KYBERNETIKA — VOLUME 30 (1994), NUMBER 1, PAGES 87–99

ASYMPTOTIC DISTRIBUTION OF THE USEFUL INFORMATIONAL ENERGY

J. A. PARDO AND M. L. VICENTE

We consider the Useful Informational Energy introduced by L. Pardo [5]. Our purpose is to study the asymptotic distribution of its analogue estimator, in a random and stratified sampling, as well as its application to testing hypotheses.

1. INTRODUCTION

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. We denote by $\mathcal{X}$ the set of all categories or classes, i.e.

$$\mathcal{X} = \{x_1, \ldots, x_M\}$$

and let

$$\Delta_M = \left\{ P = (p_1, \ldots, p_M) \Big/ \sum_{i=1}^{M} p_i = 1,\ p_i \geq 0,\ i = 1, \ldots, M \right\}$$

be the set of all probability distributions over $\mathcal{X}$. The informational energy of X is given by

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

for all $P \in \Delta_M$. This measure was introduced on Information Theory by Onicescu [3], by analogy to kinetic energy in the classical mechanics. Some interesting applications and properties of the Information Energy can be seen in L. Pardo ([5, 6, 7, 8] and [9]), L. Pardo et al. [10], Perez [11], Theodorescu [15] and Vajda [16]. Vajda [16] and Theodorescu [15] present axiomatics treatment of the expression (1).

The measure (1) depends only on the probabilities of the events and does not take into account the effectiveness of the events under consideration. In order to distinguish the elements $x_1, \ldots, x_M$ of $\mathcal{X}$ in according to their importance with respect to a given qualitative characteristic of the system, we shall ascribe to each outcome x_k a non-negative number $u_k \geq 0$ directly proportional to its importance. We call u_k the utility or weight of the element x_k. It should be pointed out that the term utility is not necessarily intended in this paper according to the classical sense (that is, the quantification of the nature of outcomes is not necessarily defined according to the von Neumann–Morgenstern axiomatic system).

In this context, Theodorescu [15] defined the following generalization of Onicesu's information energy

$$e\,U(P) = \sum_{i=1}^{M} u_i p_i^2 \tag{2}$$

where

$$U = (u_1, \ldots, u_M).$$

In this line Pardo [5] defined the Useful Informational Energy as follows

$$e\,U(P) = \frac{1}{E_P[U]} \sum_{i=1}^{M} u_i p_i^2 \tag{3}$$

where

$$E_P[U] = \sum_{i=1}^{M} u_i p_i$$

and he analyzed some of its properties. Also, J. A. Pardo [4] gave an axiomatic characterization of this measure.

In this paper, we obtain the asymptotic distribution of the analogue estimate of the expression (3) in a random and stratified sampling. The knowledge of this asymptotic distribution allows us to construct tests of hypotheses. We also analyze the important case when $\mathcal{X}$ has two elements.

If all the weights are the same then the expression (3) becomes a particular case of various Schur concave entropies

$$H(P) = \Psi\left(\sum_{i=1}^{M} \phi(p_i)\right),$$

where ϕ is concave (convex) and Ψ increasing (decreasing). For example of the order 2 entropy

$$H(P) = -\log\left(\sum_{i=1}^{M} p_i^2\right)$$

of Rényi [11], or the quadratic entropy

$$H(P) = 1 - \left(\sum_{i=1}^{M} p_i^2\right)$$

of Vajda [17]. For most of these entropies similar asymptotic results have been derived.

2. ASYMPTOTIC DISTRIBUTION OF $eU(\hat{P})$ IN A RANDOM SAMPLING

Consider a sample of n members drawn at random with replacement from the population. If $\hat{p}_i$ $(i = 1, \ldots, M)$ denotes the relative frequencies of the class x_i in the sample of size n, $i = 1, \ldots, M$, the analogue estimate for the useful informational energy is given by

$$eU(\hat{P}) = \frac{1}{E[U]} \sum_{i=1}^{M} u_i \hat{p}_i^2$$

where

$$E[U] = \sum_{i=1}^{M} u_i \hat{p}_i.$$

In this situation, the random vector $(n\hat{p}_1, \ldots, n\hat{p}_M)$ has a multinomial distribution with parameters $(np_1, \ldots, np_M)$. The following theorem establishes the asymptotic distribution of $eU(\hat{P})$.

