Pergamon

Appl. Math. Lett. Vol. 11, No. 4, pp. 99–104, 1998

Printed in Great Britain
0893-9659/98 $19.00 + 0.00

PII: S0893-9659(98)00064-0

Estimators Based on Sample Quantiles using (h, ϕ)-Entropy Measures

M. L. MENÉNDEZ
Department of Applied Mathematics
Technical University of Madrid
Madrid, Spain

J. A. PARDO AND M. C. PARDO
Department of Statistics and O.R.
Complutense University of Madrid
Madrid, Spain

(Received and accepted June 1997)

Communicated by R. Ahlswede

Abstract—A point estimation procedure based on the maximum entropy principle for (h, ϕ) entropies is proposed using sample quantiles. These estimators are efficient and asymptotically normal under standard regularity conditions. A test for goodness-of-fit is constructed, being the corresponding statistic asymptotically distributed chi-squared. These results generalize the results obtained in [1]. © 1998 Elsevier Science Ltd. All rights reserved.

Keywords—Maximum entropy principle, (h, ϕ)-entropy, Point estimator, Goodness-of-fit.

1. INTRODUCTION

Let $X_1, \dots, X_n$ be a random sample and we observe only the order statistics

$$X_{([n/M]+1)}, X_{([2n/M]+1)}, \dots, X_{([(M-1)/Mn]+1)},$$

where $[x]$ is the greatest integer in x. We form M cells having boundaries

$$-\infty = c_{0,n} < c_{1,n} < \dots < c_{M-1,n} < c_{M,n} = \infty,$$

where $c_{i,n} = X_{([\text{in}/M]+1)}$, $1 < i < M-1$. The observed frequency N_{in} in the i^{th} cell $(c_{i-1,n}, c_{i,n}]$ is nonrandom, $N_{\text{in}} = [\text{in}/M] - [(i-1)n/M]$. The vectors of sample and population quantiles of orders $1/M$, $2/M, \dots, (M-1)/M$, are $c_n = (c_{1,n}, \dots, c_{M-1,n})^t$ and $c = (c_1, \dots, c_{M-1})^t$, respectively. The i^{th} "estimated cell probability" is therefore, $p_i(c_n; \theta) = F(c_{\text{in}}; \theta) - F(c_{(i-1)n}; \theta)$, where $\theta = (\theta_1, \dots, \theta_s)^t$. These are random, unlike the cell frequencies. For $\theta = \theta_0$ (true value of the parameter), we write $p_i = p_i(c; \theta) = i/M - (i-1)/M = 1/M$, $i = 1, \dots, M$. In this context, in [1] there is given an estimation procedure on the basis of Shannon's entropy. In that paper, it

This work was supported by Grants DGICYT PB94-0308 and PB93-0022.

Typeset by $\mathcal{A}_\mathcal{M}\mathcal{S}$-TEX can be seen also, an interesting justification of it on the basis of the maximum entropy principle. This estimator is defined as

$$\bar{\theta} = \arg\max_{\theta\in\Theta} H(\theta),$$

being

$$H(\theta) = -\sum_{i=1}^{M} p_i(c_n;\theta)\log p_i(c_n;\theta).$$

The approach that takes some but not all the sample quantiles, has been proposed by Bofinger [2] for problems of testing goodness-of-fit with χ^2-statistic. In [1], this idea is adapted to problems of point estimation. In [3], there are many references in which only a relatively small number of order statistics are used.

In the literature of Statistical Information Theory, there are many important entropy measures different from Shannon's entropy. A systematic attempt to develop a generalization of Shannon's entropy was carried out by Rényi [4], who characterized an entropy of order r. For operational purposes, it seems more natural to consider the slar expression $\sum_{i=1}^{n} p_i^s$ as an entropy measure instead of Rényi's entropy of order r. So, Havrda and Charvat [5] proposed the entropy of order s. Arimoto [6], the entropy of kind t; Sharma and Mittal [7] introduced and characterized the entropy of order 1 and degree s, as well as the entropy of order r and degree s. Later, other authors have introduced more entropy measures.

