ON A CLASS OF PERMUTATION TESTS FOR STOCHASTIC INDEPENDENCE, II

by

Pranab Kumar Sen

University of North Carolina

Institute of Statistics Mimeo Series No. 533

July 1967

Work supported by the National Institute of Health, Public Health Service, Grant GM-12868

DEPARTMENT OF BIOSTATISTICS

UNIVERSITY OF NORTH CAROLINA

Chapel Hill, N. C.

ON A CLASS OF PERMUTATION TESTS FOR STOCHASTIC INDEPENDENCE, II*

By

Pranab Kumar Sen
University of North Carolina at Chapel Hill

SUMMARY. Extending the ideas of Hájek (1962) and Sen (1967a, 1967b), a class of asymptotically most powerful permutation (rank order) tests is offered for testing independence of two stochastic variates when the observable random variables correspond to a finite or countable set of contiguous cells having an underlying continuous distribution. The theory is illustrated by some examples.

1. INTRODUCTION

Let $\underset{\sim}{X}_i = (X_{1i}, X_{2i})$, i=1,...,N be N independent and identically distributed random vectors (i.i.d.r.v.) having a continuous (bivariate) cumulative distribution function (c.d.f.) $F(\underset{\sim}{x})$, $\underset{\sim}{x} \in R^2$, the real plane. On R^2, a finite or countable set of contiguous cells is defined by

$$I_{jk} = \{\underset{\sim}{x}: \; a_{j-1} < x_1 \le a_j, \; b_{k-1} < x_2 \le b_k\} \text{ for } j,k=1,2,\dots \; ; \tag{1.1}$$

where $\{-\infty = a_o < a_1 < a_2 < \dots < \infty\}$ and $\{-\infty = b_o < b_1 < b_2 \dots < \infty\}$ are any two sets of ordered points on the real line $(-\infty,\infty)$. Let then

$$Z_{ijk} = \begin{cases} 1, & \underset{\sim}{X}_i \in I_{jk}, \\ 0, & \text{otherwise}; \end{cases} \quad \underset{\sim}{Z}_i = (Z_{i11}, Z_{i21}, Z_{i12}, Z_{i31}, Z_{i22}, Z_{i13}, \dots) \tag{1.2}$$

for all i=1,...,N. Our observable random variables are $\underset{\sim}{Z}_1, \dots, \underset{\sim}{Z}_N$, and having observed them, we want to test the null hypothesis

$$H_o: \; F(x_1,x_2) = F(x_1,\infty)F(\infty,x_2) \text{ for all } \underset{\sim}{x} \in R^2 \tag{1.3}$$

---

* Work supported by the National Institute of Health, Public Health Service, Grant GM-12868.

A class of permutation tests based on appropriate rank order statistics is offered for testing H_o in (1.3). This generalizes some earlier results of Sen (1967a, Section 5, (5.23) through (5.26)), and the proposed tests are shown to be asymptotically most powerful for certain specific alternative hypotheses which are generalizations of the <u>regression alternative</u> of Hájek (1962) to the case when (i) both the variables are stochastic and (ii) the regression is not necessarily linear. Assuming X_{2i}, i=1,...,N to be non-stochastic, Hájek (1962) has considered the model specified by

$$P\{X_{1i} \leq x | X_{2i}\} = G([x-\alpha-\beta X_{2i}]/\sigma) \quad i=1,\ldots,N, \tag{1.4}$$

where α, β and σ (> 0) are unknown parameters and we desire to test H_o^*: $\beta=0$. Sen (1967b) has considered the same model but the case when X_{1i}, i=1,...,N, are not observable; the observable random variables correspond to a set of finite or countable number of contiguous class intervals on $(-\infty,\infty)$. In the present paper, the general case of both the variates being stochastic is considered, and the tests are based on the observable random variables $\underset{\sim}{Z}_1,\ldots,\underset{\sim}{Z}_N$, defined by (1.2).

