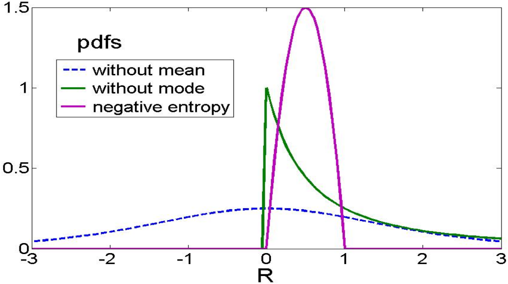
New view on continuous probability distributions

Zdeněk Fabián

December 5, 2020

Continuous random variables $X \sim (\mathcal{R}, f)$



$$f(x) \geq 0 \qquad \mathcal{X} \in \mathcal{R} \text{ support}$$

Pearson's system:

$$\frac{f'(x)}{f(x)} = \frac{a + yx}{c_0 + c_1 x + c_2 x^2}$$

Some basic concepts

- I. Probability: numerical characteristics are moments

$$EX^k = \int_{\mathcal{X}} x^k f(x)\, dx$$

mean value: EX
variance: $VarX = E(X - EX)^2$

Some basic concepts

- I. Probability: numerical characteristics are moments

  $$EX^k = \int_{\mathcal{X}} x^k f(x)\, dx$$

  mean value: EX
  variance: $VarX = E(X - EX)^2$

- II. Information and uncertainty:

  Information theory: mean uncertainty of a distribution is

  differential entropy $H_F = -\int_{\mathcal{X}} f(x) \log f(x)\, dx$

  Statistics: Fisher information (defined for parameters of parametric distributions only)

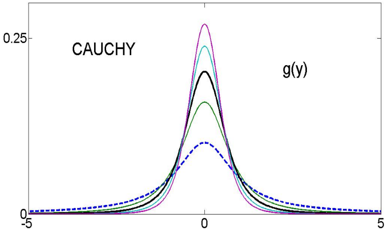
But there are some problems:

Problem I: Moments

There are heavy-tailed distributions having neither variance nor even mean value

Cauchy distribution

$$f(x) = \frac{1}{\pi\sigma(1 + \frac{x^2}{\sigma^2})}, \qquad EX \sim x^{-1}$$

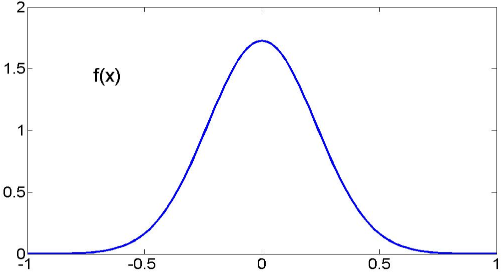


Problem II: Differential entropy

even in the case of the normal distribution

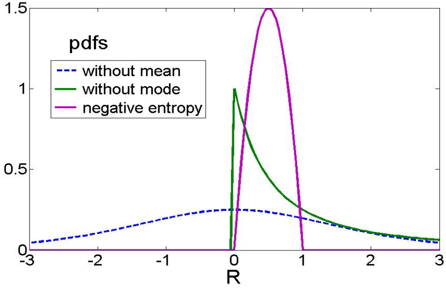
$$H_F = -\int_{\mathcal{X}} f(x) \log f(x)\, dx = \log\left(\sigma\sqrt{2\pi e}\right)$$

is negative if $\sigma < 1/\sqrt{2\pi e}$



Description of continuous distributions by classical probability theory is useful for distributions with neither too small nor too large variability

Problem III. Typical value: mean, mode or median ?



Point estimation

$X: \quad (x_1, ..., x_n)$ z $\mathcal{F}_\theta = \{F_\theta : \theta \in \Theta \subseteq \mathcal{R}^m\}, \theta = (\theta_1, ..., \theta_m)$

Moment method

$$\hat{\theta}_n : \quad \frac{1}{n}\sum_{i=1}^{n} x_i^k = EX^k(\theta) \qquad k = 1, ..., m$$

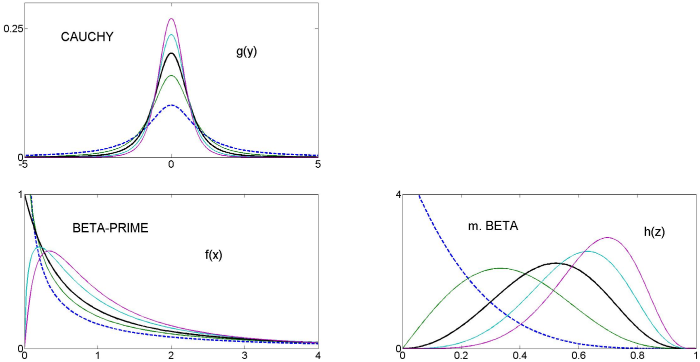
Inference function $\psi(x;\theta)$

$$\frac{1}{n}\sum_{i=1}^{n} \psi_k(x_i;\theta) = E\psi_k(\theta) \qquad k = 1, ..., m$$

classical: ML $\quad \psi_k = \frac{\partial}{\partial\theta_k}\log f(x;\theta)$ Fisher score, $E\psi_k^2$ FI

robust: $\quad \tilde{\mu}, \tilde{\sigma}$: $\quad \psi(x;\tilde{\mu},\tilde{\sigma})$ a bounded function

