María Del Carmen Pardo; Julio A. Pardo

Statistical applications of order α-β weighted information energy

Applications of Mathematics, Vol. 40 (1995), No. 4, 305–317

Persistent URL: http://dml.cz/dmlcz/134296

STATISTICAL APPLICATIONS OF ORDER α-β WEIGHTED INFORMATION ENERGY

M. C. PARDO, J. A. PARDO, Madrid*

(Received December 28, 1993)

Summary. A statistic using the concept of order α-β weighted information energy introduced by Tuteja et al. (1992) is considered and its asymptotic distribution in a stratified random sampling is obtained. Some special cases are also discussed.

Keywords: order α-β weighted information energy, asymptotic distribution, testing of hypotheses

AMS classification: 62B10, 62E20

1. INTRODUCTION

Onicescu (1966) introduced the concept of *information energy* in *information theory*. This measure, for a discrete random variable having a finite number of values $x_1, \ldots, x_M$ with probabilities $p_1, \ldots, p_M$, respectively, is given by

$$\mathfrak{E}(P) = \sum_{i=1}^{M} p_i^2 \tag{1}$$

where $P = (p_1, \ldots, p_M)$. Some interesting applications and properties of this expression can be found in Pardo (1981, 1983, 1987), Pardo et al. (1985, 1988, 1989), Pérez (1966) and Theodorescu (1977), Vajda (1967) and Theodorescu (1977) present an axiomatic treatment of the expression (1).

* The research in this paper was supported in part by DGICYT Grants No. PB91-0387 and No. PB91-0155. Their financial support is gratefully acknowledged.

In order to distinguish the elements $x_1, \ldots, x_M$ according to their importance with respect to a given qualitative characteristic of the system, we associate a positive number u_i with each outcome x_i. In this context Theodorescu (1977) presented a generalization of Onicescu's information energy given by

$$(2) \qquad \mathfrak{E}(P,U) = \frac{\sum_{i=1}^{M} u_i p_i^2}{\sum_{i=1}^{M} p_i u_i} \quad \text{where } U = (u_1, \ldots, u_M).$$

Pardo (1981) defined the *useful informational energy*, Pardo (1985) gave an axiomatic characterization and Pardo et al. (1994) obtained the asymptotic distribution of the analogue estimator of this measure, in a random and stratified sampling.

Aggarwal and Picard (1978) and Sharman et al. (1978) introduced and characterized a generalized measure of useful information, called *useful information of degree β*, given by

$$(3) \qquad H_\beta(P) = \frac{\sum_{i=1}^{M} u_i p_i (1 - p_i^{\beta-1})}{1 - 2^{1-\beta}}, \quad \beta > 0 \text{ and } \beta \neq 1.$$

In this line, Singh (1983) introduced the measure

$$(4) \qquad I_{(\alpha,\beta)}(P,U) = \frac{\sum_{i=1}^{M} u_i (p_i^\alpha - p_i^\alpha)}{2^{1-\alpha} - 2^{1-\beta}}, \quad \alpha, \beta > 0,\ \alpha \neq 1,\ \beta \neq 1 \text{ and } \alpha \neq \beta$$

that includes interesting particular and limiting cases. For example, when $\alpha = 1$ and $\beta \neq 1$, (4) reduces to (3); when $u_i = 1$ for all $i = 1, \ldots, M$, (4) reduces to entropy of type (α, β) (Sharma and Taneja, 1975); when $\alpha = 1$ and $u_i = 1$ for all $i = 1, \ldots, M$, (4) reduces to entropy of degree β (Havrda and Charvát, 1967); when $\alpha = 1$ and $\beta \to 1$, (4) reduces to weighted entropy (Belis and Guiasu, 1968) and when $\alpha = 1$, $\beta \to 1$ and $u_i = 1$ (4) reduces to Shannon's entropy (Shannon, 1948). Also, Singh (1983) gave a characterization of the expression (4).

In this paper, we consider a new concept of weighted information energy that depends upon two parameters, α and β, introduced and characterized by Tuteja et al. (1992). This measure, called *order α-β weighted information energy*, is defined as

$$(5) \qquad \mathfrak{E}_{(\alpha,\beta)}(P,U) = \frac{\sum_{i=1}^{M} u_i (p_i^\alpha - p_i^\beta)}{(\alpha - 1) \sum_{i=1}^{M} p_i u_i}, \quad \alpha, \beta > 0,\ \alpha \neq 1,\ \beta \neq 1 \text{ and } \alpha \neq \beta.$$

When $\beta \to \infty$ the measure (5) reduces to the order *α-weighted information energy* given by Pardo (1986) and when $\beta = 1$ and $\alpha \to 1$ this expression reduces to the useful Shannon entropy introduced by Picard (1972, 1979). Its asymptotic behaviour analyzed by Pardo (1993).

