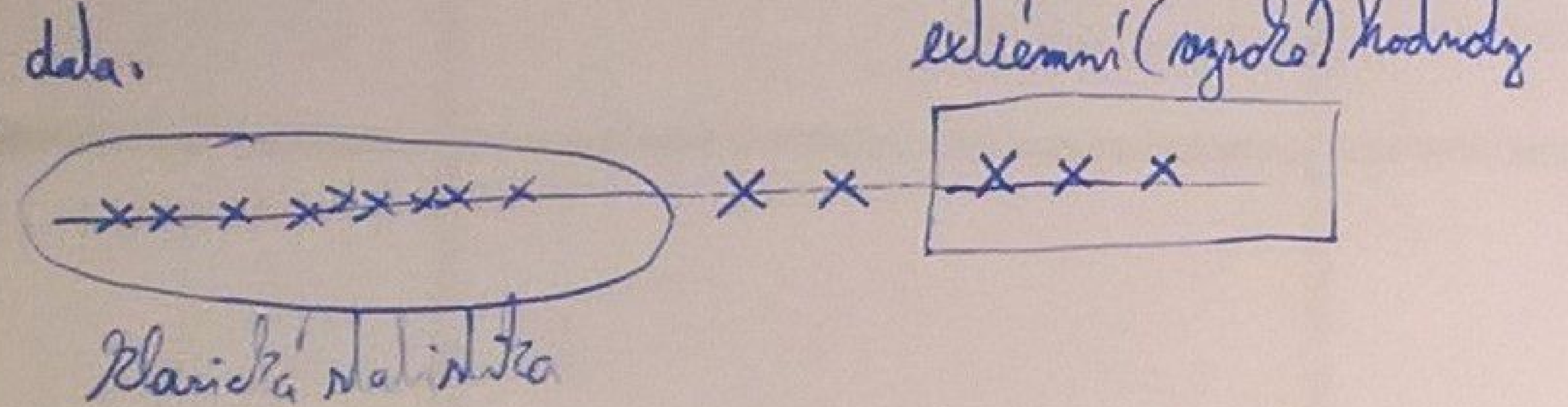
Teorie extrémních hodnot (EVT)



- průměrná škoda, výše plnění
- průměrný klient

Cíl: odhadnout $P(X > x)$ pro x velké (pravd. extrémních událostí)

odhadnout $F^{-1}(\alpha)$ pro α velké (vysoké kvantily)

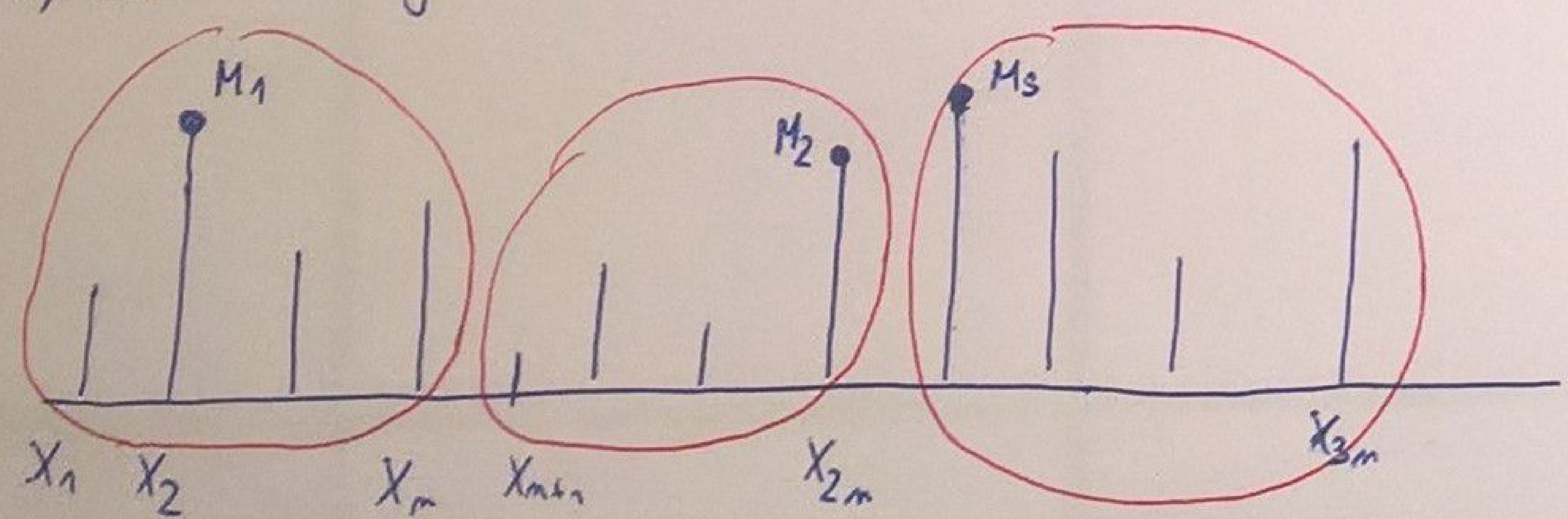
- určit výši plnění, kterou nárokuje jen malé procento klientů
- určit hranici, která bude překonána v průměru jednou za x let

