第 48 卷 第 7 期 西南师范大学学报(自然科学版) 2023 年 7 月
Vol. 48 No. 7 Journal of Southwest China Normal University (Natural Science Edition) Jul. 2023

DOI:10.13718/j.cnki.xsxb.2023.07.010

极值指数的分布式矩率估计量①

罗心艺， 彭作祥

西南大学 数学与统计学院，重庆 400715

摘要：基于分治算法和矩率估计量构造了分布式矩率估计量并证明其相合性和渐近正态性.

关 键 词：分治算法；分布式矩率估计量；渐近正态性

中图分类号：**O211.3** **文献标志码**：**A** **文章编号**：1000-5471(2023)07-0067-06

Distributed Moment Ratio Estimator for Extreme Value Index

LUO Xinyi， PENG Zuoxiang

School of Mathematics and Statistics, Southwest University, Chongqing 400715, China

Abstract: Based on a divide-and-conquer algorithm and the moment ratio estimator, we proposed a distributed moment ratio estimator and prove its consistency and asymptotic normality.

Key words: divide-and-conquer algorithm; distributed moment ratio estimator; asymptotic normality

设$\{X_n, n \geqslant 1\}$为独立同分布的随机变量序列，其公共分布函数为$F(x)$. 若存在常数$a_n > 0$，$b_n \in \mathbb{R}$使得对所有$1+\gamma x > 0$，有

$$\lim_{n\to\infty} F^n(a_n x + b_n) = G_\gamma(x) \tag{1}$$

其中$G_\gamma(x) = \exp\{-(1+\gamma x)^{-\frac{1}{\gamma}}\}$，$\gamma \in \mathbb{R}$，则称$F$属于极值分布$G_\gamma$的吸引场，记为$F \in D(G_\gamma)$，$\gamma$为极值指数. 当$\gamma > 0$时，(1) 式等价于

$$\lim_{t\to\infty} \frac{U(tx)}{U(t)} = x^\gamma,\ x > 0 \tag{2}$$

其中$U(t) = \dfrac{1}{(1-F)^{\leftarrow}(t)}$.

当分布函数F未知时，对极值指数γ的估计是极值理论的一个重要组成部分，受到了学者的广泛关注，常用于金融、保险、自然灾害等领域. 在分布函数形式未知的情况下，文献[1]提出了著名的 Hill 估计量，推断分布函数的尾部表现；文献[2-4]在一定条件下证明了 Hill 估计量的相合性和渐近正态性；文献[5]提出了矩率估计量，并给出了其分布表示；文献[6]提出了一系列基于二阶参数的外部估计得到的渐近无偏

① 收稿日期：2022-09-05
作者简介：罗心艺，硕士研究生，主要从事极值统计分析研究.
通信作者：彭作祥，教授.

估计量，并证明了其渐近性质；文献[7]证明了包含 Hill 估计量和矩率估计量在内的一系列尾指数估计量的渐近正态性.关于尾指数估计量的更多研究，见文献[8-10].

在大数据时代，估计极值指数时，常常会遇到被分开存储的数据，例如分析来自不同保险公司的保险索赔时，为了保护客户的隐私，保险公司不能向外部分享具体的数据，甚至不能分享任何索赔结果，此时前文所提的 Hill 估计量和矩率估计量等都不可用.与大部分尾指数估计量的相关文献一样，Hill 估计量等只使用了一部分秩序较高的统计量.文献[11]和文献[12]基于块方法提出了 DPR 估计量.当数据被分组储存且每组只有少数几个最大的样本可用于分析时，DPR 方法是可行的，但是它仅使用了每块中最大的两个样本，很可能并不是尾指数的充分统计量.

