TAMKANG JOURNAL OF MATHEMATICS
Volume 23, Number 2, Summer 1992

A TEST OF INDEPENDENCE BASED ON THE (r,s)-DIRECTED DIVERGENCE

D. MORALES, L. PARDO, M. SALICRU AND M. L. MENÉNDEZ

Abstract. (r,s)-Directed divergence statistics quantifies the divergence between a joint probability measure and the product of its marginal probabilities on the basis of contingency tables. Asymptotic properties of these statistics are investigated either considering random sampling or stratified random sampling with proportional allocation and independence among strata. To finish some tests of hypotheses of independence are presented.

1. Introduction

Let (X,Y) be a random variable of the discrete type taking on pairs of values (x_i,y_j), $i=1,\dots,M$, and $j=1,2,\dots,K$. We denote

$$P_{XY}=(p_{ij})_{\substack{i=1,\dots,M\\ j=1,\dots,K}} = (P(X=x_i,Y=y_j))_{\substack{i=1,\dots,M\\ j=1,\dots,K}}$$

the joint probability mass function of (X,Y) and by $P_X=(p_{i.})_{i=1,\dots,M}$ and $P_Y=(p_{.j})_{j=1,\dots,K}$ the corresponding marginal probability distributions, respectively, i.e. $p_{i.}=\sum_{j=1}^K p_{ij}$ and $p_{.j}=\sum_{i=1}^M p_{ij}$.

Kullback and Leibler (1951) first introduced a measure of information con-

---

This work was partially supported by the direccion general de investigacion cientifica y tecnica (Dgicyt) under the contract ps89-0019.
Received September 30, 1991.

cerning to P_{XY} and $P_X * P_Y = (p_{i.}p_{.j})_{\substack{i=1,\dots,M\\ j=1,\dots,K}}$, as

$$D(P_{XY} \,||\, P_X * P_Y) \;=\; \sum_{i=1}^{M}\sum_{j=1}^{K} p_{ij} \log \frac{p_{ij}}{p_{i.}p_{.j}} \tag{1}$$

Renyi (1961) first presented a generalization of (1), as

$$D_r^1(P_{XY} \,||\, P_X * P_Y) = (r-1)^{-1} \log\Big[\sum_{i=1}^{M}\sum_{j=1}^{K} p_{ij}^r (p_{i.}p_{.j})^{1-r}\Big], \; r \neq 1,\; r > 0 \tag{2}$$

Another well known generalization of (1) is given by

$$D_s^s(P_{XY} \,||\, P_X * P_Y) = (s-1)^{-1} \Big[\sum_{i=1}^{M}\sum_{j=1}^{K} p_{ij}^s (p_{i.}p_{.j})^{1-s} - 1\Big] \tag{3}$$

The following limits are easy to check

$$\lim_{r\to 1} D_r^1(P_{XY} \,||\, P_X * P_Y) = \lim_{s\to 1} D_s^s(P_{XY} \,||\, P_X * P_Y) = D(P_{XY} \,||\, P_X * P_Y)$$

Sharma and Mittal (1975) studied the following two generalizations

$$D_1^s(P_{XY} \,||\, P_X * P_Y) = (s-1)^{-1} \Big[\exp\big[(s-1) D(P_{XY} \,||\, P_X * P_Y)\big] - 1\Big] \; s \neq 1 \tag{4}$$

$$D_r^s(P_{XY} \,||\, P_X * P_Y) = (s-1)^{-1} \Big[\big[\sum_{i=1}^{M}\sum_{j=1}^{K} p_{ij}^r (p_{i.}p_{.j})^{1-r}\big]^{\frac{s-1}{r-1}} - 1\Big]$$

$$r \neq 1,\; s \neq 1,\; r > 0 \tag{5}$$

Again, we can easily verify the following limits:

$$\lim_{s\to 1} D_r^s(P_{XY} \,||\, P_X * P_Y) \;=\; D_r^1(P_{XY} \,||\, P_X * P_Y)$$
$$\lim_{s\to 1} D_1^s(P_{XY} \,||\, P_x * P_Y) \;=\; D(P_{XY} \,||\, P_X * P_Y)$$

when $r = s$ in (5) we have

$$D_r^s(P_{XY} \,||\, P_X * P_Y) \;=\; D_s^s(P_{XY} \,||\, P_X * P_Y)$$

In this paper we analyze the properties of the analogue estimate of $D_r^s(P_{XY}\| P_X * P_Y)$ in a random sampling as well as its application to testing statistical hypotheses. Taking limits the results obtained are also valid for the divergences $D_1^s(P_{XY}\|P_X * P_Y)$, $D_r^1(P_{XY}\|P_X * P_Y)$ and $D(P_{XY}\|P_X * P_Y)$.

