C. R. Acad. Sci. Paris, Ser. I 335 (2002) 279–282

Statistique/*Statistics*
(Probabilités/*Probability Theory*)

A new approach on estimation of the tail index

Dimitris N. Politis

Department of Mathematics, University of California–San Diego, La Jolla, CA 92093, USA

Received 21 December 2001; accepted after revision 31 May 2002

Note presented by Paul Deheuvels.

Abstract A new approach on tail index estimation is proposed based on studying the in-sample evolution of appropriately chosen diverging statistics. The resulting estimators are simple to construct, and they can be generalized to address other rate estimation problems as well. ***To cite this article: D.N. Politis, C. R. Acad. Sci. Paris, Ser. I 335 (2002) 279–282.*** © 2002 Académie des sciences/Éditions scientifiques et médicales Elsevier SAS

Une nouvelle méthode pour l'estimation de l'index d'une queue

Résumé Une nouvelle méthode est proposée pour l'estimation de l'index d'une queue de distribution. Elle est basée sur l'étude de statistiques divergentes. Les estimateurs résultants sont simples à construire et peuvent être utilisés pour résoudre d'autres problèmes d'estimation. ***Pour citer cet article : D.N. Politis, C. R. Acad. Sci. Paris, Ser. I 335 (2002) 279–282.*** © 2002 Académie des sciences/Éditions scientifiques et médicales Elsevier SAS

1. Introduction

Let $X_1, \ldots, X_n$ be an observed stretch of a linear time series satisfying $X_t = \sum_{j\in\mathbb{Z}} \psi_j Z_{t-j}$, for all $t \in \mathbb{Z}$, where $\{Z_t\}$ is i.i.d. from some distribution F, and the filter coefficients $\{\psi_j\}$ are absolutely summable; the case where $\psi_j = 0$ for $j \neq 0$ is the special case of $\{X_t\}$ being i.i.d. We assume that F belongs to $D(\alpha)$, the domain of attraction of an α-stable law; however, the heavy tail index α is unknown and must be estimated from the data. In this context, there exist sequences a_n and b_n such that $a_n^{-1}(\sum_{t=1}^n Z_t - b_n) \stackrel{\mathcal{L}}{\Longrightarrow} S_\alpha$, where S_α denotes a generic α-stable law with unspecified scale, location and skewness, and $\alpha \in (0, 2]$; recall that $a_n = n^{1/\alpha} L(n)$ for some slowly-varying function $L(\cdot)$.

Tail index estimators typically are based upon a number q of extreme order statistics; see Csörgő et al. [1] for a general class of estimators that includes many such estimators, for example, the well-known Hill estimator H_q, as special cases. A challenging problem lies in choosing the number of order statistics q to be used in practice; see e.g. Embrechts et al. [3] and the references therein.

In this report, we construct a new type of tail index estimator not necessarily based on order statistics which is analyzed in the case where $L(n)$ is constant. The new approach is very simple and intuitive, and can be easily generalized to rate estimation settings other than the heavy tail problem. We also propose some ways to improve upon the basic form of the new estimators, and give some finite-sample simulation results.

E-mail address: dpolitis@ucsd.edu (D.N. Politis).

S1631-073X(02)02450-0/FLA

2. Motivation for the new estimation approach

Define the second sample moment $S_n^2 = \frac{1}{n}\sum_{t=1}^n X_t^2$; as it turns out, S_n^2 diverges with rate a_n^2/n. Since $a_n = n^{1/\alpha}L(n)$, the rate of divergence of S_n^2 may give crucial information about α. For concreteness, consider the standard case where $a_n = n^{1/\alpha}$, i.e., $L(n)$ is constant. Define $Y_k := \log S_k^2$, and $U_k := Y_k - \gamma \log k$ for $k = 1, \ldots, n$, where $\gamma = -1 + 2/\alpha$.

