Skórová funkce rozdělení a možné aplikace

Zdeněk Fabián
Ústav informatiky AVČR Praha

January 19, 2014

Starověk

- $x_1, ..., x_n$ data

průměry

Starověk

- $x_1, ..., x_n$ data

průměry

- aritm., geom., harm.

Novověk

Model F a skórová funkce Ψ_F

inferenční funkce:

náhodný výběr $(x_1, ...x_n)$ z $F \in \mathcal{F}_\theta$:

$$\sum_{k=1}^{n} (x_i - \mu) = 0 \qquad \sum_{k=1}^{n} \Psi_F(x_k; \theta) = 0$$

vlivová funkce:

$\Psi_F(x_k; \theta)$ pro pevné k: relativní vliv pozorovaného x_k na odhadovanou charakteristiku výběru

2 typy skórových funkcí

Klasická statistika:

Model F_θ s hustotou $f(x;\theta)$

$$\Psi_{clasic}(x;\theta) = \frac{\partial}{\partial\theta} \log f(x;\theta)$$

Fisherova skórová funkce

Robustní statistika:

'Struktura' s parametry polohy a měřítka

$$\Psi_{robust}(x;\mu,\sigma) = \psi_{bounded}\left(\frac{x-\mu}{\sigma}\right)$$

Třetí typ skórové funkce na $\mathbb{R}$

- Základní identita

$$\frac{\partial}{\partial\mu}\log g(y-\mu) = -\frac{1}{g(y-\mu)}\frac{d}{dy}g(y-\mu)$$

Třetí typ skórové funkce na $\mathbb{R}$

- Základní identita

$$\frac{\partial}{\partial \mu} \log g(y-\mu) = -\frac{1}{g(y-\mu)} \frac{d}{dy} g(y-\mu)$$

- $$S_G(y) = -\frac{g'(y)}{g(y)}$$

Třetí typ skórové funkce na $\mathbb{R}$

- Základní identita

$$\frac{\partial}{\partial \mu} \log g(y-\mu) = -\frac{1}{g(y-\mu)} \frac{d}{dy} g(y-\mu)$$

- $$S_G(y) = -\frac{g'(y)}{g(y)}$$

- $S_G(y)$ je charakteristikou **rozdělení**

Pearson (1895) Typy rozdělení $S_G(y) = \frac{ay+b}{cy^2+dy+v}$
Hampel et al. (1986) S_G vlivová fce rozdělení
Jurečková (2012) skórová funkce rozdělení

Typy funkcí s definičním oborem reálná přímka

typ	funkce
NE	$\sinh(u)$
NP	u
ON	$e^u - 1$
NO	$1 - e^{-u}$
OO	$\tanh(u/2)$
OR	$\frac{2u}{1+u^2}$

jako skórové funkce rozdělení

typ	$S_G(u)$	$g(u)$	rozdělení
NE	$\frac{e^u-e^{-u}}{2}$	$\frac{1}{2K_1(1)}e^{-\frac{1}{2}(e^u+e^{-u})}$	hyperbolic
NP	u	$\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}u^2}$	normal
ON	e^u-1	$e^u e^{-e^u}$	Gumbel ?
NO	$1-e^{-u}$	$e^{-u}e^{-e^{-u}}$	extreme value
OO	$\frac{e^u-1}{e^u+1}$	$\frac{e^u}{(1+e^u)^2}$	logistic
OR	$\frac{2u}{1+u^2}$	$\frac{1}{\pi(1+u^2)}$	Cauchy

$$S_G(u) = -\frac{g'(u)}{g(u)}$$

Parametrické skórové funkce rozdělení

- zobecnění s podmínkou $S_G(u;\alpha,\nu) = 0$

typ	$S_G(u;\alpha,\nu)$	$\sim g(u;\alpha,\nu)$	rodina
NE	$\frac{\alpha}{2}(e^u - \nu e^{-u}) + \rho$	$e^{-\frac{\alpha}{2}(e^u+\nu e^{-u})}$	hyperbolic
NP	u	$e^{-\frac{1}{2}u^2}$	normal
ON	$\alpha(e^u - 1)$	$e^{\alpha u} e^{-\alpha e^u}$	gamma
NO	$\alpha(1 - e^{-u})$	$e^{-\alpha u} e^{-\alpha e^{-u}}$	extreme val.
OO	$\alpha \frac{e^u - 1}{e^u + 1/\nu}$	$\frac{e^{\nu\alpha u}}{[e^u+1/\nu]^{(1+\nu)\alpha}}$	beta
OR	$\frac{2\alpha u}{1+u^2}$	$\frac{1}{(1+u^2)^\alpha}$	Cauchy

Parametrické skórové funkce rozdělení

- zobecnění s podmínkou $S_G(u;\alpha,\nu)=0$

typ	$S_G(u;\alpha,\nu)$	$\sim g(u;\alpha,\nu)$	rodina
NE	$\frac{\alpha}{2}(e^u-\nu e^{-u})+\rho$	$e^{-\frac{\alpha}{2}(e^u+\nu e^{-u})}$	hyperbolic
NP	u	$e^{-\frac{1}{2}u^2}$	normal
ON	$\alpha(e^u-1)$	$e^{\alpha u}e^{-\alpha e^u}$	gamma
NO	$\alpha(1-e^{-u})$	$e^{-\alpha u}e^{-\alpha e^{-u}}$	extreme val.
OO	$\alpha\frac{e^u-1}{e^u+1/\nu}$	$\frac{e^{\nu\alpha u}}{[e^u+1/\nu]^{(1+\nu)\alpha}}$	beta
OR	$\frac{2\alpha u}{1+u^2}$	$\frac{1}{(1+u^2)^\alpha}$	Cauchy