New view



Every model has a finite center and variability and their estimates are the center and variance of a random sample from them

Starting position

- We have find a "natural" inference function $S_F(x)$ of the model F and study random variables $S_F(X)$ instead of X

Starting position

- We have find a "natural" inference function $S_F(x)$ of the model F and study random variables $S_F(X)$ instead of X

- Let $G : g(y - \mu)$

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

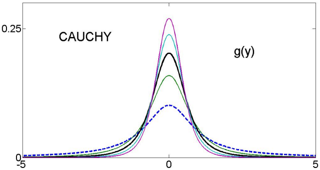
Starting position

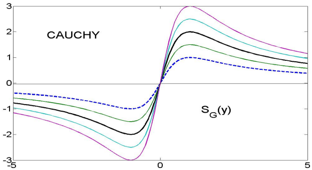
- We have find a "natural" inference function $S_F(x)$ of the model F and study random variables $S_F(X)$ instead of X

- Let $G : g(y-\mu)$

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

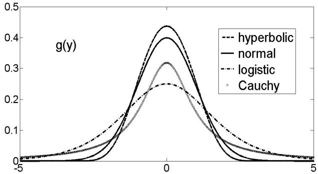
- **score function** $S_G(y) = -\frac{g'(y)}{g(y)}$ $\qquad S_{Cauchy}(y) = \frac{1}{\sigma}\frac{2y/\sigma}{1+(y/\sigma)^2}$

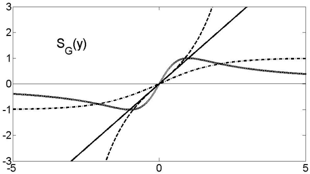




6 simple score functions: 6 different types

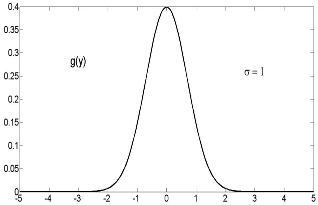
type	$S_G(y)$	$g(y)$	distribution
UE	$\sinh y = \frac{e^y - e^{-y}}{2}$	$\frac{1}{K} e^{-\frac{1}{2}(e^y + e^{-y})}$	hyperbolic
UP	y	$\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}$	normal
BU	$e^y - 1$	$e^y e^{-e^y}$	Gumbel
UB	$1 - e^{-y}$	$e^{-y} e^{-e^{-y}}$	extreme value
BB	$\tanh \frac{y}{2} = \frac{e^y - 1}{e^y + 1}$	$\frac{e^y}{(1+e^y)^2}$	logistic
BR	$\frac{2y}{1+y^2}$	$\frac{1}{\pi(1+y^2)}$	Cauchy

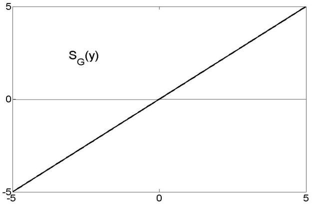




Normal distribution

$$g(y;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{y-\mu}{\sigma})^2} \qquad S_G(y;\mu,\sigma) = \frac{y-\mu}{\sigma^2}$$

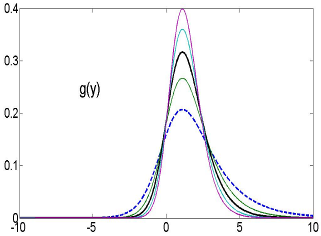


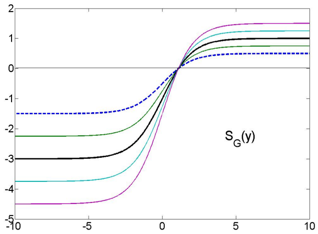


$$EX = \mu, \; VarX = \frac{1}{ES_G{}^2} = \sigma^2$$

'Prototype beta'

