María Luisa Menéndez; Domingo Morales; Leandro Pardo; Miquel Salicrú
(h, Φ)-entropy differential metric

Applications of Mathematics, Vol. 42 (1997), No. 2, 81–98

Persistent URL: `http://dml.cz/dmlcz/134347`

(h, Φ)-ENTROPY DIFFERENTIAL METRIC

M. L. Menéndez, D. Morales, L. Pardo, Madrid, M. Salicrú, Barcelona[1]

(Received May 10, 1995)

Abstract. Burbea and Rao (1982a, 1982b) gave some general methods for constructing quadratic differential metrics on probability spaces. Using these methods, they obtained the Fisher information metric as a particular case. In this paper we apply the method based on entropy measures to obtain a Riemannian metric based on (h, Φ)-entropy measures (Salicrú et al., 1993). The geodesic distances based on that information metric have been computed for a number of parametric families of distributions. The use of geodesic distances in testing statistical hypotheses is illustrated by an example within the Pareto family. We obtain the asymptotic distribution of the information matrices associated with the metric when the parameter is replaced by its maximum likelihood estimator. The relation between the information matrices and the Cramér-Rao inequality is also obtained.

Keywords: (h, Φ)-entropy measures, information metric, geodesic distance between probability distributions, maximum likelihood estimators, asymptotic distributions, Cramér-Rao inequality.

MSC 2000: 62B10, 62H12

1. Introduction

Rao (1945) introduced a Riemannian metric as a measure of distance between two probability distributions. Burbea and Rao (1982a, 1982b) gave some general methods for constructing Riemannian metrics on probability spaces, of which the Fisher information metric belonged to special class. In view of the rich variety of possible metrics, it was found desirable to lay down some criteria for the choice of an appropriate metric for a given problem. Amari (1983) has stated that a metric should reflect the stochastic and statistical properties of the family of probability

[1] The research in this paper was supported in part by DGICYT Grants PB94-0308 and PB93-0022. Their financial support in gratefully acknowledged.

distributions. In particular he emphasized the invariance of the metric under transformations of the variable as well as of the parameters. Cencov (1982) showed that the Fisher information metric is unique under some conditions including invariance. Burbea and Rao (1982a) showed that the Fisher information metric is the only metric associated with divergence measures of the type introduced by Csiszar (1967). However, there exist other types of invariant metrics as Rao (1987) showed.

In Burbea and Rao (1982a) a metric based on the Hessian of the Φ-entropy functional is obtained. Salicrú et al (1993) established the necessity of introducing a more general entropy functional than the one introduced by Burbea and Rao (1982c), because there exist many entropy and uncertainty measures that are not particular cases of the Φ-entropy functional. To solve this problem they introduced the (h,Φ)-entropy in the following terms: Let $(\mathfrak{X}, \beta_{\mathfrak{X}}, P_\theta; \theta \in \Theta)$ be a statistical space, where Θ is an open subset of $\mathbb{R}^M$. Assume that there exists a probability density function (p.d.f.) $f(x,\theta)$ for the probability P_θ with respect to a σ-finite measure μ. Then the (h,Φ)-entropy associated with $f(x,\theta)$ is given by

$$H_\Phi^h(\theta) = h\Big[\int_{\mathfrak{X}} \Phi\big(f(x,\theta)\big)\,\mathrm{d}\mu(x)\Big], \tag{1}$$

where either $\Phi\colon [0,\infty) \to \mathbb{R}$ is concave and $h\colon \mathbb{R} \to \mathbb{R}$ is increasing or $\Phi\colon [0,\infty) \to \mathbb{R}$ is convex and $h\colon \mathbb{R} \to \mathbb{R}$ is decreasing. Note also that if h is increasing and Φ is convex or h is decreasing and Φ is concave, $H_\Phi^h(\theta)$ plays the role of a certainty function (e.g. van der Lubbe, 1977). In what follows, we assume that $H_\Phi^h(\theta)$ is an entropy function and that $h(x)$ and $\Phi(x)$ are two real valued C^3-functions defined on $\mathbb{R}$ and $[0,\infty)$, respectively. In the important particular case when the family $\{P_\theta; \theta \in \Theta\}$ is discrete, the entropies $H_\Phi^h(\theta)$ defined in this way have been considered by many authors, e.g. Vajda and Vasek (1985), where arbitrary Schur-concave entropies have been studied, and other references therein. In Tables 1 and 2 we present some examples of certainty and (h,Φ)-entropy measures:

Table 1. Certainty Measures

$\mathbf{h(x)}$	$\mathbf{\Phi(x)}$	**Certainty Measures**
x	x^2	Information Energy (Onicescu, 1966)
$x^{1/r}$	x^r	r-Norm (Van der Lubbe, 1977)
$x^{1/(r-1)}$	x^r	r-Mean (Van der Lubbe, 1981)
x^s	x^r	Generalized Measure of average Certainty (Van der Lubbe, 1981)

Table 2. (h,Φ)-Entropies

$\mathbf{h(x)}$	$\mathbf{\Phi(x)}$	$(\mathbf{h},\mathbf{\Phi})$**-Entropies**
x	$-x\log x$	Shannon (1948)
$(1-r)^{-1}\log x$	x^r	Renyi (1961)
$(m-r)^{-1}\log x$	x^{r-m+1}	Varma (1966)
$[m(m-r)]^{-1}\log x$	$x^{r/m}$	Varma (1966)
x	$(1-s)^{-1}(x^s-x)$	Havrda and Charvat (1967)
$(t-1)^{-1}(x^t-1)$	$x^{1/t}$	Arimoto (1971)
$(1-s)^{-1}\big[\exp\{(s-1)x\}-1\big]$	$x\log x$	Sharma and Mittal (1975)
$(1-s)^{-1}(x^{s-1/r-1}-1)$	x^r	Sharma and Mittal (1975)
x	$-x^r\log x$	Taneja (1975)
$(s-r)^{-1}x$	x^r-x^s	Sharma and Taneja (1975)
$\big(1+\frac{1}{\lambda}\big)\log(1+\lambda)-\frac{x}{\lambda}$	$(1+\lambda x)\log(1+\lambda x)$	Ferreri (1980)

In Section 2, we obtain a metric based on the Hessian of the (h,Φ)-entropy as well as the geodesic distances induced by the (h,Φ)-entropy for a particular selection of h and Φ and some probability distributions. The use of geodesic distances in testing statistical hypotheses is illustrated by an example wihin the Pareto family. In Section 3 we obtain the asymptotic distribution of the information matrices associated with the metric based on the Hessian of the (h,Φ)-entropy when the parameter is replaced by its maximum likelihood estimator. In Section 4 we obtain the relation between the information matrices and the Cramér-Rao inequality.

