University of Siegen
Department of Mathematics

When the POT–approach fails: super heavy tailed distributions and covariate information

Ulf Cormann

NCAR, Boulder

25.02.2009

Inhalt

Overview

1. Motivation
2. Modeling using Covariate Information
3. Super–Heavy Tailed Distributions
4. Literature

Consider random variables Y and random vectors $\boldsymbol{X}$ with values in $\mathbb{R}$ and $S \subset \mathbb{R}^m$ and the pertaining conditional distribution function

$$F(y|\boldsymbol{x}) := P(Y \leq y|\boldsymbol{X} = \boldsymbol{x}).$$

Assume that $F(\cdot|\boldsymbol{x})$ is in the domain of attraction of some GPD which parameters depend on $\boldsymbol{x}$, thus

$$F(y|\boldsymbol{x}) \approx W_{\gamma(\boldsymbol{x}),\mu(\boldsymbol{x}),\sigma(\boldsymbol{x})}(y), \quad y > u.$$

where

$$W_{\gamma,\mu,\sigma}(y) = 1 - \left(1 + \gamma\frac{y-\mu}{\sigma}\right)^{-1/\gamma}.$$

Consider for example the case $\boldsymbol{X} : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathbb{B}(\mathbb{R}))$

$$F(y|x) = W_{1/x,0,\gamma/x}(y)$$

and X has a gamma distribution with parameter γ then the distribution of Y is a log–Pareto distribution

$$L(y) = 1 - (\log(1 + x/\sigma))^{-1/\gamma}\,.$$

Note that the tail of L cannot be approximated by any GPD.

Considering iid copies $(\boldsymbol{X}_1, Y_1), \ldots, (\boldsymbol{X}_n, Y_n)$ of the random vector $(\boldsymbol{X}, Y)$ and assuming

$$F(y|\boldsymbol{x}) = W_{\gamma(\boldsymbol{x}),\mu(\boldsymbol{x}),\sigma(\boldsymbol{x})}(y), \quad y > u$$

one has two possible situations:

1. the covariates $\boldsymbol{X}_i$ can be observed, so one can make inference for the conditional distribution
2. the covariates $\boldsymbol{X}_i$ cannot be observed, so one has to take distributions different from GPDs into account

Overview

Point Processes

A point process is a mapping $N : (\Omega, \mathcal{A}, P) \to (\mathbb{M}(T), \mathcal{M}(T))$, where $\mathbb{M}(T)$ is the space of point measures $\mu = \sum_{i=1}^{k} \varepsilon_{x_i}$, $x_i \in T$, on some space T.
Let $\mathcal{B}$ be a σ-algebra on T.
Then $N(A)$ is for $A \in \mathcal{B}$ a $\mathbb{N}_0$-valued random variable.
Furtheron $\nu(A) := E(N(A))$ defines a measure on $(T, \mathcal{B})$, the intensity measure of N.
Example: empirical process

$$N_n = \sum_{i=1}^{n} \varepsilon_{X_i}$$

for iid random variables X_i on $(T, \mathcal{B})$.

Poisson Processes

A point process N with finite intensity measure ν ($\nu(T) < \infty$) is a Poisson process, if

- for all disjoint $A_i \in \mathcal{B}$, $i = 1, ..., n$, $N(A_i)$ are independent.
- $N(A)$ is for $A \in \mathcal{B}$ distributed according to a Poisson distribution with parameter $\nu(A)$

Remark: The distribution of a Poisson process is uniquely determined by its intensity measure

Representation of Poisson Processes

A Poisson process N with finite intensity measure ν has the representation

$$N = \sum_{i=1}^{\tau} \varepsilon_{X_i}$$

where

- τ is distributed according to Poisson distribution with parameter $\nu(T)$, independent of
- $X_1, X_2, \dots$ iid random variables on $(T, \mathcal{B})$ with $P\{X_i \in B\} = \frac{\nu(B)}{\nu(T)}$.