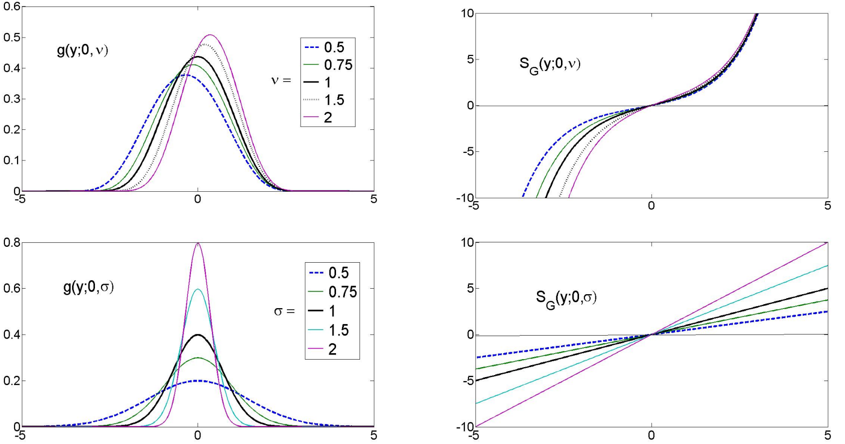
$$g(y) = \frac{e^{py}}{(e^y+1)^{p+q}} \qquad S_G(y) = \frac{qe^y - p}{e^y + 1}$$



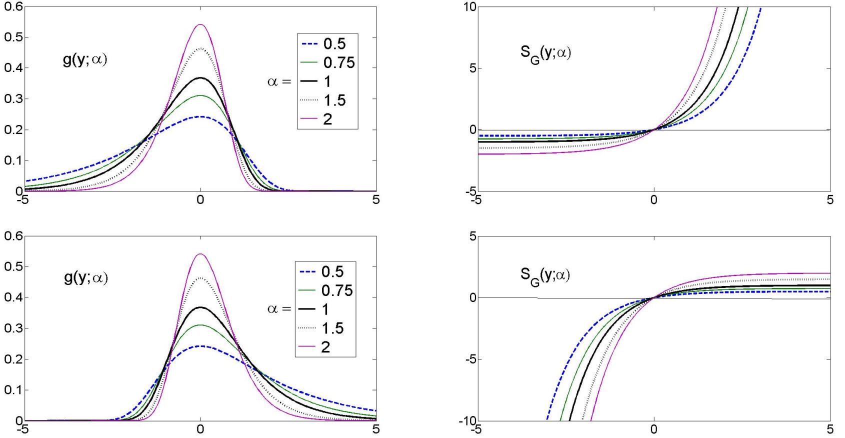


$EX = \log \frac{p}{q-1}$ $\qquad y^* = \log \frac{p}{q}$

Hyperbolic and normal



Gumbel and extreme value



Score random variable $S_G(Y)$

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

  typical value y^* : $S_G(y) = 0$ (mode)

  score moments: $ES_G^k = \int_{\mathcal{X}} S_G^k(y) g(y)\, dy$ are finite

  $ES_G = 0, ES_G^2$ Fischer information for y^*

Score random variable $S_G(Y)$

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

  typical value y^* : $S_G(y) = 0$ (mode)

  score moments: $ES_G^k = \int_{\mathcal{X}} S_G^k(y) g(y)\, dy$ are finite

  $ES_G = 0, ES_G^2$ Fischer information for y^*

- suggested measure of variability: score variance

  $$\omega_G^2 \equiv \frac{1}{ES_G^2}$$

  parametric estimates: SM method $(y_1, ..., y_n)$ from G

$$\frac{1}{n} \sum_{i=1}^{n} S_G^k(y_i; \theta) = ES_G^k(\theta), \qquad k = 1, ..., m$$

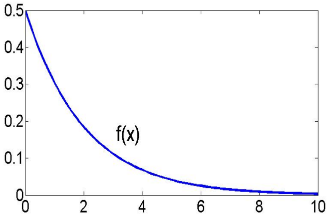
Why this approach is not used in statistics ?

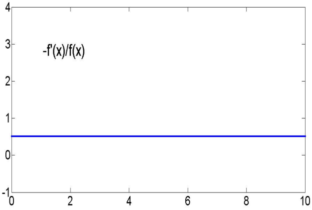
- Why the score function is not taken as inference function ?

Why this approach is not used in statistics ?

- Why the score function is not taken as inference function ?
- The reason is that $-\frac{f'(x)}{f(x)}$ for F on $\mathcal{X} \neq \mathcal{R}$ does not work:

$f(x) = \frac{1}{2}e^{-x/2}$





Scalar score function on $\mathcal{X} \neq \mathcal{R}$

- **Idea: Scalar score (influence) function of F on $\mathcal{X} \neq \mathcal{R}$ exists. It is given by different formulas on different $\mathcal{X}$.**

Key word: Transformation

Scalar score function on $\mathcal{X} \neq \mathcal{R}$

- **Idea: Scalar score (influence) function of F on $\mathcal{X} \neq \mathcal{R}$ exists. It is given by different formulas on different $\mathcal{X}$.**