Salicrú *et al.* [8] presented an unification of entropy measures by means of (h,ϕ)-entropy measures. In our context, the (h,ϕ)-entropy associated to $p(c_n,\theta) = (p_1(c_n;\theta),\dots,p_M(c_n;\theta))^t$ is given by

$$H_\phi^h(p(c_n,\theta)) = h\left(\sum_{i=1}^{M}\phi(p_i(c_n,\theta))\right), \tag{1}$$

where either $\phi:[0,\infty)\longrightarrow R$ is concave and $h:R\longrightarrow R$ is increasing, or $\phi:[0,\infty)\longrightarrow R$ is convex and $h:R\longrightarrow R$ is decreasing. In the rest of the paper, we shall suppose that ϕ is concave and h is increasing. In the following table, we present some examples of the (h,ϕ)-entropy measures.

$\phi(x)$	$h(x)$	(h,ϕ)-entropies
$-x\log x$	x	Shannon
x^r	$(1-r)^{-1}\log x$	Rényi (1961)
x^{r-m+1}	$(m-r)^{-1}\log x$	Varma (1966)
$x^{r/m}$	$m(m-r)\log x$	Varma (1966)
$(1-s)^{-1}(x^s-x)$	x	Havrda and Charvat (1967)
$x^{1/t}$	$\left(2^{t-1}-1\right)^{-1}\left(x^t-1\right)$	Arimoto (1971)
$x\log x$	$(1-s)^{-1}\left(\exp_2((s-1)x)-1\right)$	Sharma and Mittal (1975)
x^r	$(1-s)^{-1}\left(x^{(s-1)/(r-1)-1}\right)$	Sharma and Mittal (1975)
$x^r\log x$	$-2^{r-1}x$	Taneja (1975)
x^r-x^s	$\left(2^{1-r}-2^{1-s}\right)^{-1}x$	Sharma and Taneja (1975)
$(1+\lambda x)\log(1+\lambda x)$	$\left(1+\frac{1}{\lambda}\right)\log(1+\lambda)-\frac{x}{\lambda}$	Ferreri (1980)

In [8], the asymptotic distribution of the (h,ϕ)-entropy measures is studied in multinomial populations. In [9], the problem is studied in general populations in which the parameter is estimated by a consistent and asymptotically normal estimator of the unknown parameter. In [10], 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 are studied.

In this paper, we consider formula (1) as a simple way to summarize previously defined entropy in such a way that when we talk about (h,ϕ)-entropy measures is because we have in mind an already existing and studied entropy measure. The final purpose is to save time and work. In Section 2, we present the maximum (h,ϕ)-entropy estimator as well as its properties. Finally, in Section 3, we see that the maximum (h,ϕ)-entropy estimator can be used for testing if data comes from a given parametric model on the basis of sample quantiles.

2. MAXIMUM (h,ϕ)-ENTROPY ESTIMATOR

We will begin defining the estimator obtained when we maximize the (h,ϕ)-entropy. This estimator is given in the following definition.

DEFINITION 2.1. *The estimation procedure is to maximize on $\theta \in \Theta$ the (h,ϕ)-entropy*

$$H_\phi^h\left(p(c_n,\theta)\right) = h\left(\sum_{i=1}^{M} \phi\left(p_i(c_n,\theta)\right)\right),$$

and the maximum (h,ϕ)-entropy estimator is

$$\overline{\theta} = \arg\max_{\theta\in\Theta} H_\phi^h\left(p(c_n,\theta)\right).$$

Now we will establish an important result that we will use in the following section.

PROPOSITION 2.1. *Let $\widehat{\theta}$ be an estimator of θ_0 such that $\widehat{\theta} = \theta_0 + O_P(n^{-1/2})$, if $F(x,\theta)$ has a density function f which is continuous in (x,θ) in a neighborhood of (c_i,θ_0), $i = 1,\dots,M$, and $\frac{\partial F(x,\theta)}{\partial\theta_j}$ exists and is continuous in (x,θ) in those neighborhoods, then*

$$H_\phi^h\left(p\left(c_n,\overline{\theta}\right)\right) = h\left(M\phi\left(\frac{1}{M}\right)\right) + o_P\left(n^{-1/2}\right),$$

where $\overline{\theta}$ is the maximum (h,ϕ)-entropy estimator.

