Informace a neurčitost

Zdeněk Fabián
Ústav informatiky AVČR Praha

Rybník
ROBUST 2018

Podivné úkazy v matematické statistice

$X, F, f(x)$

střední hodnota EX

X na $(0, \infty)$ střed: mean, mode, median

$\sum X_i$

Dá se to spravit

$X \to S_F(X)$ kde S_F je vlivová funkce F

střední hodnota $ES_F(X) = 0$

$$ES_F^2 < \infty$$

X na $(0, \infty)$ střed: $x^* : S_F(x;\theta) = 0$

$\sum S_F(X_i)$ se chová podle CLT

Vlivová funkce exponeciálního F

$$\mathcal{R}, Y, G, g(y) \quad \rightarrow \quad (0,\infty), X, F, f(x)$$

$$y = \log x$$

$$g(y) = e^y e^{-e^y} \quad \rightarrow \quad f(x) = xe^{-x}\frac{1}{x} = e^{-x}$$

$$S_G(y) = -\frac{g'(y)}{g(y)} = e^y - 1 \quad \rightarrow \quad S_F(x) = x - 1$$

$$\mu = \log \tau$$

$$g(y) = e^{y-\mu} e^{-e^{y-\mu}} \quad \rightarrow \quad f(x) = \frac{x}{\tau} e^{-\frac{x}{\tau}} \frac{1}{x} = \frac{1}{\tau} e^{-x/\tau}$$

$$S_G(y-\mu) = e^{y-\mu} - 1 \quad \rightarrow \quad T_F(x;\tau) = \frac{x}{\tau} - 1$$

likelihood funkce pro F: $\frac{\partial}{\partial \tau} \log f(x) = \frac{1}{\tau}(\frac{x}{\tau} - 1)$

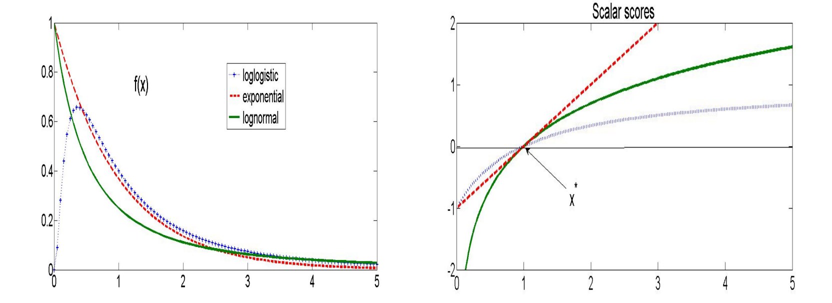
Definice. F rozdělení na $\mathcal{X} \subseteq \mathcal{R}$ s hustotou $f(x)$, $\eta(x)$. Funkce

$$T_F(x) = -\frac{1}{f(x)} \frac{d}{dx} \left[\frac{1}{\eta'(x)} f(x) \right] \tag{1}$$

je transformation-based score (t-score) rozdělení F

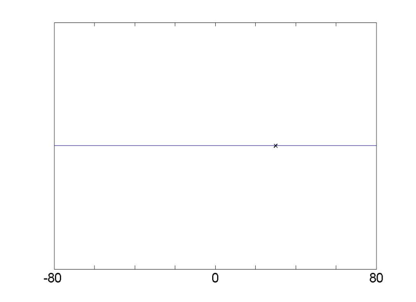
$$T_F(x) = \begin{cases} -x\frac{f'(x)}{f(x)} - 1 & (0,\infty) \quad \eta(x) = \log x \\ -1 + 2x - x(1-x)\frac{f'(x)}{f(x)} & (0,1) \quad \eta(x) = \log \frac{x}{1-x} \end{cases}$$

$x^* : T_F(x) = 0 \qquad S_F(x) = \eta'(x^*) T_F(x)$ vlivová funkce


f(x)
loglogistic
exponential
lognormal
0
0.2
0.4
0.6
0.8
1
0
1
2
3
4
5
Scalar scores
2
1
0
-1
-2
x*
0
1
2
3
4
5

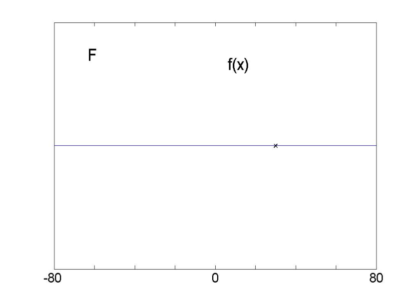
$S_F(x)$ skórová funkce rozdělení

Realizace spojité náhodné veličiny

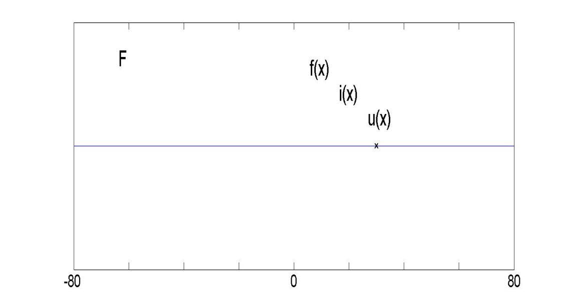


informace a neurčitost v x ?

