INFORMATION AND CONTROL 51, 193–208 (1981)

New Parametric Measures of Information

K. FERENTINOS AND T. PAPAIOANNOU

Department of Mathematics, University of Ioannina, Ioannina, Greece

In this paper methods are presented for obtaining parametric measures of information from the non-parametric ones and from information matrices. The properties of these measures are examined. The one-dimensional parametric measures which are derived from the non-parametric are superior to Fisher's information measure because they are free from regularity conditions. But if we impose the regularity conditions of the Fisherian theory of information, these measures become linear functions of Fisher's measure.

1. INTRODUCTION

Several functionals have been proposed in the literature as measures of information and each definition enjoys certain axiomatic and/or heuristic properties. A convenient way to differentiate among the various measures of information is to classify them in the following three categories: parametric, non-parametric and entropy-type measures of information.

Parametric measures of information measure the amount of information supplied by the data about an unknown parameter θ and are functions of θ. In this case the best known measure is Fisher's measure of information.

Non-parametric measures express the amount of information supplied by the data for discriminating in favour of a distribution f_1 against another f_2 or measure the distance or affinity between f_1 and f_2. The best known measure of this type is the Kullback–Leibler measure.

Measures of entropy give the amount of information contained in a distribution, that is, the amount of uncertainty concerning the outcome of an experiment. The classical measures of this type are Shannon's and Rényi's.

The measures of information proposed in the literature have various properties which establish their applicability in various sciences. The importance of an amount which has been proposed or can be proposed as a measure of information emanates from these properties. Many properties have appeared in the literature in various contexts. If I is a measure of information of whatsoever category, X a r.v. and $T(X)$ a measurable transformation of X (statistic), then the main properties of I are summarized as follows:

0019-9958/81/120193–16$02.00/0

(1) Non-negativity: $I_X \geqslant 0$.

(2) Additivity:

$$I_{X,Y} = I_X + I_Y, \qquad I_{X,Y} = I_X + I_{Y|X}, \qquad I_{X,Y} \leqslant I_X + I_Y.$$

(weak if X ind Y) (strong) (subadditivity)

(3) Conditional inequality: $I_Y \geqslant I_{Y|X}$.

(4) Maximal information: $I_X \geqslant I_{T(X)}$.

(5) Invariance under sufficient transformations: $I_X = I_{T(X)}$ iff $T(X)$ is a sufficient statistic.

(6) Convexity: Let $\alpha_1 + \alpha_2 = 1$, $\alpha_1, \alpha_2 \geqslant 0$ and f_1, f_2 pdf's, then

$$I_X(\alpha_1 f_1 + \alpha_2 f_2) \leqslant \alpha_1 I_X(f_1) + \alpha_2 I_X(f_2).$$

(7) Loss of information: Let G be the set of all partitions g of R^k and I_g the measure of information of the r.v. g, then $\sup_{g \in G} I_g = I_X$.

(8) If the experiment $\mathscr{E}_X$ is sufficient for the experiment $\mathscr{E}_Y$ according to Blackwell's definition with the same parameter space Θ, then $I_X \geqslant I_Y$.

The construction of measures of information is not an easy problem. Ideally one would like to find measures of information which satisfy all of the previous properties for any family of probability distributions. Such measures do not exist. So one has to be less demanding and require the fulfillment of as many properties as possible. The non-negativity, maximal information and invariance under sufficient transformations properties constitute a minimum set of properties to be required of a measure of information. Thus our target is the determination of amounts which satisfy these three properties. Then we examine the fulfillment of other properties as well.

In this paper methods are presented for obtaining parametric measures of information by two ways. First, new one-dimensional parametric measures of information are derived from general information matrices. The case of Fisher's information matrix is examined separately (Section 2). Second new parametric measures are derived from the non-parametric ones by either taking the limit of f_1 to f_2 or by using special parametric transformations (Section 3). The relationship of these new measures of information with Fisher's classical measure is examined in Section 4. The results lead us to make the conjecture that if certain regularity conditions are satisfied Fisher's measure of information is unique up to a constant. Finally, in Section 5 we discuss the behaviour of our parametric measures of information with respect to the Cramér–Rao bounds.

2. One-Dimensional Parametric Measures from Information Matrices

Parametric measures of information refer to parametric families $\mathscr{M} = \{P_\theta, \theta \in \Theta\}$ of probability measures on a measurable space $(\mathscr{X}, \mathscr{A})$ dominated by a finite or σ-finite measure μ. The parameter space Θ can either be an open subset of the real line or an open subset of the k-dimensional Euclidean space R^k. Let also $f(x, \theta) = dP_\theta/d\mu$. The pure parametric measures of information are the measures of

$$\text{Fisher (1925): } I_X^{\mathrm{F}}(\theta) = E_\theta \left[\frac{\partial}{\partial\theta} \log f(X, \theta)\right]^2, \qquad \text{if } \theta \text{ univariate,}$$

$$= \left\| E_\theta \left[\frac{\partial}{\partial\theta_i} \log f(X, \theta) \frac{\partial}{\partial\theta_j} \log f(X, \theta)\right]\right\|_{k\times k}, \qquad \text{if } \theta \text{ } k\text{-variate,}$$

where $\| \; \|_{k\times k}$ denotes a $k \times k$ matrix and, for θ univariate, the measures of

$$\text{Vajda (1973): } I_X^{\mathrm{V}}(\theta) = E_\theta \left|\frac{\partial}{\partial\theta} \log f(X, \theta)\right|^\alpha, \qquad \alpha \geqslant 1,$$

$$\text{Mathai (1967): } I_X^{\mathrm{Mat}}(\theta) = \left[E_\theta \left|\frac{\partial}{\partial\theta} \log f(X, \theta)\right|^\alpha\right]^{1/\alpha}, \qquad \alpha \geqslant 1,$$

$$\text{Boekee (1977): } I_X^{\mathrm{Bo}}(\theta) = \left[E_\theta \left|\frac{\partial}{\partial\theta} \log f(X, \theta)\right|^{\frac{s}{s-1}}\right]^{s-1}, \qquad 1 < s < \infty.$$

If θ is k-variate Fisher's information matrix is the only parametric measure of information available for this case. Its diagonal elements are Fisher's univariate measures of information about a single parameter given the values of all other nuisance parameters. There is no simple information theoretic interpretation for the off-diagonal elements of this matrix. It seems awkward to measure information about a k-variate parameter by means of a $k \times k$ matrix, some elements of which have no information connotation. Thus it is desirable to measure information by a single quantity.