Theorem 1. If we consider the analogue estimate $eU(\hat{P})$ obtained by replacing p_i's by the observed frequencies $\hat{p}_i$'s $(i = 1, \ldots, M)$ then if $\sigma^2 > 0$

$$n^{\frac{1}{2}} \left(eU(\hat{P}) - eU(P) \right) \xrightarrow[n \uparrow \infty]{L} N(0, \sigma^2)$$

where

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

with

$$t_i = \frac{2u_i p_i E[U] - u_i \left(\sum_{i=1}^{M} u_i p_i^2 \right)}{(E[U])^2}.$$

Moreover, if $\hat{\sigma}^2$ is the estimate of σ^2 obtained by replacing p_i's by $\hat{p}_i$'s,

$$\left(n^{1|2} \big/ \hat{\sigma}^2 \right) \left(eU(\hat{P}) - eU(P) \right) \xrightarrow[n \uparrow \infty]{L} N(0, 1).$$

Proof. We define the function $g : \mathbb{R}^{M-1} \to \mathbb{R}$ given by

$$g(x_1, \ldots, x_{M-1}) = \frac{1}{E[U]} \sum_{i=1}^{M-1} u_i x_i^2 + \frac{1}{E[U]} u_M \left(1 - \sum_{i=1}^{M-1} x_i \right)^2$$

where

$$E[U] = \sum_{k=1}^{M-1} x_k u_k + \left(1 - \sum_{k=1}^{M-1} x_k \right) u_M.$$

Consider Taylor's expansion of $g(x_1,\ldots,x_{M-1})$ at point $\hat{P}_0=(\hat{p}_1,\ldots,\hat{p}_{M-1})$ in a neighbourhood of $P_0=(p_1,\ldots,p_{M-1})$, which is given by

$$g(\hat{P}_0)=g(P_0)+\sum_{i=1}^{M-1}\left(\frac{\partial g(x_1,\ldots,x_{M-1})}{\partial x_i}\right)_{(p_1,\ldots,p_{M-1})}(\hat{p}_i-p_i)+R_n$$

where R_n is the Lagrange rest.

Observe that

$$g(\hat{P}_0)=eU(\hat{P}) \quad \text{and} \quad g(P_0)=eU(P)$$

where

$$\hat{P}=(\hat{p}_1,\ldots,\hat{p}_M) \quad \text{and} \quad P=(p_1,\ldots,p_M).$$

Then,

$$eU(\hat{P})=eU(P)+\sum_{i=1}^{M}t_i(\hat{p}_i-p_i)+R_n$$

where

$$t_i=\frac{2u_ip_iE[U]-u_i\left(\sum_{i=1}^{M}u_ip_i^2\right)}{(E[U])^2}.$$

Therefore, the random variables

$$eU(\hat{P})-eU(P) \quad \text{and} \quad \sum_{i=1}^{M}t_i(\hat{p}_i-p_i)$$

converge in law to the same distribution because R_n converges in probability to zero.

As

$$n^{\frac{1}{2}}(\hat{p}_1-p_1,\ldots,\hat{p}_M-p_M)\ \underset{n\uparrow\infty}{\xrightarrow{L}}\ N(0,\Sigma),$$

where

$$\Sigma=(p_i(\delta_{ij}-p_j))_{\substack{i=1,\ldots,M\\ j=1,\ldots,M}}$$

we get

$$n^{\frac{1}{2}}\sum_{i=1}^{M}t_i(\hat{p}_i-p_i)\ \underset{n\uparrow\infty}{\xrightarrow{L}}\ N(0,T^t\Sigma T)$$

with

$$T=(t_1,\ldots,t_M)^t.$$

Therefore,

$$n^{\frac{1}{2}}\left(eU(\hat{P})-eU(P)\right)\ \underset{n\uparrow\infty}{\xrightarrow{L}}\ N(0,\sigma^2)$$

where

$$\sigma^2=T^t\Sigma T=\sum_{i=1}^{M}p_it_i^2-\left(\sum_{i=1}^{M}p_it_i\right)^2.$$

