Information-Theoretical Considerations on Estimation Problems

SUGURU ARIMOTO

Faculty of Engineering Science, Osaka University, Toyonaka, Osaka, Japan

A new definition of generalized information measures is introduced so as to investigate the finite-parameter estimation problem. This definition yields a class of generalized entropy functions which is useful for treating the error-probability of decision and the other equivocation measures such as Shannon's logarithmic measure in the same framework and, in particular, deriving upper bounds to the error-probability. A few of inequalities between these equivocation measures are presented including an extension of Fano's inequality.

1. INTRODUCTION

Let $X = \{x_1, \ldots, x_n\}$ be a finite set of parameters and $(Y, \mathcal{Y})$ be a measurable space and let us consider a random variable $(\xi, \eta)$ defined by

$$P[\xi = x_k, \eta \in E] = p_k P_k(E),$$

where $E \in \mathcal{Y}$ and $p = (p_1, \ldots, p_n)$ or $P_k$ are probability distributions on $X$ or $\mathcal{Y}$, respectively.

The finite-parameter estimation problem consists in that a decision maker is faced with the problem of selecting only one true value of the parameter $X$ after observing a sample $\eta = y \in Y$, referring to the knowledge of probability measures $P_k, (k = 1, \ldots, n)$. Hence $p$ may be interpreted as an a priori distribution of $\eta$. In order to approach the problem, it is usual to evaluate a posteriori distribution $p^* = (p_1^*(y), \ldots, p_n^*(y))$ corresponding to the sample $\eta = y$ using Bayes' theorem and compare it with the a priori distribution $p$. From the information-theoretical point of view, this is based upon the fact that, roughly speaking, an amount of equivocation concerning the parameter decreases through observations. In fact, Shannon's measure of entropy leads to

$$H(\xi) \geq H(\xi | \eta), \quad (1.1)$$

© 1971 by Academic Press, Inc.
where

\[ H(\xi) = \sum_{k=1}^{n} - p_k \log p_k, \quad (1.2) \]

\[ H(\xi | \eta) = \mathcal{E} \sum_{k=1}^{n} - p_k^* \log p_k^*. \quad (1.3) \]

Here \( \mathcal{E} \) denotes the expectation operator with respect to the marginal distribution \( P = \sum_k p_k p_k \) of \( \eta \). Hence the decrease of uncertainty \( I[\xi, \eta] = H(\xi) - H(\xi | \eta) \) is called an average amount of information contained in the observation concerning the parameter \( X \) (Lindley [6], Rényi [9], and Vajda [10]).

Let \( f(\cdot) \) be a real-valued scalar function and let us replace \( -p_k \log p_k \) and \( -p_k^* \log p_k^* \) by \( f(p_k) \) and \( f(p_k^*) \) in (1.2) and (1.3), respectively. Then it is obvious that the inequality (1.1) is valid as long as \( f(\cdot) \) is a concave function. This fact motivated a more general approach to the definition of information. Thus a number of entropy functions have been proposed (Rényi [8], Csiszár [3], and Perez [7]), and their usefulness for finite parameter estimation problems has been developed (Perez [7], Rényi [9], and Vajda [10]).

The most general type of such entropy functions is defined with an arbitrary continuous concave function \( f(\cdot) \) as

\[ H'(\xi) = \sum_{k=1}^{n} f(p_k) \quad (1.4) \]

or with a continuous decreasing function \( I(\cdot) \) on the interval \((0,1]\) as

\[ H'(\xi) = \sum_{k=1}^{n} p_k I(p_k). \quad (1.5) \]

This paper has its starting point at a new definition of generalized information measures slightly different from (1.4) or (1.5) and shows a number of properties to which the definition leads. The introduction of this kind of information measure clarifies, as a result, interrelations among a class of information quantities and, in particular, the error-probability and the logarithmic equivocation measure can be treated in the same framework of a family of entropy functions.
2. Generalized Entropy Function and Its Properties

Let \( f(u) \) be a real-valued scalar function defined and nonnegative on \((0, 1]\) with a continuous derivative on \((0, 1]\) and \(f(1) = 0\). Differently from definition (1.4) or (1.5), let us define

\[
H_f(\xi) = F_f(p_1, \ldots, p_n) = \inf_{\tilde{p}} \sum_{k=1}^{n} p_k f(\tilde{p}_k),
\]

where the operation \(\inf\) is taken over all probability distributions such as \(\tilde{p} = (\tilde{p}_1, \tilde{p}_2, \ldots, \tilde{p}_n), \sum_k \tilde{p}_k = 1\) and \(\tilde{p}_k > 0\). This definition yields immediately the following properties:

