Problémy s extrémy

Jan Dienstbier
dienstbi@karlin.mff.cuni.cz

Tento poster je věnován některým aspektům teorie extremálních hodnot (extreme value theory). Začněme stručným uvedením problému:

Data: $X_i \sim F(x)$, $i = 1, \dots, n$ $\longrightarrow$ **Chceme**: odhad psti. vzácných událostí $P(X > z_n)$, $z_n \to 1$ $\longrightarrow$

Fisher-Tippet: Pokud standardizovaná maxima $X_{n,n}$ konvergují, limitní rozdělení může být pouze generalizované extrémní rozdělení

$$G_\gamma(x) = \begin{cases} \exp\left(-(1+\gamma x)^{-1/\gamma}\right) & \gamma \neq 0 \\ \exp\left(-e^{(-x)}\right) & \gamma = 0 \end{cases}.$$

$\longrightarrow$

Centrální role γ

1. To jak jsou chvosty těžké určuje γ.
2. Je důležité odhadnout dobře γ.
3. Chvost F pak lze aproximovat G_γ.

Fisher-Tippetovu větu lze ekvivalentně napsat v teorii pravidelně se měnících funkcí:

kvantilová funkce chvostů $Q(t) := F^{-1}\left(1 - \frac{1}{t}\right)$ $\Updownarrow$ **vysoké pořádkové statistiky** $X_{n-[k_n t]:n}, t \in [0,1]$ $\longrightarrow$

Podmínka prvního řádu (rozepsáno pro těžké chvosty t.j. $\gamma > 0$, jinak obdobně)

$F \in \mathcal{D}(G_\gamma)$ a $\gamma > 0$ právě tehdy, když

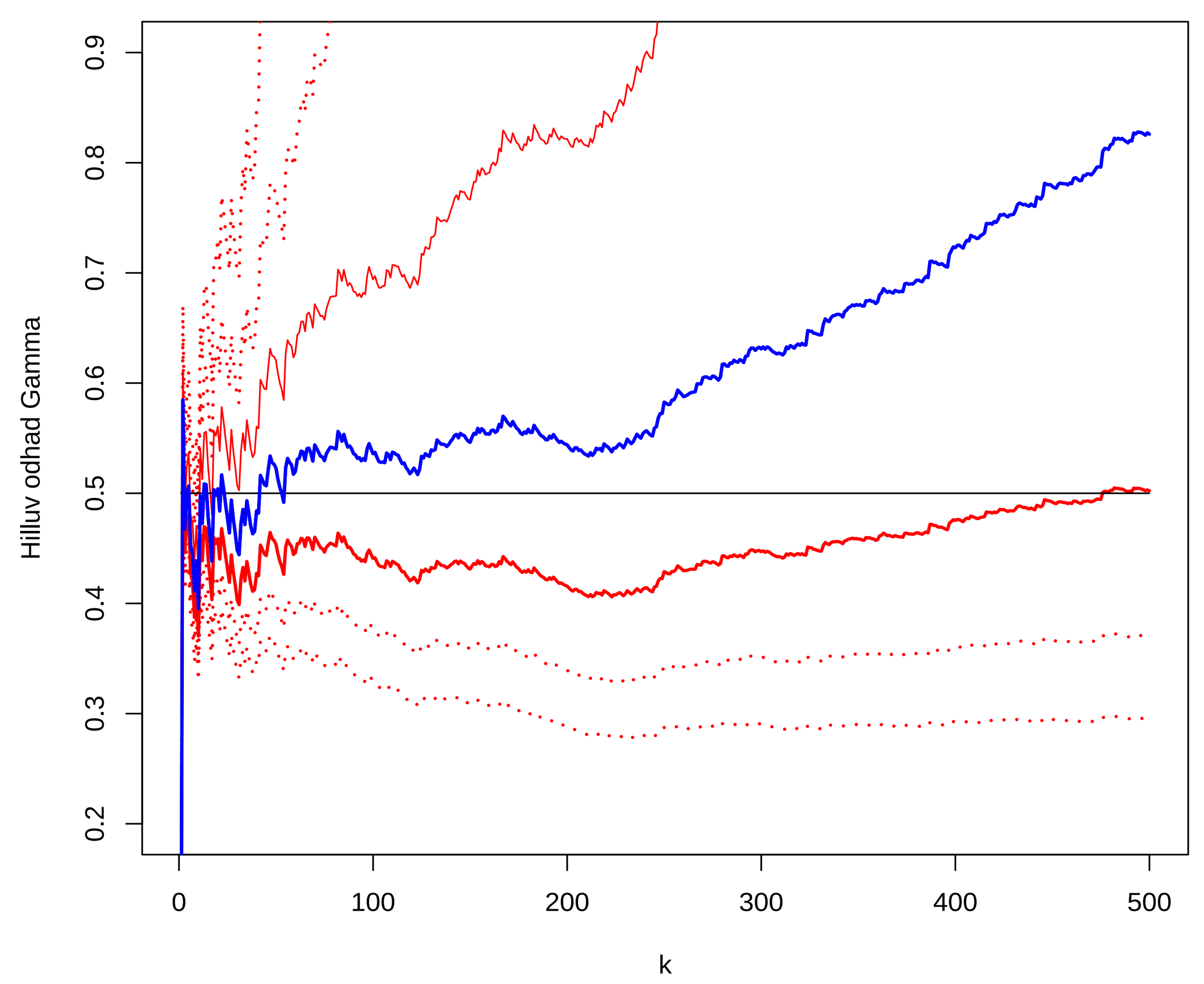
$$\lim_{t \Downarrow 0} \frac{Q(tx)}{Q(t)} = x^\gamma, \qquad \forall x > 0.$$

