ABOUT SOME PROPERTIES OF THE EXPONENTIAL FAMILIES AND FISHER INFORMATION

Ion Mihoc and Cristina-Ioana Fătu

Abstract. Fisher information is a fundamental concept of statistical theory and plays an important role in many areas of statistical analysis. Importance of Fisher information as a measure of the information in a distribution is well known. In classical inference with a random sample, the Fisher information appears in the Cramer-Rao lower bound which is a fundamental limit of the variance of an unbiased or biased estimator.

Let $\mathbf{X} = (X_1, X_2, ..., X_n)$ be a sample from the population $P = \{P_\theta : \theta \in D_\theta\}-$ a parametric family, where D_θ is called the parameter space, $D_\theta \subset \mathbb{R}^k$ (k is some fixed positive integer, $k \geq 1$) and let $f(x;\theta)$ be the probability density function for some model of the data, which has parameter vector $\theta = (\theta_1, \theta_2, ..., \theta_k)$.

In this article, under certain regularity conditions, we discuss some properties of the Fisher information then we have in view a random variable X which belongs to the class of exponential dispersion models. This class, introduced by Jorgensen [2], include as a special case, the generalized liniar model families of Nelder and Wedderburn [4] as well as many standard families such as Normal, Gamma, Inverse Gaussian and others. Also, using a weight function, $w(x) \geq 0$, we analyse the Fisher information associeaed to $f_w(x;\theta)-$ the probability density function of the weighted distribution corresponding to the random variable X.

2000 *Mathematics Subject Classification*: 62B10, 62F10, 94A17.

Keywords and phases. Statistical estimation, score function, Fisher information, exponential dispersion family, weighted distributions.

Introduction

We recall that, in statistical inference, the data can be represented as a random element X (known as a theoretical random variable) with values in measurable space (Ω, K), where K is a $\sigma-$algebra. If the distribution P_θ of X is assumed to belong to a parametric family $P = \{P_\theta : \theta \in D_\theta\}$, then the triple (Ω, K, P) will be a statistical model. Also, the data set, $\mathbf{x} = (x_1, x_2, ..., x_n)$, is viewed as a realization of this random element defined on a probability space (Ω, K, P) or as a realization of the random sample vector $\mathbf{X} = (X_1, X_2, ..., X_n)$. Because, generally, the goal is to use the **data x** to get information on the unknown value of the parameter θ or on $g(\theta)$, where $g : D_\theta \subset \mathbb{R}^k$ is a parametric function, in the next, we consider a family of probability density functions $\{f(x; \theta) : \theta \in D_\theta\}$, where $D_\theta \subset \mathbb{R}^k$, $(k \geq 1)$.

In the next, the **parameter space** D_θ can either be an **open subset** of the **real line** $\mathbb{R}$ if $k = 1$ or an open subset of $n-$**dimensional Euclidian space** $\mathbb{R}^k$.

Let $\mathbf{X} = (X_1, X_2, ..., X_n)$ be a random sample of the size n from a population characterized by the parameter θ and density function $f(x; \theta)$, $\theta \in D_\theta$, where D_θ is an open subset of the **real line** $\mathbb{R}$ (*i.e.*, $k = 1$) and a **statistic**

$$T(\mathbf{X}) = T(X_1, X_2, ..., X_n) \tag{1.1}$$

which is a function of the random sample variables $X_1, X_2, ..., X_n$ that does not depend upon any **unknown** parameter θ. Evidently, in **the**"**suppositional optics**", this statistic $T(\mathbf{X})$ is a random variable what can be used as an "**approximation**" for the parametric function $g(\theta)$.

Thus, if $x_1, x_2, ..., x_n$ are the observed experimental values of $X_1, X_2, ..., X_n$, then the real number $y = T(\mathbf{x}) = T(x_1, x_2, ..., x_n)$ can be a good **point estimate** of θ or of the parametric function $g(\theta)$ and $T(\mathbf{X})$ can be a good **point estimator** of θ or $g(\theta)$.

From a probabilistic point of view, the "**information**" within the statistic $T(\mathbf{X})$ (concerning the unknown distribution of $\mathbf{X}$, i.e., with respect to the unknown parameter θ) is contained in $\sigma(T(\mathbf{X})-$the $\sigma-$field generated by the statistic $T(\mathbf{X})$.

