Two applications of max–stable distributions: Random sketches and Heavy–tail exponent estimation

Stilian Stoev (sstoev@umich.edu)
Department of Statistics
University of Michigan

Boston University Probability Seminar
March 1, 2007

Based on joint works with Marios Hadjieleftheriou, George Kollios, George Michailidis and Murad Taqqu

Outline:

• Max–stable distributions: background

• Max–stable sketches: norm, distance and dominance norms

• Max–spectrum: heavy tail exponent estimation

Max–stable distributions

Th *(Fisher-Tippett-Gnedenko) Let* $X_1, \ldots, X_n, \ldots$ *be iid. If*

$$\frac{1}{a_n}(X_1 \vee \cdots \vee X_n) - b_n \xrightarrow{d} Z, \quad (n \to \infty) \tag{1}$$

then the non–degenerate Z *has one of the three types of distributions:*

(i) Frechet: $G_\alpha(x) := \exp\{-x^{-\alpha}\},\ x > 0$, *with* $\alpha > 0$.

(ii) Negative Frechet: $\Psi_\alpha(x) := \exp\{-(-x)^\alpha\},\ x < 0$, *with* $\alpha > 0$.

(iii) Gumbel: $\Lambda(x) := \exp\{-e^{-x}\},\ x \in \mathbb{R}$.

- Analog of the CLT when $\vee$ replaces $+$.
- Roughly: the tails of X_i determine the type of limit.

Def A rv X is max–stable if, $\forall a, b > 0$, $\exists c > 0, d$:

$$aX' \vee bX'' \stackrel{d}{=} cX + d,$$

where X, X', X'' iid.

Th *The limit* Z *in (1) are max–stable. Conversely any max–stable* Z *appears in a limit of iid maxima as in (1).*

More on Frechet distributions

Def Z is $\alpha-$Fréchet ($\alpha > 0$) if

$$\mathbb{P}\{Z \leq x\} = \exp\{-cx^{-\alpha}\}, \quad x > 0, \quad (c > 0).$$

Properties

- *scale family:* For all $a > 0$,

$$\mathbb{P}\{aZ \leq x\} = \exp\{-(a^{\alpha}c)x^{-\alpha}\}, \quad x > 0$$

- *Notation:* For the scale $c^{1/\alpha}$:

$$\|Z\|_{\alpha} := c^{1/\alpha}, \quad \textit{so} \quad \|aZ\|_{\alpha} = a\|Z\|_{\alpha},\ a > 0.$$

- *heavy tails:* As $x \to \infty$,

$$\mathbb{P}\{Z > x\} = 1 - e^{-cx^{-\alpha}} \sim cx^{-\alpha}.$$

- *moments:* $\mathbb{E}Z^p < \infty$ iff $p < \alpha$ and

$$\mathbb{E}Z^p = \int_0^{\infty} x^p de^{-cx^{-\alpha}} = c^p\Gamma(1 - p/\alpha).$$

Max–stability: For iid $Z, Z_1, \ldots, Z_n$:

$$Z_1 \vee \cdots \vee Z_n \stackrel{d}{=} n^{1/\alpha} Z.$$

Indeed:

$$\mathbb{P}\{Z_i \leq x,\ i = 1, \ldots, n\} = \mathbb{P}\{Z \leq x\}^n = e^{-ncx^{-\alpha}}, \quad x > 0.$$

Application I: Random Sketches

First, mathematical concepts, then explanation & use.

Max–stable sketches

'Signal': $f(i) \geq 0$, $i = 1, \ldots, n$, with *large* n.

Def For $\alpha > 0$, and $k \in \mathbb{N}$, let

$$E_j(f) := \bigvee_{i=1}^{n} f(i) Z_j(i), \quad 1 \leq j \leq k,$$

for iid *standard* $\alpha-$Fréchet $Z_j(i)$'s

$$\mathbb{P}\{Z_j(i) \leq x\} = \exp\{-x^{-\alpha}\}, \quad x > 0 \;\; i.e. \; \|Z_j(i)\|_\alpha = 1.$$

- *Max–stable sketch of* f: $E_j(f)$, $1 \leq j \leq k$.
- *Idea:* use $\{E_j(f)\}$ with $k << n$ as a proxy to f!
- More on this later...

Properties

∘ $E_j(f)$ are $\alpha-$Fréchet with

$$\|E_j(f)\|_\alpha = (\sum_{i=1}^{n} f(i)^\alpha)^{1/\alpha} =: \|f\|_{\ell_\alpha}.$$

Indeed, for $x > 0$:

$$\mathbb{P}\{E_j(f) \leq x\} = \prod_i \mathbb{P}\{Z \leq x/f(i)\} = \exp\{-(\sum_i f(i)^\alpha) x^{-\alpha}\}.$$

∘ *max–linearity*: For $a, b > 0$ and another 'signal' $g(i)$

$$E_j(af \vee bg) = aE_j(f) \vee bE_j(g), \quad 1 \leq j \leq k,$$

where $(af \vee bg)(i) = af(i) \vee bg(i)$.

Max–stable sketches: estimation

Observations

- The $E_j(f)$'s are iid in $j = 1, \ldots, k$
- The scale $\|E_j(f)\|_\alpha$ equals the norm $\|f\|_{\ell_\alpha}$.

Estimators:

- *(moment)* For $0 < p < \alpha$, define

$$L_p(f) := \left(\frac{c_{p,\alpha}}{k} \sum_{j=1}^{k} E_j(f)^p \right)^{1/p},$$

with $c_{p,\alpha} := \Gamma(1 - p/\alpha)^{-1}$.

Motivation:

$$\frac{1}{k} \sum_j E_j(f)^p \approx \mathbb{E} E_j(f)^p = \Gamma(1 - p/\alpha) \|E_j(f)\|_\alpha^p.$$

- *(median)* Define,

$$M(f) := \ln(2)^{1/\alpha} \text{med}\{E_j(f),\ 1 \le j \le k\}.$$

Motivation:

$$e^{-x^\alpha} = 1/2 \quad \Leftrightarrow \quad x = \ln(2)^{-1/\alpha}.$$

Norms and distances

Given the max–stable sketch $\{E_j(f)\}$ of a 'signal' f.

- For 'large' k's, we can estimate the **norm**:

$$\|f\|_{\ell_\alpha} \approx L_p(f) \quad \text{or} \quad \|f\|_{\ell_\alpha} \approx M(f).$$

- Performance guarantees later...

- Given another 'signal' g, define the **distance**:

$$\rho_\alpha(f,g) := \sum_{i=1}^{n} |f(i)^\alpha - g(i)^\alpha|.$$

Note that $|a-b| = 2(a \vee b) - a - b$, $a, b \geq 0$, so:

$$\rho_\alpha(f,g) = 2\sum_i f(i)^\alpha \vee g(i)^\alpha - \sum_i f(i)^\alpha - \sum_i g(i)^\alpha,$$

and hence:

$$\rho_\alpha(f,g) = 2\|f \vee g\|_{\ell_\alpha}^\alpha - \|f\|_{\ell_\alpha}^\alpha - \|g\|_{\ell_\alpha}^\alpha.$$

- Thus, given the sketches $E_j(f)$ and $E_j(g)$, we can estimate $\rho_\alpha(f,g)$ by

$$\rho_\alpha(f,g) \approx 2L_p(f \vee g)^\alpha - L_p(f)^\alpha - L_p(g)^\alpha$$

or by

$$\rho_\alpha(f,g) \approx 2M(f \vee g)^\alpha - M(f)^\alpha - M(g)^\alpha$$

- The key is: $E_j(f \vee g) = E_j(f) \vee E_j(g)$ so $E_j(f \vee g)$ is available from $E_j(f)$ and $E_j(g)$.

Dominance norms

Given are several 'signals' $f_t(i)$, $1 \le i \le n$, indexed by $t = 1, \ldots, T$.