Recalling that σ in (1.4), in general, depends on β, we may consider the slightly more general model

$$P\{X_{1i} \leq x | X_{2i}\} = G([x-\alpha-\beta X_{2i}]/\sigma(\beta)), \tag{1.5}$$

where $\sigma(\beta)$ is continuous (in β) in the neighbourhood of $\beta=0$ and attains a maximum at $\beta=0$, (e.g., bivariate normal distribution). In this paper, we consider the following more general model.

Let $F(\infty,x_2) = H(x_2)$ and $F(x_1|x_2) = G(x_1|x_2)$. We assume that there is some regression or association parameter β, such that

$$G(x_1|x_2) = G_\beta(x_1|x_2), \tag{1.6}$$

where β is such that $G_o(x_1|x_2) = G_o(x_1)$ is independent of x_2, i.e., $G_o(x_1)$ is the marginal c.d.f. of x_1. Then, we shall make the following assumptions:

(I) $H(x)$ and $G_\beta(x_1|x_2)$ have continuous density functions $h(x)$ and $g_\beta(x_1|x_2)$, respectively.

(II) There exists a $\delta > 0$, such that for all $|\beta| \leq \delta$

$$\left|\frac{\partial}{\partial\beta} \log \{g_\beta(x_1|x_2\}\right| \leq t(x_1,x_2),\ \underset{\sim}{x} \in R^2, \tag{1.7}$$

where

$$\iint_{R^2} t^2(x_1,x_2)dH(x_2)dG_\beta(x_1|x_2) < \infty. \tag{1.8}$$

(III) $$\frac{\partial}{\partial\beta} \log \{g_\beta(x_1|x_2)\}\Big|_{\beta=0} = C(F)\ell_1(x_1) \cdot \ell_2(x_2), \tag{1.9}$$

where $C(F)$ is a constant (which may depend on F) and $\ell_i(x_i)$ is a sole function of x_i, $i=1,2$. By virtue of (1.7) and (1.8), we obtain that

$$\gamma^2 = [\int_{-\infty}^{\infty} \ell_1^2(x)dG_o(x)][\int_{-\infty}^{\infty} \ell_2^2(x)dH(x)] < \infty . \tag{1.10}$$

Our proposed rank order tests will be asymptotically most powerful for testing H_o: $\beta=0$, against H: $\beta > 0$ when the assumptions in (1.6) through (1.10) hold.

2. ASYMPTOTICALLY MOST POWERFUL PARAMETRIC TEST

Let us denote by

$$P_{jk} = P\{\underset{\sim}{X} \in I_{jk}\},\ P_{jo} = \sum_k P_{jk},\ P_{ok} = \sum_j P_{jk}; \tag{2.1}$$

$$\mu_j = \int_{a_{j-1}}^{a_j} \ell_1(x)dG_o(x)/P_{jo},\ \nu_k = \int_{b_{k-1}}^{b_k} \ell_2(x)dH(x)/P_{ok}, \tag{2.2}$$

for all $j,k=1,2,\ldots$;

$$A^2 = \sum_j \mu_j^2 P_{jo} \text{ and } B^2 = \sum_k \nu_k^2 P_{ok} . \tag{2.3}$$

As in Sen (1967b), it is seen that

$$A^2 \leq \int_{-\infty}^{\infty} \ell_1^2(x)dG_o(x),\ B^2 \leq \int_{-\infty}^{\infty} \ell_2^2(x)dH(x), \tag{2.4}$$

uniformly in $\{a_o, a_1, \ldots\}$ and $\{b_o, b_1, \ldots\}$. Thus, by (1.10), both A^2 and B^2 are finite; they are also assumed to be positive. Now, in the sequel, for testing H_o: $\beta=0$ against $\beta > 0$, we shall conceive of a sequence $\{H_N\}$ of alternative hypotheses, where

$$H_N:\quad \beta = \beta_N = N^{-\frac{1}{2}}\lambda;\quad \lambda \text{ is real and finite.} \tag{2.5}$$

