EVT and Angular Distributions

Peter Trubey
UCSC Department of Statistics

Bruno's Lab Meeting
12/17-18/2020

Overview

Intruction

- Our goal is a framework for anomaly detection using extreme value theory
- blah

EVT - A brief introduction

- Why EVT?
- Block Maxima - GEV
- Thresholding - GP

Block Maxima - GEV

- GEV (Generalized Extreme Value Distribution) is a unification of three distributions: Frechet, Gumbel, and Weibull.
- For i.i.d. $x_1, \ldots, x_n$ $x_i \sim F$, $M_n = \max_i x_i$, if there exists some sequence $a_n > 0$, b_n such that:

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

Then $G(z)$ takes the form:

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

Then F falls into the *domain of attraction* of said GEV.

Thresholding - GP

- We look now at the distribution of excesses y over a threshold u.

  $$\Pr[X > u + y \mid X > u] = \frac{1 - F(u+y)}{1 - F(u)}, \qquad y > 0$$

  If F is in the domain of attraction of a GEV, then $\Pr[X > u + y \mid X > u]$ converges to

  $$H(y) = 1 - \left(1 + \frac{\xi y}{\sigma}\right)^{-\xi^{-1}},$$

  the Generalized Pareto distribution.
- We set a high threshold by some criteria, then model excesses over that threshold using GP.

Multivariate EVT

- Useful to standardize each X_i according to its marginal distribution
- Standardization occcurs as:

$$z_j = \left(1 + \xi \frac{x_j - b_{t,j}}{a_{t,j}}\right)_+^{1/\xi} \tag{1}$$

  where $b_{t,j} = \hat{F}^{-1}(1 - 1/t)$.
- Note that $Z_j > 1 \implies X_j > b_{t,j}$
- $\max_j Z_j \sim$ Pareto
- a, ξ are currently found by MLE. Not ideal, I know.

Multivariate EVT

- Assume the existence of a limit measure μ on $\mathbf{Z}$ such that:

$$n\Pr\left(\frac{V_1}{n} \geq v_1 \text{ or } \ldots \text{ or } \frac{V_d}{n} \geq v_d\right) \to \mu\left([\mathbf{0}, \mathbf{v}]^C\right)$$

- μ is the asymptotic distribution of $\mathbf{Z}$ in extreme regions.
- μ features the homogeneity property, $\mu(t\cdot) = t^{-1}\mu(\cdot)$.

Pseudo-Polar Representation

Given the homogeneity property, we can decompose $\mathbf{Z}$ into two components, radial and angular:

$$R(\mathbf{Z}) = \|\mathbf{Z}\|_\infty = \max_i v_i$$

$$\mathbf{V} = \frac{\mathbf{Z}}{R} \in S_\infty^{d-1}$$

where S_∞^{d-1} is the positive orthant of the unit hypercube in R^d.

Spectral Measure

For $B \subset S_\infty^{d-1}$, define the *Spectral Measure*:

$$\Omega(B) = \mu[\mathbf{z} : R(\mathbf{z}) > 1, \mathbf{V} \in B].$$

Then,

$$\mu[\mathbf{z} : R(z) > t, \mathbf{V} \in B] = t^{-1}\Omega(B).$$

Thus t is independent of Ω, the spectral measure. This is completed as a probability measure as:

$$\Pr\left(\mathbf{V} \in B \mid r > 1\right) = \frac{\Omega(B)}{\Omega(S_\infty^{d-1})}.$$

So conditional on at least one dimension exceeding its threshold, we can hold the angular measure independent of the magnitude.

Polar Coordinate Transformation

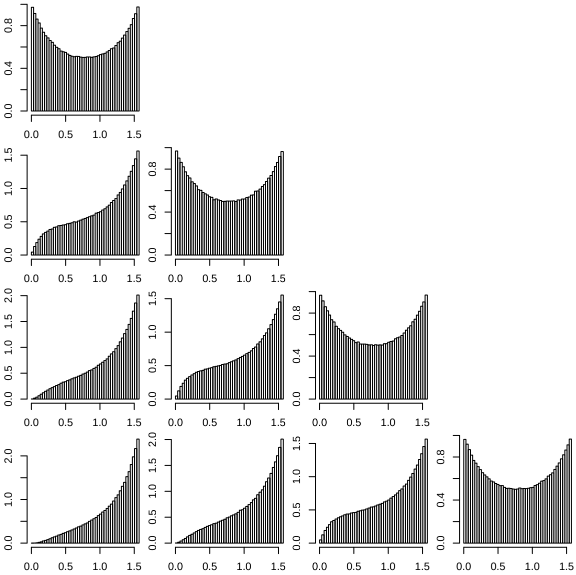
- One way

$$
\begin{aligned}
y_1 &= r\cos\theta_1, \\
y_2 &= r\sin\theta_1\cos\theta_2 \\
&\vdots \\
y_{d-1} &= r\sin\theta_1\ldots\sin\theta_{d-2} \\
y_d &= r\sin\theta_1\ldots\sin\theta_{d-1}
\end{aligned}
$$

