第 43 卷第 3 期 西南大学学报(自然科学版) 2021 年 3 月
Vol. 43 No. 3 Journal of Southwest University (Natural Science Edition) Mar. 2021

DOI: 10.13718/j.cnki.xdzk.2021.03.013

一类重尾极值指数估计

张绿云, 陈守全

西南大学 数学与统计学院, 重庆 400715

摘要: 提出了如下一类重尾极值指数估计

$$\tilde{\gamma}_n^{(\beta_1,\beta_2)}(k)=\left\{(\beta_2-\beta_1)\left[\frac{\hat{x}^{(\beta_1)}(k)}{\hat{x}^{(\beta_2)}(k)}-1\right]^{-1}+1-\beta_1\right\}^{-1}$$

其中

$$\hat{x}^{(\beta)}(k):=\frac{1}{k}\sum_{i=0}^{k-1}\left(\frac{X_{n-i,n}}{X_{n-k,n}}\right)^{1-\beta}$$

证明了其弱相合性和渐近正态性, 并在均方误差最小的情况下, 给出了 $\tilde{\gamma}_n^{(\beta_1,\beta_2)}(k)$ 中调节参数 k 的最优选择.

关 键 词: 渐近正态性; 极值指数; 二阶正则变换

中图分类号: O211.4 **文献标志码:** A **文章编号:** 1673-9868(2021)03-0089-06

设 $\{X_n, n\geqslant 1\}$ 是一列独立同分布的随机变量序列, 其共同的分布函数为 $F(x)$, $X_{1,n}\leqslant\cdots\leqslant X_{n,n}$ 为 $X_1, \cdots, X_n$ 的顺序统计量. 若对非退化分布函数 $G(x)$, 存在规范化常数 $a_n>0$, $b_n\in\mathbb{R}$, 对一切 $G(x)$ 的连续点有

$$\lim_{n\to\infty}P(X_{n,n}\leqslant a_n x+b_n)=\lim_{n\to\infty}F^n(a_n x+b_n)=G(x)$$

成立, 则称 F 属于 G 的吸引场, 记为 $F\in D(G)$. 文献[1-2] 证明 $G(x)$ 的形式必为

$$G(x)=\exp\{-(1+\gamma x)^{-\frac{1}{\gamma}}\}\qquad \gamma\in\mathbb{R},\ 1+\gamma x\geqslant 0$$

极值理论的应用非常广泛, 比如在自然灾害、巨额保险赔付以及其他领域如水文学或工程中的罕见事件都时常用到. 形状参数 γ 被称作极值指数(EVI), 它衡量右尾函数 $\overline{F}(x):=1-F(x)$ 的尾部性质. 当分布函数未知时, 对极值指数 γ 的估计是极值理论的重要组成部分. 因此, 对于极值指数 γ 的估计受到学者的广泛关注[3-8].

在本文中, 我们针对极值指数 $\gamma>0$ 的分布函数进行研究, 这种分布函数也称作重尾分布函数. 如果 $1-F$ 是指数为 $-\frac{1}{\gamma}$ 的正则变换(记为 $1-F\in RV_{-\frac{1}{\gamma}}$), 即

$$\lim_{t\to\infty}\frac{1-F(tx)}{1-F(x)}=x^{-\frac{1}{\gamma}}\tag{1}$$

则称分布函数 F 是重尾分布函数.

对于极值指数的研究, 前人已经得到较好的结论. 当 $\gamma>0$ 时, 文献[3] 提出了著名的的 Hill 估计量.

收稿日期: 2019-12-18

基金项目: 国家自然科学基金项目(11571283).

作者简介: 张绿云, 硕士研究生, 主要从事极值统计的研究.

通信作者: 陈守全, 博士, 副教授.

而作为 Hill 估计的一个推广，文献[9] 提出了如下定义的调和矩估计

$$H_{n,k}^{(\beta)} := \frac{1}{\beta-1}\left\{\left[\frac{1}{k}\sum_{i=1}^{k}\left(\frac{X_{n-k,n}}{X_{n-i+1,n}}\right)^{\beta-1}\right]^{-1}-1\right\}$$

其中 $1 \leqslant k \leqslant n-1$ 且 $\beta > 0$ 是调谐参数.