2. Score functions and Fisher's information measures

Let X be a continuous random variable with the probability density function $f(x; \theta)$,where $\theta = (\theta_1, \theta_2, ..., \theta_k)$, $\theta \in \mathbf{D}_\theta \subset \mathbb{R}^k, k \geq 1$.

In the next, we consider the random vector

$$S_n(X) = \mathbf{X} =(X_1, X_2, ..., X_n), \tag{2.1}$$

which represents a random sample of size n, where the components $X_i, i = \overline{1, n}$, are **random variables statistically independent and identically**

distributed as the theoretical random variable X, that is, we have

$$f(x;\theta) = f(x_i;\theta);\ \ i=\overline{1,n}, \theta \in \mathbf{D}_\theta. \tag{2.2}$$

Let

$$L_n(x_1,x_2,...,x_n;\theta_1,\theta_2,...,\theta_k) = L_n(\theta;\mathbf{x}) = \prod_{i=1}^{n} f(x_i;\theta_1,\theta_2,...,\theta_k) \tag{2.3}$$

be the joint probability density for the random sample $S_n(X)$ which, viewed as a function of the unknown parameter θ given $\mathbf{x}$, is the **likelihood function** and

$$\ln L_n(\theta;\mathbf{x}) = \sum_{i=1}^{n} \ln f(x_i;\theta_1,\theta_2,...,\theta_k) \tag{2.4}$$

is the **log-likelihood function** corresponding to $L_n(\theta;\mathbf{x})$, where $\mathbf{x} = (x_1,x_2,...,x_n)$ belongs to the **selection space** $\mathbb{R}^n$ and $\theta = (\theta_1,\theta_2,...,\theta_k) \in \mathbf{D}_\theta$.

Remark 2.1. Because the parametric measures of information are applicable to regular families of probability distributions, in the next, for the function $f(x;\theta)$, as well as for the likelihood function $L_n(\theta_1,\theta_2,...,\theta_k;\mathbf{x})$, which are the probability density functions, we assume that are satisfied the following **Fisher information regularity conditions (FIRCs)**:

$\mathbf{R}_1$) The set $\{x : f(x,\theta) > 0\}$ is the same for all $x \in \Omega$ and all $\theta \in D_\theta$;

$\mathbf{R}_2$) $\frac{\partial}{\partial\theta_i}[f(x,\theta)]$ exists for all $x \in \Omega$, all $\theta \in D_\theta$ and all $i = \overline{1,k}$;

$\mathbf{R}_3$) $\frac{\partial}{\partial\theta_i}\left[\int_A f(x,\theta)dx\right] = \int_A \frac{\partial}{\partial\theta_i}[f(x,\theta)]dx$ for any $A, A \subset K$, all $\theta \in D_\theta$ and all $i = \overline{1,k}$;

$\mathbf{R}_4$) $\int_A \frac{\partial}{\partial\theta_i\partial\theta_j}\left[f(x,\theta)\right]dx < \infty$ for any $A, A \subset K$, all $\theta \in D_\theta$ and all $i = \overline{1,k}$.

Definition 2.1. *If for the probability distribution $f(x;\theta)$ (either discrete or continuous), that depends on the parameter θ, $\theta \in D_\theta \subset \mathbb{R}^k, (k \geq 1)$, are satisfied the* **FIRCs**, *then $L_n(\mathbf{x};\theta)$ is a differentiable function and the function $\mathbf{U} : \mathbb{R}^n \longrightarrow \mathbb{R}$ as*

$$\mathbf{U}(\theta;\mathbf{x}) = \mathbf{U} = \frac{\partial \log L_n(\theta;\mathbf{x})}{\partial\theta} = \frac{1}{L_n(\theta;\mathbf{x})}\frac{\partial L_n(\theta;\mathbf{x})}{\partial\theta} \tag{2.5}$$

is called **the score of the sample** $S_n(X)$ *or*, *simply*, **the score function with respect to** $\theta \in D_\theta \subset \mathbb{R}^k$.