We obtain the asymptotic distribution of the analogue estimate of the expression (5) in a stratified random sampling. The knowledge of this asymptotic distribution allows us to construct different tests of hypotheses. Some special cases are also discussed.

Other contributions to measures of useful information have also been made by Gurdial and Pesson (1973), Hooda (1984), Kannappan (1980), Mohan and Mitter (1978) and Sharma and Shing (1983).

2. Asymptotic distribution of the order α-β weighted information energy

Consider a population with N individuals which can be classified into M classes or categories, $x_1, \ldots, x_M$ according to a certain process X, and let

$$\Delta_M = \left\{ P = (p_i)_{i=1,\ldots,M} \middle| \sum_{i=1}^{M} p_i - 1, p_i \geqslant 0, i = 1, \ldots, M \right\}$$

be the set of all probability distributions over $\chi = \{x_1, \ldots, x_M\}$. Now we suppose that the population with N individuals can be divided into r non-overlapping sub-populations, called strata, as homogeneous as possible with respect to X. Let N_k be the number of individuals in the kth stratum, p_{ik} the probability that a randomly selected member belongs to the kth stratum and to the class x_i, $p_{i.}$ the probability that a randomly selected member in the whole population belongs to the class x_i, and $p_{.k}$ the probability that it belongs into the kth stratum. Then one obtains

$$\sum_{k=1}^{r} N_k = N, \quad \sum_{i=1}^{M} \sum_{k=1}^{r} p_{ik} = 1,$$
$$p_{i.} = \sum_{k=1}^{r} p_{ik} \quad p_{.k} = \sum_{i=1}^{M} p_{ik}$$

and we denote by W_k the relative size of the kth stratum, i.e., $W_k = N_k/N = p_{.k}$. Finally, let u_i be the utility of the class x_i.

In order to obtain an estimate for the order α-β weight information energy in the population, we shall draw at random a stratified sample of size n, independently of the other strata. Assume that the sample is chosen by a specified allocation w_k, $k = 1, \ldots, r$, so that a sample of size n_k is drawn independently at random with replacement from the kth stratum, where $w_k = n_k/n$. For example, if w_k is constant we get a constant allocation, if $w_k = N_k/N$ we get a proportional allocation and if $w_k = \sigma_k/c_k$ we get an optimum allocation where σ_k is the variance and c_k is the cost per unit of sampling in the kth stratum. If $\hat{p}_{ik}$ denotes the relative frequency, in the size n sample, of individuals belonging to the class x_i in the kth stratum, and we define

$$\hat{p}_{i.} = \sum_{k=1}^{r} \frac{W_k}{w_k} \hat{p}_{ik},$$

then $\mathfrak{E}_{(\alpha,\beta)}(P,U)$ can be estimated by

$$\mathfrak{E}_{(\alpha,\beta)}(\hat{P},U) = \frac{\sum_{i=1}^{M} u_i(\hat{p}_{i.}^{\alpha} - \hat{p}_{i.}^{\beta})}{(\alpha-1)\sum_{i=1}^{M} \hat{p}_{i.} u_i}, \quad \alpha, \beta > 0,\ \alpha \neq 1,\ \beta \neq 1 \text{ and } \alpha \neq \beta$$

where $\hat{P} = (\hat{p}_{1.}, \ldots, \hat{p}_{M.})$.

The following theorem establishes the asymptotic behavior of $\mathfrak{E}_{(\alpha,\beta)}(\hat{P},U)$ is a stratified random sampling.

