STUDIA UNIV. "BABEŞ–BOLYAI", MATHEMATICA, Volume **XLVII**, Number 4, December 2002

RELATION BETWEEN THE AMOUNT OF INFORMATION AND THE LIKELIHOOD FUNCTION

ION MIHOC AND CRISTINA IOANA FĂTU

Dedicated to Professor D.D. Stancu on his 75^{th} *birthday*

Abstract. The objective of this paper is to give some properties for the Fisher information measure and as well as some relations and informational characterizations.

1. Introduction

The notion of information plays a central role both in the life of the person and of society, as well as in all kinds of scientific research. The notion of information is so universal, it penetrates our everyday life so much that from this point of view, it can be compared only with the notion of energy [5],[6].

The information theory is an important branch of probability theory and it has very much applications in mathematical statistics. The notion of information plays a central role in the fundamental statistical works of R.A.Fisher. Thus, e.g., Fisher characterized a sufficient statistical function by the fact that it exhausts all the information on the estimated parameter, contained by the sample.

Let X be a random variable on the probability space (Ω, K, P). A statistical problem arises when the distribution of X is not known and we want to draw some inference concerning the unknown distribution of X on the basis of a limited number of observations on X. A general situation may be described as follows: The functional form of the distribution function is known and merely the values of a finite number of parameters,involved in the distribution function, are unknown; i.e., the probability density function of the random variable X is known except for the value of a finite number of parameters. In general, the parameters $\theta_1, \theta_2, ..., \theta_k$ will not be subject to

Received by the editors: 29.10.2002.
2000 *Mathematics Subject Classification.* 94A17.
Key words and phrases. gain of information, Fisher information, local parameter.

any a priori restrictions; i.e., they may take any values. However, the parameters may in some cases be restricted to certain intervals. In the next we shell restrict ourselves to the case of a single parameter θ.

2. **Fisher's information measure**

Let X be a continuous random variable and its probability density function $f(x;\theta)$ depends on a parameter θ which values in a specified parameter space $D_\theta, D_\theta \subseteq \mathbf{R}$.Thus we are confronted, not with one distribution of probability, but with a family of distributions. To each value of θ, $\theta \in D_\theta$,there corresponds one member of the family. A family of probability density functions will be denoted by the symbol $\{f(x;\theta);\theta \in D_\theta\}$ .Any member of this family of probability density functions will be denoted by the symbol $f(x;\theta), \theta \in D_\theta$.

Let $S_n(X) = (X_1, X_2, ..., X_n)$ denote a random sample from a distribution that has a probability density function which is one member (but which member we do not known) of the family $\{f(x;\theta);\theta \in D_\theta\}$ of the probability density functions. That is, our sample arises from a distribution that has the probability distribution $f(x;\theta), \theta \in D_\theta$.Our problem is that of defining a statistic $T = T(X_1, X_2, ..., X_n)$, so that if $x_1, x_2, ..., x_n$ are the observed experimental values of $X_1X_2, ..., X_n$,then the number $t = t(x_1, x_2, ..., x_n)$ will be a good point estimate of θ.

In the next we suppose that the parameter θ is unknown and we estimate a specified function of $\theta, g(\theta)$ with the help of statistic $T = T(X_1, X_2, ..., X_n)$ which is based on a random sample $S_n(X) = (X_1, X_2, ..., X_n)$, where X_i are independent and identically distributed (*i.i.d.*) random variable with density $f(x;\theta), \theta \in D_\theta$.

A well known means of measuring the quality of the statistic

$$T = T(X_1, X_2, ..., X_n)$$

is to use the inequality of Cramér-Rao which states that, under certain regularity conditions for $f(x;\theta)$ (more particularly, it requires the possibility of differentiating under the integral sign) any unbiased estimator of $g(\theta)$ has variance which satisfies the following inequality [4]

$$VarT \geq \frac{[g^{'}(\theta)]^2}{n.I_X(\theta)} = \tag{2.1}$$

$$= \frac{[g^{'}(\theta)]^2}{I_n(\theta)}, \tag{2.1a}$$

where

$$I_X(\theta) = \int_{\Omega} \left(\frac{\partial \ln f(x;\theta)}{\partial \theta} \right)^2 f(x;\theta) dx = \tag{2.2}$$

$$= \int_{\Omega} \frac{1}{f(x;\theta)} \left(\frac{\partial f(x;\theta)}{\partial \theta} \right)^2 dx, \tag{2.3}$$

and

$$I_n(\theta) = E\left[\left(\frac{\partial \ln L(x_1, x_2, ..., x_n; \theta)}{\partial \theta} \right)^2 \right] = \tag{2.4}$$

$$= \int_{\Omega} ... \int_{\Omega} \left(\frac{\partial L(x_1, x_2, ..., x_n; \theta)}{\partial \theta} \right)^2 L(x_1, x_2, ..., x_n; \theta) dx_1 ... dx_n = \tag{2.5}$$