本文受文献[9] 中估计量的启发，提出定义如下的一类关于重尾极值指数的新估计

$$\tilde{\gamma}_n^{(\beta_1,\beta_2)}(k) = \left\{(\beta_2-\beta_1)\left[\frac{\hat{x}^{(\beta_1)}(k)}{\hat{x}^{(\beta_2)}(k)}-1\right]^{-1}+1-\beta_1\right\}^{-1} \tag{2}$$

其中 $\beta_1 \neq \beta_2$ 是调谐参数 β 的两个取值，且

$$\hat{x}^{(\beta)}(k) := \frac{1}{k}\sum_{i=0}^{k-1}\left(\frac{X_{n-i,n}}{X_{n-k,n}}\right)^{1-\beta}$$

当序列 $k = k(n) \to \infty$ 满足 $\frac{k}{n} \to 0$，此时容易得到 $\tilde{\gamma}_n^{(\beta_1,\beta_2)}(k)$ 的弱相合性. 此外，存在 $A(t)$ 使得当 $t \to \infty$

$$\frac{\frac{U(tx)}{U(t)}-x^{\gamma}}{A(t)} \to x^{\gamma}\frac{x^{\rho}-1}{\rho} \tag{3}$$

成立时，我们考虑 $\tilde{\gamma}_n^{(\beta_1,\beta_2)}(k)$ 的渐近性质，其中 $\rho < 0$ 是二阶参数且 $|A(t)| \in RV_{\rho}$(参见文献[10]，推论 2.3.5).

1 主要结果

在下文中，我们假设 $F \in D(G)$，$\gamma > 0$，这等价于 $U := \left(\frac{1}{1-F(x)}\right)^{\leftarrow}$ 是指数为 γ 的正则变换(记为 $U \in RV_{\gamma}$)，其中 $F^{\leftarrow}(x) := \inf\{y : F(y) \geqslant x\}$ 是 F 的广义逆函数.

$F \in D(G)$ 当且仅当

$$\lim_{t\to\infty}\frac{U(tx)-U(t)}{a(t)} = \frac{x^{\gamma}-1}{\gamma} \qquad \gamma \in \mathbb{R} \tag{4}$$

对所有 $x > 0$ 和某些正可测函数 $a(t)$ 成立. 注意，如果 $\gamma = 0$，则(4) 式的右边为 $\ln x$.

为简化定理的表示，本文规定以下记号：

$$V_{\beta_1,\beta_2}(\gamma) = \frac{\gamma^2(1-\beta_1)^2}{1-2\gamma(1-\beta_1)} + \frac{\gamma^2(1-\beta_2)^2}{1-2\gamma(1-\beta_2)} - \frac{2\gamma^2(1-\beta_1)(1-\beta_2)}{1-\gamma(1-\beta_1)-\gamma(1-\beta_2)} \tag{5}$$

下面给出本文的主要结果.

定理 1 假设条件(4) 成立，序列 $\{k(n)\}$ 满足 $k = k(n) \to \infty$，$\frac{k(n)}{n} \to 0 (n \to \infty)$. 则

$$\tilde{\gamma}_n^{(\beta_1,\beta_2)}(k) \xrightarrow{\mathrm{p}} \gamma$$

此外，若二阶条件(3) 成立，且存在 $\lambda \in \mathbb{R}$ 使得

$$\lim_{n\to\infty}\sqrt{k}A\left(\frac{n}{k}\right) = \lambda$$

则对 $\beta_1 > 1-\frac{1}{2\gamma}$，$\beta_2 > 1-\frac{1}{2\gamma}$，有

$$\sqrt{k}\,(\tilde{\gamma}_n^{(\beta_1,\beta_2)}(k)-\gamma) \xrightarrow{\mathrm{d}} N(\lambda\mu, \sigma^2)$$

其中

$$\mu := \frac{(1-\gamma(1-\beta_1))(1-\gamma(1-\beta_2))(1-\rho)}{(1-\rho-\gamma(1-\beta_1))(1-\rho-\gamma(1-\beta_2))}$$

$$\sigma^2 := \frac{(1-\gamma(1-\beta_1))^2(1-\gamma(1-\beta_2))^2 V_{\beta_1,\beta_2}(\gamma)}{(\beta_2-\beta_1)^2}$$

$V_{\beta_1,\beta_2}(\gamma)$ 如(5)定义.

