On semiparametric tail index estimation based on exponential families

Schell, Dieter (1st author)
University of Konstanz, Department of Mathematics and Statistics
78457 Konstanz, Germany
E-mail: dieter.schell@uni-konstanz.de

Beran, Jan
University of Konstanz, Department of Mathematics and Statistics
78457 Konstanz, Germany
E-mail: jan.beran@uni-konstanz.de

Introduction

We consider heavy-tailed distributions F defined by

$$\overline{F}(x) = 1 - F(x) = x^{-\alpha} L_F(x),\ \alpha > 0,$$

where L_F is slowly varying. The most popular estimator of the tail index α, introduced by Hill (1975), is given by

$$\hat{\gamma}_n^{(H)}(k) = \frac{1}{k} \sum_{i=1}^{k} \log\left(\frac{X_{n-i+1,n}}{X_{n-k,n}}\right), \quad \text{for} \quad k = 1, ..., n-1,$$

where $\gamma = \frac{1}{\alpha}$ and $X_{1,n} \leq X_{2,n} \leq ... \leq X_{n,n}$ are order statistics based on an iid sample. For $k \to \infty$, $n \to \infty$, $k/n \to 0$, $\hat{\gamma}_n^{(H)}$ is consistent (Mason 1982) and asymptotically normal (under additional regularity conditions). The estimator performs well in the class of Pareto-like distributions (see e.g. Peng 1998). Outside this class the bias-variance tradeoff can be addressed by choosing a sequence k_n^{opt} that minimizes the AMSE (Hall 1982). However, k_n^{opt} is only asymptotically optimal, providing little guidance for finite samples (Drees et al. 2000). Moreover, using k_n^{opt} leads to a slower rate of convergence compared to parametric estimation. (cf. Beran and Schell 2010).

Tail Index Estimation by an exponential family of Pareto-Spline distributions

The new approach is based on an exponential family generated by a hyperbola and a truncated series expansion. The idea is related to Neyman (1937), Stone and Koo (1986), Clutton-Brock (1990). Let

$$f(x;\theta) = \frac{\exp(\sum_{j=1}^{\infty} \theta_j B_j(x)) x^{-\alpha-1}}{\int \exp(\sum_{j=1}^{\infty} \theta_j B_j(x)) x^{-\alpha-1} dx}, \quad (x \geq 1,\ \alpha > 0),$$

where $(B_i)_{i\in\mathbb{N}}$ is a sequence of basis functions with compact support $[1, W]$ and $\theta = (\alpha, \theta_1, \dots) \in \Theta = (0, \infty) \times \mathbb{R}^{\mathbb{N}}$ an infinite dimensional parameter vector consisting of the tail index and the coefficients of the basis functions.

We will focus on cubic B-splines as basis functions, since for a given compact interval $[a, b] \in \mathbb{R}$, there exists some constant c_3 such that for all sequences of knots $\mathbf{t} = (t_i)_{i=1}^{n+3}$ with

$$t_1 = t_2 = t_3 = a < t_4 \leq \cdots < b = t_{n+1} = t_{n+2} = t_{n+3}$$

and for all $g \in C^{(2)}([a, b])$,

$$dist(g, S_{3,\mathbf{t}}) \leq c_3 |\mathbf{t}|^3 ||D^3 g||,$$

where $S_{3,\mathbf{t}}$ is the linear space of cubic B-splines associated with the knot sequence $\mathbf{t}$, $dist(g,S) = \inf_{f\in S}\{||g-f||\}$, $||f|| := \max_{x\in[a,b]}|f(x)|$, D is the differential operator and $|\mathbf{t}| = \max_i \Delta t_i$ (de Boor 2001). Under the regularity conditions

(1)
$$|\theta_j| < \infty, \quad \sum_{j=1}^{\infty}\theta_j < \infty, \quad \sum_{j=1}^{\infty}\theta_j B_j(W) = 0, \quad \sum_{j=1}^{\infty}\theta_j B_j'(W) = 0,$$
$$\text{and} \quad \int \exp\left(\sum_{j=1}^{\infty}\theta_j B_j(x)\right) x^{-\alpha-1}dx < \infty$$

$f(x;\theta)$ is a well defined, smooth density. Considering the restricted parameter space $\Theta^{(p)} = (0,\infty)\times\mathbb{R}^p$, we obtain

$$f(x;\theta^{(p)}) = \exp\left[\sum_{j=1}^{p}\theta_j B_j(x) + \alpha(-\log(x)) - A(\theta)\right] h(x),$$

where $h(x) = \frac{1}{x}$, $\theta^{(p)} = (\alpha,\theta_1,\ldots,\theta_p) \in \Theta^{(p)}$ and

$$A(\theta^{(p)}) = \log\left(\int \exp\left(\sum_{j=1}^{p}\theta_j B_j(x) + \alpha(-\log(x))\right)\frac{1}{x}dx\right).$$