Theorem 1. *Consider the estimate $\mathfrak{E}_{(\alpha,\beta)}(\hat{P},U)$, obtained by replacing p_{ij}, $p_{i.}$ and $p_{.j}$ by $\hat{p}_{ij}$, $\hat{p}_{i.}$ and $\hat{p}_{.j}$, $(i = 1, \ldots, M; j = 1, \ldots, r)$ in a stratified random sample of size n and allocation $(w_1, \ldots, w_r)$. Then we have*

$$n^{1/2}\left(\mathfrak{E}_{(\alpha,\beta)}(\hat{P},U) - \mathfrak{E}_{(\alpha,\beta)}(P,U)\right) \xrightarrow[n\to\infty]{L} N(0, {}^{st}\sigma^2)$$

where

$${}^{st}\sigma^2 = \sum_{k=1}^{r} \frac{W_k}{w_k} \sum_{i=1}^{M} p_{ik} t_{i.}^2 = \sum_{k=1}^{r} \frac{1}{w_k} \Big(\sum_{i=1}^{M} p_{ik} t_{i.}\Big)^2$$

and

$$t_{i.} = \frac{u_i(\alpha p_{i.}^{\alpha-1} - \beta p_{i.}^{\beta-1}) \sum_{i=1}^{M} p_{i.} u_i - u_i \sum_{i=1}^{M} u_i(p_{i.}^{\alpha} - p_{i.}^{\beta})}{(\alpha-1)\Big(\sum_{i=1}^{M} p_{i.} u_i\Big)^2}$$

whenever ${}^{st}\sigma^2 > 0$.

P r o o f. Consider the Taylor expansion of $\mathfrak{E}_{(\alpha,\beta)}(\hat{P},U)$ around the point $P = (p_{i.}, i = 1, \ldots, M)$, which is given

$$\mathfrak{E}_{(\alpha,\beta)}(\hat{P},U) = \mathfrak{E}_{(\alpha,\beta)}(P,U) + \sum_{i=1}^{M} t_{i.}(\hat{p}_{i.} - p_{i.}) + R_n^{(1)}$$

where $R_n^{(1)}$ is the Lagrange remainder and

$$t_{i.} = \frac{\partial \mathfrak{E}_{(\alpha,\beta)}(P,U)}{\partial p_{i.}} = \frac{u_i(\alpha p_{i.}^{\alpha-1} - \beta p_{i.}^{\beta-1}) \sum_{i=1}^{M} p_{i.} u_i - u_i \sum_{i=1}^{M} u_i (p_{i.}^{\alpha} - p_{i.}^{\beta})}{(\alpha - 1)\left(\sum_{i=1}^{M} p_{i.} u_i\right)^2}.$$

Therefore, we obtain that the random variables

$$n^{1/2}\left(\mathfrak{E}_{(\alpha,\beta)}(\hat{P},U) - \mathfrak{E}_{(\alpha,\beta)}(P,U)\right) \quad \text{and} \quad n^{1/2}\left(\sum_{i=1}^{M} t_{i.}(\hat{p}_{i.} - p_{i.})\right)$$

have asymptotically the same distribution because $n^{1/2} R_n^{(1)}$ converges in probability to zero.

Finally, applying the Central Limit Theorem in each stratum, we have

$$n^{1/2}(\hat{p}_{1.} - p_{1.}, \ldots, \hat{p}_{M.} - p_{M.}) \xrightarrow[n\to\infty]{L} N\left(0, \sum_{k=1}^{r} \frac{W_k^2}{w_k} \Sigma(k)\right)$$

with

$$\Sigma(k) = \left(\frac{p_{ik}}{W_k}\left(\delta_{ij} - \frac{p_{jk}}{W_k}\right)\right)_{\substack{i=1,\ldots,M \\ j=1,\ldots,M}}, \quad k = 1, \ldots, r.$$

Therefore the result required follows. □

Remark 1.

1) If $\beta \to \infty$ and $\alpha = 2$ is immediate that

$$t_{i.} = \frac{2u_i p_{i.} \sum_{i=1}^{M} p_{i.} u_i - u_i \sum_{i=1}^{M} u_i p_{i.}^2}{\left(\sum_{i=1}^{M} p_{i.} u_i\right)^2},$$

i.e., we have the result obtained by Pardo et al. (1993).

2) If $r = 1$ and we denote $p_i = p_{i.}$, $i = 1, \ldots, M$, we have that

$$n^{1/2}\left(\mathfrak{E}_{(\alpha,\beta)}(\hat{P},U) - \mathfrak{E}_{(\alpha,\beta)}(P,U)\right) \xrightarrow[n\to\infty]{L} N(0, \sigma^2)$$

with

$$\sigma^2 = \sum_{i=1}^{M} p_i t_{i.}^2 - \left(\sum_{i=1}^{M} p_i t_{i.}\right)^2$$

where $t_{i.}$ is given in the previous theorem. In this case we obtain the result for simple random sampling.