Key word: Transformation

- Y na $\mathcal{R}$, G, g, S_G. $\eta^{-1} : \mathcal{R} \to \mathcal{X}$ strictly increasing continuous

  Transf. r.v. $X = \eta^{-1}(Y)$ has density
  $(y \to \log x, x \to e^y)$

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

Scalar score function on $\mathcal{X} \neq \mathcal{R}$

- **Idea: Scalar score (influence) function of F on $\mathcal{X} \neq \mathcal{R}$ exists. It is given by different formulas on different $\mathcal{X}$.**

Key word: Transformation

- Y na $\mathcal{R}$, G, g, S_G. $\eta^{-1} : \mathcal{R} \to \mathcal{X}$ strictly increasing continuous

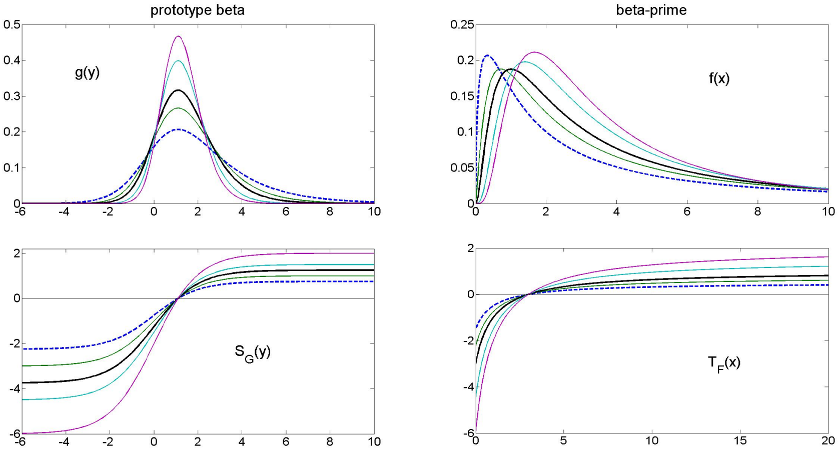
  Transf. r.v. $X = \eta^{-1}(Y)$ has density
  $(y \to \log x, x \to e^y)$

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

- Definition: **t-score** of $F(x) = G(\eta(x))$

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

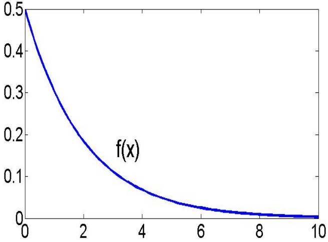
Transformed distributions

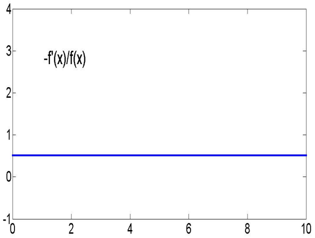


Exponential distribution $Y = \eta(X) = \log X$

$$g(y) = e^y e^{-e^y} \qquad S_G(y) = e^y - 1$$

$$f(x) = xe^{-x}\frac{1}{x} \qquad T_F(x) = \frac{d}{dx}xe^{-x} = x - 1$$





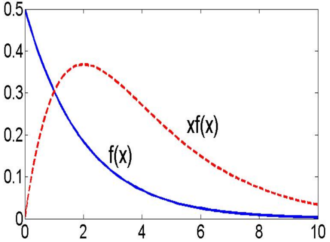
Exponential distribution $Y = \eta(X) = \log X$

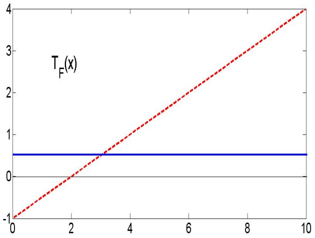
'Prototype'

$$g(y) = e^y e^{-e^y} \qquad S_G(y) = e^y - 1$$

transformed

$$f(x) = xe^{-x}\frac{1}{x} \qquad T_F(x) = \frac{d}{dx}(xe^{-x}) = x - 1$$





Transformed distributions

type	$S_G(y)$	$T_F(x)$	$f(x)$	
UE	$\frac{e^y - e^{-y}}{2}$	$\frac{1}{2}(x - \frac{1}{x})$	$\frac{1}{Kx} e^{-(x+1/x)}$	inv. Gaussian
UP	y	$\log x$	$\frac{1}{\sqrt{2\pi}x} e^{-\frac{1}{2}\log^2 x}$	lognormal
BU	$e^y - 1$	$x - 1$	e^{-x}	exponential
UB	$1 - e^{-y}$	$1 - 1/x$	$\frac{1}{x^2} e^{-1/x}$	Fréchet
BB	$\frac{e^y - 1}{e^y + 1}$	$\frac{x-1}{x+1}$	$\frac{1}{(1+x)^2}$	loglogistic
BR	$\frac{2y}{1+y^2}$	$\frac{2\log x}{1+\log^2 x}$	$\frac{1}{\pi x(1+\log^2 x)}$	log-Cauchy