The second part follows from (6a.2.7.) of Rao [12], pp. 387. □

When $u_i=u, i=1,\ldots,M$ and $p_i=1/M,\ i=1,\ldots,M,\ \sigma^2=0$. Also, if $M=2$ and we consider, for example, $p_1=1/3$ and $p_2=2/3$ if $u_2=\frac{1}{8}u_1$ it is verified that $\sigma^2=0$. In this situations we have obtained the following result.

Theorem 2. If we consider the analogue estimate $eU(\hat{P})$ obtained by replacing p_i's by the observed frequencies $\hat{p}_i$'s $(i = 1, \dots, M)$ then, if $\sigma^2 = 0$

$$n\left(eU(\hat{P}) - eU(P)\right) \xrightarrow[n\uparrow\infty]{L} \sum_{i=1}^{M} \beta_i \chi_1^2,$$

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

$$A = \frac{1}{E[U]}\left(\begin{pmatrix} u_1 & & \\ & \ddots & \\ & & u_M \end{pmatrix} - \left(\frac{u_i u_j (p_i + p_j)}{E[U]} - \frac{u_i u_j \sum_{k=1}^{M} u_k p_k^2}{(E[U])^2}\right)_{\substack{i=1,\dots,M \\ j=1,\dots,M}}\right)$$

and

$$\Sigma = \left(p_i\left(\delta_{ij} - p_j\right)\right)_{\substack{i=1,\dots,M \\ j=1,\dots,M}}.$$

Proof. By considering Taylor's expansion of function g given in Theorem 1, including the term corresponding to the second partial derivatives we get

$$eU(\hat{P}) = eU(P) + (\hat{p}_1 - p_1, \dots, \hat{p}_M - p_M) A \begin{pmatrix} \hat{p}_1 - p_1 \\ \dots \\ \hat{p}_M - p_M \end{pmatrix} + R_n$$

where

$$A = \frac{1}{E[U]}\left(\begin{pmatrix} u_1 & & \\ & \ddots & \\ & & u_M \end{pmatrix} - \left(\frac{u_i u_j (p_i + p_j)}{E[U]} - \frac{u_i u_j \sum_{k=1}^{M} u_k p_k^2}{(E[U])^2}\right)_{i,j}\right)$$

and R_n is the Lagrange rest. Therefore, the random variables

$$eU(\hat{P}) - eU(P) \qquad \text{and} \qquad (\hat{P} - P)^t A (\hat{P} - P)$$

with

$$(\hat{P} - P) = (\hat{p}_1 - p_1, \dots, \hat{p}_M - p_M)^t$$

converge in law to the same distribution because R_n converges in probability to zero.

As,

$$n^{\frac{1}{2}}(\hat{p}_1 - p_1, \dots, \hat{p}_M - p_M) \xrightarrow[n\uparrow\infty]{L} N(0, \Sigma)$$

we have (see Mardia et al. [2], pp. 68)

$$n(\hat{P} - P)^t A (\hat{P} - P) = n\left(eU(\hat{P}) - eU(P)\right) \xrightarrow[n\uparrow\infty]{L} \sum_{i=1}^{M} \beta_i \chi_1^2,$$

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

$$\Sigma = \left(p_i\left(\delta_{ij} - p_j\right)\right)_{\substack{i=1,\dots,M \\ j=1,\dots,M}}.$$

