Center of Gravity and Information of Continuous Distribution

Fabián, Zdeněk
2002

Institute of Computer Science
Academy of Sciences of the Czech Republic

Center of Gravity and information of continuous distribution

Zdeněk Fabián

Technical report No. 870

July 2002

Pod Vodárenskou věží 2, 182 07 Prague 8, phone: 6605 3210, fax: (+4202) 86 58 57 89,
e-mail:zdenek@cs.cas.cz

Institute of Computer Science
Academy of Sciences of the Czech Republic

Center of Gravity and information of continuous distribution

Zdeněk Fabián

Technical report No. 870

July 2002

Abstract:

The recently proposed core function of absolutely continuous distributions are re-introduced and their moments are proved to exist. The core moments can be used as numerical characteristics of distributions. Particularly, it is shown that the first core moment defines a point (we call it the center of gravity) charakterizing the central tendency of the distribution besides the mode, mean and median. The core function can be interpreted as the likelihood score for the center of gravity. The second core moment expresses the information of the distribution.

Keywords:
Core function; Moments; Numerical characteristics of distributions; Estimation.

CENTER OF GRAVITY AND INFORMATION OF CONTINUOUS DISTRIBUTION

Zdeněk Fabián

Institute of Computer Science, Academy of Sciences of the Czech republic, Pod vodárenskou věž 2, 18200 Prague, zdenek@cs.cas.cz

ABSTRACT

The recently proposed core function of absolutely continuous distributions are re-introduced and their moments are proved to exist. The core moments can be used as numerical characteristics of distributions. Particularly, it is shown that the first core moment defines a point (we call it the center of gravity) characterizing the central tendency of the distribution besides the mode, mean and median. The core function can be interpreted as the likelihood score for the center of gravity. The second core moment expresses the information of the distribution.

Key words: Core function; Moments; Numerical characteristics of distributions; Estimation.

1. INTRODUCTION

Let F be the distribution function and f the density of a continuous random variable X supported by interval $S \subseteq R$. The commonly accepted numerical characteristics of distributions are the absolute moments

$$m_k = E(X^k) = \int_S x^k f(x)\, dx, \qquad k = 1, 2... \tag{1}$$

and the central moments $\mu_k = E(X - m_1)^k$. The integrals may not converge, however, even for regular distribution.

In Table 1 we present the densities of three parametric families $\{f_\theta, \theta \in \Theta\}$ where $\Theta = \{(\beta, \alpha) : \beta, \alpha \in (0, \infty)\}$ and the values of their absolute moments $m_k(\theta)$ together with the ranges of the values of the parameters which they exist for.

Table 1. Absolute Moments of Three Parametric Families

F_θ	f_θ	$m_k(\theta)$	exist for
1	$\frac{\beta\alpha^\alpha}{\Gamma(\alpha)} x^{-\beta\alpha-1} e^{-\alpha x^{-\beta}}$	$\alpha^{k/\beta} \frac{\Gamma(\alpha-k/\beta)}{\Gamma(\alpha)}$	$k < \beta\alpha$
2	$\frac{\beta\Gamma(2\alpha)}{\Gamma^2(\alpha)} \frac{x^{\beta\alpha-1}}{(x^\beta+1)^{2\alpha}}$	$\frac{\Gamma(\alpha+k/\beta)\Gamma(\alpha-k/\beta)}{\Gamma^2(\alpha)}$	$k < \beta\alpha$
3	$\frac{\beta}{B(1/2,\alpha-1/2)} \frac{1}{(1+\beta \ln^2 x)^\alpha}$	$\frac{\Gamma(k+1/2)\Gamma(\alpha-k-1/2)}{\Gamma(1/2)\Gamma(k-1/2)}$	$k < \alpha - \frac{1}{2}$

Parents of families (with parameters $\alpha = \beta = 1$) are the standard extreme value II (1), standard log-logistic (2) or standard log-Cauchy (3) distributions, respectively. For these simple distributions there is even no mean. The same is true in the case of *symmetric* Cauchy distribution. Table 1 illustrates the fact that the traditional moments are of no use for both characterizing distributions and for estimation of parameters not only in some exceptional cases but for large regular families. The infiniteness of moments is rather mysterious, since the aim of them is to characterize densities with a unit area under the curve. A straightforward consequence is that there is no common agreement in statistic which is the point characterizing central tendency of a distribution: the mean, mode or perhaps median ?