For the remaining part of this section θ will be k-variate. In general, by the term information matrix we shall mean a parametric measure of information, which for every $\theta \in \Theta$ is a $k \times k$ positive semidefinite real symmetric matrix which satisfies the property $I_X(\theta) \geqslant I_T(\theta)$ for every statistic $T(X)$ and every $\theta \in \Theta$ with equality iff $T(X)$ is sufficient for θ. Throughout this paper the matrix inequality $A \geqslant B$ will mean that the matrix $A-B$ is positive semidefinite. Information matrices provide a way to construct new one-dimensional measures of information by considering special functions of their eigenvalues. The methodology is based on the following theorems.

THEOREM 2.1. *Let $\mathscr{I} = \{I_X(\theta), \theta \in \Theta\}$ be a set of information matrices about a k-variate parameter θ. Then a real-valued function $f[I_X(\theta)]$ defined on $\mathscr{I}$ is non-negative $\{f[I_X(\theta)] \geqslant 0$ for every $\theta \in \Theta\}$ and satisfies the maximal property $\{f[I_X(\theta)] \geqslant f[I_{T(X)}(\theta)]$ for every statistic T and for every $\theta \in \Theta\}$ if the function $f[I_X(\theta)]$ is a function of the eigenvalues of $I_X(\theta)$ which is strictly increasing in each argument, with $f(0) = 0$, where 0 is the null matrix.*

A converse of the above theorem is the following.

THEOREM 2.2. *Let $\mathscr{I} = \{I_X(\theta), \theta \in \Theta\}$ be a set of information matrices about a k-variate parameter θ. If a real-valued function $f[I_X(\theta)]$ defined on $\mathscr{I}$ is non-negative $\{f[I_X(\theta)] \geqslant 0$ for every $\theta \in \Theta\}$, satisfies the maximal property $\{f[I_X(\theta)] \geqslant f[I_T(\theta)]$ for every statistic T$\}$ and is invariant under orthogonal transformations $\{f[P^{\mathscr{T}} I_X(\theta) P] = f[I_X(\theta)]$ for any orthogonal matrix P$\}$, then $f[I_X(\theta)]$ is identical to some function of the eigenvalues of $I_X(\theta)$, which is strictly increasing in each argument.*

Proof of Theorems 2.1 *and* 2.2. The proof of Theorems 2.1 and 2.2 is based on the following lemma of Okamoto and Kanazawa (1968).

LEMMA. *Let $\mathscr{C}$ be the set of all positive semidefinite real symmmetric matrices and $\lambda_1(A) \geqslant \lambda_2(A) \geqslant \cdots \geqslant \lambda_k(A)$ be the eigenvalues of $A \in \mathscr{C}$. Then a real-valued function $f(A)$ defined on $\mathscr{C}$ is*

(i) *strictly increasing $[f(A) > f(B)$, if $A > B]$ and*

(ii) *invariant under orthogonal transformations $[f(P^{\mathscr{T}} AP) = f(A)$ for any orthogonal matrix $P]$*

iff $f(A)$ is identical to some function $g(\lambda_1(A),..., \lambda_k(A))$ of the eigenvalues of A, which is strictly increasing in each argument.

For Theorem 2.1 we have that $f(I_X(\theta))$ satisfies the assumptions of the inverse part of the above lemma. Therefore $f[I_X(\theta)] > f[I_T(\theta)]$ if $I_X(\theta) > I_T(\theta)$. This, however, is true because $I_X(\theta)$ is an information matrix. If T is sufficient for θ, $I_X = I_T$ and hence $f[I_X(\theta)] = f[I_T(\theta)]$. This proves the maximal property. The non-negativity property is valid, since $f[I_X(\theta)]$ is strictly increasing, $I_X(\theta) \geqslant 0$ and $f(0) = 0$.

The proof of Theorem 2.2 follows easily from the above lemma.

2.1. *Applications*

Let $I_X(\theta) = I_X^F(\theta)$ be Fisher's information matrix and $\theta = (\theta_1, ..., \theta_k)$. Let the following regularity conditions be satisfied:

A. $f(x, \theta) > 0$ for all $x \in \mathscr{X}$, $\theta \in \Theta$;

B. $\dfrac{\partial}{\partial \theta_i} f(x, \theta)$ exists for all $x \in \mathscr{X}$, all $\theta \in \Theta$ and all i $(i = 1, ..., k)$;

C. for any $A \in \mathscr{A}$, $\dfrac{d}{d\theta_i} \int_A f(x, \theta)\, d\mu = \int_A \dfrac{\partial f(x, \theta)}{\partial \theta_i}\, d\mu$ for all $\theta \in \Theta$ and all i.

Let also similar regularity conditions be satisfied for the density of $T(X)$. These conditions will be referred to below as regularity conditions of Fisher's information measure. The following theorem supplements Theorem 2.2 for Fisher's information matrix.