- neparametrické metody – nejsou vhodné, protože nás zajímají hodnoty, které se v datech objevují zřídka nebo vůbec
- parametrické metody – příliš citlivé na vhodnost modelu

=> v praxi se uplatňují jiné přístupy

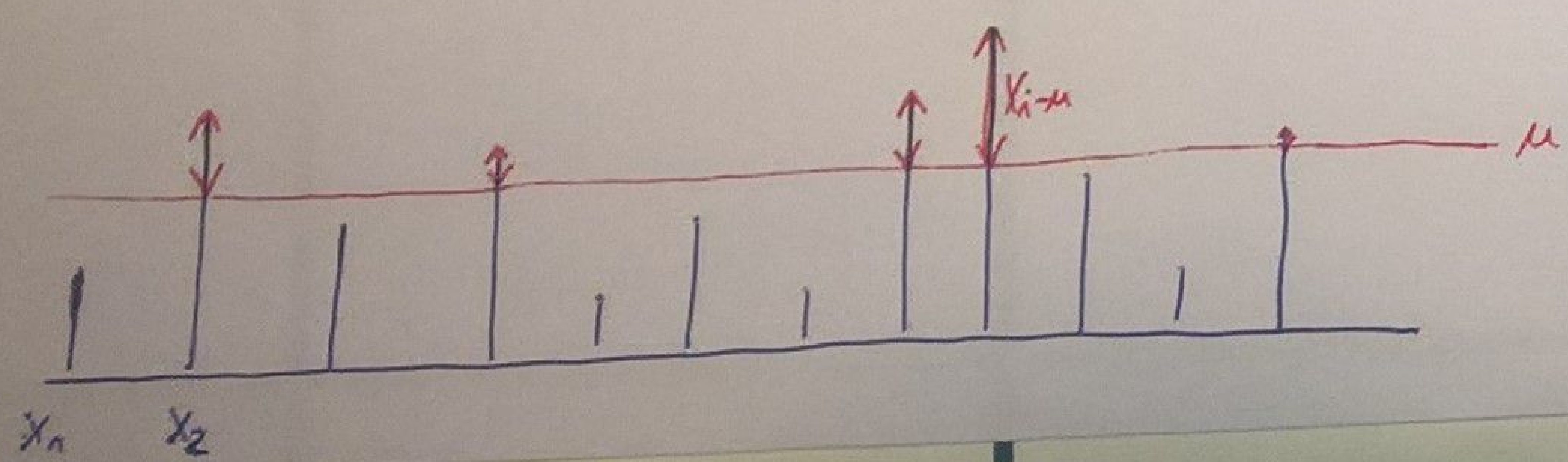
1.) Metoda blokových maxim

- pozorování rozdělíme na bloky
- v každém bloku najdeme maximální hodnotu
- modelujeme jen bloková maxima



2.) Metoda POT (peaks-over-threshold) – překročení hranice

- zvolíme hranici u
- najdeme ta pozorování, která překročí hranici u
- modelujeme velikosti překročení této hranice $X_i - u$ pro pozorování, která ji překročila



Chování maxima náh. výběru

$X_1, \dots, X_n$ je náh. výběr s distribuční funkcí F

označme $M_n = \max\{X_1, \dots, X_n\}$ maximum náh. výběru o rozsahu n.