LEMMA 2.1. – *Let $a_n = n^{1/\alpha}$ for some $\alpha \in (0, 2]$. Then, $U_n \stackrel{\mathcal{L}}{\Longrightarrow}$ some probability law (that depends on α), as $n \to \infty$.*

Proof. – Straightforward calculation gives $U_n = \log(n^{-2/\alpha}\sum_{t=1}^n X_t^2)$. First assume $\alpha = 2$. The fact that F is in $D(2)$ with $a_n = n^{1/2}$ implies that $EZ_t^2 < \infty$ by a theorem of Giné and Zinn [4]); thus, $EX_t^2 < \infty$ as well, and the lemma follows from a law of large numbers. Now assume $\alpha < 2$. An extension of Lemma 2 of McElroy and Politis [5] gives $a_n^{-2}\sum_{t=1}^n X_t^2 = \mathrm{o}_P(1) + \Psi_2^2 a_n^{-2}\sum_{t=1}^n Z_t^2$, with $\Psi_2^2 = \sum_{j\in\mathbb{Z}} \psi_j^2$. Since $a_n^{-2}\sum_{t=1}^n Z_t^2 \stackrel{\mathcal{L}}{\Longrightarrow} S_{\alpha/2}$ the lemma is proven. □

From Lemma 2.1 it follows that $U_n = \mathrm{O}_P(1)$. Thus, from the relation $Y_k = \gamma \log k + U_k$ it is suggested that γ could plausibly be estimated as the slope of a regression of Y_k on $\log k$, with a resulting estimator for α; note that an intercept should also be included in this regression to account for the nonzero large-sample expectation of U_n. We formalize the above by defining:

$$\hat{\alpha} = \frac{2}{\hat{\gamma} + 1}, \quad \text{and} \quad \hat{\gamma} = \frac{\sum_{k=1}^n (Y_k - \overline{Y})(\log k - \overline{\log n})}{\sum_{k=1}^n (\log k - \overline{\log n})^2}, \tag{1}$$

where $\overline{Y} = \frac{1}{n}\sum_{k=1}^n Y_k$, and $\overline{\log n} = \frac{1}{n}\sum_{k=1}^n \log k$. Our main result follows.

THEOREM 2.1. – *Let $a_n = n^{1/\alpha}$ for some $\alpha \in (0, 2]$. Then, $\hat{\alpha} \stackrel{P}{\longrightarrow} \alpha$, as $n \to \infty$.*

Proof. – Let $\overline{U} = \frac{1}{n}\sum_{k=1}^n U_k$, and note that since $Y_k = \gamma \log k + U_k$, we have

$$\hat{\gamma} = \gamma + \frac{\sum_{k=1}^n (U_k - \overline{U})(\log k - \overline{\log n})}{\sum_{k=1}^n (\log k - \overline{\log n})^2}.$$

Note that $c_1 \log n \leqslant \overline{\log n} \leqslant c_2 \log n$, and $c_3(\log n)^2 \leqslant n^{-1}\sum_{k=1}^n (\log k)^2 \leqslant c_4(\log n)^2$, for some constants $c_i > 0$; thus, from Lemma 2.1, it follows that $\hat{\gamma} = \gamma + \mathrm{O}_P(1/\log n)$, and the theorem is proven. □

Remark 2.1. – The rate of convergence and asymptotic distribution of $\hat{\alpha}$ will be investigated in a follow-up paper. Note that Eq. (1) corresponds to an L_2 regression estimator of slope. However, due to the approximately exponential tails of the large-sample distribution of Y_n, it is expected that an L_1 regression would be consistent as well, and perhaps give better performance than L_2. Also note that tail index estimators based on least-squares arguments have previously been considered in the literature (mainly in connection with order statistics); see e.g. the class of Zipf estimators analyzed in Csörgő and Viharos [2].