□

Remark 1.
a) If $u_i = u, i = 1, \ldots, M$, and $p_i = 1/M, i = 1, \ldots, M$, the matrix

$$A\Sigma = \begin{pmatrix} 1-\frac{1}{M} & -\frac{1}{M} & \cdots & -\frac{1}{M} \\ \cdots & \cdots & \cdots & \cdots \\ -\frac{1}{M} & -\frac{1}{M} & \cdots & 1-\frac{1}{M} \end{pmatrix}$$

has the eigenvalues $\beta_1 = 0$ with multiplicity 1 and $\beta_2 = 1$ with multiplicity $M-1$, then

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

b) If $M = 2$, we have that the eigenvalues of $A\Sigma$ are $\beta_1 = 0$ and $\beta_2 = vp_1p_2$, where

$$v = \frac{u_1 + u_2}{E[U]} - (2rp_1 - s)u_1^2 - (2rp_2 - s)u_2^2 + 2u_1u_2(r - s),$$

$$r = \frac{1}{(E[U])^2} \qquad \text{and} \qquad s = \frac{\sum_{k=1}^{2} u_k p_k^2}{(E[U])^3}.$$

Thus,

$$\frac{n}{vp_1p_2}\left(eU(\hat{P}) - eU(P)\right)$$

is distributed as a χ-square distribution with one degree of freedom.

The results obtained in Theorems 1 and 2 allow us to construct the following test of hypotheses

a) $\mathrm{H}_0 : eU(P) = eU_0$, i.e., the useful informational energy of a population equals to a specified value. Under H_0, we have to consider two situations according to the value of σ^2. If $\sigma^2 = 0$, then we must use the statistic

$$T_1 = n\left(eU(\hat{P}) - eU_0\right).$$

If H_0 is true, then T_1 will be small. Thus a large value of T_1 indicates data less compatible with the null hypothesis, hence for large n, when $T_1 = t$, one would reject H_0 at a level α if

$$P\left(\sum_{i=1}^{M} \beta_i \chi_1^2 > t\right) \leq \alpha$$

where the β_i's, $i = 1, \ldots, M$, are given by Theorem 2, and the last probability can be computed using the methods given by Kotz et al. [1]. Rao and Scott [13] suggest to consider the aproximate distribution of $\sum_{i=1}^{M} \beta_i \chi_1^2$ which is given by $\overline{\beta}\chi_M^2$, where $\overline{\beta} = \sum_{i=1}^{M} \frac{\beta_i}{M}$. In this case we can easily compute the value of $\overline{\beta}$, since $\sum_{i=1}^{M} \beta_i = \mathrm{tr}(A\Sigma)$.

If we have a Bernoulli population of parameter p unknown, $\hat{p}$ is the analogue estimate of p, u_1 denotes the weight of the result with probability p and u_2 the weight for the other result, we can use the part 2 of the remark 1 to test $\mathrm{H}_0 : p = p_0$. In this case one must reject the null hyphothesis at a level α, if the following relation is satisfied

$$\frac{n}{vp_1p_2}\left(eU(\hat{P}) - eU(P)\right) > \chi^2_{1,\alpha}.$$

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

$$Z_1 = \frac{n^{1/2}(eU(\hat{P}) - eU_0)}{\hat{\sigma}}$$

which has approximately a standard normal distribution for sufficiently large n, where $\hat{\sigma}$ is the estimate of σ obtained by replacing $p_i's$ by $\hat{p}_i's$. Intuitively, one would rejet H_0 at a level α if $|Z_1| > z_{\alpha/2}$, where z_α is the quantile of order $1 - \alpha$ for the standard normal distribution.

b) $H_0 : eU(P_1) = eU(P_2)$, i. e., the useful informational energy of two independent populations is the same. In this situation, if $\hat{\sigma}_i$ is positive $(i = 1, 2)$, the statistics to be used is

$$Z_2 = \frac{(n_1n_2)^{\frac{1}{2}}(eU(\hat{P}_1) - eU(\hat{P}_2))}{(n_2\hat{\sigma}_1 + n_1\hat{\sigma}_2)^{\frac{1}{2}}}$$

which has approximately a 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, $(i = 1, 2)$.