PROOF. Taking $a_i = 1/M$ and $x_i = p_i(c_n,\widehat{\theta})$ in the formula

$$h\left(\sum_{i=1}^{M}\phi(x_i)\right) = h\left(\sum_{i=1}^{M}\phi(a_i)\right) + \sum_{i=1}^{M} h'\left(\sum_{i=1}^{M}\phi(a_i)\right)\phi'(a_i)(x_i - a_i) + o\left\|x-a\right\|_2,$$

and taking in account that $p_i(c_n,\widehat{\theta}) = 1/M + O_P(n^{-1/2})$ (see [1]), we get

$$H_\phi^h\left(p\left(c_n,\widehat{\theta}\right)\right) = h\left(M\phi\left(\frac{1}{M}\right)\right) + o_P\left(n^{-1/2}\right).$$

From the inequalities

$$h\left(M\phi\left(\frac{1}{M}\right)\right) \ge h\left(\sum_{i=1}^{M}\phi\left(p_i\left(c_n,\overline{\theta}\right)\right)\right) \ge h\left(\sum_{i=1}^{M}\phi\left(p_i\left(c_n,\widehat{\theta}\right)\right)\right),$$

because ϕ is concave and h is increasing, it follows that

$$h\left(\sum_{i=1}^{M}\phi\left(p_i\left(c_n,\overline{\theta}\right)\right)\right) = h\left(M\phi\left(\frac{1}{M}\right)\right) + o_P\left(n^{-1/2}\right).$$

Now, we shall establish the properties of the estimator $\overline{\theta}$ and we suppose the following assumptions hold.

(i) $F(c_i,\theta_1) \ne F(c_i,\theta_2)$ for at least one i when $\theta_1 \ne \theta_2$.

(ii) The derivatives $\frac{\partial F(x;\theta)}{\partial \theta_j}$ are continuous with respect to (x, θ) in a neighborhood of (c_i, θ_0) for $i = 1, \ldots, M$.

(iii) $F(x, \theta)$ has density function which is continuous in a neighborhood of (c_i, θ_0) for $i = 1, \ldots, M$.

(iv) The Fisher information matrix $I(\theta) = A(c, \theta)^t A(c, \theta)$, where $A(c, \theta) = (p_i(c, \theta)^{-1/2} \frac{\partial p_i(c,\theta)}{\partial \theta_j})_{\substack{i=1,\ldots,M \\ j=1,\ldots,s}}$ is nonsingular at θ_0.

Under assumptions (i)–(iv), the estimator $\widehat{\theta}$, which maximizes $\log L(\theta) = \sum_{i=1}^n N_{\text{in}} \log p_i(c, \theta)$ is asymptotically efficient, i.e.,

$$\widehat{\theta} = \theta + n^{-1} B_0 \left[\frac{\partial \log L(\theta)}{\partial \theta} \right]_{\theta = \theta_0} + o_P\left(n^{-1/2}\right),$$

where B_0 is a matrix of constants which may depend on θ_0.

First, let us see that $\overline{\theta}$ converges in probability to the true value of the parameter. According with Taylor's formula, we have

$$h\left(\sum_{i=1}^M \phi\left(p_i\left(c_n, \overline{\theta}\right)\right)\right) = h\left(M\phi\left(\frac{1}{M}\right)\right) + \frac{1}{2} h'\left(M\phi\left(\frac{1}{M}\right)\right) \phi''\left(\frac{1}{M}\right) \sum_{i=1}^M \left(p_i\left(c_n, \overline{\theta}\right) - \frac{1}{M}\right)^2 + o_P\left(n^{-1}\right),$$

and taking in account Proposition 2.1, we obtain that

$$\frac{1}{2} h'\left(M\phi\left(\frac{1}{M}\right)\right) \phi''\left(\frac{1}{M}\right) \sum\_i = 1^M \left(p_i\left(c_n, \overline{\theta}\right) - \frac{1}{M}\right)^2 = o_P\left(n^{-1/2}\right).$$

As a consequence, we have

$$\sum_{i=1}^M \left(p_i\left(c_n, \overline{\theta}\right) - \frac{1}{M}\right)^2 = o_P\left(n^{-1/2}\right)$$

and

$$p_i\left(c_n, \overline{\theta}\right) = \frac{1}{M} + o_P(1) = p_i(c, \theta_0) + o_P(1).$$

By the continuity of $F(x, \theta_0)$ in a neighborhood of (c_i, θ_0) and taking in account the fact that $c_{i,n}$ converges in probability to c_i, $i = 1, \ldots, n$, and the assumption (i), we conclude that $\overline{\theta} \xrightarrow{P} \theta_0$.

Let us see that $\overline{\theta}$ is asymptotically efficient. Since θ_0 is an interior point of Θ, $\overline{\theta}$ must be a solution of the s equations $\frac{\partial H_\phi^h(p(c_n,\theta))}{\partial \theta_r} = 0$, $r = 1, \ldots, s$.