Def The *dominant* signal is:

$$f^*(i) := \max_{1 \le t \le T} f_t(i), \quad 1 \le i \le n.$$

The dominance norm of the f_t's is $\|f^*\|_{\ell_\alpha}$.

Problem: Estimate $\|f^*\|_{\ell_\alpha}$ from the sketches $\{E_j(f_t)\}$, $t = 1, \ldots, T$.

Without accessing the signals f_t *directly*!

Solution:

- Max–linearity implies

$$E_j(f^*) = E_j(f_1) \vee \cdots \vee E_j(f_T), \quad 1 \le j \le k.$$

- The sketch $E_j(f^*)$ is readily available!

Proceed as above and estimate $\|f^*\|_{\ell_\alpha}$ by:

$$\|f^*\|_{\ell_\alpha} \approx L_p(f^*) \quad \textit{or by} \quad \|f^*\|_{\ell_\alpha} \approx M(f^*).$$

Motivation: Data streams

Signals which cannot be stored and processed with conventional methods are emerging in:

- Telephone call monitoring (unmanageably large data bases).

- Internet traffic monitoring (vast, rapidly growing amounts of data).

- Sensing applications (restrictions on power, bandwidth, etc.).

- Online transactions (Banking, Stock market data, E-commerce, etc.).

The streaming model – a general framework. Let

$$f(i), \quad 1 \leq i \leq n$$

be a signal with large domain size n.

Starting with $f(i) = 0$, update f sequentially in time:

- cash register Observe $(i, a(i))$, and update f:

$$f(i) := f(i) + a(i)$$

- aggregated Observe $(i, f(i))$ directly.

Examples: Call detail records & IP monitoring

Phone calls: Each time a phone call ends, a CDR packet is sent to a main station. Think of:

$CDR = (i, a(i), \cdots),\ with\ i = (phone\ no.)\ \ a(i) = time\ used.$

• Given a CDR $(i, a(i)$, update

$$f(i) := f(i) + a(i).$$

So the 'signal' $f(i)$ contains info on **all** phone numbers!

• Multiple signals: $f_t(i)$, $t = 1, \ldots, 7$ capture the calling patterns for different days of the week.

IP monitoring: Monitor a fast Internet link during a given day (or hour). How many distinct IPs used it? How many bytes were transmitted from a given range of IPs? Can you give me estimates *in real time*???

• *Signal:* $f(i) = Bytes$ transmitted by $IP\ i$.

• *Stream items:* A packed from IP i contains info:

$$(i, b(i)), \quad i = (IP\ number), \quad b(i) = (Bytes)$$

On observing $(i, b(i))$, update:

$$f(i) := f(i) + b(i).$$

• Impossible to store and process all data *in real time*.

"Classical" (additive) random sketches

Instead of storing the signal $f(i)$, $1 \leq i \leq n$, keep:

$$S_j(f) = \sum_{i=1}^{n} f(i) X_j(i), \quad 1 \leq j \leq k,$$

where $k << n$ and where $X_j(i)$ are iid random variables.

The $S_j(f)$'s $1 \leq j \leq k$ is called the random sketch of f.

- Important: Can generate $X_j(i)$'s 'on the fly' from small $\mathcal{O}(k \log_2(n))$ set of 'seeds' in main memory.

- Can update the sketch *sequentially* in the most general, cash register model.

- Can approximate norms and inner products!

If $\text{Var}(X_j(i)) = \sigma^2$, then $\text{Var}(S_j(f)) = \|f\|_{\ell_2}^2 \sigma^2, \quad \|f\|_{\ell_2}^2 = \sum_i f(i)^2$.

- So

$$\|f\|_{\ell_2}^2 \approx \frac{1}{K} \sum_j S_j(f)^2.$$

- For two signals f and g, by linearity,

$$\|f - g\|_{\ell_2}^2 \approx \frac{1}{k} \sum_{j=1}^{k} |S_j(f) - S_j(g)|^2,$$

Back to max–stable sketches

Strengths

1. Max–stable sketches are max–linear. Natural use in: sensor data, Internet traffic monitoring, Transaction record tracking, max–sequential updates.

2. Work for any $\alpha > 0$ and not only for $0 < \alpha \leq 2$ (unlike the sum–stable sketches).

3. Tailor made for: dominance norm estimation.

Illustration: Signals: $f_t(i) = Bytes$ of IP number i during t–th hour of monitoring. Large domain: $1 \leq i \leq n$ where $n = 2^{32}$.

Problem: *Estimate the norm of the 'heaviest load scenario' during the day, i.e. for*

$$f^*(i) = \max_{1 \leq t \leq 24} f_t(i)$$

estimate the dominance norm

$$\|f^*\|_{\ell_\alpha} \approx ?$$

Problem: *Compare several network traffic scenarios* f, g, h,... *without accessing the enormous data directly. Compute pair–wise distances, norms and dominance norms.*

- Max–stable sketches provide efficient solutions.

Limitation: Max–stable sketches are *not linear* – cannot be updated in the linear cash register format!

Max–stable sketches: Performance guarantees

Norms

Given a max–stable sketch $E_j(f),\ j = 1, \ldots, k$ of a signal $f(i) \geq 0$.

Recall that $M(f)$ is the median–based estimator of $\|f\|_{\ell_\alpha}$.

Th 1(SHKT'07) *Let* $\epsilon,\ \delta \in (0, 1)$. *Then,*

$$\mathbb{P}\{|M(f)/\|f\|_{\ell_\alpha} - 1| \leq \epsilon\} \geq 1 - \delta,$$

if $k \geq C \log(1/\delta)/\epsilon^2$, *for some* $C > 0$.

For optimal value of the constant, as $\epsilon, \delta \to 0$, we have

$$k = k(\epsilon, \delta) \sim \frac{2}{\alpha^2 \ln^2(2)} \frac{\ln(1/\delta)}{\epsilon^2}.$$

Distances

Let now g be another signal and let

$$D(f, g) := 2M(f \vee g)^\alpha - M(f) - M(g) \approx \rho_\alpha(f, g) = \|f^\alpha - g^\alpha\|_{\ell_1}.$$

Th 4 (SHKT'07) *Let* $\epsilon, \delta, \eta \in (0, 1)$. *If* $\rho_\alpha(f) \geq \eta \|f \vee g\|_{\ell_\alpha}^\alpha$, *then*

$$\mathbb{P}\{|D(f, g)/\rho_\alpha(f, g) - 1| \leq \epsilon\} \geq 1 - \delta,$$

for $k \geq C \log(1/\delta)/\epsilon^2$, *with some* $C > 0$.

Sketches: summary & credits

<u>Some</u> seminal papers & algorithmic applications

Alon, Matias & Szegedy (1996)

Feingenbaum, Kannanm Strauss & Viswanathan (1999)

Fong & Strauss (2000)

Indyk (2000)

Gilbert, Kotidis, Muthukrishnan & Strauss (2001, 2002)

Cormode & Muthukrishnan (2003)

Datar, Immorlica, Indyk & Mirrokni (2004)

Review paper: Muthukrishnan (2003)

More on max–stable sketches: Stoev, Hadjielefthe-riou, Kollios & Taqqu (2007)

Code: By Marios Hadjieleftheriou (AT&T Research)

`http://sourceforge.net/projects/sketches`

A lesson: (*For probabilists and computer scientists*)

Study each other's fields!

Application II: Heavy tail exponents

Heavy tailed data

• A random variable X is said to be *heavy–tailed* if

$$\mathbb{P}\{|X| \geq x\} \sim L(x)x^{-\alpha}, \quad \text{as } x \to \infty,$$

for some $\alpha > 0$ and a slowly varying function L.