c) $H_0 : eU(P_1) = \ldots. = eU(P_s)$, i.e., the useful informational energy of s independent populations coincides. If σ_i $(i = 1, \ldots, s) > 0$ and we have a sample of size n_i from the ith population, we must consider the statistic

$$G = \sum_{i=1}^{s} n_i \frac{(eU(\hat{P}_i) - \overline{e})^2}{\hat{\sigma}_i^2}$$

where

$$\overline{e} = \frac{\sum_{i=1}^{s} n_i \frac{eU(\hat{P}_i)}{\hat{\sigma}^2}}{\sum_{i=1}^{s} \frac{n_i}{\hat{\sigma}_i^2}}$$

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

3. ASYMPTOTIC DISTRIBUTION OF $^*eU(\hat{P})$ IN A STRATIFIED SAMPLING

Now we suppose that the population with N individuals can be divided into r non-overlapping subpopulations, called strata, as homogeneous as possible with respect to X.

Let N_k be the number of individuals into the kth stratum, p_{ik} the probability that a randomly selected number belongs to the kth stratum and to the class x_i, $p_{i.}$

the probability that a randomly selected number in the whole population belongs to the class x_i, and $p_{.k}$ the probability into the kth stratum, then one obtains

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

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 Useful Informational Energy in the population, we shall draw at random a stratified sample of size n, independently from diferent 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 at random from the kth stratum, where $w_k = n_k/n$. If $\hat{p}_{ik}$ denotes the relative frequency, in the size n sample, of the value x_i into the kth stratum, and we denote by

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

$eU(P)$ can be estimated by

$${}^{*}eU(\hat{P}) = \frac{1}{E_{\hat{P}}[U]} \sum_{i=1}^{M} u_i \left(\sum_{k=1}^{r} \frac{W_k}{w_k} \hat{p}_{ik} \right)^2$$

where

$$\hat{P} = \left(\frac{W_1}{w_1}\hat{p}_{11}, \ldots, \frac{W_1}{w_1}\hat{p}_{M1}, \ldots, \frac{W_r}{w_r}\hat{p}_{1r}, \ldots, \frac{W_r}{w_r}\hat{p}_{Mr} \right)$$

and

$$E_{\hat{P}}[U] = \sum_{i=1}^{M} \left(\sum_{k=1}^{r} \frac{W_k}{w_k} \hat{p}_{ik} \right) u_i.$$

The following theorem establishes the asymptotic behavior of ${}^{*}eU(\hat{P})$.

Theorem 3. If we consider the estimate ${}^{*}eU(\hat{P})$ obtained by replacing $p_{i.}$'s by $\hat{p}_{i.}$'s $(i = 1, .., M)$ then if ${}^{*}\sigma^2 > 0$

$$n^{\frac{1}{2}} \left({}^{*}eU(\hat{P}) - {}^{*}eU(P) \right) \xrightarrow[n\uparrow\infty]{L} N(0, {}^{*}\sigma^2)$$

where

$${}^{*}\sigma^2 = \sum_{i=1}^{M}\sum_{k=1}^{r} \frac{W_k}{w_k} p_{ik} \left(T(u_i, p_{i.}) \right)^2 - \sum_{k=1}^{r} \frac{1}{w_k} \left(\sum_{i=1}^{M} p_{ik} T(u_i, p_{i.}) \right)^2$$

with

$$T(u_i, p_{i.}) = \frac{2u_i p_{i.} E_P[U] - u_i \left(\sum_{i=1}^{M} u_i p_{i.}^2\right)}{(E_P[U])^2}.$$

Moreover, if $^*\hat{\sigma}^2$ is the estimate of $^*\sigma^2$ obtained by replacing $p_{i.}$'s by $\hat{p}_{i.}$'s,

$$\left(n^{\frac{1}{2}} / {}^*\hat{\sigma}2\right) \left({}^*eU(\hat{P}) - {}^*eU(P)\right) \xrightarrow[n\uparrow\infty]{L} N(0,1).$$