Známe model



hustota informace $i(x)$ a neurčitosti $u(x)$?



Co tomu říká matematická statistika

Statistický pojem je Fisherova informace

Mějme $f(x;\theta), \theta = (\theta_1, ..., \theta_m) \in \Theta \subseteq R^m$. Fisher (likelihood) scores jsou

$$\psi_{\theta_j}(x) = \frac{\partial}{\partial \theta_j} \log f(x;\theta), \qquad j = 1, ..., m, \qquad \theta \in \Theta$$

Fisherova informace vzhledem k θ_j

$$I(\theta_j) = E\psi_j^2$$

Co tomu říká teorie informace

Střední neurčitost rozdělení je differenciální entropie

$$h_= E(-\log f) = \int_{\mathcal{X}} \log \frac{1}{f(x)} f(x)\, dx$$

Ale: h je někdy záporná:

$$f_N(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$$

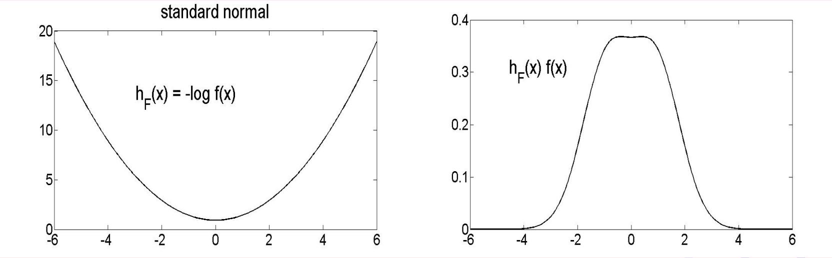
$h = \log \sqrt{2\pi e}\sigma < 0$ pro malé σ

Differenciální entropie normálního rozdělení

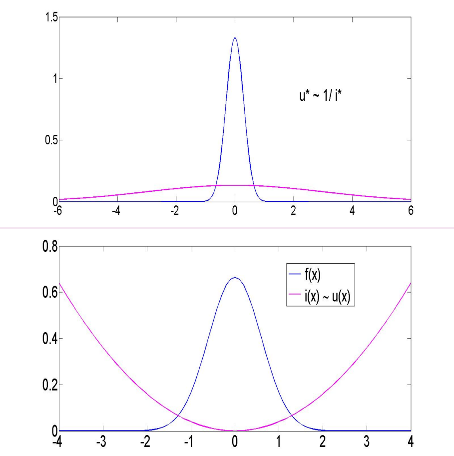
$h_F(x)$ ... hustota neurčitosti $f_N(0,1)$:

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

$$h = Eh_F = \int_R h_F(x) f(x)\, dx$$



Požadavky na hustotu informace a neurčitosti

- 

Normální rozdělení

$$f_N(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$$

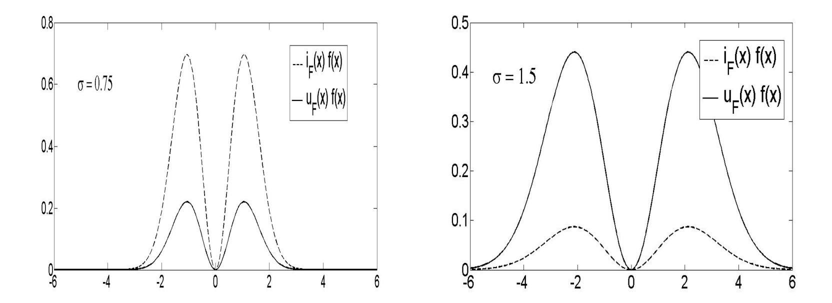
$$S_F(x) = \frac{x-\mu}{\sigma^2}$$

$$x^* = \mu$$

$$i_F(x) = S_F^2(x) = \frac{(x-\mu)^2}{\sigma^4} \qquad Ei_F = \frac{1}{\sigma^2}$$

$$u_F(x) = \frac{i_F(x)}{[Ei_F]^2} = (x-\mu)^2 \qquad Eu_F = \sigma^2$$

Normální rozdělení



Definice

vlivová funkce (skórová funkce)