而分治算法可以分析存储在多台计算机中的数据集，分别估计每台机器上的参数，并将结果发送到中央机器.中央机器通过简单的平均来结合所有结果，以获得计算可行的估计量.设独立同分布的样本 X_1，$X_2,\cdots,X_n$ 存储在 k 个机器中，每个机器存储 m 个观测，令 $M_j^{(1)}\geqslant M_j^{(2)}\geqslant\cdots\geqslant M_j^{(m)}$ 表示存储在机器 j 内的样本次序统计量，文献[13]基于 Hill 估计量提出如下分布式 Hill 估计量

$$\hat{\gamma}_{\mathrm{DH}}:=\frac{1}{k}\sum_{j=1}^{k}\frac{1}{d}\sum_{i=1}^{d}(\log M_j^{(i)}-\log M_j^{(d+1)}) \tag{3}$$

其中，d 为一个确定的整数，表示每个机器所使用的超过门限值的个数.$\hat{\gamma}_{\mathrm{DH}}$ 是极值指数的一个相合估计量，且在二阶正规变换条件下具有渐近正态性.

受文献[13]启发，本文基于矩率估计量提出如下分布式矩率估计量

$$\hat{\gamma}_{\mathrm{DMR}}:=\frac{1}{2k}\sum_{j=1}^{k}\frac{\sum_{i=1}^{d}(\log M_j^{(i)}-\log M_j^{(d+1)})^2}{\sum_{i=1}^{d}(\log M_j^{(i)}-\log M_j^{(d+1)})} \tag{4}$$

并研究其相合性和渐近正态性，中间序列 $k=k(n)$ 满足当 $n\to\infty$ 时，

$$k=k(n)\to\infty,\ m=m(n)\to\infty,\ \frac{m}{\log k}\to\infty \tag{5}$$

为了证明其渐近正态性，我们需要文献[14]中的二阶条件，即存在一个最终或正或负的函数 A，且 $\lim\limits_{t\to\infty}A(t)=0$，和一个实数 $\rho\leqslant 0$，使得

$$\lim_{t\to\infty}\frac{\dfrac{U(tx)}{U(t)}-x^{\gamma}}{A(t)}=x^{\gamma}\,\frac{x^{\rho}-1}{\rho},\ x>0 \tag{6}$$

1 相合性和渐近正态性

定理1 假设(2)式及(5)式成立.当 $n\to\infty$ 时，$\hat{\gamma}_{\mathrm{DMR}}\xrightarrow{P}\gamma$.

定理2 假设(5)式及(6)式成立.当 $n\to\infty$ 时，若 $(kd)^{\frac{1}{2}}A\left(\dfrac{m}{d}\right)=O(1)$ 成立，则

$$(kd)^{\frac{1}{2}}\left\{\hat{\gamma}_{\mathrm{DMR}}-\gamma-(1-\rho)^{-1}A\left(\frac{m}{d}\right)g(d,m,\rho)\right\}\xrightarrow{d}N(0,2\gamma^2)$$

其中

$$g(d,m,\rho)=\frac{1}{1-\rho}\left(\frac{m}{d}\right)^{-\rho}\frac{\Gamma(m+1)\Gamma(d-\rho+1)}{\Gamma(m-\rho+1)\Gamma(d+1)} \tag{7}$$

2 定理的证明

设独立同分布的样本 $X_1,X_2,\cdots,X_n$ 存储在 k 个机器中，每个机器存储 m 个观测，令 $M_j^{(1)}\geqslant M_j^{(2)}\geqslant\cdots\geqslant M_j^{(m)}$ 表示存储在机器 j 内的样本次序统计量.$Z_j^{(1)}\geqslant\cdots\geqslant Z_j^{(m)}$ 表示对应于 $M_j^{(1)}\geqslant\cdots\geqslant M_j^{(m)}$ 的服从

Pareto(1) 分布的次序统计量，则由文献[14] 知$\{M_j^{(i)}\}_{i=1}^m \overset{d}{=} \{U(Z_j^{(i)})\}_{i=1}^m$.

