The high order moments method in endpoint estimation: an overview

Gilles STUPFLER (Aix Marseille Université)
Joint work with Stéphane GIRARD (INRIA Rhône-Alpes) and Armelle GUILLOU (Université de Strasbourg)

Statistics Seminar
Aveiro, Portugal, January 14^{th}, 2015

Outline

The context

We assume that Y is a univariate random variable, with an unknown finite right endpoint θ:

$$\theta := \sup\{y \in \mathbb{R} \,|\, \mathbb{P}(Y \leq y) < 1\} < \infty.$$

We address the problem of estimating θ, which can be thought of as the maximal value that Y can take.

Practical relevance:

- Maximal temperature/wind speed at a given location;
- Best performances in athletics;
- Maximal life span in biology;
- Maximal production level/output given an input in economics.

Some existing methods

- Maximum estimator and improvements: Quenouille (1949), Miller (1964), Robson and Whitlock (1964), Cooke (1979), de Haan (1981).
- Maximum likelihood estimator: Hall (1982), Li and Peng (2009).
- Threshold estimators:
  - Deduced from extreme quantile estimators: Hosking and Wallis (1987), Smith (1987), Dekkers *et al.* (1989);
  - Specific estimators: Fraga Alves *et al.* (2013), Fraga Alves and Neves (2014).
- Miscellaneous:
  - Bootstrap method: Loh (1984) and Athreya and Fukuchi (1997);
  - Minimal-distance method: Hall and Wang (1999);
  - With random errors: Goldenshluger and Tsybakov (2004);
  - Bayesian likelihood: Hall and Wang (2005);
  - Censored likelihood: Li *et al.* (2011);
  - Empirical likelihood: Li *et al.* (2011b).

Basic idea

Assume first that Y is a positive random variable whose survival function has a polynomial decay when $y \uparrow \theta$:

$$\forall y \in [0, \theta],\ \overline{F}(y) := \mathbb{P}(Y > y) = \left(1 - \frac{y}{\theta}\right)^{\alpha},$$

where $\alpha > 0$. Then if $\mu_p := \mathbb{E}(Y^p)$, we have:

$$\forall p \geq 1,\ \frac{\mu_p}{\mu_{p+1}} = \frac{1}{\theta}\left(1 + \frac{\alpha}{p+1}\right)$$

and therefore θ can be obtained by combining a finite number of moments of the random variable Y.

Why high order moments?

Suppose now that Y is positive and has a finite right endpoint θ, without any further assumption.

Proposition 1

It holds that $\mu_p/\mu_{p+1} \to 1/\theta$ *as* $p \to \infty$.

Thus, while in this context we cannot give an expression for μ_p, we recover that μ_p/μ_{p+1} converges to $1/\theta$ as $p \to \infty$.

$\Rightarrow$ in the general case, the moments of interest (for our problem) are the high order moments of Y.

Endpoint estimation: with positive random variables

If $Y_1, \ldots, Y_n$ are i.i.d. copies of the positive random variable Y, and

$$\forall q \geq 1,\ \widehat{\mu}_q = \frac{1}{n} \sum_{k=1}^{n} Y_k^q$$

is the standard empirical version of μ_q, then recalling Proposition 1, a first estimator $\widetilde{\theta}_n$ of θ is obtained by setting

$$\widetilde{\theta}_n = \frac{\widehat{\mu}_{p_n+1}}{\widehat{\mu}_{p_n}}$$

where (p_n) is a sequence of real numbers ≥ 1 tending to infinity.

This choice however, is not a nice one. A reason why is the following: when

$$\forall y \in [0,\theta],\ \overline{F}(y) = \left(1 - \frac{y}{\theta}\right)^{\alpha},$$

then

$$\frac{\mu_{p+1}}{\mu_p} = \theta\left(1 - \frac{\alpha}{p} + \mathrm{o}\left(\frac{1}{p}\right)\right) \quad \text{as} \quad p \to \infty,$$

leading to (among others...) a $\mathrm{O}(1/p)$ bias in the estimation, which may be severe if α is large.