$$S_F(x)$$

hustota informace

$$i_F(x) = S_F^2(x)$$

hustota neurčitosti

$$u_F(x) = \frac{i_F(x)}{[Ei_F]^2}$$

varibilita (score variance)

$$\omega^2 = Eu_F = \frac{1}{Ei_F}$$

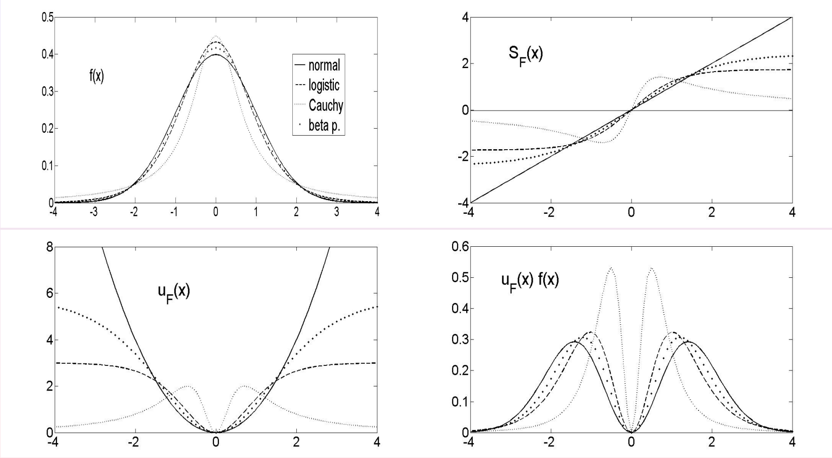
Rozdělení na R

$z = \frac{x-\mu}{\sigma}$ a $\omega^2 = Eu_F$

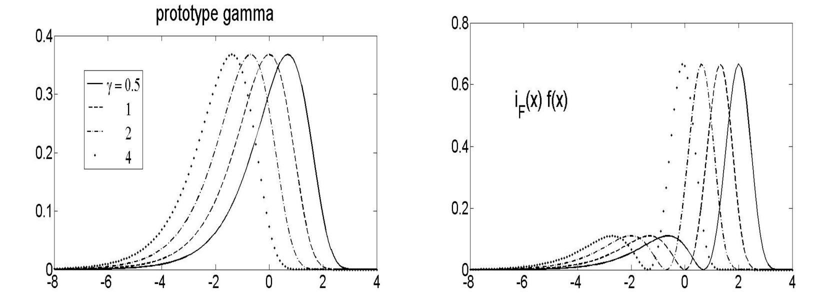
F	$f(x)$	$S_F(x)$	ω^2
Gumbel	$\frac{1}{\sigma} e^z e^{e^z}$	$\frac{1}{\sigma}(e^z - 1)$	σ^2
extreme value	$\frac{1}{\sigma} e^{-z} e^{-e^{-z}}$	$\frac{1}{\sigma}(1 - e^{-z})$	σ^2
Cauchy	$\frac{1}{\sigma\pi} \frac{1}{1+z^2}$	$\frac{2z}{1+z^2}$	$2\sigma^2$
logistic	$\frac{1}{\sigma} \frac{z}{(z+1)^2}$	$\frac{1}{\sigma} \frac{z-1}{z+1}$	$3\sigma^2$
prototype gamma	$\frac{\gamma^\alpha}{\Gamma(\alpha)} e^{\alpha x} e^{-\gamma e^x}$	$\gamma e^x - \alpha$	$1/\alpha$
prototype beta	$\frac{1}{B(p,q)} \frac{e^{px}}{(e^x+1)^{p+q}}$	$\frac{pe^x - q}{e^x+1}$	$\frac{p+q+1}{pq}$

Pro $\omega^2 = 1$ je $i_F(x) = u_F(x)$. Pro $\sigma = const$. nemá normální rozdělení maximální neurčitost