- $$u=\frac{y-\mu}{\sigma}$$

Obecná parametrická skórová funkce rozdělení na $\mathbb{R}$

- Pro $g(y;\theta)$ s nosičem $\mathbb{R}$ (Fabián, 2001)

$$S_G(y;\theta) = -\frac{1}{g(y;\theta)}\frac{d}{dy}g(y;\theta)$$

Obecná parametrická skórová funkce rozdělení na $\mathbb{R}$

- Pro $g(y;\theta)$ s nosičem $\mathbb{R}$ (Fabián, 2001)

$$S_G(y;\theta) = -\frac{1}{g(y;\theta)}\frac{d}{dy}g(y;\theta)$$

- např.

$$\begin{aligned} g(y;p,q) &= \frac{1}{B(p,q)}\frac{e^{py}}{(e^y+1)^{p+q}} \\ S_G(y;p,q) &= \frac{qe^y - p}{e^y + 1} \end{aligned}$$

$$S_G(y;p,q) = 0 \quad \Rightarrow \quad y^* = \log\frac{p}{q}$$

Vlastnosti $S_G(y; \theta)$

- momenty $ES_G^k(\theta)$ existují

Vlastnosti $S_G(y;\theta)$

- momenty $ES_G^k(\theta)$ existují
- a jsou jednoduchými funkcemi θ
  $\Rightarrow$ obecná momentová metoda

Vlastnosti $S_G(y;\theta)$

- momenty $ES_G^k(\theta)$ existují
- a jsou jednoduchými funkcemi θ
  $\Rightarrow$ obecná momentová metoda

- $g(y) = \frac{1}{B(p,q)} \frac{e^{py}}{(e^y+1)^{p+q}}, \quad S_G(y) = \frac{qe^y-p}{e^y+1}$

$$\begin{aligned} \sum_{i=1}^{n} \frac{qe^{y_i} - p}{e^{y_i}+1} &= 0 \\ \sum_{i=1}^{n} \left(\frac{qe^{y_i} - p}{e^{y_i}+1} \right)^2 &= \frac{pq}{p+q+1} \end{aligned}$$

Vlastnosti $S_G(y;\theta)$

- momenty $ES_G^k(\theta) = \int_{-\infty}{}^{\infty} S_G^k(\theta) f(y;\theta\, dy$ existují a jsou jednoduchými funkcemi θ
  $\Rightarrow$ obecná momentová metoda

Vlastnosti $S_G(y;\theta)$

- momenty $ES_G^k(\theta) = \int_{-\infty}^{\infty} S_G^k(\theta) f(y;\theta\, dy$ existují a jsou jednoduchými funkcemi θ
  $\Rightarrow$ obecná momentová metoda
- typická hodnota: mód $\quad y^* : S_G(y;\theta) = 0$

Vlastnosti $S_G(y;\theta)$

- momenty $ES_G^k(\theta) = \int_{-\infty}^{\infty} S_G^k(\theta) f(y;\theta\, dy$ existují a jsou jednoduchými funkcemi θ
  $\Rightarrow$ obecná momentová metoda
- typická hodnota: mód $\quad y^* : S_G(y;\theta) = 0$
- variabilita rozdělení (z analogie s Cramér-Rao větou): score variance
  $$\omega^2 = \frac{1}{ES_G^2}$$
  kde ES_G^2 je Fisherova informace pro mód

  normální rozdělení $S_\Phi(x;\mu,\sigma) = \frac{x-\mu}{\sigma^2}, \quad x^* = \mu, \omega = \sigma$

Proč se to takhle ve statistice nedělá

- Funkce

$$S_F(x) = -\frac{f'(x)}{f(x)}$$

je pro rozdělení F s nosičem $\mathcal{X} \neq \mathbb{R}$ divná:

$$f(x) = e^{-x}, \qquad \mathcal{X} = (0, \infty)$$

$$f(x) = 1, \qquad \mathcal{X} = (0, 1)$$

Proč se to takhle ve statistice nedělá

- Funkce
$$S_F(x) = -\frac{f'(x)}{f(x)}$$
je pro rozdělení F s nosičem $\mathcal{X} \neq \mathbb{R}$ divná:
$$f(x) = e^{-x}, \qquad \mathcal{X} = (0, \infty)$$
$$f(x) = 1, \qquad \mathcal{X} = (0, 1)$$

- Nápad: Skórová funkce rozdělení na $\mathcal{X} \neq \mathbb{R}$ existuje, ale je dána jiným vzorcem, který závisí na $\mathcal{X}$

Skórová funkce rozdělení na $\mathcal{X} \neq \mathbb{R}$

- F má nosič $\mathcal{X}$
  $\eta : \mathcal{X} \to \mathbb{R}$ je hladká rostoucí funkce

  $Y \sim \mathbb{R}$ $\quad G \quad$ 'prototyp', S_G

  X na $\mathcal{X}$: $\quad X = \eta^{-1}(Y)$, $F = G \circ \eta$

$$f(x) = g(\eta(x))\eta'(x)$$

Skórová funkce rozdělení na $\mathcal{X} \neq \mathbb{R}$

- F má nosič $\mathcal{X}$
  $\eta : \mathcal{X} \to \mathbb{R}$ je hladká rostoucí funkce

