Bivariate generalized Pareto distribution in practice

Pál Rakonczai

Eötvös Loránd University, Budapest, Hungary

Minisymposium on Uncertainty Modelling
27 September 2011, CSASC 2011, Krems, Austria
Outline

- Short summary of extreme value theory (EVT), including:
  - univariate/multivariate case
  - representation of dependence structures within EVT
  - multivariate generalized Pareto distribution of type II
- Construction of new asymmetric models
- Properties and comparison with well-known models
- Applications on environmental data (wind speed / flood peak)
There is often interest in understanding how the extremes of two different processes are related to each other.

One possible solution is using asymptotic approach from EVT.

EVT provides limits for the distribution of extremes for a single/multivariate stationary process.
Limit for Maximum

Theorem (Fisher and Tippett, 1928)

If there exist \( \{a_n\} \) and \( \{b_n\} > 0 \) sequences such that

\[
P\left( \frac{M_n - a_n}{b_n} \leq z \right) \to G(z) \quad \text{as} \quad n \to \infty
\]

for some non-degenerate distribution \( G \), then

\[
G(x) = \exp \left\{ - \left( 1 + \xi \frac{x - \mu}{\sigma} \right)^{-\frac{1}{\xi}} \right\},
\]

where \( 1 + \xi \frac{x - \mu}{\sigma} > 0 \) and in such a case random variable belongs to the max-domain of attraction of a GEV distribution.

\( \rightarrow \) Often applied for block maxima (monthly, yearly, ...)
Limit for Threshold Exceedances

How to model large values over a given high threshold?

Theorem (Balkema and de Haan, 1974)

Suppose, that \( X_1, \ldots, X_n \overset{i.i.d.}{\sim} F \), where \( F \) belongs to the max-domain of attraction of GEV for some \( \xi, \mu \) and \( \sigma > 0 \). Then

\[
P(X_i - u \leq z | X_i > u) \to H(z) = 1 - \left(1 + \frac{\xi z}{\tilde{\sigma}}\right)^{-\frac{1}{\xi}} \quad \text{as } u \to z_+,
\]

where \( \tilde{\sigma} = \sigma + \xi (u - \mu) \).

→ Often applied for threshold exceedances over a high quantile of the obs.
Remarks

- The limit distributions are **not Gaussian**
- Strong link between GEV and GPD
- See examples below:

**GEV**

- Gumbel
- Frechet
- Weibull

**GPD**

\( \xi = 0 \)
\( \xi = 0.5 \)
\( \xi = -0.3 \)
Limit for Multivariate Maxima

- Component-wise maxima of $X_i = (X_{i,1}, ..., X_{i,d})$ is

$$M_n = \left(\bigvee_{i=1}^n X_{i,1}, ..., \bigvee_{i=1}^n X_{i,d}\right).$$

- If $X \sim F$ and there exist $a_n$ and $b_n > 0$ sequences of normalizing vectors, such that

$$P\left(\frac{M_n - a_n}{b_n} \leq z\right) = F^n(b_n z + a_n) \to G(z),$$

where $G_i$ margins are non-degenerate, then $G$ is said to be a multivariate extreme value distribution (MEVD).
Characterization I. - Resnick (1987)

- MEVD with unit Fréchet margins can be written as

\[ G_{\text{Fréchet}}(t_1, \ldots, t_d) = \exp\left(-V(t_1, \ldots, t_d)\right) \]

- \( V \) is called **exponent measure**

\[ V(t_1, \ldots, t_d) = \nu\left(([0, t_1] \times \cdots \times [0, t_d])^c\right) = \int_{S_d} \bigvee_{i=1}^{d} \left(\frac{w_i}{t_i}\right) S(d\mathbf{w}), \]

- \( S \) **spectral measure** is a finite measure on the \( d \)-dimensional unit simplex satisfying the equations

\[ \int_{S_d} w_i S(d\mathbf{w}) = 1 \quad \text{for} \quad i = 1, \ldots, d. \]
Some properties of the characterization I.

