0442	0.2	-0.1289	0.0616
		(-0.25, 1.01, 1)	0.3	-0.0192	<u>0.0031</u>	0.2	0.0681	<u>0.0211</u>	0.4	-0.2226	0.0883
		(-0.15, -0.151, 1)	0.15	-0.0332	0.0054	0.4	0.0712	0.0304	-0.2	-0.0780	<u>0.0372</u>
25	0	(-0.25, 1.01, 1)	0.15	-0.0200	0.0018	0.4	0.0590	0.0242	0.2	-0.1128	0.0471
		(-0.25, 1.01, 1)	0.15	0.0743	0.0090	0.4	-0.0620	0.0222	0.2	0.0047	0.0257
		(-0.25, 1.01, 1)	0.3	-0.0145	<u>0.0015</u>	0.2	0.0558	<u>0.0118</u>	0.4	-0.1838	0.0662
		(-0.15, -0.151, 1)	0.15	-0.0110	0.0026	0.4	0.0247	0.0150	-0.2	-0.0050	<u>0.0208</u>
50	0	(-0.25, 1.01, 1)	0.15	-0.0110	<u>0.0009</u>	0.4	0.0346	0.0107	0.2	-0.0653	0.0260
		(-0.25, 1.01, 1)	0.15	-0.0175	0.0021	0.4	0.0464	0.0158	0.2	-0.0675	0.0266
		(-0.25, 1.01, 1)	0.3	-0.0289	0.0015	0.2	0.0575	0.0083	0.4	-0.1598	0.0514
	0.2	(-0.15, -0.151, 1)	0.15	-0.0090	0.0013	0.4	0.0020	<u>0.0071</u>	-0.2	0.0224	<u>0.0126</u>
		(-0.25, 2.01, 2)	0.15	-0.0305	0.0033	0.4	0.0616	0.0183	0.2	-0.0878	0.0329
		(-0.25, 2.01, 2)	0.3	-0.0401	0.0023	0.2	0.0701	0.0103	0.4	-0.1746	0.0628
100	0	(-0.25, 1.01, 1)	0.15	-0.0075	<u>0.0004</u>	0.4	0.0197	0.0058	0.2	-0.0439	0.0144
		(-0.25, 1.01, 1)	0.15	-0.0115	0.0010	0.4	0.0261	0.0068	0.2	-0.0460	0.0154
		(-0.25, 1.01, 1)	0.3	-0.0176	0.0006	0.2	0.0342	<u>0.0036</u>	0.4	-0.1113	0.0288
	0.2	(-0.15, -0.151, 1)	0.15	-0.0010	0.0007	0.4	-0.0132	0.0040	-0.2	0.0421	<u>0.0086</u>
		(-0.25, 2.01, 2)	0.15	-0.0179	0.0016	0.4	0.0347	0.0089	0.2	-0.0511	0.0170
		(-0.25, 2.01, 2)	0.3	-0.0213	0.0008	0.2	0.0369	0.0038	0.4	-0.1112	0.0302

从表 1 可以看出: (i) 对于形状参数 γ 和尺度参数 σ 来说, 对于不同的样本容量 n , 均有 $F_0 = 0.1$ 时参数估计精度最高, 且随着样本容量的增加, 均方误差逐渐减小. (ii) 对于位置参数 μ 来说, 当样本较小 ($n = 15$ 和 25) 时, 取 $F_0 = 0.1$ 估计效果最好; 当样本容量 $n = 50$ 和 100 时, 取 $F_0 = 0$ 估计效果最好. 且随着样本容量的增大, 估计精度逐渐提高. (iii) 随着 F_0 的增加, s_2 和 s_3 的最佳取值不断增加.