定理 1 的证明　由文献[14] 的定理 B.1.9 知，对 $x>1$ 和 $t\geqslant t_0$ 有，

$$(\gamma-\varepsilon)\log((1-\varepsilon)x) < \log U(tx)-\log U(t) < (\gamma+\varepsilon)\log((1+\varepsilon)x) \tag{8}$$

则

$$(\gamma-\varepsilon)^2(\log((1-\varepsilon)x))^2 < (\log U(tx)-\log U(t))^2 < (\gamma+\varepsilon)^2(\log((1+\varepsilon)x))^2 \tag{9}$$

令 $tx=Z_j^{(i)}$，$t=Z_j^{(d+1)}$，由(8) 式和(9) 式知

$$(\gamma-\varepsilon)\left(1+\frac{1}{d}\sum_{i=1}^d\left(\log((1-\varepsilon)\frac{Z_j^{(i)}}{Z_j^{(d+1)}})-1\right)\right)<$$

$$\frac{1}{d}\sum_{i=1}^d\log\frac{U(Z_j^{(i)})}{U(Z_j^{(d+1)})} < (\gamma+\varepsilon)\left(1+\frac{1}{d}\sum_{i=1}^d\left(\log((1+\varepsilon)\frac{Z_j^{(i)}}{Z_j^{(d+1)}})-1\right)\right) \tag{10}$$

$$2(\gamma-\varepsilon)^2\left(1+\frac{1}{2d}\sum_{i=1}^d\left((\log((1-\varepsilon)\frac{Z_j^{(i)}}{Z_j^{(d+1)}}))^2-2\right)\right)<$$

$$\frac{1}{d}\sum_{i=1}^d\left(\log\frac{U(Z_j^{(i)})}{U(Z_j^{(d+1)})}\right)^2 < 2(\gamma+\varepsilon)^2\left(1+\frac{1}{2d}\sum_{i=1}^d\left(\left(\log\left((1+\varepsilon)\frac{Z_j^{(i)}}{Z_j^{(d+1)}}\right)\right)^2-2\right)\right) \tag{11}$$

由(10) 式和(11) 式可得

$$\frac{1}{k}\sum_{j=1}^k\frac{(\gamma-\varepsilon)^2}{\gamma+\varepsilon}\left(1+\frac{1}{2d}\sum_{i=1}^d\left(\left(\log\left((1-\varepsilon)\frac{Z_j^{(i)}}{Z_j^{(d+1)}}\right)\right)^2-2\right)-\frac{1}{d}\sum_{i=1}^d\left(\log\left((1+\varepsilon)\frac{Z_j^{(i)}}{Z_j^{(d+1)}}\right)-1\right)\right)<$$

$$\frac{1}{k}\sum_{j=1}^k\frac{\frac{1}{d}\sum_{i=1}^d(\log M_j^{(i)}-\log M_j^{(d+1)})^2}{\frac{2}{d}\sum_{i=1}^d(\log M_j^{(i)}-\log M_j^{(d+1)})}<$$

$$\frac{1}{k}\sum_{j=1}^k\frac{(\gamma+\varepsilon)^2}{\gamma-\varepsilon}\left(1+\frac{1}{2d}\sum_{i=1}^d\left(\left(\log\left((1+\varepsilon)\frac{Z_j^{(i)}}{Z_j^{(d+1)}}\right)\right)^2-2\right)-\frac{1}{d}\sum_{i=1}^d\left(\log\left((1-\varepsilon)\frac{Z_j^{(i)}}{Z_j^{(d+1)}}\right)-1\right)\right) \tag{12}$$

由文献[15] 的引理 3.4 知

$$\frac{1}{k}\frac{1}{d}\sum_{j=1}^k\sum_{i=1}^d\log\frac{Z_j^{(i)}}{Z_j^{(d+1)}} \overset{d}{=} \frac{1}{k}\frac{1}{d}\sum_{j=1}^k\sum_{i=1}^d E_j(i)$$

$$\frac{1}{k}\frac{1}{d}\sum_{j=1}^k\sum_{i=1}^d\left(\log\frac{Z_j^{(i)}}{Z_j^{(d+1)}}\right)^2 \overset{d}{=} \frac{1}{k}\frac{1}{d}\sum_{j=1}^k\sum_{i=1}^d(E_j(i))^2$$