A common opinion in the information theory says that the information of continuous distributions is closely connected with continuous analogy of the Shannon entropy, the differential entropy

$$h(F) = \int_S -\ln f(x)\, f(x)\, dx, \tag{2}$$

which is sometimes interpreted as an expected value $E(-\ln f(x))$ [Cover and Thomas (1990), p. 13]. Actually, (2) has some properties that agree with the intuitive notion of what a measure of information should be, but it has a serious defect; unlike the discrete entropy it can be negative. The often quoted example of a distribution with negative differential entropy is the uniform distribution with $||S|| < 1$, but it should be said that (2) can be negative in the case of *any* parametric distribution with a sufficiently small scale parameter. On the other hand, in mathematical statistics it is the Fisher information that is widely used. For parametric distribution $F_\theta, \theta = (\theta_1, ..., \theta_m)$, the Fisher information about parameter $\theta_k, 1 \le k \le m$ is given by

$$J_{kk}(\theta) = \int_S \left(\frac{\partial \ln f_\theta(x)}{\partial \theta_k}\right)^2 f_\theta(x)\, dx \tag{3}$$

and represents the mean information about θ_k carried by one observation taken from F_θ. However, there is no common agreement about which value represents the information of a distribution without parameters.

Recently, a core function of an absolutely continuous probability distribution has been introduced by Fabián (2001). Although it had not been known in its general form, it was known in many particular cases as a simple and relevant characteristic of distributions. In Section 2 of the present paper, the core function is briefly re-introduced in a somewhat different way and in slightly different terms from that in the cited paper. In Section 3, moments of the core function have been proved to exist for any regular distribution. Apparently, it is the core moments and not the absolute or central moments that are to be used to characterize distributions. Particularly, in Section 4 we study the first two core moments and show that the first core moment defines a point generally characterizing the central tendency of the distribution and the second core moment expresses the mean information of distributions including distributions without parameters. Methods of an estimation of these new characteristics of distributions are discussed in Section 5.

2. CORE FUNCTION

Let for every $\emptyset \neq S = (a,b) \subseteq R$, Π_S be a class of distributions on Borel sets of real line R absolutely continuous with respect to the Lebesgue measure λ on R. Let X be a random variable with distribution $P \in \Pi_S$, distribution function F and density $f = dF/d\lambda$ satisfying the relation

$$f(x) = \begin{cases} > 0 & \text{if} \quad x \in S \\ = 0 & \text{if} \quad x \in R - S. \end{cases} \tag{4}$$

Let the derivative $f'(x) = df(x)/dx$ be well-defined. f is supposed to be regular in the sense of Hájek and Šidák, which means that the integral

$$I_F = \int_S \left(\frac{f'(x)}{f(x)} \right)^2 f(x)\, dx \tag{5}$$

is finite and positive [cf. van der Waart (1998)]. The open interval S is the support of distribution F and the sample space of random variable X.

2.1 Core Function of Simple Distributions

Let Y be a random variable with distribution $Q \in \Pi_R$, distribution function G and density g. The *core function* of random variable Y or of distribution Q with "full support" $S = R$ is the score function

$$T_G(y) = -\frac{g'(y)}{g(y)}. \tag{6}$$

Let us select some simple real functions defined on $S = R$ with different behaviour in infinity: unbounded functions (denoted in the sequel by U1 and U2), bounded functions (B1 and B2) and the mixed types (UB and BU). Considering them to be the core functions of some distributions, we computed the corresponding densities by the use of (6). The results, called here *simple distributions,* are given in Table 2.

Table 2. Core Functions and Densities of Simple Distributions

Type	$T_G(y)$	$g(y)$	Distribution
$U1$	$\sinh y$	$\frac{1}{2K_0(1)} e^{-\cosh y}$	no name
$U2$	y	$\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}$	normal
UB	$1 - e^{-y}$	$e^{-y} e^{-e^{-y}}$	extreme value I
BU	$e^y - 1$	$e^y e^{-e^y}$	Gumbel
$B1$	$\tanh \frac{y}{2}$	$\frac{e^y}{(1+e^y)^2}$	logistic
$B2$	$\frac{2y}{1+y^2}$	$\frac{1}{\pi(1+y^2)}$	Cauchy

In the table, K_0 denotes the Bessel function of the third kind. All densities except U1 are the standardized forms of densities of well-known distributions.

2.2 Core Function of Composite Distributions

Let $\Phi_S = \{\varphi : R \to S\}$ be a set of continuous bijective mappings. Let us speak for simplicity about distributions F and G instead of 'P and Q with distribution functions F and G'.

Definition 1. Any distribution $F \in \Pi_S$ which is expressed in a form $F = G\varphi^{-1}$ where $G \in \Pi_R$ and $\varphi \in \Phi_S$ will be called a *composite distribution.* Distribution G will be called its *prototype.*

Denoting $\psi = \varphi^{-1}$, the density f of the composite distribution $F = G\psi$ is obviously given by formula

$$f(x) = g(\psi(x))\ \psi'(x). \tag{7}$$

The following definition is equivalent to Definition 2 in Fabián (2001).