Remark 2.2. Because $\theta = (\theta_1, \theta_2, ..., \theta_k) \in \mathbf{D}_\theta \subset \mathbb{R}^k$ the score function $\mathbf{U} \equiv \mathbf{U}(\theta; \mathbf{x})$ will be an $k-$dimensional random vector as

$$\mathbf{U} = (U_1, U_2, ..., U_k)^\intercal, \tag{2.6}$$

where

$$U_j = \frac{\partial \ln L_n(\theta; \mathbf{X})}{\partial \theta_j} = \frac{\partial \ln L_n(\theta_1, \theta_2, ..., \theta_k; \mathbf{X})}{\partial \theta_j}, j = \overline{1,k} \tag{2.7}$$

represents the **score of the sample with respect to the parameter** θ_j, $\theta_j \in D_j \subset \mathbb{R}$, $j = \overline{1,k}$.

We can remark that **,** in the "**suppositional optics**",all these **score functions** are random variables because, for a given value of θ, the score functions depend on the sample. Also, the score function can be interpreted as: the value of the score of the sample is a measure of the sensitivity of the sample log-likelihood to small changes of the value of θ. If the value of the score is small for a given value of θ, the likelihood of the sample (that is, its probability density) will be essentially unaffected by small changes of θ.

Lemma 2.1. [4] *Under the FIRCs the expectation of the score function has the value zero* (*or the score function is centred*), *that is*

$$E_\theta[\mathbf{U}(\theta; \mathbf{X})] = E_\theta[\frac{\partial}{\partial \theta}[\ln L_n(\theta; \mathbf{X})] = \mathbf{0} \textit{ for all } \theta \in \mathbf{D}_\theta, \tag{2.8}$$

where $\theta = (\theta_1, \theta_2, ..., \theta_k) \in D_\theta \subset \mathbb{R}^k$ *and* E_θ *represents expectation with respect to the distribution determined by* θ.

Lemma 2.2.[4] *Under the FIRCs*, *the second moment of the score function*, *when* $k = 1$ (*i.e.*, $\theta \in D_\theta \subset \mathbb{R}$), *has the following property*

$$E_\theta \left\{ \left[\frac{\partial \ln L_n(\theta; \mathbf{X})}{\partial \theta} \right]^2 \right\} = -E_\theta \left\{ \frac{\partial^2 \ln L_n(\theta; \mathbf{X})}{\partial \theta^2} \right\} \textit{ for all } \theta \in D_\theta \subset \mathbb{R}. \tag{2.9}$$

Corollary 2.1. *The variance of the score function has the following expresion* :

$$Var\left[\mathbf{U}(\theta;\mathbf{X})\right]=Var\ \left[\frac{\partial\log L_n(\theta;\mathbf{X})}{\partial\theta}\right]= \tag{2.10}$$

$$=E_\theta\left[\frac{\partial\log L_n(\theta;\mathbf{X})}{\partial\theta}.\frac{\partial\log L_n(\theta;\mathbf{X})}{\partial\theta}\right]= \tag{2.11}$$

$$=-E_\theta\left[\frac{\partial^2\log L_n(\theta;\mathbf{X})}{\partial\theta\partial\theta^\intercal}\right]\ \ if\ \theta\in\mathbf{D}_\theta\subset\mathbb{R}^k,\ k>1, \tag{2.12}$$

respectively

$$Var\left[\mathbf{U}(\theta;\mathbf{X})\right]\ =Var\ \left[\frac{\partial\log L_n(\theta;\mathbf{X})}{\partial\theta}\right]= \tag{2.13}$$

$$=-E_\theta\left[\frac{\partial^2\log L_n(\theta;\mathbf{X})}{\partial\theta^2}\right]\ \ if\ \theta\in\mathbf{D}_\theta\subset\mathbb{R},\ k=1. \tag{2.14}$$

