On the time-dependent Fisher information of a density function

Omid Kharazmi[a] and Majid Asadi[a,b]

[a]*University of Isfahan*
[b]*Institute of Research in Fundamental Sciences (IPM)*

Abstract. Fisher information is a very important and fundamental criterion in statistical inference especially in optimal and large sample studies in estimation theory. It also plays a key role in physics, thermodynamic, information theory and other applications. In the literature there have been defined two forms of Fisher information: one for the parameters of a distribution function and one for the density function of a distribution. In this paper, we consider a nonnegative continuous random (lifetime) variable X and define a time-dependent Fisher information for density function of the residual random variable associated to X. We also propose a time-dependent version of Fisher information distance (relative Fisher information) between the densities of two nonnegative random variables. Several properties of the proposed measures and their relations to other statistical measures are investigated. To illustrate the results various examples are also provided.

1 Introduction

The concept of Fisher information is a fundamental concept in statistical inference and plays a crucial role in many other disciplines such as information theory and coding, derivation of physical laws in statistical physics and thermodynamics, probability theory, Kinetic theory, astronomy as well as biosciences (see, for example, Frieden (2004), Pardo, Morales and Taneja (1995), Shao (2003)). Fisher information provides a way of quantifying the amount of information that observations of a random sample carry about an unknown parameter or vector of parameters in a statistical model. Let us consider a random variable X (continuous or discrete) with a distribution function F_θ having a probability density function f_θ, where $\theta \in \Theta \subseteq R$. We assume throughout the paper that $f_\theta(x)$ is differentiable with respect to both θ and x. The Fisher information of random variable X (or distribution F_θ) about the parameter θ, based on an observation x of X, is defined

as

$$I(\theta) = E\left(\frac{\partial \log f_\theta(X)}{\partial \theta}\right)^2$$
$$= \int \left(\frac{\partial \log f_\theta(x)}{\partial \theta}\right)^2 f_\theta(x)\,dx. \tag{1.1}$$

In fact, the Fisher information $I(\theta)$ measures how high the peaks are in the log-likelihood function. It plays crucial role, in both classic and Bayesian statistical inference, in derivation of optimal estimators particularly in large sample studies when one is interested to study asymptotic properties of estimators; see, for example Lehmann and Casella (1998) and Shaked and Shanthikumar (2007). Another form of the Fisher information which is known in the literature as the *Fisher information of the density function* and has different applications is considered by authors. If a random variable X has density $f(x)$, under the condition that the derivative of f exists for all values on its support, the Fisher information of the density is defined as

$$I(f) = \int \rho^2(x) f(x)\,dx, \tag{1.2}$$

where $\rho(x) = \frac{f'(x)}{f(x)}$ is called the score function corresponding to f. The quantity $I(f)$ provides information about the density itself; (see, for example, Yáñez et al. (2008)). Recently, attention have been shown to $I(f)$ and its applications in different areas. In physics theory, $I(f)$ is used as a measure of the state of disorder of a system or phenomenon; see Frieden (2004) for more details. Brown (1982) applied $I(f)$ to prove central limit theorem based on Cramér-Rao inequality. Johnson and Barron (2004) and Johnson (2004) used $I(f)$ to prove central limit theorem under different conditions. Papaioannou and Ferentinos (2005) studied several properties of the Fisher information in form (1.2). Bercher and Vignat (2009) obtained the minimum Fisher information distributions in the class of distributions with restricted support and fixed variance. Kostal, Lansky and Pokora (2013) studied some measures of dispersion based on Fisher information. Yáñez et al. (2008) showed that the two measures in (1.1) and (1.2) are connected through a measure of gain in information from a Bayesian experiment.

Note that when θ is a location parameter, that is, $f_\theta(x) = f(x - \theta)$, by the fact that, under regularity conditions, $\partial f(x - \theta)/\partial \theta = -\partial f(x - \theta)/\partial x$, the Fisher information $I(\theta)$ in (1.1) and Fisher information of the density function $I(f)$ in (1.2) are identical, see Johnson (2004).

The Fisher information distance (or relative Fisher information) is also defined in the literature. Let X and Y be two continuous nonnegative random variables with absolutely continuous density functions f and g, respectively. The Fisher information distance between X and Y (or f and g) is defined by

$$D(f, g) = \int \left(\rho_f(x) - \rho_g(x)\right)^2 f(x)\,dx, \tag{1.3}$$

where $\rho_f(x)$ and $\rho_g(x)$ are the score functions corresponding to f and g, respectively. The Fisher information distance between Y and X is defined similarly. The Fisher information and Fisher information distance are studied and applied by different authors in various of subject of applications. We refer, among others, to Bobkov, Chistyakov and Gtze (2014), Johnson (2004), Otto and Villani (2000), Venkatesan and Plastino (2014), Walker (2016), Yamano (2013) and Zegres (2002).

The aim of the present paper is to investigate properties of Fisher information of the density function, $I(f)$, and Fisher information distance $D(f, g)$. In the first part of the paper, we propose a time-dependent version of $I(f)$ as measure of information for nonnegative random variables denoting a duration. Duration study is a subject of interest in many branches of science such as reliability, survival analysis, actuary, economics, business etc. Let X be a nonnegative random variable denoting a duration such as a lifetime where we assume that it has the distribution function F, and the probability density function f. Capturing effects of the age t of an individual or an device under study on the information about the remaining lifetime is important under different purposes. For example, in reliability when a component or a system of components is working at time t, one is interested in the study of the information of the density of the lifetime of component or system beyond t. In such case, the random variable of of interest is the *residual random variable*, $X_t = X|X > t$, on the set

$$\mathcal{S}_t = \{x : x > t\}, \qquad t \in (0, b), b \leq \infty.$$

Hence, the distribution of interest for computing information is the residual distribution with survival function

$$\bar{F}_t(x) = \begin{cases} \dfrac{\bar{F}(x)}{\bar{F}(t)} & x \in \mathcal{S}_t, \\ 1 & \text{otherwise,} \end{cases} \qquad (1.4)$$