下面考虑序列 k 的最优选择，寻找使 $\widetilde{\gamma}_n^{(\beta_1,\beta_2)}(k)$ 的均方误差 $\mathrm{MSE}_\infty(\widetilde{\gamma}_n^{(\beta_1,\beta_2)}(k))$ 可以达到最小的 k.

定理 2 1) 假设条件(3)对 $\rho<0$ 成立，中间序列 $\{k_0(n)\}$ 表示使 $\mathrm{MSE}_\infty(\widetilde{\gamma}_n^{(\beta_1,\beta_2)}(k))=A^2\left(\frac{n}{k}\right)\mu^2+\frac{\sigma^2}{k}$ 达到最小的序列.

则

$$\sqrt{k_0}\,(\widetilde{\gamma}_n^{(\beta_1,\beta_2)}(k_0)-\gamma)\xrightarrow{\mathrm{d}}N\left(\frac{\mathrm{sign}(A)}{\sqrt{-2\rho}}\sigma,\ \sigma^2\right)$$

此外，

$$k_0(n):=k_0\sim\frac{n}{s^{\leftarrow}\left(\frac{\tau}{n}\right)}$$

其中 $s^{\leftarrow}$ 是 $s\in RV_{2\rho-1}$ 的广义逆函数，且

$$A^2(t)\sim\int_t^\infty s(u)\mathrm{d}u$$

$$\tau:=\frac{\sigma^2}{\mu^2}=\frac{(1-\rho-\gamma(1-\beta_1))^2(1-\rho-\gamma(1-\beta_2))^2V_{\beta_1,\beta_2}(\gamma)}{(\beta_2-\beta_1)^2(1-\rho)^2}$$

2) 假设 $A(t)=C\,t^\rho$，则 1) 中的 k_0 为

$$k_0:=k_0(n)=\left\lfloor\left(\frac{\tau}{-2\rho C^2}\right)^{\frac{1}{1-2\rho}}n^{-\frac{2\rho}{1-2\rho}}\right\rfloor$$

其中 $\lfloor x\rfloor$ 表示不超过 x 的最大整数. 此外，

$$\sqrt{k_0}\,(\widetilde{\gamma}_n^{(\beta_1,\beta_2)}(k_0)-\gamma)\xrightarrow{\mathrm{d}}N\left(\frac{\mathrm{sign}(C)}{\sqrt{-2\rho}}\sigma,\ \sigma^2\right)$$

且

$$\mathrm{MSE}_\infty(\widetilde{\gamma}_n^{(\beta_1,\beta_2)}(k_0))=\sigma^2\left(1-\frac{1}{2\rho}\right)\left[\left(\frac{\tau}{-2\rho C^2}\right)^{\frac{1}{2\rho-1}}n^{\frac{2\rho}{1-2\rho}}\right]$$

2 定理的证明

设 $Y_1,Y_2,\cdots,Y_n$ 是独立同分布的标准 Pareto 随机变量序列，其共同的分布函数为 $1-\frac{1}{y}$，$y\geqslant1$，$Y_{1,n}\leqslant Y_{2,n}\leqslant\cdots\leqslant Y_{n,n}$ 表示 $Y_1,Y_2,\cdots,Y_n$ 的顺序统计量. 易知 $\{X_i\}_{i=1}^n\overset{\mathrm{d}}{=}\{U(Y_i)\}_{i=1}^n$，且 $\left\{\frac{Y_{n-i,n}}{Y_{n-k,n}}\right\}_{i=0}^{k-1}\overset{\mathrm{d}}{=}\{Y_{k-i,k}\}_{i=0}^{k-1}$.

下面对本文的主要结果进行证明.