  $Y \sim \mathbb{R}$ $\quad G \quad$ 'prototyp', S_G

  X na $\mathcal{X}$: $\quad X = \eta^{-1}(Y)$, $F = G \circ \eta$

$$f(x) = g(\eta(x))\eta'(x)$$

- $$S_F(x) = S_G(\eta(x))$$

Skórová funkce rozdělení na $\mathcal{X} \neq \mathbb{R}$

Dvě ekvivalentní vyjádření:

$$S_F(x) = S_G(\eta(x))$$

$$S_F(x) = -\frac{1}{f(x)}\frac{d}{dx}\left(\frac{1}{\eta'(x)}f(x)\right)$$

Věta: η je dáno jednoznačně

- Pokud vzorec pro transformovanou hustotu

$$f(x) = g(\eta(x))\eta'(x)$$

obsahuje Jacobian nějaké transformace z $\mathcal{X}$ na $\mathbb{R}$, je to ona

Věta: η je dáno jednoznačně

- Pokud vzorec pro transformovanou hustotu

$$f(x) = g(\eta(x))\eta'(x)$$

  obsahuje Jacobian nějaké transformace z $\mathcal{X}$ na $\mathbb{R}$, je to ona

- Příklady:

$$f(x) = \frac{1}{\sqrt{2\pi}\cos^2 x} e^{-\frac{1}{2}\tan^2 x}$$

$$f(x) = \frac{1}{\sqrt{2\pi}} \frac{1}{1+x^2} e^{-\frac{1}{2}\arctan^2 x}$$

η je dáno jednoznačně

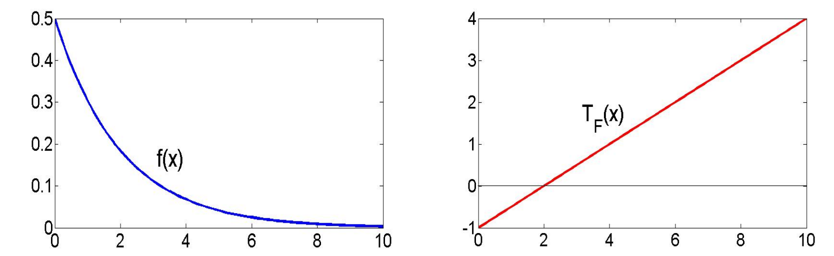
- V ostatní případech to nejfrekventovanější $\eta : (0, \infty) \to \mathbb{R} : \eta(x) = \log x$

η je dáno jednoznačně

- V ostatní případech to nejfrekventovanější $\eta : (0, \infty) \to \mathbb{R} : \eta(x) = \log x$
- Exponenciální $f(x) = \frac{1}{2}e^{-x/2} \quad f(x) = [xf(x)]\frac{1}{x}$

η je dáno jednoznačně

- V ostatní případech to nejfrekventovanější
  $\eta : (0, \infty) \to \mathbb{R} : \eta(x) = \log x$
- Exponenciální $f(x) = \frac{1}{2} e^{-x/2} \quad f(x) = [xf(x)]\frac{1}{x}$
- $$T_F(x) = -\frac{1}{f(x)} \frac{d}{dx}[xf(x)] = \frac{x}{2} - 1$$



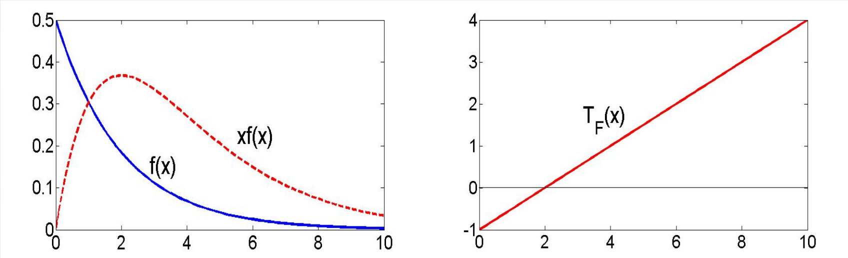
η je dáno jednoznačně

v ostatní případech to nejfrekventovanější

$\eta : (0, \infty) \to \mathbb{R} : \eta(x) = \log x$

Exponenciální $f(x) = \frac{1}{2}e^{-x/2} \quad f(x) = [xf(x)]\frac{1}{x}$

$$T_F(x) = -\frac{1}{f(x)}\frac{d}{dx}[xf(x)] = \frac{x}{2} - 1$$



Příklady

	$g(y)$	$S_G(y)$
normal	$\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}y^2}$	y
extreme value	$e^{-y}e^{-e^{-y}}$	$1-e^{-y}$
Gumbel	$e^{y}e^{-e^{y}}$	$e^{y}-1$
logistic	$\frac{e^y}{(1+e^y)^2}$	$\frac{e^y-1}{e^y+1}$

	$f(x)$	$S_F(x)$
lognormal	$\frac{1}{\sqrt{2\pi}x}e^{-\frac{1}{2}\log^2 x}$	$\log x$
Fréchet	$\frac{1}{x^2}e^{-1/x}$	$1-1/x$
exponential	$\frac{1}{x}xe^{-x}$	$x-1$
log-logistic	$\frac{1}{x}\frac{x}{(1+x)^2}$	$\frac{x-1}{x+1}$

Rozdělení na intervalu

- $\mathcal{X} = (0, 1)$ nejfrekventovanější $\eta(x) = \log \frac{x}{1-x}$

  Johnson U_B $\quad f(x) = \frac{1}{\sqrt{2\pi}x(1-x)} e^{-\frac{1}{2}\log^2 \frac{x}{1-x}}$ $\quad S_F(x) = \log \frac{x}{1-x}$

Rozdělení na intervalu

- $\mathcal{X} = (0,1)$ nejfrekventovanější $\eta(x) = \log \frac{x}{1-x}$

  Johnson U_B $\quad f(x) = \frac{1}{\sqrt{2\pi}x(1-x)} e^{-\frac{1}{2}\log^2 \frac{x}{1-x}} \quad S_F(x) = \log \frac{x}{1-x}$