Transformed location parameter on $\mathcal{R}^+$

- $$\frac{y-\mu}{\sigma} \to \frac{\log x - \log \tau}{\sigma} = \log\left(\frac{x}{\tau}\right)^c$$

Transformed location parameter on $\mathcal{R}^+$

- $$\frac{y-\mu}{\sigma} \to \frac{\log x - \log \tau}{\sigma} = \log\left(\frac{x}{\tau}\right)^c$$

- Let $\frac{1}{\tau} f(x/\tau)$

  Fisher score
  $$= \frac{\partial}{\partial \tau} \log(\tfrac{1}{\tau} f(x/\tau)) = \tfrac{1}{\tau}[-1 - \tfrac{f'}{f}\tfrac{x}{\tau}]$$

Transformed location parameter on $\mathcal{R}^+$

- $$\frac{y-\mu}{\sigma} \to \frac{\log x - \log \tau}{\sigma} = \log\left(\frac{x}{\tau}\right)^c$$

- Let $\frac{1}{\tau} f(x/\tau)$

  Fisher score
  $$= \frac{\partial}{\partial \tau} \log(\tfrac{1}{\tau} f(x/\tau)) = \tfrac{1}{\tau}[-1 - \tfrac{f'}{f}\tfrac{x}{\tau}]$$

- t-score
  $$T_F(x/\tau) = -\tfrac{\tau}{f(x/\tau)} \tfrac{d}{dx}[x \tfrac{1}{\tau} f(x/\tau)] = -1 - \tfrac{x}{\tau}\tfrac{f'}{f}$$

Transformed location parameter on $\mathcal{R}^+$

- $$\frac{y-\mu}{\sigma} \to \frac{\log x - \log \tau}{\sigma} = \log\left(\frac{x}{\tau}\right)^c$$

- Let $\frac{1}{\tau} f(x/\tau)$

  Fisher score
  $$= \frac{\partial}{\partial \tau} \log(\frac{1}{\tau} f(x/\tau)) = \frac{1}{\tau}[-1 - \frac{f'}{f}\frac{x}{\tau}]$$

- t-score
  $$T_F(x/\tau) = -\frac{\tau}{f(x/\tau)}\frac{d}{dx}[x\frac{1}{\tau}f(x/\tau)] = -1 - \frac{x}{\tau}\frac{f'}{f}$$

- t-score is proportional to the Fisher score for central parameter

For given F: which $\eta : \mathcal{X} \to \mathcal{R}$?

a) universal:

$$\eta(x) = \begin{cases} \log x & \text{pro} \quad \mathcal{X} = (0, \infty) \\ \log \frac{(x)}{(1-x)} & \text{pro} \quad \mathcal{X} = (0, 1) \end{cases}$$

b) innate:

loggamma on $\mathcal{X} = (1, \infty)$

$$f(x) = \frac{c^\alpha}{\Gamma(\alpha)} (\log x)^{\alpha-1} \frac{1}{x^{c+1}}$$

For given F: which $\eta : \mathcal{X} \to \mathcal{R}$?

a) universal:

$$\eta(x) = \begin{cases} \log x & \text{pro} \quad \mathcal{X} = (0, \infty) \\ \log \frac{(x)}{(1-x)} & \text{pro} \quad \mathcal{X} = (0, 1) \end{cases}$$

b) innate:

loggamma on $\mathcal{X} = (1, \infty)$

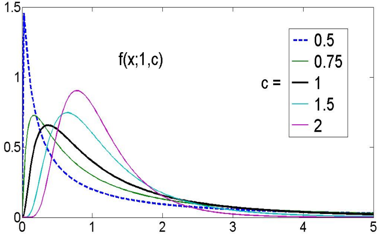
$$f(x) = \frac{c^\alpha}{\Gamma(\alpha)} (\log x)^{\alpha-1} \frac{1}{x^{c+1}} = \frac{c^\alpha}{\Gamma(\alpha)} (\log x)^\alpha \frac{1}{x^c} \frac{1}{x \log x}$$

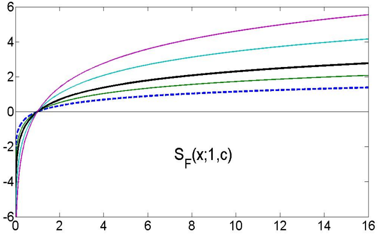
$$\eta(x) = \log \log x$$

Lognormal

$$g(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{y-\mu}{\sigma})^2} \qquad S_G(y) = \frac{y-\mu}{\sigma^2}$$