Remark 2.2. – The validity of the regression of Y_k on $\log k$ is based on Lemma 2.1, i.e., on the fact that the distribution of U_k tends to a limit. Thus, the $(Y_k, \log k)$ points will not be very informative if k is small because the distribution of U_k has not yet stabilized. Consequently, it may be advisable in practice to drop some points, much in the same manner as some points are dropped in a Markov Chain simulation. Thus, one would regress Y_k on $\log k$ for $k = m, \ldots, n$, for some m chosen either as some constant or even as a function of n but such that $n - m \to \infty$; doing so does not affect the asymptotic consistency of $\hat{\alpha}$.

Remark 2.3. – The general case $a_n = n^{1/\alpha}L(n)$ is quite more difficult to handle; however, the special case where $L(n) = (\log n)^{\beta}$, for some unknown $\beta \in \mathbb{R}$, can be dealt with in the same fashion as above. To

Table 1. – Entries represent empirical mean squared errors of different tail index estimators as computed from 100 Monte Carlo replications (with $n = 1000$).

	$\hat{\alpha}_T$	$\bar{\alpha}^*_{(N=10)}$	$\bar{\alpha}^*_{(N=30)}$	$\bar{\alpha}^*_{(N=50)}$	$H_{q_{\text{opt}}}$
Cauchy	0.133	0.081	0.068	0.066	0.009
1.5-stable	0.122	0.081	0.068	0.068	0.012
1.9-stable	0.037	0.030	0.030	0.029	0.031
Gaussian	0.012	0.003	0.002	0.001	0.034
Pareto(2, 1)	0.224	0.217	0.206	0.199	0.126
Burr(2, 1, 0.5)	0.183	0.073	0.056	0.053	0.032

do this, one would simply regress Y_k on $\log k$ and $\log\log k$ simultaneously, with resulting estimators for α and β.

3. Improving upon the basic estimator

In general, $\hat{\alpha}$ is not guaranteed to fall in the interval $(0, 2]$; thus, we may define the truncated estimator $\hat{\alpha}_T$ by $\hat{\alpha}_T = 0$ if $\hat{\alpha} \leqslant 0$, $\hat{\alpha}_T = \hat{\alpha}$ if $\hat{\alpha} \in (0, 2]$, and $\hat{\alpha}_T = 2$ if $\hat{\alpha} \geqslant 2$. It is apparent that $\hat{\alpha}_T$ is also consistent and more accurate than $\hat{\alpha}$.

Focusing on the case where the data $X_1, \ldots, X_n$ are i.i.d., we may consider using permutations of the data. There will be at most $n!$ such permutations; by some arbitrary criterion, choose N of them, order them in some fashion, and let $X_1^{*k}, \ldots, X_n^{*k}$ denote the kth permutation. Let $\hat{\alpha}^{*k}$ and $\hat{\gamma}^{*k}$ denote the estimators $\hat{\alpha}$ and $\hat{\gamma}$ as computed from the kth permutation. Finally, we define $\bar{\alpha}^* = \text{median}\{\hat{\alpha}^{*1}, \ldots, \hat{\alpha}^{*N}\}$, $\hat{\alpha}^* = 2/(\hat{\gamma}^* + 1)$, and $\hat{\gamma}^* = \text{median}\{\hat{\gamma}^{*1}, \ldots, \hat{\gamma}^{*N}\}$.

COROLLARY 3.1. – *Assume $a_n = n^{1/\alpha}$ for $\alpha \in (0, 2]$, and that $\psi_j = 0$ for $j \neq 0$, i.e., that the sequence $\{X_t\}$ is i.i.d. Then, $\bar{\alpha}^* \xrightarrow{P} \alpha$ and $\hat{\alpha}^* \xrightarrow{P} \alpha$, as $n \to \infty$.*