$$= nE\left[\left(\frac{\partial \ln f(x;\theta)}{\partial \theta} \right)^2 \right] = n \int_{\Omega} \left(\frac{\partial \ln f(x;\theta)}{\partial \theta} \right)^2 f(x;\theta) dx, \tag{2.6}$$

$$f(x;\theta) = f(x_i;\theta), i = \overline{1,n}, \tag{2.7}$$

$$L(x_1, x_2, ..., x_n; \theta) = \prod_{i=1}^{n} f(x_i;\theta) \tag{2.8}$$

is the joint probability density function of $X_1, X_2, ..., X_n$.

This joint probability density function of $X_1, X_2, ..., X_n$ may be regarded as a function of θ and it is called the likelihood function of the random sample $S_n(X) = (X_1 X_2, ..., X_n)$.

The quantity $I_X(\theta)$ is known as Fisher$'s$ information measure and it measures the information about $g(\theta)$ which is contained in an observation of X.Also, the quantity $I_n(\theta) = n.I_X(\theta)$ measures the information about $g(\theta)$ contained in a random sample $S_n(X) = (X_1 X_2, ..., X_n)$,than then X_i, $i = \overline{1,n}$ are independent and identically distributed random variables with density $f(x;\theta), \theta \in D_\theta$. An unbiased estimator of $g(\theta)$ that achieves this minimum from (2.1) is known as an efficient estimator.

3. **Some properties of Fisher's information measure**

Let $f(x;\theta), \theta \in D_\theta$ be a positive probability density function in the interval $[a,b]$ depending on the continuous parameter θ in a continuously differentiable way.

Definition 1. [6], [3]. The gain of information when a distribution with probability density function $f(x;\theta_0)$ is replaced by another one with probability density function $f(x;\theta_1)$ has the form

$$I[f(x;\theta_1)\|f(x;\theta_0)] = I(\theta_1\|\theta_0) = \tag{3.1}$$

$$= \int_a^b f(x;\theta_1)\log_2 \frac{f(x;\theta_1)}{f(x;\theta_0)}dx. \tag{3.2}$$

Theorem 1. *Let X be a continuous random variable with probability density function $f(x;\theta), \theta \in D_\theta$. Then we have the following relation*

$$k\left.\frac{d^2 I(\theta_1\|\theta_0)}{d\theta_1^2}\right|_{\theta_1=\theta_0} = I_F[f(x;\theta_0)], \tag{3.3}$$

where

$$I_F[f(x;\theta_0)] = \int_a^b \left(\frac{\partial \ln f(x;\theta_0)}{\partial \theta_0}\right)^2 f(x;\theta_0)dx, \tag{3.4}$$

$$k = \ln 2. \tag{3.6}$$

Proof. Indeed, if we have in view the form (3.1) of the gain of information and we compute the derivative, we obtain

$$\frac{dI(\theta_1\|\theta_0)}{d\theta_1} = \frac{d}{d\theta_1}\left(\int_a^b f(x;\theta_1)\log_2\frac{f(x;\theta_1)}{f(x;\theta_0)}dx\right) =$$

$$= \int_a^b \left(\frac{df(x;\theta_1)}{d\theta_1}\log_2\frac{f(x;\theta_1)}{f(x;\theta_0)} + f(x;\theta_1)\frac{d}{d\theta_1}\log_2\frac{f(x;\theta_1)}{f(x;\theta_0)}\right)dx =$$

$$= \frac{1}{k}\int_a^b \left(\frac{d\ln f(x;\theta_1)}{d\theta_1}\ln\frac{f(x;\theta_1)}{f(x;\theta_0)} + \frac{df(x;\theta_1)}{d\theta_1}\right)dx =$$

$$= \frac{1}{k}\int_a^b \left(1+\ln\frac{f(x;\theta_1)}{f(x;\theta_0)}\right)\frac{df(x;\theta_1)}{d\theta_1}dx,$$

respectively,

$$\frac{dI(\theta_1\|\theta_0)}{d\theta_1} = \frac{1}{k}\int\limits_a^b \left(1+\ln\frac{f(x;\theta_1)}{f(x;\theta_0)}\right)\frac{df(x;\theta_1)}{d\theta_1}dx. \tag{3.7}$$