provided that $\bar{F}(t) < \infty$, where $\bar{F} = 1 - F$ denotes the survival function of X (Barlow and Proschan (1981)). We denote the Fisher information of residual random variable X_t by $I(f; t)$. In Section 2, we obtain the functional form of $I(f; t)$. We show that the underlying distribution function F can be represented in terms of $I(f; t)$ and the score function $\rho(x)$. It is argued in the class of nonnegative continuous random variables, the exponential distribution can be characterized as a distribution for which $I(f; t)$ is constant. We prove, in Section 2, that when two random variables are ordered in the sense of likelihood ratio order, then under some conditions, the corresponding residual random variables are ordered in the sense of residual Fisher information. It is shown that the Fisher information of the equilibrium distribution associated to X is closely related to the hazard rate and mean residual life of X. In particular, we show that when the random variable X has increasing (decreasing) hazard rate then the Fisher information of the equilibrium is increasing (decreasing). The correlation between the cumulative hazard

rate and the score function of X are explored in Section 2. Section 3 is devoted to Fisher information distance $D(f, g)$. We propose a time-dependent version of this distance which is useful for measuring the distance between two residual lifetime distributions. We show that the Fisher information distance between the residual random variable associated to weighted distribution and its parent distribution can be represented as the conditional expectation of square of the weight function. Several examples are shown as special cases. For instance, the distance between the distribution of order statistics and the underlying distribution and also distance between different order statistics are investigated in Section 3. In this section, we also study the Fisher information distance between escort distributions which play important role in thermodynamics and physics. Throughout the paper, several illustrative examples are also provided.

2 Fisher information of the residual life distributions

Let $f_t(x)$ denote the density function corresponding to $\bar{F}_t(x)$. That is, $f_t(x) = \frac{f(x)}{\bar{F}(t)}$, $x > t$. Then we have the following definition.

Definition 2.1. Let X_t be a residual random variable with an absolutely continuous density function $f_t(x)$. The residual Fisher information of $f_t(x)$ is defined as

$$
\begin{aligned}
I(f; t) &= \int_t^b \left(\frac{\frac{df_t(x)}{dx}}{f_t(x)} \right)^2 f_t(x)\, dx \\
&= \frac{\int_t^b \rho^2(x) f(x)\, dx}{\bar{F}(t)} \\
&= E\big(\rho^2(X)|X > t\big),
\end{aligned}
\tag{2.1}
$$

where $t > 0$ and $b \le \infty$ the right extremity of the support of X, that is, $F(b) = 1$.

The idea here is to measure the relative changes of the residual density after time t; that is, how much small changes in x, $x > t$, affect the the residual density. Obviously, as $t \to 0$, $I(f; t)$ tends to the Fisher information $I(f) = \int_0^b \rho^2(x) f(x)\, dx$. It can be easily shown that $I(f, t)$ can be represented as

$$
I(f; t) = 4 \frac{\int_t^b (\frac{d}{dx} \sqrt{f(x)})^2\, dx}{\bar{F}(t)},
$$

see also, Johnson (2004), p. 24.

Let us look at the following examples.
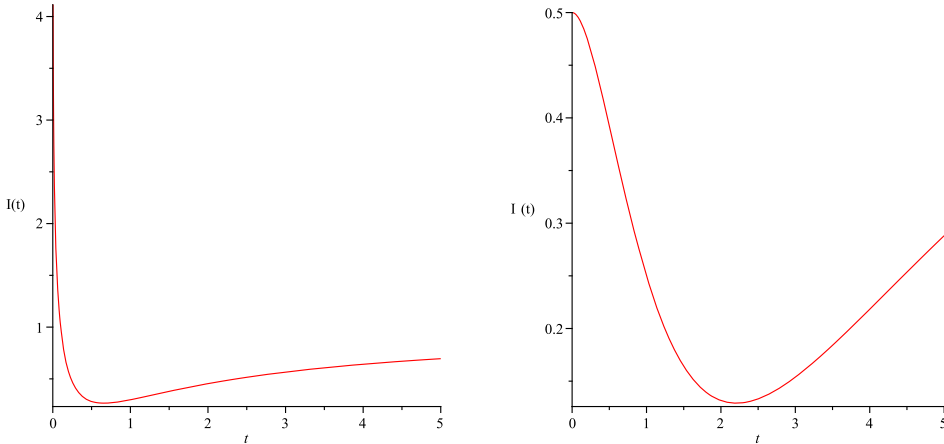
Figure 1 *The plots of* (a): $I(f;t)$ *for* Gamma(2, 1), (b): $I(f;t)$ *for* Gamma(4, 1).

Example 2.1. Let X be distributed as gamma with density

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \qquad x > 0, \alpha, \beta > 0.$$

Then we have $\rho^2(x) = (\frac{\alpha-1}{x} - \beta)^2$, and hence for $\alpha > 2$ we obtain

$$I(f;t) = \frac{\beta^2((\alpha-1)^2\Gamma(\alpha-2, \beta t) - 2(\alpha-1)\Gamma(\alpha-1, \beta t) + \Gamma(\alpha, \beta t))}{\Gamma(\alpha, \beta t)},$$

where $\Gamma(\alpha, t) = \int_t^\infty x^{\alpha-1} e^{-x}\, dx$ is incomplete Gamma function. Figure 1(a) depicts the plot of $I(f;t)$ for $\alpha = 3$ and $\beta = 1$ and Figure 1(b) depicts the plot of $I(f;t)$ for $\alpha = 4$ and $\beta = 1$. It is evident from the plots that the residual Fisher information first decreases for a period of time and then increases.

Example 2.2. Let X have beta distribution with density function

$$f(x) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1}(1 - x)^{\beta-1}, \qquad 0 < x < 1, \alpha > 0, \beta > 0,$$

where $B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1 - x)^{\beta-1}\, dx$.

Then $\rho^2(x) = (\frac{\alpha-1}{x} - \frac{\beta-1}{1-x})^2$. Hence for $\alpha, \beta > 2$

$$I(f;t) = \frac{(\alpha-1)^2 B(t, \alpha-2, \beta) - 2(\alpha-1)(\beta-1) B(t, \alpha-1, \beta-1)}{B(t, \alpha, \beta)}$$

$$+ \frac{(\beta-1)^2 B(t, \alpha, \beta-2)}{B(t, \alpha, \beta)},$$

where $B(t, \alpha, \beta) = \int_t^1 x^{\alpha-1}(1-x)^{\beta-1}\,dx$. It can be easily seen that for $\beta = 1$ and $\alpha = 2$ we get

$$I(f;t) = \frac{B(t, 0, 1)}{B(t, 2, 1)} = \frac{\int_t^1 \frac{1}{x}\,dx}{\int_t^1 x\,dx} = \frac{-2\log t}{1-t^2},$$

which is a decreasing function of t on $t \in (0, 1)$. For $\alpha = 1$ and $\beta = 3$, we get

$$I(f;t) = \frac{4B(t, 1, 1)}{B(t, 1, 3)} = \frac{12}{(1-t)^2}$$

which is an increasing function of t on $(0, 1)$.