其中$\{E_j(i),\ i=1,\ \cdots,\ d\}$ 服从独立同分布的标准指数分布，$j=1,\cdots,k$. 因此

$$\frac{1}{kd}\sum_{j=1}^k\sum_{i=1}^d\left(\log\left((1+\varepsilon)\frac{Z_j^{(i)}}{Z_j^{(d+1)}}\right)\right)^2=(\log(1+\varepsilon))^2+\frac{1}{k}\frac{1}{d}\sum_{j=1}^k\sum_{i=1}^d\left(\log\frac{Z_j^{(i)}}{Z_j^{(d+1)}}\right)^2 \xrightarrow{\mathrm{P}} 2 \tag{13}$$

$$\frac{1}{kd}\sum_{j=1}^k\sum_{i=1}^d\log\left((1-\varepsilon)\frac{Z_j^{(i)}}{Z_j^{(d+1)}}\right)=\log(1-\varepsilon)+\frac{1}{k}\frac{1}{d}\sum_{j=1}^k\sum_{i=1}^d\log\frac{Z_j^{(i)}}{Z_j^{(d+1)}} \xrightarrow{\mathrm{P}} 1 \tag{14}$$

将(13) 式及(14) 式代入不等式(12) 即可得 $\hat{\gamma}_{\mathrm{DMR}} \xrightarrow{P} \gamma$，定理证毕.

对定理 2 的证明，我们需要下面这个辅助引理.

引理 1　令 $Z^{(1)}\geqslant\cdots\geqslant Z^{(m)}$ 表示服从 Pareto(1) 分布的独立随机变量$\{Z_1,\ \cdots,\ Z_m\}$ 的次序统计量，则对任意 $\rho\leqslant 0$，有

$$\frac{1}{1-\rho}\mathrm{E}\left(\left(\frac{dZ^{(d+1)}}{m}\right)^\rho\right)=g(d,\ m,\ \rho)$$

其中 $g(d,\ m,\ \rho)$ 的定义见(2) 式. 并且，当 $m\to\infty$ 时，

$$g(d,\ m,\ \rho)\to\frac{d^\rho}{1-\rho}\frac{\Gamma(d-\rho+1)}{\Gamma(d+1)}$$

证　见文献[13] 的引理 S.3.

定理 2 的证明 由文献[14]的定理 B.2.18 知,存在一个函数 $A_0(t)$,使得当 $t\to\infty$ 时,$A_0(t)\sim A(t)$,且对所有 $\varepsilon>0$,$\delta>0$,存在一个 $t_0>0$ 使得对 $tx\geqslant t_0$,$t\geqslant t_0$,

$$\left|\frac{\log U(tx)-\log U(t)-\gamma\log x}{A_0(t)}-\frac{x^{\rho}-1}{\rho}\right|\leqslant\varepsilon x^{\rho}\max\{x^{\delta},x^{-\delta}\} \tag{15}$$

定义 $\mathscr{I}_{n,d,t_0}=\{Z_j^{(d+1)}\geqslant t_0,\ 1\leqslant j\leqslant k\}$,由文献[13]的引理 S.2 知,对任意 $t_0>0$,$\lim\limits_{n\to\infty}\mathrm{P}(\mathscr{I}_{n,d,t_0})=1$. 所以,基于集合 $\mathscr{I}_{n,d,t_0}$,我们用$\frac{m}{d}$ 和 $Z_j^{(i)}$,$i=1,2,\cdots,d+1$ 替换不等式(15)中的 t 和 tx,得到

$$\left|\frac{\log U(Z_j^{(i)})-\log U\left(\frac{m}{d}\right)-\gamma\log\left(\frac{dZ_j^{(i)}}{m}\right)}{A_0\left(\frac{m}{d}\right)}-\frac{\left(\frac{dZ_j^{(i)}}{m}\right)^{\rho}-1}{\rho}\right|\leqslant \\ \varepsilon\left(\frac{dZ_j^{(i)}}{m}\right)^{\rho}\max\left\{\left(\frac{dZ_j^{(i)}}{m}\right)^{\delta},\left(\frac{dZ_j^{(i)}}{m}\right)^{-\delta}\right\} \tag{16}$$

