

# 5. előadás, 2022. március 10.

Zempléni András

Valószínűségelméleti és Statisztika Tanszék  
Természettudományi Kar  
Eötvös Loránd Tudományegyetem

Áringadozások előadás

- Jó lenne több adatot használni
- Valódi megfigyelések
- Klasszikus modell: minden koordinátában haladja meg a magas küszöböt. Tulajdonságok:
  - Peremek: GPD
  - Paraméteres modellek egy része átvihető
  - EVD csomag használható
  - De: valójában kevés az adat

# Alternatív definíció (MGPD II)

- Minden megfigyelést figyelembe vesz, ami legalább egy koordinátában meghaladja a küszöböt
- Formálisan:  $\underline{Y} = (Y_1, \dots, Y_d)$  vektorváltozó,  $\underline{u} = (u_1, \dots, u_d)$  megfelelően magas küszöb  $\underline{X} = \underline{Y} - \underline{u} = (Y_1 - u_1, \dots, Y_d - u_d)$  a meghaladások.
- A többdimenziós általánosított Pareto eloszlás (MGPD):

$$H(x_1, \dots, x_d) = \frac{-1}{\log G(0, \dots, 0)} \log \frac{G(x_1, \dots, x_d)}{G(x_1 \wedge 0, \dots, x_d \wedge 0)},$$

ahol  $G$  MGEV eloszlású

- Standard Fréchet marginálisokat kapunk a

$$t_i = t_i(x_i) = \frac{-1}{\log \mathcal{G}_{\xi_i, \mu_i, \sigma_i}(x_i)} = (1 + \xi_i(x_i - \mu_i)/\sigma_i)^{1/\xi_i},$$

transzformációval, ahol  $1 + \xi_i(x_i - \mu_i)/\sigma_i > 0$  és  $\sigma_i > 0$ ,  $i = 1, \dots, d$ .

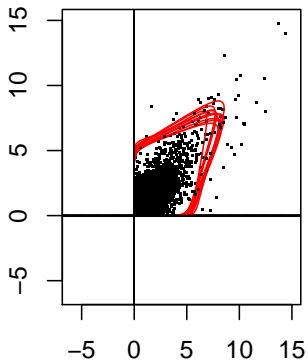
- Az MGPD sűrűségfüggvénye (ha létezik)

$$\begin{aligned}h(\underline{x}) &= \frac{\partial H}{\partial x_1 \dots \partial x_d}(\mathbf{x}) = \frac{\partial}{\partial x_1 \dots \partial x_d} \left( 1 - \frac{\log G(\underline{x})}{\log G(\underline{0})} \right) \\ &= \frac{\prod_{i=1}^d t'_i(x_i)}{V(t_1(0), \dots, t_d(0))} \times \frac{\partial V}{\partial t_1 \dots \partial t_d} (t_1(x_1), \dots, t_d(x_d)).\end{aligned}$$

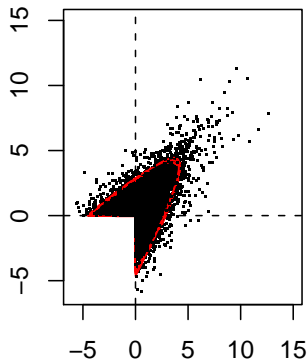
- Nem lesznek a peremek GPD-k
- Kevés a jól használható, identifikálható modell
- Becslés R-ben: mgpd csomaggal, de kettőnél több dimenzióban nem könnyű
- Összehasonlítás: több adat alapján valóban megbízhatóbb becslések adódnak

# Szimmetrikus és aszimmetrikus modellek összehasonlítása

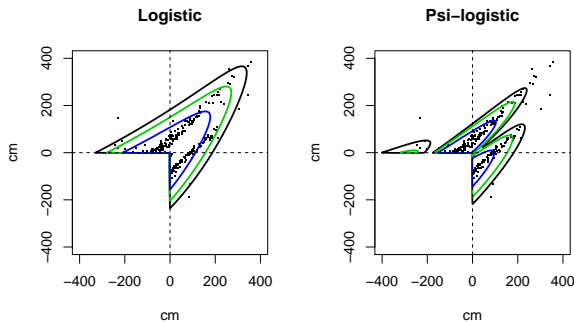
**BGPD-I**



**BGPD-II**



# Kétfajta 2D modell összehasonlítása



$A(\Psi(t)) = A(t + f(t))$ , például  $f_{\psi_1, \psi_2}(t) = \psi_1(t(1-t))^{\psi_2}$ , ha  $t \in [0, 1]$ , ahol  $\psi_1 \in \mathbb{R}$  és  $\psi_2 \geq 1$  aszimmetria paraméterek

- Itt az exponenciális peremeloszlást tételezzük fel
- Legyen  $Y_i = Z - \max(X_1, \dots, X_d) + X_i$ , ahol  $Z$  standard exponenciális eloszlású és  $\underline{X}$  független  $Z$ -től. Ekkor  $Y$  MGPD II eloszlású, standard exponenciális peremeloszlással, feltéve, hogy  $Y_i > 0$ .
- Sőt, minden MGPD eloszlás megadható így
- Bizonyos  $X_i$  generátorokra a sűrűségfüggvény explicit számolható
- ML becslés működik

