Avoiding Threshold Selection in Extremal Inference

Gennady Samorodnitsky
Cornell University
ORIE
gs18@cornell.edu

Julian Sun
Cornell University
ORIE
ys598@cornell.edu

12 May 2017

Threshold Selection

- Much of extreme value analysis involves extracting values over certain thresholds
- Idea of “optimal threshold”
- Problems with “optimality”
  - Optimal threshold in terms of marginal distribution
  - Observations below threshold not used
  - Estimators often sensitive to thresholds

Threshold Selection

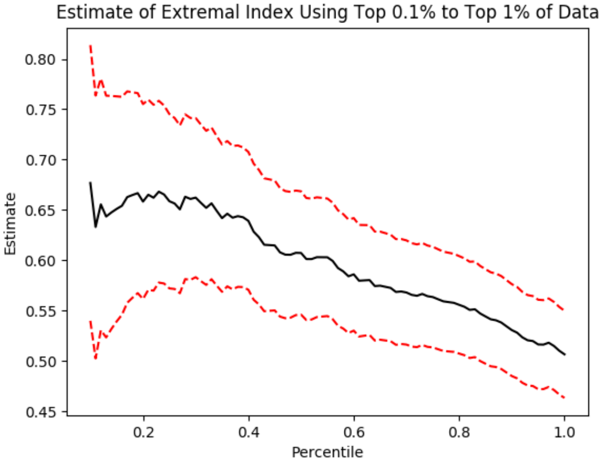

Figure: True value $\theta = 3/4$

Approach to Avoiding Thresholds

- Using data below optimal threshold;
- Giving less weight to smaller observations;
- Achieve reduced asymptotic variance, maintain small bias.

Extremal Index and Blocks Estimator

Take stationary sequence (X_i), for (u_n) satisfying $n\overline{F}(u_n) \to \tau$ for some $\tau > 0$, the exceedance process

$$N_n = \sum_{i=1}^{n} \epsilon_{i/n} I_{\{X_i > u_n\}} \to_d N = \sum_{i=1}^{\infty} \xi_i \epsilon_{\Gamma_i},$$

where (Γ_i) is a Poisson process with intensity $\theta\tau$, and (ξ_i) are the distributions of extremal cluster sizes.

- Extremal index: θ is the reciprocal of expected value of ξ_i
- Blocks estimator: exceedance per block among blocks with at least one exceedance

Blocks Estimator with Threshold u_n

Divide n observations into contiguous blocks of size $r_n = o(n)$, for suitable u_n, take

$$\hat{M}_n = \# \text{ of blocks with exceedance over } u_n,$$
$$\hat{\tau}_n = \# \text{ of exceedances over } u_n,$$

then blocks estimator $\hat{\theta}_n = \hat{M}_n / \hat{\tau}_n$.

If $h_n = o(n)$ with $r_n = o(n/h_n)$, $\frac{n}{h_n}\overline{F}(u_n) \to \tau$ for some τ, then $\hat{\theta}_n$ is consistent, and

$$\sqrt{h_n}(\hat{\theta}_n - b_n - \theta) \to_d \mathcal{N}(0, \frac{\theta(\theta^2 m_2 - 1)}{\tau}),$$

where b_n is the bias, and m_2 is second moment of cluster size distribution.

Using Observations below u_n

Take $u_n = u_n^1 > u_n^2 > \cdots > u_n^m$, and

$$\hat{M}_n(u_n^s) = \#\text{ of blocks with exceedance over } u_n^s,$$
$$\hat{\tau}_n(u_n^s) = \#\text{ of exceedances over } u_n^s,$$

use estimator

$$\tilde{\theta}_n = \frac{a_1 \hat{M}_n(u_n^1) + \cdots + a_m \hat{M}_n(u_n^m)}{a_1 \hat{\tau}_n(u_n^1) + \cdots + a_m \hat{\tau}_n(u_n^m)}$$

where $a_1, \ldots, a_m > 0$.

Similar consistency and asymptotic normality conditions hold under weak dependence assumptions.

Bias Correction

Under mild assumptions, bias term b_n is asymptotically linear in τ:

$$b_n \sim c_n \tau + \zeta_n,$$

where $E\zeta_n = 0$ and c_n is a function of r_n, h_n, θ.