Definition 2. Core function of a composite distribution $F = G\psi$ is

$$T_F(x) = T_G(\psi(x)) = -\frac{g'(\psi(x))}{g(\psi(x))}. \tag{8}$$

The general forms of core functions and densities of composite distributions F with prototypes in Table 2 are given in Table 3.

Table 3. Core Functions and Densities of Composite Distributions

Type	$T_F(x)$	$f(x)$
$U1$	$\frac{1}{2}(e^{\psi(x)} - e^{-\psi(x)})$	$\frac{1}{2K_0(1)}e^{-\frac{1}{2}(e^{\psi(x)}+e^{-\psi(x)})}\ \psi'(x)$
$U2$	$\psi(x)$	$\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\psi^2(x)}\ \psi'(x)$
UB	$1 - e^{-\psi(x)}$	$e^{-\psi(x)}e^{-e^{-\psi(x)}}\ \psi'(x)$
BU	$e^{\psi(x)} - 1$	$e^{\psi(x)}e^{-e^{\psi(x)}}\ \psi'(x)$
$B1$	$\frac{e^{\psi(x)}-1}{e^{\psi(x)}+1}$	$\frac{e^{\psi(x)}}{(1+e^{\psi(x)})^2}\ \psi'(x)$
$B2$	$\frac{2\psi(x)}{1+\psi^2(x)}$	$\frac{1}{\pi(1+\psi^2(x))}\ \psi'(x)$

Since $\psi(\cdot)$ is a continuous monotonous mapping $\psi : S \to R$, the types of both core functions and densities of composite distributions remain the same as the types of their prototypes.

According to the following proposition, the core function of a composite distribution F can be expressed without referring to its prototype by means of f and ψ only.

Proposition 1. [Fabián (2001)]. Let $F = G\psi \in \Pi_S$. Denote by f its density and by T_F its core function. It holds

$$T_F(x) = \frac{1}{f(x)}\frac{d}{dx}\left(-\frac{1}{\psi'(x)}f(x)\right). \tag{9}$$

Proof. Set $u = \psi(x)$. By (8) and (7)

$$T_F(x) = -\frac{1}{g(u)}\frac{dg(u)}{du} = -\frac{\psi'(x)}{f(x)}\frac{d}{dx}\left(-\frac{f(x)}{\psi'(x)}\right)\frac{dx}{du}$$

and $du/dx = \psi'(x)$. □

2.3 Core Function of Parametric Distributions

Let $\mu \in R$ and $\sigma \in (0, \infty)$ be the usual location and scale parameters. Random variable Y with distribution $G \in \Pi_R$ induces a family

$$Y_{(\mu,\sigma)} = \mu + \sigma Y$$

with density $g_{\mu,\sigma}(y) = \sigma^{-1} g(u)$ and core function $T_{G_{\mu,\sigma}}(y) = T_G(u)$ where

$$u = \frac{y - \mu}{\sigma}. \tag{10}$$

(10) will be called a *location and scale kernel.*

Thanks to properties of $\psi(\cdot)$ there exist a random variable $X_{(\mu,\sigma)}$,

$$X_{(\mu,\sigma)} = \psi^{-1}(Y_{(\mu,\sigma)}).$$

Obviously $X = X_{(0,1)}$.

Definition 3. Let $F = G\psi$ and μ be the location of G. Parameter τ of distribution F given by

$$\tau = \psi^{-1}(\mu) \tag{11}$$

will be called a *transformed location.*

Using τ, we can write family $X_{\mu,\sigma}$ as

$$X_{(\tau,\sigma)} = X_{(\psi(\tau),\sigma)} = \psi^{-1}(\psi(\tau) + \sigma\psi(X))$$

density and core function of which are

$$f_{\tau,\sigma}(x) = f_{\psi(\tau),\sigma}(x) = g_{\mu,\sigma}(y)\psi'(x) = \sigma^{-1} g(u)\psi'(x), \tag{12}$$

$$T_{F_{\tau,\sigma}}(x) = T_{F_{\psi(\tau),\sigma}}(x) = T_{G_{\mu,\sigma}}(y) = T_G(u), \tag{13}$$

where u is given by

$$u = \frac{\psi(x) - \psi(\tau)}{\sigma}. \tag{14}$$

In the case of a composite distribution, (14) is an equivalent of the location and scale kernel (10). It will be called a *transformed location and scale kernel.* It should be perhaps noted that transformed location τ is not a location parameter in the usual sense [cf. Bickel and Lehmann (1975)].