2. Asymptotic distribution of $D_r^s(\hat{P}_{XY}\|\hat{P}_X * \hat{P}_Y)$

Consider a sample of n members drawn at random with replacement from the population. We denote by $\hat{p}_{ij} = n_{ij}/n$, $\hat{p}_{i.} = n_{i.}/n$, $\hat{p}_{.j} = n_{.j}/n$ the sample estimators of p_{ij}, $p_{i.}$ and $p_{j.}$, where n_{ij} is the number of observations of the value (x_i, y_j) $(i = 1,\ldots,M,\ j = 1\ldots,K)$ in the sample, $n_{i.} = \sum_{j=1}^{K} n_{ij}$ and $n_{.j} = \sum_{i=1}^{M} n_{ij}$. The (r,s)-directed divergence in the sample may be quatified as follows:

$$D_r^s(\hat{P}_{XY} \,\|\, \hat{P}_X * \hat{P}_Y) = (s-1)^{-1}\Big[\big[\sum_{i=1}^{M}\sum_{j=1}^{K}\hat{p}_{ij}^{r}(\hat{p}_{i.}\hat{p}_{.j})^{1-r}\big]^{\frac{s-1}{r-1}} - 1\Big]$$

$$r \neq 1,\ s \neq 1,\ r > 0.$$

the other divergence measures in the sample are given by

$$\lim_{s\to 1} D_r^s(\hat{P}_{XY} \,\|\, \hat{P}_X * \hat{P}_Y) = D_r^1(\hat{P}_{XY} \,\|\, \hat{P}_X * \hat{P}_Y)$$

$$\lim_{r\to 1} D_r^s(\hat{P}_{XY} \,\|\, \hat{P}_X * \hat{P}_Y) = D_1^s(\hat{P}_{XY} \,\|\, \hat{P}_X * \hat{P}_Y)$$

$$\lim_{r\to 1} D_r^1(\hat{P}_{XY} \,\|\, \hat{P}_X * \hat{P}_Y) = \lim_{s\to 1} D_1^s(\hat{P}_{XY} \,\|\, \hat{P}_X * \hat{P}_Y) = D(\hat{P}_{XY} \,\|\, \hat{P}_X * \hat{P}_Y)$$

When the sample is drawn at random and with replacement from the population, the random vector $(n\hat{p}_{11},\ldots,\hat{p}_{MK})$ has a multinomial distribution with parameters $(n; p_{11},\ldots,p_{MK})$.

The asymptotic distribution of $D_r^s(\hat{P}_{XX}\|\hat{P}_X * \hat{P}_Y)$ in the random sampling is given in the following theorem:

Theorem 1. *If we consider the analogue estimate $D_r^s(\hat{P}_{X/Y}\|\hat{P}_X * \hat{P}_Y)$ obtained by replacing p_{ij}, $p_{i.}$ and $p_{.j}$ by the observed frequencies $\hat{p}_{ij}$, $\hat{p}_{i.}$ and $\hat{p}_{.j}$,* *then*

$$n^{\frac{1}{2}}\left[D_r^s(\hat{P}_{XY} \,||\, \hat{P}_X * \hat{P}_Y) - D_r^s(P_{XY} \,||\, P_X * P_Y)\right] \xrightarrow[n\uparrow\infty]{L} \mathcal{N}(0, v_{r,s}^2)$$

where

$$v_{r,s}^2 = \sum_{i=1}^{M}\sum_{j=1}^{K} p_{ij}\frac{\partial}{\partial p_{ij}} h_r^s(p_{11},\ldots,p_{MK})^2 - \Big[\sum_{i=1}^{M}\sum_{j=1}^{K} p_{ij}\frac{\partial}{\partial p_{ij}} h_r^s(p_{11},\ldots,p_{MK})\Big]^2$$

and

$$\frac{\partial}{\partial p_{ij}} h_r^s(p_{11},\ldots,p_{MK}) = \begin{cases} \frac{1}{r-1}\Big[\sum_{i=1}^{M}\sum_{j=1}^{K} p_{ij}^r (p_{i.}p_{.j})^{1-r}\Big]^{\frac{s-r}{r-1}} \Big\{ r p_{ij}^{r-1}(p_{i.}p_{.j})^{1-r} \\ \quad +(1-r)\sum_{j=1}^{K} p_{ij}^r p_{i.}^{-r} p_{.j}^{1-r} + (1-r)\sum_{i=1}^{M} p_{ij}^r p_{.j}^{-r} \\ \quad p_{i.}^{1-r}\Big\} \qquad r \neq 1,\ r > 0 \\ \exp\Big[(s-1)\sum_{i=1}^{M}\sum_{j=1}^{K} p_{ij}\log\frac{p_{ij}}{p_{i.}p_{.j}}\Big]\Big[\log\frac{p_{ij}}{p_{i.}p_{.j}} - 1\Big] \\ \qquad r = 1 \end{cases}$$