$\Longrightarrow$

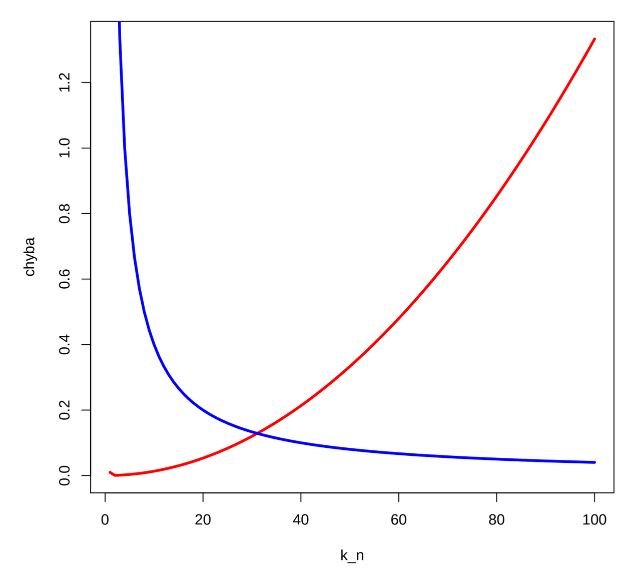
Podobně lze napsat i důležitou expanzi Q druhého řádu

Nechť existuje funkce $A(t)$, že $sign(A(t)) = C$ a $\rho \in \mathbb{R}$ tak, že

$$\lim_{t \Downarrow 0} \frac{\frac{Q(tx)}{Q(t)} - x^\gamma}{A(t)} = x^\gamma \cdot \frac{x^\rho - 1}{\rho}, \qquad \forall x > 0.$$