- The other way

$$
\begin{aligned}
\theta_1 &= \cos^{-1}\left[\frac{y_1}{\|y_{1:d}\|_2}\right] \\
\theta_2 &= \cos^{-1}\left[\frac{y_2}{\|y_{2:d}\|_2}\right] \\
&\vdots \\
\theta_{d-1} &= \cos^{-1}\left[\frac{y_{d-1}}{\|y_{(d-1):d}\|_2}\right].
\end{aligned}
$$

Polar Coordinate Transformation Effects



Projected Gamma Distribution

Consider d independent Gammas

$$f(\mathbf{y} \mid \alpha, \beta) = \prod_{j=1}^{d} \text{Ga}(y_j \mid \alpha_j, \beta_j), \tag{2}$$

Use polar coordinate transformation, then integrate out r.

$$\text{PG}(\theta \mid \alpha, \beta) = \frac{\Gamma(A)\beta_d^{\alpha_d}}{B^A\Gamma(a_d} \left(\prod_{j=1}^{d-1} \frac{\beta_j^{\alpha_j}}{\Gamma(\alpha_j)} (\cos\theta_j)^{\alpha_j - 1} (\sin\theta_j)^{(\sum_{h=j+1}^{d} \alpha_h) - 1} \right) \mathcal{I}_{(0,\pi/2)^{d-1}}(\theta) \tag{3}$$

where

$$A = \sum_{j=1}^{d} \alpha_j \qquad \text{and} \qquad B = \beta_1 \cos\theta_1 + \sum_{j=2}^{d-1} \left(\beta_j \cos\theta_j \prod_{i=1}^{j-1} \sin\theta_i \right) + \beta_d \prod_{j=1}^{d} -1 \sin\theta_j. \tag{4}$$

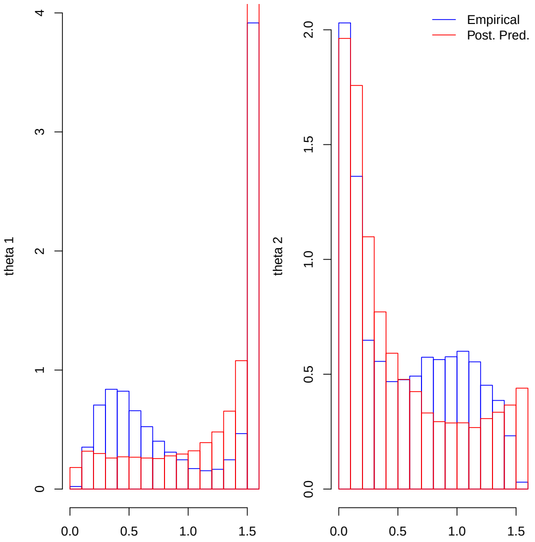
Projected Gamma, Cont.

- Let $\mathbf{y}' = \frac{\mathbf{y}}{r}$, $\mathbf{y} = r\mathbf{y}'$.
- Gibbs sampling updates to α, β possible with latent r.

$$r_i \mid \alpha, \beta \sim \text{Ga}\left(r_i \mid \sum_j \alpha_j, \sum_j \beta_j y'_{ij}\right)$$

- For Projected Gamma, information between dimensions is effectively communicated through the latent r.

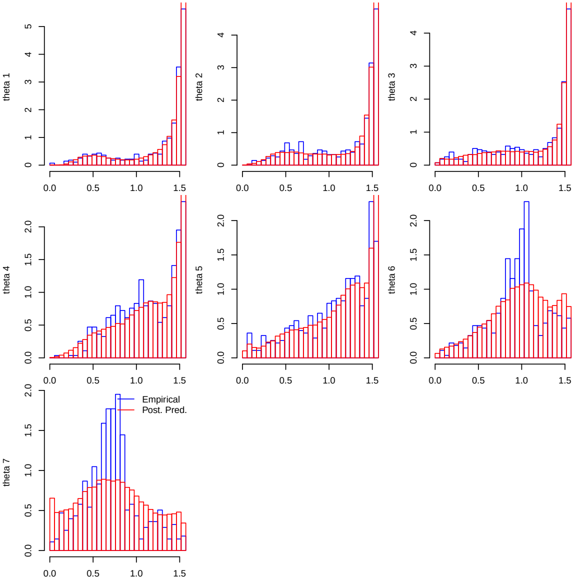
Projected Gamma, Cont.