∘ Here we focus on the simpler but important context:

$$X \geq 0, \quad \text{a.s.} \quad \text{and} \quad \mathbb{P}\{X > x\} \sim Cx^{-\alpha}, \quad \text{as } x \to \infty.$$

∘ X (*infinite moments*) For $p > 0$,

$$\mathbb{E}X^p < \infty \quad \text{if and only if} \quad p < \alpha.$$

In particular,

$$0 < \alpha \leq 2 \quad \Rightarrow \quad \text{Var}(X) = \infty$$

and

$$0 < \alpha \leq 1 \quad \Rightarrow \quad \mathbb{E}|X| = \infty.$$

• The estimation of the *heavy–tail exponent* α is an important problem with rich history.

• Why do we need heavy–tail models?

Every finite sample $X_1, \ldots, X_n$ has finite sample mean, variance and all sample moments!

Why consider heavy tailed models in practice?!

Why use heavy–tailed models?

"All models are wrong, but some are useful."

George Box

Let F and G be any two distributions with positive densities on $(0, \infty)$.

Let $\epsilon > 0$ and $x_1, \ldots, x_n \in (0, \infty)$ be arbitrary, then both:

$$\mathbb{P}_F\{X_i \in (x_i - \epsilon, x_i + \epsilon),\ i = 1, \ldots, n\} > 0$$

and

$$\mathbb{P}_G\{X_i \in (x_i - \epsilon, x_i + \epsilon),\ i = 1, \ldots, n\} > 0$$

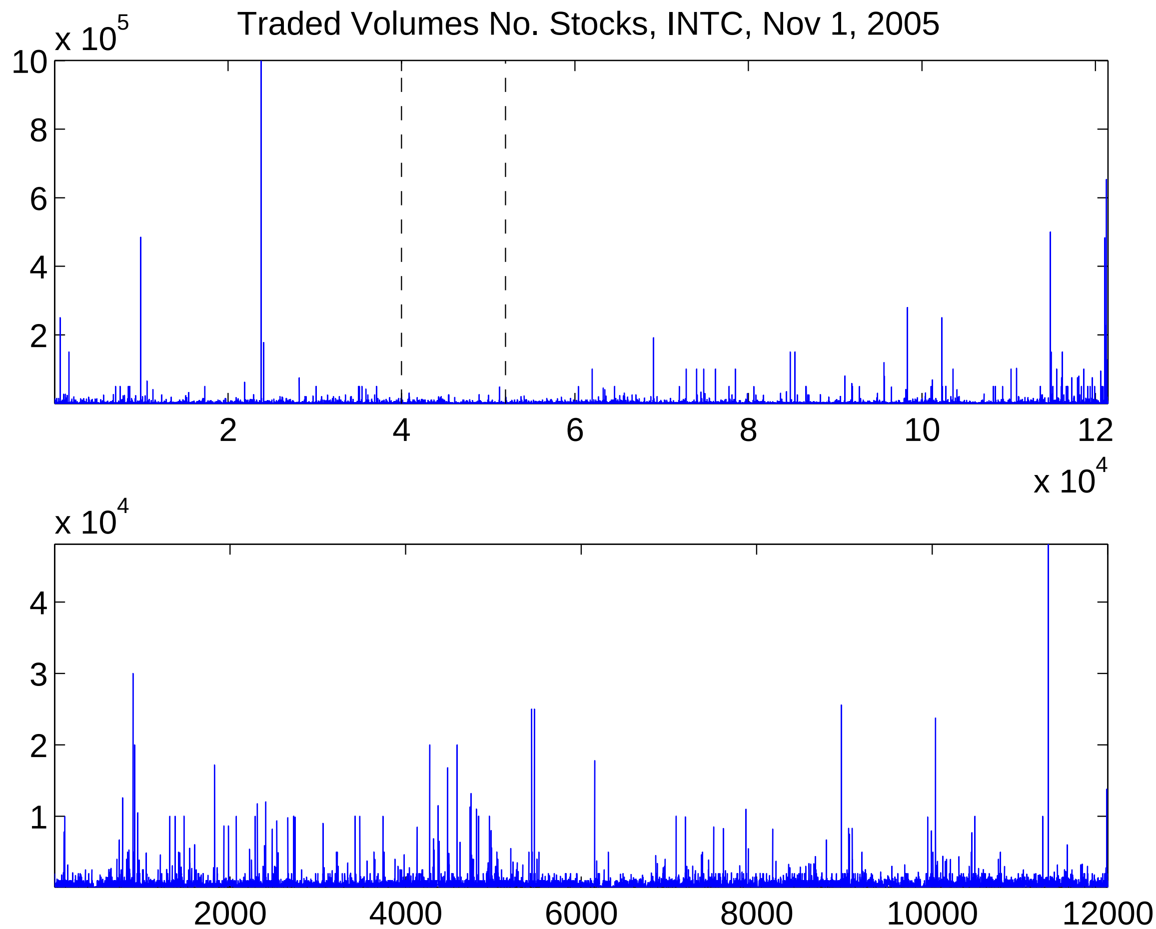
are positive!

• For a given sample, very many models apply.

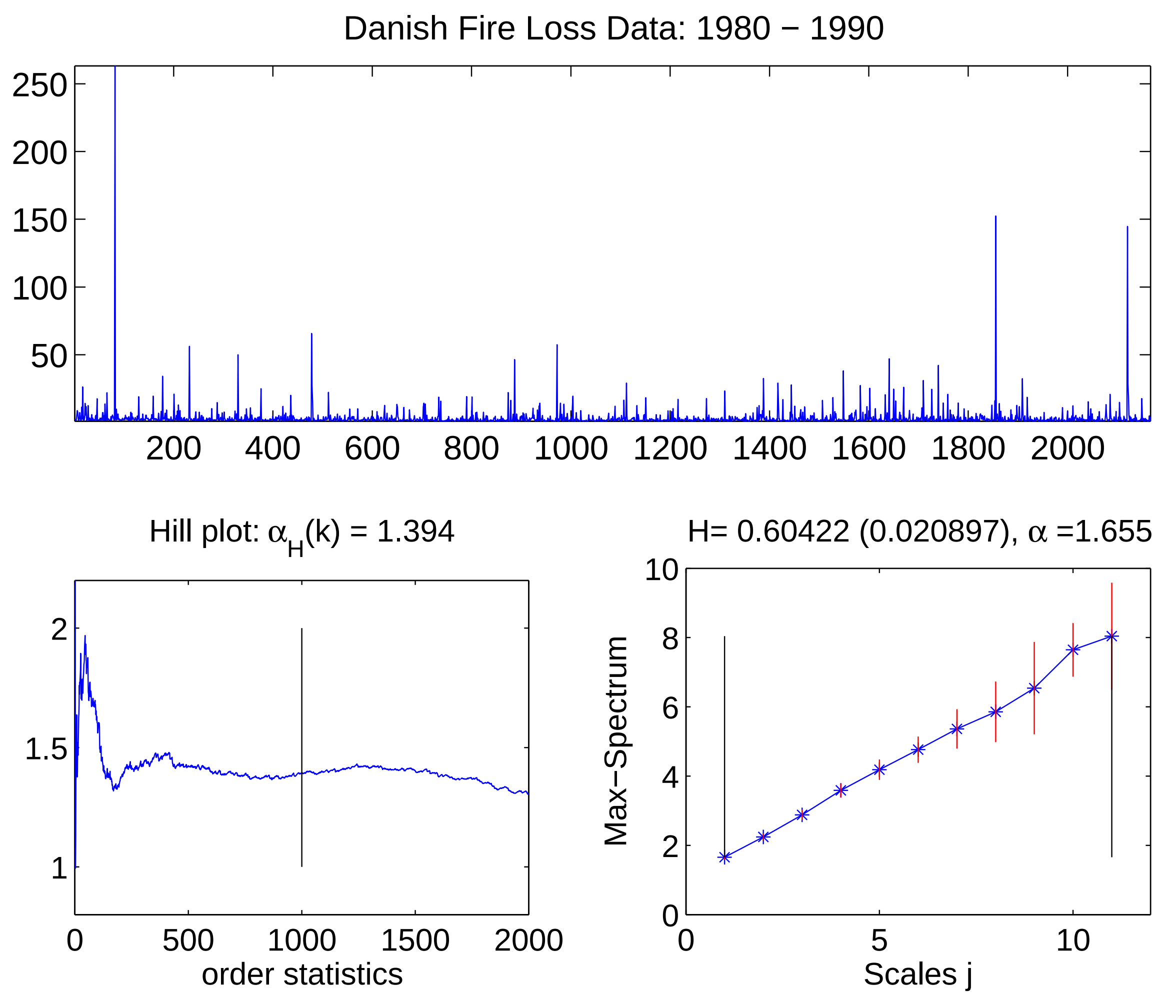
• The ones that continue to work as the sample grows are most suitable.