THEOREM 2.3. *Let $f[I_X^F(\theta)]$ be a real-valued function of Fisher's information matrix. If $f[I_X^F(\theta)]$ is non-negative $\{f[I_X^F(\theta)] \geqslant 0\}$, satisfies the maximal property $\{f[I_X^F(\theta)] \geqslant f[I_T^F(\theta)]\}$ and is invariant under orthogonal transformations of θ $\{f[I_X^F(\phi)] = f[I_X^F(\theta)]$, where $\phi = P\theta$, P any orthogonal matrix$\}$, then the $f[I_X^F(\theta)]$ is identical to some function of the eigenvalues of $I_X^F(\theta)$, which is strictly increasing in each argument.*

Proof. Let $\phi = P\theta$ be a linear orthogonal transformation of θ, where $\theta = (\theta_1, \theta_2, ..., \theta_k)$, $\phi = (\phi_1, \phi_2, ..., \phi_k)$ and P an orthogonal matrix of order k. It is known that

$$I_X^F(\phi) = A I_X^F(\theta) A^{\mathscr{T}},$$

where $\theta = g(\phi)$ is a parametric function of θ and $A = \|\partial g/\partial \phi\|_{k \times k}$ (cf. Fisher, 1956). Here $A = P^{\mathscr{T}}$. Therefore

$$f[I_X^F(\phi)] = f[P^{\mathscr{T}} I_X^F(\theta) P] = f[I_X^F(\theta)].$$

The proof now follows on account of Theorem 2.2.

Now we apply the previous theory and obtain new one-dimensional parametric measures of information as byproducts of Fisher's information matrix.

a. *The measure* $I_X^*(\theta) = \operatorname{tr}[I_X^F(\theta)]$. Let $\lambda_i(\theta)$, $i = 1, 2, ..., k$, be the eigenvalues of Fisher's information matrix $I_X^F(\theta)$. Then

$$I_X^*(\theta) = \operatorname{tr}[I_X^F(\theta)] = \sum_{i=1}^{k} \lambda_i(\theta)$$

is a strictly increasing function for every eigenvalue and $I_X^{\mathrm{F}}(\theta)$, $I_T^{\mathrm{F}}(\theta)$, $I_X^{\mathrm{F}}(\theta) - I_T^{\mathrm{F}}(\theta) \in \mathscr{T}$. Thus, by Theorem 2.1, we have

$$I_X^*(\theta) \geqslant 0, \qquad I_X^*(\theta) \geqslant I_T^*(\theta) \qquad \text{for every } \theta \in \Theta.$$

Also $I_X^*(\theta) = I_T^*(\theta)$ for every $\theta \in \Theta$ iff T sufficient for θ. At this point we remark that $\mathrm{tr}[I_X^{\mathrm{F}}(\theta)] = \mathrm{tr}[I_T^{\mathrm{F}}(\theta)]$ for every $\theta \in \Theta$ implies $I_X^{\mathrm{F}}(\theta) = I_T^{\mathrm{F}}(\theta)$ for every $\theta \in \Theta$, which is not true for every function of matrices. Also $I_X^*(\theta) = 0$ for every $\theta \in \Theta$ iff $f(x, \theta)$ does not depend on θ. Again this is not true for every function of $I_X^{\mathrm{F}}(\theta)$.

b. *The measure* $D_X(\theta) = \det[I_X^{\mathrm{F}}(\theta)]$. It is known that

$$D_X(\theta) = \det[I_X^{\mathrm{F}}(\theta)] = \prod_{i=1}^{k} \lambda_i(\theta),$$

which is a strictly increasing function for every eigenvalue. Then on account of Theorem 2.1 we have $D_X(\theta) \geqslant 0$, $D_X(\theta) \geqslant D_T(\theta)$ for every $\theta \in \Theta$. It is clear that $D_X(\theta) = D_T(\theta)$ if T sufficient for θ. In general the converse is not true. Also $D_X(\theta) = 0$ if $f(x, \theta)$ does not depend on θ. The converse again is not always true.

c. *The measure* $\lambda_X^i(\theta) = \lambda_i[I_X^{\mathrm{F}}(\theta)]$. Every eigenvalue of Fisher's information matrix $I_X^{\mathrm{F}}(\theta)$ is a satisfactory measure of information. Let $\lambda_1[I_X^{\mathrm{F}}(\theta)] \geqslant \lambda_2[I_X^{\mathrm{F}}(\theta)] \geqslant \cdots \geqslant \lambda_k[I_X^{\mathrm{F}}(\theta)]$ and $\lambda_1[I_T^{\mathrm{F}}(\theta)] \geqslant \lambda_2[I_T^{\mathrm{F}}(\theta)] \geqslant \cdots \geqslant \lambda_k[I_T^{\mathrm{F}}(\theta)]$ be the eigenvalues of the matrices $I_X^{\mathrm{F}}(\theta)$ and $I_T^{\mathrm{F}}(\theta)$, respectively. In this case $f[I_X^{\mathrm{F}}(\theta)] = \lambda_X^i(\theta)$ is a strictly increasing function of $\lambda_X^i(\theta)$ and thus by Theorem 2.1 we have $\lambda_X^i(\theta) \geqslant 0$ and $\lambda_X^i(\theta) \geqslant \lambda_T^i(\theta)$ for every $\theta \in \Theta$. It is clear that $\lambda_X^i(\theta) = \lambda_T^i(\theta)$ for every $\theta \in \Theta$ if T sufficient for θ. The converse is not in general true because the eigenvalues of a matrix do not uniquely determine the matrix. Also if $f(x, \theta)$ does not depend on θ, then $\lambda_X^i(\theta) = 0$. Evidently the converse is not true. For $i = 1$ we have the relation

$$\lambda_{X,Y}^1(\theta) \leqslant \lambda_X^1(\theta) + \lambda_Y^1(\theta),$$

where X, Y are independent. For $i \neq 1$ the above relation does not hold. A counter example is given by the bivariate normal distribution with only two unknown parameters: μ_X and μ_Y.

d. *The measure* $\Lambda_X(\theta) \equiv \lambda_w[I_X^{\mathrm{F}}(\theta)] = \sum_i^k w_i \lambda_i[I_X^{\mathrm{F}}(\theta)]$, $w_i > 0$. Here, as above, on account of Theorem 2.1, we have $\Lambda_X(\theta) \geqslant 0$ and $\Lambda_X(\theta) \geqslant \Lambda_T(\theta)$ for every $\theta \in \Theta$. Evidently $\Lambda_X(\theta) = \Lambda_T(\theta)$ for every $\theta \in \Theta$ if T sufficient for θ. The inverse is not true as in case c above. Similarly for the relation $\Lambda_X(\theta) = 0$.