Finite Mixture of Projected Gammas

$$
\begin{aligned}
\text{MPG}(\theta \mid \lambda, \alpha, \beta) &= \sum_{j=1}^{J} \lambda_i \text{PG}(\alpha_i, \beta_i) \\
&= \int \text{MPG}(\theta, \gamma \mid \alpha, \beta) d\gamma \\
&= \int \prod_{j=1}^{J} [\lambda_j \text{PG}(\theta \mid \alpha_j, \beta_j)]^{\gamma_j} \, \text{d}\gamma \\
&= \int_{\gamma} \int_{r} \lambda_j^{\gamma_j} \prod_{k=1}^{K} \text{Ga}(r\mathbf{y}' \mid \alpha_{\mathbf{j}}, \beta_{\mathbf{j}})^{\gamma_j} |\text{Jac}| \text{d}r \text{d}\gamma
\end{aligned}
$$

Finite Mixture of Projected Gammas - cont.



DP Mixture of Projected Gammas, Independent Gamma Prior

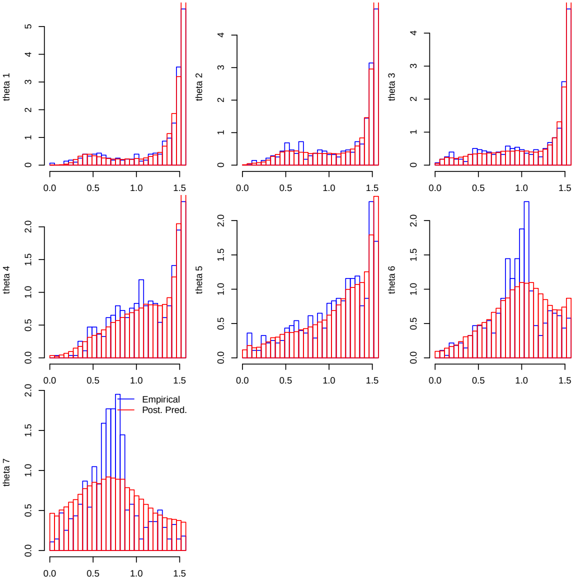
$$
\begin{aligned}
\theta_i &\sim \mathrm{PG}\left(\theta_i \mid (\alpha_i, \beta_i)\right) \\
(\alpha_i, \beta_i) &\sim G_i \\
G_i &\sim \mathrm{DP}\left(\eta, G_0\left((\alpha_i, \beta_i) \mid (\mathbf{a}_\alpha, \mathbf{b}_\alpha, \mathbf{a}_\beta, \mathbf{b}_\beta)\right)\right) \\
(\mathbf{a}_\alpha, \mathbf{b}_\alpha, \mathbf{a}_\beta, \mathbf{b}_\beta) &\sim P\left((\mathbf{a}_\alpha, \mathbf{b}_\alpha, \mathbf{a}_\beta, \mathbf{b}_\beta)\right) \\
\eta &\sim \mathrm{Ga}(a_\eta, b_\eta)
\end{aligned}
$$

with

$$
G_0\left((\alpha_i, \beta_i)\right) = \mathrm{Ga}(\alpha_1 \mid a_{\alpha_1}, b_{\alpha_1}) \prod_{j=2}^{d} \mathrm{Ga}(\alpha_j \mid a_{\alpha_j}, b_{\alpha_j}) \mathrm{Ga}(\beta_j \mid a_{\beta_j}, b_{\beta_j})
$$

$$
P((\mathbf{a}_\alpha, \mathbf{b}_\alpha, \mathbf{a}_\beta, \mathbf{b}_\beta)) = \mathrm{Ga}(a_{\alpha_1}) \mathrm{Ga}(b_{\alpha_1}) \prod_{j=2}^{d} \mathrm{Ga}(a_{\alpha_j}) \mathrm{Ga}(b_{\alpha_j}) \mathrm{Ga}(a_{\beta_j}) \mathrm{Ga}(b_{\beta_j})
$$

DP Mixture of Projected Gammas, Independent Gamma Prior - Cont.