Proof. Consider Taylor's expansion of function $g : \mathbb{R}^{(M-1)r} \to \mathbb{R}$ given by

$$g\left(x_{11}, \ldots, x_{(M-1)r}\right) = \frac{1}{E[U]} \sum_{i=1}^{M-1} \left(\sum_{k=1}^{r} x_{ik}\right)^2 u_i + \frac{1}{E[U]} \left(1 - \sum_{i=1}^{M-1} \left(\sum_{k=1}^{r} x_{ik}\right)\right)^2 u_M$$

where

$$E[U] = \sum_{i=1}^{M-1} \left(\sum_{k=1}^{r} x_{ik}\right) u_i + \left(1 - \sum_{i=1}^{M-1} \left(\sum_{k=1}^{r} x_{ik}\right)\right) u_M$$

at point $\hat{P}_0 = \left(\left(\frac{W_1}{w_1}\hat{p}_{11}, \ldots, \frac{W_1}{w_1}\hat{p}_{(M-1)1}, \ldots, \frac{W_r}{w_r}\hat{p}_{1r}, \ldots, \frac{W_r}{w_r}\hat{p}_{(M-1)r}\right)\right.$ in a neighbourhood of $P_0 = \left(p_{11}, \ldots, p_{(M-1)1}, \ldots, p_{1r}, \ldots, p_{(M-1)r}\right)$,

$$g(\hat{P}_0) = g(P_0) + \sum_{i=1}^{M-1} \sum_{k=1}^{r} \left(\frac{\partial g(x_{11}, \ldots, x_{(M-1)r})}{\partial x_{ik}}\right)_{(p_{11}, \ldots, p_{(M-1)r})} \left(\frac{W_k}{w_k}\hat{p}_{ik} - p_{ik}\right) + R_n$$

where R_n is the Lagrange rest.

Observe that

$$g(\hat{P}_0) = {}^*eU(\hat{P}) \qquad \text{and} \qquad g(P_0) = {}^*eU(P)$$

where

$$\hat{P} = \left(\frac{W_1}{w_1}\hat{p}_{11}, \ldots, \frac{W_1}{w_1}\hat{p}_{M1}, \ldots, \frac{W_r}{w_r}\hat{p}_{1r}, \ldots, \frac{W_r}{w_r}\hat{p}_{Mr}\right)$$

and

$$P = \left(p_{11}, .., p_{M1}, .., p_{1r}, .., p_{Mr}\right).$$

Then,

$${}^*eU(\hat{P}) = {}^*eU(P) + \sum_{i=1}^{M} T(u_i, p_{i.})\,(\hat{p}_{i.} - p_{i.}) + R_n$$

where

$$T(u_i, p_{i.}) = \frac{2u_i p_{i.} E_P[U] - u_i \left(\sum_{i=1}^{M} u_i p_{i.}^2\right)}{(E_P[U])^2}.$$

Therefore, the random variables

$$eU(\hat{P}) - eU(P) \qquad \text{and} \qquad \sum_{i=1}^{M} T(u_i, p_{i.})(\hat{p}_{i.} - p_{i.})$$

converge in law to the same distribution because R_n converges in probability to zero.

On the other hand, since the random vectors

$$\left(\frac{\hat{p}_{1k}}{w_k}, \ldots, \frac{\hat{p}_{Mk}}{w_k}\right), \quad k = 1, \ldots, r$$

are independent and follow a multinomial distribution of parameters

$$\left(n_k; \frac{p_{1k}}{W_k}, \ldots, \frac{p_{Mk}}{W_k}\right), \quad k = 1, \ldots, r$$