Thus,

$$\mathfrak{F}(\theta^{(p)}) = \{F(x;\theta^{(p)}) = \int_1^x f(y;\theta)dy, \; \theta^{(p)} \in \Theta^{(p)}, \}$$

is a regular exponential family, which will be referred to as the family of Pareto-Spline distributions. The statistic

$$T(x) = (-\log(x), B_1(x), \ldots, B_p(x))$$

is sufficient and $\mathfrak{F}(\theta^{(p)})$ is complete. We define an M-functional $T(F(x;\theta^{(p)}))$ associated with

$$\psi(x;\theta^{(p)}) = \nabla \log f(x;\theta^{(p)}) = \begin{pmatrix} -\log(x) + \int \log(x) f(x;\theta^{(p)})dx \\ B_1(x) - \int B_1(x) f(x;\theta^{(p)})dx \\ \cdots \\ B_p(x) - \int B_p(x) f(x;\theta^{(p)})dx \end{pmatrix}$$

as the solution $t^{(p)}$ of

$$\int \nabla \log f(x;t^{(p)})dF(x;\theta^{(p)}) = 0. \quad \text{(see e.g. Serfling 1980)}$$

The ML-estimator $\hat{\theta}_n^{(p)} := T(F_n(x))$ is defined as a solution of

(2)
$$\int \nabla \log f(x;\hat{\theta}_n^{(p)})dF_n(x) = \begin{pmatrix} -\sum_{i=1}^n \log(X_i) + n\int \log(x) f(x;\hat{\theta}_n^{(p)})dx \\ \sum_{i=1}^n B_1(X_i) - n\int B_1(x) f(x;\hat{\theta}_n^{(p)})dx \\ \cdots \\ \sum_{i=1}^n B_p(X_i) - n\int B_p(x) f(x;\hat{\theta}_n^{(p)})dx \end{pmatrix} = 0$$

where $F_n(x)$ is the empirical distribution function. Thus, $\hat{\theta}_n^{(p)}$ results from adjusting the population moments of $-\log(X), B_1(X), \ldots, B_p(X)$ to the empirical moments $-\frac{1}{n}\sum_{i=1}^n \log(X_i)$, $\frac{1}{n}\sum_{i=1}^n B_1(X_i)$, $\ldots$, $\frac{1}{n}\sum_{i=1}^n B_p(X_i)$. Standard arguments imply that $\hat{\theta}_n^{(p)}$ is consistent and asymptotically normal (see e.g. Bickel and Doksum (1977)) with covariance matrix

$$\Sigma := (\sigma_{i,j}^2)_{i,j=1,\ldots,p+1} = \left(\ddot{A}(\theta^{(p)})\right)^{-1} = (I_{\theta^{(p)}})^{-1},$$

where

$$I_{\theta^{(p)}} = \begin{pmatrix} a_{11} & c^T \\ c & Cov_{\theta^{(p)}}(B_1(X), \dots, B_p(X)) \end{pmatrix}_{(p+1)\times(p+1)},$$

$$a_{11} = E_{\theta^{(p)}}\left[\left(\frac{\partial}{\partial\alpha}\log f(x;\theta^{(p)})\right)^2\right], \; c_i = E_{\theta^{(p)}}\left[\frac{\partial}{\partial\alpha}\log f(x;\theta^{(p)}) \cdot \frac{\partial}{\partial\theta_i}\log f(x;\theta^{(p)})\right] \quad (i = 1, \dots, p)$$

and $Cov_{\theta^{(p)}}(B_1(X), \dots, B_p(X)) = (h_{i,j})_{i,j=1,\dots,p}$ with

$$h_{i,j} = E_{\theta^{(p)}}\left[\frac{\partial}{\partial\theta_i^{(p)}}\log f(x;\theta^{(p)}) \cdot \frac{\partial}{\partial\theta_j^{(p)}}\log f(x;\theta^{(p)})\right].$$

In particular, the asymptotic variance of $\hat{\alpha}_n^{(p)}$ is given by

$$\sigma_{11}^2 = \frac{|I_{\theta^{(p)}}^{(11)}|}{|I_{\theta^{(p)}}|} = \frac{|Cov_{\theta^{(p)}}(B_1(X), \dots, B_p(X))|}{|I_{\theta^{(p)}}|}$$

where $|I_{\theta^{(p)}}^{(11)}|$ is the minor of the first element in the Fisher Information matrix $I_{\theta^{(p)}}$. Note that σ_{11}^2 can be interpreted as the ratio of linear dependence of $(-\log(X), B_1(X), \dots, B_p(X))$ and linear dependence of $(B_1(X), \dots, B_p(X))$.