Definition 2.2. *The quantity*

$$\mathbf{I}_1(\theta)=E_\theta\left\{\left[\frac{\partial\ln f(\theta;X)}{\partial\theta}\right]^2\right\}=\int_{-\infty}^{+\infty}\left[\frac{\partial\ln f(\theta;x)}{\partial\theta}\right]^2 f(\theta;x)dx= \tag{2.15}$$

$$=-\int_{-\infty}^{+\infty}\left[\frac{\partial^2\ln f(\theta;x)}{\partial\theta^2}\right]f(\theta;x)dx=-E_\theta\left\{\left[\frac{\partial^2\ln f(X;\theta)}{\partial\theta^2}\right]\right\}, \tag{2.16}$$

where $\theta\in D_\theta\subset\mathbb{R}$, represents the **Fisher information** *which measures the information about the univariate unknown parameter θ which is contained in an observation of the random variable X.*

Definition 2.3. *The quantity $\mathbf{I}_n(\theta)$, defined by the relation*

$$\mathbf{I}_n(\theta)=E_\theta\left\{\left[\frac{\partial\ln L_n(\theta;X)}{\partial\theta}\right]^2\right\}=\int_{\mathbb{R}^n}\left[\frac{\partial\ln L_n(\theta;x)}{\partial\theta}\right]^2 L_n(\theta;\mathbf{x})\mathbf{dx}= \tag{2.17}$$

$$=-\int_{\mathbb{R}^n}\left[\frac{\partial^2\ln L_n(\theta;\mathbf{x})}{\partial\theta^2}\right]L_n(\theta;\mathbf{x})\mathbf{dx}=-E_\theta\left\{\left[\frac{\partial^2\ln L_n(\theta;\mathbf{x})}{\partial\theta^2}\right]\right\}, \tag{2.18}$$

where $\theta \in D_\theta \subset \mathbb{R}$, *represents the Fisher information measure which measures the information about univariate unknown parameter* θ *contained in a random sample* $S_n(X) = (X_1, X_2, ..., X_n)$.

Remark 2.3. The **Fisher information** has several **good properties**, namely:

$\mathbf{1}^0$ **– non – negativity,** *i.e.*, $\mathbf{I}(X;\theta) \geq 0$ *and is* 0 *only when* $f(X;\theta)$ *is free of* θ;

$\mathbf{2}^0$ **– additivity** *and* **subadditivity**, *i.e.*, $\mathbf{I}(X,Y;\theta) \leq \mathbf{I}(X;\theta) + \mathbf{I}(Y;\theta)$, *with equality if* X, Y *are independent*;

$\mathbf{3}^0$ **– maximal information**, *i.e.*, $\mathbf{I}[\mathbf{T}(\mathbf{X});\theta] \leq \mathbf{I}(X;\theta)$ *for any statistic* $\mathbf{T}$, *and equality holds if and only if* $\mathbf{T}(\mathbf{X})$ *is sufficient*;

$\mathbf{4}^0$ **– convexity**, *i.e., if* X_i *has probability density function* f_i, $i = 1, 2$, *and* Y *has probability density function* $f(y) = \alpha f_1(x) + (1-\alpha) f_2(x)$; $0 \leq \alpha \leq 1$, *then* $\mathbf{I}(Y;\theta) \leq \alpha \mathbf{I}(X_1;\theta) + (1-\alpha)\mathbf{I}(X_2;\theta)$.

3. Exponential family and Fisher information

3.1. Exponential dispersion family

The early development of exponential dispersion models is often attributed to Tweedie, M.C.K. (1947) although a more thorough and systematic investigation of its statistical properties was done by Jorgensen, B. (1997).

Definition 3.1. *The random variable X is sed to belong to the* **Exponential Dispersion Family** (**EDF**) *of distribution if its probability measure* $P_{\theta,\lambda}$ *is absolutely continuous with respect to some measure* Q_λ *and can represented as*

$$f(x;\theta,\lambda) = \exp\left\{\lambda[\theta x - k(\theta)]\right\} q_\lambda(x),\ \ x \in S \subset \mathbb{R}, \tag{3.1.1}$$

where :

• *the parameter* θ *is named the* **canonical parameter**, $\theta \in D_\theta = \{\theta \in \mathbb{R} \mid k(\theta) \mid < \infty\}$;

• *the parameter* λ *is a* **dispersion** (*or index*) **parameter**, $\lambda \in D_\lambda = \{\lambda \mid \lambda > 0\} = \mathbb{R}_+$;

• *the function* $k(\theta)$ *is named the* **cumulant function**;

• *the function* $q_\lambda(x)$ *is the Radon – Nikodim derivative fo the measure* Q_λ, *i.e.*, $q_\lambda(x) = \frac{dQ_\lambda}{dx} > 0$.