respectively, applying the M-dimensional Central Limit Theorem, we obtain that

$$n_k^{\frac{1}{2}} \left(\frac{\hat{p}_{1k}}{w_k} - \frac{p_{1k}}{W_k}, \ldots, \frac{\hat{p}_{Mk}}{w_k} - \frac{p_{Mk}}{W_k}\right) \xrightarrow[n \uparrow \infty]{L} N(0, \Sigma(k)), \quad k = 1, \ldots, r$$

where

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

As $n = n_k/w_k$ we have

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

with

$$T = (T(u_1, p_{1.}), \ldots, T(u_M, p_{M.}))^t.$$

Therefore,

$$n^{\frac{1}{2}} \left({}^*eU(\hat{P}) - {}^*eU(P)\right) \xrightarrow[n \uparrow \infty]{L} N(0, {}^*\sigma^2)$$

where

$${}^*\sigma^2 = \sum_{k=1}^{r} \frac{W_k^2}{w_k} T^t \Sigma(k)\, T = \sum_{i=1}^{M} \sum_{k=1}^{r} \frac{W_k}{w_k} p_{ik} \left(T(u_i, p_{i.})\right)^2 - \sum_{k=1}^{r} \frac{1}{w_k} \left(\sum_{i=1}^{M} p_{ik} T(u_i, p_{i.})\right)^2.$$

The second part follows from (6a.2.7.) of Rao [12], pp. 387. □

Remark 2.
1) In the stratified random sampling the asymptotic variance ${}^*\sigma^2$ is minimized for a fixed total size of sample n, if

$$w_k = \frac{\alpha_k^{\frac{1}{2}}}{\sum_{k=1}^{r} \alpha_k^{\frac{1}{2}}} \quad (k = 1, \ldots, r)$$

where

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

furthermore, the minimun asymptotic variance is given by

$$ {}^*\sigma^2_{\text{opt}} = \left(\sum_{k=1}^{r} \alpha_k^{\frac{1}{2}} \right)^2 . $$

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

$$ \frac{\alpha_k^{\frac{1}{2}}}{W_k}, \quad k = 1, \dots, r $$

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

$$ {}^*\sigma^2_{\text{opt}} = \left(\sum_{k=1}^{r} \alpha_k^{\frac{1}{2}} \right)^2 \leq \sum_{k=1}^{r} \frac{\alpha_k}{W_k} = {}^*\sigma^2_{\text{prop}} $$

where ${}^*\sigma^2_{\text{prop}}$ denote the asymptotic variance in the stratified random sampling with proportional allocation and the equality holds if and only if $r = 1$ or $\frac{\alpha_k^{\frac{1}{2}}}{W_k}$ does not depend on $k (k = 1, \dots, r)$.

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

$$ \sum_{i=1}^{M} \frac{1}{W_k} p_{ik} \, T(u_i, p_{i.}), \quad k = 1, \dots, r $$

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

$$ {}^*\sigma^2_{\text{prop}} \leq \sigma^2 $$

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

$$ \sum_{i=1}^{M} \frac{1}{W_k} p_{ik} \, T(u_i, p_{i.}) $$

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

Theorem 4. If we consider the estimate ${}^*eU(\hat{P})$ obtained by replacing $p_{i.}$'s by $\hat{p}_{i.}$'s $(i = 1, .., M)$ then, if ${}^*\sigma^2 = 0$

$$ n \left({}^*eU(\hat{P}) - {}^*eU(P) \right) \xrightarrow[n \uparrow \infty]{L} \sum_{i=1}^{M} \beta_i \, \chi_1^2, $$

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

$$ A = \frac{1}{E[U]} \left(\begin{pmatrix} u_1 & & \\ & \ddots & \\ & & u_M \end{pmatrix} - \frac{u_i u_j (p_{i.} + p_{j.})}{E[U]} - \left(\frac{u_i u_j \sum_{k=1}^{M} u_k p_{k.}^2}{(E[U])^2} \right)_{\substack{i=1,\dots,M \\ j=1,\dots,M}} \right) $$