2. Information matrices associated to the (h,Φ)-entropy

Various procedures have been proposed in literature to introduce information matrices. Partial lists can be found in Ferentinos and Papaioannou (1981) or in Morales et al (1993). In this section we consider a differential geometric approach to this problem

Taking into account that each population can be characterized by a particular point θ of Θ, we may interpret $\{P_\theta\colon\ \theta\in\Theta\}$ as a manifold and view $\theta=(\theta_1,\dots,\theta_M)$ as a coordinate system. In general, it is also assumed that for any fixed $\theta\in\Theta$ the M functions

$$\frac{\partial f(x,\theta)}{\partial\theta_i},\quad i=1,\dots,M,$$

are linearly independent. Thus, the tangent space T_θ at point θ is the M-dimensional vector space spanned by

$$\Big[\frac{\partial f(x,\theta)}{\partial\theta_i}\Big]_{i=1,\dots,M}.$$

In this context the derivative of $H_\Phi^h(\theta)$ at $f(x,\theta)$ in the direction to the p.d.f. $g_1(x,\theta)$ is given by

$$\frac{\mathrm{d}}{\mathrm{d}t} H_\Phi^h\Big[f(x,\theta)+tg_1(x,\theta)\Big]_{t=0} = h'\Big[\int_{\mathfrak{X}} \Phi\big(f(x,\theta)\big)\,\mathrm{d}\mu(x)\Big] \times \int_{\mathfrak{X}} \Phi'\big(f(x,\theta)\big) g_1(x,\theta)\,\mathrm{d}\mu(x)$$

and the second derivative at $f(x,\theta)$ in the direction to the p.d.f. $g_2(x,\theta)$ is given by

$$\begin{aligned}\frac{\mathrm{d}^2}{\mathrm{d}t\,\mathrm{d}s} H_\Phi^h\big(f(x,\theta)+tg_1(x,\theta)+sg_2(x,\theta)\big)_{t=s=0} &= h''\Big[\int_{\mathfrak{X}} \Phi\big(f(x,\theta)\big)\,\mathrm{d}\mu(x)\Big] \\ \times \Big[\int_{\mathfrak{X}} \Phi'\big(f(x,\theta)\big) g_1(x,\theta)\,\mathrm{d}\mu(x)\Big]&\Big[\int_{\mathfrak{X}} \Phi'\big(f(x,\theta)\big) g_2(x,\theta)\,\mathrm{d}\mu(x)\Big] \\ + h'\Big[\int_{\mathfrak{X}} \Phi\big(f(x,\theta)\big)\,\mathrm{d}\mu(x)\Big]&\int_{\mathfrak{X}} \Phi''\big(f(x,\theta)\big) g_1(x,\theta) g_2(x,\theta)\,\mathrm{d}\mu(x)\end{aligned}$$

Then, when $g_1 = g_2 = g$, the Hessian is

$$\begin{aligned}\Delta_g H_\Phi^h(\theta) = h''\Big[\int_{\mathfrak{X}} \Phi\big(f(x,\theta)\big)\,\mathrm{d}\mu(x)\Big]\Big[\int_{\mathfrak{X}} \Phi'\big(f(x,\theta)\big) g(x,\theta)\,\mathrm{d}\mu(x)\Big]^2 \\ + h'\Big[\int_{\mathfrak{X}} \Phi\big(f(x,\theta)\big)\,\mathrm{d}\mu(x)\Big]\int_{\mathfrak{X}} \Phi''\big(f(x,\theta)\big) g(x,\theta)^2\,\mathrm{d}\mu(x)\end{aligned}$$

and the Hessian along the direction of the tangent space of the parameter space Θ is obtained by replacing g by

$$\mathrm{d}f(x,\theta) = \sum_{i=1}^{M} \frac{\partial f(x,\theta)}{\partial \theta_i}\,\mathrm{d}\theta_i.$$

Thus, we get the following result:

Theorem 1. *The Hessian of the (h,Φ)-entropy along the direction of the tangent space of the parameter space Θ is given by*

$$\mathrm{d}s^2_{(h,\Phi)}(\theta) = \sum_{ij=1}^{M} g_{ij}(\theta)\,\mathrm{d}\theta_i\,\mathrm{d}\theta_j,$$

where

$$g_{ij}(\theta) = h''\Bigg[\int\limits_{\mathfrak{X}} \Phi\big(f(x,\theta)\big)\,\mathrm{d}\mu(x)\Bigg]$$
$$\times \int\limits_{\mathfrak{X}} \Phi'\big(f(x,\theta)\big)\frac{\partial f(x,\theta)}{\partial\theta_i}\,\mathrm{d}\mu(x) \int\limits_{\mathfrak{X}} \Phi'\big(f(x,\theta)\big)\frac{\partial f(x,\theta)}{\partial\theta_j}\,\mathrm{d}\mu(x)$$
$$+ h'\Bigg[\int\limits_{\mathfrak{X}} \Phi\big(f(x,\theta)\big)\,\mathrm{d}\mu(x)\Bigg] \int\limits_{\mathfrak{X}} \Phi''\big(f(x,\theta)\big)\frac{\partial f(x,\theta)}{\partial\theta_i}\frac{\partial f(x,\theta)}{\partial\theta_j}\,\mathrm{d}\mu(x),$$

provided the integrals exist and are finite.

If $\mathrm{d}s^2_{(h,\Phi)}(\theta)$ is a positive definite quadratic form on the tangent space we have a differential metric of a Riemannian geometry because $\big(g_{ij}(\theta)\big)_{ij=1,\dots,M}$ defines a second order covariant tensor, since after performing a non-singular parameter change

$$\theta = (\theta_1,\dots,\theta_M) \to \overline{\theta} = (\overline{\theta}_1,\dots,\overline{\theta}_M),$$

we have

$$g_{ij}(\overline{\theta}) = \sum_{r,s=1}^{M} g_{rs}(\theta)\frac{\partial\theta_r}{\partial\overline{\theta}_i}\frac{\partial\theta_s}{\partial\overline{\theta}_j}.$$

To get that $\mathrm{d}s^2_{(h,\Phi)}(\theta)$ is a positive definite quadratic form when $H^h_\Phi(\theta)$ plays the role of an entropy measure, it is necessary to consider $-g_{ij}(\theta)$ instead of $g_{ij}(\theta)$. The metric $\mathrm{d}s^2_{(h,\Phi)}(\theta)$ and the matrix $IM^h_\Phi(\theta) = \big(g_{ij}(\theta)\big)_{ij=1,\dots,M}$ will be called the (h,Φ)-entropy metric and the (h,Φ)-entropy matrix, respectively.