3) Following the ideas in Gil (1989, 1992) we get that the optimum allocation is given by

$$w_k = \alpha_k^{1/2} \left(\sum_{k=1}^{r} \alpha_k^{1/2} \right)^{-1}, \quad (k = 1, \ldots, r)$$

where

$$\alpha_k = \sum_{i=1}^{M} W_k p_{ik} t_{i.}^2 - \left(\sum_{i=1}^{M} p_{ik} t_{i.} \right)^2, \quad (k = 1, \ldots, r).$$

4) If we consider a random variable taking on the values

$$\alpha_k^{1/2} W_k^{-1}, \quad k = 1, \ldots, r$$

with probabilities W_k, respectively, applying Jensen's inequality to the function $\varphi(x) = x^2$ we obtain

$${}^{st}\sigma_{\text{opt}}^2 = \left(\sum_{k=1}^{r} \alpha_k^{1/2} \right)^2 \leqslant \sum_{k=1}^{r} \frac{\alpha_k}{W_k} = {}^{st}\sigma_{\text{prop}}^2$$

where ${}^{st}\sigma_{\text{prop}}^2$ denotes the asymptotic variance in the stratified random sampling with proportional allocation and the equality holds if and only if $r = 1$ or $\alpha_k^{1/2} W_k^{-1}$ does not depend on k $(k = 1, \ldots, r)$.

5) If we consider a random variable taking on the values

$$\sum_{i=1}^{M} \frac{1}{W_k} p_{ik} t_{i.}, \quad k = 1, \ldots, r$$

with probabilities W_k, respectively, applying Jensen's inequality to the function $\varphi(x) = x^2$ we obtain

$${}^{st}\sigma_{\text{prop}}^2 \leqslant \sigma^2$$

and the equality holds if and only if $r = 1$ or

$$\sum_{i=1}^{M} \frac{1}{W_k} p_{ik} t_{i.}$$

does not depend on k $(k = 1, \ldots, r)$.

In general the stratification may produce a gain in precision in the estimates of characteristics of the whole population because it provides a method of utilizing supplementary information. Auxiliary information may be used to divide the population in strata. In points 3 and 4 of this remark a comparison when we try to estimate the order α-β weighted information energy by means of a large sample is made between simple random sampling and stratified random sampling with proportional and optimum allocation. This comparison shows how the gain due to stratification is achieved.

Now, if it is verified that the first derivative order term is zero and so does ${}^{st}\sigma^2 = 0$, we must use Taylor's expansion of $\mathfrak{E}_{(\alpha,\beta)}(\hat{P}, U)$ including the second order term. In this situation we have obtained the following result.

Theorem 2. *If ${}^{st}\sigma^2 = 0$, then*

$$2n\left(\mathfrak{E}_{(\alpha,\beta)}(\hat{P}, U) - \mathfrak{E}_{(\alpha,\beta)}(P, U)\right) \xrightarrow[n\to\infty]{L} \sum_{i=1}^{M} \beta_i \chi_1^2,$$

where χ_1^2's are independent and β_i's are the eigenvalues of the matrix $A\Sigma$ where

$$A = \left(\begin{pmatrix} s_1 & & \\ & \ddots & \\ & & s_M \end{pmatrix} + \left(\left[2\sum_{i=1}^{M} u_i (p_{i.}^{\alpha} - p_{i.}^{\beta}) \right] \frac{u_i u_j}{(\alpha-1)\left(\sum_{i=1}^{M} p_{i.} u_i\right)^3} \right)_{i=1,\ldots,M} \right)$$

with

$$s_i = \frac{u_i\left(\alpha(\alpha-1)p_{i.}^{\alpha-2} - \beta(\beta-1)p_{i.}^{\beta-2}\right)}{(\alpha-1)\left(\sum_{i=1}^{M} p_{i.} u_i\right)} - \frac{2u_i^2(\alpha p_{i.}^{\alpha-1} - \beta p_{i.}^{\beta-1})}{(\alpha-1)\left(\sum_{i=1}^{M} p_{i.} u_i\right)^2}$$

nd

$$\Sigma = \sum_{k=1}^{r} \frac{W_k^2}{w_k} \Sigma(k) \quad \text{with } \Sigma(k) = \left(\frac{p_{ik}}{W_k} \left(\delta_{ij} - \frac{p_{jk}}{W_k} \right) \right)_{\substack{i=1,\ldots,M \\ j=1,\ldots,M}}.$$