and

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

Proof. In a similar way to Theorem 2, we can establish that the random variables

$$^{*}eU(\hat{P}) - {}^{*}eU(P) \qquad \text{and} \qquad (\hat{p}_{1.} - p_{1.}, \dots, \hat{p}_{M.} - p_{M.})\, A \begin{pmatrix} \hat{p}_{1.} - p_{1.} \\ \cdots \\ \hat{p}_{M.} - p_{M.} \end{pmatrix}$$

where

$$A = \frac{1}{E[U]} \left(\begin{pmatrix} u_1 & & \\ & \ddots & \\ & & u_M \end{pmatrix} - \left(\frac{u_i u_j (p_{i.} + p_{j.})}{E[U]} - \frac{u_i u_j \sum_{k=1}^{M} u_k p_{k.}^2}{(E[U])^2} \right)_{\substack{i=1,\dots,M \\ j=1,\dots,M}} \right)$$

converge in law to the same distribution.

We now that

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

hence (see Mardia et al. [2], pp. 68)

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

where β_i's are the eigenvalues of the matrix $A\Sigma$ with

$$\Sigma = \sum_{k=1}^{r} \frac{W_k^2}{w_k} \Sigma(k).$$

□

Remark 3. The results obtained in Theorems 3 and 4 can be used in a similar form as Theorems 1 and 2 for testing some hypotheses.

(Received July 31, 1992.)

REFERENCES

[1] S. Kotz, N. M. Johnson and D. W. Boid: Series representation of quadratic forms in normal variables. I. Central case. Ann. Math. Statist. *38* (1967), 823–837.

[2] K. V. Mardia, J. T. Kent and J. M. Kent: Multivariate Analysis. Academic Press, New York 1982.

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

[4] J. A. Pardo: Caracterización axiomática de la Energía Informacional Util. Estadística Española *108* (1986), 107–116.

[5] L. Pardo: La Energía Informacional como fundamento de una Teoría de la Información. Memoria Instituto Universitario de Estadística. Madrid 1977.
[6] L. Pardo: Energía Informacional Util. Trabajos de Estadística e Investigación Operativa *32* (1981), 85–94.
[7] L. Pardo: The order α information energy gain in sequential design of experiments. In: Proceedings of Third European Young Statisticians Meeting, (1983), pp. 140–147.
[8] L. Pardo: Plan de muestreo secuencial basado en la Energía Informacional del modelo de Bernouilli. Estadística española *104* (1984), 27–49.
[9] L. Pardo: Information energy of a fuzzy event and of a partition of fuzzy events. IEEE Trans. Systems Man Cybernet. *15* (1985), 139–144.
[10] L. Pardo, M. L. Menéndez and 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.
[11] A. Perez: Sur l'energie informationnelle de M. Octavio Onicescu. Rev. Roumaine Math. Pures Appl. *12* (1966), 1341–1347.
[12] C. R. Rao: Linear Statistical Inference and its Applications. J. Wiley, New York 1973.
[13] J. N. K. Rao and 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. Statist. Assoc. *76* (1981), 221–230.
[14] A. Rényi: On measures of entropy and information. In: Proceedings of 4th Berkeley Symp. on Probab. Theory Math. Statist. *1* (1961), 547–561.
[15] A. Theodorescu: Energie Informationnelle et notions Apparentes. Trabajos de Estadística e Investigación Operativa *28* (1977), 183–206.
[16] I. Vajda: Axiomatic definition of energy of complete and incomplete probability schemes. Bull. Math. Soc. Sci. Math. Roum. *11* (1967), 197–203.
[17] I. Vajda: Bounds on the minimal error probability in checking a finite or countable number of hypotheses. Inform. Transm. Probl. *4* (1968), 9–17.

Dr. J. A. Pardo, Departamento de Estadística e Investigación Operativa, Facultad de Matemáticas, Universidad Complutense de Madrid, 28040-Madrid. Spain.

Dr. M. L. Vicente, Departamento de Estadística e Investigación Operativa, Escuela Universitaria de Estadística, Universidad Complutense de Madrid, 28040-Madrid. Spain.