Odhady γ z rozdělení $F(x) = 1 - 1/log(x)$



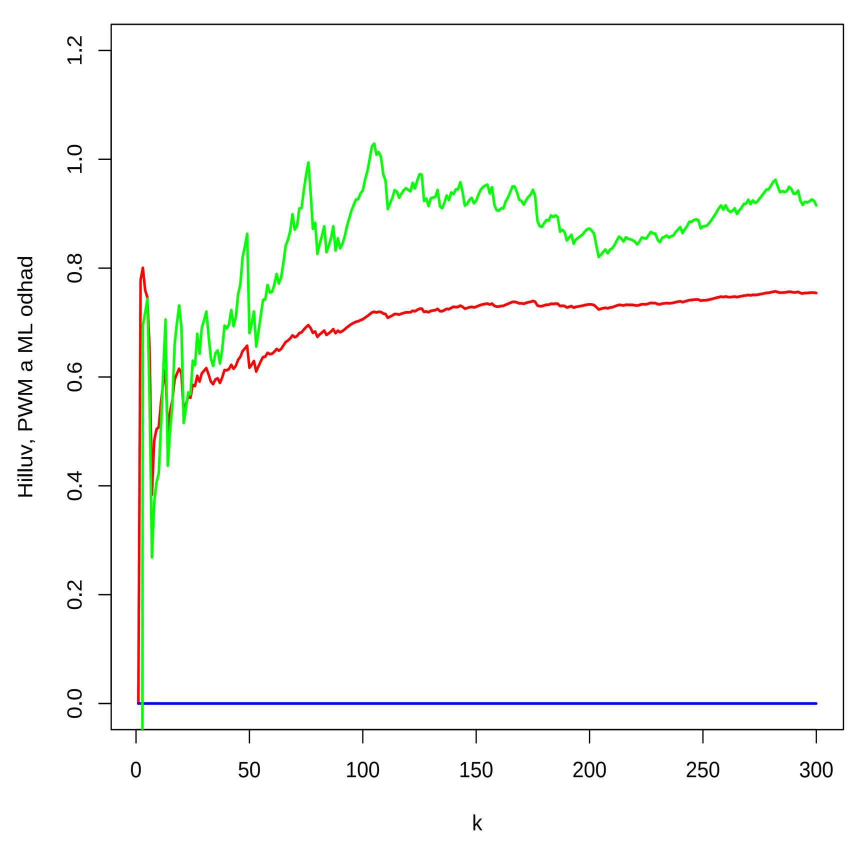
Zatímco rozptyl odhadů je funkcí γ, je vychýlení funkcí parametru druhého řádu ρ. Obrázek ilustruje situaci na příkladě Hillova odhadu (vidíme jeho vychýlení a rozptyl versus k).

#1 Vlastnosti odhadů γ

Odhady γ, které lze chápat jako funkcionály empirické kvantilové funkce chvostů Q_n, mají některé vlastnosti společné

- **vztah mezi vychýlením & rozptylem odhadu**, tj. rozptyl klesá s k_n zatímco vychýlení roste
- velká část odhadů je **invariantní pouze na měřítko** (např Hillův odhad na obrázku vlevo, data z t-rozdělení s 2 s.v., data $\pm 1, 2, 3$ – správně vždy $\gamma = 1/2$)
- vychýlení lze popsat pouze na základě ρ a **aproximace** druhého řádu, což je určující i pro normalitu

! Tyto vlastnosti jsou pro správnost odhadu γ zcela rozhodující !



#2 Porušení podmínek

Jak vypadají rozdělení, která nesplňují Fisher–Tippetovu větu?