- $\mathcal{X} = (-\pi/2, \pi/2)$ $\eta(x) = \tan x$

  $S_F(x) = -\frac{1}{f(x)} \frac{d}{dx}[\cos^2 x f(x)] = \sin 2x - \cos^2 x \frac{f'(x)}{f(x)}$

$S_F(x)$	$f(x)$	$g(y)$	$S_G(y)$
$\tan x$	$\frac{1}{\sqrt{c\cos^2 x}} e^{-\frac{1}{2}\tan^2 x}$	$\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}$	y
$\sin x$	$\frac{1}{c\cos^2 x} e^{-1/\cos x}$	$\frac{1}{2K_1(1)} e^{-\sqrt{1+y^2}}$	$\frac{y}{\sqrt{1+y^2}}$
$\sin 2x - k\cos^2 x$	$\frac{1}{c_k} e^{kx}$	$\frac{1}{c_k} \frac{1}{1+y^2} e^{k\tan^{-1} y}$	$\frac{2y-k}{1+y^2}$

t-score a S_F

- Pro parametrická rozdělení

$$T_F(x;\theta) = -\frac{1}{f(x;\theta)}\frac{d}{dx}\left(\frac{1}{\eta'(x)}f(x;\theta)\right)$$

$x^* : T_F(x;\theta) = 0$ a skórová funkce rozdělení je

$$S_F(x;\theta) = \eta'(x^*)T_F(x;\theta)$$

t-score a S_F

- Pro parametrická rozdělení

$$T_F(x;\theta) = -\frac{1}{f(x;\theta)}\frac{d}{dx}\left(\frac{1}{\eta'(x)}f(x;\theta)\right)$$

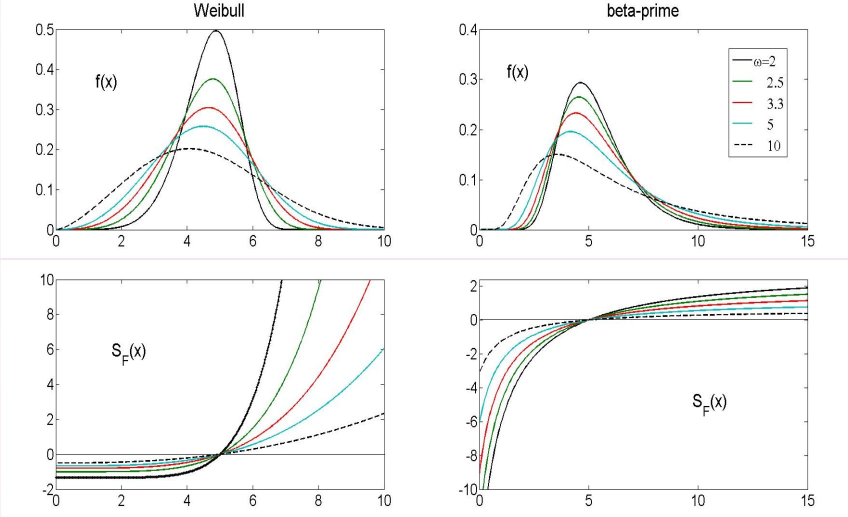
$x^* : T_F(x;\theta) = 0$ a skórová funkce rozdělení je

$$S_F(x;\theta) = \eta'(x^*)T_F(x;\theta)$$

- Pro exponenciální $f(x;\tau) = \frac{1}{\tau}e^{-x/\tau}$ je

$$S_F(x;\tau) = \frac{1}{\tau}[\frac{x}{\tau} - 1]$$

zde je $x^* = \tau = \eta^{-1}(\mu)$ a $S_F(x;\tau)$ je Fisherův skór pro τ



Weibull
beta-prime
f(x)
f(x)
$S_F(x)$
$S_F(x)$
ω=2
2.5
3.3
5
10

Vlastnosti S_F

K regulární f na $\mathcal{X}$ je S_F určena jednoznačně

Skalární funkce reflektující vlastnosti rozdělení

Typická hodnota $x^* : T_F(x;\theta) = 0$ je $x^* = \eta^{-1}(y^*)$, y^* obraz módu G

Je-li $\theta = (x^*, ...)$, je $S_F(x;\theta)$ Fisherův skór pro x^*

S_F rozdělení s těžkými chvosty omezená

Co přináší S_F nového pro popis modelu

- A. Systematika rozdělení: podle chování S_F na koncích intervalu $\mathcal{X}$

Co přináší S_F nového pro popis modelu

- A. Systematika rozdělení: podle chování S_F na koncích intervalu $\mathcal{X}$
- B. Nové numerické charakteristiky: x^*, ω^2

Co přináší S_F nového pro popis modelu

- A. Systematika rozdělení: podle chování S_F na koncích intervalu $\mathcal{X}$
- B. Nové numerické charakteristiky: x^*, ω^2
- C. Nové funkce charakterizující rozdělení:

$$S_F^2(x), S_F'(x) = \frac{dS_F(x)}{dx}$$

Co přináší S_F nového pro popis modelu

- A. Systematika rozdělení: podle chování S_F na koncích intervalu $\mathcal{X}$
- B. Nové numerické charakteristiky: x^*, ω^2
- C. Nové funkce charakterizující rozdělení:

$$S_F^2(x), S_F'(x) = \frac{dS_F(x)}{dx}$$

- D. Vzdálenosti pravděpodobnostních měr a metrika ve výběrovém prostoru

$$D(P,Q) = E_P(S_P - S_Q)^2 \qquad d(x_1,x_2) = |S_F(x_1) - S_F(x_2)|$$

B. Numerické charakteristiky

- Poloha na ose x: těžiště

$$x^* : \qquad S_F(x) = 0$$

B. Numerické charakteristiky

- Poloha na ose x: těžiště

$$x^* : \qquad S_F(x) = 0$$

- ES_F^2 (Fisherova) informace (pro x^*)