对 i 和 $i=d+1$ 应用两次不等式(16),当 $n\to\infty$ 时,

$$\frac{\log U(Z_j^{(i)})-\log U(Z_j^{(d+1)})-\gamma(\log Z_j^{(i)}-\log Z_j^{(d+1)})}{A_0\left(\frac{m}{d}\right)}= \\ \frac{\left(\frac{dZ_j^{(i)}}{m}\right)^{\rho}-1}{\rho}-\frac{\left(\frac{dZ_j^{(d+1)}}{m}\right)^{\rho}-1}{\rho}+ \\ o_p(1)\left\{\max\left\{\left(\frac{dZ_j^{(i)}}{m}\right)^{\rho+\delta},\left(\frac{dZ_j^{(i)}}{m}\right)^{\rho-\delta}\right\}+\max\left\{\left(\frac{dZ_j^{(d+1)}}{m}\right)^{\rho+\delta},\left(\frac{dZ_j^{(d+1)}}{m}\right)^{\rho-\delta}\right\}\right\} \tag{17}$$

对 $\rho<0$,存在 $\delta>0$ 使得 $\rho+\delta<0$,应用不等式

$$\frac{\max\{x^{\rho+\delta},x^{\rho-\delta}\}}{\max\{y^{\rho+\delta},y^{\rho-\delta}\}}\leqslant\max\left\{\left(\frac{x}{y}\right)^{\rho+\delta},\left(\frac{x}{y}\right)^{\rho-\delta}\right\},\ x>0,\ y>0$$

可以得到

$$\begin{aligned}
&(kd)^{\frac{1}{2}}(\hat{\gamma}_{\mathrm{DMR}}-\gamma)= \\
&\gamma(kd)^{\frac{1}{2}}\left(\frac{1}{k}\sum_{j=1}^{k}\left(\frac{1}{2d}\sum_{i=1}^{d}\left(\left(\log\left(\frac{Z_j^{(i)}}{Z_j^{(d+1)}}\right)\right)^2-2\right)-\frac{1}{d}\sum_{i=1}^{d}\left(\log\left(\frac{Z_j^{(i)}}{Z_j^{(d+1)}}\right)-1\right)\right)\right)+ \\
&(kd)^{\frac{1}{2}}\frac{A_0\left(\frac{m}{d}\right)}{\rho}\frac{1}{k}\sum_{j=1}^{k}\frac{1}{d}\sum_{i=1}^{d}\left(\left(\log\left(\frac{Z_j^{(i)}}{Z_j^{(d+1)}}\right)\right)\left(\left(\frac{dZ_j^{(i)}}{m}\right)^{\rho}-\left(\frac{dZ_j^{(d+1)}}{m}\right)^{\rho}\right)\right)- \\
&(kd)^{\frac{1}{2}}\frac{A_0\left(\frac{m}{d}\right)}{\rho}\frac{1}{k}\sum_{j=1}^{k}\frac{1}{d}\sum_{i=1}^{d}\left(\left(\frac{dZ_j^{(i)}}{m}\right)^{\rho}-\left(\frac{dZ_j^{(d+1)}}{m}\right)^{\rho}\right)+ \\
&o_p(1)(kd)^{\frac{1}{2}}A_0\left(\frac{m}{d}\right)\frac{1}{k}\sum_{j=1}^{k}\left(\max\left\{\left(\frac{dZ_j^{(d+1)}}{m}\right)^{\rho+\delta},\left(\frac{dZ_j^{(d+1)}}{m}\right)^{\rho-\delta}\right\}\times\right. \\
&\left.\frac{1}{d}\sum_{i=1}^{d}\left(\log\left(\frac{Z_j^{(i)}}{Z_j^{(d+1)}}\right)\right)\left\{\max\left\{\left(\frac{Z_j^{(i)}}{Z_j^{(d+1)}}\right)^{\rho+\delta},\left(\frac{Z_j^{(i)}}{Z_j^{(d+1)}}\right)^{\rho-\delta}\right\}+1\right\}\right)+ \\
&o_p(1)(kd)^{\frac{1}{2}}A_0\left(\frac{m}{d}\right)\frac{1}{k}\sum_{j=1}^{k}\left(\max\left\{\left(\frac{dZ_j^{(d+1)}}{m}\right)^{\rho+\delta},\left(\frac{dZ_j^{(d+1)}}{m}\right)^{\rho-\delta}\right\}\times\right. \\
&\left.\frac{1}{d}\sum_{i=1}^{d}\left\{\max\left\{\left(\frac{Z_j^{(i)}}{Z_j^{(d+1)}}\right)^{\rho+\delta},\left(\frac{Z_j^{(i)}}{Z_j^{(d+1)}}\right)^{\rho-\delta}\right\}+1\right\}\right)=: \\
&I_1+I_2+I_3+I_4+I_5
\end{aligned}$$