定理 1 的证明 由二阶条件可得

$$\frac{U(tx)}{U(t)}=x^\gamma+A(t)x^\gamma\frac{x^\rho-1}{\rho}(1+o(1))$$

因此

$$\frac{X_{n-i,n}}{X_{n-k,n}}\overset{\mathrm{d}}{=}\frac{U(Y_{n-i,n})}{U(Y_{n-k,n})}=$$

$$\left(\frac{Y_{n-i,n}}{Y_{n-k,n}}\right)^\gamma\left[1+A(Y_{n-k,n})\frac{\left(\frac{Y_{n-i,n}}{Y_{n-k,n}}\right)^\rho-1}{\rho}+o_p(A(Y_{n-k,n}))\right]$$

故

$$\hat{x}^{(\beta)}(k):=\frac{1}{k}\sum_{i=0}^{k-1}\left(\frac{X_{n-i,n}}{X_{n-k,n}}\right)^{1-\beta}\overset{d}{=}$$

$$\frac{1}{k}\sum_{i=1}^{k}Y_i^{\gamma(1-\beta)}+(1-\beta)A(Y_{n-k,n})\frac{1}{k}\sum_{i=1}^{k}Y_i^{\gamma(1-\beta)}\frac{Y_i^{\rho}-1}{\rho}+o_p(A(Y_{n-k,n}))$$

当 $a<1$ 时，$\mathrm{E}(Y^a)=\dfrac{1}{1-a}$，当 $a<\dfrac{1}{2}$ 时，$\mathrm{var}(Y^a)=\dfrac{a^2}{(1-a)^2(1-2a)}=:\sigma_a^2$ 且当 $a<1$，$a+b<1$ 时

$$\mathrm{E}\left(Y^a\frac{Y^b-1}{b}\right)=\frac{1}{(1-a)(1-a-b)}$$

那么对于 $\beta>1-\dfrac{1}{2\gamma}$，当 $n\to\infty$ 时

$$\frac{\sqrt{k}\left(\frac{1}{k}\sum_{i=1}^{k}Y_i^{\gamma(1-\beta)}-\frac{1}{1-\gamma(1-\beta)}\right)}{\sigma_{\gamma(1-\beta)}}=:P_n^{(\beta)}\to N(0,1)$$

其中 $N(0,1)$ 表示标准正态分布. 由 $P_n^{(\beta)}$ 的定义可得

$$\hat{x}^{(\beta)}(k)\overset{d}{=}\frac{1}{1-\gamma(1-\beta)}\left[1+\frac{\gamma(1-\beta)}{\sqrt{1-2\gamma(1-\beta)}\sqrt{k}}P_n^{(\beta)}+\frac{(1-\beta)A\left(\frac{n}{k}\right)}{1-\rho-\gamma(1-\beta)}+o_p\left(A\left(\frac{n}{k}\right)\right)\right]$$

那么

$$\frac{1}{k}\sum_{i=0}^{k-1}\left(\frac{X_{n-i,n}}{X_{n-k,n}}\right)^{1-\beta_1}\overset{d}{=}$$

$$\frac{1}{1-\gamma(1-\beta_1)}\left[1+\frac{\gamma(1-\beta_1)}{\sqrt{1-2\gamma(1-\beta_1)}\sqrt{k}}P_n^{(\beta_1)}+\frac{(1-\beta_1)A\left(\frac{n}{k}\right)}{1-\rho-\gamma(1-\beta_1)}+o_p\left(A\left(\frac{n}{k}\right)\right)\right]$$

$$\left[\frac{1}{k}\sum_{i=0}^{k-1}\left(\frac{X_{n-i,n}}{X_{n-k,n}}\right)^{1-\beta_2}\right]^{-1}\overset{d}{=}(1-\gamma(1-\beta_2))\left[1-\frac{\gamma(1-\beta_2)}{\sqrt{1-2\gamma(1-\beta_2)}\sqrt{k}}P_n^{(\beta_2)}-\frac{(1-\beta_2)A\left(\frac{n}{k}\right)}{1-\rho-\gamma(1-\beta_2)}\right](1+o_p(1))$$