The information measures given in a, b, c, d above satisfy additional information theoretic properties as follows: additivity (weak and strong) for I_X^*, loss of information due to groupings for I_X^*, D_X, λ_X^i, Λ_X, convexity for I_X^*, λ_X^i (if $i=1$) and information inequality for sufficient experiments according to Blackwell's definition (property 8 of Section 1) for I_X^*, $D_X(\theta)$, λ_X^i and Λ_X. Proofs and details can be found in Ferentinos and Papaioannou (1979, 1982), Goel and DeGroot (1979) and Papaioannou and Kempthorne (1971).

3. Parametric Measures from Non-Parametric

The parametric measures of information of Section 2 (Fisher, Mathai, Vajda, Boekee and trace, eigenvalues, determinant of Fisher's information matrix, etc.) are applicable to regular families of probability distributions, i.e., to families for which conditions A, B and C of Section 2.1 hold. In this section our target is to construct parametric measures of information from the non-parametric which are applicable to broader families of probability measures including the non-regular ones. The parameter θ is assumed to be univariate.

Non-parametric measures of information refer to any family of probability measures $P_i \ll \mu$ with $f_i = dP_i/d\mu$. For the remaining part of the paper any integration with respect to μ will always be understood to be over the whole space $\mathscr{X}$. The main non-parametric measures of information are the measures of

Kullback–Leibler (1951): $I_X^{\mathrm{KL}}(f_1, f_2) = \int f_1 \log(f_1/f_2)\, d\mu$,

$$\text{Rényi (1961): } I_X^{\mathrm{R}}(f_1, f_2) = \frac{1}{\alpha-1} \log \int f_1^{\alpha} f_2^{1-\alpha}\, d\mu, \quad \alpha > 0,\ \alpha \neq 1,$$

$$\text{Csiszár (1963): } I_X^{\mathrm{C}}(f_1, f_2) = \int f_2 \phi(f_1/f_2)\, d\mu, \qquad \phi \text{ convex function},$$

$$\text{Kagan (1963): } I_X^{\mathrm{Ka}}(f_1, f_2) = \int \left(1 - \frac{f_2}{f_1}\right)^2 f_1\, d\mu,$$

$$\text{Matusita (1967): } I_X^{\mathrm{M}}(f_1, f_2) = \left[\int (f_1^{1/2} - f_2^{1/2})^2\, d\mu\right]^{1/2},$$

$$\text{Vajda (1973): } I_X^{\mathrm{V}}(f_1, f_2) = \int \left|1 - \frac{f_2}{f_1}\right|^{\alpha} f_1\, d\mu, \qquad \alpha \geqslant 1.$$

Let $\mathscr{M}$ be a parametric family of probability measures and Θ be an open subset of the real line. Let also $I_X(f_\theta, f_{\theta+\Delta\theta})$ be a non-parametric measure of information for the densities $f_\theta, f_{\theta+\Delta\theta}$. There are two methods of constructing parametric measures of information from the non-parametric:

Method A.

$$I_X(\theta) = \liminf_{\Delta\theta \to 0} \frac{1}{(\Delta\theta)^2} I_X[f(x, \theta), f(x, \theta + \Delta\theta)].$$

This method has been successfully employed with some non-parametric measures of information, such as Kagan's $I_X^{\mathrm{Ka}}(f_1, f_2)$ and Csiszár's $I_X^{\mathrm{C}}(f_1, f_2)$ (cf. Kagan, 1963; Vajda, 1973; Aggarwál, 1974; and Boekee, 1979).

These measures yield parametric measures of information which enjoy several of the properties of Section 1. In general they are applicable to larger families of probability distributions than Fisher's measure.

Similar success is attained using Rényi's $I_X^{\mathrm{R}}(f_1, f_2)$ non-parametric measure of information.

DEFINITION 3.1. Rényi's parametric measure of information about θ contained in X is defined as

$$I_X^{\mathrm{R}}(\theta) = \liminf_{\Delta\theta \to 0} \frac{1}{(\Delta\theta)^2} \frac{1}{\alpha - 1} \log \int [f(x, \theta)]^\alpha [f(x, \theta + \Delta\theta)]^{1-\alpha} \, d\mu,$$
$$\alpha > 0, \ \alpha \neq 1.$$

This measure has the following properties:

$$\text{(i) if} \quad \alpha = \frac{1}{p}, \quad \frac{1}{p} + \frac{1}{q} = 1, \qquad \text{then}$$
$$I_X^R(\theta) = q I_X^B(\theta),$$

where $I_X^B(\theta)$ is the parametric measure of information obtained by Method A from the generalized Bhattacharyya measure

$$I_X^{\mathrm{B}}(f_1, f_2) = -\log \int f_1^{1/p} f_2^{1/q} \, d\mu$$

(cf. Papaioannou and Kempthorne, 1971).

$$\text{(ii)} \quad \lim_{\alpha \to 1} I_X^{\mathrm{R}}(\theta) = I_X^{\mathrm{KL}}(\theta),$$

where $I_X^{\mathrm{KL}}(\theta)$ is the parametric measure of information obtained by Method A from the Kullback–Leibler measure $I_X^{\mathrm{KL}}(f_1, f_2)$. $I_X^{\mathrm{KL}}(\theta)$ is applicable to families of probability distributions for which $P_\theta \ll P_{\theta+\Delta\theta}$ for every $\theta \in \Theta$.