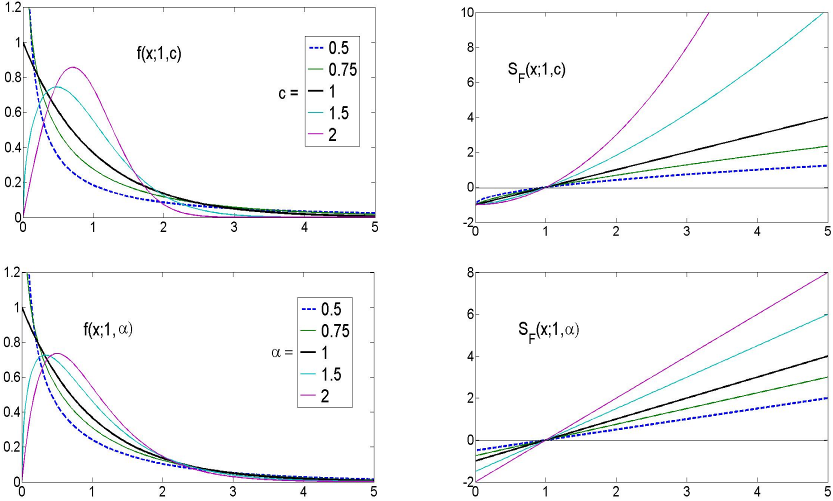
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{1}{2}\log^2(\frac{x}{\tau})^c} \qquad T_F(x) = c\log(x/\tau)^c$$



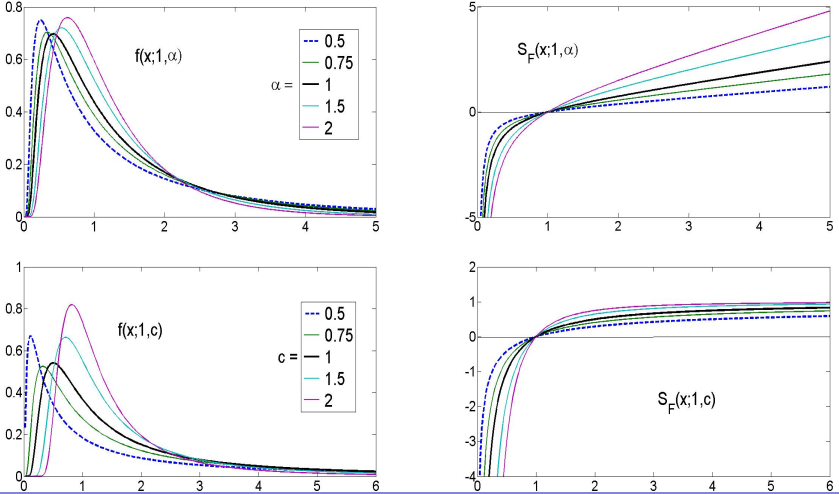


$c = 1/\sigma$

Type BU: Weibull, **Type BU**: gamma

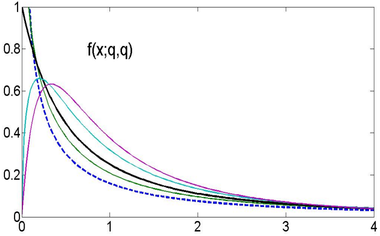


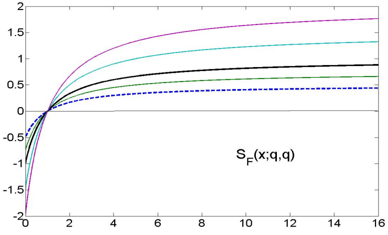
Type UE: gen. inverse Gaussian, **Type UB**: Fréchet



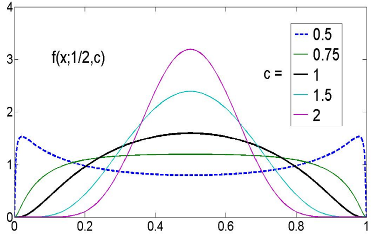
Type UB: beta-prime

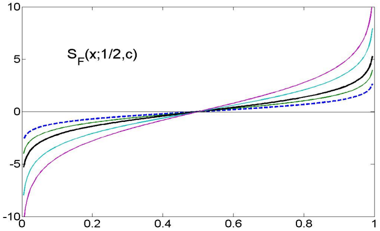
$$f(x) = \frac{1}{B(p,q)} \frac{x^{p-1}}{(x+1)^{p+q}} \qquad T_F(x) = \frac{qx-p}{x+1}$$