- Total mass of $S$ is always $S(S_d) = d$
- Density not only on the interior of $S_d$ but also on each of the lower-dimensional subspaces of $S_d$
- For independent margins $S(\{e_i\}) = 1$ for all $e_i$ vertices of the simplex
- For completely dependent margins $S((1/d, ..., 1/d)) = d$
Characterization II. (bivariate case) - Pickands (1981)

The dependence function $A(t)$ must satisfy the following three properties: (P)

1. $A(t)$ is convex
2. $\max\{(1 - t), t\} \leq A(t) \leq t$
3. $A(0) = A(1) = 1$

$$V(t_1, t_2) = \left(\frac{1}{t_1} + \frac{1}{t_2}\right)A\left(\frac{t_1}{t_1 + t_2}\right)$$
The copula of a MEVD $G$ is the distribution function of the random vector $(G_1(Y_1), \ldots, G_d(Y_d))$ and is given by

$$C(\mathbf{u}) = \exp\left(-V\left(\frac{-1}{\log u_1}, \ldots, \frac{-1}{\log u_d}\right)\right), \mathbf{u} \in [0, 1]^d.$$ 

Such a copula necessarily satisfies the stability property

$$C^s(\mathbf{u}) = C(u_1^s, \ldots, u_d^s)$$

and conversely if this property is satisfied than we get a copula of a MEVD.
Logistic Dependence (Gumbel copula)

\[ A(t) = (t^\alpha + (1 - t)^\alpha)^{1/\alpha} \]
\[ C(u, v) = \exp\left(-((-\log u)^\alpha + (-\log v)^\alpha)^{1/\alpha}\right) \]
**Logistic BEVD and its asymmetric extension**

\[ A(t) = (1 - \Phi_1)t + (1 - \Phi_2)(1 - t) + ((\Phi_1 t)^\alpha + (\Phi_2 (1 - t))^\alpha)^{1/\alpha} \]

\[ 0 \leq \Phi_1, \Phi_2 \leq 1 \]

\[ \rightarrow S(\{0, 1\}) = 1 - \Phi_2, S(\{1, 0\}) = 1 - \Phi_1 \]

![Diagram of Logistic BEVD and its asymmetric extension](image.png)
MGPD of type II

- Problem: taking maxima hides the time structure within the period.
- Might be better investigating all observations over a given high threshold

Let \( Y = (Y_1, \ldots, Y_d) \) denote a random vector, \( u = (u_1, \ldots, u_d) \) a high threshold vector and so \( X = Y - u \) the vector of exceedances. The distribution of \( X \not\leq 0 \) exceedances can be well approximated by the multivariate generalized Pareto distribution (Rootzén and Tajvidi, 2006)

\[
H(x_1, \ldots, x_d) = \frac{-1}{\log G(0, \ldots, 0)} \log \frac{G(x_1, \ldots, x_d)}{G(x_1 \wedge 0, \ldots, x_d \wedge 0)},
\]

where \( 0 < G(0, \ldots, 0) < 1 \).
What is type I. then?

- If we claim exceedance in all of the coordinates (BGPD I), we usually get simpler models, with nice properties (marginals are univariate GPD etc.)
- But we may use less data
Remarks

- $H_1(x) = H(x, \infty)$ is not a one dimensional GPD, only the conditional distribution of $X_1|X_1 > 0$ is GPD.
- The $\xi_i, \mu_i, \sigma_i$ parameters cannot be considered as the $i$th marginal parameters any more.
- Motivating problem: asymmetric logistic model does not lead to an absolutely continuous MGPD.

Theorem

Let $H$ be a MGPD represented by an absolutely continuous MEVD $G$ with spectral measure $S$. The MGPD $H$ is absolutely continuous if and only if $S(\text{int}(S_d)) = d$ holds, i.e. all mass is put on the interior of the simplex.
Overview of parametric dependence models

<table>
<thead>
<tr>
<th>Model</th>
<th>Asym.</th>
<th>Density</th>
<th>Complications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sym. Logistic</td>
<td>–</td>
<td>✓</td>
<td>–</td>
</tr>
<tr>
<td>Asym. Logistic (Tawn, 1988)</td>
<td>✓</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>$\Psi$-logistic*</td>
<td>✓</td>
<td>✓</td>
<td>convexity constr.</td>
</tr>
<tr>
<td>Sym. Neg. Logistic</td>
<td>–</td>
<td>✓</td>
<td>–</td>
</tr>
<tr>
<td>Asym. Neg. Logistic (Joe, 1990)</td>
<td>✓</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>$\Psi$-negative logistic*</td>
<td>✓</td>
<td>✓</td>
<td>convexity constr.</td>
</tr>
<tr>
<td>Bilogistic (Smith, 1990)</td>
<td>✓</td>
<td>✓</td>
<td>not explicit</td>
</tr>
<tr>
<td>Neg. Bilogistic (C. &amp; T., 1994)</td>
<td>✓</td>
<td>✓</td>
<td>not explicit</td>
</tr>
<tr>
<td>Tajvidi’s (Tajvidi, 1996)</td>
<td>–</td>
<td>✓</td>
<td>–</td>
</tr>
<tr>
<td>C-T (C. &amp; T., 1991)</td>
<td>✓</td>
<td>✓</td>
<td>only $s(w)$ is given</td>
</tr>
<tr>
<td>Mixed (Tawn, 1988)</td>
<td>–</td>
<td>if $\psi = 1$</td>
<td>–</td>
</tr>
<tr>
<td>Asym. Mixed</td>
<td>✓</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>
Construction of absolutely continuous BGPDs