Proof. Bickel and Doksum (1977, pp. 135) have shown that if $\frac{\partial}{\partial x_i}h(x_1,\ldots,x_l)$ exist and is continuous for all $i = 1,\ldots,l$, then the asymptotic distribution of $T_n = h(\hat{p}_1,\ldots,\hat{p}_l)$ in a random sampling is given by

$$n^{\frac{1}{2}}\left[T_n - h(p_1,\ldots,p_l)\right] \xrightarrow[n\uparrow\infty]{L} \mathcal{N}(0, \sigma^2)$$

where

$$\sigma^2 = \sum_{i=1}^{l} p_i\Big[\frac{\partial h}{\partial p_i}(p_1,\ldots,p_l)\Big]^2 - \Big[\sum_{i=1}^{l} p_i\frac{\partial h}{\partial p_i}(p_1,\ldots,p_l)\Big]^2$$

The derivatives are calculated treating $p_1,\ldots,p_l$ as "independent variables" not linked by $p_1 + p_2 + \ldots + p_l = 1$. If we consider the function

$$h(p_{11},\ldots,p_{MK}) \;=\; D_s^r(P_{XY} \,||\, P_X * P_Y)$$

we obtain the enunciated result, $r \neq 1$ and $s \neq 1$. By continuity of the variance we obtain the ennunciated result for $r > 0$ and $s \in (-\infty, \infty)$.

Theorem 2. *If $P_{XY} = P_X * P_Y$, then*

$$\frac{2n\, D_r^s(\hat{P}_{XY} \mid\mid \hat{P}_X * \hat{P}_Y)}{r} \underset{n\uparrow\infty}{\overset{L}{\longrightarrow}} \chi^2_{(M-1)(K-1)}$$

Proof. Consider the function $\phi(x) = x^r$. A Taylor's expansion of $\phi(x)$ around the point 1 for $x = (\hat{p}_{ij})/(\hat{p}_{i.}\hat{p}_{.j})$ yields

$$\phi[\frac{\hat{p}_{ij}}{\hat{p}_{i.}\hat{p}_{.j}}] = 1 + [\frac{\hat{p}_{ij}}{\hat{p}_{i.}\hat{p}_{.j}} - 1]r + \frac{1}{2}[\frac{\hat{p}_{ij}}{\hat{p}_{i.}\hat{p}_{.j}} - 1]^2 [r(r-1) + \varepsilon_{ijn}]$$

$i = 1, \ldots, M$, $j = 1, \ldots, K$, where $\varepsilon_{i,j,n} \underset{n\uparrow\infty}{\overset{P}{\longrightarrow}} 0$ as $n \to \infty$.

Multiplying both sides of the last expression by $\hat{p}_{i.}\hat{p}_{.j}$ and suming over $i = 1, \ldots, M$, $j = 1, \ldots, K$, we get

$$\sum_{i=1}^{M}\sum_{j=1}^{K} \hat{p}_{ij}^r (\hat{p}_{i.}\hat{p}_{.j})^{1-r} = 1 + \frac{1}{2} r(r-1) \sum_{i=1}^{M}\sum_{j=1}^{K} [\frac{\hat{p}_{ij}}{\hat{p}_{i.}\hat{p}_{.j}} - 1]^2 \hat{p}_{i.}\hat{p}_{.j} + \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{K} [\frac{\hat{p}_{ij}}{\hat{p}_{i.}\hat{p}_{.j}} - 1]^2 \hat{p}_{i.}\hat{p}_{.j}\varepsilon_{ijn} = 1 + x_n$$

Using the binomial expansion of $[1 + x_n]^{\frac{s-1}{r-1}}$, substracting 1 and multiplying by $\frac{2n}{r(s-1)}$, we obtain $\frac{2nD_r^s(\hat{P}_{XY} \mid\mid \hat{P}_X * \hat{P}_Y)}{r} = n \sum_{i=1}^{M}\sum_{j=1}^{K} \frac{1}{\hat{p}_{i.}\hat{p}_{.j}} [\hat{p}_{ij} - \hat{p}_{i.}\hat{p}_{.j}]^2 + \frac{n}{r(r-1)} \sum_{i=1}^{M}\sum_{j=1}^{K} \frac{(\hat{p}_{ij} - \hat{p}_{i.}\hat{p}_{.j})^2}{p_{i.}p_{.j}} \varepsilon_{ijn} + \frac{2(s-1)}{r(r-1)} n x_n \varepsilon'_n$, where $\varepsilon'_n \underset{n\uparrow\infty}{\overset{P}{\longrightarrow}} 0$.