We next present real data sets of Financial, Insurance and Internet data. They can be *very heavy tailed*.

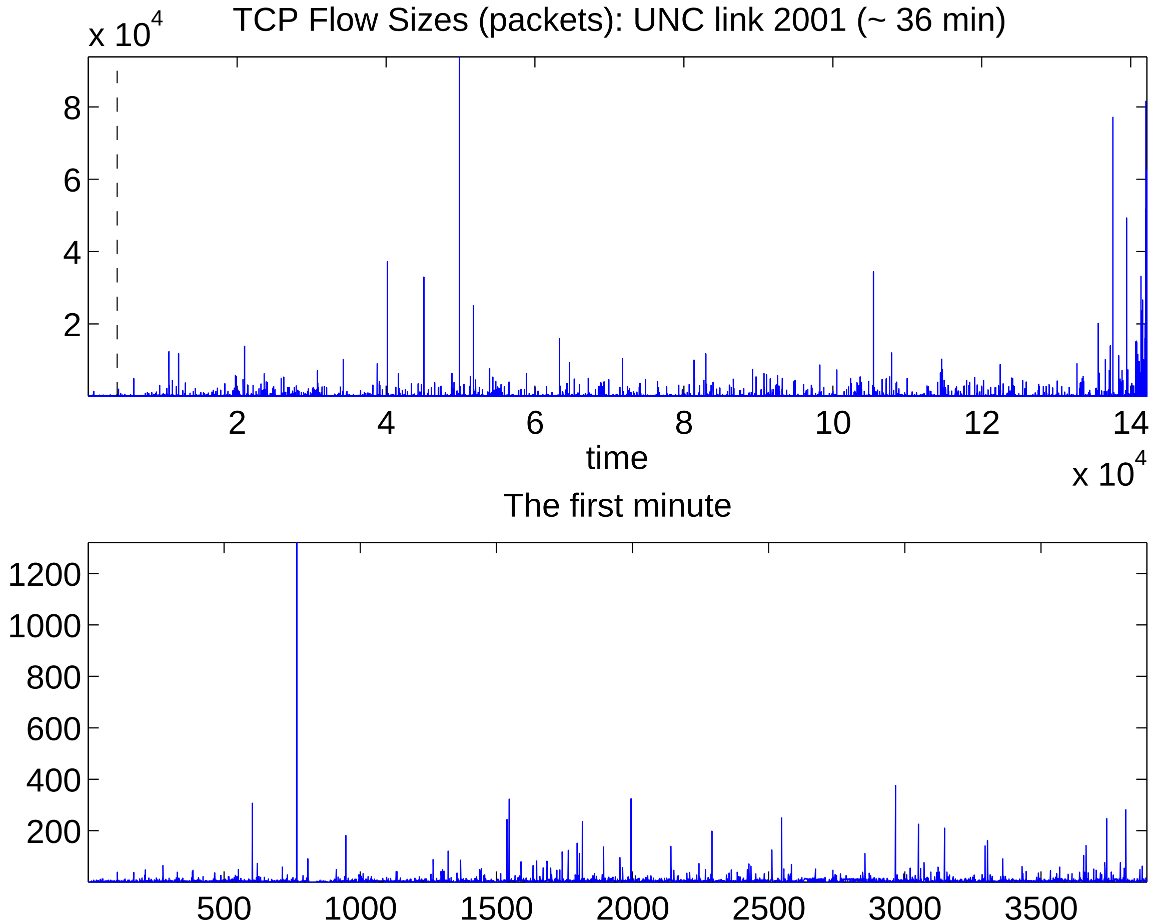
Traded volumes on the Intel stock



Insurance claims due to fire loss



TCP flow sizes (in number of packets)



History

• Hill (1975) worked out the MLE in the Pareto model $\mathbb{P}\{X > x\} = x^{-\alpha},\ x \geq 1$ and introduced the *Hill plot*:

$$\widehat{\alpha}_H(k) := (\frac{1}{k}\sum_{i=1}^{k} \log(X_{i,n}) - \log(X_{k+1,n}))^{-1},$$

where $X_{1,n} \geq X_{2,n} \geq \cdots \geq X_{k,n}$ are the *top–K order statistics* of the sample.

• How to choose k?

∘ pick k where the plot of $\widehat{\alpha}_H(k)$ vs. k stabilizes.

∘ serious problems in practice:

volatile & hard to interpret: "Hill horror plot"

confidence intervals

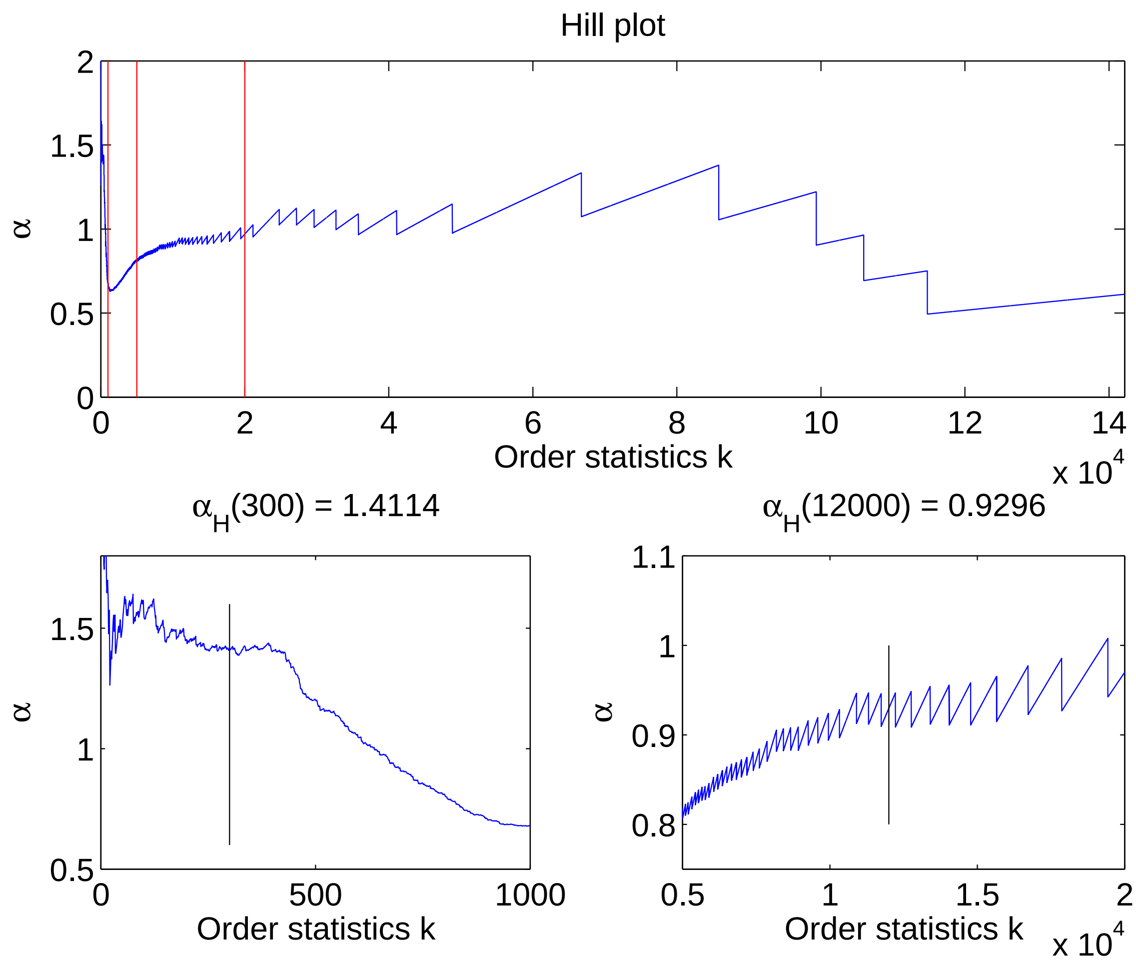
robustness

• Consistency and asymptotic normality resolved: Weissman (1978), Hall (1982) in semi–parametric setting.