1. take an absolutely continuous baseline dependence model;
2. take a strictly monotonic transformation
   \( \Psi(x) : [0, 1] \rightarrow [0, 1] \), such that \( \Psi(0) = 0, \Psi(1) = 1 \);
3. construct a new dependence model from the baseline model
   \( A_\Psi(t) = A(\Psi(t)) \);
4. check the constraints: \( A'_\Psi(0) = -1, A'_\Psi(1) = 1 \) and \( A''_\Psi \geq 0 \) (convexity).
An example for $\Psi(x)$

- Feasible transformation if

$$\Psi(t) = t + f(t) \Rightarrow (A(\Psi(t)))' = A'(\Psi(t)) \times [1 + f'(t)]$$

so choosing $f$ such that $f'(0) = f'(1) = 0$, the $A'_\Psi(0) = -1$ and $A'_\Psi(1) = 1$ are fulfilled.

- $f_{\psi_1,\psi_2}(t) = \psi_1(t(1-t))^{\psi_2}$, for $t \in [0, 1]$,

where $\psi_1 \in \mathbb{R}$ and $\psi_2 \geq 1$ are asymmetry parameters.
Graphical examples

\[ \delta^2 \eta_{\text{Log}}(\Psi(t)) \]

Constraints of convexity for psi–logistic models

\[ \delta^2 \eta_{\text{NegLog}}(\Psi(t)) \]

Constraints of convexity for psi–negative logistic models
Remarks

- Of course $f$ and or the baseline dependence function can be chosen differently.
- This transformation has particular impact beyond the BGPD models as well: applicable for BEVDs, BGPDs of type I or for copulas of any bivariate distributions.
- Although the transformations $\Psi$ was defined now for the bivariate case only, it can be extended for the higher dimensional cases as well as e.g. in the trivariate case it can have a form of

$$
\Psi(t_1, t_2) = (t_1 + \psi_{1,1}(t_1([1-t_2]-t_1))^{\psi_{2,1}}, t_2 + \psi_{2,1}(t_2([1-t_1]-t_2))^{\psi_{2,2}}).
$$
Further models - polynomial transformations

\[ f_{a,p}(t) = \frac{64a(p^5 - p^4 - 2p^3 - 2p^2 + 5p - 2)p^5}{(p + 1)^2(p - 1)^2(-1 + 2p)(1 + 2p^2 - 3p)} t^6 \]
\[ + \frac{-32a(5p^6 - 2p^5 - 13p^4 - 12p^3 + 15p^2 + 6p - 6)p^4}{(p + 1)^2(p - 1)^2(-1 + 2p)(1 + 2p^2 - 3p)} t^5 \]
\[ + \frac{32a(4p^7 + 3p^6 - 14p^5 - 15p^4 + 23p^2 - 8p - 2)p^3}{(p + 1)^2(p - 1)^2(-1 + 2p)(1 + 2p^2 - 3p)} t^4 \]
\[ - \frac{32a(p^7 + 4p^6 - 5p^5 - 10p^4 - 5p^3 + 12p^2 + 2p - 4)p^3}{(1 + 2p^2 - 3p)(p + 1)^2(4p - 1 + 2p^3 - 5p^2)} t^3 \]
\[ + \frac{32a(p^6 - 3p^4 - p^2 + 4p - 2)p^3}{(1 + 2p^2 - 3p)(p + 1)^2(4p - 1 + 2p^3 - 5p^2)} t^2. \]
Further models - polynomial transformations

Pál Rakonczai

Bivariate generalized Pareto distribution
BEVD case
Wind data

- Daily windspeed values at two locations in northern Germany (Bremerhaven and Schleswig) over the past five decades (1957-2007)
- Originally hourly windspeed was recorded, but in order to reduce the serial correlation within the series, we calculated the maxima of the original observations for each day.
- The data for the above cities are rather strongly correlated, due to the relatively short distance between them (Kendall’s correlation is 0.57) and also can be considered as asymptotically dependent.
- 95% quantiles have been used as threshold level (13.8m/s, 10.3m/s)
- Observations are considered as being stationary, non-stationarity can also be handled e.g. by choosing time-dependent model parameters.
Wind data