Proof. By considering Taylor's expansion of the function $\mathfrak{E}_{(\alpha,\beta)}(\hat{P}, U)$ including the term corresponding to the second partial derivatives we get

$$\mathfrak{E}_{(\alpha,\beta)}(\hat{P}, U) = \mathfrak{E}_{(\alpha,\beta)}(P, U) + \frac{1}{2}(\hat{p}_{1.} - p_{1.}, \ldots, \hat{p}_{M.} - p_{M.}) A \begin{pmatrix} \hat{p}_{1.} - p_{1.} \\ \vdots \\ \hat{p}_{M.} - p_{M.} \end{pmatrix} + R_n^{(2)}$$

where A is given above and $R_n^{(2)}$ is the Lagrange remainder. Therefore, the random variables

$$2\big(\mathfrak{E}_{(\alpha,\beta)}(\hat{P},U)-\mathfrak{E}_{(\alpha,\beta)}(P,U)\big) \quad \text{and} \quad (\hat{p}_{1.}-p_{1.},\ldots,\hat{p}_{M.}-p_{M.})A\begin{pmatrix}\hat{p}_{1.}-p_{1.}\\ \vdots \\ \hat{p}_{M.}-p_{M.}\end{pmatrix}$$

converge in law to the same distribution because $R_n^{(2)}$ converges in probability to zero.

Furthermore,

$$n^{1/2}(\hat{p}_{1.}-p_{1.},\ldots,\hat{p}_{M.}-p_{M.}) \xrightarrow[n\to\infty]{L} N\Big(0,\sum_{k=1}^{r}\frac{W_k^2}{w_k}\Sigma(k)\Big),$$

hence (see Mardia et al. 1982, p. 68)

$$n(\hat{p}_{1.}-p_{1.},\ldots,\hat{p}_{M.}-p_{M.})A\begin{pmatrix}\hat{p}_{1.}-p_{1.}\\ \vdots \\ \hat{p}_{M.}-p_{M.}\end{pmatrix} \xrightarrow[n\uparrow\infty]{L} \sum_{i=1}^{M}\beta_i\chi_1^2$$

where the χ_1^2's are independent and the β_i's are the eigenvalues of the matrix $A\Sigma$ with A and Σ given above. □

Remark 2.

1) If $r=1$, ${}^{st}\sigma^2=0$ and we denote $p_i=p_{i.}$, $i=1,\ldots,M$, we have

$$2n\big(\mathfrak{E}_{(\alpha,\beta)}(\hat{P},U)-\mathfrak{E}_{(\alpha,\beta)}(P,U)\big) \xrightarrow[n\to\infty]{L} \sum_{i=1}^{M}\beta_i\chi_1^2$$

where χ_1^2's are independent and β_i's are the eigenvalues of the matrix $A\Sigma$ where

$$A=\left(\begin{pmatrix}s_1 & & \\ & \ddots & \\ & & s_M\end{pmatrix}+\left(\left[2\sum_{i=1}^{M}u_i(p_i^{\alpha}-p_i^{\beta})\right]\frac{u_iu_j}{(\alpha-1)\big(\sum_{i=1}^{M}p_iu_i\big)^3}\right)_{ij}\right)$$

with

$$s_i=\frac{u_i\big(\alpha(\alpha-1)p_i^{\alpha-2}-\beta(\beta-1)p_i^{\beta-2}\big)}{(\alpha-1)\big(\sum_{i=1}^{M}p_iu_i\big)}-\frac{2u_i^2(\alpha p_i^{\alpha-1}-\beta p_i^{\beta-1})}{(\alpha-1)\big(\sum_{i=1}^{M}p_iu_i\big)^2}$$

and

$$\Sigma=\big(p_i(\delta_i-p_i)\big)_{i=1,\ldots,M}.$$

In this case we obtain the result for simple random sampling.

2) If $\beta \to \infty$ and $\alpha = 2$ it is immediate that we have the result obtained by Pardo and Vicente (1994).