Odvoďme jeho distribuční funkci: $G_n(x) = P(M_n \le x) = P(\max\{X_1, \dots, X_n\} \le x) = P(X_1 \le x, \dots, X_n \le x) = P(X_1 \le x) \cdot \ldots \cdot P(X_n \le x) = [F(x)]^n, x \in \mathbb{R}$

Hledejme limitní rozdělení maxima M_n:

$$\lim_{n \to \infty} G_n(x) = \begin{cases} 0 & , \dots x < x_F \\ 1 & , \dots x \ge x_F \end{cases}$$

, kde $x_F = \sup\{x \in \mathbb{R} : F(x) < 1\}$ je pravý koncový bod nosiče F (může být i $+\infty$)

$\Rightarrow$ limitní rozdělení M_n je degenerované v bodě x_F (případně $M_n \xrightarrow{P} \infty$ je-li $x_F = \infty$), $P(M = x_F) = 1$, $P(M = x) = 0$, jinak.

Příklad (CLV)

$X_1, \dots, X_n$ je náh. výběr, označme $\mu = EX_i$ a $\sigma^2 = DX_i$

$\Rightarrow$ jeho limitní rozdělení je degenerované v bodě μ ($\bar{X} \xrightarrow{P} \mu$).

zkoumáme chování výběrového průměru $\bar{X}$ – jeho limitní rozdělení je degenerované v bodě μ ($\bar X \xrightarrow{P} \mu$).

distribuční funkce $N(0,1)$

CLV: $\bar{X} \approx N\left(\mu, \frac{\sigma^2}{n}\right) \rightarrow \dfrac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \approx N(0,1)$, tj. $P\left(\dfrac{\bar{X} - \underbrace{\mu}_{b_n}}{\underbrace{\frac{\sigma}{\sqrt{n}}}_{a_n}} \le x\right) \xrightarrow[n \to \infty]{} \Phi(x)$, $\forall x \in \mathbb{R}$

obecně: $P\left(\dfrac{\bar{X} - b_n}{a_n} \le x\right) \xrightarrow{n \to \infty} \Phi(x)$, kde $\{b_n\}_{n=1}^{\infty}$ je posl. reálných čísel a $\{a_n\}_{n=1}^{\infty}$ posl. kladných čísel.

idea pro maxima:

Budeme hledat podobné normující konstanty $a_n > 0$ a $b_n \in \mathbb{R}$ tak, aby $\underbrace{\frac{M_n - b_n}{a_n}}_{\text{normované maximum}}$ při $n \to \infty$ konvergovalo k nějakému nedegenerovanému rozdělení.

distribuční funkce normovaných maxim: $P\left(\frac{M_n - b_n}{a_n} \le x\right) = P(M_n \le a_n x + b_n) = G_n(a_n x + b_n) = [F(a_n x + b_n)]^n \xrightarrow[n \to \infty]{\text{chceme}} H(x), \forall x \in \mathbb{R}$

pro nějakou nedegenerovanou distr. fci H.

Příklad (maximum exponenciálního rozdělení)

$X_1, \dots, X_n$ je náh. výběr z $\mathcal{E}x(1)$

$f(x) = e^{-x}, x > 0$

$F(x) = 1 - e^{-x}, x > 0$

$a_n = 1$

$b_n = \log n$

distr. fce normovaných maxim: $P\left(\frac{M_n - b_n}{a_n} \le x\right) = [F(a_n x + b_n)]^n = \left[1 - e^{-(x + \log n)}\right]^n = \left[1 - e^{-x} \cdot e^{-\log n}\right]^n = \left(1 - \frac{e^{-x}}{n}\right)^n \xrightarrow{n \to \infty} e^{-e^{-x}}, x \in \mathbb{R}$

distribuční funkce Gumbelova rozdělení

Příklad (maximum Paretova rozdělení typu II - Lomax rozdělení)