Example 1. Gamma distribution has density

$$f_{\gamma,\alpha}(x) = \frac{\gamma^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\gamma x}.$$

Reparametrizing it, one obtains

$$f_{\tau,\alpha}(x) = \frac{\alpha^\alpha}{\Gamma(\alpha)} \left(\frac{x}{\tau}\right)^\alpha e^{-\alpha \frac{x}{\tau}} \cdot \frac{1}{x}$$

where $\tau = \alpha/\gamma$ is the transformed location parameter and α is a shape parameter. It holds that $f_{\tau,\alpha}(x) = g_{\mu,\alpha}(\psi(x))\psi'(x)$ where $\psi(x) = \ln x$ and where

$$g_{\mu,\alpha}(y) = \frac{\alpha^\alpha}{\Gamma(\alpha)} e^{\alpha(x-\mu)} e^{-\alpha e^{x-\mu}}$$

is the density of the prototype distribution.

In Table 4 there are explicit forms of some mappings $\psi : S \to R$ and corresponding transformed location and scale kernels (14).

Table 4. Transformations $\psi : S \to R$ and the Forms of Corresponding Kernels u

S	$\psi(x)$	$\psi'(x)$	u
R	x	1	$\frac{x-\mu}{\sigma}$
R	$\sinh x$	$\cosh x$	$\frac{2}{\sigma} \sinh \frac{x-\tau}{2} \cosh \frac{x+\tau}{2}$
$(0, \infty)$	$\ln x$	$\frac{1}{x}$	$\ln \left(\frac{x}{\tau}\right)^{1/\sigma}$
$(0, 1)$	$\ln \frac{x}{1-x}$	$\frac{1}{x(1-x)}$	$\ln \left(\frac{x(1-\tau)}{(1-x)\tau}\right)^{1/\sigma}$
$(0, 1)$	$-\ln(-\ln x)$	$\frac{-1}{x \ln x}$	$\ln \left(\frac{\ln \tau}{\ln x}\right)^{1/\sigma}$
$(-1, 1)$	$\tanh^{-1} x$	$\frac{1}{1-x^2}$	$\frac{1}{\sigma} \tanh^{-1} \frac{x-\tau}{1-x\tau}$
$(-1, 1)$	$\tan \frac{\pi}{2} x$	$\frac{\pi}{2} (\cos \frac{\pi}{2} x)^{-2}$	$\frac{1}{\sigma} \frac{\sin \frac{\pi}{2}(x-\tau)}{\cos \frac{\pi}{2} x \cos \frac{\pi}{2} \tau}$

Transformations in 2-4 rows in Table 4 were originally introduced by Johnson (1949) and generalized for arbitrary support by Fabián (1997, 2001).

In the sequel, $(\Pi_S)_\psi$ will denote the set of distributions with support S originating from the distributions of set Π_R by a mapping $\psi^{-1} : R \to S$. For the sake of simplicity we introduce parameter t by relation

$$t = \begin{cases} \mu & \text{for simple distribution } G_{\mu,\sigma} \\ \tau & \text{for composite distribution } F_{\tau,\sigma} = G_{\mu,\sigma}\psi. \end{cases} \tag{15}$$

Finally, we note that distribution $F_{t,\sigma}$ can be further provided by other (shape) parameters $\mathbf{c} \in (0, \infty)^{m-2}$ to obtain general family F_θ with $\theta \in \Theta$ where

$$\Theta = \{(t, \sigma, \mathbf{c}) : t \in S, \sigma \in (0, \infty), \mathbf{c} \in (0, \infty)^{m-2}\}.$$

If we speak somewhere about a distribution $F_{t,\sigma}$ with parent F, we suppose that the parent contains, if necessary, the shape parameters.

2.4 Basic Theorem

Consider a distribution F_θ. Let us remind that likelihood score of F_θ for parameter $\gamma \in \theta$ is $s_\gamma(x) = \frac{\partial}{\partial \gamma} \ln f_\theta(x)$. Theorem 1 (Fabián 2001) creates a basis for an interpretation of the core function.