Note that

$$T_n = n \sum_{i=1}^{M}\sum_{j=1}^{K} \frac{1}{\hat{p}_{i.}\hat{p}_{.j}} [\hat{p}_{ij} - \hat{p}_{i.}\hat{p}_{.j}]^2 = \sum_{i=1}^{M}\sum_{j=1}^{K} \frac{[n_{ij} - \frac{n_{i.}n_{.j}}{n}]^2}{\frac{n_{i.}n_{.j}}{n}} \chi^2_{(M-1)(K-1)}$$

which is the well known limit distribution of the classical statistic to test independence in contingency tables.

As $n x_n = \frac{1}{2} r(r-1) T_n + \frac{1}{2} T_n \varepsilon_{ijn}$, then $n x_n$ converges in law to $\frac{1}{2} r(r-1)\chi^2_{(M-1)(K-1)}$ and $n x_n \varepsilon'_n \underset{n\uparrow\infty}{\overset{P}{\longrightarrow}} 0$. So the result follows.

Now we suppose that the population associated with the bidimensional random variable (X,Y) is finite, N elements, and it can divided into L non-overlapping subpopulations, called strata as homogeneous as possible. Let N_j be the number of individuals into the jth stratum (so that $\sum_{j=1}^{L} N_j = N$) and let p_{ijl} be the probability that a randomly selected member belongs to the lth stratum and takes on the value (x_i, y_j), $i = 1, \ldots, M$, $j = 1, \ldots, k$, $l = 1, \ldots, L$. Thus

$$\sum_{i=1}^{M}\sum_{j=1}^{K} p_{ijl} = \frac{N_l}{N}, \qquad \sum_{i=1}^{M}\sum_{j=1}^{K}\sum_{l=1}^{L} p_{ijk} = 1$$

Let $p_{ij.}$ be the probability that a randomly selected number in the whole population takes on the value (x_i, y_j), $p_{ij.} = \sum_{l=1}^{L} p_{ijl}\, i = 1, \ldots, M$ and $j = 1, \ldots, K$, $p_{i.l} = \sum_{j=1}^{K} p_{ijl}$ the marginal probability of the value x_i in the lth stratum and $p_{.jl} = \sum_{i=1}^{M} p_{ijl}$ the marginal probability of the value y_j in the lth stratum. The (r,s)-directed divergence in this context is given by

$$\begin{aligned}&{}^{st}D_r^s(P_{XY} \,||\, P_X * P_Y)\\ &= (s-1)^{-1}\Big\{\Big[\sum_{i=1}^{M}\sum_{j=1}^{K}\big[\sum_{l=1}^{L} p_{ijl}\big]^r \big[\sum_{l=1}^{L} p_{i.l}\big]^{1-r}\big[\sum_{l=1}^{L} p_{.jl}\big]^{1-r}\Big]^{\frac{s-1}{r-1}} - 1\Big\}\\ &\qquad r \neq 1,\ s \neq 1,\ r > 0\end{aligned}$$

The other divergence measures in the stratified sampling are giving by

$$\begin{aligned}\lim_{s\to 1}\ {}^{st}D_r^s(P_{XY} \,||\, P_X * P_Y) &= {}^{st}D_r^1(P_{XY} \,||\, P_X * P_Y)\\ \lim_{r\to 1}\ {}^{st}D_r^s(P_{XY} \,||\, P_X * P_Y) &= {}^{st}D_1^s(P_{XY} \,||\, P_X * P_Y)\end{aligned}$$

$$\lim_{r\to 1}\ {}^{st}D_r^1(P_{XY} \,||\, P_X * P_Y) = \lim_{s\to 1}\ {}^{st}D_1^s(P_{XY} \,||\, P_X * P_Y) = {}^{st}D(P_{X*Y} \,||\, P_X * P_Y)$$

Now we suppose that a stratified sample of size n is drawn at random from the population independently in different strata. We hereafter suppose that the sample is choosen by proportional allocation in each stratum. Assume also that a sample of size n_k is drawn independently at random with replacement from the lth stratum where $n_l/n = N_l/N$. If $\hat{p}_{ijk}$ denotes the relative frequency of the values (x_i, y_j) into the lth stratum (and hence $\sum_{i=1}^{M}\sum_{j=1}^{K} p_{ijk} = n_l/n$),

$\hat{p}_{i.l} = \sum_{j=1}^{K} \hat{p}_{ijl}$ and $\hat{p}_{.jl} = \sum_{i=1}^{M} \hat{p}_{ijk}$, the (r,s)-directed divergence in the sample may be quantified by ${}^{st}D_r^s(\hat{P}_{XY} \| \hat{P}_X * P_Y)$. In this context we establish the following theorem.