A better choice can be made by noting that in this case

$$\forall a > 0,\ \frac{1}{\theta} = \frac{1}{ap}\left[((a+1)p+1)\frac{\mu_{(a+1)p}}{\mu_{(a+1)p+1}} - (p+1)\frac{\mu_p}{\mu_{p+1}}\right].$$

Back to the general case, Proposition 1 yields:

$$\forall a > 0,\ \frac{1}{\theta} = \lim_{p \to \infty} \frac{1}{ap} \left[((a+1)p + 1) \frac{\mu_{(a+1)p}}{\mu_{(a+1)p+1}} - (p+1) \frac{\mu_p}{\mu_{p+1}} \right].$$

This formula is the basis for the estimator $\widehat{\theta}_n$ defined by:

$$\frac{1}{\widehat{\theta}_n} = \frac{1}{ap_n} \left[((a+1)p_n + 1) \frac{\widehat{\mu}_{(a+1)p_n}}{\widehat{\mu}_{(a+1)p_n+1}} - (p_n+1) \frac{\widehat{\mu}_{p_n}}{\widehat{\mu}_{p_n+1}} \right]$$

for some $a > 0$ and a sequence (p_n) of real numbers ≥ 1 tending to infinity.

Asymptotic properties: weak consistency

To study the asymptotic properties of $\widehat{\theta}_n$, the first step is to obtain a weak law of large numbers for the empirical high order moment $\widehat{\mu}_{p_n}$.

Lemma 1 (WLLN, first endpoint estimator)

If $n\theta^{-p_n}\mu_{p_n} \to \infty$ *then* $\widehat{\mu}_{p_n}/\mu_{p_n} \xrightarrow{\mathbb{P}} 1$.

A straightforward corollary is the consistency of our estimator:

Theorem 1 (Consistency, first endpoint estimator)

If $n\theta^{-(a+1)p_n}\mu_{(a+1)p_n} \to \infty$ *then* $\widehat{\theta}_n \xrightarrow{\mathbb{P}} \theta$.

Asymptotic properties: strong consistency

A strong asymptotic result for the estimator is obtained in the same way: we start by a strong law of large numbers for the empirical high order moment $\widehat{\mu}_{p_n}$.

Lemma 2 (SLLN, first endpoint estimator)

If $n\theta^{-p_n}\mu_{p_n}/\log n \to \infty$ *then* $\widehat{\mu}_{p_n}/\mu_{p_n} \to 1$ *almost surely.*

We use this result to get the strong consistency of our estimator:

Theorem 2 (Strong consistency, first endpoint estimator)

If $n\theta^{-(a+1)p_n}\mu_{(a+1)p_n}/\log n \to \infty$ *then* $\widehat{\theta}_n \to \theta$ *almost surely.*

Asymptotic properties: asymptotic normality

To obtain the asymptotic normality of our estimator, we assume that $\overline{F}$ decays roughly like a polynomial in a neighborhood of θ:

(H) There is $\gamma < 0$ such that

$$\overline{F}(y) = (1 - y/\theta)^{-1/\gamma} L((1 - y/\theta)^{-1}) \text{ as } y \uparrow \theta$$

with L being a Borel slowly varying function at infinity; more precisely, we assume that

$$L(y) = \exp\left(\int_1^y \frac{\eta(t)}{t} dt\right)$$

where η is a continuously differentiable, ultimately monotonic function tending to 0 at infinity, such that $|\eta'|$ is regularly varying at infinity. Besides, $y\eta'(y)/\eta(y) \to \nu \leq 0$ as $y \to \infty$.

Condition **(H)** is a second-order condition which makes it possible to control the bias of our estimator.