与定理 1 的证明类似，我们可以得到当 $n \to \infty$ 时，$I_1 \xrightarrow{d} N(0, 2\gamma^2)$.

对于 I_2，记

$$I_2 = (kd)^{\frac{1}{2}} \frac{A_0\left(\frac{m}{d}\right)}{\rho} \mathrm{E}\left(\left(\frac{dZ_1^{(d+1)}}{m}\right)^{\rho}\right) \frac{1}{k}\sum_{j=1}^{k} \frac{\left(\frac{dZ_j^{(d+1)}}{m}\right)^{\rho}}{\mathrm{E}\left(\left(\frac{dZ_1^{(d+1)}}{m}\right)^{\rho}\right)} \frac{1}{d}\sum_{i=1}^{d}\left(\log\left(\frac{Z_j^{(i)}}{Z_j^{(d+1)}}\right)\right)\left(\left(\frac{Z_j^{(i)}}{Z_j^{(d+1)}}\right)^{\rho} - 1\right)$$

由弱大数定律可知，当 $n \to \infty$ 时，

$$\frac{1}{k}\sum_{j=1}^{k} \frac{\left(\frac{dZ_j^{(d+1)}}{m}\right)^{\rho}}{\mathrm{E}\left(\left(\frac{dZ_1^{(d+1)}}{m}\right)^{\rho}\right)} \frac{1}{d}\sum_{i=1}^{d}\left(\log\left(\frac{Z_j^{(i)}}{Z_j^{(d+1)}}\right)\right)\left(\left(\frac{Z_j^{(i)}}{Z_j^{(d+1)}}\right)^{\rho} - 1\right) \xrightarrow{\mathrm{P}}$$

$$\mathrm{E}\left(\frac{1}{d}\sum_{i=1}^{d}\left(\log\left(\frac{Z_1^{(i)}}{Z_1^{(d+1)}}\right)\right)\left(\left(\frac{Z_1^{(i)}}{Z_1^{(d+1)}}\right)^{\rho} - 1\right)\right) = \frac{\rho(2-\rho)}{(1-\rho)^2}$$

故

$$I_2 = \frac{2-\rho}{(1-\rho)^2}(kd)^{\frac{1}{2}} A_0\left(\frac{m}{d}\right) \mathrm{E}\left(\left(\frac{dZ_1^{(d+1)}}{m}\right)^{\rho}\right)(1 + o_p(1))$$

由引理 1 知，当 $m \to \infty$ 时，

$$\frac{1}{1-\rho}\mathrm{E}\left(\left(\frac{dZ^{(d+1)}}{m}\right)^{\rho}\right) = g(d, m, \rho) \to \frac{d^{\rho}}{1-\rho}\frac{\Gamma(d-\rho+1)}{\Gamma(d+1)}$$

结合 $(kd)^{\frac{1}{2}} A\left(\frac{m}{d}\right) = O(1)$ 的假设可以得到，当 $n \to \infty$ 时，$I_2 = \frac{2-\rho}{1-\rho}(kd)^{\frac{1}{2}} A_0\left(\frac{m}{d}\right) g(d, m, \rho) + o_p(1)$. 类似地可以得到 $I_3 = -(kd)^{\frac{1}{2}} A_0\left(\frac{m}{d}\right) g(d, m, \rho) + o_p(1)$. 故 $I_2 + I_3 = \frac{1}{1-\rho}(kd)^{\frac{1}{2}} A_0\left(\frac{m}{d}\right) g(d, m, \rho) + o_p(1)$.