The representation in (3.1.1) is called the **reproductive form of EDF** and we shall denote by $X \sim \mathbf{ED}(\theta,\lambda)$ for a random variable belonging to this family.

Theorema 3.1.1. *If X is a random variable distributed according to $P_{\theta,\lambda}$, then*

$$\mu = \mu(\theta) = E_{\theta,\lambda}(X) = \int_S xf(x;\theta,\lambda)dx = k'(\theta), \tag{3.1.2}$$

and

$$V_{\theta,\lambda}(X) = Var(X) = \frac{1}{\lambda}k''(\theta) = Var(\mu)\sigma^2, \tag{3.1.3}$$

where

$$Var(\mu) = k''(\theta) \tag{3.1.3a}$$

is called the **variance function** *and*

$$\sigma^2 = \frac{1}{\lambda} \tag{3.1.3b}$$

is called the **dispersion parameter**.

Proof. If we consider the reproductive form of **EDF** then its cumulative generating function can be derived as follows:

$$\begin{aligned} K_X(t) = \log_e E(e^{Xt}) = \log\left\{\int_S e^{xt}e^{\lambda[\theta x - k(\theta)]}q_\lambda(x)dx\right\} = \\ = \log_e\left\{\int_S e^{\{\lambda[(\theta+t/\lambda)x-k(\theta)]\}}q_\lambda(x)dx\right\} = \\ = \log_e\left\{\int_S e^{A}q_\lambda(x)dx\right\}. \end{aligned} \tag{3.1.4}$$

Because the exponent A can be written as

$$A=\lambda\left[\left(\theta+\frac{t}{\lambda}\right)x-k(\theta)\right]=$$
$$=\lambda\left[\left(\theta+\frac{t}{\lambda}\right)x+k\left(\theta+\frac{t}{\lambda}\right)-k\left(\theta+\frac{t}{\lambda}\right)-k(\theta)\right]=$$
$$=\lambda\left[k\left(\theta+\frac{t}{\lambda}\right)-k(\theta)\right]+\lambda\left[\left(\theta+\frac{t}{\lambda}\right)x-k\left(\theta+\frac{t}{\lambda}\right)\right], \tag{3.1.4a}$$

the above function $K_X(t)$ can be expressed as

$$K_X(t)=\log_e\left\{\int_S e^{\lambda\left[k\left(\theta+\frac{t}{\lambda}\right)-k(\theta)\right]}e^{\lambda\left[\left(\theta+\frac{t}{\lambda}\right)x-k\left(\theta+\frac{t}{\lambda}\right)\right]}q_\lambda(x)dx\right\}=$$
$$=\log_e\left\{e^{\lambda\left[k\left(\theta+\frac{t}{\lambda}\right)-k(\theta)\right]}.\underbrace{\int_S e^{\lambda\left[\left(\theta+\frac{t}{\lambda}\right)x-k\left(\theta+\frac{t}{\lambda}\right)\right]}q_\lambda(x)dx}_{=1(see\ (3.1.1)}\right\}=$$
$$=\log_e e^{\lambda\left[k\left(\theta+\frac{t}{\lambda}\right)-k(\theta)\right]}=\lambda\left[k\left(\theta+\frac{t}{\lambda}\right)-k(\theta)\right],$$

that is

$$K_X(t)=\lambda\left[k\left(\theta+\frac{t}{\lambda}\right)-k(\theta)\right] \tag{3.1.5}$$

and the **moment generating function** can be written as

$$M_X(t)=e^{\lambda\left[k\left(\theta+\frac{t}{\lambda}\right)-k(\theta)\right]}=e^{K_X(t)}. \tag{3.1.6}$$