If we consider a curve in $\{P_\theta\colon\ \theta\in\Theta\}$ connecting P_{θ_a} and P_{θ_b}, i.e.

$$\theta(t) = \big(\theta_1(t),\dots,\theta_M(t)\big), \quad t_a \leqslant t \leqslant t_b$$

with $\theta(t_a) = \theta_a$ and $\theta(t_b) = \theta_b$, then the distance between the probability density functions $f(x,\theta_a)$ and $f(x,\theta_b)$ along the curve $\theta(t)$ is given by

$$S(\theta_a,\theta_b) = \left|\int\limits_{t_a}^{t_b}\left[\sum_{i=1}^{M}\sum_{j=1}^{M} g_{ij}(\theta)\frac{\mathrm{d}\theta_i}{\mathrm{d}t}\frac{\mathrm{d}\theta_j}{\mathrm{d}t}\right]^{1/2}\mathrm{d}t\right|,$$

where, for ease of exposition, we have written θ, θ_i and θ_j instead of $\theta(t)$, $\theta_i(t)$ and $\theta_j(t)$, respectively. In particular, the curve connecting θ_a and θ_b with the shortest $S(\theta_a,\theta_b)$ is of interest. Such a curve is called a geodesic and is given as the solution of differential equations, the so-called Euler-Lagrange equations,

$$\sum_{i=1}^{M} g_{ik}(\theta)\frac{\mathrm{d}^2\theta_i}{\mathrm{d}t^2} + \sum_{i,j=1}^{M}[i,j;k]\frac{\mathrm{d}\theta_i}{\mathrm{d}t}\frac{\mathrm{d}\theta_j}{\mathrm{d}t} = 0, \quad j = 1,\dots,M,$$

where $[i,j;k]$ is the Christoffer symbol of the first kind which is defined by

$$[i,j;k]=\frac{1}{2}\Big[\frac{\partial g_{ki}(\theta)}{\partial\theta_j}+\frac{\partial g_{jk}(\theta)}{\partial\theta_i}-\frac{\partial g_{ij}(\theta)}{\partial\theta_k}\Big];\quad i,j,k=1,\dots,M.$$

The geodesic distance between θ_a and θ_b was proposed by Rao to measure the distance between distributions with parameters θ_a and θ_b. In our case, the geodesic pseudo-distance (a pseudo-distance satisfies all the postulates of distance except that it may vanish for elements which are distinct) induced by $\mathrm{d}s^2_{(h,\Phi)}(\theta)$ is denoted by $S_{(h,\Phi)}$ and called the (h,Φ)-pseudo-distance.

Remark 1. If we consider Renyi's entropy, that is $h(x)=(1-r)^{-1}\log x$ and $\Phi(x)=x^r$ $(r>0)$, we obtain

$$\begin{aligned}{}^r g_{ij}(\theta)=\frac{1}{r-1}\Big[\int_{\mathfrak{X}} f(x,\theta)^r\,\mathrm{d}\mu(x)\Big]^{-2}\int_{\mathfrak{X}} rf(x,\theta)^{r-1}\frac{\partial f(x,\theta)}{\partial\theta_i}\,\mathrm{d}\mu(x)\\ \cdot\int_{\mathfrak{X}} rf(x,\theta)^{r-1}\frac{\partial f(x,\theta)}{\partial\theta_i}\,\mathrm{d}\mu(x)-\Big[\int_{\mathfrak{X}} f(x,\theta)^r\,\mathrm{d}\mu(x)\Big]^{-1}\\ \int_{\mathfrak{X}} rf(x,\theta)^{r-2}\frac{\partial f(x,\theta)}{\partial\theta_i}\frac{\partial f(x,\theta)}{\partial\theta_j}\,\mathrm{d}\mu(x).\end{aligned}$$

The metric $\mathrm{d}s^2_r(\theta)=\sum\limits_{i,j=1}^{M}{}^r g_{ij}(\theta)\,\mathrm{d}\theta_i\,\mathrm{d}\theta_j$, $r\in\mathbb{R}^+$, will be called Renyi's entropy metric, the matrix $IM_r(\theta)=\big({}^r g_{ij}(\theta)\big)_{ij=1,\dots,M}$, Renyi's entropy matrix and the geodesic pseudo-distance induced by $\mathrm{d}s^2_r(\theta)$, S_r, Renyi's entropy pseudo-distance. The special case of $r\to1$ corresponds to Shannon's entropy, which is widely used in applied research (see Burbea and Rao (1982a)). In this case we obtain the information metric of Rao (1945), while $IM^h_\Phi(\theta)$ is the Fisher information matrix $I^F_x(\theta)$.

Remark 2. If $\theta\in\Theta\subset\mathbb{R}$, we have $\mathrm{d}s^2_{(h,\Phi)}(\theta)=IM^h_\Phi(\theta)\,\mathrm{d}\theta^2$ and the distance between the distributions $f(x,\theta_a)$ and $f(x,\theta_b)$ is given by

$$S_{(h,\Phi)}(\theta_a,\theta_b)=\Big|\int_{\theta_a}^{\theta_b}\big(IM^h_\Phi(\theta)\big)^{1/2}\,\mathrm{d}\theta\Big|.$$

We present Renyi's entropy pseudo-distance for some probability distributions:

Bernoulli (θ)**:**

$$\begin{aligned}&IM_r(\theta)\\&=\frac{r\big[(r-1)\theta^{2r-2}+(r-1)(1-\theta)^{2r-2}+2(r-1)\theta^{r-1}(1-\theta)^{r-2}-\theta^{r-2}(1-\theta)^{r-2}\big]}{(r-1)\big(\theta^r+(1-\theta)^r\big)^2}\end{aligned}$$

and for $r = 2$ we have

$$S_2(\theta_a, \theta_b) = \Big| \frac{1}{2^{1/2}} \arctan \frac{A(\theta_b)}{2^{1/2}} - \frac{1}{2} \arctan \frac{A(\theta_b)}{2} - \frac{1}{2^{1/2}} \arctan \frac{A(\theta_a)}{2^{1/2}} + \frac{1}{2} \arctan \frac{A(\theta_a)}{2} \Big|,$$

where $A(\theta) = (2\theta - 1)\big(\theta(1-\theta)\big)^{-1/2}$.

Geometric (θ)**:**

$$IM_r(\theta) = \frac{r\big[(r-1)\theta^{2r-2} + (1-\theta)^{r-2}(\theta^2 + 2\theta - 2) + 1 - r\theta^2(1-\theta)^{r-2}\big]}{(r-1)\big(1 - (1-\theta)^r\big)^2\theta^2}, r \neq 1,$$

and for $r \to 1$ we have

$$S_r(\theta_a, \theta_b) = \Big| \log \frac{\theta_b}{2 - \theta_b + 2(1-\theta_b)^{1/2}} - \log \frac{\theta_a}{2 - \theta_a + 2(1-\theta_a)^{1/2}} \Big|.$$

Exponential (θ)**:**

$$S_r(\theta_a, \theta_b) = \Big| \frac{2-r}{r} \Big|^{1/2} \big| \log \theta_b - \log \theta_a \big|, \quad 0 < r < 2.$$

Pareto (θ)**: (**x_0 **fixed)**

$$IM_r(\theta) = \frac{r\theta^{r-2}}{(\theta r + r - 1)x_0^{r-1}} \Big\{ 1 - \frac{2\theta}{\theta r + r - 1} + \frac{2\theta^2}{(\theta r + r - 1)^2} \Big\}$$

and for $r = 1$ we have

$$S_1(\theta_a, \theta_b) = |\log \theta_a - \log \theta_b|.$$

Erlang (θ, n)**: (**n **fixed)**

$$S_r(\theta_a, \theta_b) = \Big| \frac{[2 - r(2-n)]}{r} \Big|^{1/2} |\log \theta_b - \log \theta_a|.$$

Normal (μ, σ^2)**: (**σ **fixed)**

$$S_r(\mu_a, \mu_b) = \frac{|\mu_b - \mu_a|}{\sigma}$$

so that in this case the Rao distance is constant for all $r > 0$.

