Kybernetika

Zdeněk Fabián; Igor Vajda

Core functions and core divergences of regular distributions

Kybernetika, Vol. 39 (2003), No. 1, [29]--42

Persistent URL: http://dml.cz/dmlcz/135506

KYBERNETIKA — VOLUME 39 (2003), NUMBER 1, PAGES 29–42

CORE FUNCTIONS AND CORE DIVERGENCES OF REGULAR DISTRIBUTIONS

ZDENĚK FABIÁN AND IGOR VAJDA

On bounded or unbounded intervals of the real line, we introduce classes of regular statistical families, called Johnson families because they are obtained using generalized Johnson transforms. We study in a rigorous manner the formerly introduced concept of core function of a distribution from a Johnson family, which is a modification of the well known score function and which in a one-to-one manner represents the distribution. Further, we study Johnson parametrized families obtained by Johnson transforms of location and scale families, where the location is replaced by a new parameter called Johnson location. We show that Johnson parametrized families contain many important statistical models. One form appropriately normalized L_2 distance of core functions of arbitrary distributions from Johnson families is used to define a core divergence of distributions. The core divergence of distributions from parametrized Johnson families is studied as a special case.

Keywords: Johnson transforms, generalized Johnson distributions, core function of distributions, core divergences of distributions
AMS Subject Classification: 62E10, 62B10

1. INTRODUCTION

Let for every open set $0 \neq S \subseteq R$, $\mathcal{Q}_S$ be the class of probability measures (distributions) Q on Borel subsets B of the real line R which are regular in the sense that they are absolutely continuous with respect to the Lebesgue measure λ on R, and a version of the density

$$g = \frac{\mathrm{d}Q}{\mathrm{d}\lambda} \tag{1}$$

is differentiable in S and satisfies the relation

$$g(x) = \begin{cases} > 0 & \text{for} \quad x \in S \\ = 0 & \text{for} \quad x \in R - S. \end{cases} \tag{2}$$

In other words, $\mathcal{Q}_S$ is the class of all Lebesgue dominated probability measures Q on R supported by $S_Q = S$ and with well defined derivatives

$$\dot{g}(y) = \frac{\mathrm{d}g(y)}{\mathrm{d}y} : S \to R \tag{3}$$

of the respective Lebesgue densities. Since $\dot{g}$ and g are Lebesgue measurable, the Lebesgue integrals

$$I_Q = \int_S \left(\frac{\dot{g}}{g}\right)^2 \mathrm{d}Q = \int_S \frac{(\dot{g}(y))^2}{g(y)} \mathrm{d}y, \quad Q \in \mathcal{Q}_S, \tag{4}$$

are well defined, with values in the extended real line interval $[0, \infty]$. The expression I_Q is a *Fisher information* of $Q \in \mathcal{Q}_S$.

We are interested in special subfamilies $\mathcal{P}_S \subset \mathcal{Q}_S$ called Johnson families. They are defined for arbitrary intervals $S = (a, b) \subseteq R$ by the "parent family" $\mathcal{Q}_R$ using the family $\Psi_S = \{\psi = \psi_{x_0} : x_0 \in S\}$ of Johnson functions, where each $\psi_{x_0} : S \to R$ is an increasing one-to-one mapping defined for all $x \in S$ by the formula

$$\psi_{x_0}(x) = \begin{cases} \sinh^{-1}(x - x_0) & \text{if} \quad (a,b) = R \\ \ln \dfrac{x-a}{x_0-a} & \text{if} \quad -\infty < a < b = \infty \\ \ln \dfrac{(x-a)(b-x_0)}{(b-x)(x_0-a)} & \text{if} \quad -\infty < a < b < \infty \\ \ln \dfrac{b-x_0}{b-x} & \text{if} \quad -\infty = a < b < \infty. \end{cases} \tag{5}$$

The Johnson functions are nothing but the reversed Johnson transformations $\psi^{-1} = \psi_{x_0}^{-1}$, $x_0 \in S_1$, which are increasing one-to-one mappings $R \to S$ defined for all $y \in R$ by the formula

$$\psi_{x_0}^{-1}(y) = \begin{cases} x_0 + \sinh y & \text{if} \quad (a,b) = R \\ (x_0 - a)e^y & \text{if} \quad -\infty < a < b = \infty \\ \dfrac{a(b-x_0) + b(x_0-a)e^y}{b - x_0 + (x_0-a)e^y} & \text{if} \quad -\infty < a < b < \infty \\ (b - x_0)e^y & \text{if} \quad -\infty = a < b < \infty. \end{cases} \tag{6}$$

(see Johnson [3], Johnson and Kotz [4] and a generalization in Fabián [1]).

Definition 1. A Johnson family $\mathcal{P}_S$ is for every $S = (a, b) \subseteq R$ defined by

$$\mathcal{P}_S = \{P = Q\psi : Q \in \mathcal{Q}_R, \psi \in \Psi_S\} \tag{7}$$

where $Q\psi(B) = Q(\psi(B))$ and $\psi(B) = \{\psi(x) : x \in B\}$ for every Borel subset $B \subseteq S$.

The Johnson families $\mathcal{P}_S$ are supported by S and they are regular in the sense that the Lebesgue densities f of $P \in \mathcal{P}_S$ are positive and differentiable on S (see Proposition 1 in Section 2).

If $Q = \mathcal{Q}_R$ (a "parent family") then the ratio

$$s_Q = -\dot{g}/g \tag{8}$$

is the well known score function of Q supported by $S = R$.

Recently, Fabián [2] introduced the core functions T_P of distributions $P \in \mathcal{P}_S$ on arbitrary supports $S = (a, b) \subseteq R$ by the formula

$$T_P = s_Q(\psi) \tag{9}$$

where $\psi \in \Psi_S$ is the Johnson function and $Q = P\psi^{-1} \in \mathcal{Q}_R$. The core functions $s_{P\psi^{-1}}(\psi)$ differ from the score functions s_Q. Fabián [2] demonstrated that the point estimation in some parametrized subfamilies of $\mathcal{P}_S$ based on the core functions leads to robust versions of the estimation based on the score functions with acceptable levels of asymptotic inefficiencies. We show that such subfamilies of $\mathcal{P}_S$ include many important models of mathematical statistics. This motivates our deeper interest in the Johnson families and their parametrized subfamilies, and in the related core functions.