Now, if we compute the second derivative of $I(\theta_1\|\theta_0)$, we get

$$\frac{d^2I(\theta_1\|\theta_0)}{d\theta_1^2} = \frac{1}{k}\int\limits_a^b \left[\frac{1}{f(x;\theta_1)}\left(\frac{df(x;\theta_1)}{d\theta_1}\right)^2 + \right.$$
$$\left. + \left(1+\ln\frac{f(x;\theta_1)}{f(x;\theta_0)}\right)\frac{d^2f(x;\theta_1)}{d\theta_1^2}\right]dx, \tag{3.8}$$

and, hence, if we consider $\theta_1=\theta_0$,we obtain

$$\left.\frac{d^2I(\theta_1\|\theta_0)}{d\theta_1^2}\right|_{\theta_1=\theta_0} = \frac{1}{k}\int\limits_a^b \frac{1}{f(x;\theta_0)}\left(\frac{df(x;\theta_0)}{d\theta_0}\right)^2 dx + \frac{1}{k}\int\limits_a^b \frac{d^2f(x;\theta_0)}{d\theta_0^2}dx =$$
$$= \frac{1}{k}I_F[f(x;\theta_0)], \tag{3.9}$$

because from the relation

$$\int\limits_a^b f(x;\theta_0)dx = 1, \tag{3.10}$$

we obtain

$$\int\limits_a^b \frac{df(x;\theta)}{d\theta_0}dx = 0, \int\limits_a^b \frac{d^2f(x;\theta)}{d\theta_0^2}dx = 0. \tag{3.11}$$

Remark 1. From this theorem it follows that the gain of information can be considered as a generating-function of the Fisher information measure [2].

Theorem 2. *Let X be a continuous random variables and $f(x;\theta)$ its probability density function which depends on a parameter θ with values in the specified parameter space D_θ and , more $f(x;\theta)$ is absolutely continuous in θ. If θ is a local parameter for X, i.e.,*

$$f(x;\theta) = f_1(x-\theta), \theta\in D_\theta, \tag{3.12}$$

then

$$I_F[f(x;\theta)] = I_F[f_1(x-\theta)], \tag{3.13}$$

where

$$I_F[f(x;\theta)] = \int_{\mathbf{R}} \left[\frac{1}{f(x;\theta)} \frac{\partial f(x;\theta)}{\partial \theta} \right]^2 f(x;\theta)dx, \tag{3.14}$$

$$I_F[f_1(x-\theta)] = \int_{\mathbf{R}} \left[\frac{1}{f_1(x-\theta)} \frac{\partial f_1(x-\theta)}{\partial \theta} \right]^2 f_1(x-\theta)dx, \tag{3.15}$$

are Fisher's information measures.

Proof. Indeed, from (3.12) and (3.14), we obtain

$$\begin{aligned} I_F[f(x;\theta)] &= \int_{-\infty}^{\infty} \left[\frac{1}{f(x;\theta)} \frac{\partial f(x;\theta)}{\partial \theta} \right]^2 f(x;\theta)dx = \\ &= \int_{-\infty}^{\infty} \left[\frac{1}{f_1(x-\theta)} \frac{\partial f_1(x-\theta)}{\partial \theta} \right]^2 f_1(x-\theta)dx = \\ &= \int_{-\infty}^{\infty} \left[-\frac{f_1'(x-\theta)}{f_1(x-\theta)} \right]^2 f_1(x-\theta)dx = \\ &= \int_{-\infty}^{\infty} \left[\frac{f_1'(u)}{f_1(u)} \right]^2 f_1(u)du = \\ &= I_F[f_1(u)], \end{aligned}$$

if we have in view the change of variables

$$u = x - \theta. \tag{3.16}$$

Corollary 3. *If the parameter θ is a scale parameter for X with center m as follows*

$$f(x;\theta) = e^{-\theta} f_2[(x-m)e^{-\theta}], -\infty < \theta < \infty, \tag{3.17}$$

then

$$I_F[f(x;\theta)] = I_F(f_2), \tag{3.18}$$

when

$$I_F(f_2) = \int_{-\infty}^{\infty} \left[1 - x\frac{f_2'(x)}{f_2(x)} \right]^2 f_2(x)dx, \tag{3.19}$$

constantly in θ and m,-$\infty < \theta, m < \infty$.

Proof. From (3.17), we obtain

$$\frac{\partial f(x;\theta)}{\partial\theta} = \frac{\partial}{\partial\theta}\left\{e^{-\theta} f_2[(x-m)e^{-\theta}]\right\} = \\ = -f(x;\theta) - (x-m)e^{-2\theta} f_2'[(x-m)e^{-\theta}], \tag{3.20}$$

where

$$f_2'(v) = \frac{df_2(v)}{dv}, v = (x-m)e^{-\theta}. \tag{3.21}$$

Then

$$I_F[f(x;\theta)] = \int_{-\infty}^{\infty} \left[\frac{1}{f(x;\theta)} \frac{\partial f(x;\theta)}{\partial\theta}\right]^2 f(x;\theta)dx = \\ = \int_{-\infty}^{\infty} \left[\frac{-f(x;\theta) - (x-m)e^{-2\theta} f_2'[(x-m)e^{-\theta}]}{f(x;\theta)}\right]^2 f(x;\theta)dx = \\ = \int_{-\infty}^{\infty} \left\{-1 - \frac{(x-m)e^{-2\theta} f_2'[(x-m)e^{-\theta}]}{f(x;\theta)}\right\}^2 f(x;\theta)dx. \tag{3.22}$$