Theorem 3 (Asymptotic normality, first endpoint estimator)

If $np_n^{1/\gamma}L(p_n) \to \infty$ *and* $np_n^{1/\gamma}L(p_n)\eta^2(p_n) \to 0$ *then*

$$\sqrt{np_n^{2+1/\gamma}L(p_n)}\left(\frac{\widehat{\theta}_n}{\theta} - 1\right) \xrightarrow{d} \mathcal{N}\left(0, V(\gamma, a)\right)$$

where $V(\gamma, a)$ *is known.*

The optimal rate of convergence of the estimator $\widehat{\theta}_n$ can actually be shown to be essentially the same as that of classical threshold estimators.

Endpoint estimation: general case

When Y is not assumed to be positive anymore, we may consider instead the random variable $Z = e^Y$, which is positive and has right endpoint e^θ. Let then

$$\forall p \geq 1, \; m_p = \mathbb{E}(Z^p) = \mathbb{E}(e^{pY})$$

and define a (slightly different) estimator $\widehat{\theta}_n$ by

$$\frac{1}{\widehat{\theta}_n} = \frac{1}{a}\left(\log\left[\frac{\widehat{\mu}_{p_n}}{\widehat{\mu}_{p_n+1}}\right] - \log\left[\frac{\widehat{\mu}_{(a+1)p_n}}{\widehat{\mu}_{(a+1)p_n+a+1}}\right]\right)$$

where again $a > 0$ and (p_n) is a sequence of real numbers ≥ 1 tending to infinity.

Asymptotic properties in the general case

The consistency properties of this estimator follow by the same type of arguments:

Theorem 4 (Consistency, second endpoint estimator)

If $ne^{-(a+1)\theta p_n} m_{(a+1)p_n} \to \infty$ *then* $\widehat{\theta}_n \xrightarrow{\mathbb{P}} \theta$.

Theorem 5 (Strong consistency, second endpoint estimator)

If $ne^{-(a+1)\theta p_n} m_{(a+1)p_n} / \log n \to \infty$ *then* $\widehat{\theta}_n \to \theta$ *almost surely.*

Assuming that $\overline{F}$ has a polynomial decay as $y \uparrow \theta$, *i.e.*

$$\overline{F}(y) = (\theta - y)^{-1/\gamma} L((\theta - y)^{-1})$$

with $\gamma < 0$ and L a "nice enough" Borel slowly varying function at infinity as in **(H)**, we have the following asymptotic normality result:

Theorem 6 (Asymptotic normality, second endpoint estimator)

If $np_n^{1/\gamma} L(p_n) \to \infty$ *and* $np_n^{1/\gamma} L(p_n)\eta^2(p_n) \to 0$ *then*

$$\sqrt{np_n^{2+1/\gamma} L(p_n)} \left(\widehat{\theta}_n - \theta \right) \xrightarrow{d} \mathcal{N}\left(0, V(\gamma, a)\right)$$

where $V(\gamma, a)$ *is known and is the same as in Theorem 3.*

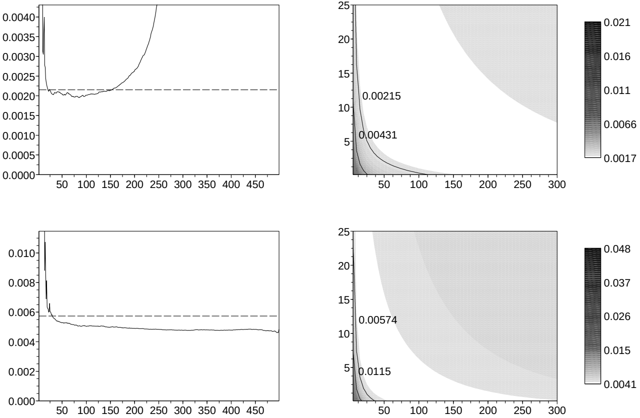

Figure 1: Mean L^1-error, case $\theta = 0$. Left: extreme-value moment estimator. Right: our estimator, $x-$axis: p_n, $y-$axis: a. Top: Burr model, bottom: transformed log-gamma model. Computations carried out on 1000 replications of a sample of size 500.