Theorem 3.

$$n^{\frac{1}{2}}\left[{}^{st}D_r^s(\hat{P}_{XY} \parallel \hat{P}_X * \hat{P}_Y) - {}^{st}D_r^s(P_{XY} \parallel P_X * P_Y)\right] \xrightarrow[n\uparrow\infty]{L} \mathcal{N}(0, {}^{st}v^2)$$

where

$${}^{st}v^2 = \sum_{i=1}^{M}\sum_{j=1}^{K} {}_s^r d_{ij} p_{ij.} - \sum_{l=1}^{L} \frac{N}{N_l}\Big[\sum_{i=1}^{M}\sum_{j=1}^{K} {}_s^r d_{ij} p_{ijl}\Big]^2$$

and

$${}_s^r d_{ij} = \begin{cases} \dfrac{1}{r-1}\Big[\sum_{i=1}^{M}\sum_{j=1}^{K} p_{ij.}^r (p_{i..}p_{.j.})^{1-r}\Big]^{\frac{s-r}{r-1}} \Big\{ r p_{ij.}^{r-1}(p_{i..}p_{.j.})^{1-r} \\ \quad + (1-r)\sum_{j=1}^{K} p_{ij.}^r p_{i..}^{-r} p_{.j.}^{1-r} + (1-r)\sum_{i=1}^{M} p_{ij.}^r p_{.j.}^{-r} p_{i..}^{1-r} \Big\}, \ r \neq 1 \\ \exp\Big\{(s-1)\sum_{i=1}^{M}\sum_{j=1}^{K} p_{ij.} \log \dfrac{p_{ij.}}{p_{i..}p_{.j.}}\Big\}\Big[\log \dfrac{p_{ij.}}{p_{i..}p_{.j.}} - 1\Big], \ r = 1 \end{cases}$$

Proof. First, we prove the result for $r \neq 1$, $s \neq 1$. Consider the vector $a = (a_{ijk},\ i = 1,\dots,M;\ j = 1,\dots,K;\ l = 1,\dots,L)$ with $a_{MKL}(l = 1,\dots,L)$ excluded and the function

$$\begin{aligned} g_r^s(a) &= (s-1)^{-1}\Big\{\Big[\sum\Big[\sum_{l=1}^{L} a_{ijl}\Big]^r \Big[\Big[\sum_{h=1}^{K}\sum_{l=1}^{L} a_{ihl}\Big]\Big[\sum_{h=1}^{M}\sum_{l=1}^{L} a_{hjl}\Big]\Big]^{1-r} \\ &\qquad (i,j) \in A = \{(i,j)/(i,j) \neq (M,K), i = 1,\dots,M, j = 1,\dots,K\} \\ &\qquad + \Big[1 - \sum_{(i,j)\in A}\Big[\sum_{l=1}^{L} a_{ijl}\Big]\Big]^r \Big[1 - \sum_{j=1}^{K-1}\Big[\sum_{l=1}^{L}\sum_{i=1}^{M} a_{ihl}\Big]\Big]^{1-r} \\ &\qquad \cdot \Big[1 - \sum_{i=1}^{M-1}\Big[\sum_{l=1}^{L}\sum_{j=1}^{K} a_{ijl}\Big]\Big]^{1-r}\Big]^{\frac{s-1}{r-1}} - 1\Big\} \ r \neq 1,\ s \neq 1 \end{aligned}$$

Let

$\hat{P}_* = (\hat{p}_{ijl},\ i = 1,\dots,M;\ j = 1,\dots,K;\ l = 1,\dots,L)$ with $\hat{p}_{MKl}(l = 1,\dots,L)$ excluded, and

$P_* = (p_{ijk},\ i = 1,\dots,M;\ j = 1,\dots,K;\ l = 1,\dots,L)$ with $p_{MKl}(l = 1,\dots,L)$ excluded.

Let us also define

$$\hat{P} = (\hat{p}_{ijl},\ i = 1,\dots,M;\ j = 1,\dots,K;\ l = 1,\dots,L)$$

and

$$P = (p_{ijl},\ i = 1,\dots,M;\ j = 1,\dots,K;\ l = 1,\dots,L).$$

If we consider the Taylor's expansion of $g_r^s(\hat{P}_*)$ in a neigbourhood of P_*, we obtain that

$$n^{\frac{1}{2}}[g_r^s(\hat{P}) - g_r^s(P)] = n^{\frac{1}{2}}[{}^{st}D_r^s(\hat{P}_{XY} \mid\mid \hat{P}_X * \hat{P}_Y) - {}^{st}D_r^s(P_{XY} \mid\mid P_X * P_Y)]$$

and

$$n^{\frac{1}{2}} \sum_{i=1}^{M} \sum_{j=1}^{K} \sum_{l=1}^{L} \frac{\partial g_r^s(P)}{\partial p_{ijl}} [\hat{p}_{ijl} - p_{ijl}]$$

have asymptotically the same p.d.f..