与 I_2 的计算类似，可以得到 $I_4 \xrightarrow{P} 0$，$I_5 \xrightarrow{P} 0$.

综上，当 $n \to \infty$ 时，$(kd)^{\frac{1}{2}}\left\{\hat{\gamma}_{\mathrm{DMR}} - \gamma - (1-\rho)^{-1} A\left(\frac{m}{d}\right) g(d, m, \rho)\right\} \xrightarrow{d} N(0, 2\gamma^2)$，$\rho < 0$ 时的结论得证.

对于 $\rho = 0$，(16) 式等价于当 $n \to \infty$ 时，

$$\frac{\log U(Z_j^{(i)}) - \log U(Z_j^{(d+1)}) - \gamma(\log Z_j^{(i)} - \log Z_j^{(d+1)})}{A_0\left(\frac{m}{d}\right)} =$$

$$\log Z_j^{(i)} - \log Z_j^{(d+1)} +$$

$$o_p(1)\left\{\max\left\{\left(\frac{dZ_j^{(i)}}{m}\right)^{\delta}, \left(\frac{dZ_j^{(i)}}{m}\right)^{-\delta}\right\} + \max\left\{\left(\frac{dZ_j^{(d+1)}}{m}\right)^{\delta}, \left(\frac{dZ_j^{(d+1)}}{m}\right)^{-\delta}\right\}\right\}$$

其后的证明方法与 $\rho < 0$ 的情况类似，此处省略，定理证毕.

3　模拟研究

本文提出了分布式矩率估计量，下面将其与分布式 Hill 估计量进行有限样本表现的比较. 以 $\gamma = 1$，$\rho = -1$，机器数量一定时，Burr 分布的表现为例，其分布函数为 $F(x) = 1 - (1 + x)^{-1}$. 随机生成 n 个来自 Burr 分布的样本，存储在 k 个机器中，每个机器有 m 个观测，对每一个机器，从 m 个观测中选取 d 个超过数(d 为自变量)，分别计算分布式 Hill 估计量和分布式矩率估计量的估计均值和均方误差，每个实验重复 s 次并取平均值.

设置 $n = 1\ 000$，$k = 20$，$m = 50$，$s = 100$，d 的取值范围为 1,…,30，模拟结果如图 1 所示.