**Theorem 1.**

1. \(F_f(p_1, \ldots, p_n)\) is continuous and symmetric with respect to its arguments \(p_1, \ldots, p_n\).
2. \(F_f(p_1, \ldots, p_n) = F_f(p_1, \ldots, p_n, 0)\).
3. \(F_f(p) \equiv F_f(p_1, \ldots, p_n)\) is a concave function with respect to \(p\).
4. \(0 \leq F_f(p_1, \ldots, p_n) \leq f(1/n)\).
5. If \(f(u)\) is convex, then \(F_f(p_1, \ldots, p_n) \leq F_f(1/n, \ldots, 1/n) = f(1/n)\).
6. In general,

\[
F_f(p_1, \ldots, p_n) \leq \sum_{k=1}^{n} p_k f(\tilde{p}_k).\]

When \(n \geq 3\), the equality sign holds for any probability distribution \(p\) if and only if the scalar function \(f\) is represented by

\[
f(u) = c \cdot \log u,
\]

where \(c\) is a nonpositive constant. When \(n = 2\), the equality holds for an infinite family of \(f\) including \(f(u) = c \cdot \log u\).

**Proof.** The proof of (1), (2) and (4) is clear. To prove (3), let \(\lambda\) be an arbitrary constant such that \(\lambda \in [0, 1]\) and \(p, q\) be arbitrary probability distributions. Then

\[
F_f(\lambda p + (1 - \lambda) q) = \inf_{\tilde{p}} \sum_{k=1}^{n} (\lambda p_k + (1 - \lambda) q_k) f(\tilde{p}_k)
\]

\[
\geq \inf_{\tilde{p}} \sum_{k=1}^{n} \lambda p_k f(\tilde{p}_k) + \inf_{\tilde{p}} \sum_{k=1}^{n} (1 - \lambda) q_k f(\tilde{p}_k)
\]

\[
= \lambda F_f(p) + (1 - \lambda) F_f(q).
\]
To prove (5), note that for an arbitrarily given $\epsilon > 0$ there exists a probability distribution $q = (q_1, \ldots, q_n)$ such that

$$F_f\left(\frac{1}{n}, \ldots, \frac{1}{n}\right) \geq -\epsilon + \frac{1}{n} f(q_k).$$

Taking into account the convexity of $f$, we have

$$F_f\left(\frac{1}{n}, \ldots, \frac{1}{n}\right) \geq -\epsilon + f\left(\frac{1}{n} \sum q_k\right) = -\epsilon + f\left(\frac{1}{n}\right).$$

Since $\epsilon > 0$ is arbitrary, this implies $F_f(1/n, \ldots, 1/n) \geq f(1/n)$ which, together with (4), completes the proof of (5). To prove (6), suppose that it holds

$$\inf_{\tilde{p}} \sum_{k=1}^{n} p_k f(\tilde{p}_k) = \sum_{k=1}^{n} p_k f(p_k) \quad (2.4)$$

for any $p$. Then according to the well-known theorem of Lagrange’s multiplier rule for a class of nonlinear programming problems, the form (2.4) means

$$p_k f'(p_k) - \lambda = 0, \quad \text{for any } p_k > 0, \quad (2.5)$$

where $\lambda$ is a constant multiplier and $f'$ is the derivative of $f$. In view of the assumption that $n > 3$, (2.5) implies

$$u \cdot f'(u) = \text{const.} \quad \text{for } 0 < u < 1. \quad (2.6)$$

By integration, this reduces to

$$f(u) = c \cdot \log u + c_1,$$

and on account of $f(0) \geq 0$ and $f(1) = 0$ the function $f$ should be of the form (2.3) with a nonpositive constant $c$. When $n = 2$, (2.5) does not generally imply (2.6) but the following equality:

$$uf'(u) = (1 - u)f'(1 - u), \quad 0 < u < 1. \quad (2.7)$$

For example, if the derivative of $f$ is represented by

$$f'(u) = g[u(1 - u)]/u, \quad \text{for } 0 < u < 1,$$

where $g[v]$ is an arbitrary continuous function defined on $0 \leq v \leq 1/4$, then $f(u)$ satisfies (2.5) or (2.7). Hence it is easy to see that there exists an infinite
family of $f$'s which satisfies (2.4) for any $p$. Indeed, a class of functions $g[v] = -v^e$ with $0 \leq e \leq 1$ generates