Using the relations (3.1.5), respectively (3.1.6), we obtain relations

$$\frac{\partial K_X(t)}{\partial t}=\frac{\partial}{\partial t}\left\{\lambda\left[k\left(\theta+\frac{t}{\lambda}\right)-k(\theta)\right]\right\}=$$
$$=\lambda\left[k'\left(\theta+\frac{t}{\lambda}\right).\frac{1}{\lambda}\right]=k'\left(\theta+\frac{t}{\lambda}\right), \tag{3.1.5a}$$

$$\frac{\partial^2 K_X(t)}{\partial t^2}=k\text{''}\left(\theta+\frac{t}{\lambda}\right).\frac{1}{\lambda}, \tag{3.1.5b}$$

respectively relations

$$\frac{\partial M_X(t)}{\partial t} = \frac{\partial}{\partial t}\left[\exp\left\{K_X(t)\right\}\right] = \frac{\partial K_X(t)}{\partial t}\exp\left\{K_X(t)\right\} \tag{3.1.6a}$$

$$\begin{aligned}\frac{\partial^2 M_X(t)}{\partial t^2} &= \frac{\partial}{\partial t}\left[\frac{\partial K_X(t)}{\partial t}\exp\left\{K_X(t)\right\}\right] = \\ &= \left[\frac{\partial^2 K_X(t)}{\partial t^2} + \left(\frac{\partial K_X(t)}{\partial t}\right)^2\right]\exp\left\{K_X(t)\right\}.\end{aligned} \tag{3.1.6b}$$

Now, using the property

$$\left.\frac{\partial^r M_X(t)}{\partial t^r}\right|_{t=0} = E(X^r), r \in \{0, 1, 2, ...\}, \tag{3.1.7}$$

and the fact that $K_X(0) = 0$, from the last four relations, we get

$$\left.\frac{\partial M_X(t)}{\partial t}\right|_{t=0} = \left.\frac{\partial K_X(t)}{\partial t}\right|_{t=0} = k'\left.\left(\theta + \frac{t}{\lambda}\right)\right|_{t=0} = k'\left(\theta\right) = E(X) = \mu, \tag{3.1.7a}$$

that is, the mean of X has the form

$$\mu = E(X) = k'\left(\theta\right). \tag{3.1.8}$$

Also, we obtain

$$\begin{aligned}\left.\frac{\partial^2 M_X(t)}{\partial t^2}\right|_{t=0} &= E(X^2) = \left.\left[\frac{\partial^2 K_X(t)}{\partial t^2} + \left(\frac{\partial K_X(t)}{\partial t}\right)^2\right]\right|_{t=0} = \\ &= \left.\frac{\partial^2 K_X(t)}{\partial t^2}\right|_{t=0} + \left.\left(\frac{\partial K_X(t)}{\partial t}\right)^2\right|_{t=0} = \\ &= \left.\left[k''\left(\theta + \frac{t}{\lambda}\right)\cdot\frac{1}{\lambda}\right]\right|_{t=0} + \left.\left(k'\left(\theta + \frac{t}{\lambda}\right)\right)^2\right|_{t=0} = \\ &= \frac{1}{\lambda}k''\left(\theta\right) + \left(k'\left(\theta\right)\right)^2,\end{aligned}$$

that is, we have a new relation

$$E(X^2) = \frac{1}{\lambda}k''(\theta) + (k'(\theta))^2 = \frac{1}{\lambda}k''(\theta) + \mu^2, \tag{3.1.9}$$

respectively relation

$$Var(X) = \frac{1}{\lambda}k''(\theta) = \sigma^2 k''(\theta), \ where\ \sigma^2 = \frac{1}{\lambda}. \tag{3.1.10}$$

Remark 3.1.1. Notice that we can view the mean μ as a function of θ, *i.e.*,

$$\mu = E(X) = \tau(\theta) = k'(\theta) \tag{3.1.11}$$

so that

$$\theta = \tau^{-1}(\mu). \tag{3.1.12}$$

and if we define the **unit variance function** (or the **variance function**) as

$$Var(\mu) = k"(\theta) = k"[\tau^{-1}(\mu)], \tag{3.1.13}$$

then $Var(X)$ can be represented as

$$Var(X) = \sigma^2 V(\mu), \tag{3.1.10a}$$

where $\sigma^2 = \frac{1}{\lambda}$ is the **dispersion parameter.** Therefore, in the next, we will can write that $X \sim$ **EDF**(θ,σ^2).