Under the assumption that $I(f;t)$ is differentiable in terms of t, we have

$$I'(f;t) = r(t)\big(I(f;t) - \rho^2(t)\big), \tag{2.2}$$

where $r(t) = \frac{f(t)}{\bar{F}(t)}$ is the hazard rate of X. The following conclusions are immediate from equation (2.2). First, note that $I(f;t)$ is a decreasing (increasing) function of t if and only if $I(f;t) \leq (\geq) \rho^2(t)$. If $I(f;t)$ is constant, then we have $(I(f;t) - \rho^2(t)) = 0$. This, in turn, implies after differentiating both sides of this last equation that $\rho(x) = c$, where c is constant. After some algebraic manipulations, we get $f(t) = ke^{ct}$, where k is normalizing constant. If c is positive, f is increasing in its support and hence the right extremity b should be finite. If c is negative, f is decreasing in its support and we arrive at the exponential distribution. Assuming that $I(f;t) \neq \rho^2(t)$, we get

$$r(t) = \frac{I'(f;t)}{I(f;t) - \rho^2(t)},$$

which implies, based on relation

$$\bar{F}(t) = e^{-\int_0^t r(x)\,dx}, \tag{2.3}$$

that for all $t > 0$, the survival function $\bar{F}$ can be represented as

$$\bar{F}(t) = e^{-\int_0^t \frac{I'(f,x)}{I(f,x) - \rho^2(x)}\,dx}.$$

Definition 2.2. Let X and Y be two random variables with residual Fisher information functions $I(f;t)$ and $I(f,t)$ respectively. X is said to be less than Y in residual Fisher information, denoted by $X \leq_{\text{DFI}} Y$, if $I(f;t) \leq I(g;t)$, for all t.

By definition, we can easily see that, for $b = \infty$, if $X \leq_{\text{DFI}} Y$ and $I(f;t)$ (or $I(g;t)$) is increasing then $a_1 X + b_1 \leq_{\text{DFI}} a_2 Y + b_2$, when $a_2 \leq a_1$ and $b_2 \leq b_1$.

Theorem 2.1. *Let X be a nonnegative random variable with an absolutely continuous density function f and ϕ be a nonnegative, increasing, twice differentiable*

and invertible function. Then the residual Fisher information of $\phi(X)$, denoted by $I_{\phi(X)}(t)$, is given as

$$I_{\phi(X)}(t) = \frac{1}{\bar{F}(\phi^{-1}(t))} \int_{\phi^{-1}(t)}^{\phi^{-1}(b)} \frac{1}{\phi'(x)^2}\left[\frac{f'(x)}{f(x)} - \frac{\phi''(x)}{\phi'(x)}\right]^2 f(x)\,dx. \qquad (2.4)$$

Proof. Using the definition of residual Fisher information we have

$$I_{\phi(X)}(t) = \frac{1}{\bar{F}(\phi^{-1}(t))} \int_t^b \left[\frac{\left(\frac{f(\phi^{-1}(x))}{\phi'(\phi^{-1}(x))}\right)'}{\frac{f(\phi^{-1}(x))}{\phi'(\phi^{-1}(x))}}\right]^2 \frac{f(\phi^{-1}(x))}{\phi'(\phi^{-1}(x))}\,dx$$

$$= \frac{1}{\bar{F}(\phi^{-1}(t))} \int_{\phi^{-1}(t)}^{\phi^{-1}(b)} \frac{1}{\phi'(x)^2}\left[\frac{f'(x)}{f(x)} - \frac{\phi''(x)}{\phi'(x)}\right]^2 f(x)\,dx. \qquad \square$$

Example 2.3. Suppose that X has density $f(x) = \alpha x^{\alpha-1}$, $0 < x < 1$. Then we have $\frac{f'(x)}{f(x)} = \frac{\alpha-1}{x}$. Let $\phi(X) = X^n$, $n < \frac{\alpha}{2}$, then we easily obtain

$$I_{\phi(X)}(t) = \frac{\alpha(1 - \frac{\alpha}{n})^2}{\alpha - 2n} \frac{1 - t^{\frac{\alpha-2n}{n}}}{1 - t^{\frac{\alpha}{n}}}.$$

The following theorem shows that if X and Y are ordered in the sense of like-lihood ratio order, then they are also ordered in DFI order under some conditions. Recall that for two random variables X and Y with densities f and g, respectively:

- X is said to be less than Y in likelihood ratio order, $X \leq_{\mathrm{lr}} Y$, if $\frac{g(x)}{f(x)}$ is increasing in x for all x in the union of supports of X and Y.
- X is said to be less than Y in hazard rate order, $X \leq_{\mathrm{hr}} Y$, if $r_X(x) \geq r_Y(x)$, for all x in the union of supports of X and Y, where $r_X(x)$ $(r_Y(x))$ is the hazard rate of $X(Y)$.

Theorem 2.2. *Let X and Y have densities f and g, respectively. Assume that f is increasing and g is log-convex. If $X \leq_{\mathrm{lr}} Y$, then $X \leq_{\mathrm{DFI}} Y$.*

Proof. $X \leq_{\mathrm{lr}} Y$ implies that $\frac{g(x)}{f(x)}$ is increasing in x, or equivalently $\frac{f'(x)}{f(x)} \leq \frac{g'(x)}{g(x)}$ for all x. Since f is assumed to be increasing we get $(\rho_f(x))^2 \leq (\rho_f(x))^2$. From this, we have

$$\frac{\int_t^b (\rho_f(x))^2 f(x)\,dx}{\bar{F}(t)} \leq \frac{\int_t^b (\rho_g(x))^2 f(x)\,dx}{\bar{F}(t)}. \qquad (2.5)$$