B. Numerické charakteristiky

- Poloha na ose x: těžiště

$$x^* : \qquad S_F(x) = 0$$

- ES_F^2 (Fisherova) informace (pro x^*)
- Míra variability: score variance $\omega^2 = \frac{1}{ES_F^2}$

B. Numerické charakteristiky

- Poloha na ose x: těžiště

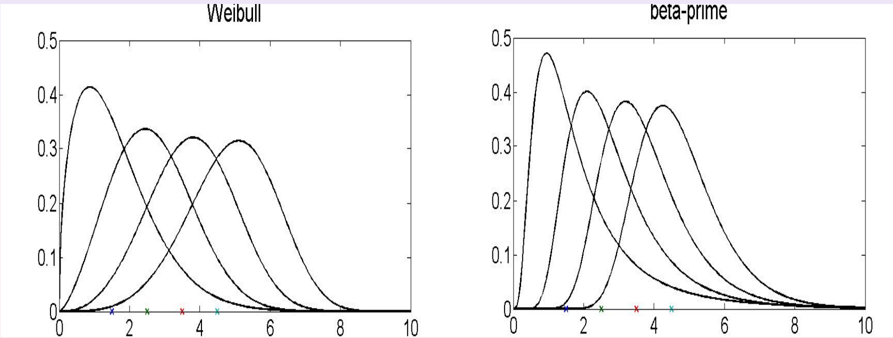
$$x^* : \qquad S_F(x) = 0$$

- ES_F^2 (Fisherova) informace (pro x^*)
- Míra variability: score variance $\omega^2 = \frac{1}{ES_F^2}$

F	$f(x)$	m_1	x^*	ω^2
Weibull	$\frac{c}{x}(\frac{x}{\tau})^c e^{-(\frac{x}{\tau})^c}$	$\tau\Gamma(\frac{1}{c}+1)$	τ	τ^2/c^2
Pareto	c/x^{c+1}	$\frac{c}{c-1}$	$\frac{c+1}{c}$	$\frac{c+2}{c^3}$
beta-prime	$\frac{1}{B(p,q)}\frac{x^{p-1}}{(x+1)^{p+q}}$	$\frac{p}{q-1}$	$\frac{p}{q}$	$\frac{p(p+q+1)}{q^3}$
Fréchet	$\frac{c}{x}(\frac{\tau}{x})^c e^{-(\frac{\tau}{x})^c}$	$\tau\Gamma(1-\frac{1}{c})$	τ	τ^2/c^2

Variabilita pomocí ω^2

Rozdělení s různými x^* a týmž $\omega^2 = 1$



Co přináší S_F nového pro statistiku

Málo dat:

- Neznáme model, čistá data: neparametrické metody

Co přináší S_F nového pro statistiku

Málo dat:

- Neznáme model, čistá data: neparametrické metody
- Neznáme model, kontaminovaná data: metody robustní statistiky

Co přináší S_F nového pro statistiku

Málo dat:

- Neznáme model, čistá data: neparametrické metody
- Neznáme model, kontaminovaná data: metody robustní statistiky
- Známe model, čistá data: parametrické metody klasické statistiky

Co přináší S_F nového pro statistiku

Málo dat:

- Neznáme model, čistá data: neparametrické metody
- Neznáme model, kontaminovaná data: metody robustní statistiky
- Známe model, čistá data: parametrické metody klasické statistiky
- Známe model, kontaminovaná data: parametrické metody založené na skalárním skóru

S_F jako inferenční funkce pro bodové odhady

- Score moment rovnice

$$\hat{\theta}_{SM}: \qquad \frac{1}{n}\sum_{i=1}^{n} S_F^k(x_i;\theta) = ES_F^k(\theta) \qquad k = 1, ..., m$$

Estimátor robustní pro všechny parametry když S_F je omezená

S_F jako inferenční funkce pro bodové odhady

- Score moment rovnice

$$\hat{\theta}_{SM}: \quad \frac{1}{n}\sum_{i=1}^{n} S_F^k(x_i;\theta) = ES_F^k(\theta) \quad k = 1, ..., m$$

  Estimátor robustní pro všechny parametry když S_F je omezená

- Výsledky pro rozdílné modely jdou porovnávat pomocí

$$\hat{x}_F^* = x_F^*(\hat{\theta}), \quad \hat{\omega}_F^* = \omega_F(\hat{\theta})$$

S_F jako inferenční funkce pro bodové odhady

- Score moment rovnice

$$\hat{\theta}_{SM}: \qquad \frac{1}{n}\sum_{i=1}^{n} S_F^k(x_i;\theta) = ES_F^k(\theta) \qquad k = 1, ..., m$$

  Estimátor robustní pro všechny parametry když S_F je omezená

- Výsledky pro rozdílné modely jdou porovnávat pomocí

$$\hat{x}_F^* = x_F^*(\hat{\theta}), \qquad \hat{\omega}_F^* = \omega_F(\hat{\theta})$$