$I_X^R(\theta)$ satisfies many of the properties of Section 1, such as non-negativity, weak additivity, maximal information, invariance, loss of information and information inequality in sufficient experiments according to Blackwell's definition. These properties follow from the corresponding properties of $I_X^R(f_1, f_2)$ (cf. Rényi, 1961, 1970; and Ferentinos and Papaioannou, 1982). $I_X^R(\theta)$ is applicable to broad families of probability measures including the non-regular ones, since for $0 < \alpha < 1$ all of its properties require no regularity conditions. If $\alpha > 1$ we must have $P_\theta \ll P_{\theta + \Delta\theta}$ for every $\theta \in \Theta$. In comparison with $I_X^{KL}(\theta)$, $I_X^R(\theta)$ is applicable to larger families of probability distributions for $0 < \alpha < 1$. For example, the latter is applicable to the uniform family of distributions over $(\theta - 1, \theta + 1)$, $\theta > 0$ while the former fails.

Method B. Let $k(\theta)$ be a one-to-one transformation of the parameter space Θ onto itself with $k(\theta) \neq \theta$. The quantity

$$I_X[\theta, k(\theta)] = I_X[f(x, \theta), f(x, k(\theta))]$$

can be considered as a parametric measure of information based on $k(\theta)$. This method has been successfully employed with the Kullback–Leibler and the generalized Bhattacharya non-parametric measures of information (cf. Papaioannou and Kempthorne, 1971). For the former measure the mild condition $P_\theta \ll P_{k(\theta)}$ for every $\theta \in \Theta$ is also required. The latter measure is free of regularity conditions. These measures yield parametric measures of information which enjoy the following properties: non-negativity, weak and strong additivity for the Kullback–Leibler measure and weak additivity for the Bhattacharya measure, maximal information, invariance, loss of information and invariance under parametric transformations (cf. Papaioannou and Kempthorne, 1971; Kale, 1964; Ferentinos and Papaioannou, 1979).

In a similar manner we have the following definition.

DEFINITION 3.2. The modified Rényi and Csiszár measures of information about θ contained in X and based on $k(\theta)$ are, respectively,

$$I_X^R[\theta, k(\theta)] = \frac{1}{\alpha - 1} \log \int [f(x, \theta)]^\alpha [f(x, k(\theta))]^{1-\alpha} \, d\mu, \qquad \alpha > 0, \alpha \neq 1,$$

$$I_X^C[\theta, k(\theta)] = \int f(x, \theta) \varphi \left(\frac{f(x, k(\theta))}{f(x, \theta)} \right) d\mu.$$

In the following $I_X^{KL}[\theta, k(\theta)]$, $I_X^{Ka}[\theta, k(\theta)]$, $I_X^M[\theta, k(\theta)]$, $I_X^V[\theta, k(\theta)]$, $I_X^B[\theta, k(\theta)]$ will refer to the parametric measures of information obtained by Method B from the non-parametric measures of Kullback–Leibler, Kagan, Matusita, Vajda and Bhattacharya, respectively.

The measures $I_X^{\mathrm{C}}[\theta, k(\theta)]$ and $I_X^{\mathrm{R}}[\theta, k(\theta)]$ have the following properties:

(i) If $\varphi(u) = u \log u$ then

$$I_X^{\mathrm{C}}[\theta, k(\theta)] = I_X^{\mathrm{KL}}[\theta, k(\theta)];$$

(ii) if $\varphi(u) = (1-u)^2$ then

$$I_X^{\mathrm{C}}[\theta, k(\theta)] = I_X^{\mathrm{Ka}}[\theta, k(\theta)];$$

(iii) if $\varphi(u) = (1-\sqrt{u})^2$ then

$$I_X^{\mathrm{C}}[\theta, k(\theta)] = [I_X^{\mathrm{M}}[\theta, k(\theta)]]^2;$$

(iv) if $\varphi(u) = |1-u|^\alpha$, $\alpha \geqslant 1$, then

$$I_X^{\mathrm{C}}[\theta, k(\theta)] = I_X^{\mathrm{V}}[\theta, k(\theta)];$$

(v)
$$\lim_{\alpha \to 1} I_X^{\mathrm{R}}[\theta, k(\theta)] = I_X^{\mathrm{KL}}[\theta, k(\theta)]$$

and substituting $\alpha = 1/p$ in Rényi's measure of information we obtain

$$I_X^{\mathrm{R}}[\theta, k(\theta)] = q I_X^{\mathrm{B}}[\theta, k(\theta)].$$

On account of the properties of the non-parametric measures of Csiszár and Rényi, $I_X^{\mathrm{C}}[\theta, k(\theta)]$ and $I_X^{\mathrm{R}}[\theta, k(\theta)]$ satisfy many of the properties of Section 1, such as maximal information, loss of information due to groupings, invariance and information inequality for sufficient experiments according to Blackwell's definition. In addition, $I_X^{\mathrm{R}}[\theta, k(\theta)]$ satisfies the non-negativity and weak additivity properties (cf. Csiszár, 1963, 1967; Rényi, 1961, 1970; and Ferentinos and Papaioannou, 1982). The measures $I_X^{\mathrm{C}}[\theta, k(\theta)]$ and $I_X^{\mathrm{R}}[\theta, k(\theta)]$ are applicable to broad families of probability distributions including the non-regular ones, since their properties require no regularity conditions.