In Section 2 we study more rigorously the concept of core function. In particular, for $P \in \mathcal{P}_S, S \neq R$ we study the pairs $\psi \in \Psi_S$ and $Q \in \mathcal{Q}_R$ satisfying the relation $P = Q\psi$ considered in (7) and prove that (9) defines T_P unambiguously in the sense that it does not depend on the particular choice of the pair ψ and Q. We also prove rigorously that the core functions T_P are related in a one-to-one manner to the score functions s_P of Johnson distributions $P \in \mathcal{P}_S$ and, consequently, to the distributions themselves. This justifies the terminology "core function of P".

In Section 3 we introduce a core divergence $D(P_1, P_2)$ of distributions $P_1, P_2 \in \mathcal{P}_S$ and study its basic properties. In a number of examples we evaluate the core divergence and the well known Kullback divergence of distributions. In some of them we compare these two divergences and analyze differences between them from the point of view of statistical applications.

2. CORE FUNCTION

The first proposition summarizes for references later some properties of the Johnson families introduced by Definition 1.

Proposition 1. For every $S = (a, b) \subseteq R$, the Johnson family $\mathcal{P}_S$ is a subset of the regular family $\mathcal{Q}_S$. For every $P \in \mathcal{P}_S$ there exist $Q \in \mathcal{Q}_R$ with a differentiable density g on R, and a Johnson function $\psi = \psi_{x_0}$ from Ψ_S, such that the density $f = \mathrm{d}P/\mathrm{d}\lambda$ satisfies the relation

$$f(x) = g(\psi(x))\dot{\psi}(x), \quad x \in S, \tag{10}$$

where

$$\dot{\psi}(x) = \frac{\mathrm{d}}{\mathrm{d}x}\psi(x) = \begin{cases} \dfrac{1}{\sqrt{1+x^2}} & \text{if} \quad (a,b) = R \\ \dfrac{1}{x-a} & \text{if} \quad -\infty < a < b = \infty \\ \dfrac{(b-a)}{(x-a)\,(b-x)} & \text{if} \quad -\infty < a < b < \infty \\ \dfrac{1}{b-x} & \text{if} \quad -\infty = a < b < \infty \end{cases} \tag{11}$$

is the derivative of $\psi = \psi_{x_0}$ on S which is independent of $x_0 \in S$. The density (10) is differentiable on S too, with the derivative

$$\dot{f}(x) = \frac{\mathrm{d}f(x)}{\mathrm{d}x} = \dot{g}(\psi(x))(\dot{\psi}(x))^2 + g(\psi(x))\ddot{\psi}(x), \quad x \in S, \tag{12}$$

where $\dot{g}$ is the derivative of g on R and

$$\ddot{\psi}(x) = \frac{\mathrm{d}^2\psi(x)}{\mathrm{d}x^2} = \begin{cases} \left(\dot{\psi}(x)\right)^2 x\dot{\psi}(x) & (a,b) = R \\ \left(\dot{\psi}(x)\right)^2 - \dfrac{2x\dot{\psi}(x)}{(x-a)\,(b-x)} & \text{if } -\infty < a < b < \infty \\ \left(\dot{\psi}(x)\right)^2 & \text{otherwise.} \end{cases} \tag{13}$$

Proof. By definition, $P = Q\psi$ for some $Q \in \mathcal{Q}_R$ and $\psi \in \Theta_S$. Since ψ is strictly monotone and continuous on S, the image $\psi(B)$ of a Lebesgue null set $B \subset S$ is a Lebesgue null set. Therefore any $P \in \mathcal{P}_S$ is absolutely continuous with respect to the Lebesgue measure and its density f satisfies (10). The formula (11) is easily verified and implies $\dot{\psi} > 0$ on S, and even

$$\min_{x\in S} \dot{\psi}(x) = \dot{\psi}\left((a+b)/2\right) = \frac{4(b-a)}{(a+b)^2} > 0 \tag{14}$$

if $-\infty < a < b < \infty$. Consequently, f is positive on S. Since the differentiability of f and formulas (12) and (13) are obvious, one can conclude that P belongs to $\mathcal{Q}_S$. This completes the proof. □

In the next proposition, and in the sequel, we denote by $B + c$ translations of subsets $B \subset R$ by constants $c \in R$, i.e.

$$B + c = \{y + c : y \in B\}.$$

Proposition 2. Let $P \in \mathcal{P}_S$ where $S = (a,b) \subseteq R$. For every $x_0 \in S$ there exists unique $Q = Q_{x_0}$ in $\mathcal{P}_R$ with the property $P = Q_{x_0}\psi_{x_0}$ where ψ_{x_0} is the Johnson function corresponding to x_0. If $P = Q_{x_0}\psi_{x_0}$ then $P = Q_{x_1}\psi_{x_1}$ for some $x_1 \in S$ if and only if every Borel set $B \subset R$ satisfies the relation

$$Q_{x_1}(B) = Q_{x_0}(B + c), \tag{15}$$

where

$$c = \begin{cases} \sinh^{-1}(x - x_0) & \text{if} \quad (a,b) = R \\ \ln \dfrac{x_1 - a}{x_0 - a} & \text{if} \quad -\infty < a < b = \infty \\ \ln \dfrac{(b-x_0)\,(x_1-a)}{(x_0-a)\,(b-x_1)} & \text{if} \quad -\infty < a < b < \infty \\ \ln \dfrac{b-x_0}{b-x_1} & \text{if} \quad -\infty = a < b < \infty. \end{cases} \tag{16}$$

Proof. By Definition 1, for P under consideration there exist $x_0 \in S$ and $Q_{x_0} \in \mathcal{P}_R$ with the property $P = Q_{x_0}\psi_{x_0}$. The equality $Q_{x_1}\psi_{x_1} = Q_{x_0}\psi_{x_0}$ for any given $x_1 \in S$ is equivalent to

$$Q_{x_1} = Q_{x_0}(\psi_{x_0}\psi_{x_1}^{-1}).$$

As is easy to verify from (5) and (6), the composed mapping $\psi_{x_0}\psi_{x_1}^{-1}$ is a translation on the real line by the constant c given by (16),

$$\psi_{x_0}\psi_{x_1}^{-1}(x) = x + c, \quad x \in R.$$

Therefore $P = Q_{x1}\psi_{x_1}$ if and only if (15) holds for c given by (16). This proves the second assertion. The second assertion implies that for every $x_1 \in S$ (including $x_1 = x_0$) there exists unique Q_{x_1} (namely, the c-translated version of Q_{x_0}, defined by (15)) such that $P = Q_{x_1}\psi_{x_1}$, which proves the first assertion. □