Normal (μ, σ^2)**: (**μ **fixed)**

$$S_r(\sigma_a^2, \sigma_b^2) = \Big| \frac{3-r}{4r} \Big|^{1/2} |\log \sigma_b - \log \sigma_a|, \quad 0 < r < 3.$$

Normal (μ, σ^2)**: (**μ **and** σ **variable)**

Renyi's entropy metric is given by

$$\mathrm{d}s_r^2(\theta) = \mathrm{d}\mu^2/\sigma^2 + \big((3-r)/4r\big)\,\mathrm{d}\sigma^2/\sigma^2, \quad 0 < r < 3.$$

For $\mu^* = \big((3-r)/4r\big)^{-1/2}\mu$ and $\sigma^* = \sigma$ we have

$$\mathrm{d}s_r^2(\theta) = \frac{3-r}{4r}\left[\frac{(\mathrm{d}\mu^*)^2}{(\sigma^*)^2} + \frac{(\mathrm{d}\mu^*)^2}{(\sigma^*)^2}\right],$$

which is the Poincaré metric. Following the method described by Burbea (1986), Renyi's entropy pseudodistance between (μ_1, σ_1^2) and (μ_2, σ_2^2), or equivalently between $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, is

$$S_r\big[(\mu_1, \sigma_1^2), (\mu_2, \sigma_b^2)\big] = \left(\frac{3-r}{4r}\right)^{1/2} \log\left(\frac{1+\delta}{1-\delta}\right) = \left(\frac{3-r}{4r}\right)^{1/2} \tanh^{-1}(\delta),$$

where

$$\delta = \left(\frac{(\mu_1^* - \mu_2^*)^2 + (\sigma_1^* - \sigma_2^*)^2}{(\mu_1^* - \mu_2^*)^2 + (\sigma_1^* + \sigma_2^*)^2}\right)^{1/2} = \left(\frac{(\mu_1 - \mu_2)^2 + \frac{3-r}{4r}(\sigma_1 - \sigma_2)^2}{(\mu_1 - \mu_2)^2 + \frac{3-r}{4r}(\sigma_1 + \sigma_2)^2}\right)^{1/2}.$$

If $\mu_1 = \mu_2$ then geodesic curve connecting (μ_1, σ_1^2) and (μ_1, σ_1^2) lies on the straight line $\mu =$ constant, and the distance is $S_r(\sigma_1^2, \sigma_2^2)$ which coincides with the distance obtained above. On the other hand, if $\sigma_1^2 = \sigma_2^2$, then the present distance differs from the distance obtained above, since $\sigma^2 =$ constant is not a geodesic curve of the present metric.

Multivariate normal (μ, Σ)**: (**Σ **fixed)**

$$S_r(\mu_1, \mu_2) = \Big((\mu_1 - \mu_2)^t \Sigma^{-1} (\mu_1 - \mu_2)\Big)^{1/2},$$

which is the square root of the well-known Mahalanobis distance.

E x a m p l e 1. When dealing with parametric distributions, statistical tests based on Renyi's entropy pseudo-distance can be constructed by substituting one or two parameters by convenient estimators. To clarify this idea, let us consider the Pareto distribution (x_0 fixed) and the Rao distance. The test of the hypothesis $H_0\colon\ \theta = \theta_0$ is equivalent to the test of the hypothesis $S_1(\theta_1, \theta_2) = 0$, so that we can use the statistics $T = S_1(\hat{\theta}, \theta_0) = |\log\hat{\theta} - \log\theta_0|$, where $\hat{\theta}$ is a suitable estimator of θ. We reject the null hypothesis, at a level α, if $T > c_\alpha$, where $P_{\theta_0}(T > c_\alpha) = \alpha$.

For the maximum likelihood estimator (M.L.E.)

$$\hat{\theta} = \frac{n}{\sum\limits_{i=1}^{n} \log \frac{x_i}{x_0}},$$

a straightforward calculus yields the following Rao distance test (R.D.T.):

$$\Phi(x_1, \dots, x_n) = \begin{cases} 0 & \text{if} \quad c_1 < S(x_1, \dots, x_n) < c_2 \\ 1 & \text{if} \quad S(x_1, \dots, x_n) < c_1 \text{ or } S(x_1, \dots, x_n) > c_2 \end{cases}$$

where

$$F_{\chi^2_{2n}}(c_2) - F_{\chi^2_{2n}}(c_1) = 1 - \alpha, \tag{1}$$

$$c_1 c_2 = 4n^2, \tag{2}$$

and

$$S(x_1, \dots, x_n) = 2\theta_0 \sum_{i=1}^{n} \log \frac{x_i}{x_0}.$$

The size condition (1) is obtained form the fact that S is chi-square distributed with $2n$ degrees of freedom under the null hypothesis. Condition (2) follows from the relations $c_1 = 2n \exp\{-c_\alpha\}$ and $c_2 = 2n \exp\{c_\alpha\}$.

From the uniformly minimum variance unbiased estimator (U.M.V.U.E.)

$$\widetilde{\theta} = \frac{n-1}{\sum\limits_{i=1}^{n} \log \frac{x_i}{x_0}},$$

the modified Rao distance test (M.R.D.T.) is obtained if we replace (2) by

$$c_1 c_2 = 4(n-1)^2. \tag{2$'$}$$

Finally, the likelihood ratio test (L.R.T.) is in this case the uniformly most powerful unbiased test (U.M.P.U.T.) and coincides with the above two tests except for condition (2), which has to be replaced by

$$2n \log \frac{c_2}{c_1} = c_2 - c_1. \tag{2$''$}$$

To compare the above three decision rules (R.D.T., M.R.D.T. and L.R.T.), the exact powers are illustrated in Figure 1 for $H_0\colon \theta = 1$, $H_1\colon \theta \neq 1$, $\alpha = 0.05$ and $n = 10$.