Theorem 1. Let $F_{t,\sigma} \in \Pi_S$. It holds

$$\frac{\partial}{\partial t} \ln f_{t,\sigma}(x) = \frac{1}{\sigma} \psi'(t)\ T_{F_{t,\sigma}}(x).$$

Proof. Since

$$\frac{\partial}{\partial t} \ln f_{t,\sigma}(x) = \frac{1}{f_{t,\sigma}(x)} \frac{df_{t,\sigma}(x)}{du} \frac{\partial u}{\partial t}$$

and $f_{t,\sigma}(x) = \sigma^{-1} g(u)\psi'(x)$ by (12) and $\frac{\partial u}{\partial t} = \sigma^{-1}\psi'(t)$ by (14), it holds that

$$\frac{\partial}{\partial t} \ln f_{t,\sigma}(x) = \frac{1}{\sigma}\psi'(t) T_G(u) = \frac{1}{\sigma}\psi'(t) T_{F_{t,\sigma}}(x).$$

□

The core function of a distribution with partial support $S \neq R$ and with parameter t (for example: exponential, Weibull, log-logistic, Johnson's U_B, log-normal) is thus the *inner part of the likelihood score for t.*

3. CORE MOMENTS

Following Fabián (2001), the core functions can be used for a definition of alternative moments of continuous probability distributions.

Definition 4. Let $F \in \Pi_S$ with core function T_F. The $k-$th order core moment of F is

$$M_k(F) = ET_F^k = \int_S T_F^k(x)\ dF(x), \qquad k = 1, 2, ... \tag{16}$$

Proposition 2. Let $G \in \Pi_R$, $\varphi \in \Phi_S$ and $F = G\varphi^{-1}$. It holds that $M_k(F) = M_k(G)$ whenever one of these integrals exist.

Proof see Fabián (2001).

Proposition 3. Core moments of regular distributions exist.

Proof. Due to Proposition 2, it suffices to examine only distributions with support $S = R$. Let $G \in \Pi_R$ and $T = T_G$ be its core (score) function. By (6), the density of G is $g(y) = ce^{-Q(y)}$ where $Q(y) = \int T(y)\ dy$ and c is an integration constant. Let us denote $g_k(y) = cT^k(y)e^{-Q(y)}$. It is to prove a relation $\int_q^\infty g_k(y)\ dy < \infty$ for some $q > 0$.

The derivative of g_k is

$$g_k'(y) = ce^{-Q(y)}T(y)^{k-1}[kT'(y) - T^2(y)].$$

If there exists such y_1 that it holds that $[kT'(y) - T^2(y)] < 0$ for $y > y_1$, g_k is a decreasing function and we can find such y_2 that for $y > y_2$ it holds that $g_k(y) < ck^k$.

Set $q = \max(1, y_1, y_2)$. For $y > q$ it holds that $g_k(y)/ck^k < 1$ and

$$\frac{1}{ck^k} g_k(y) = \left(-\frac{d}{dy} e^{-\frac{1}{k}Q(y)}\right)^k < -\frac{d}{dy} e^{-\frac{1}{k}Q(y)}.$$

Thus, $\int_q^\infty g_k(y)\, dy < ck^k e^{-\frac{1}{k}Q(y)}|_q^\infty < \infty$. The existence of y_1 follows from the existence of finite $M_2(F)$, guaranteed by regularity condition (5). □

Proposition 4. Core moments are independent of the transformed location and scale parameters.

Proof. By (13) and (12),

$$M_k(F_\theta) = \int_R T_G^k(u) g(u) \sigma^{-1} \psi'(x)\, dx = \int_R T_G^k(u) g(u)\, du.$$

□

Core moments thus describe the shape of distributions. In Table 5 are given values of core moments for parametric families listed in Table 1.

Table 5. Core Moments of Families from Table 1

F_θ	$M_2(F_\theta)$	$M_3(F_\theta)$	$M_4(F_\theta)$
1	α	-2α	$3\alpha(\alpha+2)$
2	$\frac{\alpha^2}{2\alpha+1}$	0	$\frac{3\alpha^4}{(2\alpha+1)(2\alpha+3)}$
3	$\frac{\alpha(2\alpha-1)}{\alpha+1}$	0	$\frac{12\alpha^3(2\alpha-1)(2\alpha+1)}{(\alpha+1)(2\alpha+4)(2\alpha+6)}$

Core moments of families from Table 1 not only exist but, in contrast with moments (1), are expressed only by means of parameters and not of functions of parameters. Moreover, they depend only on shape parameter (in accordance with Proposition 4). We conjecture that they are the true numerical characteristics.

4. CENTER OF GRAVITY AND INFORMATION OF DISTRIBUTIONS

Proposition 5. For any $F \in \Pi_S$, it holds that

$$M_1(F) = \int_S T_F(x)\, dF(x) = 0. \tag{17}$$

Proof see Fabián (2001).

According to this proposition, core moments are the central moments around the point $x^* : T_F(x^*) = 0$. This point appears to be generally different from the mean, mode or median.

Definition 5. Let $F \in \Pi_S$. Point $x^* : T_F(x^*) = 0$ will be called the center of gravity of distribution F.