Proposition 3. Consider $P \in \mathcal{P}_S$ with a Lebesgue density f. Then the score function $s_P(x) = -\mathrm{d}(\ln f(x))/\mathrm{d}x$ on the support S of P is given by the formula

$$s_P = -\dot{\psi}\frac{\dot{g}(\psi)}{g(\psi)} - \frac{\ddot{\psi}}{\dot{\psi}}, \tag{17}$$

where ψ is any Johnson function from Ψ_S, g is the Lebesgue density of $Q = P\psi^{-1} \in \mathcal{P}_R$, and $\dot{g}$, $\dot{\psi}$, $\ddot{\psi}$ are the derivatives introduced above. The ratio

$$T_P = -\frac{\dot{g}(\psi)}{g(\psi)} \tag{18}$$

does not depend on the particular choice of $\psi \in \Psi_S$, i.e. if $P = Q_{x_0}\psi_{x_0} = Q_{x_1}\psi_{x_1}$ for different $x_0, x_1 \in S$ then, for every $x \in S$,

$$\frac{\dot{g}_{x_0}(\psi_{x_0}(x))}{g_{x_0}(\psi_{x_0}(x))} = \frac{\dot{g}_{x_1}(\psi_{x_1}(x))}{g_{x_1}(\psi_{x_1}(x))}. \tag{19}$$

Further, if $-\infty < a < b < \infty$ then the ratio $\ddot{\psi}/\dot{\psi}$ in decomposition (17) is a score function of a probability distribution supported by $S = (a, b)$, namely

$$-\frac{\ddot{\psi}(x)}{\dot{\psi}(x)} = s_{P_{a,b}}(x) = \frac{b + a - 2x}{(x-a)(b-x)}, \quad x \in (a, b), \tag{20}$$

where $P_{a,b}$ is absolutely continuous on the support (a, b) with the Lebesgue density

$$f_{a,b}(x) = \frac{6(x-a)(b-x)}{(b-a)^3}, \quad x \in (a, b). \tag{21}$$

The equality (20) remains valid also for $-\infty = a < b < \infty$ or $-\infty < a < b = \infty$ if the function $s_{P_{a,b}}(x)$ is extended by continuity to $a = -\infty$ or $b = \infty$, respectively.

Proof. The decomposition (17) follows directly from formulas (10) and (12) for f and $\dot{f}$ by taking the ratio $s_P = -\dot{f}/f$. The second assertion (19) follows from the fact that if c is defined by (16) then

$$\psi_{x_1}(x) = \psi_{x_0}(x) - c, \quad x \in S,$$

and that, by (15),

$$g_{x_1}(x) = g_{x_0}(x + c), \quad x \in S.$$

Indeed, then for every $x \in S$ also $\dot{g}_{x_1}(x) = \dot{g}_{x_0}(x + c)$ so that

$$\frac{\dot{g}_{x_1}(\psi_{x_1}(x))}{g(\psi_{x_1}(x))} = \frac{\dot{g}_{x_1}(\psi_{x_0}(x) - c)}{g_{x_1}(\psi_{x_0}(x) - c)} = \frac{\dot{g}_{x_0}(\psi_{x_0}(x))}{g_{x_0}(\psi_{x_0}(x))}.$$

The third assertion (20) follows from formulas (11) and (13) for $\dot{\psi}(x)$ and $\ddot{\psi}(x)$, and from the easily verifiable fact that the function $f_{a,b}$ defined by (21) is a probability density on the bounded interval $S = (a, b)$. If this interval is unbounded below or above then the validity of (20) for

$$\lim_{a \to -\infty} s_{P_{a,b}} \qquad \text{or} \qquad \lim_{b \to \infty} s_{P_{a,b}}$$

follows again from formulas (11) and (13). □

Note that the distributions $P_{-\infty,b}$ or $P_{a,\infty}$ figuring in the continuous extensions of $s_{P_{a,b}}$ to $a = -\infty$ or $b = \infty$ cannot be defined by a similar extension of the density (21).

If $P \in \mathcal{P}_R$ then the score function on R is

$$s_P = -\frac{\dot{f}}{f}, \tag{22}$$

where f is the Lebesgue density of P. If $P \in \mathcal{P}_S$ where $S = (a, b) \subseteq R$ then Proposition 3 guarantees a canonical decomposition

$$s_P = \dot{\psi} T_P + s_{P_{a,b}} \tag{23}$$

of the score function on the support S where $\dot{\psi}, T_P$ and $s_{P_{a,b}}$ are given by (11), (18) and (20) (with the corresponding limits if $a = -\infty$ or $b = \infty$). We see from (11) and (20) that $\dot{\psi}$ and $s_{P_{a,b}}$ depend only on the support S and not on the density f of P defined on this support. Thus a complete information about the score function s_P is contained in the function T_P. This leads to the following definition.

Definition 2. The *core function* of $P \in \mathcal{P}_S$, $S \subseteq R$, is defined on the support interval S by formula

$$T_P = -\frac{\dot{g}(\psi)}{g(\psi)} \tag{24}$$

where g is the Lebesgue density of $Q = P\psi^{-1}$ on R and ψ is an arbitrary Johnson function defined by (5).

Remark 1. Obviously, the definition (24) agrees with the more concise form used in (9). If $P \in \mathcal{P}_S$ then, as said above, T_P specifies in a simple one-to-one manner the score function s_P for all Johnson distributions $P \in \mathcal{P}_S$, $S \subseteq R$. Since the score function $s_P(x) = -\mathrm{d}(\ln f(x))/\mathrm{d}x$, $x \in S$, uniquely specifies the Lebesgue density f of any Johnson distribution $P \in \mathcal{P}_S$, $S \subseteq R$, the interpretation of T_P as a core function of Johnson distribution $P \in \mathcal{P}_S$ in Definition 2 is fully justified for all $S \subseteq R$.