Let

$$N = \sum_{i=1}^{\tau} \varepsilon_{(\boldsymbol{x}_i, Y_i)},$$

$\tau \sim \mathrm{P}_\lambda$ the Poisson point process of the observed data (on $T = S \times \mathbb{R}$). Define

$$N^{[S,u]} = N(\cdot \cap S \times (u, \infty))$$

the point process of exceedances over the threshold u and the pertaining covariates.

The Process of Exceedances

Now it holds that $N^{[S,u]}$ is again a Poisson Process

$$N^{[S,u]} \overset{d}{=} \sum_{i=1}^{\tau^*} \varepsilon_{(\boldsymbol{X}_i^*, Y_i^*)} \tag{1}$$

where τ^* and $\left(\boldsymbol{X}_i^*, Y_i^*\right)$, $i \in \mathbb{N}$ are independent, τ^* is a Poisson random variable with parameter $\lambda^* = \lambda P\left\{Y > u\right\}$,

$$P\left(Y^* \leq y \,|\boldsymbol{X}^* = \boldsymbol{x}\right) = F^{[u]}(u|\boldsymbol{x}) \tag{2}$$

and

$$P\left\{\boldsymbol{X}^* \in B\right\} = P\left(\boldsymbol{X} \in B\,|Y > u\right). \tag{3}$$

Parametric Modeling

We assume that

$$F(y|\boldsymbol{x}) = W_{\gamma_\theta(\boldsymbol{x}),\mu_\theta(\boldsymbol{x}),\sigma_\theta(\boldsymbol{x})}(y), \quad y > u$$

where $\theta \in \Theta \subset \mathbb{R}^d$ is a parameter, for example if $S = \mathbb{R}$ one may choose

$$\gamma(x) \equiv \theta_1 \in \mathbb{R},$$

$$\sigma(x) = \exp(\theta_2 + \theta_3 x), \quad \theta_2, \theta_3 \in \mathbb{R} \tag{4}$$

as well as

$$\mu(x) = \theta_4 + \theta_5 x, \quad \theta_4, \theta_5 \in \mathbb{R}. \tag{5}$$

Densities of Poisson Processes

Let N be N_0 arbitrary Poisson Processes on the space $(\mathbb{M}(T), \mathcal{M}(T))$ with finite intensity measures ν and ν_0.
Let h be a ν_0–density of ν.
Then $\mathcal{L}(N)$ has the $\mathcal{L}(N_0)$–density

$$g(\mu) = \left(\prod_{i=1}^{k} h(x_i)\right) \exp\left(\nu_0(T) - \nu(T)\right)$$

for

$$\mu = \sum_{i=1}^{k} \varepsilon_{x_i}.$$

The intensity measure $\nu^{[S,u]}$ of $N^{[S,u]}$ is given by

$$\nu^{[S,u]}(B \times (u,z]) = \lambda \int_B \int_u^z f(y|\boldsymbol{x}) dy d\mathcal{L}(\boldsymbol{X})(\boldsymbol{x})$$

where $f(\cdot|\boldsymbol{x})$ denotes the Lesbegue density of $F(\cdot|\boldsymbol{x})$.

Let N_0 be a Poisson process on $S \times \mathbb{R}$ with intensity measure $\nu_0 = \lambda \cdot (\mathcal{L}(\boldsymbol{X}) \times P)$, where P is a probability measure with positive density p on (u, ∞).
The intensity measure $\nu^{[S,u]}$ of $N^{[S,u]}$ has the ν_0-density

$$h(y, \boldsymbol{x}) = \frac{w_{\gamma_\theta(\boldsymbol{x}),\mu_\theta(\boldsymbol{x}),\sigma_\theta(\boldsymbol{x})}(y)}{p(y)}, \quad y > u.$$