• Many other estimators: *kernel based* Csörgő, Deheuvels and Mason (1985), *moment* Dekkers, Einmahl and de Haan (1989), among many others.

• Most estimators exploit the *top order statistics*.

Hill horror plots: TCP flow sizes



Fréchet max–stable laws

Consider i.i.d. X_i's with $\mathbb{P}\{X_1 > x\} \sim Cx^{-\alpha},\ x \to \infty$. By the Fisher–Tippett–Gnedenko theorem:

$$\frac{1}{n^{1/\alpha}} \max_{1\le i\le n} X_i \equiv \frac{1}{n^{1/\alpha}} \bigvee_{i=1}^{n} X_i \stackrel{d}{\longrightarrow} Z, \quad \text{as } n \to \infty,$$

where Z is $\alpha-$Fréchet *extreme value distribution*:

$$\mathbb{P}\{Z \le x\} = \exp\{-Cx^{-\alpha}\}, \quad x > 0.$$

• The extreme value distributions are *max–stable*. In particular, for i.i.d. $\alpha-$Fréchet $Z, \&\ Z_i$'s:

$$Z_1 \vee \cdots \vee Z_n \stackrel{d}{=} n^{1/\alpha} Z.$$

• A time series of i.i.d. $\alpha-$Fréchet $\{Z_k\}$ is $1/\alpha-$max-self-similar:

$$\forall m \in \mathbb{N} : \{\vee_{1\le i\le m} Z_{m(k-1)+i}\}_{k\in\mathbb{N}} \stackrel{d}{=} m^{1/\alpha}\{Z_k\}_{k\in\mathbb{N}}.$$

∘ Block–maxima of size m have the same distribution as the original data, rescaled by the factor $m^{1/\alpha}$.

• Any heavy–tailed $\{X_k\}$ (i.i.d.) data set is *asymptotically max self–similar*.

Max–spectrum

Given a positive sample $X_1, \ldots, X_n$, consider the *dyadic* block–maxima:

$$D(j,k) := \max_{1 \le i \le 2^j} X_{2^j(k-1)+i} \equiv \bigvee_{i=1}^{2^j} X_{2^j(k-1)+i},$$

with $k = 1, \ldots, n_j = [n/2^j]$.

• In view of the *asymptotic scaling*:

$$\frac{1}{2^{j/\alpha}} D(j,k) \xrightarrow{d} Z, \quad (j \to \infty),$$

observe that

$$Y_j := \frac{1}{n_j} \sum_{j=1}^{n_j} \log_2 D(j,k) \simeq j/\alpha + c, \quad (j,\ n_j \to \infty)$$

where $c := \mathbb{E} \log_2 Z$.

∘ The last asymptotics "follow" from the LLN since $D(j,k)$'s are independent in k.

• The *max–spectrum* of the data is defined as the statistics:

$$Y_j,\ j = 1, \ldots, [\log_2(n)].$$

∘ Can identify α from the *slope* of the Y_j's vs. j, *for large* j's.

Max–spectrum estimators of α

Given the *max–spectrum* $Y_j,\ j = 1, \ldots, [\log_2(n)]$, define

$$\widehat{\alpha}(j_1, j_2) := (\sum_{j=j_1}^{j_2} w_j Y_j)^{-1},$$

where $\sum_j j w_j = 1$ and $\sum_j w_j = 0$.

• That is, use *linear regression* to estimate the slope $1/\alpha$.

• Issues:

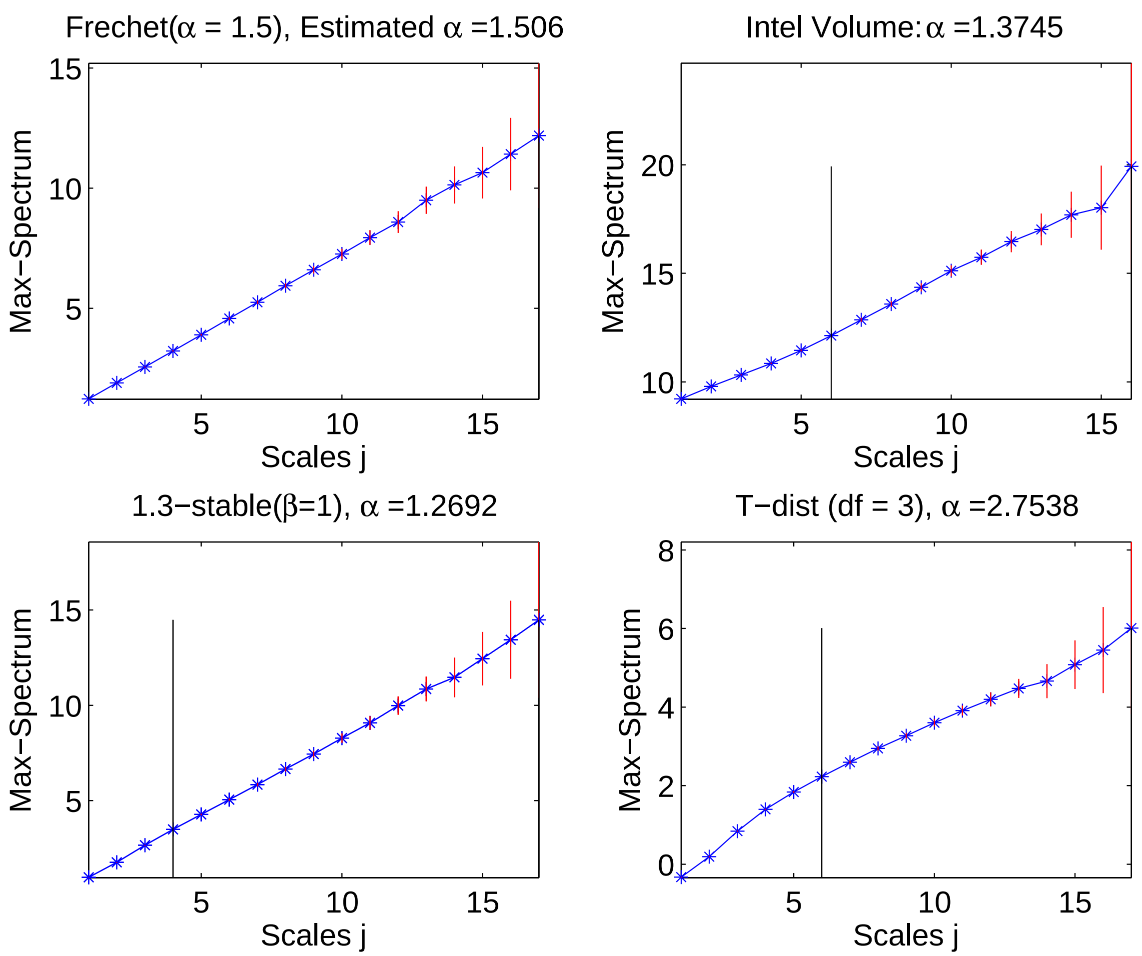
∘*weighted* or *generalized least squares* must be used, since