In the following proposition we study the second moments $E_P(T_P)^2$ of the core functions of Johnson distributions. This proposition refers to the Fisher information defined by (4).

Proposition 4. For all Johnson distributions $P \in \mathcal{P}_S$, $S \subseteq R$

$$E_P(T_P)^2 = I_Q$$

where I_Q is the Fisher information of the parent distribution $Q = P\psi^{-1}$ for arbitrary Johnson function $\psi \in \Psi_S$.

Proof. If $P \in \mathcal{P}_S$ then $P = Q\psi, Q \in \mathcal{Q}_R$ and

$$E_P(T_P)^2 = \int_S \left(\frac{\dot{g}(\psi)}{g(\psi)}\right)^2 \mathrm{d}P = \int_R \left(\frac{\dot{g}(\psi)}{g(\psi)}\right)^2 \mathrm{d}(Q\psi) = \int_R \left(\frac{\dot{g}}{g}\right)^2 \mathrm{d}Q,$$

where the last equality follows from the substitution rule in Lebesgue integrals. □

In Table 1 are listed some Johnson distributions P defined by their densities $f(x)$ for $S = (0,\infty)$ and $S = (0,1)$, densities of their parent distributions and corresponding core functions.

Table 1. Johnson distributions $P = Q\psi_{x_0} \in \mathcal{P}_S, S \neq R$, with densities $f(x)$, parent densities $g = \mathrm{d}Q/\mathrm{d}\lambda$ and corresponding core functions T_Q, T_P.

Name	$f(x), x \in (0,\infty)$	$g(y), y \in R$	Name	$T_Q(y)$	$T_P(x)$
Lognormal	$\frac{1}{\sqrt{2\pi}x}e^{-\frac{1}{2}\ln^2 x}$	$\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}y^2}$	Normal	y	$\ln x$
Exponential	e^{-x}	$e^y e^{-e^y}$	Gumbel	$e^y - 1$	$x-1$
Extr. v. II	$\frac{1}{x^2}e^{-1/x}$	$e^{-y}e^{-e^{-y}}$	Extr. v. I	$1-e^{-y}$	$1-1/x$
Wald-type	$\frac{1}{Kx}e^{-\frac{1}{2}(x+1/x)}$	$\frac{1}{K}e^{-\cosh y}$	no name	$\sinh y$	$\frac{1}{2}(x-1/x)$
Log-logistic	$\frac{1}{(x+1)^2}$	$\frac{e^y}{(e^y+1)^2}$	Logistic	$\frac{e^y-1}{e^y+1}$	$\frac{x-1}{x+1}$
Beta-prime	$\frac{1}{B}\frac{x^{\alpha-1}}{(x+1)^{\alpha+\beta}}$	$\frac{1}{B}\frac{e^{\alpha y}}{(e^y+1)^{\alpha+\beta}}$	no name	$\frac{\alpha e^y-\beta}{e^y+1}$	$\frac{\alpha x-\beta}{x+1}$
Gamma(α,γ)	$\frac{\gamma^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\gamma x}$	$\frac{\gamma^\alpha}{\Gamma(\alpha)}e^{\alpha y}e^{-\gamma e^y}$	no name	$\alpha e^y - \gamma$	$\alpha x - \gamma$
Johnson	$\frac{1_{0,1}(x)}{\sqrt{2\pi}x(1-x)}e^{-\frac{1}{2}\ln^2\frac{x}{1-x}}$	$\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}y^2}$	Normal	y	$\ln\frac{x}{1-x}$
Beta(α,β)	$\frac{1_{0,1}(x)}{B}x^{\alpha-1}(1-x)^{\beta-1}$	$\frac{1}{B}\frac{e^{\alpha y}}{(e^y+1)^{\alpha+\beta}}$	no name	$\frac{\alpha e^y-\beta}{e^y+1}$	$(\alpha+\beta)x-\alpha$

Here $K = 2K_0(1)$ and $B = B(\alpha,\beta)$, where $K_\nu(u)$ is the Bessel function of the third kind and $B(u,v)$ is the beta function.

3. JOHNSON LOCATION

In this section we describe a method leading to parametrized families $\mathcal{P} \subset \mathcal{P}_S$ of Johnson distributions. We show that these families include many important parametric statistical models such as lognormal, Weibull and gamma distributions and therefore they play an important role in statistical applications. Distributions from these families will be used in the next section.

If $Q \in \mathcal{P}_R$ then it is well known that the automorphisms $[\mu,\sigma]$: $R \to R$ defined for $(\mu,\sigma) \in R \times (0,\infty)$ by

$$[\mu,\sigma]\,(y) = \mu + \sigma y, \quad y \in R \tag{25}$$

specify a *location and scale family* $\mathcal{Q} = \{Q_{\mu,\sigma} = Q\,[\mu,\sigma]^{-1} : (\mu,\sigma) \in R \times (0,\infty)\}$. The distribution Q is a *parent* of $\mathcal{Q}$ and it holds $\mathcal{Q} \subset \mathcal{P}_R$ and

$$g_{\mu,\sigma}(y) = \frac{\mathrm{d}Q_{\mu,\sigma}}{\mathrm{d}\lambda}(y) = \frac{1}{\sigma} g\left(\frac{y-\mu}{\sigma}\right), \quad y \in R, \tag{26}$$

for the Lebesgue parent density $g = \mathrm{d}Q/\mathrm{d}\lambda$.

Definition 3. Define for $(\tau,\sigma) \in S \times (0,\infty)$ one-to-one mappings $\{\tau,\sigma\} : S \to R$ by

$$\{\tau,\sigma\} = \psi(\tau) + \sigma\psi(x), \qquad x \in S. \tag{27}$$

Then for any $P \in \mathcal{P}_S$ the mapping (27) defines a Johnson location and scale family

$$\mathcal{P} = \left\{P_{\tau,\sigma} = P\,\{\tau,\sigma\}^{-1} : (\tau,\sigma) \in S \times (0,\infty)\right\} \tag{28}$$

The distribution P is a parent of $\mathcal{P}$ and the parameter $\tau \in S$ is a *Johnson location.*

Proposition 5. All distributions $P_{\tau,\sigma}$ from the above defined Johnson location and scale family $\mathcal{P}$ satisfy the relation

$$P_{\tau,\sigma} = Q_{\psi(\tau),\sigma}\psi, \tag{29}$$

where $\psi \in \Psi_S$ is a Johnson function and $Q_{\psi(\tau),\sigma}$ is the element of the location and scale family $\mathcal{Q}$ with the parent $Q = P\psi^{-1}$. It holds $\mathcal{P} \subset \mathcal{P}_S$ and the Lebesgue densities $f_{\tau,\sigma} = \mathrm{d}P_{\tau,\sigma}/\mathrm{d}\lambda$ of distributions $P_{\tau,\sigma} \in \mathcal{P}$ are given by formulas

$$f_{\tau,\sigma}(x) = \frac{1}{\sigma} g\left(\frac{\psi(x)-\psi(\tau)}{\sigma}\right)\dot{\psi}(x), \quad x \in S, \tag{30}$$

for the density $g = \mathrm{d}Q/\mathrm{d}\lambda$.