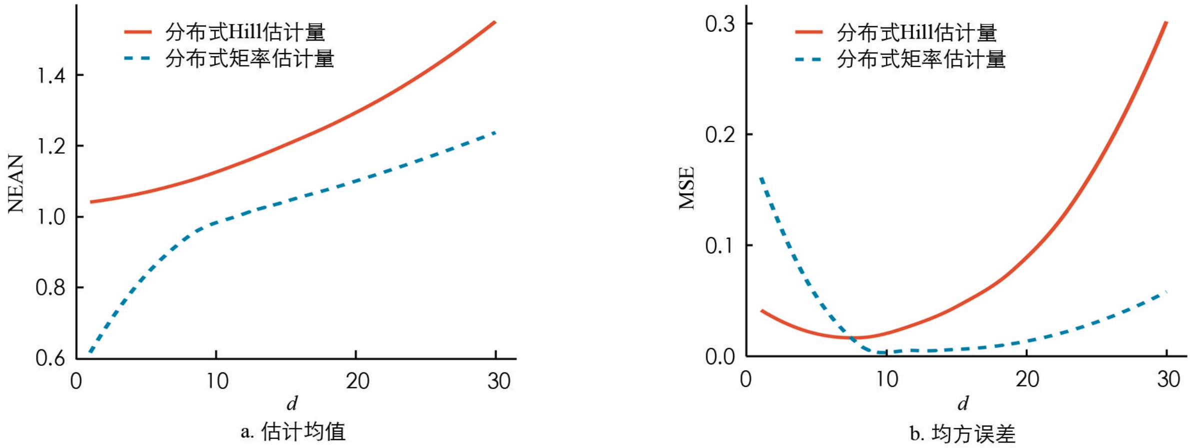

图 1 Burr(1) 的分布式 Hill 估计量和分布式矩率估计量的估计均值及均方误差

由图 1 可知，当 k 一定时，在 $\gamma=1$，$\rho=-1$ 的 Burr 分布下，随着 d 的增加，$\hat{\gamma}_{\mathrm{DH}}$ 和 $\hat{\gamma}_{\mathrm{DMR}}$ 的均值都存在上升的趋势，但当 d 较大时，$\hat{\gamma}_{\mathrm{DMR}}$ 的均值更接近真实值；同时，$\hat{\gamma}_{\mathrm{DH}}$ 和 $\hat{\gamma}_{\mathrm{DMR}}$ 的均方误差随着 d 的增加先下降后上升，并且当 d 较大时，$\hat{\gamma}_{\mathrm{DMR}}$ 的均方误差低于 $\hat{\gamma}_{\mathrm{DH}}$ 的均方误差，表现更为稳定. 因此，在最小均方误差准则下，当 d 较大时，$\hat{\gamma}_{\mathrm{DMR}}$ 的估计效果明显优于 $\hat{\gamma}_{\mathrm{DH}}$.

参考文献：

[1] HILL B M. A Simple General Approach to Inference about the Tail of a Distribution [J]. The Annals of Statistics, 1975, 3(5): 1163-1174.
[2] MASON D M. Laws of Large Numbers for Sums of Extreme Values [J]. The Annals of Probability, 1982, 0(3): 754-764.
[3] DAVIS R, RESNICK S. Tail Estimates Motivated by Extreme Value Theory [J]. The Annals of Statistics, 1984, 12 (4): 1467-1487.
[4] GOLDIE C M, SMITH R L. Slow Variation with Remainder: Theory and Applications [J]. The Quarterly Journal of Mathematics, 1987, 38(1): 45-71.
[5] GOMES M I, MARTINS M J, NEVES M. Alternatives to a Semi-Parametric Estimator of Parameters of Rare Events—The Jackknife Methodology [J]. Extremes, 2000, 3(3): 207-229.
[6] GOMESA M I, MARTINS M J. "Asymptotically Unbiased" Estimators of the Tail Index Based on External Estimation of the Second Order Parameter [J]. Extremes, 2002, 5(1): 5-31.
[7] DE HAAN L, PENG L. Comparison of Tail Index Estimators [J]. Statistica Neerlandica, 1998, 52(1): 60-70.
[8] PENG L. Asymptotically Unbiased Estimators for the Extreme-Value Index [J]. Statistics & Probability Letters, 1998, 38(2): 107-115.
[9] GOMES M I, MARTINS M J. Bias Reduction and Explicit Semi-Parametric Estimation of the Tail Index [J]. Journal of Statistical Planning and Inference, 2004, 124(2): 361-378.
[10] IVETTE GOMES M, DE HAAN L, RODRIGUES L H. Tail Index Estimation for Heavy-Tailed Models: Accommodation of Bias in Weighted Log-Excesses [J]. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 2008, 70(1): 31-52.
[11] DAVYDOV Y, PAULAUSKAS V, RAČKAUSKAS A. More on P-Stable Convex Sets in Banach Spaces [J]. Journal of Theoretical Probability, 2000, 13(1): 39-64.
[12] PAULAUSKAS V. A New Estimator for a Tail Index [J]. Acta Applicandae Mathematica, 2003, 79(1): 55-67.
[13] CHEN L J, LI D Y, ZHOU C. Distributed Inference for the Extreme Value Index [J]. Biometrika, 2022, 109(1): 257-264.
[14] DE HAAN L, FERREIRA A. Extreme Value Theory [M]. New York: Springer New York, 2006.
[15] DEKKERS A L M, EINMAHL J H J, DE HAAN L. A Moment Estimator for the Index of an Extreme-Value Distribution [J]. The Annals of Statistics, 1989, 17(4): 1833-1855.

责任编辑 张栒