Frontier estimation

In practice, it often happens that Y is recorded along with a covariate X:

- Temperature/wind speed along with 2D/3D coordinates;
- Performances in athletics along with age;
- Life span along with socioeconomic status;
- Production level along with input.

The problem then becomes the estimation of the conditional right endpoint of Y given $X = x$.

To be more specific, assume that (X, Y) is a random pair whose distribution has a support

$$S = \{(x, y) \in \Omega \times \mathbb{R} \,|\, 0 \leq y \leq g(x)\}$$

where

- X has a pdf f on the compact subset Ω of $\mathbb{R}^d$ having nonempty interior;
- g is a positive Borel function on Ω.

The problem we consider is the pointwise estimation of the frontier function g of S, given i.i.d. replications of (X, Y).

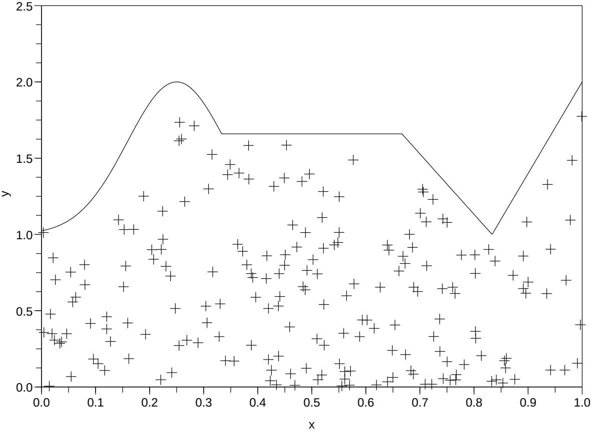

Figure 2: Frontier g (solid line), data points (+). The sample size is 200.

Some existing methods

- Extreme-value based estimators: Gardes (2002), Geffroy (1964), Girard and Jacob (2003a, 2003b, 2004), Girard and Menneteau (2005), Menneteau (2008);
- Piecewise polynomial estimators: Korostelev and Tsybakov (1993), Korostelev *et al.* (1995), Härdle *et al.* (1995);
- Projection estimators: Jacob and Suquet (1995).

Under the further hypothesis that g is nondecreasing and concave (think about the input-output example!):

- DEA/FDH estimators and improvements: Deprins *et al.* (1984), Farrell (1957), Gijbels *et al.* (1999);
- Robust estimators: Aragon *et al.* (2005), Cazals *et al.* (2002), Daouia and Simar (2005), Daouia *et al.* (2012);
- With random noise: local maximum likelihood estimators by Aigner *et al.* (1976), kernel generalizations by Fan *et al.* (1996), Kumbhakar *et al.* (2007), Simar and Zelenyuk (2011).

The estimator

We can generalize to this context our (first) endpoint estimator $\widehat{\theta}_n$.

To this end, remark that if $\mu_p(x) := \mathbb{E}(Y^p|X = x)$ is the conditional pth order moment of Y given $X = x$, then clearly

$$\frac{1}{g(x)} = \lim_{p\to\infty} \frac{1}{ap}\left[((a+1)p+1)\frac{\mu_{(a+1)p}(x)}{\mu_{(a+1)p+1}(x)} - (p+1)\frac{\mu_p(x)}{\mu_{p+1}(x)}\right]$$

for any $a > 0$.

The question is then:

How to estimate a conditional high order moment $\mu_p(x)$, $p \to \infty$?

Let (p_n) be a sequence of numbers ≥ 1 tending to infinity and let K be a kernel function, *i.e.* a bounded pdf on $\mathbb{R}^d$ with support included in the unit Euclidean ball of $\mathbb{R}^d$.