Distributions on $(0, 1)$: **Type UP**: Johnson

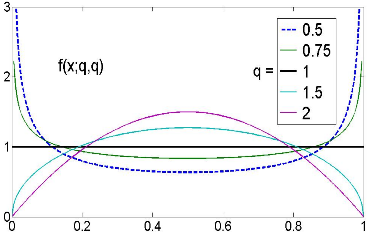


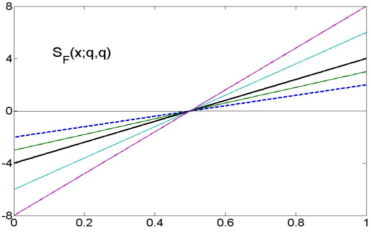


Type BB: beta

$$h(x;p,q) = \frac{1}{B(p,q)} x^{p-1}(1-x)^{q-1} \qquad T_H(x;p,q) = (p+q)x - p$$

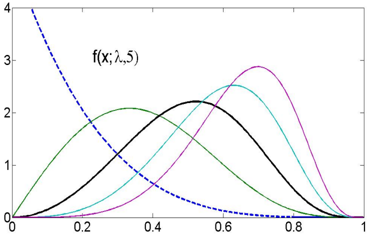
with $x^* = p/(p+q)$

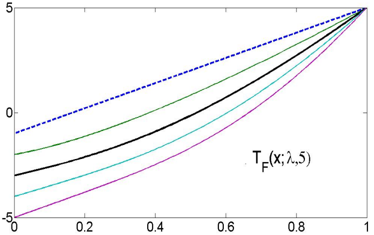




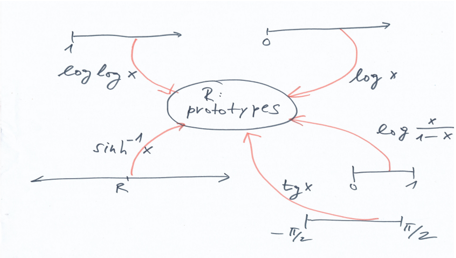
Kumaraswamy

$$h(x) = \lambda\varphi x^{\lambda-1}(1-x^{\lambda})^{\varphi-1} \qquad T_H(x) = (1+\lambda)x - \lambda + \lambda(\varphi-1)\frac{(1-x)x^{\lambda}}{1-x^{\lambda}}$$



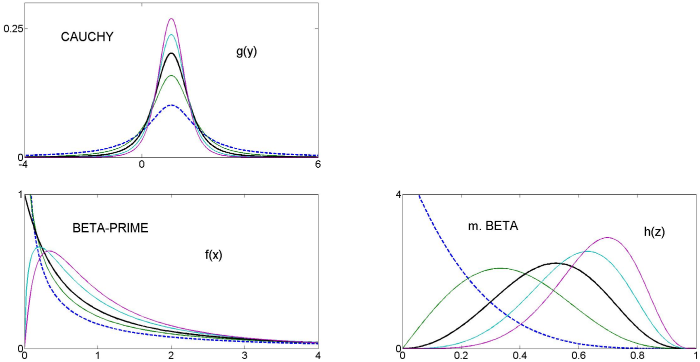


Systematics of distributions



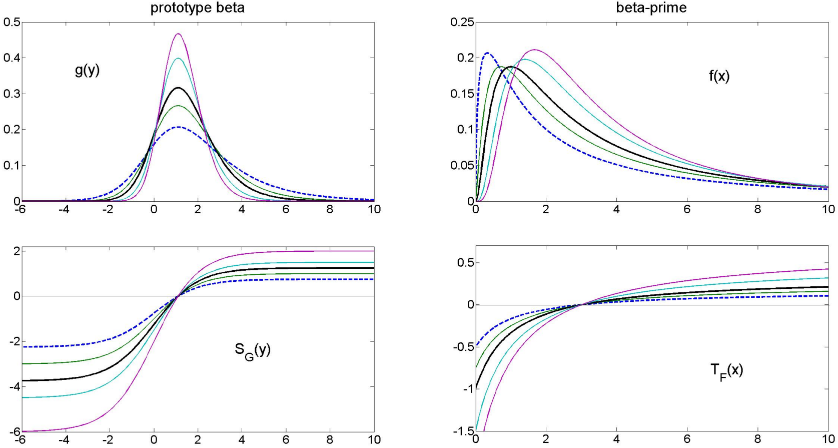
Mean, mode or median ?