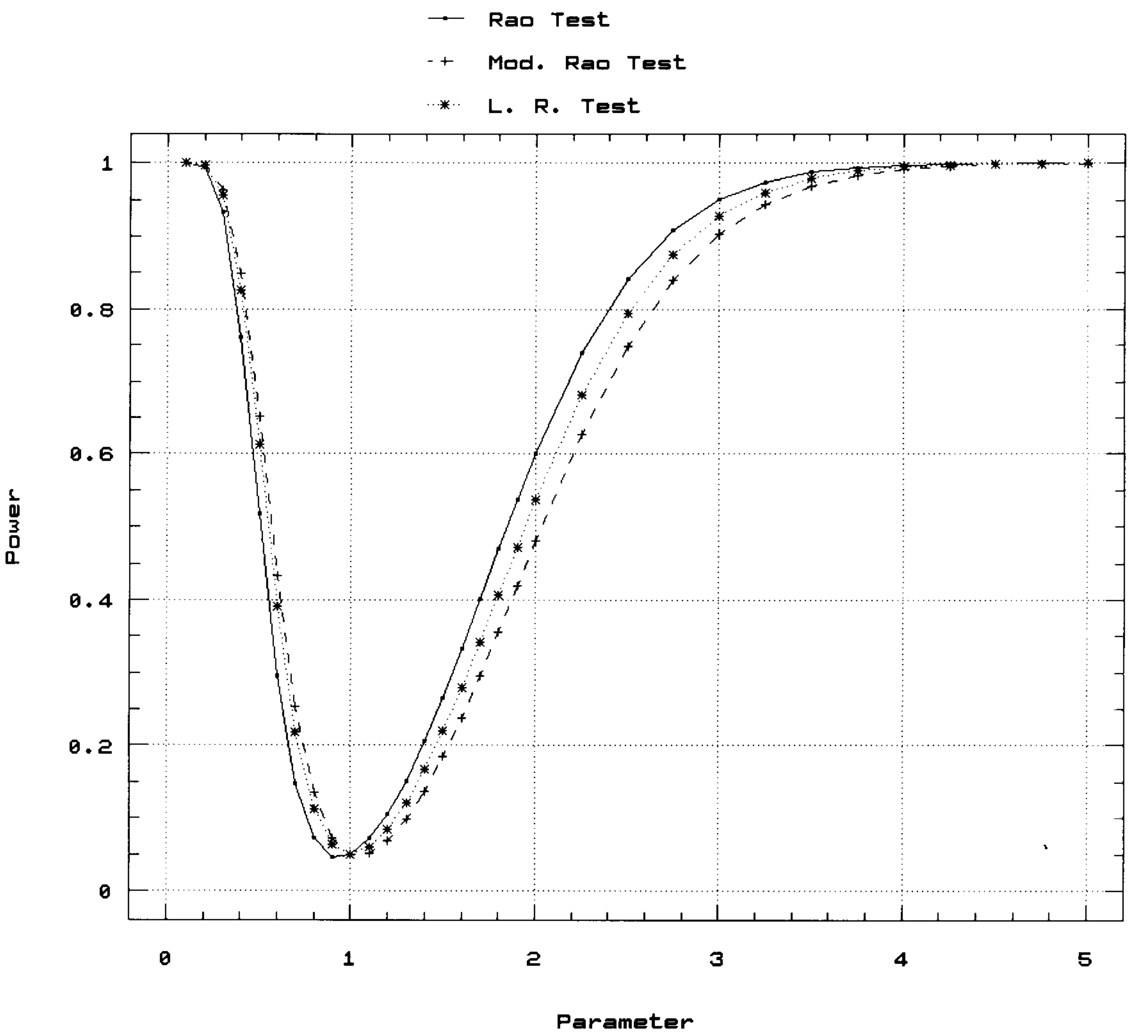

Figure 1

3. Asymptotic distribution of information matrices

When dealing with parameter distributions, statistical tests based on information matrices can be constructed by substituting the parameter θ by an estimator $\hat{\theta}$. In general, it will not be possible to get the exact distribution of the statistic $IM^{h}_{\underline{\Phi}}(\hat{\theta}) = \left(g_{ij}(\hat{\theta})\right)_{ij}$ and we will have to use its asymptotic distribution. In the next theorem we obtain the asymptotic distribution of $IM^{h}_{\underline{\Phi}}(\hat{\theta})$, where $\hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_M)$ is the maximum likelihood estimator of θ based on a random sample of size n. We suposse that the following regularity assumptions hold:

(i) The set $A = \{x \in \mathfrak{X}/f(x,\theta) > 0\}$ does not depend on θ and for all $x \in A$, $\theta \in \Theta$

$$\frac{\partial f(x,\theta)}{\partial \theta_i}, \quad \frac{\partial^2 f(x,\theta)}{\partial \theta_i \partial \theta_j}, \quad \frac{\partial^3 f(x,\theta)}{\partial \theta_i \partial \theta_j \partial \theta_k}, \quad i,j,k = 1,\dots,M$$

exist and are finite.

(ii) There exist real valued functions $F(x)$ and $H(x)$ such that

$$\left|\frac{\partial f(x,\theta)}{\partial \theta_i}\right| < F(x), \quad \left|\frac{\partial^2 f(x,\theta)}{\partial \theta_i \partial \theta_j}\right| < F(x), \quad \left|\frac{\partial^3 f(x,\theta)}{\partial \theta_i \partial \theta_j \partial \theta_k}\right| < H(x),$$

where F is finitely integrable and $E\big[H(X)\big] < M$ with M independent of θ.

(iii) $I_X^F(\theta) = \left[E\left\{\frac{\partial \log f(X,\theta)}{\partial \theta_i}\frac{\partial \log f(X,\theta)}{\partial \theta_j}\right\}\right]_{i,j=1,\dots,M}$ is finite and positive definite.

Theorem 2. *If assumptions* (i), (ii) *and* (iii) *hold then*

$$n^{1/2}\big(\operatorname{Vec} IM_\Phi^h(\hat\theta) - \operatorname{Vec} IM_\Phi^h(\theta)\big) \xrightarrow[n\to\infty]{L} N\big(0, A(\theta) I_X^F(\theta)^{-1} A(\theta)^t\big),$$

where

$$\begin{aligned}
\operatorname{Vec} IM_\Phi^h(\hat\theta) &= \big(g_{11}(\hat\theta),\dots,g_{MM}(\hat\theta)\big)^t\\
\operatorname{Vec} IM_\Phi^h(\hat\theta) &= \big(g_{11}(\theta),\dots,g_{MM}(\theta)\big)^t\\
A(\theta) &= \begin{bmatrix} \frac{\partial g_{11}(\theta)}{\partial \theta_1}, & \dots, & \frac{\partial g_{11}(\theta)}{\partial \theta_M}\\ \vdots & & \vdots \\ \frac{\partial g_{MM}(\theta)}{\partial \theta_1}, & \dots, & \frac{\partial g_{MM}(\theta)}{\partial \theta_M}\end{bmatrix}
\end{aligned}$$

P r o o f. The Taylor expansion of $IM_{\underline{\Phi}}^h(\hat\theta) = \big(g_{ij}(\hat\theta)\big)_{ij}$ at the point θ yields

$$\operatorname{Vec} IM_\Phi^h(\hat\theta) = \operatorname{Vec} IM_\Phi^h(\theta) + \left[\sum_{i=1}^M \frac{\partial g_{11}}{\partial \theta_i}(\hat\theta_i - \theta_i),\dots,\sum_{i=1}^M \frac{\partial g_{MM}}{\partial \theta_i}(\hat\theta_i - \theta_i)\right]^t + R_n.$$

Since $n^{1/2} R_n \xrightarrow[n\to\infty]{L} 0$, the random variables

$$n^{1/2}\big(\operatorname{Vec} IM_\Phi^h(\hat\theta) - \operatorname{Vec} IM_\Phi^h(\theta)\big) \quad \text{and} \quad A(\theta) n^{1/2}(\hat\theta - \theta)$$

have asymptotically the same distribution, where the matrix $A(\theta)$ is given above. Hence the result follows.

We now consider the generalized inverse for any matrix $A(m \times n)$. As is well-known $A_g(n \times m)$ is called the g-inverse (generalized inverse) of A if

(i) $AA_gA = A$,
(ii) AA_g is symmetric,
(iii) A_gA is symmetric,
(iv) $A_gAA_g = A_g$.