- We first approximate $\mu_{p_n}(x)$ by its smoothed counterpart

$$M_{p_n,h_n}(x) := \frac{1}{h_n^d}\mathbb{E}\left[Y^{p_n} K\left(\frac{x-X}{h_n}\right)\right],$$

  where the bandwidth h_n defines a positive nonrandom sequence which converges to 0;

- Estimate $\mu_{p_n}(x)$ by the empirical counterpart of $M_{p_n,h_n}(x)$,

$$\widehat{\mu}_{p_n,h_n}(x) := \frac{1}{nh_n^d}\sum_{k=1}^{n} Y_k^{p_n} K\left(\frac{x-X_k}{h_n}\right).$$

Our estimator is then the quantity $\widehat{g}_n(x)$ defined by the equality

$$\frac{ap_n}{\widehat{g}_n(x)} = ((a+1)p_n+1)\frac{\widehat{\mu}_{(a+1)p_n,h_n}(x)}{\widehat{\mu}_{(a+1)p_n+1,h_n}(x)} - (p_n+1)\frac{\widehat{\mu}_{p_n,h_n}(x)}{\widehat{\mu}_{p_n+1,h_n}(x)}.$$

It can be seen as a kernel estimator of $g(x)$.

Kernel estimators are generally easy to study and they inherit the asymptotic properties of their equivalents when there is no covariate. They are however very sensitive to the choice of the parameter h_n:

- choosing a small h_n leads to a very wiggly estimator whose variance is high;
- choosing a large h_n leads to an oversmoothed estimator having a large bias.

Of course, some regularity conditions are needed to ensure that the bias introduced by taking into account the observations whose covariate is near x can be controlled.

To make things simpler, assume in what follows that we work in the parametric setting

(P) $\forall y \in [0, g(x)]$, $\overline{F}(y|x) = (1 - y/g(x))^{-1/\gamma(x)}$, with $\gamma(x) < 0$.

The condition we require is then

(A) f, g and γ are positive and Hölder continuous with respective exponents η_f, η_g and η_α.

Asymptotic properties of the frontier estimator

Theorem 7 (Pointwise consistency, frontier estimator)

Assume that **(P)** *and* **(A)** *hold. Pick* x *in the interior of* Ω. *If* $np_n^{1/\gamma(x)} h_n^d \to \infty$ *and* $p_n\, h_n^{\eta_g} \to 0$, *then* $\widehat{g}_n(x) \xrightarrow{\mathbb{P}} g(x)$.

Theorem 8 (Asymptotic normality, frontier estimator)

Assume that **(P)** *and* **(A)** *hold. Pick* x *in the interior of* Ω. *If* $np_n^{1/\gamma(x)} h_n^d \to \infty$, $np_n^{2+1/\gamma(x)} h_n^{d+2\eta_g} \to 0$ *and* $np_n^{1/\gamma(x)} h_n^{d+2\eta_\alpha} \to 0$, *then*

$$\sqrt{np_n^{2+1/\gamma(x)} h_n^d} \left(\frac{\widehat{g}_n(x)}{g(x)} - 1 \right) \xrightarrow{d} \mathcal{N}\left(0, \frac{\int_B K^2}{f(x)} V(\gamma(x), a)\right)$$

where $V(\gamma, a)$ *is known and is the same as in Theorem 3.*

Uniform consistency results on any compact subset E of the interior of Ω are also available:

Theorem 9 (Uniform strong consistency, frontier estimator)

Assume that **(P)** *and* **(A)** *hold. Let*

$$\overline{\gamma} = \sup_{x \in E} \gamma(x).$$

If $n p_n^{1/\overline{\gamma}} h_n^d / \log n \to \infty$ *and* $p_n\, h_n^{\eta_g} \to 0$ *as* $n \to \infty$*, then*

$$\sup_{x \in E} |\widehat{g}_n(x) - g(x)| \to 0 \quad \textit{almost surely.}$$

The $\log n$ factor in the hypothesis of Theorem 9 is classical when considering uniform convergence results. It basically comes from the application of exponential inequalities to control the large deviations of $\widehat{g}_n(x) - g(x)$ uniformly in $x \in E$.