$X_1, \dots, X_n$ je náh. výběr z Pareto(α, θ), $\alpha, \theta > 0$

$f(x) = \frac{\alpha}{\theta}\left(1 + \frac{x}{\theta}\right)^{-(\alpha+1)}, x \ge 0$

$$F(x) = 1-\left(1+\frac{x}{\theta}\right)^{-\alpha}, \quad x \geq 0$$

$$a_n = \frac{\theta \cdot n^{\frac{1}{\alpha}}}{\alpha} \qquad b_n = \theta \cdot n^{\frac{1}{\alpha}} - \theta$$

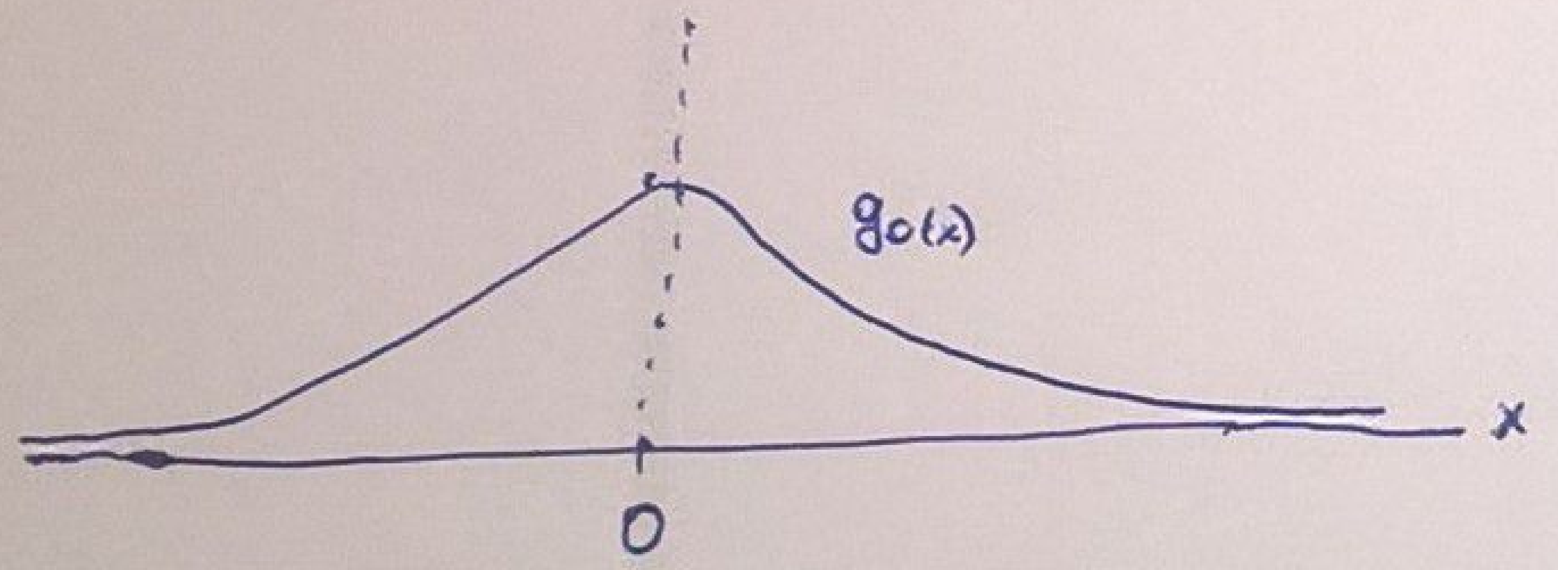
$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) = \left[F(a_n x + b_n)\right]^n = \left[1-\left(1+\frac{\frac{\cancel{\theta} n^{\frac{1}{\alpha}}}{\alpha}\cdot x + \cancel{\theta}\cdot n^{\frac{1}{\alpha}} - \cancel{\theta}}{\cancel{\theta}}\right)^{-\alpha}\right]^n = \left[1-\left(\cancel{1} + n^{\frac{1}{\alpha}}\left(\frac{x}{\alpha}+1\right) - \cancel{1}\right)^{-\alpha}\right]^n = \left[1 - \frac{1}{n}\left(1+\frac{x}{\alpha}\right)^{-\alpha}\right]^n \xrightarrow{n\to\infty}$$

$$e^{-\left(1+\frac{x}{\alpha}\right)^{-\alpha}}$$ ... distribuční funkce Fréchetova rozdělení