$$f(u) = \int_u^1 v^{e-1}(1 - v)^e \, dv$$

which clearly satisfies (2.4) as long as $n = 2$. The sufficiency part of the proof follows from the inequality (throughout this paper the natural logarithm is employed)

$$\sum_{k=1}^n c \cdot p_k \log \hat{p}_k - \sum_{k=1}^n c \cdot p_k \log p_k = \sum_{k=1}^n c \cdot p_k \log \frac{\hat{p}_k}{p_k} \geq \sum_{k=1}^n c \cdot p_k \left( \frac{\hat{p}_k}{p_k} - 1 \right) = 0.$$

**Example 1.** Consider the case $n = 2$, and let $f(u) = \sqrt{(1 - u)/u}$. Then it is easy to see that

$$F_1(p_1, p_2) = 2 \sqrt{p_1 p_2}.$$  \hspace{1cm} (2.8)

From noting the inequalities

$$1 - u \leq -\log u \leq \sqrt{(1 - u)/u} \quad \text{for} \quad 0 < u \leq 1,$$

and using the fact that $H_{f_1}(\xi) \leq H_{f_2}(\xi)$ whenever $f_1(u) \leq f_2(u)$, it follows immediately that

$$\min(p_1, p_2) \leq H(p_1, p_2) = -p_1 \log p_1 - p_2 \log p_2 \leq 2 \sqrt{p_1 p_2}. \quad (2.9)$$

These inequalities play implicitly an important role in the paper given by Rényi [9].

**Example 2.** It is convenient to introduce a class of scalar functions which are defined as

$$f^\beta(u) = \frac{1 - u^{1-\beta}}{1 - \beta}, \quad \text{for} \quad \beta \neq 1 \quad \text{and} \quad \beta \geq 0,$$

$$f^1(u) = -\log u = \lim_{\beta \to 1} f^\beta(u). \quad (2.10)$$
Then it follows from an easy calculation that

\[ F_\beta(p) = \inf_\beta \sum_{k=1}^n p_k f_\beta(p_k) \]

\[
\begin{cases} 
1 - \max_k p_k & \text{for } \beta = 0, \\
1 - \left( \frac{\sum p_k^{1/\beta}}{1 - \beta} \right) & \text{for } \beta \neq 1 \text{ and } \beta > 0, \\
\sum_k - p_k \log p_k & \text{for } \beta = 1.
\end{cases}
\]

(2.11)

The entropy \( F_\beta(p) \) is closely related with Renyi's entropy of order \( \alpha \) (see Rényi[8]) which is defined as

\[ H_\alpha(p) = \frac{1}{1 - \alpha} \log \left( \sum p_k^\alpha \right). \]

(2.12)

By letting \( \beta = \alpha^{-1} \) and noting the inequality \( \log u \leq u - 1 \), we have the following relations:

\[ F_\beta(p) \leq H_\alpha(p) \leq H(p) \quad \text{for } 0 \leq \beta = \alpha^{-1} < 1, \]

\[ F_\beta(p) = H_\alpha(p) = H(p) \quad \text{for } \beta = \alpha = 1, \]

\[ F_\beta(p) \geq H_\alpha(p) \geq H(p) \quad \text{for } \beta = \alpha^{-1} > 1. \]

(2.13)

Moreover, it is quite easy to see that

\[ f_\beta(u) \geq f_\alpha(u) \quad \text{and} \quad F_\beta(p) \geq F_\alpha(p) \quad \text{for } \beta_1 \geq \beta_2. \]

(2.14)

Another interesting special case of \( \beta \)'s is \( \beta = 2 \) in which the following equality is obtained:

\[ F_2(p) = \left( \sum p_k^{1/2} \right)^2 - 1 = \sum_{k \neq j} \sqrt{p_k p_j}. \]

(2.15)

It should be noted that the quantity (2.15), together with (2.8), is closely related to Chernoff's bound (Chernoff [1]). On the other hand, the measure \( F_\alpha(p) \) corresponds eventually to the probability of error of decision making.
Thus, in order to investigate interrelations between these quantities it is useful to introduce the following relation

\[ F_0(p) \leq F_1(p) = H(p) \leq F_2(p) \] (2.16)

which is a special case of (2.14).