Proof. Let $P = Q\psi$ for some $\psi \in \Psi_S$. It is easy to verify from (25) and (27) that $[\psi(\tau), \sigma]^{-1} = \psi\left(\{\tau, \sigma\}^{-1}\right)$. Therefore $P_{\tau,\sigma} = Q\left(\psi\{\tau), \sigma\}^{-1}\right) = Q\,[\psi(\tau), \sigma]^{-1} = Q_{\psi(\tau),\sigma}$, which proves (29). The inclusion $\mathcal{P} \subset \mathcal{P}_S$ follows from the relation $P_{\tau,\sigma} = Q_{\psi(\tau),\sigma}\psi$, where $Q_{\psi(\tau),\sigma} \in \mathcal{P}_R$ by the definition of the Johnson class $\mathcal{P}_S$ in Definition 1. Relation (30) follows from equality $P_{\tau,\sigma} = Q_{\psi(\tau),\sigma}\psi$ and from the fact that

$$g_{\psi(\tau),\sigma}(y) = \frac{1}{\sigma} g\left(\frac{y - \psi(\tau)}{\sigma}\right), \quad y \in R$$

(see (26) for $g = \mathrm{d}Q/\mathrm{d}\lambda$). □

Remark 2. The last proposition implies that $f_{\tau,\sigma}$ is the density of P if (and only if) $(\tau, \sigma) = (\psi^{-1}(0), 1)$. This is a neutral element of the group $S \times (0, \infty)$ under the associative multiplication

$$(\tau, \sigma)(\tilde{\tau}, \tilde{\sigma}) = \left(\psi^{-1}\left[\psi(\tau) + \sigma\psi(\tilde{\tau})\right], \sigma\tilde{\sigma}\right)$$

with the inverse element $(\tau, \sigma)^{-1} = \left(\psi^{-1}(-\psi(\tau)/\sigma), 1/\sigma\right)$. This group structure of S does not define the equivariance structure on the family P in the common sense considered e. g. in Chapter 7 of Zaks [5].

Proposition 6. The core functions $T_{P_{\tau,\sigma}}$ of distributions $P_{\tau,\sigma}$ with density $f_{\tau,\sigma}$ from the above defined Johnson location and scale family P are given by formula

$$T_{P_{\tau,\sigma}}(x) = -\frac{\dot{g}\left(\frac{\psi(x)-\psi(\tau)}{\sigma}\right)}{g\left(\frac{\psi(x)-\psi(\tau)}{\sigma}\right)} = T_Q\left(\frac{\psi(x) - \psi(\tau)}{\sigma}\right), \quad x \in S, \tag{31}$$

where Q and g are the same as in the previous proposition. Moreover, a relation between the efficient score $\frac{\partial}{\partial\tau} \ln f_{\tau,\sigma}(x)$ and the core function (31) is

$$\frac{\partial}{\partial\tau} \ln f_{\tau,\sigma}(x) = \frac{1}{\sigma}\dot{\psi}(\tau)\, T_{P_{\tau,\sigma}}(x).$$

Proof. The first part is clear from (29) and from Definition 2. Put $u = \frac{\psi(x)-\psi(\tau)}{\sigma}$. Since

$$\frac{\partial}{\partial\tau} \ln f_{\tau,\sigma}(x) = \frac{1}{f_{\tau,\sigma}(x)} \frac{\mathrm{d}f_{\tau,\sigma}(x)}{\mathrm{d}u} \frac{\partial u}{\partial\tau}$$

and $f_{\tau,\sigma}(x) = \frac{1}{\sigma} g(u)\dot{\psi}(x)$ by (30), and $\frac{\partial u}{\partial\tau} = \frac{1}{\sigma}\dot{\psi}(\tau)$, it holds by (31) that

$$\frac{\partial}{\partial\tau} \ln f_{\tau,\sigma}(x) = \frac{1}{\sigma}\dot{\psi}(\tau) T_Q(u) = \frac{1}{\sigma}\dot{\psi}(\tau) T_{P_{\tau,\sigma}}(x).$$ □

In Table 2 are listed densities and core functions of some Johnson location and scale families and their parents.

Table 2. Johnson families $P_{\tau,\sigma}$ with densities $f_{\tau,\sigma}(x)$, parent densities $f(x)$ and corresponding core functions T_P and $T_{P_{\tau,\sigma}}$. The scale σ is reparametrized by $\beta = 1/\sigma$.