$$\text{Var}(Y_j) \propto 1/n_j \propto 2^j,$$

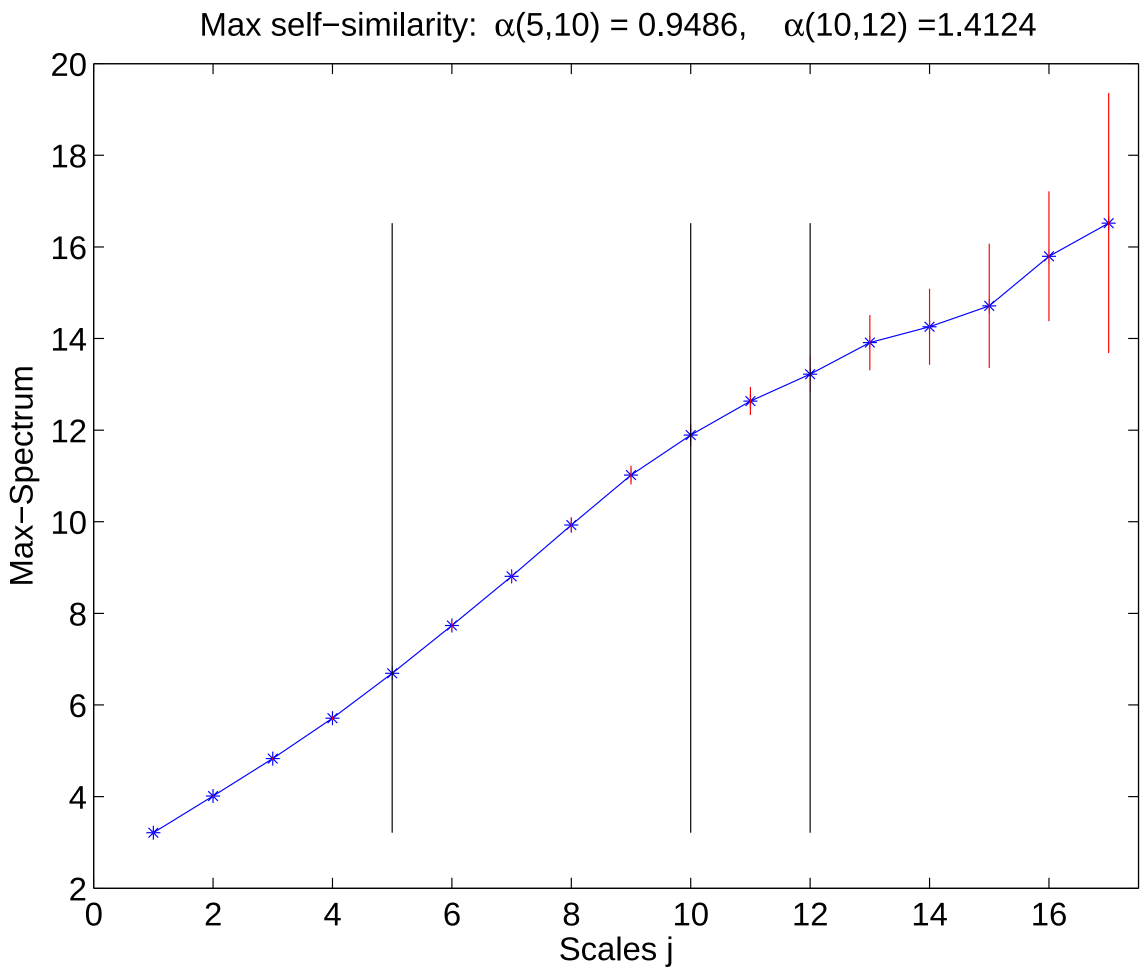
∘ The choice of the *scales* j_1 and j_2 is critical.

• These issues and confidence intervals, are addressed in Stoev, Michailidis and Taqqu (2006).

Examples of max–spectra

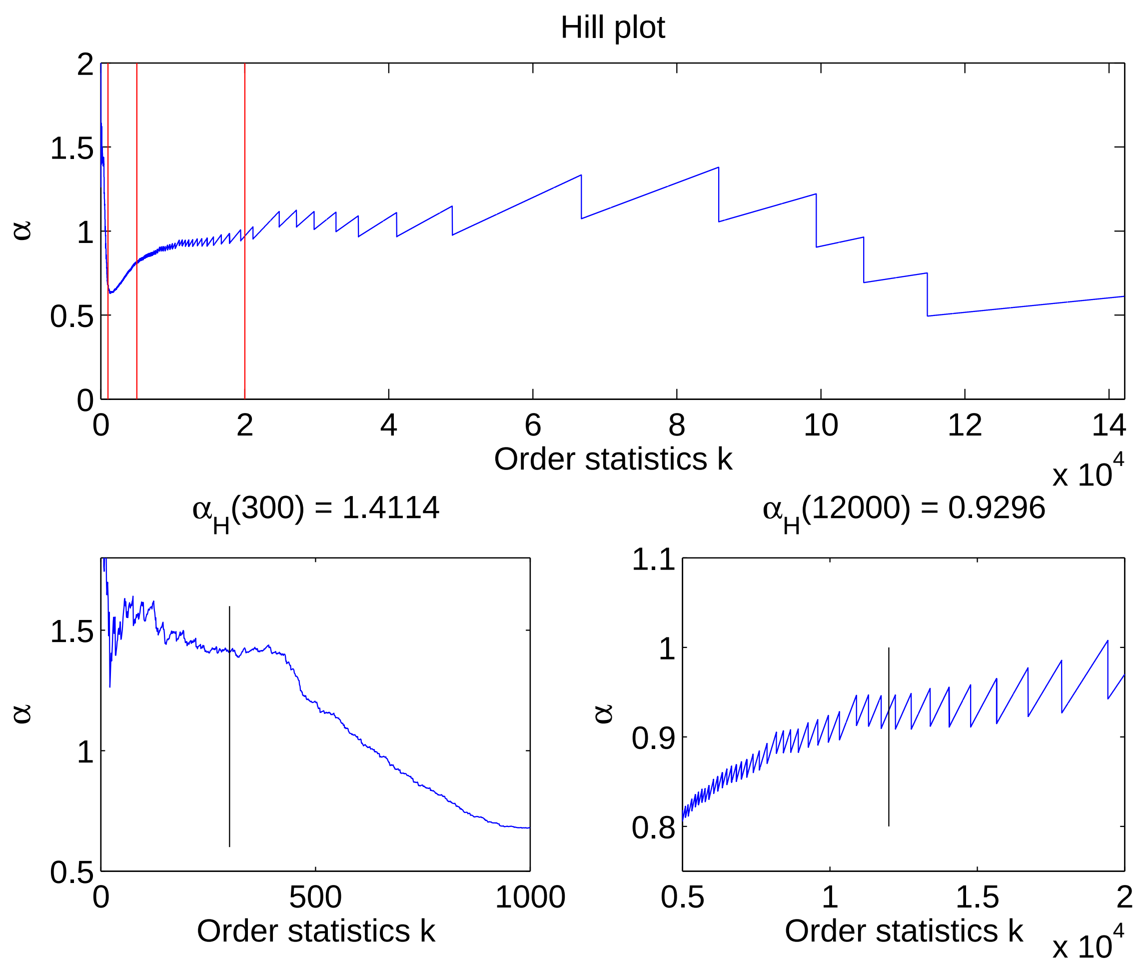


Robustness of the max–spectrum



Compare with the Hill horror plot of the Internet flow–size data (**next slide)**.

Hill horror plots: TCP flow sizes



Compare with the max–spectrum plot of the Internet flow–size data (**previous slide**).