The random vectors

$$[n\,\hat{p}_{11l},\dots,n\,\hat{p}_{Mkl}], \quad l = 1,\dots,L$$

are independent and multiomial distributed with parameters

$$[n_l;\ \frac{N}{N_l}p_{11l},\dots,\frac{N}{N_l}p_{Mkl}], \quad l = 1,\dots,L$$

Applying the MK-dimensional Central Limit Theorem, we obtain

$$n_l^{\frac{1}{2}}[[\frac{n}{n_l}\hat{p}_{11l} - \frac{N}{N_l}p_{11l}],\dots,[\frac{n}{n_l}\hat{p}_{MKl} - \frac{N}{N_l}p_{MKl}]] \xrightarrow[n\uparrow\infty]{L} \mathcal{N}(0,\Sigma(l))$$
$$l = 1,\dots,L$$

where

$$\sum(l) = [\frac{N}{N_l}p_{(i_1,j_1)l}(\delta_{(i_1,j_1)(i_2,j_2)} - \frac{N}{N_l}p_{(i_2,j_2)l})]_{\substack{i_1,i_2=1,\dots,M \\ j_1,j_2=1,\dots,K}}$$

and

$\delta_{(i_1,j_1)(i_2,j_2)} = 1$ if $(i_1 = i_2, j_1 = j_2)$ and $\delta_{(i_1,j_1)(i_2,j_2)} = 0$ otherwise

As $n/n_l = N/N_l$ and $n_l^{1/2} = n^{1/2}(N_l/N)^{1/2}$, we have

$$X_l^t = n^{\frac{1}{2}}\left[\frac{N}{N_l}\right]^{\frac{1}{2}}\left[(\hat{p}_{11l} - p_{11l}), \ldots, (\hat{p}_{MKl} - p_{MKl})\right] \xrightarrow[n\uparrow\infty]{L} \mathcal{N}(0, \Sigma(l)).$$

$$l = 1, \ldots, L$$

Therefore, the asymptotic probability distribution function of the linear function $b_l^t X_l$, where

$$b_l^t = \left[\frac{\partial g_r^s(P)}{\partial p_{11l}}, \ldots, \frac{\partial g_r^s(P)}{\partial p_{MKl}}\right], \qquad l = 1, \ldots, L$$

is normal with mean zero and variance $b_l^t \Sigma(l) b_l$.

As $X_1, X_2, \ldots, X_L$ are independent vectors,

$$\sum_{l=1}^{L}\left[\frac{N_l}{N}\right]^{\frac{1}{2}} b_l^t X_l^t = n^{\frac{1}{2}} \sum_{l=1}^{L}\sum_{i=1}^{M}\sum_{j=1}^{K} \frac{\partial g_r^s(P)}{\partial p_{ijl}}(\hat{p}_{ijl} - p_{ijl})$$

has a normal asymptotic probability distribution function with mean zero and variance

$$^{st}v^2 = \frac{1}{N}\sum_{l=1}^{L} N_l b_l^t \Sigma(l)\, b_l$$

Now we calculate $^{st}v^2$ explicitly. As

$$\frac{\partial g_r^s(P)}{\partial p_{ijl}} = \frac{1}{r-1}\left[\sum_{i=1}^{M}\sum_{j=1}^{K} p_{ij.}^r (p_{i..}p_{.j.})^{1-r}\right]^{\frac{s-r}{r-1}} \Big\{ r\, p_{ij.}^{r-1}(p_{i..}p_{.j.})^{1-r}$$
$$+ (1-r)\sum_{j=1}^{K} p_{ij.}^r p_{i..}^{-r} p_{.j.}^{1-r} + (1-r)\sum_{i=1}^{M} p_{ij.}^r p_{.j.}^{-r} p_{i..}^{1-r} \Big\}$$

we obtain the expression of $^{st}v^2$ for $r \neq 1$.

Now, by contynuity of ^{st}v we obtain the enunciated result.