If we make the following change of variables

$$v = (x-m)e^{-\theta}, \tag{3.23}$$

then we obtain

$$I_F[f(x;\theta)] = \int_{-\infty}^{\infty} \left[-1 - v\frac{f_2'(v)}{f_2(v)}\right]^2 f_2(v)dv, \tag{3.24}$$

because we have

$$\left\{-1 - \frac{(x-m)e^{-2\theta} f_2'[(x-m)e^{-\theta}]}{f(x;\theta)}\right\}^2 f(x;\theta)dx = \left[-1 - v\frac{f_2'(v)}{f_2(v)}\right]^2 f_2(v)dv. \tag{3.25}$$

4. **Application**

Let X be a continuous random variable which follows a normal distribution, that is, its probability density function is defined by

$$f(x;m,\sigma^2)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right\}, x\in\mathbf{R}, \tag{4.1}$$

where $\sigma>0$ and $m\in\mathbf{R}$ are the two parameters of the distribution, namely, m is a location parameter and σ^2 is a scale parameter.

Then for the function

$$g(x;m,\sigma^2)=-\ln f(x;m,\sigma^2)= \tag{4.2}$$

$$=\ln\sqrt{2\pi}\;+\frac{1}{2}\ln\sigma^2+\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2, \tag{4.3}$$

we obtain

$$\frac{\partial g(x;m,\sigma^2)}{\partial x}=\frac{x-m}{\sigma}, \tag{4.4}$$

$$\frac{\partial^2 g(x;m,\sigma^2)}{\partial x^2}=\frac{1}{\sigma^2}>0,\forall x\in\mathbf{R}, \tag{4.5}$$

and from here it follows that the probability density (4.1) is a strongly unimodal function and more it is an absolute continuous function, if we have in view the following remark.

Remark 2. [1] Let X be a continuous random variable on the probability space (Ω,K,P) and $f(x)$, $x\in(a,b),a<b,(a,b)\subset\mathbf{R}$ its probability density function. If the function g,defined as

$$g(x)=-\ln f(x),x\in(a,b) \tag{4.6}$$

is a convex function, than f is called strongly unimodal.

Such strongly unimodal probability density function is absolutely continuous within (a,b) and more

$$g'(x)=-\frac{f'(x)}{f(x)},(f(x)\neq o,x\in(a,b)) \tag{4.7}$$

is a non-decreasing function.

Also, we say that X is absolutely continuous random variable if its probability density $f(x)$ is an absolutely continuous function.

Then, according to the relation (2.2), when $\theta = m$, we obtain

$$I_F(x;\theta) = I_F(x;m) = \frac{1}{\sigma^2}. \tag{4.8}$$

Now, we consider the relation

$$f(x;\theta) = e^{-\theta} f_2[(x-m)e^{-\theta}], -\infty < \theta < +\infty \tag{4.9}$$

and if

$$e^{\theta} = \sigma, \tag{4.10}$$

then

$$e^{-\theta} = \frac{1}{\sigma}, \theta = \ln \sigma \tag{4.11}$$

and from (4.9), we obtain

$$f(x;\theta) = e^{-\theta} f(v) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2 \right\} = f(x;m,\sigma^2), \tag{4.12}$$

where

$$v = \frac{x-m}{\sigma}. \tag{4.13}$$

Also, according to the relation (2.2), when $\theta = \sigma^2$,we obtain

$$I_F[f(x;\theta)] = I_F[f(x;\sigma^2)] = \frac{1}{2\sigma^4}. \tag{4.14}$$

References

[1] Ibragimov, J. A., *On the composition on unimodal distributions,* Teorija veroyatnostey, 1 (1956), 283-288 (In Russian).
[2] Golomb, S. W., *The information – generating function of a probability distribution*, Transactions of the Fourth Prague Conference on Information Theory, Prague, 1967,291-295.
[3] Kullback, S., *Information Theory and Statistics*,Wiley, New York, 1959.
[4] Rao, C. R., Linear Statistical Inference and Its Applications, John Wiley and Sons, Inc., New York, 1965.
[5] Rényi, A., *Some fundamental questions of information theory*, MTA III, Oszt. Közl., 10 (1960), 251-282.
[6] Rényi, A., *Probability Theory*, Akadémiai Kiado, Budapest, 1970.

Faculty of Mathematics and Computer Sciences,
Babeş-Bolyai University of Cluj-Napoca, Str.M. Kogălniceanu 1,
3400 Cluj-Napoca, Romania
E-mail address: imihoc@math.ubbcluj.ro