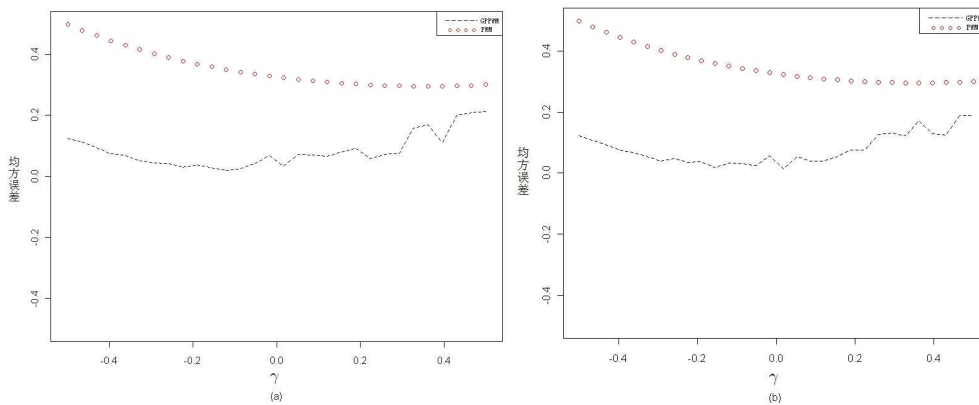


图 1 $n = 10$ 时 GPPWM 和 PWM 法 γ 的均方误差对比图

(a) $n = 10, F_0 = 0.$ (b) $n = 10, F_0 = 0.1.$

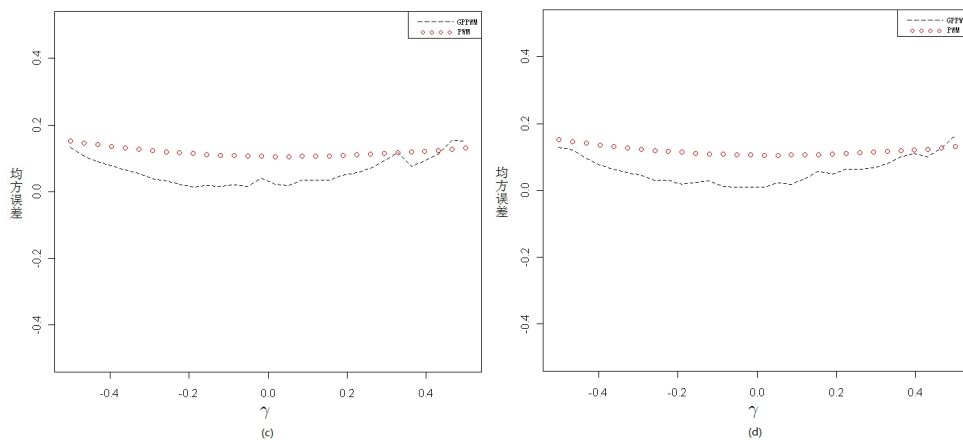


图 2 $n = 25$ 时 GPPWM 和 PWM 法 γ 的均方误差对比图

(c) $n = 25, F_0 = 0$. (d) $n = 25, F_0 = 0.1$.

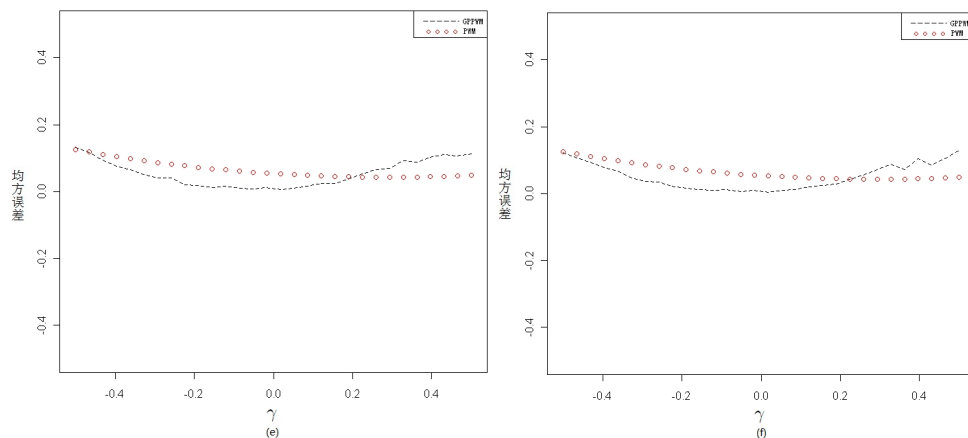


图 3 $n = 50$ 时 GPPWM 和 PWM 法 γ 的均方误差对比图

(e) $n = 50, F_0 = 0$. (f) $n = 50, F_0 = 0.1$.

