γ_1

A new kernel estimator for the tail index under random censoring

Louiza Soltane[1], Abdelhakim Necir [2]
[1,2] Mohamed Khider University, Biskra, Algeria

Abbreviated abstract: To our knowledge, when the data are randomly censored, the kernel estimator for the tail index is not yet extreme addressed in the value literature. For this, we review the existing tail index estimators and then propose a new kernel estimator to γ_1 for censored data.

Related publications:

– Beirlant et al, Statist. Plann. Inference 202, 31–56, (2019)
– Worms, J., and Worms, R., Extremes 17, 337–358, (2014)

louiza_stat@yahoo.com

Previous work, challenge, and approach

Complete data

Let $X_1, X_2, ..., X_n$ be independent and identically distributed (iid) of non-negative random variables (rv's) as **n** copies of a rv **X**; defined over some probability space $(\Omega, \mathcal{A}, \mathbf{P})$ with cumulative distribution function (cdf) **F**: We assume that the distribution tail $\overline{F} := 1 - F$ is regularly varying at infinity, with index $(-1/\gamma_1)$ notation: $\overline{F} \in \mathcal{RV}_{(-1/\gamma_1)}$ that is

$$\lim_{t\to\infty} \frac{\overline{F}(tx)}{\overline{F}(t)} = x^{-1/\gamma_1}, \text{ for any } x > 0, \quad (1.1)$$

where $\gamma_1 > 0$ is called the shape parameter or the tail index or the extreme value index (EVI).

- Hill's estimator Hill (1975) →
$$\hat{\gamma}_{1,k}^{(H)} := \frac{1}{k}\sum_{i=1}^{k} \log \frac{X_{n-i+1:n}}{X_{n-k:n}} = \sum_{i=1}^{k} \frac{i}{k} \log \frac{X_{n-i+1:n}}{X_{n-i:n}}, \quad (1.2)$$

- Csörgo et al. (1985) →
$$\hat{\gamma}_{1,k}^{(CDM)}(K) := \sum_{i=1}^{k} \frac{i}{k+1} K\left(\frac{i}{k+1}\right) \log \frac{X_{n-i+1:n}}{X_{n-i:n}},$$

where K is a kernel function satisfying the following assumptions:

- [A1] is non increasing and right-continuous on $\mathbb{R}$.
- [A2] $K(s) = 0$ for $s \notin (0, 1]$ and $K(s) \geq 0$ for $s \in (0, 1]$.
- [A3] $\int_{\mathbb{R}} K(s)\, ds = 1$.
- [A4] K and its first and second Lebesgue derivatives K' and K'' are bounded.

Techniques and Methods

Censored data

In many real situations the variable of interest **X** is not always available. An appropriate way to model this matter, is to introduce a non-negative rv **Y**; called censoring rv, independent of **X** and then to consider the rv $Z := \min(X, Y)$ and the indicator variable $\delta := \mathbf{1}\{X \leq Y\}$, which determines whether or not **X** has been observed.

A new kernel estimator for γ_1

1. By using Potter's inequalities, see e.g. Proposition B.1.10 in de Haan and Ferreira (2006), to the regularly varying function **F** together with assumptions **[A1]- [A4]**;

$$\lim_{u\to\infty} \int_u^\infty g_K'\left(\frac{\overline{F}(x)}{\overline{F}(u)}\right) \log\frac{x}{u} d\frac{F(x)}{\overline{F}(u)} = \gamma_1 \int_0^\infty K(x)\,dx = \gamma_1.$$

   Where $g_K(x) := xK(x)$ and g' denotes the Lebesgue derivative of **g**.

2. By letting $u = Z_{n-k:n}$ and substituting **F** by Kaplan-Meier estimator

$$\overline{F}_n^{KM}(t) := \begin{cases} \prod_{i=1}^{n}\left(1 - \frac{\delta_{(i)}}{n-i+1}\right)^{\mathbf{1}\{Z_{i:n}\leq t\}} & \text{if } t < Z_{n:n} \\ 0 & \text{otherwise} \end{cases},$$

Results and Conclusion

3. We derive a kernel estimator to the tail index γ_1 defined by

$$\tilde{\gamma}_{1,k}(K) := \int_{Z_{n-k:n}}^{\infty} g_K'(\mathcal{F}_n(x)) \log\left(\frac{x}{Z_{n-k:n}}\right) \frac{dF_n^{KM}(x)}{\overline{F}_n^{KM}(Z_{n-k:n})},$$

where

$$\mathcal{F}_n(x) := \theta_{x,n} \frac{\overline{F}_n^{KM}(x^-)}{\overline{F}_n^{KM}(Z_{n-k:n})} + (1-\theta_{x,n}) \frac{\overline{F}_n^{KM}(x)}{\overline{F}_n^{KM}(Z_{n-k:n})},$$

and $\theta_{x,n} := \theta_{x,Z_{n-k:n}}$ (arbitrary):

4. In view of the mean value theorem, we may choose the sequence of constants $\theta_{j,n}$, and recall that $g_K(x) = xK(x)$, we end up to the final form of our new kernel estimator given by

$$\tilde{\gamma}_{1,k}(K) = \sum_{j=1}^{k} \frac{\overline{F}_n^{KM}(Z_{n-j:n})}{\overline{F}_n^{KM}(Z_{n-k:n})} K\left(\frac{\overline{F}_n^{KM}(Z_{n-j:n})}{\overline{F}_n^{KM}(Z_{n-k:n})}\right) \log \frac{Z_{n-j+1:n}}{Z_{n-j:n}}.$$

We proposed a smoothed (or a kernel) version of Worms's estimator (Worms and Worms, 2014) of the tail index of a Pareto-type distribution for randomly censored data. This estimator is a generalization of the well-known kernel estimator of the extreme value index for complete data introduced by Csörgo et al. (1985).