Since g is log-convex the score function $\rho_g(x)$ is an increasing function of x. Hence from the assumption $0 \leq \rho_f(x) \leq \rho_g(x)$, and that $X \leq_{\mathrm{lr}} Y \Rightarrow X \leq_{\mathrm{hr}} Y$, we

get

$$\frac{\int_t^b (\rho_g(x))^2 f(x)\,dx}{\bar{F}(t)} \leq \frac{\int_t^b (\rho_g(x))^2 g(x)\,dx}{\bar{G}(t)}, \tag{2.6}$$

where the inequality follows from the fact that $X \leq_{\mathrm{hr}} Y$ implies $E(\phi(X)|X > t) \leq E(\phi(Y)|Y > t)$ for any increasing function $\phi(x)$. (see Salicrú and Taneja (1993)). From (2.5) and (2.6), we get the result. $\square$

Example 2.4. Let X be distributed as $f(x) = \frac{2x}{3}$, $1 < x < 2$ and Y be distributed as $g(x) = \frac{e^x}{e^2 - e}$, $1 < x < 2$. clearly $f(x)$ is increasing and $g(x)$ is log-convex. Also, as the ratio e^x/x is increasing on $(1, 2)$, we have $X \leq_{\mathrm{lr}} Y$. Hence using above theorem, we have $X \leq_{\mathrm{DFI}} Y$.

Assume that $\bar{F}(x)$ is the survival function of a nonnegative continuous random variable X with finite mean μ. The random variable X_e is said to be the equilibrium random variable corresponding to the random variable X, if the density function of X_e is given by

$$f_e(x) = \frac{\bar{F}(x)}{\mu}, \qquad x > 0. \tag{2.7}$$

The equilibrium distributions arise in renewal theory as the asymptotic distributions of the waiting time till the next event and the time since the last event at time t. It is also known that a delayed renewal process has stationary increments if and only if the distribution of the actual remaining life is (2.7). The random variable X_e with density (2.7) is also referred to as the "asymptotic age" by Salicrú and Taneja (1993). The cited monograph gives several ordering results on these distributions. The following theorem shows a representation result for the residual Fisher information of equilibrium distribution. Before presenting the theorem, we recall that the mean residual lifetime (MRL) of continuous random variable X with survival function $\bar{F}$ is defined at time t as as

$$m(t) = E(X - t|X > t) = \frac{\int_t^b \bar{F}(x)\,dx}{\bar{F}(t)},$$

provided that $\bar{F}(t) > 0$. Note that $m(0) = \mu$ is the mean of X.

Theorem 2.3. *The residual Fisher information of X_e can be represented as*

$$I(f_e; t) = \frac{E(r(X)|X > t)}{m(t)},$$

where $r(t)$ and $m(t)$ denote the hazard rate and the MRL of X, respectively.

Proof. We have

$$I(f_e; t) = E\left(\left(\frac{f'_e(X_e)}{f_e(X_e)}\right)^2 \middle| X_e > t\right)$$

$$= \frac{\int_t^b (\frac{f'_e(x)}{f_e(x)})^2 \bar{F}(x)\, dx}{\int_t^b \bar{F}(x)\, dx}$$

$$= \frac{\int_t^b (r(x))^2 \bar{F}(x)\, dx}{\int_t^b \bar{F}(x)\, dx} \qquad (2.8)$$

$$= \frac{\int_t^b r(x) f(x)\, dx}{\int_t^b \bar{F}(x)\, dx}$$

$$= \frac{E(r(X)|X > t)}{m(t)}. \qquad \square$$

Corollary 2.1. *From* (2.8), *we get that the Fisher information of X_e is given by $I(f_e; 0) = I(f_e) = \frac{E(r(X))}{\mu}$, where μ is the mean of X. Also, it should be pointed out that if $r(t)$ is increasing (decreasing) then $E(r(X)|X > t)$ is increasing (decreasing). Hence from the fact that a random variable with increasing (decreasing) hazard rate has decreasing (increasing) MRL (see* Barlow and Proschan (1981)), *the representation* (2.8) *implies that if X has increasing (decreasing) hazard rate then the residual Fisher information $I(f_e; t)$ is increasing (decreasing) in t.*

Corollary 2.2. *From Theorem* 2.3, *we have $I(f_e; t) = c$, where c is positive constant if and only if X is exponential distribution. The proof of "if" part is trivial. To prove the "only if" part note that when $I(f_e; t) = c$, we get, for all $t > 0$,*

$$\int_t^\infty r(x) f(x)\, dx = c \int_t^\infty \bar{F}(x)\, dx.$$

Derivative of both sides of this equation implies that $r(t) = \sqrt{c}$, which in turn implies that X is exponential with mean $1/\sqrt{c}$.

The following lemma gives a useful property of the score function which is in fact the time-dependent version of Stein identity for nonnegative random variables (see Johnson (2004), p. 22).

Lemma 2.1. *Let X be a nonnegative random variable with density f and score function $\rho(x) = \frac{f'(x)}{f(x)}$. For any function g such that $\lim_{x \to b} g(x) f(x) = 0$, we have*

$$E(g(X)\rho(X)|X > t) = -(r(t)g(t) + E(g(X)'|X > t)),$$

where $r(t)$ is the hazard rate of X.

Proof.

$$E\big(g(X)\rho(X)|X > t\big) = \frac{1}{\bar{F}(t)} \int_t^b g(x)\rho(x)f(x)\,dx$$

$$= \frac{1}{\bar{F}(t)} \int_t^b g(x)f'(x)\,dx$$

$$= -\big(r(t)g(t) + E\big(g'(X)|X > t\big)\big). \qquad \square$$

The lemma gives the following immediate lower bound for $I(f; t)$ in terms of hazard rate.

Theorem 2.4. *Let the conditions of Lemma* 2.1 *are met. Given a random variable X with residual Fisher information $I(f; t)$ and hazard rate $r(t)$, we have*

$$I(f; t) \geq r^2(t),$$

where equality holds if and only if X is exponential.

Proof. Lemma 2.1 with $g(x) = 1$, gives $E[\rho(X)|X > t] = -r(t)$. Hence, we get

$$\mathrm{Var}\big(\rho(X)|X > t\big) = E\big[\rho^2(X)|X > t\big] - E^2\big[\rho(X)|X > t\big]$$

$$= I(f; t) - r^2(t) \geq 0.$$

Using equation (2.3) an immediate consequence of the theorem is that, for all $t > 0$,

$$\bar{F}(t) \geq e^{-\int_0^t \sqrt{I(f;x)}\,dx},$$

where, under the conditions of theorem, the equality holds if and only if X is exponential. $\qquad \square$

The quantity $R(x) = \int_0^x r(t)\,dt = -\log \bar{F}(x)$ in known in reliability engineering as the cumulative hazard rate and plays important role in the study of aging properties of lifetime random variables. In the following, we get some results which show, under some smooth conditions, that the Fisher information can be represented through a function of the the hazard rate and the covariance between the score function and cumulative hazard.