Thus the density of $\mathcal{L}\left(N^{[S,u]}\right)$ with respect to $\mathcal{L}\left(N_0\right)$ is given by

$$g\left(\boldsymbol{\eta}\right) = \prod_{i=1}^{k} \frac{w_{\gamma(\boldsymbol{x}_i^*),\mu(\boldsymbol{x}_i^*),\sigma(\boldsymbol{x}_i^*)}(y_i^*)}{p(y_i^*)} \cdot \exp\left(\lambda - \lambda \int W_{\gamma(\boldsymbol{x}),\mu(\boldsymbol{x}),\sigma(\boldsymbol{x})}(u) d\mathcal{L}(\boldsymbol{X})(\boldsymbol{x})\right)$$

for

$$\boldsymbol{\eta} = \sum_{i=1}^{k} \varepsilon_{(\boldsymbol{x}_i^*, y_i^*)}.$$

log–likelihood

The resulting log–Likelihood is given by

$$
\begin{aligned}
l_{\boldsymbol{\eta}}(\theta) \quad = \quad & \sum_{i=1}^{k} \log \left(W_{\gamma_\theta(\boldsymbol{x}_i^*),\mu_\theta(\boldsymbol{x}_i^*),\sigma_\theta(\boldsymbol{x}_i^*)}(y_i^*) \right) \\
& \qquad - n \int \left(1 - W_{\gamma_\theta(\boldsymbol{x}),\mu_\theta(\boldsymbol{x}),\sigma_\theta(\boldsymbol{x})}(u)\right) d\mathcal{L}(\boldsymbol{X})(\boldsymbol{x}),
\end{aligned}
$$

for $\boldsymbol{\eta} = \sum_{i=1}^{k} \varepsilon_{(\boldsymbol{x}_i^*, y_i^*)}$ and n the size of the original sample.

A Conditional Approach

Let N and $N^{[S,u]}$ be as before. Let π_1, be the projection mapping

$$\pi_1\left(\sum_{i=1}^{n}\varepsilon_{(\boldsymbol{x}_i,y_i)}\right)=\sum_{i=1}^{n}\varepsilon_{\boldsymbol{x}_i}$$

and define

$$N_1=\pi_1(N)$$

as well as

$$N_1^{[S,u]}=\pi_1\left(N^{[S,u]}\right)$$

A Conditional Approach

1. N_1 carries no relevant information for the estimation problem
2. distribution of the covariates should be kept out of the considerations as far as possible
3. distribution of N and $N^{[S,u]}$ strongly depends on the distribution of the covariates

Solution: We consider the conditional distribution

$$P\left(N^{[S,u]} \in \cdot | N_1 = \mu\right), \ \mu = \sum_{i=1}^{n} \varepsilon_{\mathbf{x}_i}$$

Advantage: This distribution carries all information about the exceedances and the exceedance probability, yet the distribution of the covariates has not to be accounted for.

Conditional Likelihood

- Let $(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_n, y_n)$ be the observed data
- $(\boldsymbol{x}_1^*, y_1^*), \ldots, (\boldsymbol{x}_k^*, y_k^*)$ the pertaining “exceedances”
- $\tilde{\boldsymbol{x}}_1, \ldots, \tilde{\boldsymbol{x}}_{n-k}$ the covariates belonging to y-values smaller then u.
- let $\boldsymbol{\eta} = \sum_{i=1}^{k} \varepsilon_{(\boldsymbol{x}_i^*, y_i^*)}$

The conditional Likelihood is given by

$$l_{\boldsymbol{\eta},\mu}(\theta) = \prod_{i=1}^{n-k} W_{\gamma_\theta(\tilde{\boldsymbol{x}}_i),\mu_\theta(\tilde{\boldsymbol{x}}_i),\sigma_\theta(\tilde{\boldsymbol{x}}_i)}(u) \prod_{i=1}^{k} w_{\gamma_\theta(\boldsymbol{x}_i^*),\mu_\theta(\boldsymbol{x}_i^*),\sigma_\theta(\boldsymbol{x}_i^*)}(y_i^*)$$

Theorie und Praxis für Karrieren von morgen

Sketch of a Proof

First show that

$$P\left(\left(N^{[S,u]}, N_1\right) \in A \times B \,\middle|\, N_1^{[S,u]} = \mu^*\right)$$
$$= P\left(N_1 \in B \,\middle|\, N_1^{[S,u]} = \mu^*\right) P\left(N^{[S,u]} \in A \,\middle|\, N_1^{[S,u]} = \mu^*\right),$$

for all $A \in \mathcal{M}(S \times \mathbb{R})$ and $B \in \mathcal{M}(S)$.
Therefore

$$P\left(N^{[S,u]} \in A \,|N_1 = \mu\right)$$
$$= \int \int 1_A(\eta) dP\left(N^{[S,u]} \in d\eta \,\middle|\, N^{[S,u]} = \mu^*\right)$$
$$dP\left(N_1^{[S,u]} \in d\mu^* \,|N_1 = \mu\right)$$

for all $A \in \mathcal{M}(S \times \mathbb{R})$.