对于不同的形状参数 γ , 图 1-3 给出 GPPWM 和 PWM 两种估计方法下 γ 的均方误差对比图. 具体方法是首先给定 s_1 和 s_2 , 它们的取值主要依赖于形状参数 γ , 通过模拟得出 γ 取不同值时对应的 s_1 和 s_2 的最佳组合. 而后分别取样本容量 $n = 10, 25, 50$, 不失一般性, 取 $\mu = 0, \sigma = 0.3, F_0 = 0, 0.1, F_1 = 0$. 模拟 γ 在 $(-0.5, 0.5)$ 内两种估计方法的均方误差.

从图 1-3 可以看出: (i) 对不同的样本容量 n , 当形状参数 $\gamma < 0.2$ 时, GPPWM 方法的均方误差小, 即本文方法对形状参数的估计精度高于 PWM 方法. (ii) 随着样本容量的增加, 特别是 $n = 50$ 时, $F_0 = 0.1$ 时的均方误差基本低于 $F_0 = 0$, 即 GPPWM 方法的估计效果好于 PWM 方法.

表 2 分别取 $n = 20, 50$ 和 100 , $\mu = 0$, $\sigma = 0.3$, $\gamma = 0.4$ 以及不同的 F_0, F_1 , 将本文方法与传统 PWM 和 GPWM 估计法进行比较. 通过 1000 次 Monte Carlo 模拟, 得出不同估计方法的偏差及均方误差, 具体结果见表 2.

表 2 不同估计方法的偏差及均方误差

样本容量	估计方法	γ		σ	
		偏差	均方误差	偏差	均方误差
20	GPPWM				
	$(F_0, F_1) = (0, 0.1)$	-0.532	0.315	-0.035	0.013
	$(F_0, F_1) = (0.05, 0)$	-0.022	<u>0.0005</u>	-0.130	0.020
	$(F_0, F_1) = (0.05, 0.1)$	-0.018	0.0015	-0.196	0.039
	GPWM	-0.282	0.159	0.068	0.024
	PWM	-0.103	0.086	0.030	0.016
50	GPPWM				
	$(F_0, F_1) = (0, 0.15)$	-0.579	0.378	-0.054	0.009
	$(F_0, F_1) = (0.05, 0)$	-0.022	0.00050	-0.096	0.011
	$(F_0, F_1) = (0.05, 0.1)$	-0.016	0.00025	-0.188	0.036
	$(F_0, F_1) = (0.05, 0.15)$	-0.008	<u>0.00011</u>	-0.210	0.044
	GPWM	-0.153	0.051	0.0304	0.007
PWM	-0.055	0.031	0.012	0.004	
100	GPPWM				
	$(F_0, F_1) = (0, 0.15)$	-0.642	0.456	-0.043	0.006
	$(F_0, F_1) = (0.05, 0)$	-0.022	0.00050	-0.075	0.007
	$(F_0, F_1) = (0.05, 0.1)$	-0.016	0.00025	-0.185	0.034
	$(F_0, F_1) = (0.05, 0.15)$	-0.007	<u>6.46e-05</u>	-0.209	0.044
	GPWM	-0.103	0.024	0.021	0.003
PWM	-0.030	0.017	0.007	0.002	

从表 2 可以看出: (i) 对形状参数 γ 来说, 当样本容量较小 ($n = 20$), 与传统的 PWM 和 GPWM 方法相比, 本文方法在 $(F_0, F_1) = (0.05, 0)$ 时估计精度最高; 当样本容量较大 ($n = 50, 100$), GPPWM 在 $(F_0, F_1) = (0.05, 0.15)$ 时估计均方误差最小, 即估计效果最好. (ii) 随着样本容量的增加, 本文方法对形状参数 γ 及尺度参数 σ 的估计精度大部分均有所提高. 对于大样本情形 ($n = 100$), 本文方法尤其对形状参数 γ 具有较高的估计精度.

5 结论

本文利用截尾样本研究三参数 GPD 的估计问题, 基于 PWM 估计思想, 提出 GP-PWM 估计方法, 是传统 PWM 和 GPWM 估计的推广和广义形式, 有效地减弱异常值的影响. 为检验本文估计方法的精确性, 我们将 GPPWM 估计所得的结果与 PWM 和 GPWM 估计结果进行比较, Monte Carlo 模拟给出不同参数估计方法的偏差和均方误差, 通过比较说明本文方法尤其对形状参数具有较高的估计精度.