Name	$f_{\tau,\sigma}(x), x \in (0,\infty)$	$f(x), x \in (0,\infty)$	T_P	$T_{P_{\tau,\sigma}}$
Lognormal	$\frac{\beta}{\sqrt{2\pi}x} e^{-\frac{1}{2}\log^2(\frac{x}{\tau})^\beta}$	$\frac{1}{\sqrt{2\pi}x} e^{-\frac{1}{2}\ln^2 x}$	$\ln x$	$\ln(\frac{x}{\tau})^\beta$
Weibull	$\frac{\beta}{x}(\frac{x}{\tau})^\beta e^{-(\frac{x}{\tau})^\beta}$	$\frac{1}{x} x e^{-x}$	$x-1$	$(\frac{x}{\tau})^\beta - 1$
Extr. v. II	$\frac{\beta}{x}(\frac{x}{\tau})^{-\beta} e^{-(\frac{x}{\tau})^{-\beta}}$	$\frac{1}{x^2} e^{-1/x}$	$1-1/x$	$1-(\frac{x}{\tau})^{-\beta}$
Wald-type	$\frac{1}{Kx} e^{-\frac{1}{2}[(\frac{x}{\tau})^\beta + (\frac{x}{\tau})^{-\beta}]}$	$\frac{1}{Kx} e^{-\frac{1}{2}(x+1/x)}$	$\frac{1}{2}(x-1/x)$	$\frac{1}{2}[(\frac{x}{\tau})^\beta - (\frac{x}{\tau})^{-\beta}]$
Log-logistic	$\frac{\beta}{x}\frac{(\frac{x}{\tau})^\beta}{(1+(\frac{x}{\tau})^\beta)^2}$	$\frac{1}{(x+1)^2}$	$\frac{x-1}{x+1}$	$\frac{(\frac{x}{\tau})^\beta - 1}{(\frac{x}{\tau})^\beta + 1}$
Gamma(α)	$\frac{\beta\alpha^\alpha}{\Gamma(\alpha)x}(\frac{x}{\tau})^{\beta\alpha} e^{-\alpha(\frac{x}{\tau})^\beta}$	$\frac{\alpha^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\alpha x}$	$\alpha(x-1)$	$\alpha\left[(\frac{x}{\tau})^\beta - 1\right]$

Formulas for the Gamma(α) distribution follows from the relation Gamma(α,γ) = $\frac{\gamma^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\alpha x} = \frac{\alpha^\alpha}{\Gamma(\alpha)x}\left(\frac{x}{\tau}\right)^\alpha e^{-\alpha\frac{x}{\tau}} =$ Gamma(α) where $\tau = \gamma/\alpha$.

Proposition 6 shows the significance of the core function for statistical inference. The core function is the inner part of the efficient score for the Johnson location parameter and appears to be the most important characteristic of distributions $P \in \mathcal{P}_S$. As is shown below, it also provides an interesting characterization of divergence in parametric families which is analogous but often simpler than the Kullback divergence.

4. CORE DIVERGENCE

By Remark 1 in the previous section, the core functions T_P characterize in a one-to-one manner all distributions $P \in \mathcal{P}_S, S = (a,b) \subseteq R$. Therefore any measure of divergence (dissimilarity, or distance if the metric axioms hold) in the space $\mathcal{T}_S = \{T_P : P \in \mathcal{P}_S\}$ will serve as a measure of divergence in the space $\mathcal{P}_S$ itself.

The most natural of the distances between measurable functions $T_P, T_{\tilde{P}}$ defined on S is the common L_2-norm $|T_P - T_{\tilde{P}}| = (\int_S (T_P - T_{\tilde{P}})^2 \mathrm{d}\mu)^{1/2}$ where the integral is taken with respect to a measure μ defined on Borel subsets of S. To achieve a better comparability with the asymmetric Kullback divergence

$$K(P,\tilde{P}) = \int_S \ln(\mathrm{d}P/\mathrm{d}\tilde{P})\,\mathrm{d}P, \qquad P,\tilde{P} \in \mathcal{P},$$

we propose to take $\mu = P$ and to normalize the resulting norm by $\|T_P\| = (\int_S T_P^2\,\mathrm{d}P)^{1/2}$ $= (I_P)^{1/2}$, where I_P is the Fisher information of P, see Proposition 3 above. This motivates the following definition, as well as the fact that, in order to avoid undefined expressions, we restrict ourselves to the subspaces

$$\mathcal{P}_S^0 = \{P \in \mathcal{P}_S : 0 < I_P < \infty\} \quad \text{and} \quad \mathcal{T}_S = \{T_P : P \in \mathcal{P}_S^0\}. \tag{32}$$

Definition 4. For every $S = (a, b) \subseteq R$, the core divergence $D(P, \tilde{P})$ of ordered pairs $P, \tilde{P}$ of distributions from $\mathcal{P}_S^0$ is defined as a divergence of the corresponding core functions $T_P, T_{\tilde{P}} \in \mathcal{T}_S$, namely

$$D(P, \tilde{P}) = \frac{1}{2I_P} \int_S (T_P - T_{\tilde{P}})^2 \mathrm{d}P. \tag{33}$$

Remark 3. Since $\int (T_P - T_{\tilde{P}})^2 \,\mathrm{d}P$ may be infinite when $P \neq \tilde{P}$, the core divergence takes on in general the values from the extended real line interval $[0, \infty]$.

The following proposition simplifies evaluation of the core divergence. It also implies (cf. Proposition 8) that the core divergence is a squared distance in the important Johnson location families with fixed scales.

Proposition 7. If $S = (a, b) \neq R$ then for every $P, \tilde{P} \in \mathcal{P}_S^0$ and $\psi \in \Psi_S$ it holds

$$D(P, \tilde{P}) = D(Q, \tilde{Q}), \tag{34}$$

where $Q, \tilde{Q}$ are elements of $\mathcal{P}_R^0$ defined by $Q = P\psi^{-1}$ and $\tilde{Q} = \tilde{P}\psi^{-1}$.

Proof. By Proposition 4, $I_P = I_Q$ for $Q = P\psi^{-1}$. If $\tilde{Q} = \tilde{P}\psi^{-1}$ then (24) implies that $T_P = T_Q(\psi)$ and $T_{\tilde{P}} = T_{\tilde{Q}}(\psi)$. Therefore, by the substitution rule for integrals,

$$\int_S (T_P - T_{\tilde{P}})^2 \mathrm{d}P = \int_S (T_Q(\psi) - T_{\tilde{Q}}(\psi))^2 \,\mathrm{d}Q\psi = \int_R (T_Q - T_{\tilde{Q}})^2 \,\mathrm{d}Q$$

which completes the proof. □

In the following assertion we consider the location and scale families $\mathcal{Q}$ and $\mathcal{P}$ with respective parents Q and P defined in Section 3.