- Konfidenční intervaly pro $\hat{x}^*$ pomocí $|S_F(x_0^*) - S_F(\hat{x}_F^*)|$

Těžiště: $\sum_i S_F(x_i;\theta) = 0$

exponential	$\sum(\frac{x_i}{\tau} - 1) = 0$	$\hat\tau = \bar x$
lognormal	$\sum \log(x_i/\tau) = 0$	$\hat\tau = \bar x_g$
Weibull	$\sum((x_i/\tau)^c - 1) = 0$	$\hat\tau = (\sum x_i^c)^{1/c}$
Pareto	$\sum(1 - x^*/x_i) = 0$	$\hat x^* = x_h$
beta-prime	$\sum \frac{qx_i - p}{x_i+1} = 0$	$\hat x^* = \frac{\sum \frac{x_i}{x_i+1}}{\sum \frac{1}{x_i+1}}$
Johnson	$\sum \log\left(\frac{x_i(1-\tau)}{(1-x_i)\tau}\right) = 0$	$\hat x^* = \Pi_{i=1}^n \frac{x_i}{1-x_i}$

$\hat{x}^*$ a $std(\hat{x}^*)$ Pareto

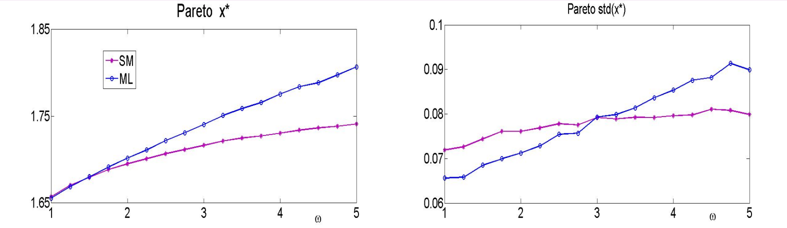
- $$f(x;c) = \frac{c}{x^{c+1}} \qquad S_F(x;c) \sim (1 - \frac{x^*}{x})$$

  kde $x^* = (c+1)/c$

$\hat{x}^*$ a $std(\hat{x}^*)$ Pareto

- $$f(x;c) = \frac{c}{x^{c+1}} \qquad S_F(x;c) \sim (1 - \frac{x^*}{x})$$

  kde $x^* = (c+1)/c$

- 

F s neomezenou S_F

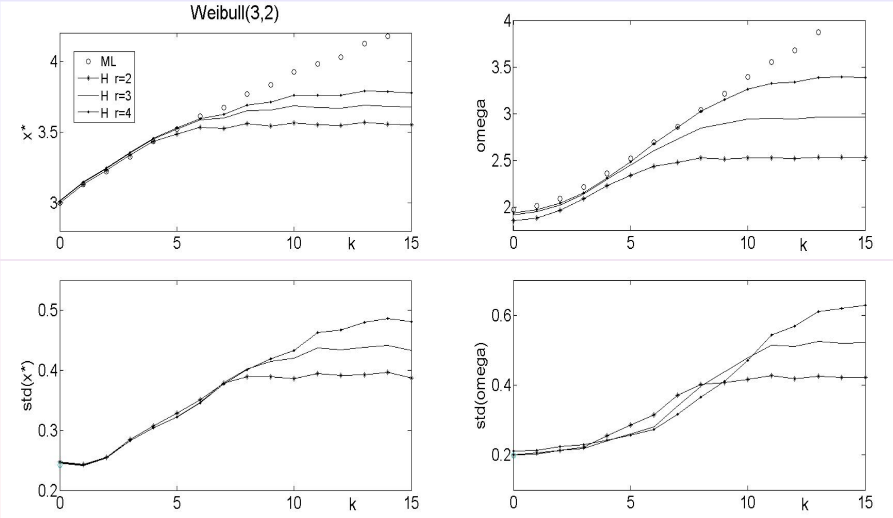
Pro $G(y-\mu)$ z $\mathbb{R}$ (Huber, Ronchetti, 2007)

$$\psi(x) = \begin{cases} u & \text{pro} \quad x \leq u \\ S_G(x) & \text{pro} \quad u < x < v \\ v & \text{pro} \quad x \geq v \end{cases}$$

Nejde pro vektorový parametr, ale jde pro $S_F(x)$ při libovolném $\mathcal{X}$

$$v = \hat{x}^* + r\, MADN(x)$$

$$F = 0.9F_0(\omega = 1) + 0.1F(\omega = k)$$


Weibull(3,2)
ML
H r=2
H r=3
H r=4
x*
k
omega
std(x*)
std(omega)

B. Složitější úlohy

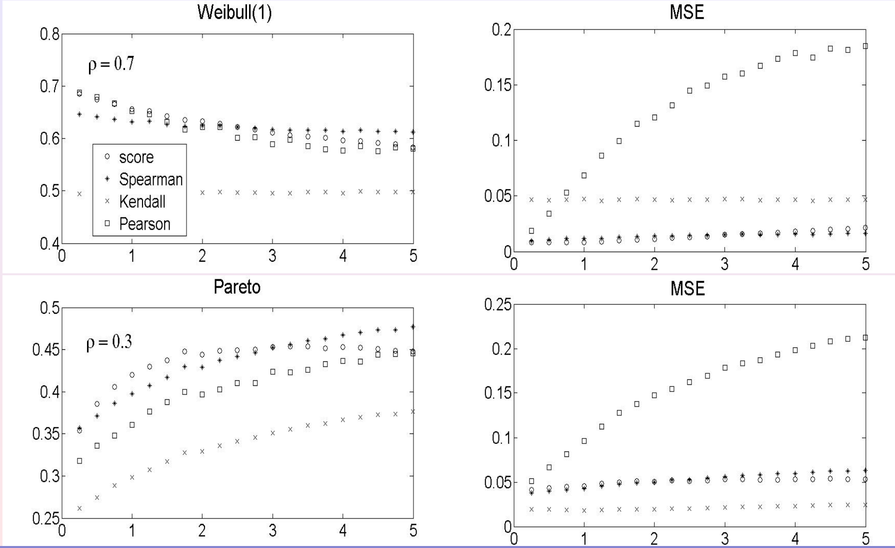
Kovarianční koeficienty: $r_S = \sum_{i=1}^{n} S_X(x_i; \hat{\theta}) S_Y(y_i; \hat{\theta})$