By (6) and Definition 4, the center of gravity y^* of distribution G is the mode or, in the parametric case, the location μ according to (10). The center of gravity of a composite distribution $F_{\tau,\sigma} = G_{\mu,\sigma}\psi$ is, according to (14), $\tau = \psi^{-1}(\mu)$. Parameter t introduced in (15) is thus the *center of gravity* of the distribution. It remains to determine the center of gravity of composite distribution $F = G\psi$ without parameter t: it is apparently the point $x^* = \psi^{-1}(y^*)$. The notion of the center of gravity is thus independent of the fact whether the distribution is parametric or without parameters.

Now we are prepared to give an interpretation of the core function:

The core function of a regular continuous distribution is the likelihood score for the center of gravity of the distribution (*whether expressed as a parameter or not*). Remembering that the influence function of the estimated parameter is proportional to the likelihood score for this parameter [Hampel et al. (1986)], we suggest: The core function of a regular distribution describes the local influence of value $x \in S$ on the position of the center of gravity or, if suitably normed, *it is the influence function of the center of gravity of the distribution.*

It is interesting that the square of the core function has a reasonable interpretation as well.

A common sense of the notion information is that the information contained in an more or less expected observation is low and, on the contrary, the unexpected observation carries on a lot of information. In the next two propositions, a relationship between the first two core moments and the information is established.

Proposition 6. Let $F \in \Pi_S$. The center of gravity is the least informative point of the distribution.

Proof. Let X be distributed by $F = G\psi$. For a given ψ there exists a large class $(\Pi_S)_\psi \subset \Pi_S$ of composite distributions with given ψ. The density of F is $f(x) = g(\psi(x))\ \psi'(x)$. Term $\psi'(x)$ is common to all $F \in (\Pi_S)_\psi$ and, therefore, does not carry any information about X. All information contained in X is thus condensed in term $g(\psi(x))$. This is minimal at a point $\tilde{x} : \frac{d}{dx}g(\psi(x)) = 0$. By using (7), one obtains $\tilde{x} : \frac{d}{dx}(f(x)/\psi'(x)) = 0$, which reduces using (9) into $\tilde{x} : -T_F(x) = 0$. The solution of the last equation is just the center of gravity $\tilde{x} = x^*$. □

Proposition 7. Let $F_{t,\sigma} \in \Pi_S$ be a parametric distribution with parent F and $J_{tt}(\theta)$ be the Fisher information (2) for its center of gravity t. Then

$$J_{tt}(\theta) = \frac{1}{\sigma^2}\psi'(t)^2 M_2(F), \tag{18}$$

Proof. By (3) and Theorem 1

$$J_{tt}(\theta) = \frac{1}{\sigma^2}\psi'(t) \int_s T^2_{F_\theta}(x) f_\theta(x)\ dx.$$

By (13) and (12),

$$\int_s T^2_{F_\theta}(x) f_\theta(x)\ dx = \int_R T^2_G(u) g(u)\ du.$$

Using Proposition 2 we obtain (18). □

Corollary: Fisher information J_F of a parent distribution F without parameters is $J_F = \psi'(\psi^{-1}(0)) M_2(F)$.

Consider now function $i_F(x) = T^2_F(x)$. It is a non-negative function, attaining its minimum $i_F(x^*) = 0$ in the least informative point of the distribution, from which it increases either quickly in the cases of distributions with unbounded cores (and sharply to zero tending densities) - for which an outlier has an immense informative value: it indicates the necessity of a change of the model - or slowly in cases of distributions with bounded cores (and heavy-tailed densities) - where the values far from the 'bulk' of the data are more or less expected. Moreover, its mean value is the Fisher information.

Our opinion is that the real function $i_F(x) = T_F^2(x)$ is the *information function* of distribution F, expressing the local information about the center of gravity of F, carried by value x. Its mean value is simply the mean information about the center of gravity of F. We conjecture that it can be considered to represent the *mean information of* F.

In Table 6 we compare functions $-\ln f(x)$ and $i_F(x)$ for some simple distributions with support $S = (0, \infty)$.

Table 6. Information Functions of some Distributions

Name	$-\ln f(x)$	$i_F(x)$
Wald	$\ln(Kx) + \frac{1}{2}(x - 1/x)$	$\frac{1}{4}(x - 1/x)^2$
Lognormal	$\ln(\sqrt{2\pi}x) + \frac{1}{2}\ln^2 x$	$\ln^2 x$
Exponential	x	$(x-1)^2$
Extreme val. II	$2\ln x + 1/x$	$(1 - 1/x)^2$
Log-logistic	$2\ln(1+x)$	$\left(\frac{x-1}{x+1}\right)^2$
Log-Cauchy	$\ln\pi + \ln(1 + \ln^2 x)$	$\frac{4\ln^2 x}{(1+\ln^2 x)^2}$

Note that in the cases of parametric distributions, the expressions of $-\ln f(x)$ are even more complicated. For example, for the relatively simple gamma distribution we have $-\ln f(x) = -\alpha\ln\gamma + \ln\Gamma(\alpha) - (\alpha - 1)\ln x + \gamma x$, whereas $i_F(x) = (\alpha x - \gamma)^2$.