Calculate these derivatives:

$$\frac{\partial H_\phi^h\left(p\left(c_n, \overline{\theta}\right)\right)}{\partial \theta_r} = h'\left(\sum_{i=1}^M \phi\left(p_i\left(c_n, \overline{\theta}\right)\right)\right) \sum_{i=1}^M \phi'\left(p_i\left(c_n, \overline{\theta}\right)\right) \frac{\partial p_i\left(c_n, \overline{\theta}\right)}{\partial \theta_r}.$$

Using a Taylor's expansion of $\phi'(p_i(c_n, \overline{\theta}))$ around θ_0, we have

$$\phi'\left(p_i\left(c_n, \overline{\theta}\right)\right) = \phi'\left(p_i\left(c_n, \theta_0\right)\right) + \sum_{u=1}^s \left(\overline{\theta}_u - \theta_{u,0}\right) \phi''\left(p_i\left(c_n, \theta^*\right)\right) \frac{\partial p_i\left(c_n, \theta^*\right)}{\partial \theta_u},$$

where $\|\theta^* - \theta_0\|_2 < \|\overline{\theta} - \theta_0\|_2$, then

$$\frac{\partial H_\phi^h\left(p\left(c_n, \overline{\theta}\right)\right)}{\partial \theta_r} = h'\left(M\phi\left(\frac{1}{M}\right)\right) \sum_{i=1}^M \phi'\left(p_i(c_n, \theta_0)\right) \frac{\partial p_i\left(c_n, \overline{\theta}\right)}{\partial \theta_r} + h'\left(M\phi\left(\frac{1}{M}\right)\right) J_r^*\left(\overline{\theta} - \theta_0\right) + o_P\left(n^{-1/2}\right),$$

where

$$J_r^* = \left(\sum_{i=1}^{M} \phi'' \left(p_i \left(c_n, \theta^* \right) \right) \frac{\partial p_i \left(c_n, \theta^* \right)}{\partial \theta_u} \frac{\partial p_i \left(c_n, \overline{\theta} \right)}{\partial \theta_r} \right)_{u=1,\ldots,s}.$$

We can observe that the matrix $J^* = (J_1^*, \ldots, J_s^*)^t$ verifies

$$J^* \xrightarrow{P} \frac{1}{M} \phi'' \left(\frac{1}{M} \right) I(\theta_0).$$

As

$$\frac{\partial p_i \left(c_n, \overline{\theta} \right)}{\partial \theta_r} = \frac{\partial p_i(c, \theta_0)}{\partial \theta_r} + o_P(1),$$

we get

$$\begin{aligned} \sum_{i=1}^{M} \phi' \left(p_i(c_n, \theta_o) \right) \frac{\partial p_i \left(c_n, \overline{\theta} \right)}{\partial \theta_r} &= \sum_{i=1}^{M} \left(\phi' \left(\frac{1}{M} \right) \right. \\ &\quad \left. + \phi'' \left(\frac{1}{M} \right) \left(p_i \left(c_n, \theta_0 \right) - \frac{1}{M} \right) o_P \left(n^{-1/2} \right) \right) \frac{\partial p_i \left(c_n, \overline{\theta} \right)}{\partial \theta_r} \\ &= \frac{1}{M} \phi'' \left(\frac{1}{M} \right) \left(-\frac{1}{n} \left[\frac{\partial \log L(\theta)}{\partial \theta_r} \right]_{\theta=\theta_0} \right) + o_P \left(n^{-1/2} \right). \end{aligned}$$

In the previous calculations, we have applied [2, Theorem 1]. If we return to the equations $\frac{\partial H_\phi^h(p(c_n,\theta))}{\partial \theta_r} = 0$, $r = 1, \ldots, s$, we have

$$0 = \frac{1}{M} \phi'' \left(\frac{1}{M} \right) \left(-\frac{1}{n} \left[\frac{\partial \log L(\theta)}{\partial \theta} \right]_{\theta=\theta_0} \right) + \frac{1}{M} \phi'' \left(\frac{1}{M} \right) I(\theta_0) \left(\overline{\theta} - \theta_0 \right) + o_P \left(n^{-1/2} \right)$$

and

$$\overline{\theta} - \theta_0 = I(\theta_0)^{-1} n^{-1} \left[\frac{\partial \log L(\theta)}{\partial \theta} \right]_{\theta=\theta_0} + o_P \left(n^{-1/2} \right),$$

i.e., $\overline{\theta}$ is asymptotically efficient.