Lemma 2.2. *We have*

$$E\big[\rho^k(X)R(X)\big] = \int_0^b f(u)E\big[\rho^k(X)|X > u\big]\,du.$$

Proof.

$$E[\rho^k(X)R(X)] = \int_0^b \rho^k(x)R(x)f(x)\,dx$$

$$= \int_0^b \left(\int_0^x r(u)\,du \right) \rho^k(x)f(x)\,dx$$

$$= \int_0^b r(u) \left(\int_u^b \rho^k(x)f(x)\,dx \right) du$$

$$= \int_0^b f(u)E[\rho^k(X)|X > u]\,du. \qquad \square$$

The lemma gives the following special cases. For $k = 1$, under the condition of Lemma 2.1 as $E[\rho(X)|X > u] = -r(u)$, we get

$$E[\rho_f(X)R(X)] = -E[r(X)].$$

In particular, under the condition $\lim_{x\to 0} f(x) = 0$, we get $\mathrm{Var}(\rho_f(X)) = I(f)$. On the other hand since $E(R(X)) = \mathrm{Var}(R(X)) = 1$, we have

$$\sqrt{I(f)} = -\frac{E[r(X)]}{\mathrm{Corr}(\rho(X), R(X))},$$

where $\mathrm{Corr}(\rho(X), R(X))$ denotes the correlation between $\rho(X)$ and $R(X)$. This representation also shows that $\rho(X)$ and $R(X)$ are negatively correlated. For $k = 2$, based on the definition of $I(f;t)$, we obtain

$$E[\rho_f^2(X)R(X)] = \int_0^b f(u)E[\rho^2(X)|X > u]\,du$$

$$= E[I(f;X)].$$

Example 2.5. Let X have Weibull distribution with density

$$f(x) = \beta\lambda^\beta x^{\beta-1}\exp(-(\lambda x)^\beta), \qquad x > 0, \beta, \lambda > 0.$$

Then for $\beta > 2$, $I(f) = (\beta - 1)^2\lambda^{\frac{2}{\beta}}\Gamma(1 - \frac{2}{\beta})$ and also, $E[r(X)] = \beta\lambda E[X^{\beta-1}] = (\beta - 1)\lambda^{\frac{1}{\beta}}\Gamma(1 - \frac{1}{\beta})$, the correlation coefficient between R and ρ is

$$\mathrm{Corr}(\rho_f(X), R(X)) = -\frac{E[r(X)]}{\sqrt{I(f)}} = -\frac{\Gamma(1 - \frac{1}{\beta})}{\sqrt{\Gamma(1 - \frac{2}{\beta})}}.$$

3 Fisher information distance of residual life distributions

Assuming that $f_t(x)$ and $g_t(x)$ denote the density functions corresponding to residual random variables $(X|X > t)$ and $(Y|Y > t)$, respectively. We define the residual Fisher information distance (RFID) between the two residual distributions as:

Definition 3.1. The RFID between f_t and g_t is defined as

$$D(f, g; t) = \int_t^b \left(\frac{\frac{df_t(x)}{dx}}{f_t(x)} - \frac{\frac{dg_t(x)}{dx}}{g_t(x)} \right)^2 f_t(x)\, dx$$

$$= \frac{\int_t^b (\rho_f(x) - \rho_g(x))^2 f(x)\, dx}{\bar{F}(t)} \qquad (3.1)$$

$$= E\big((\rho_f(X) - \rho_g(X))^2 | X > t\big).$$

Clearly, RFID reduces to $D(f, g)$ in (1.3), if $t \to 0$.

Remark 3.1. One can easily see that there are two equivalent forms of representations for $D(f, g; t)$ as follows:

$$D(f, g; t) = 4 \frac{\int_t^b (\frac{d}{dx} \sqrt{\frac{f(x)}{g(x)}})^2 g(x)\, dx}{\bar{F}(t)} = \frac{\int_t^b (\frac{d}{dx} \log \frac{f(x)}{g(x)})^2 f(x)\, dx}{\bar{F}(t)}.$$

See also Johnson (2004).

In the following we give an example.

Example 3.1. Let X_i, $i = 1, 2$, be distributed as gamma with density

$$f_i(x) = \frac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)} x^{\alpha_i - 1} e^{-\beta_i x}, \qquad x > 0, \alpha_i, \beta_i > 0, i = 1, 2.$$

Then, it can be easily verified that for $\alpha_1 = \alpha_2$, we have $D(f_1, f_2; t) = (\beta_2 - \beta_1)^2$ which is free of t. In the case when, $\beta_1 = \beta_2$ and $\alpha_1 > 2$, we have

$$D(f_1, f_2; t) = (\alpha_2 - \alpha_1)^2 E\left(\frac{1}{X_1^2} \Big| X_1 > t \right)$$

$$= \beta_1^2 \frac{(\alpha_2 - \alpha_1)^2 \Gamma(\alpha_1 - 2, \beta_1 t)}{\Gamma(\alpha_1, \beta_1 t)},$$

where $\Gamma(\alpha_1, t) = \int_t^\infty x^{\alpha_1 - 1} e^{-x}\, dx$, is incomplete gamma function. It is clear that $D(f_1, f_2; t)$ is always a decreasing function of t. If $t = 0$, we obtain

$$D(f_1, f_2) = \beta_1^2 \frac{(\alpha_2 - \alpha_1)^2 \Gamma(\alpha_1 - 2)}{\Gamma(\alpha_1)}.$$

Weighted distributions arise naturally in different areas of statistics; (see Rao (1997)). Let f is a density function with support S and w is a function such that $E(w(X)) < \infty$. The weighted distribution associated to f is

$$g(x) = \frac{w(x) f(x)}{E(w(X))}, \qquad x \in S. \qquad (3.2)$$

The following theorem shows how the distance between the density f and corresponding weighted distribution g is related to weight w.