It should be noted that the generalized entropy function defined in (1.4) or (1.5) can not include any type of \( F_0 \)'s except \( F_1 \).

**Example 3.** Let us consider the function

\[ f(u) = [(1 - u)/u]^\gamma, \quad 0 \leq \gamma \leq 1. \] (2.17)

From definition (2.1) and the inequality \([(1 - u)/u]^\gamma \geq 1 - u\), it is obvious that, for every probability distribution \( \hat{p} \),

\[ F_0(p) \leq F_f(p) \leq \sum_{k=1}^{n} p_k[(1 - \hat{p}_k)/\hat{p}_k]^\gamma. \] (2.18)

In particular, if we put

\[ \hat{p}_k = \frac{\hat{p}_k^{1/(1+\gamma)}}{\sum_{j=1}^{n} \hat{p}_j^{1/(1+\gamma)}} \]

and substitute this into (2.18), we obtain

\[ F_0(p) \leq F_f(p) \leq \sum_{k=1}^{n} \hat{p}_k^{1/(1+\gamma)} \left[ \sum_{j \neq k} \hat{p}_j^{1/(1+\gamma)} \right]^\gamma. \] (2.19)

This type of inequality plays a vital role in the simple derivation of Shannon’s coding theorem by Gallager [5].

### 3. Equivocation and Probability of Error

According to Theorem 1, a generalized entropy function \( F_f(p) \) associated with a scalar function \( f \) becomes, of itself, concave with respect to the probability vector \( p \). Hence, if we define a generalized equivocation quantity, similarly to (1.3), by

\[ H_f(\xi \mid \eta) = \delta F_f(\hat{p}^\pm), \] (3.1)
then it holds obviously that
\[ H_f(\xi | \eta) = \mathcal{E} F_f(p^*) \leq F_f(\mathcal{E} p^*) = F_f(p) = H_f(\xi). \] (3.2)

The generalized equivocation measure defined as (3.1) includes the probability of error as a special case. To show this, let \( f(u) = 1 - u \) and suppose that one made a decision by choosing a decision function \( \delta(y) \) from \( Y \) to \( X \), or, equivalently, let us define

**Definition 1.** A measurable \( n \)-vector-valued function

\[ \delta(y) = (\delta_1(y), \ldots, \delta_n(y)), \]

whose \( k \)-th coordinate takes 1 or 0, correspondingly to \( \delta(y) = x_k \) or \( \delta(y) \neq x_k \), is called a decision.

Then the average probability of error corresponding to the decision \( \delta \) can be defined as

\[ P_e(\delta) = \mathcal{E} \sum_{k=1}^{n} p_k^*(1 - \delta_k(y)). \] (3.3)

In particular, it is possible to choose one of the optimum decisions \( \delta^*(y) \) such that

\[ P_e(\delta^*) = \inf_{\delta} P_e(\delta) = \mathcal{E} F_{1-n}(p^*) = P_e, \] (3.4)

where the operation \( \inf \) is taken over all measurable decision functions, and \( P_e \) in this paper is called the Bayes probability of decision error. For example, let us define

\[ \delta_k^*(y) = \begin{cases} 1 & \text{if } k \text{ is the largest integer such that } \\ p_k^*(y) = \max\{p_1^*(y), \ldots, p_n^*(y)\}, \\ 0 & \text{otherwise}, \end{cases} \]

then clearly \( \delta^*(y) \) satisfies (3.4).

In view of the definition (3.1) and the relation (3.4) it is quite easy to see

**Theorem 2.** If \( f(u) \geq 1 - u \) on \( (0, 1] \), then \( H_f(\xi | \eta) \geq P_e \).
Example 4. In relation to Example 1, let us examine the case \( n = 2 \) and choose \( f(u) = \frac{1}{2} \sqrt{1 - u} \). Then it follows from Theorem 2 that

\[
H_f(\xi \mid \eta) = \mathbb{E} \sqrt{p_1^*(y) p_2^*(y)} = \sqrt{p_1 p_2} \int_y \sqrt{p_1(y) p_2(y)} \, dy \geq P_e, \tag{3.5}
\]

where \( p_i(y) \) is the \( \mathcal{Y} \)-measurable density function of the probability distribution \( p_i(E), E \in \mathcal{Y} \). A generalization of this inequality to the cases \( n \geq 3 \) will be given in the next example.