参 考 文 献

- [1] Pickands J. Statistical Inference Using Extreme Order Statistics. *J. Ann. Stat.*, 1975, 3(1): 119–131
- [2] Grimshaw S D. Computing Maximum Likelihood Estimates for the Generalized Pareto Distribution. *J. Technometrics*, 1993, 35(2): 185–191
- [3] Hosking J R M, Wallis J R. Parameter and Quantile Estimation for the Generalized Pareto Distribution. *J. Technometrics*, 1987, 29(3): 339–349
- [4] Greenwood J A, Landwehr J M, Matalas N C, Wallis J R. Probability Weighted Moments: Definition and Relation to Parameters of Several Distributions Expressible in Inverse Form. *J. Water Resour. Res.*, 1979, 15(5): 1049–1054
- [5] Hosking J R M. L-moments: Analysis and Estimation of Distributions Using Linear Combinations of order Statistics. *J. R. Statist Soc. B*, 1990, 52(1): 105–124
- [6] Wang Q J. LH Moments for Statistical Analysis of Extreme Events. *J. Water Resour. Res.*, 1997, 33(12): 2841–2848
- [7] Haktanir T. Self-determined Probability-weighted Method and its Application to Various Distributions. *J. Hydrol.*, 1997, 194(1–4): 180–200
- [8] Whalen T M, Savage G T, Jeong G D. The Method of Self-determined Probability Weighted Moments Revisited. *J. Hydrol.*, 2002, 268(1–4): 177–191
- [9] Rasmussen P F. Generalized Probability Weighted Moments: Application to the Generalized Pareto Distribution. *J. Water Resour. Res.*, 2001, 37(6): 1745–1751
- [10] Wang Q J. Estimation of the GEV Distribution from Censored Samples by Method of Partial Probability Weighted Moments. *J. Hydrol.*, 1990, 120(1-4): 103–114
- [11] Wang Q J. Using Partial Probability Weighted Moments to Fit the Extreme Value Distributions to Censored Samples. *J. Water Resour. Res.*, 1996, 32(6): 1767–1771
- [12] Bhattarat K P. Partial L-moments for the Analysis of Censored Flood Samples. *J. Hydrological Sciences Journal des Sciences Hydrologiques*, 2004, 49(5): 855–868
- [13] Davison A C, Smith R L. Models for Exceedances over High Thresholds. *J. R. Statist Soc. B.*, 1990, 52(3): 393–442
- [14] Landwehr J M, Matalas N C, Wallis J R. Probability Weighted Moments Compared with Some Traditional Techniques for Estimating Gumbel Parameters and Quantiles. *J. Water Resour. Res.*, 1979, 15(5): 1055–1064

Parameter Estimation of Generalized Partial Probability Weighted Moments for the Generalized Pareto Distribution

ZHAO XU CHENG WEIHU LI JINGLAN

(College of Applied Science, Beijing University of Technology, Beijing 100124)

(E-mail: barbiezx@sina.com)

Abstract The generalized Pareto distribution (GPD) is one of the most important distribution in statistics analysis that has been widely used in finance, insurance, hydrology and meteorology applications and so on. While traditional estimation methods, such as maximum likelihood (ML), methods of moments (MOM) and probability weighted moments (PWM) methods have been extensively applied, the use of these methods are often restricted. Alternative approaches (e.g., generalized probability weighted moments, L-moments and LH-moments) exist but they use complete or non-censored samples. However, censored samples are often encountered in hydrology and meteorology fields. In this article, we propose a computationally easy method from censored data for fitting the GPD, which is resistant against extremely small or large outliers, i.e., they will be robust with the lower and upper breakdown points. This method is based on probability weighted moments. Firstly, we solve shape parameter estimator which has high estimated precision, then the location and scale parameters are given for the GPD. Simulation studies show that the proposed method performs well compared to traditional techniques.

Key words generalized Pareto distribution;
generalized partial probability weighted moments;
order statistics; censored samples; probability weighted moments

MR(2000) Subject Classification 62F12

Chinese Library Classification O212.1