Theorem 3.1. *Let f be a density function and g be weighted distribution associated to f with weight w. Then*

$$D(f, g; t) = E\left(\left(\frac{\partial \log w(X)}{\partial X}\right)^2 \bigg| X > t\right). \tag{3.3}$$

Proof. The result follows from the fact that the score function of g is given by $\rho_g(x) = \frac{w'(x)}{w(x)} + \rho_f(x)$. □

Example 3.2. Let $g(x) = \frac{xf(x)}{E(X)}$ be the density of size-biased distribution corresponding to f. Then $D(f, g; t) = E(\frac{1}{X^2}|X > t)$. Also $D(g, f; t) = E(\frac{1}{X_g^2}|X_g > t)$, where X_g is distributed as g.

The following theorem shows a relation between Fisher information and Fisher information distance under some conditions.

Theorem 3.2. *Let X and Y have densities f and g and distribution functions F and G, respectively. If $I_{G(X)}(G(t))$ is the residual Fisher information of $G(X)$ at $G(t)$.*

(a) *If g is increasing then $I_{G(X)}(G(t)) \leq \frac{1}{g^2(t)} D(f, g; t)$.*
(b) *If g is decreasing then $I_{G(X)}(G(t)) \geq \frac{1}{g^2(t)} D(f, g; t)$.*

Proof. (a) From Theorem 2.1, and the assumption that g is increasing, we have

$$I_{G(X)}(G(t)) = \frac{1}{\bar{F}(t)} \int_t^{G^{-1}(b)} \frac{1}{g^2(x)}\left[\frac{f'(x)}{f(x)} - \frac{g'(x)}{g(x)}\right]^2 f(x)\, dx$$

$$\leq \frac{1}{\bar{F}(t)} \int_t^{G^{-1}(b)} \frac{1}{g^2(t)}\left[\frac{f'(x)}{f(x)} - \frac{g'(x)}{g(x)}\right]^2 f(x)\, dx$$

$$\leq \frac{1}{g^2(t)} D(f, g; t).$$

(b) Part (b) follows similar to part (a) from the assumption that g is decreasing. □

3.1 Distance between order statistics

Order statistics have applications in various directions such as statistical inference, reliability engineering, quality control, etc. If $X_1, \ldots, X_n$ is a random sample from an absolutely continuous distribution F with density f and survival function $\bar{F}$,

the density function of the kth order statistics, $X_{k:n}$, associated to sample is given by

$$f_{k:n}(x) = \frac{n!}{(k-1)(n-k)} f(x) F^{k-1}(x) \bar{F}^{n-k}(x).$$

This distribution can be considered as a weighted distribution of the form (3.2) with $w(x) = F^{k-1}(x) \bar{F}^{n-k}(x)$. Hence, the score function of $X_{k:n}$ is

$$\rho_{k:n}(x) = \rho_f(x) + (k-1)r_1(x) - (n-k)r_2(x),$$

where $\rho(x)$ is the score function, $r_1(x) = \frac{f(x)}{F(x)}$ is revered hazard rate and $r_2(x) = \frac{f(x)}{\bar{F}(x)}$ is the hazard rate of F, respectively. Then it can be shown that

- The RFID of the underlying distribution f and the distribution of kth order statistic is

$$D(f, f_{k:n}; t) = E\big([(k-1)r_1(X) - (n-k)r_2(X)]^2 | X > t\big).$$

  In particular, the RFID of f and smallest order statistic, $X_{1:n}$, is

$$D(f, f_{1:n}; t) = (n-1)^2 E\big([r_2(X)]^2 | X > t\big).$$

  If for example X has Weibull distribution with parameters β and λ, then for $\beta > \frac{2}{3}$

$$D(f, f_{1:n}; t) = (n-1)^2 E\big([r_2(X)]^2 | X > t\big) = \frac{(n-1)^2 \beta^2 \lambda^{\frac{2}{\beta}} \Gamma(3 - \frac{2}{\beta}, \lambda t^\beta)}{\Gamma(1, \lambda t^\beta)}.$$

  In this case $D(f, f_{1:n}; t)$ is decreasing for $\beta < 1$ and is increasing for $\beta > 1$. In special case when $\beta = 1$, f is exponential with parameter λ and we get $D(f, f_{1:n}; t) = [\lambda(n-1)]^2$ which is free of t.

  The RFID of f and largest order statistic, $X_{n:n}$, is

$$D(f, f_{n:n}; t) = (n-1)^2 E\big([r_1(X)]^2 | X > t\big).$$

  If, for example, X has power distribution with density $f(x) = \alpha x^{\alpha-1}$, $0 < x < 1$, then for $\alpha > 2$

$$D(f, f_{n:n}; t) = (n-1)^2 E\big([r_1(X)]^2 | X > t\big) = (n-1)^2 \frac{\alpha^3}{\alpha-2} \frac{1 - t^{\alpha-2}}{1 - t^\alpha}.$$

- The RFID of k_1th and k_2th order statistics in a sample of size n is

$$D(f_{k_1:n}, f_{k_2:n}; t) = (k_2 - k_1)^2 E\big([r_1(X_{k_1:n}) + r_2(X_{k_1:n})]^2 | X_{k_1:n} > t\big).$$

If we take $\phi(x) = \log \frac{F(x)}{\bar{F}(x)}$, then $\phi(x)$ is known as the log of odds corresponding to F. One can easily verify that $D(f_{k_1:n}, f_{k_2:n}; t)$ can also be represented as

$$D(f_{k_1:n}, f_{k_2:n}; t) = (k_2 - k_1)^2 E\left(\left(\frac{d \log \phi(X_{k_1:n})}{d X_{k_1:n}}\right)^2 \bigg| X_{k_1:n} > t\right).$$

- The RFID of kth order statistic in two samples of sizes n_1 and n_2 is

$$D(f_{k:n_1}, f_{k:n_2}; t) = (n_2 - n_1)^2 E\left(\left[r_2(X_{k:n_1})\right]^2 | X_{k:n_1} > t\right).$$

3.2 Results on escort distributions

Let X be a random variable with density function f. The random variable Y is said to be the escort random variable corresponding to X if for any positive real number c, the density function of Y is given by

$$g_c(x) = \frac{f^c(x)}{\int_0^b f^c(x)\, dx}, \qquad 0 < x < b, \tag{3.4}$$

provided that $\int_0^b f^c(x)\, dx < \infty$. The escort distribution $g_c(x)$ can also be represented as

$$g_c(x) = \frac{w(x)f(x)}{E[w(X)]},$$

where $w(x) = f^{c-1}(x)$. That is, $g_c(x)$ is a weighted distribution associated to f. The escort distributions are proved to have applications on different areas. Beck and Schlögl (1993) showed that escorts distributions play the key role for various analogies between chaos theory and thermodynamics. Bercher (2015) showed that escorts distribution can be applied in coding theory to modify the weights of the words when the words with low probabilities are important. The cited author has also proved that the escort distributions are appearing as solution of a minimum Kullback-liebler discrimination associated to state transition model framework.