4. Relationship with Fisher's Measure of Information

In this section we relate Rényi's parametric measure $I_X^{\mathrm{R}}(\theta)$ to Fisher's information measure $I_X^{\mathrm{F}}(\theta)$. Similar results can be derived from the work of Aggarwál (1974) for the Kullback–Leibler, Kagan and Matusita parametric measures of information. In particular, if certain regularity conditions are satisfied and θ is univariate, we have

$$I_X^{\mathrm{KL}}(\theta) = \tfrac{1}{2} I_X^{\mathrm{F}}(\theta), \qquad I_X^{\mathrm{Ka}}(\theta) = I_X^{\mathrm{F}}(\theta), \qquad I_X^{\mathrm{M}}(\theta) = \tfrac{1}{2}\sqrt{I_X^{\mathrm{F}}(\theta)}$$

(cf. Aggarwál, 1974; Boekee, 1979). We consider first the case of θ being univariate.

THEOREM 4.1. *Let θ be univariate and the regularity conditions of Fisher's information measure be satisfied. Also suppose that $\int |(\partial^2/\partial\theta^2) f(x, \theta)| \, d\mu < \infty$ for all $\theta \in \Theta$ and that the third partial derivative of $f(x, \theta)$ with respect to θ exists for all $\theta \in \Theta$ and for all $x \in \mathscr{X}$. Then*

$$I_X^{\mathrm{R}}(\theta) = \frac{\alpha}{2} I_X^{\mathrm{F}}(\theta).$$

Proof. A Taylor series expansion of $[f(x, \theta + \Delta\theta)]^{1-\alpha}$ with respect to the argument $\theta + \Delta\theta$ and around the point θ yields

$$\begin{aligned}
[f(x, \theta + \Delta\theta)]^{1-\alpha} &= [f(x, \theta)]^{1-\alpha} + \Delta\theta(1-\alpha)[f(x, \theta)]^{-\alpha} \frac{\partial}{\partial\theta} f(x, \theta) \\
&\quad + \frac{1}{2} (\Delta\theta)^2 (1-\alpha)[f(x, \theta)]^{-\alpha} \frac{\partial^2}{\partial\theta^2} f(x, \theta) \\
&\quad + \frac{1}{2} (\Delta\theta)^2 \, \alpha(\alpha - 1)[f(x, \theta)]^{-1-\alpha} \\
&\quad \times \left[\frac{\partial}{\partial\theta} f(x, \theta)\right]^2 + 0(\Delta\theta)^3.
\end{aligned}$$

Multiplying both sides by $[f(x, \theta)]^{\alpha}$, integrating over $\mathscr{X}$ and interchanging the integration and the partial differentiation signs, we have

$$\int [f(x, \theta)]^{\alpha} [f(x, \theta + \Delta\theta)]^{1-\alpha} \, d\mu = 1 + \frac{\alpha(\alpha - 1)}{2} (\Delta\theta)^2 \, I_X^{\mathrm{F}}(\theta) + 0(\Delta\theta)^3.$$

By the definition of $I_X^{\mathrm{R}}(\theta)$ we have

$$\begin{aligned}
I_X^{\mathrm{R}}(\theta) &= \lim_{\Delta\theta \to 0} \frac{1}{(\Delta\theta)^2} \frac{1}{\alpha - 1} \log \left[1 + \frac{\alpha(\alpha - 1)}{2} (\Delta\theta)^2 \, I_X^{\mathrm{F}}(\theta) + 0(\Delta\theta)^3\right] \\
&= \frac{\alpha(\alpha - 1)}{2(\alpha - 1)} I_X^{\mathrm{F}}(\theta) \\
&= \frac{\alpha}{2} I_X^{\mathrm{F}}(\theta).
\end{aligned}$$

COROLLARY 4.1.

$$I_X^{\mathrm{B}}(\theta)=\frac{1}{2pq}I_X^{\mathrm{F}}(\theta),$$

$$I_X^{\mathrm{KL}}(\theta)=\frac{1}{2}I_X^{\mathrm{F}}(\theta).$$

Now let θ be k-variate. In this case the non-parametric measures of information produce interesting parametric matrices by means of Method A. These matrices are not in general information matrices except in the case of Kagan's measure of information and for regular families of probability distributions. There are easily constructed counterexamples which prove the previous assertion. Kagan's case is covered in Theorem 4.2 below. Now we give the following definition.

DEFINITION 4.1. Let $\Theta\in R^k$ and $\theta=(\theta_1,..., \theta_k)$, $\theta+t_1e_1,..., \theta+t_ke_k$ be points in Θ, where $t_1, t_2,..., t_k$ are k scalars and $e_1,..., e_k$ are the unit vectors $(1, 0,..., 0),..., (0,..., 0, 1)$, respectively. Then Rényi's parametric matrix is defined by the quantity

$$R_X(\theta)=\|R_{ij}\|_{k\times k},$$

where

$$R_{ij}=\liminf_{\substack{|t|\to 0\\ t_i=t_j}}\frac{1}{t_it_j}\frac{1}{\alpha-1}\log\int[f(x,\theta)]^{\alpha}$$
$$\times[f(x,\theta+t_ie_i)f(x,\theta+t_je_j)]^{(1-\alpha)/2}\,d\mu,\qquad \alpha>0, \alpha\neq 1,$$

and $t=(t_1,..., t_k)$. Kagan's parametric matrix is defined by

$$K_X(\theta)=\|K_{ij}\|_{k\times k},$$

where

$$K_{ij}=\liminf_{|t|\to 0}\frac{1}{t_it_j}\int\left[1-\frac{f(x,\theta+t_ie_i)}{f(x,\theta)}\right]$$
$$\times\left[1-\frac{f(x,\theta+t_je_j)}{f(x,\theta)}\right]f(x,\theta)\,d\mu.$$