3) If $r = 1$, $\beta = 1$, $\alpha \to 1$, $u_1 = \ldots = u_M$ and $p_1 = \ldots = p_M$ it is immediate to establish that the matrix $A\Sigma$ has the eigenvalues $\beta_1 = 0$ and $\beta_2 = 1$ with multiplicity $M - 1$, hence

$$\sum_{i=1}^{M} \beta_i \chi_1^2 \stackrel{d}{\equiv} \chi_{M-1}^2.$$

3. Applications on testing hypotheses

The results obtained in the previous sections can be used in various settings to test statistical hypotheses based on one sample.

a) We can test that the order α-β weighted information energy of a population equals specified value, i.e., $H_0\colon \mathfrak{E}_{(\alpha,\beta)}(P,U) = E_0$. In this case, under H_0, we have to consider two situations according to the value of ${}^{st}\sigma^2$. If ${}^{st}\sigma^2 = 0$, then we must use the statistic

$$T_1 = 2n\big(\mathfrak{E}_{(\alpha,\beta)}(\hat{P},U) - E_0\big)$$

which is approximately distributed as a linear form in chi square variables for sufficiently large n. Then a test criterion would be to reject H_0 at a level α, when $T_1 > t_\alpha$, if

$$P\bigg(\sum_{i=1}^{M} \beta_i \chi_1^2 > t_\alpha\bigg) = \alpha$$

where the β_i's, $i = 1, \ldots, M$, are given by Theorem 2, and the last probability can be composed using the methods given by Kotz et al. (1967). Rao and Scott (1981) suggested to consider the approximate distribution of $\sum_{i=1}^{M} \beta_i \chi_1^2$ which is given by $\beta\chi_M^2$, where $\beta = \sum_{i=1}^{M} \frac{\beta_i}{M}$. In this case we can easily compute the value of β, since $\sum_{i=1}^{M} \beta_i = tr(A\Sigma)$. In this case Theorem 1 can be used to evaluate the asymptotic power of the previous test. If $H_1\colon \mathfrak{E}_{(\alpha,\beta)}(P,U) = E_1$ is the alternative hypothesis, then the asymptotic power is given by

$$\beta_n(E_1) = P_{E_1}(T_1 > t_\alpha) = 1 - \Psi\Big(\frac{t_\alpha + 2n(E_0 - E_1)}{2n^{1/2}\,{}^{st}\sigma(Q)}\Big)$$

where ${}^{st}\sigma(Q)$ is the expression of ${}^{st}\sigma$ given in Theorem 1 with $\mathfrak{E}_{(\alpha,\beta)}(Q,U) = E_1$ and $\Psi(x)$ denotes the standard normal distribution function. Also note that

$$\lim_{n\to\infty} \beta_n(E_1) = 1$$

so the test is asymptotically consistent in the sense of Fraser (1957).

If ${}^{st}\sigma^2 > 0$, we can use the statistic

$$Z_1 = \frac{n^{1/2}\left(\mathfrak{E}_{(\alpha,\beta)}(\hat{P},U) - E_0\right)}{{}^{st}\hat{\sigma}}$$

which has approximately the standard normal distribution for sufficiently large n and ${}^{st}\hat{\sigma}$ is obtained by replacing p_{ik}'s by $\hat{p}_{ik}$'s in ${}^{st}\sigma$. In this context an approximate $1-\alpha$ level confidence interval for $\mathfrak{E}_{(\alpha,\beta)}(P,U)$ is given by

$$\mathfrak{E}_{(\alpha,\beta)}(\hat{P},U) \pm \frac{z_{\alpha/2}{}^{st}\hat{\sigma}}{n^{1/2}}$$

where z_α is a real number such that $P(X > z_\alpha) = \alpha$ when X is normally distributed with mean zero and variance one.

b) We can test that the order α-β weighted information energy of s independent populations equals a specified value, i.e., $H_0\colon \mathfrak{E}_{(\alpha,\beta)}(P_1,U) = \ldots = \mathfrak{E}_{(\alpha,\beta)}(P_s,U) = E_0$. In this case we can use the statistic

$$T_2 = \sum_{i=1}^{S} n_i \frac{\left(\mathfrak{E}_{(\alpha,\beta)}(\hat{P},U) - E_0\right)^2}{{}^{st}\hat{\sigma}_i^2}$$

which is asymptotically chi-square distributed with s degrees of freedom.

c) Test for equality of the order α-β weighted information energy of s independent populations, i.e., $H_0\colon \mathfrak{E}_{(\alpha,\beta)}(P_1,U) = \ldots = \mathfrak{E}_{(\alpha,\beta)}(P_s,U)$. If ${}^{st}\sigma_i > 0$ $(i = 1, \ldots, s)$ then we have a sample of size n_i from the ith population. We must consider the statistic

$$T_3 = \sum_{i=1}^{S} n_i \frac{\left(\mathfrak{E}_{(\alpha,\beta)}(\hat{P}_i,U) - E\right)^2}{{}^{st}\hat{\sigma}_i^2}$$

where

$$E = \left(\sum_{i=1}^{s} n_i \frac{\left(\mathfrak{E}_{(\alpha,\beta)}(\hat{P}_i,U)\right)}{{}^{st}\hat{\sigma}_i^2}\right)\left(\sum_{i=1}^{s} \frac{n_i}{{}^{st}\hat{\sigma}_i^2}\right)^{-1}$$

which, under H_0, has approximately a Chi-square distribution with $s-1$ degrees of freedom.