Exceedances (m/s)

Copula sample

Bivariate generalized Pareto distribution
Parameter estimation

- We found that maximum likelihood methods perform well in general
- However there is strong pairwise correlation between the parameters $\mu_i, \sigma_i, i = 1, 2$
- This can be extinguished in practice by keeping e.g. one location parameter fixed during the optimization process
- Estimation within this subclass gives stable convergence results

Prediction regions (Hall and Tajvidi, 2000) have been computed - the smallest region containing an observation with a given (usually high) probability.
Prediction regions for the 4 best models (having the smallest $\chi^2$-statistics)

The effect of asymmetry parameters - relevant torsion in the regions:

- **NegLog**
- **Psi–NegLog**
- **Coles–Tawn**
- **Tajvidi’s**
### Goodness-of-fit: Bremerhaven és Schleswig

<table>
<thead>
<tr>
<th>$\gamma$</th>
<th>0.99</th>
<th>0.95-0.99</th>
<th>0.9-0.95</th>
<th>0.75-0.9</th>
<th>0.5-0.75</th>
<th>0.5</th>
<th>$\chi^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exp.</td>
<td>12.6</td>
<td>50.3</td>
<td>62.8</td>
<td>188.6</td>
<td>314.2</td>
<td>628.5</td>
<td></td>
</tr>
<tr>
<td>Log</td>
<td>7</td>
<td>30</td>
<td>66</td>
<td>216</td>
<td>350</td>
<td>588</td>
<td>21.5</td>
</tr>
<tr>
<td>$\psi$-L</td>
<td>9</td>
<td>33</td>
<td>60</td>
<td>215</td>
<td>349</td>
<td>591</td>
<td>16.9</td>
</tr>
<tr>
<td>NL</td>
<td>11</td>
<td>31</td>
<td>66</td>
<td>215</td>
<td>334</td>
<td>600</td>
<td>14.0</td>
</tr>
<tr>
<td>$\psi$-NL</td>
<td>11</td>
<td>34</td>
<td>63</td>
<td>210</td>
<td>337</td>
<td>602</td>
<td>10.7</td>
</tr>
<tr>
<td>Mix</td>
<td>12</td>
<td>46</td>
<td>71</td>
<td>247</td>
<td>331</td>
<td>550</td>
<td>30.3</td>
</tr>
<tr>
<td>C-T</td>
<td>12</td>
<td>34</td>
<td>65</td>
<td>209</td>
<td>330</td>
<td>607</td>
<td>9.1</td>
</tr>
<tr>
<td>Tajv.</td>
<td>11</td>
<td>35</td>
<td>62</td>
<td>212</td>
<td>340</td>
<td>597</td>
<td>11.5</td>
</tr>
<tr>
<td>BL</td>
<td>6</td>
<td>29</td>
<td>62</td>
<td>220</td>
<td>354</td>
<td>586</td>
<td>25.6</td>
</tr>
<tr>
<td>NBL</td>
<td>13</td>
<td>32</td>
<td>61</td>
<td>220</td>
<td>337</td>
<td>594</td>
<td>15.5</td>
</tr>
</tbody>
</table>
Introduction

Extreme value models

Applications

Wind data

Flood data

Software

Results

▶ C-T ($\chi^2_{CT} = 9.1$) and $\Psi$-Neglog models ($\chi^2_{\Psi-NegLog} = 10.7$) perform the best

▶ In general the new asymmetric ($\Psi$) models are better than their symmetric baseline versions, $\chi^2_{Log} = 21.5$ reduces to $\chi^2_{\Psi-Log} = 16.9$ and $\chi^2_{NegLog} = 14$ reduces to $\chi^2_{\Psi-NegLog} = 10.7$

▶ The difference between the log-likelihood values of the Neglog and $\Psi$-NegLog models is around 10, which shows a significant improvement in the likelihood ratio as well

▶ The Tajvidi’s generalized symmetric logistic model ($\chi^2_{Tajvidi’s} = 11.5$) is the best among the symmetric ones.
Fitted BGPD models for flood data from Budapest and Tass (threshold was chosen at about the 95% quantiles. Black region covers 95%, green: 90%, blue: 75%).
R package available

- Graphs and computations have been made by the add-on package mgpd of R
- Wind data are also available
- Updated version is coming up very soon (mgpd v2.0)

**Acknowledgements:** The European Union and the European Social Fund have provided financial support to the project under the grant agreement no. TÁMOP 4.2.1./B-09/KMR-2010-0003.
Thank you for your attention!