THEOREM 4.2. *Let the regularity conditions of Fisher's information measure be satisfied for* θ *k-variate and, in addition, suppose that* $\int|(\partial^2/\partial\theta_i\,\partial\theta_j)f(x,\theta)|\,d\mu<\infty$ *for all* $\theta\in\Theta$ *and* $i,j=1,2,...,k$ *and the third* *partial derivative of $f(x, \theta)$ with respect to θ exists for all $\theta \in \Theta$ and for all $x \in \mathscr{X}$. Then*

$$\text{(i)} \quad R_X(\theta) = \frac{(\alpha - 1)}{4} I_X^F(\theta) + \frac{(\alpha + 1)}{8} [J_X(\theta) + J_X^{\mathscr{T}}(\theta)],$$

where

$$J_X(\theta) = \begin{Vmatrix} I_{11} & I_{11} & & I_{11} \\ I_{22} & I_{22} & & I_{22} \\ \vdots & \vdots & & \vdots \\ I_{kk} & I_{kk} & \cdots & I_{kk} \end{Vmatrix}$$

with $I_{ii} = I_{ii}^F(\theta) = E[\partial/\partial\theta_i \log f(x, \theta)]^2$,

$$\text{(ii)} \quad K_X(\theta) = I_X^F(\theta).$$

Proof. A Taylor series expansion of $[f(x, \theta + t_\lambda e_\lambda)]^{(1-\alpha)/2}$ with respect to the argument $\theta_\lambda + t_\lambda$ and around the point θ_λ yields

$$\begin{aligned} &[f(x, \theta + t_\lambda e_\lambda)]^{(1-\alpha)/2} \\ &= [f(x, \theta)]^{(1-\alpha)/2} + t_\lambda \frac{(1-\alpha)}{2} [f(x, \theta)]^{-(1+\alpha)/2} \frac{\partial}{\partial\theta_\lambda} f(x, \theta) \\ &- \frac{1}{2} t_\lambda^2 \frac{(1-\alpha^2)}{4} [f(x, \theta)]^{(-(1+\alpha)/2)-1} \left[\frac{\partial}{\partial\theta_\lambda} f(x, \theta)\right]^2 \\ &+ \frac{1}{2} t_\lambda^2 \frac{(1-\alpha)}{2} [f(x, \theta)]^{-(1+\alpha)/2} \frac{\partial^2}{\partial\theta_\lambda^2} f(x, \theta) + 0(t_\lambda)^3. \end{aligned}$$

Multiplying both sides of the above relationship for $\lambda = i$ and $\lambda = j$ by $[f(x, \theta)]^\alpha$, integrating over $\mathscr{X}$ and interchanging the integration and the partial differentiation signs, after some algebra we obtain

$$R_{ij} = \frac{(\alpha - 1)}{4} I_{ij}^F(\theta) + \frac{(\alpha + 1)}{8} [I_{ii}^F(\theta) + I_{jj}^F(\theta)].$$

Thus

$$R_X(\theta) = \frac{(\alpha - 1)}{4} I_X^F(\theta) + \frac{(\alpha + 1)}{8} [J_X(\theta) + J_X^{\mathscr{T}}(\theta)].$$

The proof of (ii) follows by a similar argument.

COROLLARY 4.2.

$$\operatorname{tr}[R_X(\theta)] = \frac{\alpha}{2}\operatorname{tr}[I_X^{\mathrm{F}}(\theta)] = \frac{\alpha}{2} I_X^{*}(\theta).$$

Similar relationships can be derived for the parametric matrices generated by the other non-parametric measures of information.

The previous results lead us to the following conjecture. The one-dimensional parametric measures of information which are derived from the non-parametric are defined without any regularity conditions. If we impose the regularity conditions of Fisher's measure of information these measures become linear functions of $I_X^{\mathrm{F}}(\theta)$. This provides evidence to conjecture that, if we require the information about θ contained in the data from regular families of distributions to enjoy certain properties and axioms (non-negativity, maximal information and invariance), the quantity $I_X^{\mathrm{F}}(\theta)$ is unique up to a positive constant. This is analogous to Shannon's entropy.

5. RELATIONSHIP WITH CRAMÉR–RAO BOUNDS

Parametric measures of information play an important role in Cramér–Rao inequalities. The classical example is given by Fisher's measure of information, which appears in the lower bound on the variance of unbiased estimators. In this section we discuss the behaviour of our parametric measures of information with respect to the Cramér–Rao bounds.

Let Θ be univariate and $\{f(x, \theta), \theta \in \Theta\}$ be a regular family of probability distributions, i.e., a family for which conditions A, B and C of Section 2.1 are satisfied. The parametric measures of information obtained from the non-parametric by Method A of Section 3 are related with Fisher's measure of information as it was established in Section 4 of this paper. Thus they appear in Cramér–Rao type inequalities as, for instance,

$$E[T(X) - g(\theta)]^2 \geqslant \frac{\alpha}{2}\frac{[g'(\theta)]^2}{I_X^{\mathrm{R}}(\theta)},$$

provided that the following additional regularity conditions are satisfied:

D. $0 < I_X^{\mathrm{F}}(\theta) < \infty$ for all $\theta \in \Theta$;

E. $\int T(x) f(x, \theta)\, d\mu$ is finite and can be differentiated under the integral sign with respect to θ, where $T(X)$ is an unbiased estimator of the parametric function $g(\theta)$ (cf. Wijsman, 1973). If regularity conditions A, B, C, D and E are not satisfied the behaviour of our parametric measures of information with respect to lower bounds for the variance of unbiased estimators is an open problem except for Kagan's measure $I_X^{\mathrm{Ka}}(\theta)$. For this measure the following analog of Cramér–Rao inequality holds:

$$E[T(X)-g(\theta)]^2 \geqslant \frac{[g'(\theta)]^2}{I_X^{\mathrm{Ka}}(\theta)}$$

(cf. Kagan, 1963).