Now, we propose the following definition.

Definition 3.2. Two random variables X and Y with score functions $\rho_f(x)$ and $\rho_g(x)$ are said to have proportional scores if $\rho_g(x) = c\rho_f(x)$, where c is positive constant.

Theorem 3.3. *Let X and Y be two absolutely continuous nonnegative random variables with score functions $\rho_f(x))$ and $\rho_g(x)$, respectively. Then X and Y have proportional scores if and only if Y is escort random variable associated to X.*

Proof. The proof of "if" part is straightforward. We prove the "only if" part. Assume that X and Y have proportional scores. That is, for $c > 0$,

$$\frac{g'(x)}{g(x)} = c\frac{f'(x)}{f(x)}.$$

Integrating both sides of this differential equation gives

$$\log g(x) = c\left[\log f(x) + k\right],$$

where we assume that k is a finite constant. This implies that

$$g(x) = f^c(x)e^{kc}.$$

Since f and g are density functions we should have $e^{kc} = (\int_0^b f^c(x)\,dx)^{-1}$. This completes the proof. □

It is easy to show that if g_c is escort distribution associated to f of the form of (3.4) then

$$D(f, g_c; t) = (c-1)^2 I(f, t).$$

This is so, because if (3.4) holds, $\rho_{g_c}(x) = c\rho_f(x)$, which in turn implies that

$$D(f, g_c; t) = \frac{\int_t^b (\rho_f(x) - \rho_{g_c}(x))^2 f(x)\,dx}{\bar{F}(t)}$$

$$= (c-1)^2 I(f, t),$$

where $\bar{F}(t)$ is the survival function of X. Similarly, it can be easily seen that $D(g_c, f; t) = \frac{(c-1)^2}{c^2} I(g_c, t)$.

Example 3.3. Suppose that the underlying distribution is exponential with density function $f(x) = \lambda \exp(-\lambda x)$, $x > 0$, $\lambda > 0$. Then the escort density function $g_c(x)$ is again exponential with density

$$g_c(x) = c\lambda \exp(-c\lambda x), \qquad x > 0. \tag{3.5}$$

We have $I(f, t) = \lambda^2$, and hence $I(g_c, t) = c^2\lambda^2$. Also $D(f, g_c; t) = c^2 D(g_c, f; t) = (c-1)^2\lambda^2$.

Example 3.4. Suppose that f is Gamma(α, β) with density

$$f(x) = \frac{1}{\Gamma(\alpha)}\beta^\alpha x^{\alpha-1}e^{-\beta x}, \qquad x > 0, \beta > 0, \alpha > 0.$$

It can be easily shown that the escort distribution g_c is Gamma$((c(\alpha-1)+1), c\beta)$ with density function

$$g_c(x) = \frac{1}{\Gamma(c(\alpha-1)+1)}(c\beta)^{c(\alpha-1)+1}x^{c(\alpha-1)}e^{-c\beta x}. \tag{3.6}$$

Hence for $\alpha > 2$ and $c(\alpha-1) > 1$, we have

$$I(f; t) = \frac{\beta^2((\alpha-1)^2\Gamma(\alpha-2, \beta t) - 2(\alpha-1)\Gamma(\alpha-1, \beta t) + \Gamma(\alpha, \beta t))}{\Gamma(\alpha, \beta t)},$$

$$I(g_c; t) = \frac{c^2\beta^2((c(\alpha-1))^2\Gamma(c(\alpha-1)-1, c\beta t) - 2(c(\alpha-1))\Gamma(c(\alpha-1), c\beta t))}{\Gamma(c(\alpha-1)+1, c\beta t)}$$

$$+ \frac{\Gamma(c(\alpha-1)+1, c\beta t)}{\Gamma(c(\alpha-1)+1, c\beta t)},$$

and

$$D(f, g_c; t)$$

$$= (c-1)^2 \frac{\beta^2((\alpha-1)^2\Gamma(\alpha-2, \beta t) - 2(\alpha-1)\Gamma(\alpha-1, \beta t) + \Gamma(\alpha, \beta t))}{\Gamma(\alpha, \beta t)},$$

In particular, at $t = 0$, we get $I(f) = \frac{\beta^2}{\alpha-2}$ and $I(g_c) = \frac{(c\beta)^2}{c(\alpha-1)-1}$. Also $D(f, g_c) = (c-1)^2 \frac{\beta^2}{\alpha-2}$.

In the following theorem, we give conditions under which the residual Fisher Information of f and its associated escort distribution are ordered in the sense of residual Fisher information.

Theorem 3.4. *Let X and Y be two random variables such that $\rho_g(x) = c\rho_f(x)$.*

(a) *If $f(x)$ and $\rho_f^2(x)$ are both increasing then, for $c > 1$, $I(f; t) \leq I(g; t)$.*
(b) *If $f(x)$ and $\rho_f^2(x)$ are both decreasing then, for $0 < c < 1$, $I(f; t) \geq I(g; t)$.*

Proof. (a) First note that the assumptions that $f(x)$ is increasing and $c > 1$ imply

$$\frac{g(x)}{f(x)} = \frac{f^{c-1}(x)}{\int_0^b f^c(x)},$$

is increasing in x, that is, $X \leq_{lr} Y$. This in turn, implies that $X \leq_{hr} Y$. Now, we have

$$I(f; t) = \frac{\int_t^b \rho_f^2(x) f(x)\, dx}{\bar{F}(t)}$$

$$\leq \frac{\int_t^b \rho_g^2(x) f(x)\, dx}{\bar{F}(t)}, \qquad c > 1$$

$$\leq \frac{\int_t^b \rho_g^2(x) g(x)\, dx}{\bar{G}(t)} = I(g; t),$$

where the second inequality follows from the assumptions that $\rho_g^2(x)$ is increasing and $X \leq_{hr} Y$ (see Theorem 2.2). Under the conditions that imposed, part (b) can be proved similar to part (a). $\square$

Example 3.5.