In this situation if $s = 2$, the statistic to be used is

$$Z_2 = \frac{(n_1 n_2)^{1/2}\left(\mathfrak{E}_{(\alpha,\beta)}(\hat{P}_1,U) - \mathfrak{E}_{(\alpha,\beta)}(\hat{P}_2,U)\right)}{(n_2{}^{st}\hat{\sigma}_1 + n_1{}^{st}\hat{\sigma}_2)^{1/2}}$$

which has approximately the standard normal distribution for sufficiently large n, where subscript i has been used to denote population i and n_i denotes the sample size in population i, $(i = 1, 2)$.

4. Example

The purpose of this section is to show some applications of the above results when $\alpha = 2$ and $\beta \to \infty$.

How to choose the right weights is a delicate problem in general. Vector U may depend on P and/or on some other information about the events involved. Here we consider the situations analyzed by Guiasu (1991) in Example 2, i.e., we suppose that taking a random sample of size $n = 300$ from a discrete probability distribution we obtain the relative frequencies $\hat{P}$ given in the second column of table 1 and the corresponding weights, U, given in the third column of Table 1. If we want to test the null hypothesis that $\hat{P}$ comes from the probability distribution mentioned in the last column of Table 1, the critical region test is

$$|Z_1| = \left| \frac{n^{1/2}\left(\mathfrak{E}_{(\alpha,\beta)}(\hat{P}, U) - E_0\right)}{\hat{\sigma}} \right| > z_{0.025} = 1.96$$

where $z_{0.025}$ is the value veryfing $P(|Z| > z_{0.025}) = 1.96$, provided Z is a normal random variable with mean zero and variance 1.

Table 1

	$\hat{P}$	U	P
1	0.2000	0.18	0.2097
2	0.2167	0.16	0.1751
3	0.1033	0.12	0.1234
4	0.0667	0.09	0.0869
5	0.3367	0.39	0.3538
6	0.0766	0.06	0.0511

Now, $n = 300$,

$$\hat{\sigma}^2 = 0.09587589,$$
$$E_0 = \mathfrak{E}_{(\alpha,\beta)}(P, U) = 0.2802601$$

and

$$\mathfrak{E}_{(\alpha,\beta)}(\hat{P}, U) = 0.2709551.$$

So we obtain $Z_1 = -1.681006$ and thus we can not reject the null hypothesis.

References

[1] *N.I. Aggarwal, C.F. Picard*: Functional equations and information measures with preference. Kybernetika *14* (1978), 174–181.

[2] *M. Belis, G. Guiasu*: A quantitative-qualitative measure of information in cybernetic systems. IEEE Trans. Inform. Theory *14* (1968), 593–594.

[3] *D.A.S. Fraser*: Nonparametric Methods in Statistics. John Wiley, New York, 1957.

[4] *M.A. Gil*: A note on stratification and gain in precision in estimating diversity from large samples. Communications in Statistics. Theory and Methods *18* (1989), no. 4, 1521–1526.

[5] *M.A. Gil*: On the asymptotic optimum allocation in estimating inequality from complete data. Kybernetika *28* (1992), 325–337.

[6] *S. Guiasu*: The least weighted deviation. Information Sciences *53* (1991), 271–284.

[7] *F. Gurdial, F. Pesson*: On useful information of order α. JCISS *3* (1973), 158–162.

[8] *J. Havrda, F. Charvát*: Quantification method of classification processes: Concepts of structural entropy. Kybernetika *3* (1967), 30–35.

[9] *D.S. Hooda*: A non-additive generalized measure of relative useful information. Pure App. Math. Sci. *22* (1984), no. 1, 2, 143–151.

[10] *P.L. Kannappan*: On some functional equations for additive and nonadditive measures. Stochastica *1* (1980), 15–22.