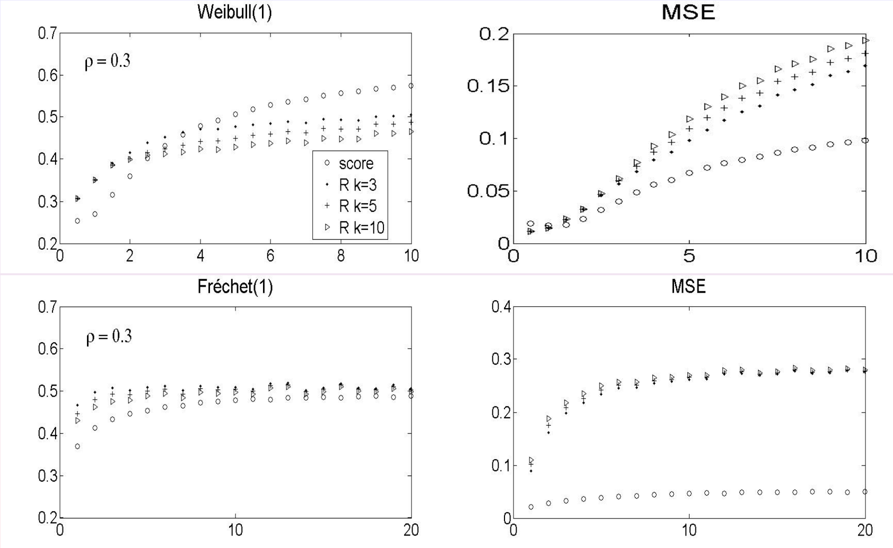
Lineární regrese: $\varepsilon_i = y_i - (\alpha_0 + \alpha_1 x_i)$

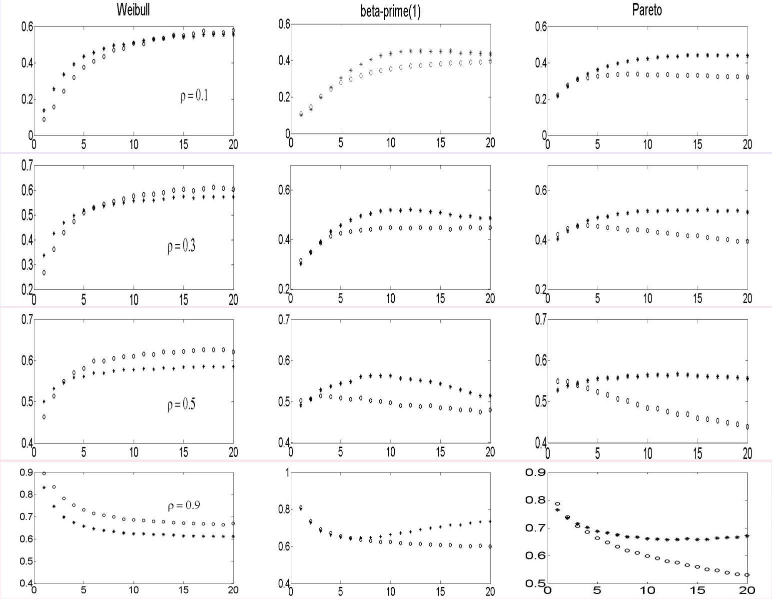
$$\frac{1}{n} \sum_{i=1}^{n} S_{\varepsilon}^2(\varepsilon_i; \hat{\theta}) = min.$$

Skalární skór časové řady $\{S_F(X_t)\}$

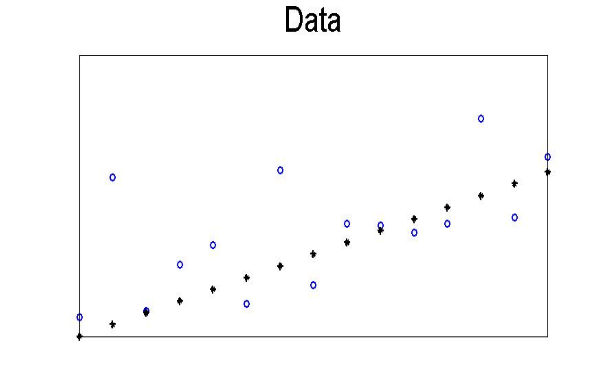
Korelační koeficient



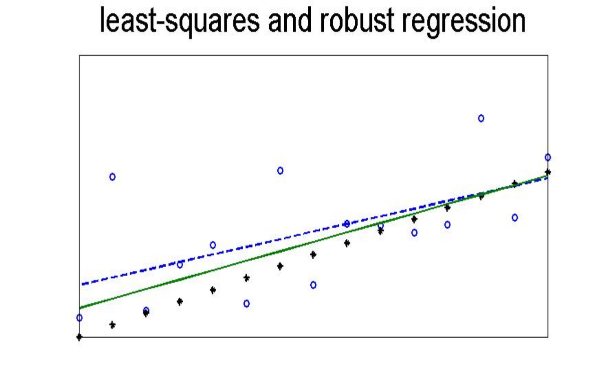

Weibull(1)
$\rho = 0.3$
score
R k=3
R k=5
R k=10
MSE
Fréchet(1)
$\rho = 0.3$
MSE


Weibull
beta-prime(1)
Pareto
ρ = 0.1
ρ = 0.3
ρ = 0.5
ρ = 0.9

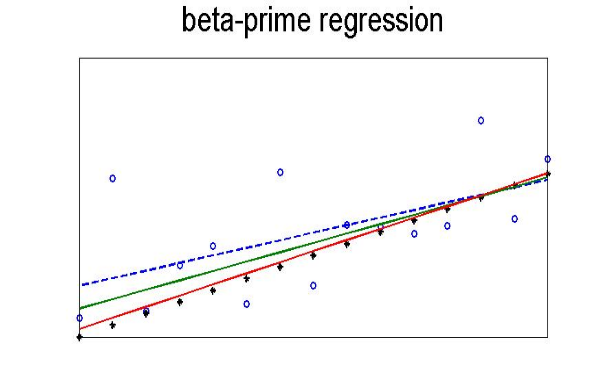
Lineární regrese pro data z R_+ (beta-prime)