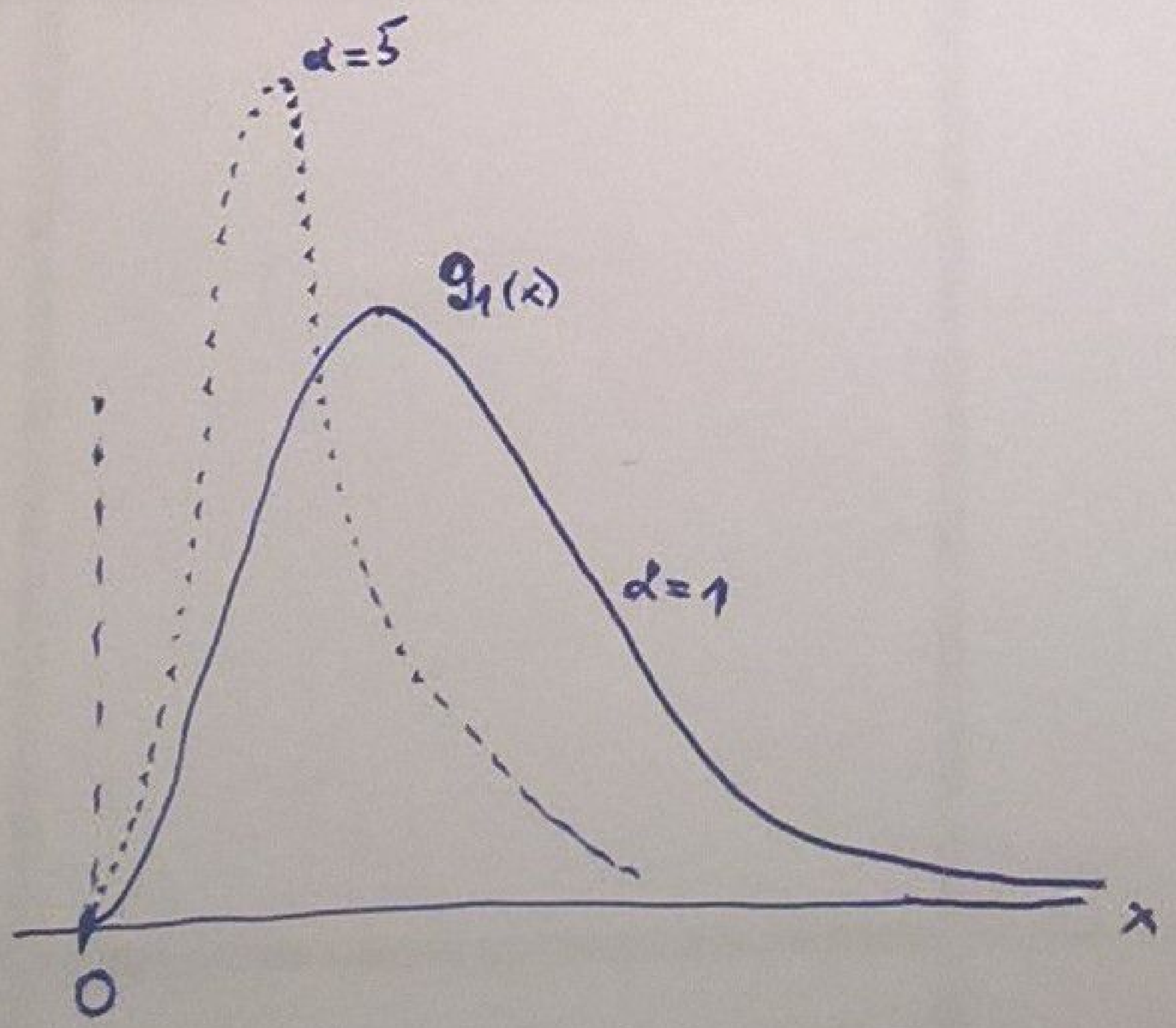
Rozdělení extrémních hodnot

1) Gumbelovo rozdělení

$$G_0(x) = e^{-e^{-x}}, \quad x \in \mathbb{R}$$



$$g_0(x) = e^{-(x+e^{-x})}, \quad x \in \mathbb{R}$$



2) Fréchetovo rozdělení s parametrem tvaru $\alpha > 0$

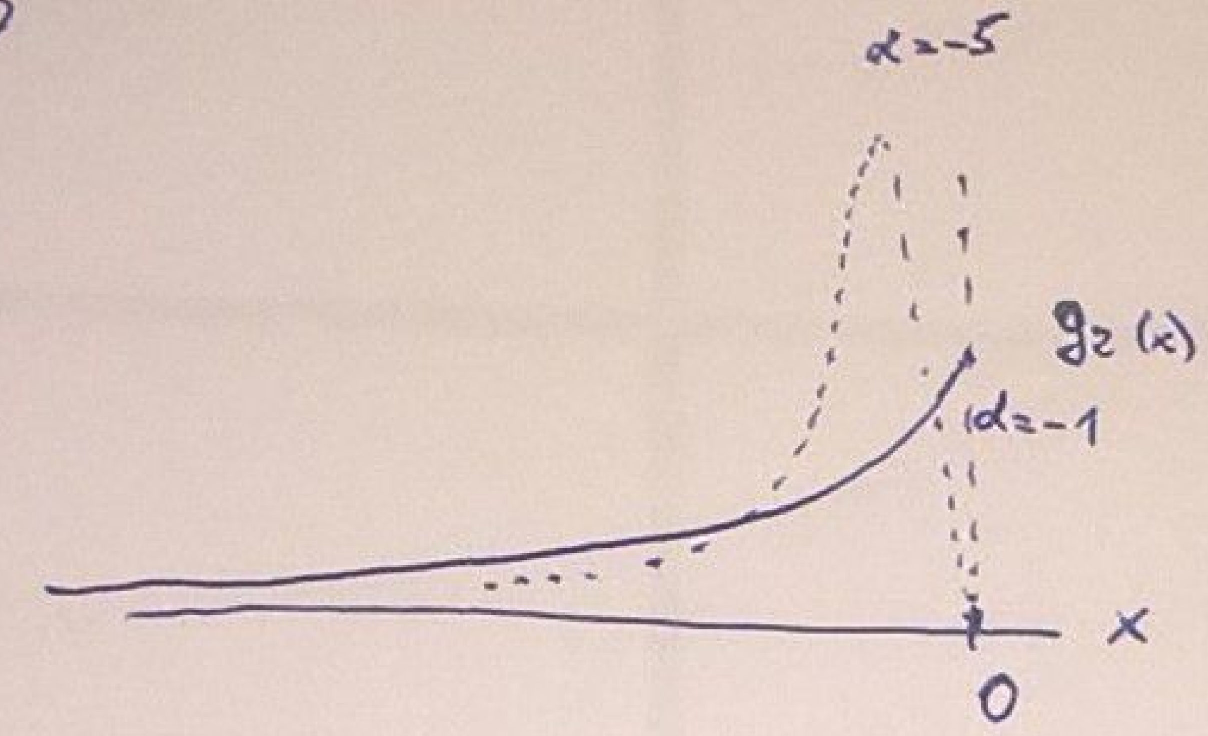
$$G_1(x) = e^{-x^{-\alpha}}, \quad x \geq 0$$

$$g_1(x) = \frac{\alpha}{x^{\alpha+1}} \cdot e^{-x^{-\alpha}}, \quad x > 0$$

3) Weibullovo rozdělení (extremální) s parametrem $\alpha < 0$

$$G_2(x) = e^{-(-x)^{-\alpha}}, \quad x \leq 0$$

$$g_2(x) = -\alpha \cdot (-x)^{-\alpha-1} \cdot e^{-(-x)^{-\alpha}}, \quad x \leq 0$$



Předchozí 3 rozdělení jsou standardizované (mají parametr polohy roven 0 a parametr měřítka roven 1).