- Hill becslés
- Pickands becslés
- Momentum becslés
- ML becslés az exponenciális regresszióra



- Vastag szélű eloszlásra:  $P(X > z) = z^{-\alpha}L(x)$ .
- $\log(X)$  eloszlása:  $P(\log X > u) = e^{-\alpha u}L(e^u)$ .
- Kvantilis függvény:  $Q(1 - p) = p^{-\gamma}L^*(1/p)$ , tehát

$$\log Q(1 - p) = -\gamma \log p + \log L^*(1/p).$$

- Az  $X_1, X_2, \dots, X_n$  rendezett mintája:  $X_1^* \leq X_2^* \leq \dots \leq X_n^*$ .
- A rendezett minta elemei konzisztens becslések a megfelelő kvantilisre.
- Ábrázolás:  $\log X_{n-j+1}^*$  vs  $\log \frac{j}{n+1}$ .
- Aszimptotikusan lineáris, meredekség  $\gamma$ .

- A meredekség becslése:

$$\frac{\frac{1}{k} \sum_{j=1}^k \left( \log(X_{n-j+1}^*) - \log(X_{n-k+1}^*) \right)}{-\frac{1}{k} \sum_{j=1}^k \left( \log\left(\frac{j}{n+1}\right) - \log\left(\frac{k}{n+1}\right) \right)}$$

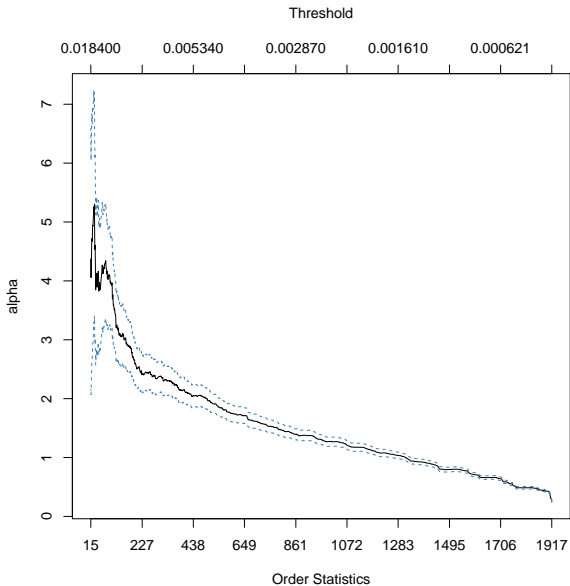
- A nevező közel 1, ha a  $k$  nagy
- Ebből a Hill becslés:

$$H_{k,n} = \frac{1}{k} \sum_{j=1}^k \left( \log(X_{n-j+1}^*) - \log(X_{n-k+1}^*) \right).$$

- Tul.:
  - függ a  $k$ -tól
  - $k$  kicsi: nagy szórás
  - $k$  nagy: nagy torzítás
  - Kompromisszumot kell találni

- $\hat{\xi}$  pontosan akkor konzisztens becslése az alakparaméternek, ha  $k \rightarrow \infty$  és  $k/n \rightarrow 0$  ha  $n \rightarrow \infty$ .
- A double bootstrap módszer azt a  $k$  értéket adja meg, amelyre az aszimptotikus négyzetes hiba minimális, bootstrap mintavétellel számolva.
- Egy másik módszer minimalizálja a távolságot a tapasztalati eloszlás és az illesztett Pareto eloszlás "széle" között. A minimalizálásra a kvantilisok Kolmogorov-Smirnov távolsága alkalmas.

# Példa: GBP vs DM napi hozam



- Hasznos eszköz, könnyű a modellek általánosítása a segítségükkel
- A minta véletlen pontok sorozata, össz-mérték most  $n$
- Az extrémumok érdekelnek:  $(t, \infty)$ -be eső pontokkal foglalkozunk
- Poisson pontfolyamat: Az egydimenziós Poisson folyamat általánosítása
  - $N(B)$ : a  $B$  Borel halmazba eső megfigyelések száma.  
Tulajdonságok:
  - Tetszőleges diszjunkt  $A, B$  halmazokra  $N(A)$  és  $N(B)$  független
  - $\lambda \geq 0$  adott mérték a Borel halmazokon (intenzitás), ezzel  $N(B)$  éppen  $\lambda(B)$  paraméterű Poisson eloszlású
  - Homogén a folyamat, ha  $\lambda$  a Lebesgue mérték (ekkor a korlátos  $B$  halmazokon nézzük)

- A  $\lambda$  intenzitás feltételezésünk szerint egy paraméteres családból származik

$$L(\vartheta; x_1, \dots, x_n) = \exp\{-\Lambda(A; \vartheta)\} \prod_{i=1}^n \lambda(x_i; \vartheta)$$

adja meg a likelihood értékét, ha egy  $A$  tartományból vannak megfigyeléseink.