从而

$$\frac{\hat{x}^{(\beta_1)}(k)}{\hat{x}^{(\beta_2)}(k)}\overset{d}{=}\frac{1-\gamma(1-\beta_2)}{1-\gamma(1-\beta_1)}\Bigg[1+\frac{1}{\sqrt{k}}\left(\frac{\gamma(1-\beta_1)}{\sqrt{1-2\gamma(1-\beta_1)}}P_n^{(\beta_1)}-\frac{\gamma(1-\beta_2)}{\sqrt{1-2\gamma(1-\beta_2)}}P_n^{(\beta_2)}\right)+$$

$$\frac{(\beta_2-\beta_1)(1-\rho)A\left(\frac{n}{k}\right)}{(1-\rho-\gamma(1-\beta_1))(1-\rho-\gamma(1-\beta_2))}\Bigg]+o_p\left(A\left(\frac{n}{k}\right)\right)+o_p\left(\frac{1}{\sqrt{k}}\right)$$

现定义

$$Q_n^{(\beta_1,\beta_2)}(k):=\frac{\gamma(1-\beta_1)}{\sqrt{1-2\gamma(1-\beta_1)}}P_n^{(\beta_1)}-\frac{\gamma(1-\beta_2)}{\sqrt{1-2\gamma(1-\beta_2)}}P_n^{(\beta_2)}$$

$$T_n^{(\beta_1,\beta_2)}(k):=\frac{Q_n^{(\beta_1,\beta_2)}(k)}{\sqrt{\mathrm{var}(Q_n^{(\beta_1,\beta_2)}(k))}}$$

设 $f_k(t)$ 表示 $Q_n^{(\beta_1,\beta_2)}(k)$ 的特征函数. 由 $P_n^{(\beta)}$ 的表达式，有

$$f_k(t)=\mathrm{E}\exp\left\{it\frac{\gamma(1-\beta_1)}{\sqrt{1-2\gamma(1-\beta_1)}}P_n^{(\beta_1)}-it\frac{\gamma(1-\beta_2)}{\sqrt{1-2\gamma(1-\beta_2)}}P_n^{(\beta_2)}\right\}=$$

$$\prod_{j=1}^{k}\left\{1-\frac{t^2}{2k}\mathrm{E}[(1-\gamma(1-\beta_1))Y_j^{\gamma(1-\beta_1)}-(1-\gamma(1-\beta_2))Y_j^{\gamma(1-\beta_2)}]^2+o_p\left(\frac{1}{k}\right)\right\}=$$

$$\left\{1-\frac{t^2}{2k}V_{\beta_1,\beta_2}(\gamma)+o_p\left(\frac{1}{k}\right)\right\}^k \to$$
$$\exp\left\{-t^2\,\frac{V_{\beta_1,\beta_2}(\gamma)}{2}\right\}$$

其中 $\beta_1>1-\frac{1}{2\gamma}$，$\beta_2>1-\frac{1}{2\gamma}$，$V_{\beta_1,\beta_2}(\gamma)$ 如前定义.

因此

$$\frac{\hat{x}^{(\beta_1)}(k)}{\hat{x}^{(\beta_2)}(k)} \overset{d}{=} \frac{1-\gamma(1-\beta_2)}{1-\gamma(1-\beta_1)}+\frac{1-\gamma(1-\beta_2)}{1-\gamma(1-\beta_1)}\,\frac{\sqrt{V_{\beta_1,\beta_2}(\gamma)}}{\sqrt{k}}T_n^{(\beta_1,\beta_2)}+$$
$$\frac{1-\gamma(1-\beta_2)}{1-\gamma(1-\beta_1)}\,\frac{(\beta_2-\beta_1)(1-\rho)A\left(\frac{n}{k}\right)}{(1-\rho-\gamma(1-\beta_1))(1-\rho-\gamma(1-\beta_2))}+$$
$$o_p\left(A\left(\frac{n}{k}\right)\right)+o_p\left(\frac{1}{\sqrt{k}}\right)$$

若存在 $\lambda\in\mathbb{R}$ 满足 $\sqrt{k}A\left(\frac{n}{k}\right)\to\lambda$，则当 $n\to\infty$ 时

$$\sqrt{k}\left(\frac{\hat{x}^{(\beta_1)}(k)}{\hat{x}^{(\beta_2)}(k)}-\frac{1-\gamma(1-\beta_2)}{1-\gamma(1-\beta_1)}\right)\xrightarrow{d}N(\lambda\tilde{\mu},\tilde{\sigma}^2)$$