The recommendation is to take N as big as computationally feasible. Table 1 presents the results of a small simulation where i.i.d. samples of size $n = 1000$ were generated and $m = 1$ (from Remark 2.2) for concreteness. The choices for F were the symmetric α-stable distributions for $\alpha = 1, 1.5, 1.9$ and 2, as well as the Burr (a, k, τ) distribution given by $F(x) = 1 - k^a/(k + x^\tau)^a$, for $x > 0$; note that the Burr $(a, k, 1)$ is a Pareto (a, k) distribution. The simulation results are quite informative. Firstly, it was confirmed that the distribution of $\hat{\alpha}$ inherits a certain degree of heavy tails from F making a strong case in favor of the truncated estimator $\hat{\alpha}_T$. Secondly, there is a significant effect of taking even a small number N of (randomly selected) permutations—in fact, $N = 10$ yields most of the benefits; in addition, the median in $\bar{\alpha}^*$ clips all outlying values so a truncation is not necessary. In many cases $\bar{\alpha}^*$ seems roughly comparable to $H_{q_{\text{opt}}}$, i.e., the Hill estimator with an empirically optimized choice of q; see e.g. Embrechts et al. [3, Ch. 6.4]. It seems that $\bar{\alpha}^*$ may have an improved performance in cases where α is close to 2 but underperform in the other cases. Interestingly, $\bar{\alpha}^*$ does not require a fine-tuning similar to the choice of q in H_q.

4. The general rate estimation approach

The estimator $\hat{\alpha}$ was based on the fact that S_n^2 diverges at a rate depending on α. Nevertheless, this is a general approach that is not limited to the second sample moment; for example, we could look at the $2r$th sample moment for some $r \in \mathbb{N}$. We could also consider sample extrema as our diverging statistics, e.g., the maximum $M_n = \max\{X_1, \ldots, X_n\}$, or the range $K_n := M_n - L_n$, where $L_n = \min\{X_1, \ldots, X_n\}$. One could even look simultaneously at a number of such diverging statistics, obtain an estimator $\hat{\alpha}$ from each of them, and appropriately combine those estimators (say, by taking their median) to get an improved one.

As a matter of fact, the proposed ideas go beyond the heavy tail problem and indicate a remarkably general method of estimating a parameter associated with the rate by which an arbitrary statistic diverges. We outline below the general rate estimation approach; here the data $X_1, \ldots, X_n$ represent a stretch from an *arbitrary* time series $\{X_t\}$ that is not necessarily linear, nor stationary.

(a) Let $T_n = T_n(X_1, \ldots, X_n)$ be some positive statistic diverging to ∞ whose rate of divergence depends on some unknown parameter λ.

(b) Let $Y_k := \log T_k$, and $U_k := \log(k^{-g(\lambda)} T_k)$, where $g(\cdot)$ is a known invertible function, continuous at λ, and such that $E|U_n| = \mathrm{O}(1)$ as $n \to \infty$.

(c) Estimate $g(\lambda)$ by $\hat{g} = \big(\sum_{k=1}^n (Y_k - \overline{Y})(\log k - \overline{\log n})\big) / \sum_{k=1}^n (\log k - \overline{\log n})^2$, and λ by $\hat{\lambda} = g^{-1}(\hat{g})$.

THEOREM 4.1. – *If statements* (a)–(c) *are true, then* $\hat{\lambda} \xrightarrow{P} \lambda$, *as* $n \to \infty$.

References

[1] S. Csörgő, P. Deheuvels, D.M. Mason, Kernel estimates of the tail index of a distribution, Ann. Statist. 13 (1985) 1050–1077.

[2] S. Csörgő, L. Viharos, Estimating the tail index, in: B. Szyszkowicz (Ed.), Asymptotic Methods in Probability and Statistics, North-Holland, Amsterdam, 1998, pp. 833–881.

[3] P. Embrechts, C. Klüppelberg, T. Mikosch, Modelling Extremal Events, Springer, Berlin, 1997.

[4] E. Giné, J. Zinn, Necessary conditions for the bootstrap of the mean, Ann. Statist. 17 (1990) 684–691.

[5] T. McElroy, D.N. Politis, Robust inference for the mean in the presence of serial correlation and heavy tailed distributions, Econometric Theory, 2002, forthcoming.