Remark 1. (a) Applying Jensen's inequality to the convex function $\gamma(x) = x^2$, we obtain $^{st}v^2 \leq v^2$. Equality holds if and only if $L = 1$ or

$$\frac{Nl}{N}\sum_{i=1}^{M}\sum_{j=1}^{K} p_{ijl}\frac{\partial g_r^s(P)}{\partial p_{ijl}}$$

does not depend on l ($l = 1, \ldots, L$).

(b) As the random variables

$$2n\,{}^{st}D_r^s(\hat{P}_{XY} \,||\, P_X * P_Y) \quad \text{and} \quad n\sum_{i=1}^{M}\sum_{j=1}^{K}\frac{1}{p_{i..}p_{.j.}}(\hat{p}_{ij.} - p_{i..}p_{.j.})^2$$

converge in law to the same distribution under the hypothesis of independence and

$$Y' = n^{\frac{1}{2}}\left[(\hat{p}_{11.} - p_{1..}p_{.1.}), \ldots, (\hat{p}_{MK.} - p_{M..}p_{.K.})\right] \xrightarrow[n\uparrow\infty]{L} N\left[0, \sum_{l=1}^{L}\frac{N_l}{N}\Sigma(l)\right]$$

If $P_{XY} = P_X * P_Y$ and $P_X = P_1$, $P_Y = P_2$ with P_1 and P_2 known, then

$$T_3 = n\,Y'CY = 2\,n\,{}^{st}D_r^s(\hat{P}_{XY} \,||\, P_X * P_Y) \xrightarrow[n\uparrow\infty]{L} \sum_{h=1}^{MK}\beta_h\chi_1^2$$

where β_i's are the eigenvalues of the matrix $C\Sigma^*$, being

$$C = \text{diag}_{MK\times MK}\left[\frac{1}{p_{i..}p_{.j.}}\right]_{\substack{i=1,\ldots,M\\ j=1,\ldots,K}}, \quad \Sigma^* = \sum_{l=1}^{L}\frac{N_K}{N}\Sigma(l)$$

and the χ_1^2's are independent (see Mardia et al, 1979, pp. 68).

3. Tests of Independence

Let X and Y be two random variables with joint distribution function $F(x,y)$, and let F_X and F_Y be the marginal distribution function of X and Y, respectively. In this section we study some tests of the hypotheses of independence, namely, $H_0 : F(x,y) = F_X(x)F_Y(y)$ for all $(x,y) \in \mathbb{R}^2$, against the alternative $H_1 : F(x,y) \neq F_X(x)F_Y(y)$ for some (x,y).

We suppose that we have n observations on (X,Y). Let us divide the space of values assumed by X (the real line) into M mutually intervals $I_1, \ldots, I_M$. Similarly, the space of values of Y is divided into K disjoints intervals $J_1, \ldots, J_K$. Let $\hat{p}_{ij}$ be the observed frequency of cell (i,j), and let $p_{ij} = P((X,Y) \in$

$I_i \times J_j) = P(X \in I_i$ and $Y \in J_j)$, where $i = 1,\ldots,M$ and $j = 1,\ldots,K$. Then the random vector $(n\hat{p}_{11},\ldots,n\hat{p}_{MK})$ has a multinomial distribution with parameters $(n; p_{11},\ldots,p_{MK})$. The hypothesis to be tested is $H_0 : p_{ij} = p_{i.}p_{.j}$ $i = 1,2,\ldots,M$, $j = 1,\ldots,K$, where $p_{i.} = P(X \in I_i)$ and $p_{.j} = P(Y \in J_j)$.

We consider the statistics given in theorem 2

$$T_1 = \frac{2n\, D_r^s(\hat{P}_{XY} \,||\, \hat{P}_X * \hat{P}_Y)}{r}$$

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 a level α test is given y

$$\Phi(\hat{p}_{11},\ldots,\hat{p}_{MK}) = \begin{cases} 1 & \text{if } T_1 > \chi^2_{(M-1)(K-1),\alpha} \\ 0 & \text{otherwise} \end{cases}$$

The theorem 1 can be used to evaluate the asymptotic power of the previous test when $(p_{11},\ldots,p_{MK})$ is not equal to $(p_{1.}p_{.1},\ldots,p_{M.}p_{.k})$.

A second possibility appears when the hypothetical probabilities $p_{i.}$ and $p_{.j}$ specifying the marginal distributions may be known, in which case we are required to examine whether the probabilities p_{ij} [$= P(X = x_i, Y = y_j)$] could be constructed by the law

$$p_{ij} = p_{i.}\,p_{.j}$$

Let X and Y be two random variables with joint distribution function $F(x,y)$, and let F_X and F_Y be the known marginal distribution of X and Y respectively. A test of hypothesis to test $H_0 : F(x,y) = F_X(x)\,F_Y(y)$ for all $(x,y) \in \mathbb{R}^2$ against the alternative $H_1 : F(x,y) \neq F_X(x)\,F_Y(y)$ for some $(x,y) \in \mathbb{R}^2$ can be constructed

If we have n observations on (X,Y) obtained by random sampling, we consider the statistics

$$T_2 = \frac{2n\, D_r^s(\hat{P}_{XY} \,||\, P_X * P_Y)}{r}$$

If H_0 is true, then T_2 is asymptotically chi-square distributed with $MK-1$ degrees of freedom. Hence for large n, one must reject H_0 at a level α if $T_2 > \chi^2_{MK-1,\alpha}$.