5. ESTIMATES

Let us have random sample $X_1, ..., X_n$ from distribution $F_{t,\sigma}$ with support S, parent F and unknown parameters t, σ.

5.1 Sample from Normal Distribution

Let us assume that the sample is from normal distribution $\Phi_{\mu,\sigma}$ with density $\phi(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$. The variance of normal distribution is $\sigma^2 = \int_{-\infty}^{\infty}(x-\mu)^2\phi(x)\,dx$, Fisher information about μ

$$J_{\mu\mu}(\theta) = \int_{-\infty}^{\infty}\left(\frac{x-\mu}{\sigma^2}\right)^2\phi(x)\,dx = \frac{1}{\sigma^2}, \tag{19}$$

core function is $T_\Phi(x) = (x-\mu)/\sigma$ and the second core moment

$$M_2(\Phi) = \int_{-\infty}^{\infty}\left(\frac{x-\mu}{\sigma}\right)^2\phi(x)\,dx = 1.$$

By Proposition 7, $J_{\Phi_{\mu,\sigma}} = 1/\sigma^2$, which is in agreement with (19). The sample characteristics are thus the sample center of gravity $\hat{\mu}$ and the sample information $\hat{J}_{\Phi_{\mu,\sigma}} = 1/s^2$ where s^2 is an estimate of the variance.

5.2 Sample from General Distribution

For estimation of parameter θ, the maximum likelihood method or some of its robust modifications [Hampel et al. (1986), Jurečková and Sen (1998)] can be used. In Fabián (2001), a different way has been suggested: the core moment method. In a particular case of estimating parameters t and σ, it consists in the solution of a system of equations

$$\sum_{i=1}^{n} T_F(u_i) = 0 \tag{20}$$

$$\frac{1}{n}\sum_{i=1}^{n} T_F^2(u_i) = M_2(F) \tag{21}$$

where u_i are values of the transformed location and scale kernel (14) (see Table 4) corresponding to the given S. For example,

$$u_i = \begin{cases} \frac{x_i-\mu}{\sigma} & \text{if } \ S = R \ \text{ and } \ \psi(u) = u \\ \ln\left(\frac{x_i}{\tau}\right)^{1/\sigma} & \text{if } \ S = (0,\infty) \ \text{ and } \ \psi(u) = \ln u. \end{cases} \tag{22}$$

Thus the parameters are estimated so that the sample core moments match the theoretical core moments. It has been shown in Fabián (2001) that the core moment estimates are consistent and asymptotically normal, and that a suitable compromise between the efficiency and the robustness of the core moment estimates is easily obtainable.

For some distributions, it is possible to estimate the center of gravity without knowledge of the scale parameter from the first core moment equation (20). In Table 7 we present densities $f_\theta(x)$, core functions $T_{F_\theta}(x)$ and the centers of gravity t of some distributions F_θ and estimates $\hat{t}_n$ of t.

Table 7. Estimates of the center of gravity of some distributions

Distribution	$f_\theta(x)$	$T_{F_\theta}(x)$	t	$\hat{t}_n$
Normal	$\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$	$\frac{x-\mu}{\sigma}$	μ	$\bar{x} = \frac{1}{n}\sum x_i$
Lognormal	$\frac{1}{\sqrt{2\pi}x}e^{-\frac{1}{2}\ln^2(x/\tau)}$	$\ln x/\tau$	τ	$\bar{x}_G = \sqrt[n]{x_1 \dots x_n}$
Exponential	$\tau^{-1}e^{-x/\tau}$	$x/\tau - 1$	τ	$\bar{x}$
Extr. val. II	$\frac{\tau}{x^2}e^{-\tau/x}$	$1 - \tau/x$	τ	$\bar{x}_H = \frac{n}{\sum 1/x_i}$
GIG	$\frac{1}{Kx}e^{-\frac{1}{2}(x/\tau+\tau/x)}$	$\frac{1}{2}(x/\tau - \tau/x)$	τ	$\bar{x} - \bar{x}_H$
Gumbel	$e^{x-\mu}e^{-e^{x-\mu}}$	$e^{x-\mu} - 1$	μ	$\ln(\frac{1}{n}\sum e^{x_i})$
Lomax	$\frac{\alpha}{(1+x)^{\alpha+1}}$	$\frac{x-1/\alpha}{x+1}$	$1/\alpha$	$\frac{\sum x_i/(1+x_i)}{\sum 1/(1+x_i)}$
Gamma	$\frac{\gamma^\alpha}{\Gamma(\alpha)x}x^{\alpha-1}e^{-\gamma x}$	$\alpha(\frac{x}{\alpha/\gamma} - 1)$	α/γ	$\bar{x}$
Beta	$\frac{1}{B(p,q)}x^{p-1}(1-x)^{q-1}$	$p(\frac{x}{p/(p+q)} - 1)$	$\frac{p}{p+q}$	$\bar{x}$