其中

$$\tilde{\mu}:=\frac{1-\gamma(1-\beta_2)}{1-\gamma(1-\beta_1)}\,\frac{(\beta_2-\beta_1)(1-\rho)}{(1-\rho-\gamma(1-\beta_1))(1-\rho-\gamma(1-\beta_2))}$$
$$\tilde{\sigma}^2:=\left(\frac{1-\gamma(1-\beta_2)}{1-\gamma(1-\beta_1)}\right)^2V_{\beta_1,\beta_2}(\gamma)$$

设

$$g(\theta)=\left[\frac{\beta_2-\beta_1}{\theta-1}+1-\beta_1\right]^{-1}$$

由于

$$g\left(\frac{1-\gamma(1-\beta_2)}{1-\gamma(1-\beta_1)}\right)=\gamma$$

故现在利用映射 $g(\theta)$ 来构造一个关于 γ 的估计量. 应用 Delta 方法，可得

$$\sqrt{k}\left(\left\{(\beta_2-\beta_1)\left[\frac{\hat{x}^{(\beta_1)}(k)}{\hat{x}^{(\beta_2)}(k)}-1\right]^{-1}+1-\beta_1\right\}^{-1}-\gamma\right)\xrightarrow{d}N(\lambda\mu,\sigma^2)$$

其中 μ，σ^2 如前定义.

定理 2 的证明　类似文献[10]中的定理可得结果.

参考文献：

[1] FISHER R A，TIPPETT L H C. Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample [J]. Mathematical Proceedings of the Cambridge Philosophical Society，1928，24(2)：180-190.

[2] JENKINSON A F. The Frequency Distribution of the Annual Maximum (or Minimum) Values of Meteorological Elements [J]. Quarterly Journal of the Royal Meteorological Society，1955，81(348)：158-171.

[3] 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.

[4] PICKANDS J. Statistical Inference Using Extreme Order Statistics [J]. The Annals of Statistics，1975，3(1)：119-131.

[5] DEKKERS A L M，EINMAHL J H J，DE HAAN L. A Moment Estimator for the Index of an Extreme-Value Distribu-

tion [J]. The Annals of Statistics, 1989, 17(4): 1833-1855.
[6] BEIRLANT J, VYNCKIER P, TEUGELS J L. Excess Functions and Estimation of the Extreme-Value Index [J]. Bernoulli, 1996, 2(4): 293-318.
[7] FRAGA ALVES M I, GOMES M I, DE HAAN L, et al. Mixed Moment Estimator and Location Invariant Alternatives [J]. Extremes, 2009, 12(2): 149-185.
[8] 王嫣然, 彭作祥. 基于分块的重尾指数估计量的推广 [J]. 西南师范大学学报(自然科学版), 2019, 44(1): 25-28.
[9] BERAN J, SCHELL D, STEHLÍK M. The Harmonic Moment Tail Index Estimator: Asymptotic Distribution and Robustness [J]. Annals of the Institute of Statistical Mathematics, 2014, 66(1): 193-220.
[10] DE HAAN L, FERREIRA A. Extreme Value Theory [M]. New York: Springer, 2006.

A Class of Extreme Value Index Estimators for Heavy-Tailed Distributions

ZHANG Lyu-yun, CHEN Shou-quan
School of Mathematics and Statistics, Southwest University, Chongqing 400715, China

Abstract: In this paper, a new class of estimators is given by

$$\tilde{\gamma}_n^{(\beta_1,\beta_2)}(k)=\left\{(\beta_2-\beta_1)\left[\frac{\hat{x}^{(\beta_1)}(k)}{\hat{x}^{(\beta_2)}(k)}-1\right]^{-1}+1-\beta_1\right\}^{-1}$$

where

$$\hat{x}^{(\beta)}(k):=\frac{1}{k}\sum_{i=0}^{k-1}\left(\frac{X_{n-i,n}}{X_{n-k,n}}\right)^{1-\beta}$$

Its weak consistency and asymptotic normality are presented, and the optimal choice of sample fraction k in $\tilde{\gamma}_n^{(\beta_1,\beta_2)}(k)$ by minimum mean square error is discussed.

Key words: asymptotic normality; extreme value index; second-order regular variation

责任编辑 张 枸