On the other hand, if we have n observations on (X,Y) obtained by stratified random sampling with proportional allocation, we consider the statistics T_3 given in remark $1(b)$. If H_0 is true, then T_3 will be small. Thus a large value of T indicates data less compatible with the null hypothesis. Hence for large n, when $T_3 = t$, one must reject h_0 at level α if

$$P[\sum_{h=1}^{MK} \beta_h \chi_1^2 > t] \le \alpha$$

where β_h's are given in remark $1(b)$. This probability can be computed using the methods given by Kotz et al (1967).

References

[1] P. J. Bickel, and K. A. Doksum, *Mathematical Statistics* Holden-Day, Inc. Oakland, California, (1977).

[2] M. A. Gil, "A note on stratification and gain in precision in estimating diversity from large samples", *Commun. Statist.* (*Theory and Methods*) 18(4), (1989), 1521-1526.

[3] S. Kotz, N. M. Johnson, and D. W. Boid, "Series representation of distributions of quadratic forms in normal variables", *I. Central Case*, AMS 823-837, (1967).

[4] S. Kullback, and A. Leibler, "On the Information and Sufficiency", *Ann. Math. Statist.*, 27, (1951), 986-1005.

[5] K. V. Mardia, J. T. Kent, and J. M. Bibby, "Multivariate analysis", *Academic Press*, (1979).

[6] M. L. Menendez, D. Morales, L. Pardo, and M. Salicru, "Some Applications of (r,s)-Directed divergences", To appear in Utilitos Mathcmatica.

[7] D. Morales, L. Pardo, "Statistical Application of Hypoentropy Measure to Populations Diversity Analisis", To appear in Trabajos de Estadística, (1991).

[8] T. K. Nayak, "Applications of entropy functions in measurement and analysis of diversity", *Ph. D. thesis. Un. Pittsburgh. Ed. Un. Microfilm International*, (1983).

[9] T. K. Nayak, "On Diversity Measures based on Entropy Functions", *Commun. Statis.-Theory and Methods*, 14, 1, (1985), 203-215.

[10] L. Pardo, and D. Morales, "An Index of Diversity in Stratified Random Sampling based on the Hypoentropy Measure", Qüestiio, 14, 1,2,3,(1991), 11-25

[11] L. Pardo, J. A. Pardo, M. L. Menendez, "Unified (r,s)-Entropy as an index of Diversity",To appear in Journol of the FranKlin Institute

[12] C. R. Rao, "Linear Statistcal Inference and its Applications", *John Wiley*, New York, (1973).

[13] A. Renyi, "On Measures of Entropy and Information". Procd. 4th *Berkeley Symp. Math. Statist. and Prob.*, 1, (1961), 547-561.

[14] M. Salicrú, "Asymptotic properies of (r,s)-Directed Divergence in a Stratified Sampling", (Communicated), (1991).

[15] R. J. Serfling, "Approximation theorems of Mathematical statistics", *Wiley*, New York, (1980).

[16] C. Shannon, "A mathematical Theory of Comunications", *Bell. System. Tech. J.* 27, (1948), 379-423.

[17] B. D. Sharma and D. P. Mittal, "New Nonadditive Measures of Entropy for Discrete Probability Distributions", *J. Math. Sci.*, 10, (1975), 28-40.

[18] K. Zografos, K. Ferentinos, and T. Papaioannou, "φ-Divergence Statistics: Sampling Properties and Multinomial Goodness of fit and Divergence Tests", *Commun. Statist. (Theory Meth.)*, 19(5), (1990), 1785-1802.

Departamento de Estadística e I.O., Facultad de Matemáticas, Universidad Complutense de Madrid, 28040 MADRID (SPAIN).

Departamento de Estadística e I.O., Facultad de Matemáticas, Universidad Complutense de Madrid, 28040 MADRID (SPAIN).

Departamento de Estadística, Universidad de Barcelona, 08028 BARCELONA (SPAIN).

Departamento de Matemática aplicada, E. T. S. de Arquitectura, Universidad Politécnica de Madrid, 28040 MADRID (SPAIN).