Further,

$$n^{1/2} \left(\overline{\theta} - \theta_0 \right) \xrightarrow[n\to\infty]{L} N \left(0, I(\theta_0)^{-1} \right),$$

because

$$n^{-1} \left[\frac{\partial \log L(\theta)}{\partial \theta} \right]_{\theta=\theta_0} \xrightarrow{L} N \left(0, I \left(\theta_0 \right) \right).$$

Therefore, we get the following results.

THEOREM 2.1. *Under assumptions (i)–(iv), the estimator $\overline{\theta}$ which maximizes (1) verifies:*

(a) *$\overline{\theta}$ converges in probability to θ_0;*
(b) *$\overline{\theta}$ is asymptotically efficient;*
(c) $n^{1/2}(\overline{\theta} - \theta_0) \xrightarrow[n\to\infty]{L} N(0, I(\theta_0)^{-1})$.

3. TESTING FOR GOODNESS-OF-FIT

The results given in the preceding section can be used to test if data comes from a given parametric model.

THEOREM 3.1. *Under the conditions of Theorem 2.1,*

$$T = \frac{2nM\left(H_\phi^h\left(p\left(c_n,\overline{\theta}\right)\right) - h\left(M\phi\left(1/M\right)\right)\right)}{h'\left(M\phi\left(1/M\right)\right)\phi''\left(1/M\right)} \xrightarrow{L} \chi^2_{M-s-1}.$$

PROOF. It is clear that

$$H_\phi^h\left(p\left(c_n,\overline{\theta}\right)\right) = h\left(M\phi\left(\frac{1}{M}\right)\right) + \frac{1}{2}h'\left(M\phi\left(\frac{1}{M}\right)\right)\phi''\left(\frac{1}{M}\right)\sum_{i=1}^{M}\left(p_i\left(c_n,\overline{\theta}\right) - \frac{1}{M}\right)^2 + o_P\left(n^{-1}\right),$$

then

$$T = nM\sum_{i=1}^{M}\left(p_i\left(c_n,\overline{\theta}\right) - \frac{1}{M}\right)^2 + o_P\left(n^{-1}\right),$$

i.e., the statistic T and $nM\sum_{i=1}^{M}(p_i(c_n,\overline{\theta}) - 1/M)^2$ have asymptotically the same distribution. But applying Theorem 3 given in [2], we have that

$$nM\sum_{i=1}^{M}\left(p_i\left(c_n,\overline{\theta}\right) - \frac{1}{M}\right)^2 \xrightarrow{L} \chi^2_{M-s-1}.$$

REFERENCES

1. M.L. Menéndez, D. Morales and L. Pardo, Maximum entropy principle and statistical inference on condensed ordered data, *Statistics and Probability Letters* (to appear).
2. E. Bofinger, Goodness of fit using sample quantiles, *Journal of the Royal Statistical Society, Series B* **35**, 277–284, (1973).
3. T.R.C. Read and N.A.C. Cressie, *Goodness of Fit Statistics for Discrete Multivariate Data*, Springer-Verlag, (1988).
4. A. Rényi, On measures of entropy and information, In *Proc. 4^{th} Berkeley Symp. Math. Statist. and Prob.*, Volume 1, pp. 547–561, (1961).
5. J. Havrda and F. Charvat, Quantification of classification processes. Concept of structural α-entropy, *Kybernetika* **2**, 30–55, (1967).
6. S. Arimoto, Information theoretical considerations on estimation problems, *Information and Control* **19**, 181–194, (1971).
7. B.D. Sharma and D.P. Mittal, New nonadditive measures of entropy for discrete probability distributions, *J. Math. Sci.* **10**, 28–40, (1977).
8. M. Salicrú, M.L. Menéndez, D. Morales and L. Pardo, Asymptotic distribution of (h,ϕ) entropies, *Communications in Statitstics; Theory and Methods* **22** (7), 2015–2031, (1993).
9. L. Pardo, D. Morales, M. Salicrú and M.L. Menéndez, Large sample behavior of entropy measures when parameters are estimated, *Communications in Statitstics; Theory and Methods* **26** (2), 483–501, (1997).
10. M.L. Menéndez, D. Morales, L. Pardo and M. Salicrú, (h,ϕ)-entropy differential metric, *Applications of Mathematics* **42** (2), 81–98, (1997).