One can use linear regression to estimate the bias.

Threshold Selection

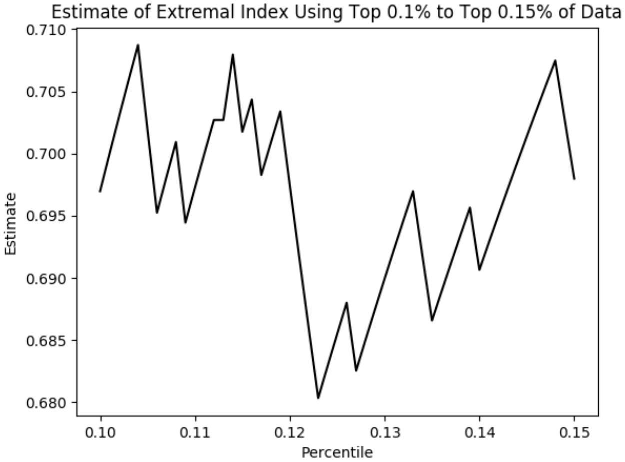

Figure: True value $\theta = 3/4$

Threshold Selection

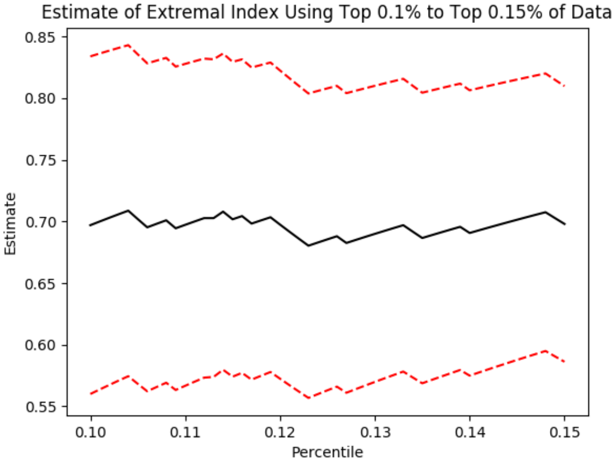

Figure: True value $\theta = 3/4$

Same Example

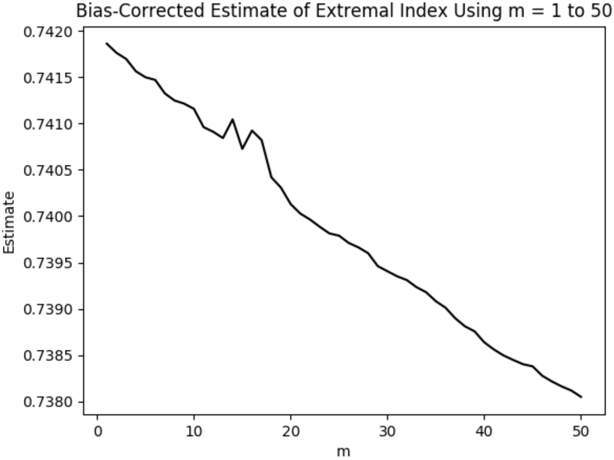

Figure: True value $\theta = 3/4$

Same Example

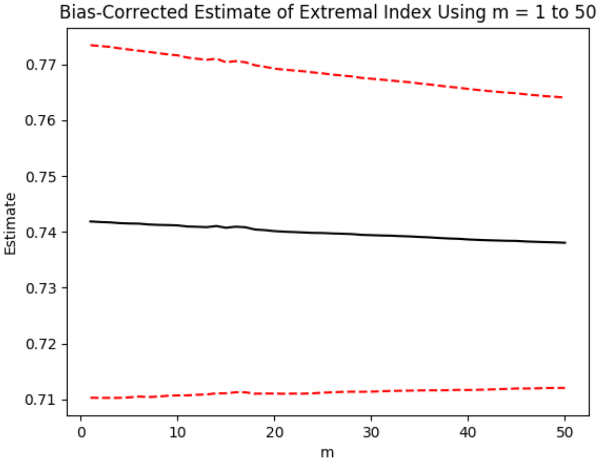

Figure: True value $\theta = 3/4$