Some properties of generalized inverse matrices are stated below.

(1) There exists only one matrix A_g satisfying (i)–(iv),
(2) $(A^t)_g = (A_g)^t$,
(3) $(A_g)_g = A$. □

Remark 3. If the random variable $X = (X_1, \dots, X_m)^t$ has a multivariate normal distribution with vector of means zero and variance-covariance matrix V, where V has rank r ($r \leqslant m$) and V_g is the generalized inverse of V, then X^tV_gX, is chi-square distributed with r degrees of freedom (see Muirhead 1982, pp. 30, Theorem 1.4.4.).

Now on the basis of Remark 3 and using Theorem 2 we can obtain the following tests of hypotheses:

a) Goodness of fit test

$$H_0\colon\ IM_\Phi^h(\theta) = IM_\Phi^h(\theta_0).$$

In this case we consider the statistic

$$T_1 = n\big(\operatorname{Vec} IM_\Phi^h(\hat\theta) - \operatorname{Vec} IM_\Phi^h(\theta_0)\big)^t \big(AI_X^F(\theta)^{-1}A^t\big)_g \big(\operatorname{Vec} IM_\Phi^h(\hat\theta) - \operatorname{Vec} IM_\Phi^h(\theta_0)\big),$$

where, for ease of exposition, we write A instead of $A(\theta)$. By Theorem 2 and Remark 3, T_1 is asymptotically chi-square distributed with s degrees of freedom, where $s = \operatorname{rank}\big(AI_X^F(\theta)^{-1}A^t\big)$. We reject the null hypothesis if $T_1 \geqslant \chi^2_{s,\alpha}$.

If we collect r independent samples of sizes $n_1, \dots, n_r$ ($n = n_1 + \ldots + n_r$) from the population associated with $f(x,\theta_1), \dots, f(x,\theta_r)$, respectively, then we can test the homogeneity of information matrices (with or without a specified common information matrix).

b) Test of homogeneity with a known information matrix.

$$H_0\colon\ IM_\Phi^h(\theta_1) = IM_\Phi^h(\theta_2) = \ldots = IM_\Phi^h(\theta_r) = IM_\Phi^h(\theta_0).$$

In this case we consider the statistic

$$T_2 = \sum_{i=1}^r n_i \big[\operatorname{Vec} IM_\Phi^h(\hat\theta_i) - \operatorname{Vec} IM_\Phi^h(\theta_0)\big]^t \big[A_iI_X^F(\theta_i)^{-1}A_i^t\big]g \times \big[\operatorname{Vec} IM_\Phi^h(\hat\theta_i) - \operatorname{Vec} IM_\Phi^h(\theta_0)\big]$$

where, for ease of exposition, we write A_i instead of $A_i(\theta)$. By Theorem 2, Remark 3 and the independence of the r random samples, T_2 is asymptotically distributed as a chi-square distribution with s degrees of freedom, where $s=\sum_{i=1}^{r}\operatorname{rank}\left(A_i I_X^F(\theta_i)^{-1}A_i^t\right)$. We reject the null hypothesis if $T_2\geqslant\chi^2_{s,\alpha}$.

c) Test of homogeneity

$$H_0\colon\ IM^{h}_{\underline{\Phi}}(\theta_1)=IM^{h}_{\underline{\Phi}}(\theta_2)=\ldots=IM^{h}_{\underline{\Phi}}(\theta_r).$$

This test is based on a lemma. First, we introduce the following notation:

$$\Sigma_i=\left(A_i I_X^F(\theta_i)^{-1}A_i^t\right)_g,\quad B=\sum_{i=1}^{r}n_i\Sigma_i,\quad \hat{Y}_i=B_gB\operatorname{Vec}IM^h_{\Phi}(\theta_i),$$

$$Y_0=B_gB\operatorname{Vec}IM^h_{\Phi}(\theta_0),\quad \overline{Y}=B_g\left[\sum_{i=1}^{r}n_i\Sigma_iB_gB\hat{Y}_j\right],$$

$$Y=(Y_1,\ldots,Y_r)^t,\quad \hat{Y}=(\hat{Y}_1,\ldots,\hat{Y}_r)^t$$

and

$$C=\begin{pmatrix} I-n_1B_g\Sigma_1B_gB & -n_2B_g\Sigma_2B_gB & \ldots & -n_rB_g\Sigma_rB_gB\\ -n_2B_g\Sigma_2B_gB & I-n_2B_g\Sigma_2B_gB & \ldots & -n_rB_g\Sigma_rB_gB\\ \vdots & & \ldots & \vdots\\ -n_1B_g\Sigma_1B_gB & -n_2B_g\Sigma_2B_gB & \ldots & I-n_rB_g\Sigma_rB_gB\end{pmatrix}_{(M^2r)\times(M^2r)}$$

For

$$\lambda_i=\lim_{n_i\to\infty}\frac{n_i}{n}\in(0,1),\quad i=1,\ldots,r$$

let us consider the $(M^2r\times M^2r)$ block diagonal matrix

$$\Delta=\frac{\prod_{i=1}^{r}\lambda_i}{\lambda_1}\left(B_gB(\Sigma_1)_gBB_g\right)\oplus\ldots\oplus\frac{\prod_{i=1}^{r}\lambda_i}{\lambda_r}\left(B_gB(\Sigma_r)_gBB_g\right).$$

Lemma 1. *If* (i), (ii) *and* (iii) *hold and*

$$\lambda_i=\lim_{n_i\to\infty}\frac{n_i}{n}\in(0,1),\quad i=1,\ldots,r$$

then the statistic

$$T_3=\left[\frac{\prod_{i=1}^{r}n_i}{n^{r-1}}\right](\hat{Y}_1-\overline{Y},\ldots,\hat{Y}_r-\overline{Y})^t\left[C\Delta C^t\right]_g(\hat{Y}_1-\overline{Y},\ldots,\hat{Y}_r-\overline{Y})^t$$

is asymptotically chi-square distributed with

$$s = \operatorname{rank} C\Delta C^t$$

degrees of freedom.

P r o o f. We know that

$$n^{1/2}(\hat{Y}_i - Y_i) = B_g B n_i^{1/2}\left[\operatorname{Vec} IM_\Phi^h(\hat{\theta}_i) - \operatorname{Vec} IM_\Phi^h(\theta_i)\right] \xrightarrow[n_i\to\infty]{L} N\big(0, B_g B(\Sigma_i) g B B_g\big),$$

and therefore

$$\left[\frac{\prod_{i=1}^r n_i}{n^{r-1}}\right](\hat{Y}_1 - Y, \ldots, \hat{Y}_r - Y)^t \xrightarrow[n_1,\ldots,n_r\to\infty]{L} N(0, \Delta).$$

Now, taking into account that

$$\big(C(\hat{Y} - Y)\big)^t = (\hat{Y}_1 - \overline{Y}, \hat{Y}_2 - \overline{Y}, \ldots, \hat{Y}_r - \overline{Y})^t$$

we obtain the announced result. □

On the basis of this lemma we reject the null hypothesis if $T_3 > \chi^2_{s,\alpha}$.