Asymptotic normality of the max–spectrum

Let $\mathbb{P}\{X_k \le x\} =: F(x)$, where

$$1 - F(x) = Cx^{-\alpha}(1 + \mathcal{O}(x^{-\beta})), \quad (x \to \infty), \tag{2}$$

with some $\beta > 0$.

• Consider the max–spectrum $\{Y_j\}$ for the range of scales:

$$r(n) + j_1 \le j \le r(n) + j_2,$$

with *fixed* $j_1 < j_2$ and $r(n) \to \infty,\ n \to \infty$.

Theorem Under (2) and another technical condition,

$$\sup_{x\in\mathbb{R}} \Big| \mathbb{P}\{\sqrt{n_{j_2+r}}((\vec{\theta}, \vec{Y}_r) - (\vec{\theta}, \vec{\mu}_r)) \le x\} - \Phi(x/\sigma_{\vec{\theta}}) \Big|$$
$$\le C_{\vec{\theta}}(1/2^{r\beta/\alpha} + r2^{r/2}/\sqrt{n}), \tag{3}$$

where Φ is the standard Normal c.d.f.

Here $\vec{Y}_r = (Y_{r+j})_{j=j_1}^{j_2}$ and

$$(\vec{\theta}, \vec{Y}_r) = \sum_{j=j_1}^{j_2} \theta_j Y_{r+j}, \qquad \mu_r(j) = (r + j)/\alpha + c$$

and

$$\sigma^2_{\vec{\theta}} = (\vec{\theta}, \Sigma_\alpha \vec{\theta}) := \sum_{i,j=j_1}^{j_2} \theta_i \Sigma_\alpha(i, j)\theta_j > 0.$$

The covariance structure of the max–spectrum

• The asymptotic covariance matrix Σ is the same as if the data were i.i.d. $\alpha-$Fréchet!

◦ The intuition is that the block–maxima

$$\{D(r+j,k),\ k=1,\ldots,n_{r+j}\}$$

behave like i.i.d. $\alpha-$Fréchet variables, as $r = r(n) \to \infty$.

• The covariance entries are given by:

$$\Sigma_\alpha(i,j) = \alpha^{-2} 2^{i\vee j - j_2} \psi(|i-j|),$$

with

$$\psi(a) := \mathrm{Cov}(\log_2(Z_1), \log_2(Z_1 \vee (2^a - 1)Z_2), \quad a \ge 0,$$

for i.i.d. $1-$Fréchet Z_1 & Z_2.

◦ Note that α appears only as a factor in Σ_α:

$$\Sigma_\alpha = \alpha^{-2}\Sigma_1.$$

It does not affect the *correlation structure* of the max–spectrum.

• GLS estimators for α use the matrix Σ_α.

• Asymptotic normality for $\widehat{\alpha}(r+j_1, r+j_2)$ follows from the last theorem.

• Stoev, Michailidis and Taqqu (2006) has details on *confidence intervals* for α and *automatic selection* of j_1&j_2.

Automatic selection of scales

'Null hypothesis': $\vec{Y} = \{Y_j\}_{j=1}^{J}$ **is** multivariate Normal with covariance matrix $\alpha^{-2}\Sigma_1$ and means

$$\mathbb{E}Y_j = j/\alpha + \text{const}, 1 \le j \le J.$$

Given a range of scales $j_1 \le j \le j_2$, set

$$\widehat{H}(j_1, j_2) \equiv 1/\widehat{\alpha}(j_1, j_2) := \vec{w}'_{j_1,j_2}\vec{Y},$$

for the GLS estimator of the slope $H = 1/\alpha$ of Y_j vs j.

Algorithm:

1. Set $j_2 \approx J = [\log_2(n)]$ and $j_1 := j_2 - 1$.

2. Compute $\widehat{H}_{+1} := \widehat{H}(j_1 - 1, j_2)$ and $\widehat{H}_0 := \widehat{H}(j_1, j_2)$.

3. Compute a *conf int* for $\widehat{H}_{+1} - \widehat{H}_0$ (based on the *null hypothesis*).

4. If the conf int does not contain 0, then *stop*. Else, set $j_1 := j_1 - 1$ and repeat Step 2.

Important details:

- Plug–in for α used in the covariance matrix

- Level of conf int to be set by the user

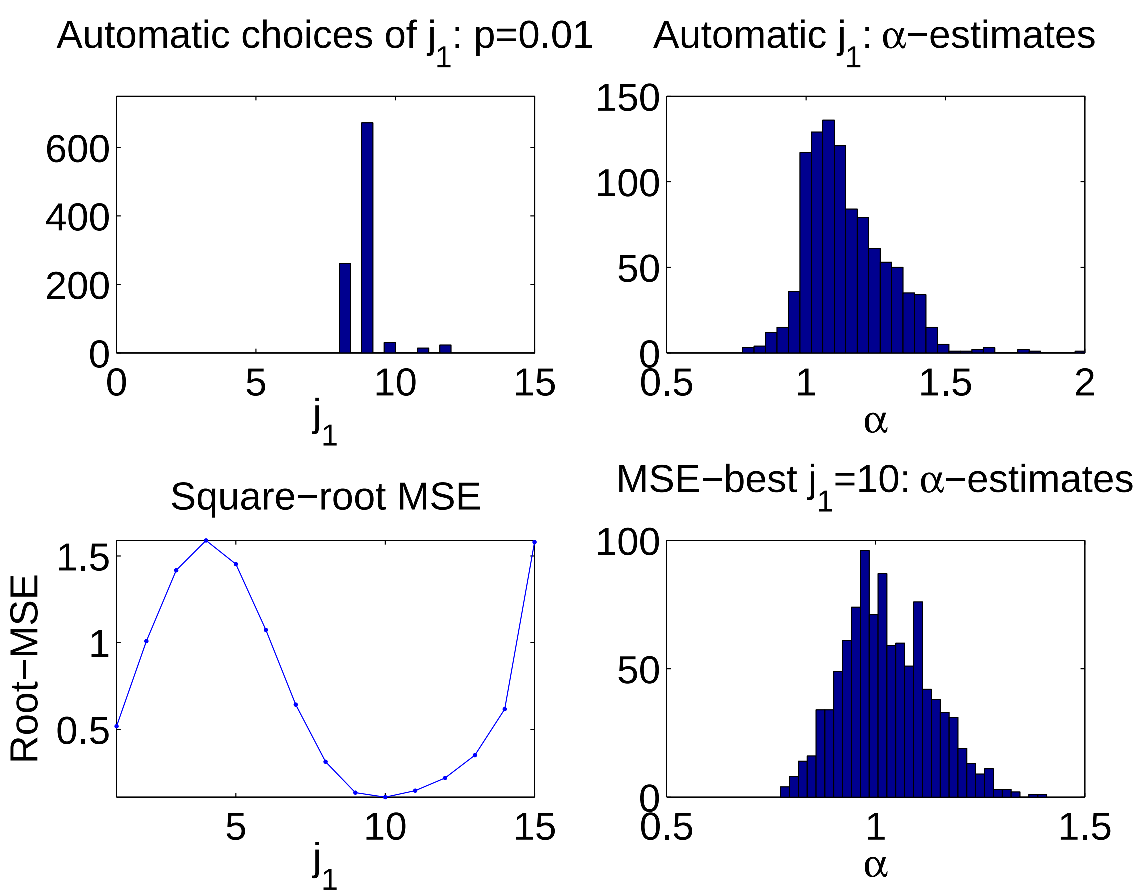
- Multiple testing issue not addressed

The method works very well in practice!

Automatic selection of scales: an illustration

Simulated: a mixture of 1−Fréchet (10%) and Exponential with mean 5 (90%).

Sample size: $n = 2^{17} = 131,072$.



Rates for moment functionals of maxima

Let $X_1, \ldots, X_n$ be i.i.d. from F and recall that

$$M_n := \frac{1}{n^{1/\alpha}} \bigvee_{1 \le i \le n} X_i \xrightarrow{d} Z, \quad (n \to \infty).$$

• Are there results on the rate of convergence

$$\mathbb{E}f(M_n) \longrightarrow \mathbb{E}f(Z), \quad (n \to \infty)$$

for "reasonable" f's?