- **Příliš "těžká" rozdělení**

  $$F(x) = 1 - \frac{1}{\log(x)}$$

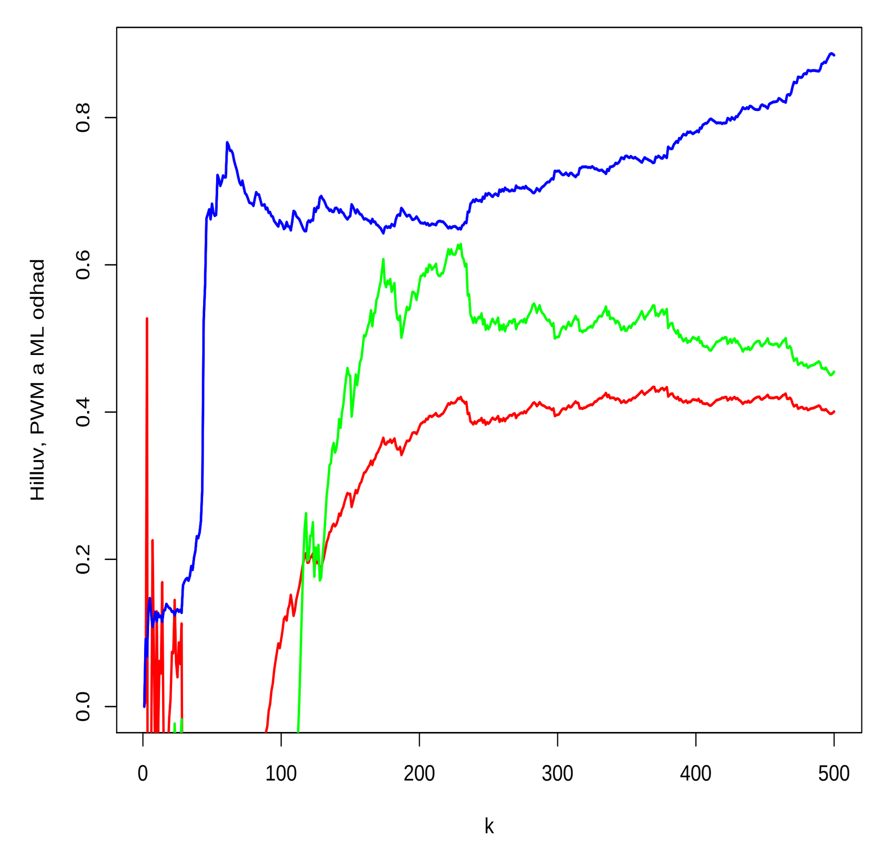
- **Hladká ale v chvostech oscilující**

  $$F(x) = 1 - \exp(-x - \sin(x)), \qquad x > 0$$

Podobně existují i rozdělení, která "pouze" nesplňují podmínku druhého řádu, např. $F(x) = 1 + x^{-1}\exp(\sin\log x)$.

? Jak se odhady γ vyrovnají s daty z takových rozdělení ?