R e m a r k 4. When Σ_i, $i = 1, \ldots, r$ and B are non-singular matrices, the statistic

$$\begin{aligned} T_3' = \sum_{i=1}^r n_j &\left[\operatorname{Vec} IM_\Phi^h(\hat{\theta}) - \overline{\operatorname{Vec} I} M_\Phi^h(\theta)\right]^t \left[A_i I_X^F(\theta_i)^{-1} A_i^t\right]_g \\ &\times \left[\operatorname{Vec} IM_\Phi^h(\hat{\theta}_i) - \overline{\operatorname{Vec} IM}_\Phi^h(\theta)\right] \end{aligned}$$

is asymptotically chi-square distributed with

$$s = \sum_{i=1}^r \operatorname{rank}(\Sigma_i) - \operatorname{rank}(B) = rM^2 - M^2 = M^2(r-1)$$

degrees of freedom, and

$$\overline{\operatorname{Vec} I} M_\Phi^h(\theta) = B_g\left[\sum_{i=1}^r n_i \Sigma_i \operatorname{Vec} IM_\Phi^h(\theta_i)\right].$$

To this end, it is easy to see that

$$\sum_{i=1}^{r} n_i \Big[\operatorname{Vec} IM_\Phi^h(\hat{\theta}_i) - \operatorname{Vec} IM_\Phi^h(\theta) \Big]^t \Sigma_i \Big[\operatorname{Vec} IM_\Phi^h(\hat{\theta}_i) - \operatorname{Vec} IM_\Phi^h(\theta) \Big]$$

$$= \sum_{i=1}^{r} n_i \Big[\operatorname{Vec} IM_\Phi^h(\hat{\theta}_i) - \overline{\operatorname{Vec} I} M_\Phi^h(\theta) \Big]^t \Sigma_i \Big[\operatorname{Vec} IM_\Phi^h(\hat{\theta}_i) - \overline{\operatorname{Vec} IM}_\Phi^h(\theta) \Big]$$

$$+ \sum_{i=1}^{r} n_i \Big[\overline{\operatorname{Vec} IM}_\Phi^h(\theta) - \operatorname{Vec} IM_\Phi^h(\theta) \Big]^t \Sigma_i \Big[\overline{\operatorname{Vec} IM}_\Phi^h(\theta) - \operatorname{Vec} IM_\Phi^h(\theta) \Big].$$

Also, if $\lim\limits_{n\to\infty} \frac{n_j}{n} \in (0,1)$, $i = 1, \dots, r$, then

$$\sum_{i=1}^{r} n_j \Big[\operatorname{Vec} IM_\Phi^h(\hat{\theta}_i) - \operatorname{Vec} IM_\Phi^h(\theta) \Big]^t \Sigma_i \Big[\operatorname{Vec} IM_\Phi^h(\hat{\theta}_i) - \operatorname{Vec} IM_\Phi^h(\theta) \Big] \xrightarrow{L} \chi_a^2$$

with $a = \sum\limits_{i=1}^{r} \operatorname{rank}(\Sigma_i) = rM^2$, and

$$\sum_{i=1}^{r} n_i \Big[\overline{\operatorname{Vec} IM}_\Phi^h(\theta) - \operatorname{Vec} IM_\Phi^h(\theta) \Big]^t \Sigma_i \Big[\overline{\operatorname{Vec} IM}_\Phi^h(\theta_i) - \operatorname{Vec} IM_\Phi^h(\theta) \Big] \xrightarrow{L} \chi_b^2$$

with $b = \operatorname{rank}(B) = M^2$. This implies the announced result.

4. Relation to Cramer-Rao's inequality

Let $X_1, \dots, X_n$ be a sample from a random variable X with p.d.f. $f(x, \theta)$. Let $\Psi_1(\theta), \dots, \Psi_s(\theta)$ be functions with continuous first partial derivatives and let $T_1 = T_1(X_1, \dots, X_n), \dots, T_s = t_s(X_1, \dots, X_n)$ be unbiased estimators of $\Psi_1(\theta), \dots, \Psi_s(\theta)$, respectively. We denote by Σ the variance-covariance matrix of $(T_1, \dots, T_s)$. If $IM_\Phi^h(\theta)$ is positive definite, it can be taken to be the variance-covariance matrix of some random variables $Y_1, \dots, Y_n$. If we define $\delta_{ij} = \operatorname{Cov}(T_i, Y_j)$ and $\delta = (\delta_{ij})_{ij=1,\dots,M}$, then we get the following inequality.

Theorem 4. $\Sigma - \delta\big(IM_\Phi^h(\theta)\big)^{-1}\delta^t$ *is positive semidefinite.*

Proof. The matrix

$$\begin{pmatrix} \Sigma & \vdots & \delta \\ \dots & \dots & \dots \\ \delta^t & \vdots & IM_\Phi^h(\theta) \end{pmatrix}$$

is positive semidefinite because it is the variance-covariance matrix of the random variable $(T_1, \dots, T_s, Y_1, \dots, Y_n)$. Then

$$\begin{pmatrix} I_s & \vdots & -\delta IM_\Phi^h(\theta)^{-1} \\ \cdots & \cdots & \cdots \\ 0 & \vdots & IM_\Phi^h(\theta)^{-1} \end{pmatrix} \begin{pmatrix} \Sigma & \vdots & \delta \\ \cdots & \cdots & \cdots \\ \delta^t & \vdots & IM_\Phi^h(\theta) \end{pmatrix} \begin{pmatrix} I_s & \vdots & 0 \\ \cdots & \cdots & \cdots \\ -\big(\delta IM_\Phi^h(\theta)^{-1}\big)^t & \vdots & IM_\Phi^h(\theta)^{-1} \end{pmatrix}$$
$$= \begin{pmatrix} \Sigma - \delta\big(IM_\Phi^h(\theta)\big)^{-1}\delta^t & \vdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \vdots & IM_\Phi^h(\theta)^{-1} \end{pmatrix}$$

is a positive semidefinite matrix, and so is $\Sigma - \delta\big(IM_\Phi^h(\theta)\big)^{-1}\delta^t$. □

Corollary 1. *If $\Phi(x) = -x \log x$ and $h(x) = x$ then the Cramer-Rao inequality is obtained so that is this case*

$$\Sigma - \delta I_X^F(\theta)^{-1}\delta^t \text{ is positive semidefinite.}$$

Here $I_X^F(\theta)$ is the Fisher information matrix and $Y_i = \frac{\partial}{\partial\theta_i} \log f(x, \theta)$.