In the table, GIG means the generalized inverse Gaussian distribution [Johnson, Kotz and Ballakrishnan (1994)], Γ and B are the gamma and beta functions and $K = 2K_0(1)$. The estimate of the center of gravity (the *core sample mean*) of lognormal distribution is the geometric mean, the core sample mean of the extreme value II distribution is the harmonic mean. There are some distributions with linear core functions (normal distribution with $S = R$, gamma with $S = (0, \infty)$ and beta with $S = (0, 1)$). Their centres of gravity are the means and their sample core means are equal to the arithmetic means of the observed values.

Denote by $J_n(t)$ the information about the center of gravity of distribution F_θ contained in a random sample $X_1, ..., X_n$ taken from F_θ. According to the Corollary of Proposition 7, in the cases of distributions generated by Johnson's ψ, for which $\psi'(\psi^{-1}(0)) = 1$, information of F_θ is

$$J_{F_\theta} = J_{tt}(\theta) \tag{23}$$

where $J_{tt}(\theta)$ is the Fisher information for the center of gravity. The Cramér-Rao theorem says that

$$Var(\hat{t}_n) = \frac{1}{nJ_{tt}(\theta)}. \tag{24}$$

Since $\hat{t}_n$ is asymptotically normal,

$$J_n(t) = 1/Var(\hat{t}_n), \tag{25}$$

so that

$$J_n(t) = nJ_{F_\theta}.$$

The Cramér-Rao theorem can thus be interpreted in the sense that the information about the center of gravity of a distribution F_θ contained in a random sample taken from F_θ equals to the information of distribution F_θ multiplied by the length of the samples. According to Proposition 7, a natural estimate of $J_n(t)$ is thus, in case of a distribution with support $S = (0, \infty)$,

$$\hat{J}_n(t) = \frac{n}{(\hat{\tau}_n \hat{\sigma}_n)^2} M_2(F).$$

Example 2. Let $X_1, ..., X_n$ be random sample from gamma distribution (see Example 1). It holds (Table 7) that $\tau = \alpha/\gamma$ and, by Fabián (2001), $M_2(\text{gamma}) = \alpha$, so that $J_n(t) = \frac{n}{(\alpha/\gamma)^2}\alpha = n\gamma^2/\alpha$. Let n_1 be the length of a sample from the exponential distribution $(\tau = 1, \sigma = 1, M_2 = 1)$, for which it can be supposed that the asymptotic relations are approximately valid. Since $J_{n_1} = n_1$, from the requirement $J_n(t) = J_{n_1}$ it follows that $n = \hat{\alpha} n_1 / \hat{\gamma}^2$.

AKNOWLEDGEMENTS

The work was supported by grant GA AV A1075101.

REFERENCES

1. Bickel, P.J. and Lehmann, E.L. (1975). Descriptive statistics for nonparametric models I, II, *Ann. Statist.*, **3**, 1038-69.

2. Cover, T.M. a Thomas, J.A. (1991). *Elements of Information Theory,* Wiley.

3. Fabián, Z. (1997). Information and entropy of continuous random variables. *IEEE Trans. on Information Theory*, 43, 1080-1083.

4. Fabián, Z. (2001). Induced cores and their use in robust parametric estimation. *Commun. Stat., Theory Methods*, 30, 3, 537-556.

5. Hampel F. R., Rousseew P. J., Ronchetti E. M. and Stahel W. A. (1986). *Robust Statistic. The Approach Based on Influence Functions,* Wiley, New York.

6. Johnson, N.L. (1949). Systems of Frequency Curves Generated by Methods of Translations. *Biometrika* **36**, 149-176.

7. Johnson, N.L., Kotz, S. and Balakrishnan, N. (1994, 95). *Continuous univariate distributions 1, 2*, Wiley.

8. Jurečková, J. and Sen, P. K. (1996). *Robust Statistical Procedures: Asymptotics and Interrelations,* Wiley.

9. van der Waart, A. W. (1998). *Asymptotic Statistics,* Cambridge Univ. Press.