Now let θ be a k-variate parameter and $G(\theta)=(g_1(\theta),..., g_r(\theta))$ be a vector valued function on Θ such that the matrix of partial derivatives $\Delta(\theta)=\|\partial g_i(\theta)/\partial\partial_j\|$ exists for all $\theta\in\Theta$. If $T(X)$ is an unbiased estimator of $G(\theta)$ and Σ_T is its variance–covariance matrix, the generalized Cramér–Rao inequality is

$$\Sigma_T \geqslant \Delta(\theta)[I_X^{\mathrm{F}}(\theta)]^{-1}\,\Delta(\theta)^{\mathscr{T}},$$

provided that conditions analogous to those for the univariate case are satisfied (cf. Zacks, 1971). This inequality yields the following inequalities for the trace, the determinant (generalized variance) and the largest eigenvalue of Σ_T:

$$\operatorname{tr}(\Sigma_T) \geqslant \operatorname{tr}\{\Delta(\theta)[I_X^{\mathrm{F}}(\theta)]^{-1}\,\Delta(\theta)^{\mathscr{T}}\},$$

$$\det(\Sigma_T) \geqslant \frac{\{\det[\Delta(\theta)]\}^2}{\det I_X^{\mathrm{F}}(\theta)} \quad \text{if} \quad \det[\Delta(\theta)] \neq 0,$$

$$\lambda_1(\Sigma_T) \geqslant \lambda_1\{\Delta(\theta)[I_X^{\mathrm{F}}(\theta)]^{-1}\,\Delta(\theta)^{\mathscr{T}}\}.$$

The second of these inequalities provides a Cramér–Rao-type inequality for the determinant of $I_X^{\mathrm{F}}(\theta)$.

Acknowledgment

The authors wish to thank the referee for many useful comments.

RECEIVED: May 8, 1981; REVISED: April 30, 1982

References

AGGARWÁL, J. (1974), Sur l'information de Fisher, *in* "Théories de l'Information" (J. Kampé de Fériet, Ed.), pp. 111–117, Springer-Verlag, Berlin/New York.

BOEKEE, D. E. (1977), An extension of the Fisher information measure, *in* "Topics in Information Theory" (I. Csiszár and P. Elias, Eds.), pp. 113–123, North-Holland, Amsterdam.

BOEKEE, D. E. (1979), The D_f-information of order s, *in* "Transactions of the 8th Prague Conference on Information Theory, Statistical Decision Functions and Random Processes, 1978," C, pp. 55–66.

CSISZÁR, I. (1963), Eine informationstheoretische ungleichung und ihre anwendung auf den beweis der ergodizität von Markoffschen ketten, *Mayar Tud. Akad. Mat. Kutató Int. Kozl.* **8**, 85–108.

CSISZÁR, I. (1967), Information-type measures of difference of probability distributions and indirect observations, *Studia Sci. Math. Hungar.* **2**, 299–318.

FERENTINOS, K., AND PAPAIOANNOU, T. (1979), Loss of information due to groupings, *in* "Transactions of the 8th Prague Conference on Information Theory, Statistical Decision Functions and Random Processes, 1978," C, pp. 87–94.

FERENTINOS, K., AND PAPAIOANNOU, T. (1982), Information in experiments and sufficiency, *J. Statist. Plann. Inference* **6**, 309–317.

FISHER, R. A. (1925), Theory of statistical estimation, *Proc. Cambridge Philos. Soc.* **22**, 700–725.

FISHER, R. A. (1956), "Statistical Methods and Scientific Inference," Oliver & Boyd, Edinburgh.

GOEL, P., AND DEGROOT, M. (1979), Comparison of experiments and information measures, *Ann. Statist.* **7**, 1066–1077.

KAGAN, A. M. (1963), On the theory of Fisher's amount of information, *Sov. Math. Dokl.* **4**, 991–993.

KALE, K. (1964), A note on the loss of information due to grouping of observations, *Biometrika* **51**, 495–497.

KULLBACK, S., AND LEIBLER, A. (1951), On information and sufficiency, *Ann. Math. Statist.* **22**, 79–86.

MATHAI, A. (1967), Dispersion and information, *Metron* **26**, 1–12.

MATUSITA, K. (1967), On the notion of affinity of several distributions and some of its applications, *Ann. Inst. Statist. Math.* **19**, 181–192.

OKAMOTO, M., AND KANAZAWA, M. (1968), Minimization of eigenvalues of a matrix and optimality of principal components, *Ann. Math. Statist.* **39**, 859–863.

PAPAIOANNOU, T., AND KEMPTHORNE, O. (1971), "On Statistical Information Theory and Related Measures of Information," Aerospace Research Laboratories Report, ARL 71-0059, Wright-Patterson A.F.B., Ohio.

RÉNYI, A. (1961), On measures of entropy and information, *in* "Berkeley Symposium on Mathematical Statistics and Probability," Vol. I, pp. 547–561, Univ. of California Press, Berkeley.

RÉNYI, A. (1970), "Probability Theory," North-Holland, Amsterdam.

VAJDA, I. (1973), $\mathscr{X}^2$-divergence and generalized Fisher's information, *in* "Transactions of the 6th Prague Conference on Information Theory, Statistical Decision Functions and Random Processes, 1971," pp. 873–886.

WIJSMAN, R. (1973), On the attainment of the Cramér–Rao lower bound, *Ann. Statist.* **1**, 538–592.

ZACKS, S. (1971), "The Theory of Statistical Inference," Wiley, New York.