Appendix A
(Pseudo-distance for Erlang distributions)

The Erlang family of distributions is defined by p.d.f.

$$f(x, \theta) = \frac{\theta^n}{\Gamma(n)} x^{n-1} \exp(-\theta x), \quad x \in \mathbb{R}^+ \quad (\theta > 0,\ n \in \mathbb{N}).$$

For a fixed $n \in N$, Renyi's entropy metric is given by

$$\mathrm{d}s_r^2(\theta) = {}^r g_{11}(\theta)\, \mathrm{d}\theta^2$$

where ${}^r g_{11}(\theta)$ is defined in Remark 1 by

$${}^r g_{11}(\theta) = \frac{1}{r-1}\Bigg\{\bigg[\int_0^\infty f(x,\theta)^r \,\mathrm{d}x\bigg]^{-2}\bigg[\int_0^\infty r f(x,\theta)^{r-1}\frac{\partial f(x,\theta)}{\partial\theta}\,\mathrm{d}x\bigg]^2$$
$$- \bigg[\int_0^\infty f(x,\theta)^r \,\mathrm{d}x\bigg]^{-1}\int_0^\infty r f(x,\theta)^{r-2}\Big[\frac{\partial f(x,\theta)}{\partial\theta}\Big]^2 \mathrm{d}x\Bigg\}.$$

After some calculations obtain

$$\int_0^\infty f(x,\theta)^r \,\mathrm{d}x = \frac{\Gamma(rn-r+1)}{\Gamma(n)^r \theta^{1-r} r^{rn-r+1}},$$

$$\int_0^\infty r f(x,\theta)^{r-1} \Big[\frac{\partial f(x,\theta)}{\partial\theta}\Big] \,\mathrm{d}x = \frac{(r-1)\Gamma(rn-r+1)}{\Gamma(n)^r r^{(n-1)r+1} \theta^{-r+2}}$$

and

$$\int_0^\infty r f(x,\theta)^{r-2} \Big[\frac{\partial f(x,\theta)}{\partial\theta}\Big]^2 \,\mathrm{d}x = \frac{\Gamma(nr-r+1)(r^2+rn-3r+2)}{r^{nr-r+2}\theta^{3-r}\Gamma(n)^r}.$$

This obviously leads to

$$g_{11}(\theta) = \frac{r(n-2)+2}{\theta^2 r}.$$

Now Renyi's entropy pseudodistance between two Erlang distributions, $f(x,\theta_a)$ and $f(x,\theta_b)$, is given by

$$S_r(\theta_a,\theta_b) = \left| \left[\frac{r(n-2)+2}{r}\right]^{1/2} \int_{\theta_a}^{\theta_b} \theta^{-1}\,\mathrm{d}\theta \right| = \left|\frac{r(n-2)+2}{r}\right|^{1/2} |\log\theta_b - \log\theta_a|.$$

References

[1] *S. I. Amari*: A foundation of information geometry. Electronics and Communications in Japan *66-A* (1983), 1–10.
[2] *C. Atkinson, A. F. S. Mitchell*: Rao's distance measure. Sankhya *43* (1981), 345–365.
[3] *S. Arimoto*: Information-theoretical considerations on estimation problems. Information and Control *19* (1971), 181–194.
[4] *J. Burbea*: Informative geometry of probability spaces. Expositions Mathematicae *4* (1986), 347–378.
[5] *J. Burbea, C. R. Rao*: Entropy differential metric, distance and divergence measures in probability spaces: A unified approach. J. Multivariate Analysis *12* (1982a), 575–596.
[6] *J. Burbea, C. R. Rao*: On the convexity of some divergence measures based on entropy functions. IEEE Trans. on Information Theory *IT-28* (1982b), no. 3, 489–495.
[7] *J. Burbea, C. R. Rao*: On the convexity of higher order Jensen differences based on entropy functions. IEEE Trans. on Information Theory *IT-28* (1982c), no. 6, 961–963.
[8] *N. N. Cencov*: Statistical Decision Rules and Optimal Inference. Trans. of Mathematical Monographs 53, Amer. Math. Soc., Providence (1982).
[9] *I. Csiszár*: Information-type measures of difference of probability distributions and indirect observations. Studia Scientiarum Mathematicarum Hungarica *2* (1967), 299–318.
[10] *K. Ferentinos, T. Papaioannou*: New parametric measures of information. Information and Control *51* (1981), 193–208.
[11] *C. Ferreri*: Hypoentropy and related heterogeneity divergence measures. Statistica *40* (1980), 55–118.

[12] *J. Havrda, F. Charvat*: Concept of structural α-entropy. Kybernetika *3* (1967), 30–35.
[13] *D. Morales, L. Pardo, L. Salicrú, M. L. Menéndez*: New parametric measures of information based on generalized R-divergences. Multivariate Analysis. Future directions 2. Editors C. M. Cuadras and C. R. Rao. North Holland (1993), 473–488.
[14] *R. J. Muirhead*: Aspect of Multivariate Statistical Theory. John Wiley and Sons. Inc. (1982).
[15] *O. Onicescu*: Energie Informationnelle. C. R. Acad. Sci. Paris. Ser A *263* (1966), 841–842.
[16] *C. R. Rao*: Information and accuracy attainable in the estimation of statistical parameters. Bull. Calcuta Math. Soc. *37* (1945), 81–91.
[17] *C. R. Rao*: Differential Metrics in probability spaces. Department of Mathematics and Statistics, University of Pittsburgh, Pittsburgh, PA (1987).
[18] *A. Rényi*: On measures of entropy and information. Proc. 4th Berkeley Symp. Math. Statist. and Prob. *1* (1961), 547–561.
[19] *M. Salicrú, M. L. Menéndez, D. Morales, L. Pardo*: Asymptotic distribution of (h, Φ)-entropies. Communications in Statistics (Theory and methods) *22(7)* (1993), 2015–2031.
[20] *C. E. Shannon*: A mathematical theory of communication. Bell Syst. Tech. J. *27* (1948), 379–423.
[21] *B. D. Sharma, I. J. Taneja*: Entropy of type (α, β) and other generalized measures in information theory. Metrika *22* (1975), 205–215.
[22] *B. D. Sharma, P. Mittal*: New non-additive measures of relative information. J. Comb. Inform. and Syst. Sci. *2* (1975), 122–133.
[23] *I. J. Taneja*: A study of generalized measures in information theory. Ph. D. Thesis. University of Delhi (1975).
[24] *I. J. Taneja*: On generalized information measures and their applications. Ad. Electronics and Electron Physics *76* (1989), 327–413.
[25] *I. Vajda, K. Vašek*: Majorization, concave entropies and comparison of experiments. Probl. of Control and Inform. Theory *14* (1985), 105–115.
[26] *J. C. A. Van der Lubbe*: R-norm information and a general class of measures for certainty and information. M. Sc. Thesis, Delf University of Technology, Dept. E.E., (1977). (In Dutch.)
[27] *J. C. A. Van der Lubbe*: A generalized probabilistic theory of the measurement of certainty and information. Ph. D. Thesis, Delf University of Technology, Dept. E.E., (1981).
[28] *R. S. Varma*: Generalizations of Renyi's entropy of order α. J. Math. Sci. *1* (1966), 34–48.

Authors' addresses: *M. L. Menéndez*, Departamento de Matemática Aplicada, E.T.S de Arquitectura, Universidad Politécnica de Madrid, 28040 Madrid, Spain; *D. Morales*, *L. Pardo*, Departamento de Estadística e I.O., Facultad de Matemáticas, Universidad Complutense de Madrid, 28040 Madrid, Spain; *M. Salicrú*, Departamento de Estadística, Avd. Diagonal 645, Universidad de Barcelona, 08028 Barcelona, Spain.