Sketch of a Proof

Therefore it remains to determine the conditional distributions

$$P\left(N_1^{[S,u]} \in \cdot \,\middle|\, N_1 = \mu\right) \tag{6}$$

and

$$P\left(N^{[S,u]} \in \cdot \,\middle|\, N_1^{[S,u]} = \mu^*\right). \tag{7}$$

Simulations

Recall the model above

$$F(y|\boldsymbol{x}) = W_{\gamma_\theta(x),\mu_\theta(x),\sigma_\theta(x)}(y), \quad y > u$$

where $\theta \in \Theta \subset \mathbb{R}^5$

$$\gamma(x) \equiv \theta_1 \in \mathbb{R},$$

$$\sigma(x) = \exp(\theta_2 + \theta_3 x), \quad \theta_2, \theta_3 \in \mathbb{R} \tag{8}$$

as well as

$$\mu(x) = \theta_4 + \theta_5 x, \quad \theta_4, \theta_5 \in \mathbb{R}. \tag{9}$$

We choose $\theta_1 = 0.6$, $\theta_2 = 0.2$, $\theta_3 = 0.8$, $\theta_4 = \theta_5 = 0.5$

Results

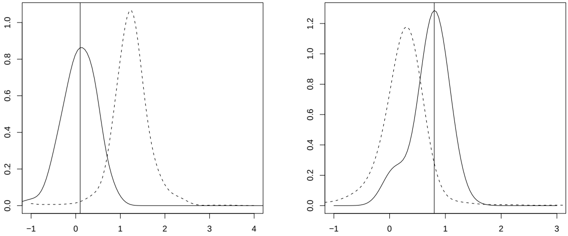

Figure: Kernel densities of the conditional likelihood estimator (solid) and the maximum–likelihood estimator (dashed) for θ_2 and θ_3.

Overview

1. Motivation
2. Modeling using Covariate Information
3. Super–Heavy Tailed Distributions
4. Literature

Recall the example of the log–Pareto df

$$L(y) = 1 - (\log(1 + x/\sigma))^{-1/\gamma}.$$

1. all moments are infinite
2. the existence of log–moments depends on the parameter γ
3. the tail of L is slowly varying

$$\lim_{t\to\infty} \frac{1 - L(tx)}{1 - L(t)} = 1$$

The log–Pareto Model

Question: What is an adequate model for such distributions?

$$L_{\gamma,\beta,\sigma}(x) = 1 - \left(1 + \frac{\gamma}{\beta}\log(1 + \frac{\beta}{\sigma}x)\right)^{-1/\gamma}$$

$x > 0$, β, $\sigma > 0$, $\gamma \in \mathbb{R}$,
the generalized log–Pareto distribution (GLPD).

Relations to GPDs

The GPD can be obtained from the GLPD by varying the parameters γ and β:

$$L_{\gamma,\beta,\sigma}(x) \xrightarrow[\gamma\to 0]{} W_{\beta,\sigma}(x) =: L_{0,\beta,\sigma}(x)$$

$$L_{\gamma,\beta,\sigma}(x) \xrightarrow[\beta\to 0]{} W_{\gamma,\sigma}(x) =: L_{\gamma,0,\sigma}(x)$$

Furthermore, if $X \sim W_{\beta,\sigma}$ then $Y := \exp(\beta/\gamma X) - 1$ is distributed according to $L_{\gamma,\beta,\sigma}$.

Exceedances under Power–Normalization

Let F and L be dfs such that

$$F^{[u]}\left(\text{sign}(x)\alpha(u)\,|x|^{\lambda(u)}\right) \xrightarrow[u\to\omega(F)]{} L(x), \tag{10}$$

where

$$F^{[u]}(x) = \frac{F(x)-F(u)}{1-F(u)}, \quad x \geq u.$$

We have

$$L(x) = W(\log(x))$$

or

$$L(x) = W(-\log(-x))$$

where W is a POT-stable df.