◦ Pickands (1975) shows only the convergence of moments (no rates).

• Our approach: consider

$$F(x) = \exp\{-\sigma^\alpha(x) x^{-\alpha}\}, \quad x \in \mathbb{R}, \tag{4}$$

with

$$\sigma^\alpha(x) \longrightarrow C, \quad (x \to \infty).$$

◦ Note: $1 - F(x) \sim Cx^{-\alpha}$, $x \to \infty$ is equivalent to (4).

• Extra assumption:

$$|\sigma^\alpha(x) - C| \le Dx^{-\beta}, \quad \text{for large } x.$$

Rates (cont'd)

But (4) is very convenient to handle rates!

- Note that

$$\begin{aligned}\mathbb{P}\{M_n \le x\} &= \mathbb{P}\{X_1 \le n^{1/\alpha}x\}^n = F(n^{1/\alpha}x)^n \\ &= \exp\{-\sigma^\alpha(n^{1/\alpha}x)x^{-\alpha}\}. \qquad (5)\end{aligned}$$

∘ Thus,

$$\mathbb{E}f(M_n) = \int_0^\infty f(x)dF_n(x) = \int_0^\infty f(x)d\exp\{-\sigma^\alpha(n^{1/\alpha}x)x^{-\alpha}\},$$

and also

$$\mathbb{E}f(Z) = \int_0^\infty f(x)dG(x) = \int_0^\infty f(x)d\exp\{-Cx^{-\alpha}\}.$$

- Now, *integration by parts* yields:

$$\mathbb{E}(f(M_n) - f(Z)) = \int_0^\infty (G(x) - F_n(x))f'(x)dx.$$

∘ However $G(x)$ and $F_n(x)$ are of the same "exponential" form!

Rates (cont'd)

By the mean value theorem:

$$\begin{aligned}
|F_n(x) - G(x)| &= |\exp\{-\sigma^\alpha(n^{1/\alpha}x)x^{-\alpha}\} - \exp\{-Cx^{-\alpha}\}| \\
&\le |\sigma^\alpha(n^{1/\alpha}x) - C|x^{-\alpha}\exp\{-cx^{-\alpha}\} \\
&\le Dn^{-\beta/\alpha}x^{-(\alpha+\beta)}\exp\{-cx^{-\alpha}\}, \quad \text{as } x \to \infty,
\end{aligned}$$

- Thus, for any $\epsilon > 0$, as $n \to \infty$,

$$\int_\epsilon^\infty |F_n(x) - G(x)||f'(x)|dx$$

$$\le Dn^{-\beta/\alpha}\int_0^\infty |f'(x)|x^{-(\alpha+\beta)}e^{-cx^{-\alpha}}dx = \mathcal{O}(n^{-\beta/\alpha}).$$

- By taking $\epsilon = \epsilon_n \to 0$, and using a mild technical condition we can also bound the integral near zero

$$\int_0^{\epsilon_n} |F_n(x) - G(x)||f'(x)|dx.$$

Th *If* $\sigma^\alpha(x) \sim Dx^{-\beta}$, *as* $x \to \infty$, *then as* $n \to \infty$,

$$n^{\beta/\alpha}(\mathbb{E}f(M_n) - \mathbb{E}f(Z)) \longrightarrow D\int_0^\infty x^{-(\alpha+\beta)}f'(x)e^{-Cx^{-\alpha}}dx,$$

provided mild technical conditions on $\sigma(x)$ *at* 0 *and on* f *at* 0 *and* ∞ *hold.*

Rates (cont'd)

• We have thus obtained *exact rates* for *moment functionals*

$$\mathbb{E} f(\frac{1}{n^{1/\alpha}} \max_{1\le i\le n} X_i)$$

∘ They are valid for a large class of *absolutely continuous* f, including:

$$f(x) = \log_2(x), \quad \text{and} \quad f(x) = x^p,\ p \in (0, \alpha),$$

for example.

∘ More details can be found in Stoev, Michailidis and Taqqu (2006).

• As a corollary, we also get rates of convergence for

$$\mathrm{Cov}(\log_2 D(r+j_1, k), \log_2 D(r+j_2, k)), \quad \text{as } r = r(n) \to \infty.$$

• These tools and the Berry–Esseen Theorem, yield the *uniform asymptotic normality* of the *max–spectrum*.

Thank you!

References

Alon, N., Matias, Y. & Szegedy, M. (1996), The space complexity of approximating the frequency moments, *in* 'Proceedings of the 28th ACM Symposium on Theory of Computing', pp. 20–29.

Cormode, G. & Muthukrishnan, S. (2003), *Algorithms - ESA 2003*, Vol. 2832 of *Lecture Notes in Computer Science*, Springer, chapter Estimating Dominance Norms of Multiple Data Streams, pp. 148 –160.

Csörgő, S., Deheuvels, P. & Mason, D. (1985), 'Kernel estimates of the tail index of a distribution', *Annals of Statistics* **13**(3), 1050–1077.

Datar, M., Immorlica, N., Indyk, P. & Mirrokni, V. S. (2004), Locality–sensitive hashing scheme based on p–stable distributions, *in* 'SCG '04: Proceedings of the twentieth annual symposium on Computational geometry', ACM Press, New York, NY, USA, pp. 253–262.

Dekkers, A., Einmahl, J. & de Haan, L. (1989), 'A moment estimator for the index of an extreme–value distribution', *Ann. Statist.* **17**(4), 1833–1855.

Feigenbaum, J., Kannan, S., Strauss, M. & Viswanathan, M. (1999), An approximate l1-difference algorithm for massive data streams, *in* 'FOCS '99: Proceedings of the 40th Annual Symposium on Foundations of Computer Science', IEEE Computer Society, Washington, DC, USA, p. 501.

Fong, J. H. & Strauss, M. (2000), An approximate lp-difference algorithm for massive data streams, *in* 'STACS '00: Proceedings of the 17th Annual Symposium on Theoretical Aspects of Computer Science', Springer-Verlag, London, UK, pp. 193–204.

Gilbert, A., Kotidis, Y., Muthukrishnan, S. & Strauss, M. (2001), Surfing wavelets on streams: one-pass summaries for approximate aggregate queries, *in* 'Procedings of VLDB, Rome, Italy'.

Gilbert, A., Kotidis, Y., Muthukrishnan, S. & Strauss, M. (2002), 'How to summarize the universe: Dynamic maintenance of quantiles'.

Hall, P. (1982), 'On some simple estimates of an exponent of regular variation', *J. Roy. Stat. Assoc.* **44**, 37–42. Series B.

Hill, B. M. (1975), 'A simple general approach to inference about the tail of a distribution', *The Annals of Statistics* **3**, 1163–1174.

Indyk, P. (2000), Stable distributions, pseudorandom generators, embeddings and data stream computation, *in* 'IEEE Symposium on Foundations of Computer Science', pp. 189–197.

Muthukrishnan, S. (2003), Data streams: algorithms and applications, Preprint. `http://www.cs.rutgers.edu/~muthu/`.

Pickands, J. (1975), 'Statistical inference using extreme order statistics', *Ann. Statist.* **3**, 119–131.

Stoev, S., Hadjieleftheriou, M., Kollios, G. & Taqqu, M. (2007), Norm, point and distance estimation over multiple signals using max–stable distributions, *in* '23rd International Conference on Data Engineering', ICDE, IEEE. To appear.

Stoev, S., Michailidis, G. & Taqqu, M. (2006), Estimating heavy–tail exponents through max self–similarity, Preprint.

Weissman, I. (1978), 'Estimation of parameters and large quantiles based on the k largest observations', *Journal of the American Statistical Association* **73**, 812–815.