(a) Let X be distributed as $f(x) = ke^{x^2}$, $0 < x < 1$, where $k = (\int_0^1 e^{x^2}\, dx)^{-1}$ is normalizing constant. Then clearly $f(x)$ is increasing on $(0, 1)$. We have

$$g(x) = k(c)e^{cx^2}, \qquad 0 < x < 1,$$

where $k(c)$ is normalizing constant. Then $\rho_g^2(x) = (2cx)^2$, which is an increasing function of x. Hence from part (a) of the theorem we have, for $c > 1$, $I(f; t) \leq I(g_c, t)$.

(b) Let X be distributed as $f(x) = \frac{e^{-\sqrt{x}}}{2}$, $x > 0$. Then $f(x)$ is a decreasing function on $(0, \infty)$. The escort function corresponding to f is with $\rho_f^2(x) = \frac{1}{4x}$ which is decreasing on $(0, \infty)$. Hence for $c \in (0, 1)$, we have $I(f; t) \geq I(g; t)$.

4 Conclusions

In this paper, we considered a continuous nonnegative random variable X and introduced a time-dependent Fisher information, $I(f, t)$, for the density function of the residual random variable $X|X > t$, $t > 0$. We also proposed a time-dependent Fisher information distance, $I(f, g; t)$, between the densities of two residual lifetime random variables. We obtained the functional form of $I(f; t)$ and showed that the underlying distribution function F can be represented in terms of $I(f; t)$ and the score function associated to X. It was shown that in the class of nonnegative continuous random variables, the exponential distribution can be characterized as a distribution with constant $I(f; t)$. It was proved that when two random variables are ordered in the sense of likelihood ratio order then, under some conditions, the corresponding time-dependent Fisher information of the two random variables are ordered. We showed that the time-dependent Fisher information of the equilibrium distribution associated to X is closely related to the hazard rate and the mean residual life of X. The correlation coefficient between the cumulative hazard rate and the score function of X is investigated. It was shown that the time-dependent Fisher information distance $I(f, g; t)$ between a density f and associated weighted density g can be represented as the conditional expectation of square of the weight function. The Fisher information distance $I(f, g; t)$ between the distributions of order statistics were investigated. We also studied $I(f, g; t)$ in a class of distributions called escort distributions.

Acknowledgments

References

Barlow, R. E. and Proschan, F. (1981). *Statistical Theory or Reliability and Life Testing: Probability Models*. New York-Montreal, Que.-London: Holt, Rinehart and Winston, Inc. MR0438625

Beck, C. and Schlögl, F. (1993). Thermodynamics of chaotic systems: an introduction. *Cambridge Nonlinear Science Series* **4**. Cambridge: Cambridge University Press. MR1237638

Bercher, J. F. (2015). Entropies and entropic criteria, theory. In *Regularization and Bayesian Methods for Inverse Problems in Signal and Image Processing* (J. F. Giovanneli and J. Idier, eds.). New York: John Wiley.

Bercher, J. F. and Vignat, C. (2009). On minimum Fisher information distributions with restricted support and fixed variance. *Information Sciences* **179**, 3832–3842. MR2640775

Bobkov, S. G., Chistyakov, G. P. and Gtze, F. (2014). Fisher information and the central limit theorem. *Probability Theory and Related Fields* **159**, 1–59. MR3201916

Brown, L. D. (1982). A proof of the central limit theorem motivated by the Cramer–Rao inequality. *Statistics and Probability: Essays in Honor of C. R. Rao*, 141–148. MR0659464

Frieden, B. R. (2004). *Science from Fisher Information: A Unification*. Cambridge: Cambridge University Press. MR2069674

Johnson, O. (2004). *Information Theory and the Central Limit Theorem*. London: Imperial College Press. MR2109042

Johnson, O. and Barron, A. (2004). Fisher information inequalities and the central limit theorem theory. *Probability Theory and Related Fields* **129**, 391–409. MR2128239

Kostal, L., Lansky, P. and Pokora, O. (2013). Measures of statistical dispersion based on Shannon and Fisher information concepts. *Information Sciences* **235**, 214–223. MR3042297

Lehmann, E. L. and Casella, G. (1998). *Information Theory and the Central Limit Theorem*. New York: Springer.

Otto, F. and Villani, C. (2000). Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *Journal of Functional Analysis* **173**, 361–400. MR1760620

Papaioannou, T. and Ferentinos, K. (2005). On two forms of Fisher's measure of information. *Communications in Statistics Theory and Methods* **34**, 1461–1470. MR2168022

Pardo, L., Morales, D. and Taneja, I. J. (1995). Generalized Jensen difference divergence measures and Fisher measure of information. *Kybernetes* **24**, 15–28. MR1332449

Rao, C. R. (1997). *Statistics and Truth: Putting Chance to Work*. River Edge: World Scientific. MR1474730

Salicrú, M. and Taneja, I. J. (1993). Connections of generalized divergence measures with Fisher information matrix. *Information Sciences* **72**, 251–269. MR1230757

Shaked, M. and Shanthikumar, J. G. (2007). *Stochastic Orders*. New York: Springer Science and Business Media.

Shao, J. (2003). *Mathematical Statistics*. New York: Springer. MR2002723

Venkatesan, R. C. and Plastino, A. (2014). Legendre transform structure and extremal properties of the relative Fisher information. *Physics Letters A* **378**, 1341–1345. MR3191530

Walker, S. G. (2016). Bayesian information in an experiment and the Fisher information distance. *Statistics & Probability Letters* **112**, 5–9. MR3475481

Yamano, T. (2013). Phase space gradient of dissipated work and information: A role of relative Fisher information. *Journal of Mathematical Physics* **54**, 113301. MR3137037

Yáñez, R. J., Sánchez-Moreno, P., Zarzo, A. and Dehesa, J. S. (2008). Fisher information of special functions and second-order differential equations. *Journal of Mathematical Physics* **49**, 082104. MR2440691

Zegres, P. (2002). Some new results on the architecture, training process, and estimation error bounds for learning machines. Ph.D. dissertation, Univ. Arizona.

Department of Statistics
University of Isfahan
Isfahan 81744
Iran
E-mail: Omid.Kharazmi@vru.ac.ir

Department of Statistics
University of Isfahan
Isfahan 81744
Iran
and
School of Mathematics
Institute of Research in Fundamental
    Sciences (IPM)
P.O. Box 19395-5746
Tehran
Iran
E-mail: m.asadi@sci.ui.ac.ir