- Kapcsolat a GPD modellel: az intenzitás-mérték az  $u_n$  meghaladásaira

$$\Lambda(A) = (t_2 - t_1) \left[ 1 + \xi \frac{z}{\sigma} \right]^{-1/\xi}$$

a  $(t_1; t_2) \times (z, \infty)$  tartományra, ahol  $z > u_n$

- Ez a megközelítés bonyolultabb modelleket is lehetővé tesz, pl. a paraméterek időfüggését.

- Legyen  $T \subset \mathbb{R}^d$  egy Borel-halmaz.  $\{Y_t : t \in T\}$  pontosan akkor max-stabilis folyamat, ha előáll folytonos trajektóriájú folyamatok koordinátánkénti (standardizált) extrémumaiként.
- Ezekre definíció szerint teljesül a max-stabilitás
- Példa:  $(r_j, s_j)$  Poisson pontfolyamat  $(0, \infty) \times S$  halmazon, intenzitásmértéke

$$\frac{dr}{r^2} dH(\omega).$$

$S$  tetszőleges Borel halmaz,  $H$  mérték  $S$ -en.

- Legyen  $f$  olyan, hogy  $\int_S f(s, t) dH(s) = 1$  minden  $t$ -re, és

$$Y_t = \max_i \{r_i f(s_i, t)\}, t \in T$$

- $r_i$  az  $i$ -edik vihar erőssége,  $s_i$  pedig a helye.

$$P(Y_t < y_t \forall t \in T) = \exp \left\{ - \int_S \max_t \left\{ \frac{f(s_i, t)}{y_t} \right\} H(ds) \right\}.$$

- Ebből:
  - $Y$  peremeloszlása standard Frechet
  - $Y$  max-stabilis



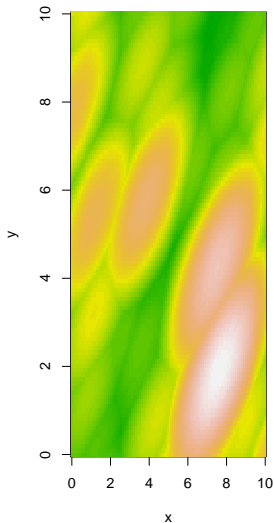
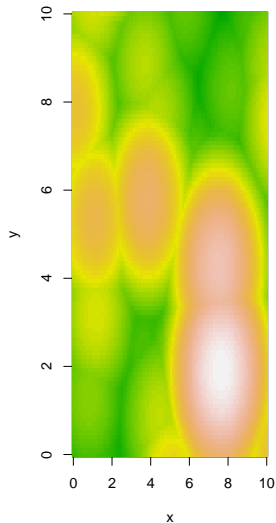
- $|T| = 1$ : egydimenziós max-stabilis eloszlás
- $T = \{1, 2\}$ ,  $S = [0, 1]$ ,  $H$ : Lebesgue mérték,

$$f(s, t) = \begin{cases} (1 - \alpha)s^{-\alpha}, & \text{hat} = 1 \\ (1 - \alpha)(1 - s)^{-\alpha}, & \text{hat} = 2 \end{cases}$$

éppen a 2 dimenziós logisztikus modell

- Gauss folyamat:  $f(s, t)$   $t$ -ben az  $s$  várható értékű,  $\Sigma$  kovariancia-mátrixú normális eloszlás sűrűségfüggvénye

# Példa: szimulált Smith-féle extrémális folyamatok



- 1 dimenziós peremek becslése
- 2 dimenziós összefüggőség becslése (extremális összefüggőségi függvény):  $\vartheta(z_1 - z_2)$ , ahol

$$P(Y(z_1) < y, Y(z_2) < y) = P(Y(z_1) < y)^{\vartheta(z_1 - z_2)}.$$

- Paraméteres (pl. Gauss) modellre közelítő (páronkénti) maximum likelihood számolható. Később még visszatérünk rá

- Danielsson, J., de Haan, L., Peng, L. and de Vries, C. G. (2001). Using a bootstrap method to choose the sample fraction in tail index estimation
- Danielsson, J., Ergun, L. M., De Haan, L. and de Vries, C. G. (2016). Tail Index Estimation: Quantile Driven Threshold Selection
- R.L. Smith (1990) Max-Stable Processes and Spatial Extremes. [www.stat.unc.edu/postscript/rs/spatex.pdf](http://www.stat.unc.edu/postscript/rs/spatex.pdf)
- Schlather, M. and Tawn, J. (2003) A dependence measure for multivariate and spatial extreme values: Properties and inference.
- Rootzén, H. and Tajvidi, N. (2006) The multivariate generalized Pareto distribution. *Bernoulli* 12, p.917-930.
- A. Kirilouk et al. (2018) Peaks Over Thresholds Modeling With Multivariate Generalized Pareto Distributions.
- Rakonczai, P.: Multivariate Threshold Models with Applications to Wind Speed Data (Ph.D. thesis, 2012)