Remark 3.1.2. The interest in the **EDF** was given by Jorgensen, who outline the **EDF** as one of the main classes of dispersion models, which includes most standard distribution families such that as Normal (an example of a symmetric distribution), Gamma, Inverse Gaussian (examples of non-symmetric and no-negative defined distributions), for the absolutely continuous case, and the Poisson, Binomial, and Negative Binomial for the discrete case.

Thus, if $X \sim \mathbf{N}(\mu,\sigma^2)$ is a normal random variable with mean μ and variance σ^2 then its probability density function can be written as

$$\begin{aligned} f(x;\mu,\sigma^2) &= \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] = \\ &= \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{x^2}{2\sigma^2}\right).\exp\left[\frac{1}{\sigma^2}(\mu x - \frac{1}{2}\mu^2)\right]. \end{aligned} \tag{3.1.14}$$

One can easily see that it belongs to the additive **EDF** by choosing:

$$\mu = \theta, \lambda = \frac{1}{\lambda}\sigma^2, k(\theta) = \frac{1}{2}\mu^2 = \frac{1}{2}\theta^2 \text{ and } q_\lambda(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{x^2}{2\sigma^2}\right). \tag{3.1.14a}$$

3.2 Fisher information in weighted distributions

Let $\mathbf{X} = (X_1, X_2, ..., X_n)$ be a sample from the population $P = \{P_\theta : \theta \in D_\theta\}-$ a parametric family, where D_θ is called the parameter space, $D_\theta \subset \mathbb{R}$. Standard inference procedures assume a such random sample from the population P with the probability density function $f(x;\theta)$ for estimating the parameter unknown θ.Using a weight function , $w(x) > 0$, to model ascertainment bias, Fisher (1934) constructed a weighted distribution with a probability density function $f_w(x)$ that is proportional to $w(x)f(x;\theta)$. In this paper, we study some properties of the Fisher information about the parameter θ using the observations obtained from some weighted distributions.

Definition 3.2.1.[1] *If the probability density function $f(x;\theta)$ of the random variable X belongs to the following exponential family of distributions*

$$f(x;\theta) = a(x).\exp\{\theta T(x) - C(\theta),\ \theta \in D_\theta \subset \mathbb{R}, \tag{3.2.1}$$

then, using a known weight function, $w(x;\theta) > 0$, a random variable Y will have a **weighted distribution** *if its probability density function, denoted by $f^w(y;\theta)$, has the form*

$$f^w(y;\theta) = \left[\frac{w(y;\theta)}{E_\theta[w(X;\theta)]}\right] f(y;\theta) = \tag{3.2.2}$$

$$= \left[\frac{w(y;\theta).a(y).\exp\{\theta T(y) - C(\theta)}{E_\theta[w(X;\theta)]}\right], \theta \in D_\theta \subset \mathbb{R}, \tag{3.2.3}$$

where the expectation $E_\theta[w(X;\theta)]$ is assumed to exist, i.e.,

$$E_\theta[w(X;\theta)] = \int_0^\infty w(x;\theta)f(x;\theta)dx < \infty. \tag{3.2.4}$$

The extension proof of the next theorem is based both on the above definition, on the sketck which was presented in the paper [1] as well as on the regularity conditions which was mentioned in the Remark 2.1 of this paper.

Theorem 3.2.1.The Fisher information $I_X(\theta)$ *based on single observation* X, *with the probability density function* $f(x;\theta)$, *defined by the relation*

$$I_X(\theta) = -E_\theta\left[\frac{d^2\log f(X;\theta)}{d\theta^2}\right], \tag{3.2.5}$$

can be expressed in the form

$$I_X(\theta) = \frac{d^2C(\theta)}{d\theta^2} = C''(\theta), \tag{3.2.6}$$

respectively in the form

$$I_Y(\theta) = I_X(\theta) + \frac{d^2}{d\theta^2}\left\{\log E_\theta[w(X;\theta)]\right\} - E_\theta\left\{\frac{d^2}{d\theta^2}\left[\log_\theta w(Y;\theta)\right]\right\}, \tag{3.2.7}$$

in the case of the random variable X, *where* E_θ *represents expectation with respect to the distribution determined by* θ.