Proposition 8. If $Q \in \mathcal{P}_R^0$ and $\mathcal{Q}$ is the location and scale family with parent Q, then $\mathcal{Q} \subset \mathcal{P}_R^0$ and for every $Q_{\mu,\sigma}$ and $Q_{\tilde{\mu},\tilde{\sigma}}$ from $\mathcal{Q}$

$$D(Q_{\mu,\sigma}, Q_{\tilde{\mu},\tilde{\sigma}}) = \frac{1}{2I_Q} \int_R \left(T_Q(y) - T_Q \left(\frac{\sigma}{\tilde{\sigma}} y + \frac{\mu - \tilde{\mu}}{\tilde{\sigma}} \right) \right)^2 \mathrm{d}Q(y). \tag{35}$$

If $S \neq R$, and $P = Q\psi^{-1}$ for some $P \in \mathcal{P}_S^0, Q \subset \mathcal{P}_R^0$ and $\psi \in \Psi_S$, and if $\mathcal{P}$ is the Johnson location and scale family with parent P, then $\mathcal{P} \subset \mathcal{P}_S^0$ and for every $P_{\tau,\sigma}$ and $P_{\tilde{\tau},\tilde{\sigma}}$ from $\mathcal{P}$

$$D(P_{\tau,\sigma}, P_{\tilde{\tau},\tilde{\sigma}}) = \frac{1}{2I_Q} \int_S \left(T_Q(y) - T_Q \left(\frac{\sigma}{\tilde{\sigma}} y + \frac{\psi(\tau) - \psi(\tilde{\tau})}{\tilde{\sigma}} \right) \right)^2 \mathrm{d}Q(y). \tag{36}$$

Proof. We shall prove (36). Proof of (35) is simpler. Fix arbitrary (τ,σ) and $(\tilde\tau,\tilde\sigma)$ from $S \times (0,\infty)$. By (29) and Proposition 7,

$$\begin{aligned} D(P_{\tau,\sigma}, P_{\tilde\tau,\tilde\sigma}) &= D(Q_{\psi(\tau),\sigma}, Q_{\psi(\tilde\tau),\tilde\sigma}) \\ &= \frac{1}{2I_Q}\int_R (T_{Q_{\psi(\tau),\sigma}} - T_{Q_{\psi(\tilde\tau),\tilde\sigma}})^2 \, \mathrm{d}Q_{\psi(\tau),\sigma}. \end{aligned}$$

The assumption $Q \in \mathcal{P}_R^0$ together with Proposition 3 implies that $I_P = I_Q$ is finite and nonzero. Further, (31) and (26) imply the relation

$$T_{Q_{\psi(\tau),\sigma}}(u) = -\frac{\dot g(\frac{u-\psi(\tau)}{\sigma})}{g(\frac{u-\psi(\tau)}{\sigma})} = T_Q\left(\frac{u-\psi(\tau)}{\sigma}\right), \quad u \in R, \tag{37}$$

so that it suffices to apply in the last integral the substitution $y = (u-\psi(\tau))/\sigma$ to get the desired equality (35). □

In Tables 3 – 5, we compare formulas of Kullback divergences and core divergences in some Johnson families. In these tables, C is the Euler constant, $\psi(u) = \Gamma'(u)/\Gamma(u)$ is the psi function, and for the Bessel function of the third kind $K_\nu(u)$ it holds $\alpha = K_2(1)/K_0(1) - 1 \doteq 2.68$, $C_1 = \frac{1}{2}K_1(1)/K_0(1) \doteq 0.72$, $C_2 = K_2(1)/(4K_0(1)\alpha) \doteq 0.34$, $C_3 = 1/2\alpha \doteq 0.174$.

Table 3. Kullback divergences and core divergences in some Johnson families from Table 2, reparametrized by $\omega = \sigma/\tilde\sigma = \tilde\beta/\beta$ and $\gamma = (\tau/\tilde\tau)^{1/\tilde\sigma}$.

Name	$K(P_{\tau,\sigma}, P_{\tilde\tau,\tilde\sigma})$	$D(P_{\tau,\sigma}, P_{\tilde\tau,\tilde\sigma})$
Lognormal	$\frac{1}{2}[-\ln\omega^2 + (\ln\gamma)^2 + \omega^2 - 1]$	$\frac{1}{2}[(\ln\gamma)^2 + (\omega-1)^2]$
Weibull	$\gamma\Gamma(\omega+1) - \ln(\omega\gamma) + (\omega-1)C - 1$	$\gamma^2\Gamma(2\omega) - \gamma\Gamma(\omega+2) + 1$
Extr. v. II	$\frac{1}{\gamma}\Gamma(\omega+1) - \ln\frac{\omega}{\gamma} + (\omega-1)C - 1$	$\frac{1}{\gamma^2}\Gamma(2\omega) - \frac{1}{\gamma}\Gamma(\omega+2) + 1$

Table 4. Similar as in Table 4 for the special case $\sigma = \tilde\sigma = 1$ and also different Johnson families.

Name	$K(P_{\tau,1}, P_{\tilde\tau,1})$	$D(P_{\tau,1}, P_{\tilde\tau,1})$
Lognormal	$\frac{1}{2}(\ln\gamma)^2$	$\frac{1}{2}(\ln\gamma)^2$
Weibull	$\gamma - \ln\gamma - 1$	$(\gamma-1)^2$
Wald-type	$4C_1(\gamma + 1/\gamma - 2)$	$C_2[(\gamma-1)^2 + (1-1/\gamma)^2]$ $+C_3(\gamma+1/\gamma-2)/$
Log-logistic	$(\gamma+1)\ln\gamma/(\gamma-1) - 2$	$2[(\gamma-1)(\gamma^2+10\gamma+1)$ $-6\gamma(\gamma+1)\ln\gamma]/(\gamma-1)^3$
Gamma(α)	$-\tilde\alpha\ln\tilde\alpha - (\tilde\alpha-\alpha)[\psi(\alpha) - \ln\alpha]$ $+\alpha\ln\alpha + \ln\frac{\Gamma(\tilde\alpha)}{\Gamma(\alpha)}$ $-\tilde\alpha\ln\gamma + \tilde\alpha\gamma - \alpha$	$\frac{1}{2}\left[\frac{\tilde\alpha^2}{\alpha^2}(\gamma-1)^2\alpha + (\frac{\tilde\alpha}{\alpha}\gamma-1)^2\right]$

Table 5. Kullback divergences and core divergences in some special cases of gamma and beta distributions.