Lineární regrese pro data z R_+



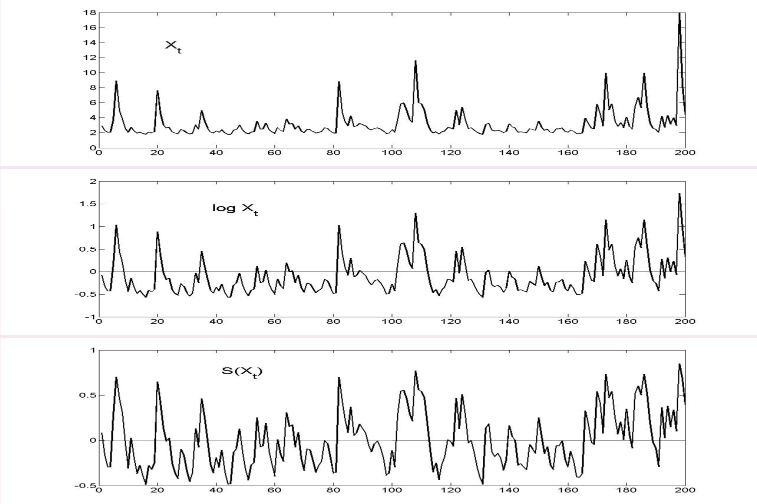
Lineární regrese pro data z R_+



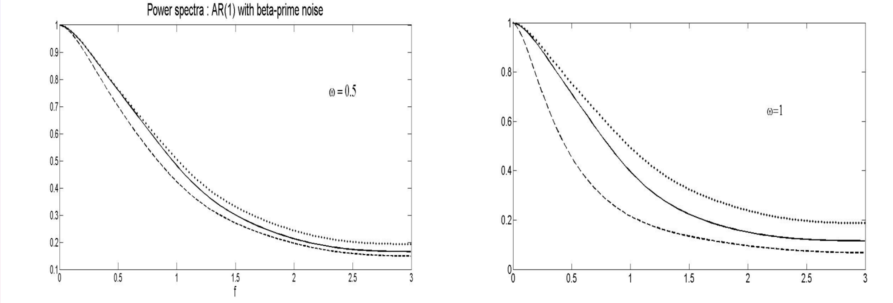
$$y_i = a_0 + a_1 x_i + \varepsilon_i \qquad a_1 = 0.25, a_0 = -2.5$$
$$\varepsilon_i \sim \text{beta-prime}$$

		Weibull		
method	$\hat{a}_1$	$\hat{a}_0$	$std(\hat{a}_1)$	$std(\hat{a}_0)$
least-squares	0.25619	-1.51736	0.07274	1.04505
Huber	0.25104	-2.57570	0.01836	0.32221
score	0.250017	-2.39620	0.00223	0.28503
		Pareto		
Huber	0.24860	-1.94217	0.01525	0.30225
score	0.25031	-2.25081	0.00489	0.07948

Skalární skór $S_F(X_t)$ časové řady X_t

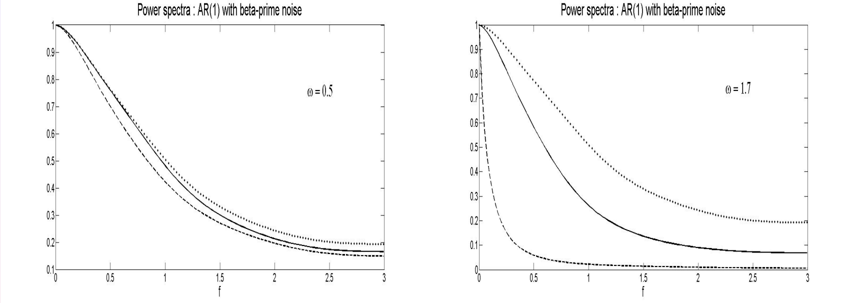


Odhad spektrální hustoty



$X_t = 0.4X_{t-1} + Z_t$ with beta-prime noise with different ω
Full line: $S_F(X_t)$, dashed line: $\log X_t$, dotted line: true

Odhad spektrální hustoty



$X_t = 0.4X_{t-1} + Z_t$ with beta-prime noise with different ω

Full line: $S_F(X_t)$, dashed line: $\log X_t$, dotted line: normal Z_t

Statistické metody při kontaminaci známého modelu

$(x_1, ..., x_n)$ z F_θ

$\psi_F(x;\theta)$ je $S_F(x;\theta)$ nebo její huberizovaná verze

data jako latentní hodnoty $(\psi_F(x_1, \theta), ..., \psi_F(x_n, \theta))$

odhady $\hat{\theta}_n, \hat{x}^* = x^*(\hat{\theta}_n), ...$

Upravená data pro další operace

$$[\psi_F(x_1; \hat{\theta}_n), ..., \psi_F(x_n; \hat{\theta}_n)]$$

nový model F_θ^2 ...

Reference.

Fabián Z.: Score function of distribution and revival of the moment method (accepted in Communication in Statistics, Theory-Method)

Děkuji za pozornost