[11] *S. Kotz, N.M. Johnson, D.W. Boid*: Series representation of quadratic forms in normal variables. I. Central Case. AMS (1967), 823–837.

[12] *K.V. Mardia, J.T. Kent, J.M. Bibby*: Multivariate Analysis. Academic Press, 1982.

[13] *M. Mohan, J. Mitter*: On bounds of useful information measures. Information and Control *39* (1978), 233–236.

[14] *O. Onicescu*: Energie informationelle. C. R. Acad. Sci. Paris, Ser. A *263* (1966), 841–842.

[15] *J.A. Pardo*: Caracterización axiomática de la energía informacional util. Estadística Española *108* (1985), 107–116.

[16] *J.A. Pardo*: On the asymptotic distribution of useful Shannon entropy in a stratified sampling. Metron, LI (1993), no. 1,2, 119–137.

[17] *J.A. Pardo, M.L. Vicente*: Asymptotic distribution of the useful informational energy. Kybernetika *30* (1994), no. 1, 87–99.

[18] *L. Pardo*: Energía informacional util. Trabajos de Estadística e investigación operativa *32* (1981), no. 2, 85–94.

[19] *L. Pardo*: The order α information energy gain in sequential design of experiments. Proceedings Third European Young Statisticians Meeting. 1983, pp. 140–147.

[20] *L. Pardo, D. Morales, V. Quesada*: Plan de muestreo secuencial basado en la energía informacional para una población exponencial. Trabajos de Estadística e I.O. *36* (1985), 233–242.

[21] *L. Pardo*: Order α weighted information energy. Information Science *40* (1986), 155–164.

[22] *L. Pardo*: La energía informacional en el muestreo secuencial: Aplicación a las poblaciones normales. Real Academia de Ciencias Exactas, Físicas y Naturales de Madrid *LXXXI* (1987), 103–115.

[23] *L. Pardo, M.L. Menéndez, J.A. Pardo*: A sequential selection method of a fixed number of fuzzy information systems based on the information energy gain. Fuzzy Sets and Systems *25* (1988), 97–105.

[24] *L. Pardo, M.L. Menéndez*: Applications of the informational energy to the design and comparison of regression experiment in a bayesian context. Journal of Combinatiorics, Information & System Sciences *14* (1989), no. 4, 163–171.

[25] *A. Pérez*: Sur l'energie informationelle de M. Octavio Onicescu. Rev. Roumaine Math. Pures Appli. *12* (1966), 1341–1347.

[26] *C.F. Picard*: Graphs et Questionnaires. Gauthier–Villars, Paris, 1972.
[27] *C.F. Picard*: Weighted probabilistic information measures. J. Comb. & Syst. Sci. *4* (1979), 343–356.
[28] *J.N.K. Rao, A.J. Scott*: The analysis of categorical data from complex sample surveys: chi-squared tests for goodness of fit and independence in two way tables. J. Amer. Stat. Assoc. *76* (1981), 221–230.
[29] *C.E. Shannon*: The Mathematical theory of communications. Bell System Techn. J. *27* (1948), 379–423.
[30] *B.D. Sharma, J. Mitter, M. Mohan*: On measures of useful information. Information and Control *39* (1978), 123–136.
[31] *B.D. Sharma, R.P. Shings*: On generating information measures with preference. JCISS *8* (1983), 61–72.
[32] *B.D. Sharma, I.J. Taneja*: Entropy of type (α, β) and other generalized measures in information theory. Metrika *22* (1975), 205–215.
[33] *R.P. Singh*: On information measure of type (α, β) with preference. Caribb. J. Math. *2* (1983), no. 1 & 2, 25–37.
[34] *A. Theodorescu*: Energie informationnelle et notions apparentes. Trabajos de Estadística e Investigación Operativa *28* (1977), 183–206.
[35] *R.K. Tuteja, S. Chaudhary*: Order α-β weighted information energy. Information Sciences *66* (1992), 53–61.
[36] *I. Vajda*: Axiomatic definition of energy of complete and incomplete probability schemes. Bull. Math. Soc. Sci. Math. Roum. *11* (1967), 197–203.

Authors' addresses: *M.C. Pardo*, Escuela Universitaria de Estadística, Departamento de Estadística e I.O., Universidad Complutense de Madrid, 28040 – Madrid, Spain; *J.A. Pardo*, Facultad de Matemáticas, Departamento de Estadística e I.O., Universidad Complutense de Madrid, 28040 – Madrid, Spain.