Name	$K(P_\alpha, P_{\tilde{\alpha}})$	$D(P_\alpha, P_{\tilde{\alpha}})$
Gamma(α)	$-\tilde{\alpha}\ln\tilde{\alpha} + \alpha\ln\alpha$	
($\tau = 1$)	$-(\tilde{\alpha} - \alpha)[\psi(\alpha) - \ln\alpha - 1]$	$\frac{1}{2}(\frac{\tilde{\alpha}}{\alpha} - 1)^2$
Beta ($\alpha, 1$)	$\frac{\tilde{\alpha}}{\alpha} - \ln\frac{\tilde{\alpha}}{\alpha} - 1$	$\frac{\alpha}{\alpha+1}\left(\frac{\tilde{\alpha}}{\alpha} - 1\right)^2$
Beta (α, α)	$-\ln\frac{\Gamma(2\tilde{\alpha})}{\Gamma(2\alpha)} + 2\ln\frac{\Gamma(\tilde{\alpha})}{\Gamma(\alpha)}$	
	$+2\lvert\tilde{\alpha} - \alpha\rvert(\psi(2\alpha) - \psi(\alpha))$	$\frac{1}{2}(\tilde{\alpha} - \alpha)^2$

By Proposition 7, if $Q_{\mu,\sigma} = P_{\tau,\sigma}\psi^{-1}$ where $\mu = \psi^{-1}(\tau)$, it holds $D(Q_{\mu,\sigma}, Q_{\tilde{\mu},\tilde{\sigma}}) = D(P_{\tau,\sigma}, P_{\tilde{\tau},\tilde{\sigma}})$. In the case of the normal distribution, for example, one can use the formulas for the lognormal distribution with $\ln\gamma = (\mu - \tilde{\mu})/\tilde{\sigma}$.

Comparison of Kullback (K) and core (D) divergences for logistic distributions with different location parameters is given in Figure 1. Similar comparison for Weibull (exponential) distributions with different Johnson locations is given in Figure 2. This figures are typical in the sense that for small deviation of parameters both divergences almost coincide, but they differ in the sensitivity to large deviations of the parameters. Figure 1 illustrates that for distributions with heavy tails, the core divergence is much less sensitive to large deviations of the parameters than the Kullback divergence.

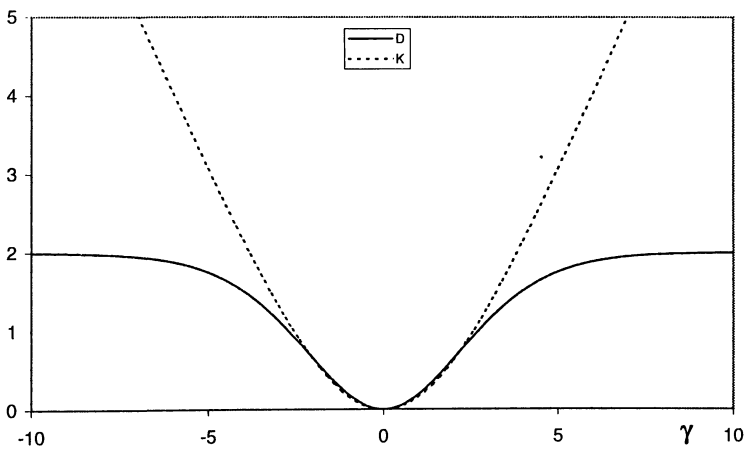

Fig. 1. $K(Q_{\mu,\sigma}, Q_{\hat{\mu},\sigma})$ (dotted line) and $D(Q_{\mu,\sigma}, Q_{\hat{\mu},\sigma})$ (full line) of logistic distributions as functions of $\gamma = \exp((\mu - \hat{\mu})/\sigma)$.

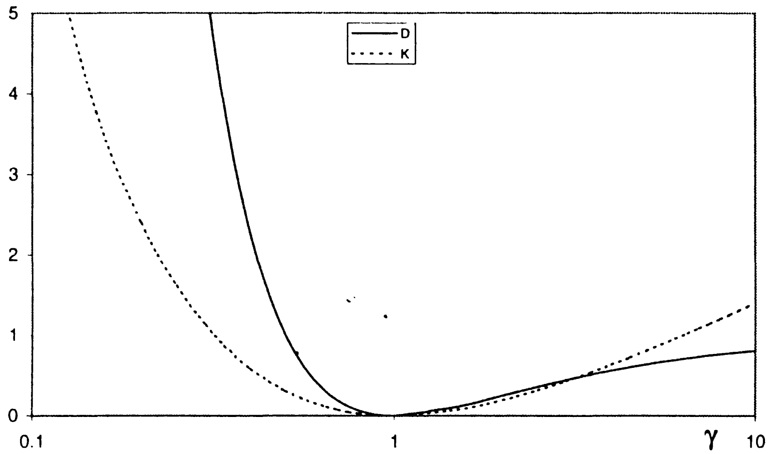

Fig. 2. $K(Q_{\mu,\sigma}, Q_{\hat{\mu},\sigma})$ (dotted line) and $D(Q_{\mu,\sigma}, Q_{\hat{\mu},\sigma})$ (full line) of extreme value II distributions as functions of $\gamma = (1/\hat{\tau})^{1/\sigma}$.

ACKNOWLEDGEMENT

This research was supported by the Grant Agency of the Academy of Science of the Czech Republic through Grant A1075101.

(Received December 14, 2001.)

REFERENCES

[1] Z. Fabián: Information and entropy of continuous random variables. IEEE Trans. Inform. Theory *43* (1997), 1080–1083.

[2] Z. Fabián: Induced cores and their use in robust parametric estimation. Commun. Statist. A – Theory Methods *30* (2001), 3, 537–556.

[3] N. L. Johnson: Systems of frequency curves generated by methods of translations. Biometrika *36* (1949), 149–176.

[4] N. L. Johnson and S. Kotz: Continuous Univariate Distributions 1, 2. Houghton Mifflin, Boston 1970.

[5] S. Zacks: The Theory of Statistical Inference. Wiley, New York 1971.

RNDr. Zdeněk Fabián, CSc., Institute of Computer Science – Academy of Sciences of the Czech Republic, Pod vodárenskou věží 2, 182 07 Praha 8. Czech Republic.
e-mail: zdenek@cs.cas.cz

Ing. Igor Vajda, DrSc., Institute of Information Theory and Automation – Academy of Sciences of the Czech Republic, Pod vodárenskou věží 4, 182 08 Praha 8. Czech Republic.
e-mail: vajda@utia.cas.cz