Theorem 10 (Rate of uniform strong consistency, frontier estimator)

Assume that **(P)** *and* **(A)** *hold. If* $np_n^{1/\overline{\gamma}} h_n^d / \log n \to \infty$, $np_n^{2+1/\overline{\gamma}} h_n^{d+2\eta_g} / \log n \to 0$ *and* $np_n^{1/\overline{\gamma}} h_n^{d+2\eta_\alpha} / \log n \to 0$, *then*

$$\sup_{x \in \Omega} \left| \sqrt{np_n^{2+1/\gamma(x)} h_n^d} (\widehat{g}_n(x) - g(x)) \right| = O(\log n) \quad \textit{almost surely.}$$

The rate of uniform convergence of the estimator on E is thus the infimum over E of the rate of pointwise convergence, provided the conditions of Theorem 8 are "uniformly satisfied" on E.

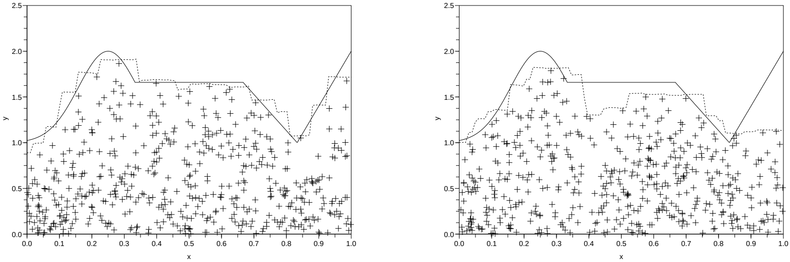

Figure 3: Case $\gamma(x) = -2[2.5 + |\cos(2\pi x)|]^{-1}$: frontier g (solid line), high-order moments estimate $\widehat{g}_n$ (dotted line). Left: best situation, right: worst situation for 500 replications of a sample of size 500.

Conclusion

- We study the asymptotic behavior of high order moments of a positive random variable having a finite right endpoint;
- We then construct an estimator of the right endpoint, which we extend to tackle the case of a random variable not being necessarily positive;
- A kernel regression makes it possible to estimate the frontier function of a covariate-and-response pair;
- The asymptotic and finite-sample properties of our estimators are generally satisfying.

Forthcoming studies

- To study the asymptotic properties of our endpoint estimators in the Gumbel extreme-value domain of attraction;
- To construct and study high order moments estimators of the extreme-value index γ in the general max-domain of attraction;
- To develop and test data-driven choices of the parameters involved in a wide array of situations;
- To develop outlier-resistant high order moments procedures and compare them with the basic ones on real data examples.

References: endpoint estimation

High-order moments estimator in the positive case:

Girard, S., Guillou, A., Stupfler, G. (2012) Estimating an endpoint with high-order moments, *TEST* **21**: 697–729.

High-order moments estimator in the general case:

Girard, S., Guillou, A., Stupfler, G. (2012) Estimating an endpoint with high order moments in the Weibull domain of attraction, *Statistics and Probability Letters* **82**: 2136–2144.

References: frontier estimation

High-order moments estimator for a frontier function:

Girard, S., Guillou, A., Stupfler, G. (2013) Frontier estimation with kernel regression on high order moments, *Journal of Multivariate Analysis* **116**: 172–189.

Uniform asymptotic properties for the high-order moments frontier estimator:

Girard, S., Guillou, A., Stupfler, G. (2014) Uniform strong consistency of a frontier estimator using kernel regression on high order moments, *ESAIM: Probability and Statistics* **18**: 642–666.

Thanks for listening!