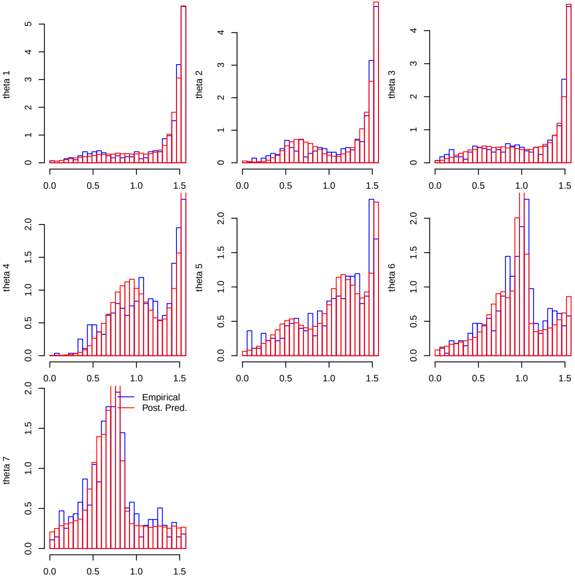


DP Mixture of Multivariate Normals on Probit Space

- $\theta \in [0, \pi/2]$
  - Jitter and rescale to $\theta' \in [\epsilon, 1-\epsilon]$
  - Then $W = \Phi^{-1}(\theta')$
- Then develop a hierarchical multivariate normal, with length of response $d-1$.

$$
\begin{aligned}
W_i &\sim \mathcal{N}_d\left(\mu_i, \Sigma_i\right) \\
\mu_i, \sigma_i &\sim G_i \\
G_i &\sim \mathrm{DP}(\eta, G_0(\mu_i, \Sigma_i \mid \mu_0, \Sigma_0)) \\
G_0(\mu_i, \Sigma_i \mid \mu_0, \Sigma_0) &= \mathcal{N}_d(\mu_i \mid \mu_0, \Sigma_0)\mathrm{IW}(\Sigma \mid \nu, \psi) \\
\mu_0 &\sim \mathcal{N}_d\left(\mathbf{u}, \mathbf{S}\right) \\
\Sigma_0 &\sim \mathrm{IW}(\nu_0, \psi_0) \\
\eta &\sim \mathrm{Ga}(\alpha, \beta)
\end{aligned}
$$

DP Mixture of Multivariate Normals on Probit Space - cont.

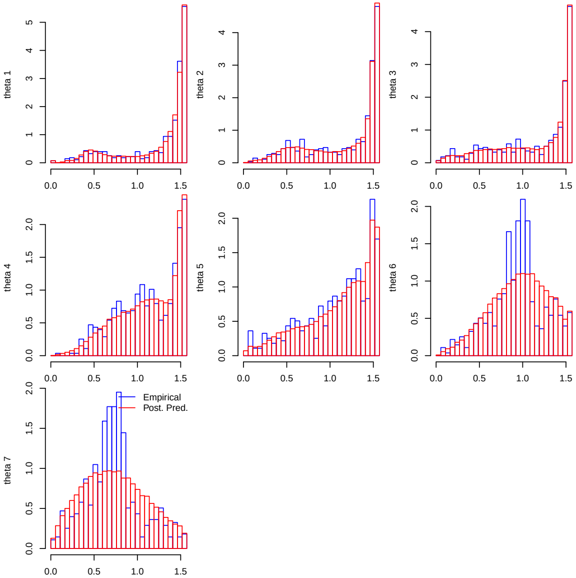


DP Mixture of Projected Gammas, LogNormal Prior

- The Normal Normal model operated in $d-1$-dimensional space without considering the transformation that brought it there.
- If we want to operate in d-dimensional space with $d-1$ degrees of freedom, then Projected Gamma is a good base distribution.
- Placing a multivariate log-normal prior on α, we can *potentially* communicate information across dimensions more effectively than with r alone.

$$
\begin{aligned}
\theta_i &\sim \text{PG}(\theta_i \mid \alpha_i, \beta_i) \\
r_i &\sim \text{Ga}(r_i \mid \alpha_i, \beta_i) \\
(\alpha_i, \beta_i) &\sim \text{DP}\left((\alpha_i, \beta_i) \mid \eta, G_0\right) \\
G_0 &= \text{Log}\mathcal{N}(\alpha_i \mid \mu, \Sigma) \prod_{j=2}^{d} \text{Ga}(\beta_{ij} \mid a, b) \\
\mu &\sim \mathcal{N}(\mu \mid \mu_0, \Sigma_0) \\
\Sigma &\sim \text{IG}(\Sigma \mid \nu, \psi)
\end{aligned}
$$

DP Mixture of Projected Gammas, LogNormal Prior - cont.



Conclusion

- Conclusions?