Super-Heavy tailed Exceedances

If $\omega(F) > 0$ and L is continuous, then GLPDs of the form

$$\hat{L}_{\gamma,\beta,\sigma}(x) = 1 - \left(1 + \frac{\gamma}{\beta}\log\left(x/\sigma\right)\right)^{-1/\gamma}, \quad x > \sigma,\ \beta, \sigma > 0, \gamma \in \mathbb{R}$$

are the only possible non degenerate limiting dfs in (10). Note, that Pareto dfs are included for $\gamma = 0$. The case $\gamma > 0$ yields super–heavy tails.

P–Pot Domains of Attraction

Let

$$\hat{L}_\gamma = \hat{L}_{\gamma,1,1}. \tag{12}$$

The p–pot domain of attraction of $\hat{L}_\gamma$, $\mathcal{D}_{p-pot}(\hat{L}_\gamma)$, is defined by the property

$$F \in \mathcal{D}_{p-pot}(\hat{L}_\gamma) \tag{13}$$

if (10) holds for F and $L = \hat{L}_\gamma$.
It can be shown that $F \in \mathcal{D}_{p-pot}(\hat{L}_\gamma)$ iff $F^*(\cdot) = F(\exp(\cdot))$ is in the pot domain of attraction of a GPD with parameter γ under linear normalization.

Domains of Attraction of log–Pareto dfs

Consider $\mathcal{D}_{p-pot}(\hat{L}_\gamma)$ for some $\gamma > 0$.

For $F \in \mathcal{D}_{p-pot}(\hat{L}_\gamma)$ F^* is in the pot–domain of attraction of a Pareto df.

In particular $1 - F^*$ is regular varying at infinity.

Thus the tail of F is slowly varying at infinity, and F is not in the pot–domain of attraction of any GPD.

P–Pot Stability

The log–Pareto df $\hat{L}_{\gamma,\beta,\sigma}$ is p–pot stable:

$$\hat{L}^{[u]}_{\gamma,\beta,\sigma}\left(\alpha_u x^{\lambda_u}\right) = \hat{L}_{\gamma,\beta,\sigma}(x), \quad x > 0 \tag{14}$$

for $\alpha_u = u\sigma^{-\lambda_u}$ and $\lambda_u = 1 + \gamma/\beta \log(u/\sigma)$ if $u > \sigma$ and $\lambda_u > 0$.
Alternatively

$$\hat{L}^{[u]}_{\gamma,\beta,\sigma}(x) = \hat{L}_{\gamma,\beta_u,u}(x), \quad x > u, \tag{15}$$

with $\beta_u = \beta + \gamma \log(u/\sigma)$. The dfs of excesses is given by

$$\hat{L}^{[u]}_{\gamma,\beta,\sigma}(x+u) = 1 - \left(1 - \frac{\gamma}{\beta_u} \log\left(1 + x/u\right)\right)^{-1/\gamma} \tag{16}$$

which is a GLPD $L_{\gamma,\beta_u,u}$ with scale parameter u.

Overview

Literature

[1] Coles, S. (2001). *An Introduction to the Statistical Modelling of Extreme Values*, Springer, London.

[2] Falk, M., Hüsler, J. and Reiss, R.–D. (1994). *Laws of Small Numbers: Extremes and Rare Events,* DMV–Seminar Bd 23, Birkhäuser, Basel, (2. Auflage 2004).

[3] Reiss, R.–D. (1993). *A Course on Point Processes,* Springer, New York.

Literature

[4] Reiss, R.–D. and Thomas, M. (1994). *Statistical Analysis of Extreme Values,* Birkhäuser, Basel, (2. Auflage 2001, 3. Auflage 2007).

[5] Tsay, R.S. (2005). *Analysis of Financial Time Series,* Wiley, New York.

[6] Tsay, R. S. (1999). Extreme Value Analysis of Financial Data. Manuscript, Univ. Of Chicago.