$$EX = \int x f(x) \, dx$$



Typical value is $x^* : T_F(x) = 0$

score mean $x^* = \eta^{-1}(y^*)$ is the image of the mode of the prototype



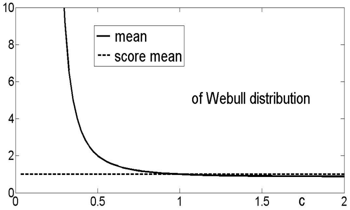
Mean EX and score mean x^*

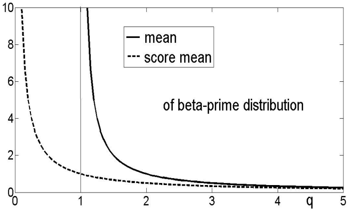
$$f(x) = cx^{c-1}e^{-x^c}$$

$$EX = \Gamma(1+1/c) \qquad x^* = 1$$

$$f(x) = \frac{q}{(1+x)^{q+1}}$$

$$EX = \frac{1}{q-1} \qquad x^* = \frac{1}{q}$$





Score average

$$\bar{S}_F = \frac{1}{n}\sum_{i=1}^{n} S_F(X_i)$$

$$S_F(\hat{x}^*) = \bar{S}_F \qquad \text{so that} \qquad \hat{x}^* = S_F^{-1}(\bar{S}_F)$$

In case of some distributions $\hat{x}^*$ is a known statistic:

distribution	$\hat{x}^*$
normal, gamma, beta	$\bar{x}$
lognormal	$\bar{x}_{Geometric}$
Weibull (c const.)	$\frac{1}{n}(\sum x_i^c)^{1/c}$
heavy tails	$\bar{x}_{Harmonic}$

Score variance $\omega_F^2 = 1/ES_F^2$

G	$g(y)$	$S_G(y)$	y^*	ω_G^2	$VarX$
normal	$\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}$	u/σ	μ	σ^2	σ^2
logistic	$\frac{1}{\sigma}\frac{u}{(u+1)^2}$	$\frac{1}{\sigma}\frac{u-1}{u+1}$	μ	$3\sigma^2$	$\pi^2\sigma^2/3$
Cauchy	$\frac{1}{\sigma\pi}\frac{1}{1+u^2}$	$\frac{1}{\sigma}\frac{2u}{1+u^2}$	μ	$2\sigma^2$	–

$$u = (y - \mu)/\sigma$$

F	$f(x)$	$T_F(x)$	x^*	ω_F^2
gamma	$\frac{\gamma^\alpha}{x\Gamma(\alpha)} x^\alpha e^{-\gamma x}$	$\gamma x - \alpha$	$\frac{\alpha}{\gamma}$	$\frac{\alpha}{\gamma^2}$
Weibull	$\frac{c}{x}(\frac{x}{\tau})^c e^{-(\frac{x}{\tau})^c}$	$c((\frac{x}{\tau})^c - 1)$	τ	$\frac{\tau^2}{c^2}$

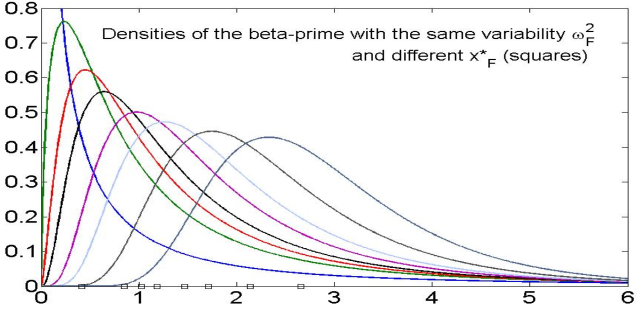
Weibull $\quad EX = \tau\Gamma(1/c) \quad VarX = \tau^2\Gamma(2/c)\Gamma^2(1/c)$

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

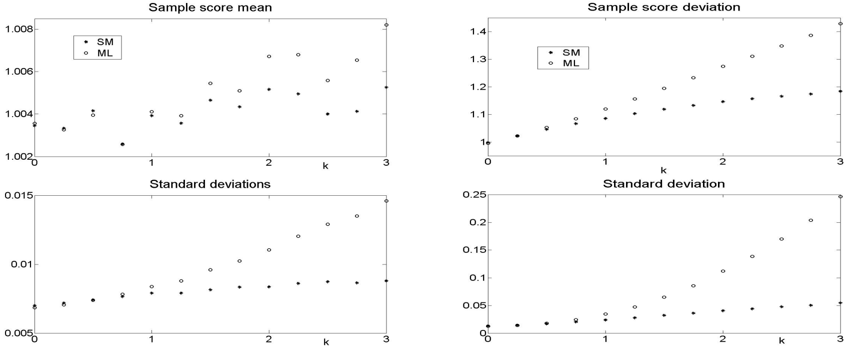
Central and variability characteristics:

Euclidean $EX = \frac{p}{q-1}$, $VarX = \frac{p(p+q-1)}{(q-2)(q-1)^2}$

Scalar-valued score $x^* = \frac{p}{q}$ $\omega_F^2 = \frac{p(p+q+1)}{q^3}$



Estimates of the score mean and score variance in a contaminated beta-prime model



maximum likelihood   SM score moment estimates

New characteristics of continuous random variables

scalar-valued score: likelihood score for
the typical value of distribution, the **score mean**

and **score variance**, representing variability