Rozdělení s jednotkovou variabilitou: $\omega^2 = 1$



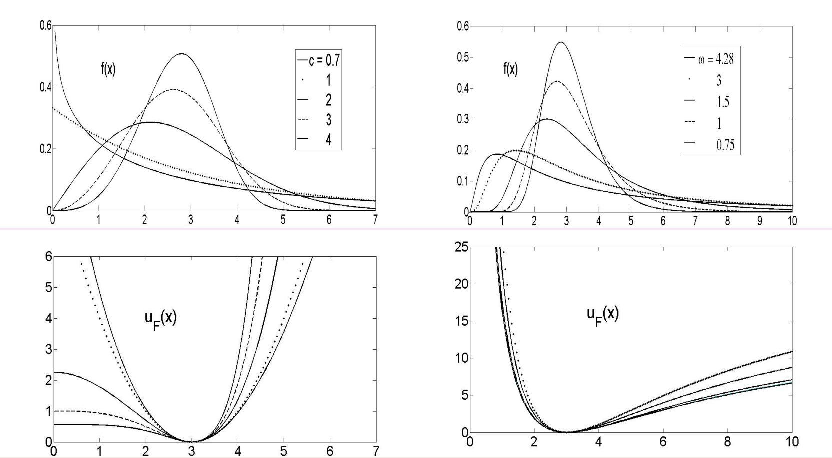
Prototype gamma



$\mathcal{X} = (0, \infty)$

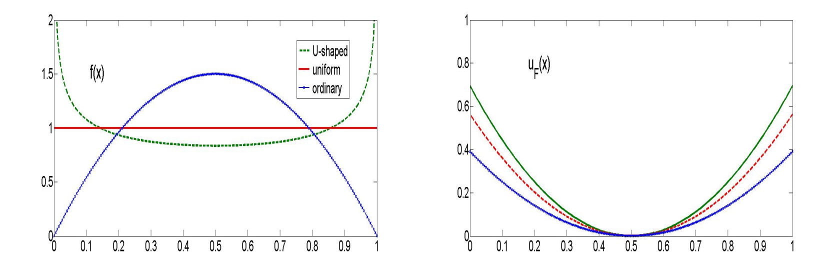
F	$f(x)$	$S_F(x)$	x^*	ω^2
lognormal	$\frac{c}{\sqrt{2\pi}x} e^{-\frac{1}{2}\log^2(\frac{x}{\tau})^c}$	$\frac{c}{\tau}\log(\frac{x}{\tau})^c$	τ	$\frac{\tau^2}{c^2}$
exponential	$\frac{1}{\tau}e^{-x/\tau}$	$\frac{1}{\tau}(\frac{x}{\tau}-1)$	τ	τ^2
Weibull	$\frac{c}{x}(\frac{x}{\tau})^c e^{-(\frac{x}{\tau})^c}$	$\frac{c}{\tau}[(\frac{x}{\tau})^c - 1]$	τ	$\frac{\tau^2}{c^2}$
loglogistic	$\frac{c}{\tau}\frac{(x/\tau)^{c-1}}{[(x/\tau)^c+1]^2}$	$\frac{c}{\tau}\frac{(x/\tau)^c-1}{(x/\tau)^c+1}$	τ	$\frac{3\tau^2}{c^2}$
gamma	$\frac{\gamma^\alpha}{x\Gamma(\alpha)}x^\alpha e^{-\gamma x}$	$\frac{\gamma^2(x-x^*)}{\alpha}$	$\frac{\alpha}{\gamma}$	$\frac{\alpha}{\gamma^2}$
beta-prime	$\frac{1}{B(p,q)}\frac{x^{p-1}}{(x+1)^{p+q}}$	$\frac{q^2}{p}\frac{x-x^*}{x+1}$	$\frac{p}{q}$	$\frac{p(p+q+1)}{q^3}$

Weibull a beta-prime



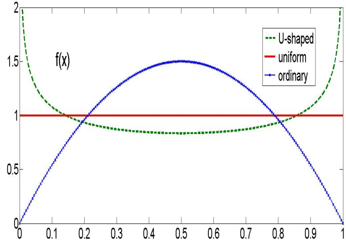
Beta na $(0,1)$

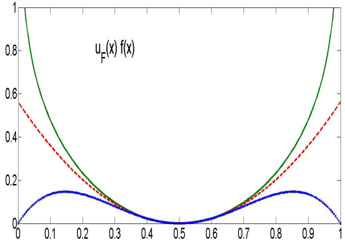
$$f(x) = \frac{1}{B(p,q)} x^{p-1}(1-x)^{q-1} \qquad T_F(x) = (p+q)x - p$$



Beta na $(0,1)$

$$f(x) = \frac{1}{B(p,q)} x^{p-1}(1-x)^{q-1} \qquad T(x) = (p+q)x - p$$





ω a diferenciální entropie

F	e^h	ω
normal	$\sqrt{2\pi e}\sigma$	σ
Cauchy	$4\pi\sigma$	$\sqrt{2}\sigma$
exponential	$e\tau$	τ
lognormal	$\sqrt{2\pi e}\,\tau/c$	τ/c
Weibull	$e^{(1+\epsilon(1-1/c))}\ \tau/c$	τ/c
uniform$_{(a,b)}$	$b-a$	$\frac{\sqrt{3}}{4}(b-a)$

Fabián Z. (2016): Score function of distribution and revival of the moment method. Communication in Statistics, Theory-Methods 45: 1118-1136.

Díky za pozornost