=> přidáme parametr polohy μ a parametr měřítka σ a dostaneme distribuční funkce:

$$G_{i,\mu,\sigma}(x) = G_i\left(\frac{x-\mu}{\sigma}\right) \quad \text{pro } i = 0,1,2$$

Poznámka

speciálně dostaneme: $G_{0,\mu,\sigma}(x) = e^{-e^{-\left(\frac{x-\mu}{\sigma}\right)}}$, $G_{1,\mu,\sigma}(x) = e^{-\left(\frac{x-\mu}{\sigma}\right)^{-\alpha}}$, $G_{2,\mu,\sigma}(x) = e^{-\left(-\frac{x-\mu}{\sigma}\right)^{-\alpha}}$.

Otázka: Nemohli bychom předchozí 3 distribuční funkce zapsat jedním vzorcem?

Zobecněné rozdělení extrémních hodnot (GEV rozdělení)

ve standardizovaném tvaru: $G_\gamma(x) = \exp\left\{-(1+\gamma x)^{-\frac{1}{\gamma}}\right\}$, $1+\gamma x > 0$ ($\gamma \in \mathbb{R}$ parametr)

Poznámka

pro $\gamma = 0$ chápeme $G_\gamma(x)$ jako $\lim_{\gamma \to 0} G_\gamma(x) = e^{-e^{-x}}$ ... Gumbelovo rozdělení

$\gamma > 0$ ... Fréchetovo rozdělení

$\gamma < 0$ ... Weibullovo rozdělení

úplnou třídu dostaneme opět přidáním parametrů μ a σ : $G_{\gamma,\mu,\sigma}(x) = \exp\left\{-\left(1+\frac{\gamma(x-\mu)}{\sigma}\right)^{-\frac{1}{\gamma}}\right\}$, $1+\frac{\gamma(x-\mu)}{\sigma} > 0$.

<u>Věta</u> (Fisher - Tippett)

Nechť $X_1, \ldots, X_n$ je náhodný výběr s distribuční funkcí F. Označme $M_n = \max\{X_1, \ldots, X_n\}$. Nechť existují posl. konstant $a_n > 0$ a $b_n \in \mathbb{R}$ takové, že

$\lim\limits_{n\to\infty} P\left(\frac{M_n - b_n}{a_n} \leq x\right) = H(x)$, $\forall x \in \mathbb{R}$ pro nějakou nedegenerovanou distr. fci H. Pak H je distribuční funkce GEV rozdělení.

<u>Poznámky</u>

- podle předchozí věty je jediné možné limitní rozdělení GEV rozdělení $\Rightarrow$ maxima budeme modelovat pomocí GEV rozdělení
- jednoznačnost a_n, b_n? Lze nalézt více posloupností $\{a_n\}$ a $\{b_n\}$ takových, že $\frac{M_n - b_n}{a_n}$ konvergují. Výsledná limitní rozdělení se liší jen posunutím a měřítkem (parametry μ a σ).

Modelování: Nechť ex. posl. a_n, b_n :

$P\left(\frac{M_n - b_n}{a_n} \leq x\right) \to G_\gamma(x) \quad \rightsquigarrow P(M_n \leq \underbrace{a_n x + b_n}_{y}) \xrightarrow{n\to\infty} G_\gamma(x)$

pro n velké $P(M_n \leq y) \approx G_\gamma\left(\frac{y - b_n}{a_n}\right) = G_{\gamma, b_n, a_n}(y)$. $\Rightarrow$ Maximum modelujeme pomocí GEV rozdělení se 3 parametry $\gamma, \mu, \sigma \Rightarrow$ ty odhadneme z dat

<u>Poznámky</u>

Pokud ... distr. fcí G_γ, píšeme $F \in MDA(G_\gamma)$ (maximum domain of attraction).

γ-kladná podmínka F.-T. věty pro F, říkáme, že F patří do oblasti přitažlivosti distr. fcí G_γ, ...

Podobně říkáme, že F patří do oblasti přitažlivosti Gumbelova, Fréchetova nebo Weibullova rozd. (dají se odvodit pravidla, podle nichž zjistíme, které rozdělení to je).