Proof. Indeed, from (3.2.1), we obtain relation

$$\log f(x;\theta) = \log a(x) + \theta T(x) - C(\theta),\ \theta \in D_\theta \subset \mathbb{R}, \tag{3.2.8}$$

respectively, relations

$$u_X(\theta) = \frac{d}{d\theta}\left[\log f(x;\theta)\right] = T(x) - C'(\theta) \tag{3.2.8a}$$

and

$$\frac{d}{d\theta}[u_X(\theta)] = \frac{d^2}{d\theta^2}\left[\log f(x;\theta)\right] = -C''(\theta) \tag{3.2.8b}$$

which represent the first and the second derivative of the log-likelihood function (3.2.1).

Then, using (3.2.5), from (3.2.8b), we get

$$-E_\theta\left[\frac{d^2\log f(X;\theta)}{d\theta^2}\right] = I_X(\theta = C''(\theta). \tag{3.2.8c}$$

Analogous, using the probability density function $f^w(y;\theta)$, defined in (3.2.3), we obtain for the log-likelihood function the following form

$$\log f^w(y;\theta) = \log w(y;\theta) + \log a(y) + \theta T(y) - C(\theta) - \log E_\theta[w(X;\theta)]. \quad (3.2.9)$$

Also, the first and the second derivatives of this function can be expressed as

$$u_Y(\theta) = \frac{d}{d\theta}\left[\log f^w(y;\theta)\right] = \\ = \frac{d}{d\theta}\left[\log w(y;\theta)\right] + T(y) - C'(\theta) - \frac{d}{d\theta}\left\{\log E_\theta\left[w(X;\theta)\right]\right\} \quad (3.2.9a)$$

and

$$\frac{d}{d\theta}[u_Y(\theta)] = \frac{d^2}{d\theta^2}\left[\log f^w(y;\theta)\right] = \\ = \frac{d^2}{d\theta^2}\left[\log w(y;\theta)\right] - C''(\theta) - \frac{d^2}{d\theta^2}\left\{\log E_\theta\left[w(X;\theta)\right]\right\}. \quad (3.2.9b)$$

Now, using this last relation, one can easily see that

$$-E_\theta\left\{\frac{d}{d\theta}[u_Y(\theta)]\right\} = -E_\theta\left\{\frac{d^2}{d\theta^2}\left[\log f^w(y;\theta)\right]\right\} = I_Y(\theta) =$$

$$= \underbrace{C''(\theta)}_{I_X(\theta)} + E_\theta\left(\underbrace{\frac{d^2}{d\theta^2}\left\{\log E_\theta\left[w(X;\theta)\right]\right\}}_{a\ constant}\right) - E_\theta\left\{\frac{d^2}{d\theta^2}\left[\log w(Y;\theta)\right]\right\} = \\ = I_X(\theta) + \frac{d^2}{d\theta^2}\left\{\log E_\theta\left[w(X;\theta)\right]\right\} - E_\theta\left\{\frac{d^2}{d\theta^2}\left[\log w(Y;\theta)\right]\right\},$$

that is, the property (3.2.7) is holds.

References

[1] Iyengar, S., Kvam, P., Singh, H., *Fisher information in weighted distributions*, Canad. J. Statist. 27, 1999, pp. 833-841.

[2] Jorgensen, B., *Some properties of exponential dispersion models*, Journal of Royal Statistical Society, Series B, 49, 1987, pp. 127-162.

[3] Landsman, Z., Markov,U., *Exponential dispersion models and credibility*, Scandinavian Actuarial Journal, 1998, 1, pp.89-96.

[4] Mihoc, I., Fătu, C.I., *About some properties of the Fisher information in the sample theory*, Automation Computers Applied Mathematics, ISSN 1221-437X, Vol. 19 (2010) No. 2, pp.97-108.

[5] Rao, C.R., *Linear statistical inference and its applications*, John Wiley & Sons, Inc., New-York, 1965.

[6] Wilks, S.S., *Mathematical statistics*, John Wiley & Sons, Inc. New-York, London, 1962.

Ion Mihoc, Cristina-Ioana Fătu
Faculty of Economics, Christian University "Dimitrie Cantemir",
Cluj-Napoca,Str.T.Mihali nr.56, Romania
Email:*ion.mihoc@cantemircluj.ro*