Example 5. By using the function \( f^2(u) = u^{-1} - 1 \) and corresponding generalized entropy given in (2.15) and taking into account (2.16), the inequalities

\[
P_e = H_{p^2}(\xi \mid \eta) \leq H_{p^2}(\xi \mid \eta) = H(\xi \mid \eta) \leq H_{p^2}(\xi \mid \eta) \tag{3.6}
\]

are obtained.

As usual, in many practical situations of parameter estimation problems, the random variable \( \eta \) consists of successive independent observations \( (\eta_1, \ldots, \eta_T) \). At that time the generalized entropy (2.15) is quite suitable for calculating a bound of the probability of decision error since the integral part of (3.6) can be decomposed into

\[
\int_y \sqrt{p_k(y) p_\eta(y)} \, dy = \prod_{t=1}^T \int_{Y_t} \sqrt{p_k(t) p_\eta(t)} \, dy_t,
\]

where \( Y = Y_1 \times \cdots \times Y_T \), and it is assumed that

\[
p_k(y) = \prod_{t=1}^T p_k^t(y_i).
\]
4. Bounds for Probability of Error and Equivocation

In this section we present various kinds of upper or lower bounds to the generalized entropies or equivocation measures.

**Theorem 3.** Let $H_f(\xi \mid \eta)$ be an equivocation measure with a scalar function $f$ defined by (3.1) and $\delta(y)$ be an arbitrary decision. Then the following inequality holds:

$$H_f(\xi \mid \eta) \leq [1 - P_e(\delta)] f(1 - \epsilon) + P_e(\delta) f(\epsilon/(n - 1)),$$

where $\epsilon$ is an arbitrary number such that $0 < \epsilon < 1$.

**Proof.** Let $\varphi(y) = (\varphi_1(y), \ldots, \varphi_n(y))$ be an $n$-vector-valued function such that

$$\varphi_k(y) = \begin{cases} 1 - \epsilon & \text{if } \delta_k(y) = 1, \\ \epsilon/(n - 1) & \text{if } \delta_k(y) = 0. \end{cases}$$

Then it follows immediately from (2.1) and (3.2) that

$$H_f(\xi \mid \eta) = \mathcal{E} F_f(\varphi^*) \leq \mathcal{E} \sum_{k=1}^n p_k^* [\delta_k(y) f(1 - \epsilon) + (1 - \delta_k(y)) f(\epsilon/(n - 1))]$$

$$= (1 - P_e(\delta)) f(1 - \epsilon) + P_e(\delta) f(\epsilon/(n - 1)).$$

The inequality (4.1) is an extension of Fano's inequality (Fano [4]). Indeed, letting $f(u) = -\log u$, $\epsilon = P_e$ and $\delta = \delta^*$ (which is one of optimum decisions), and substituting these into (4.1), we obtain

$$H(\xi \mid \eta) \leq H(P_e, 1 - P_e) + P_e \log(n - 1).$$

**Theorem 4.** If $f'(u) < 0$ on $[0, 1]$, then the following inequalities hold:

1. $F_f(p) \geq F_f(F_0(p), 1 - F_0(p))$,
2. $F_f(p) \geq f(\frac{1}{2}) F_0(p)$,
3. $H_f(\xi \mid \eta) \geq f(\frac{1}{2}) P_e$,
where $F_o(p)$ denotes the function

$$F_o(p) = 1 - \max_k p_k,$$

Proof. The proof of (1) follows from property (1) of Theorem 1 and the following more general inequality:

$$F_F(p_1, \ldots, p_n) = \inf_{\hat{p}} \sum_{k=1}^{n} p_k f(\hat{p}_k)$$

$$= \inf_{\hat{p}} \sum_{k=1}^{n-1} p_k \left[ \frac{p_n}{1-p_n} f(\hat{p}_n) + f(\hat{p}_k) \right]$$

$$\geq \sum_{k=1}^{n-1} \inf_{\hat{p}_n} \frac{p_n}{1-p_n} \left[ p_n f(\hat{p}_n) + (1-p_n) f(1-\hat{p}_n) \right]$$

$$= F_F(p_n, 1-p_n).$$

To prove (2), we need the following lemma:

**Lemma 1.** Let $p = (p_1, \ldots, p_n)$ be an arbitrary probability distribution such that

$$p_1 > p_2 > \cdots > p_n$$

and $q = (q_1, \ldots, q_n)$ be a corresponding probability distribution satisfying

$$\inf_{\hat{p}} \sum_{k=1}^{n} p_k f(\hat{p}_k) = \sum_{k=1}^{n} p_k f(q_k). \quad (4.2)$$

Then the distribution $q$ has the following descending order:

$$q_1 > q_2 \geq \cdots \geq q_n.$$

Proof. Without loss of generality we at first suppose that $p_1 > p_2 > \cdots > p_n$ and $q_k < q_{k+1}$. Then, by letting

$$\bar{p} = (p_1, \ldots, p_n) = (q_1, \ldots, q_{k-1}, q_k, \ldots, q_n),$$
one obtains

\[ \sum_{k=1}^{n} p_k f(q_k) - \sum_{k=1}^{n} p_k f(\bar{p}_k) = (p_k - p_{k+1})[f(q_k) - f(q_{k+1})] > 0. \]

This contradicts (4.2). Let us now suppose that \( p_1 > p_2 > \cdots > p_n \) and \( q_1 = q_2 \). Then, putting

\[ \bar{p} = (\bar{p}_1, \ldots, \bar{p}_n) = (q_1 + \epsilon, q_1 - \epsilon, q_3, \ldots, q_n), \]

where \( \epsilon \) is a positive number, we have

\[ \sum_{k=1}^{n} p_k f(q_k) - \sum_{k=1}^{n} p_k f(\bar{p}_k) = -f'(q_1)(p_1 - p_2) + o(\epsilon) > 0 \]

for a sufficiently small \( \epsilon \) (where \( o(\epsilon) \) means that \( o(\epsilon)/\epsilon \to 0 \) as \( \epsilon \to 0 \)). This contradicts (4.2).

Returning to the proof of (2), let us assume that for an arbitrarily given and fixed \( p = (p_1, \ldots, p_n) \) it holds that \( p_k = \max(p_1, \ldots, p_n) \). Then, according to Lemma 1, one can choose a probability distribution \( q = (q_1, \ldots, q_n) \) such that it satisfies simultaneously

\[ q_k = \max(q_1, \ldots, q_n), \quad (4.3) \]

\[ F_j(p) = \inf_{\bar{q}} \sum_{j=1}^{n} p_j f(\bar{q}_j) = \sum_{j=1}^{n} p_j f(q_j). \quad (4.4) \]

Since (4.3) means \( q_j \leq 1/2 \) for \( j \neq k \), it follows from (4.4) that

\[ F_j(p) = p_k f(q_k) + \sum_{j \neq k} p_j f(q_j) \geq \sum_{j \neq k} p_j f(\frac{1}{2}) = f(\frac{1}{2}) F_0(p). \]

The proof of (3) follows directly from expectation over (2).

The inequalities obtained in this theorem are closely related to those presented by Chu and Chueh [2], Rényi [9], and Vajda [10].
THEOREM 5. If \( f(u) \) is a convex function with \( f'(u) < 0 \) on \((0, 1]\), then it holds that

\[
(1) \quad F_f(p) \geq f\left(1 - \frac{F_0(p)}{2}\right)
\]

\[
(2) \quad H_f(\xi \mid \eta) \geq f\left(1 - \frac{p_{\xi}}{2}\right).
\]

Proof. Let \( p = (p_1, \ldots, p_n) \) be an arbitrarily given and fixed probability distribution and \( q = (q_1, \ldots, q_n) \) be one of probability distributions satisfying (4.3) and (4.4), simultaneously. Then, owing to the convexity of \( f \) and \( q_j \leq 1/2 \) for \( j \neq k \), it is easy to see that

\[
F_f(p) \geq f\left(p_k q_k + \sum_{j \neq k} p_j q_j\right) \geq f\left(p_k + \frac{1}{2} \sum_{j \neq k} p_j\right) = f\left(1 - \frac{1 - p_k}{2}\right) = f\left(1 - \frac{F_0(p)}{2}\right),
\]

and

\[
H_f(\xi \mid \eta) = \varepsilon F_f(p^*) \geq f\left(1 - \frac{F_0(p^*)}{2}\right) \geq f\left(1 - \frac{p_{\xi}}{2}\right).
\]

REFERENCES