Odhady γ z rozdělení $F(x) = 1 - e^{-x-\sin(x)}$

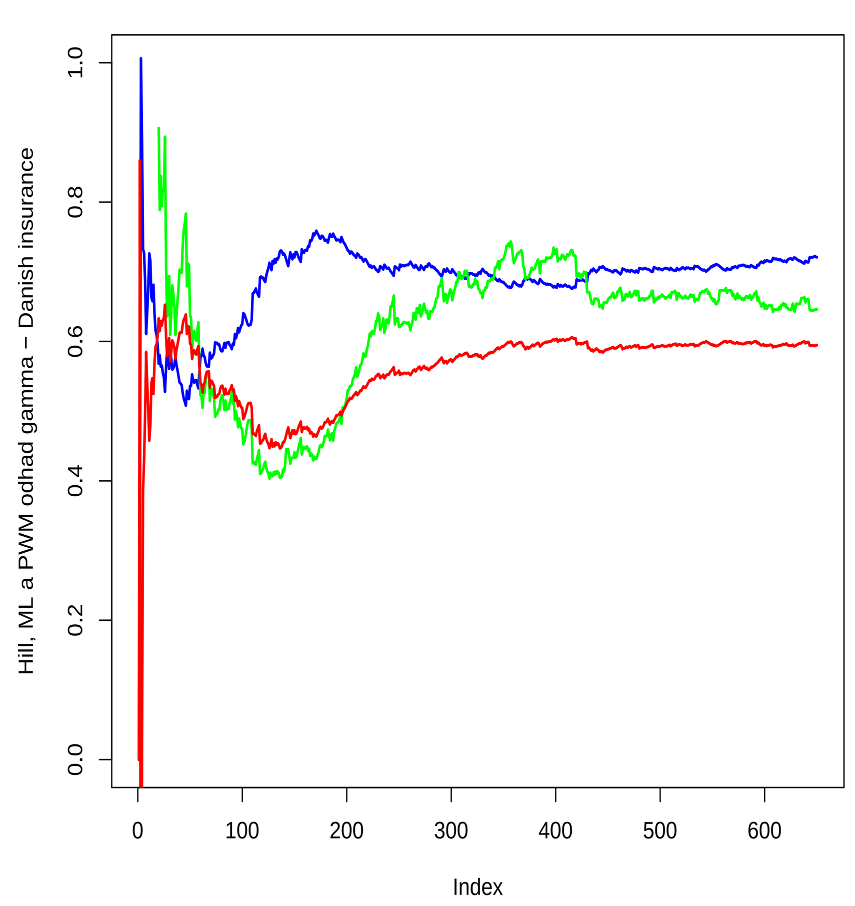


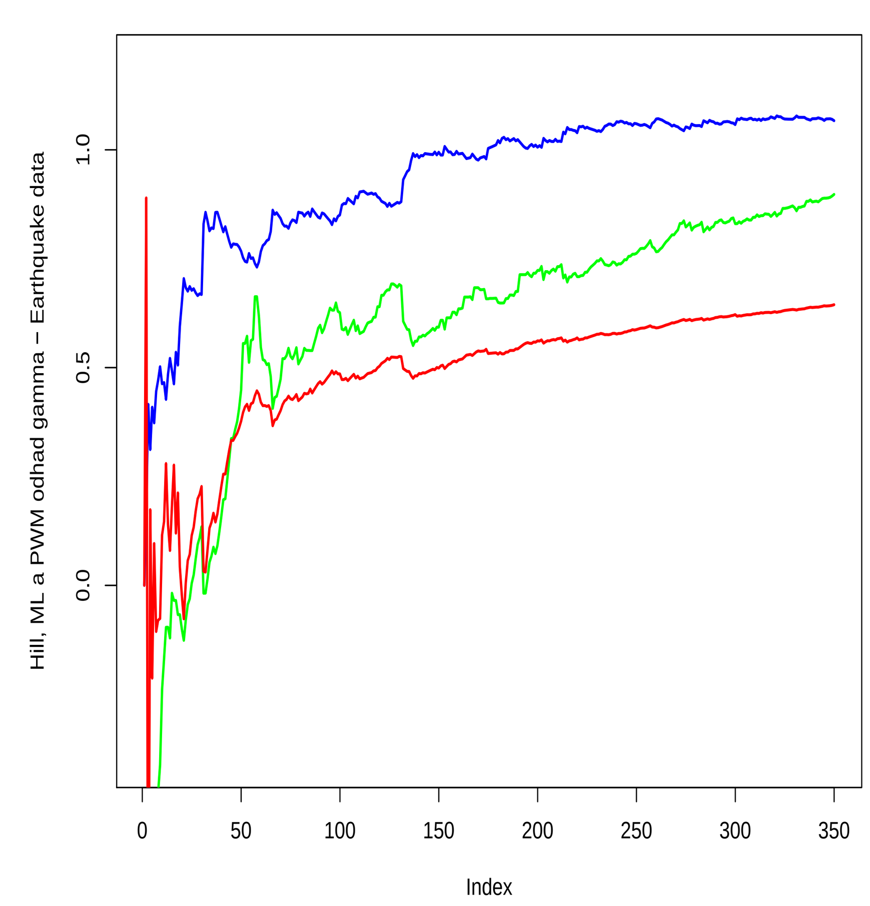
#3 Pokus o shrnutí

Při modelování s EVT bychom měli dávat pozor na řadu věcí.

- Je kvantilová funkce chvostů opravdu pomalu se měnící? (lze i testovat, viz např. Dietrich, de Haan, Hüsler (2002))
- Co aproximace druhého druhu? Jak se odhad mění s k_n?
- Skutečně používáme správný odhad pro naši situaci? (např. není odhad příliš vychýlený díky hodnotě ρ?)

? Nakolik umíme využít tyto znalosti při analýze reálných dat ?





dienstbi@karlin.mff.cuni.cz
KPMS MFF UK Praha

Reference
http://artax.karlin.mff.cuni.cz/~diky/rob08refence.pdf

Autor děkuje Luboši Prchalovi & Martinu Betincovi za LaTeX-ovský design, který použil při přípravě tohoto posteru. Poster vznikl za podpory grantu GAČR 201/05/H007.