Asymptotic properties outside the exponential family

Let $F_{\alpha_0}(x)$, $(\alpha_0 > 0)$ be an arbitrary heavy-tailed distribution with density $f_{\alpha_0}(x)$. We define

$$\theta^{(p)} := \arg \max_{\tilde{\theta}^{(p)} \in \Theta^{(p)}} L_{F_{\alpha_0}}(\tilde{\theta}^{(p)}),$$

where

$$L_F(\theta) := \int \log f(x;\theta) dF(x)$$

is assumed to be strictly concave in θ. The corresponding minimum contrast estimator $\hat{\theta}_n^{(p)}$ is defined by

$$\hat{\theta}_n^{(p)} := \arg \max_{\theta\in\Theta} L_{F_n}(\theta).$$

and is asymptotically normal under some regularity conditions. Assuming that the moments of $\hat{\theta}_n^{(p)}$ converge, the mean squared error can be approximated by

$$AMSE_{\alpha_0}(\hat{\alpha}_n) = (\alpha_0 - \alpha)^2 + O(n^{-1}).$$

For small and moderate sample sizes, the stability of $\hat{\alpha}_n$ due to the $\sqrt{n}$-convergence to α often outweighs the lack of consistence. This is illustrated in figure 1.

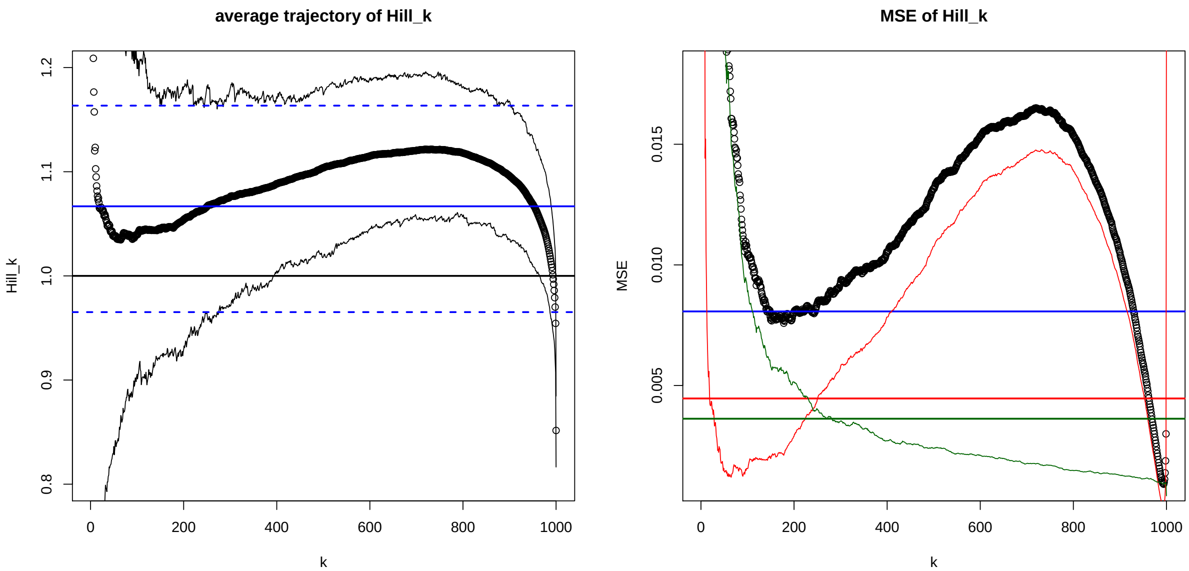

Figure 1: (Left) Average Hill estimator (with 95% confidence intervals) as a function of k (black) and $\hat{\alpha}_n$ with $W = q_{Fr}(0.75)$ and 8 equidistant knots together with a 95% confidence interval (blue). (Right) MSE of Hill's estimator for different values of k (black) and MSE for $\hat{\alpha}_n$ (blue), as well as the variances (green) and the squared bias (red) for both estimators.

REFERENCES

de Boor, C. (2001). A Practical Guide to Splines. *Springer, New York*

Beran, J. and Schell, D. (2010). On robust tail-index estimation. *Computational Statistics & Data Analysis, Article in Press*

Bickel, P. J. and Doksum, D. A. (1977). Mathematical Statistics. *San Francisco: Holden-Day Inc.*

Clutton-Brock, M. (1990). Density estimation using exponential orthogonal series. *Journal of the American Statistical Association 85, 760–764*

Drees, H., de Haan, L. and Resnick, S. (2000) How to make a Hill Plot. *The Annals of Statistics 28, 254–274*

Hall, P. (1982). On some simple estimates of an exponent of regular variation. *Journal of Royal Statistical Society B 44, 37-42*

Hill, B. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Probability 10, 754–764*

Mason, D.M. (1982). Laws of Large Numbers for Sums of Extreme Values. *John Whiley & Sons, New York*

Neyman, J. (1937). 'Smooth Test' for Goodness of Fit. *Skandinaviske Aktuartidskrifte 20, 149–199*

Peng, L. (1998). Asymptotically unbiased estimators for the extreme-value index. *Statistics & Probability Letters 38, 107–115*

Serfling, R.J. (1980). Approximation Theorems of Mathematical Statistics. *John Whiley & Sons, New York*

Stone, C.J. and Koo C.-Y. (1986). Logspline density estimation. *Contemporary Mathematics 59, 1–15*