Choosing N sufficiently large (so that $|\beta_N| \leq \delta$ for all $N \geq N_o$), writing $f(x_1,x_2) = h(x_2)g_\beta(x_1|x_2)$ and making use of (1.7), (1.8) and (1.9), we obtain that under H_N:

$$P_{jk} = \int_{a_{j-1}}^{a_j} \int_{b_{k-1}}^{b_k} f(x_1,x_2)dx_1dx_2 = P_{jo}P_{ok}[1+c(F)\beta_N\mu_j\nu_k + o(\beta_N)] \tag{2.6}$$

for all $j,k=1,2,\ldots$. Let us denote by $L(\underset{\sim}{Z}_1,\ldots,\underset{\sim}{Z}_N|\lambda)$ the likelihood function of the sample observations for $\beta_N = N^{-\frac{1}{2}}\lambda$. Then from (2.6), we obtain

$$\begin{aligned} &\log\{L(\underset{\sim}{Z}_1,\ldots,\underset{\sim}{Z}_N|\lambda)/L(\underset{\sim}{Z}_1,\ldots,\underset{\sim}{Z}_N|0)\} \\ &= C(F)\lambda[N^{-\frac{1}{2}} \sum_{i=1}^{N} \sum_j \sum_k Z_{ijk}\mu_j\nu_k] + o(\lambda). \end{aligned} \tag{2.7}$$

Thus, on writing

$$T_N = N^{-\frac{1}{2}}\Sigma_{i=1}^{N} \Sigma_j\Sigma_k Z_{ijk}\mu_j\nu_k , \tag{2.8}$$

it follows from Neyman-Pearson's fundamental lemma [cf. Rao (1965, pp. 375-377)] that for testing H_o: $\lambda = 0$ against $\lambda > 0$ (referred to (2.5)), the asymptotically most powerful size α $(0 < \alpha < 1)$ test is given by

$$\psi_1(\underset{\sim}{Z}_1,\ldots,\underset{\sim}{Z}_N) = \begin{cases} 1, & T_N > T_{N,\alpha} \\ a_{N,\alpha}, & T_N = T_{N,\alpha} \\ 0, & T_N < T_{N,\alpha} \end{cases} \tag{2.9}$$

where $T_{N,\alpha}$ and $a_{N,\alpha}$ $(0 \le a_{N,\alpha} \le 1)$ are so chosen that

$$E\{\psi_1(\underset{\sim}{Z}_1,\ldots,\underset{\sim}{Z}_N)|H_o: \lambda=0\} = \alpha. \tag{2.10}$$

Now, on making use of the fact that $\frac{\partial}{\partial\beta}\iint_{R^2} dF(x_1,x_2)\Big|_{\beta=0} = 0$, we obtain from (1.9) that

$$(\Sigma_j \mu_j P_{jo})(\Sigma_k \nu_k P_{ok}) = 0. \tag{2.11}$$

Let then $Y_i = \Sigma_j\Sigma_k Z_{ijk}\mu_j\nu_k$ for $i=1,\ldots,N$, so that $Y_1,\ldots,Y_n$ are i.i.d.r.v. Using (2.6), (2.11) and following a few simple steps, we obtain that

$$E\{Y_i|H_N\} = C(F)N^{-\frac{1}{2}}\lambda A^2B^2[1+o(\lambda)], \tag{2.12}$$

$$E\{Y_i^2|H_N\} = \Sigma_j\Sigma_k\mu_j^2\nu_k^2P_{jk} < \infty, \text{ by (1.8)}. \tag{2.13}$$

Hence, making use of (2.6), we obtain from (2.13) that

$$V(Y_i|H_N) = A^2B^2 + o(1). \tag{2.14}$$

Thus, using the classical central limit theorem for i.i.d.r.v. [cf. Rao (1965), p. 107)], we obtain the following.

THEOREM 2.1. Under (1.6) through (1.10) and $\{H_N\}$ in (2.5)

$$\mathcal{L}(T_N|AB - C(F)\lambda AB) \rightarrow \mathcal{N}(0,1),$$

i.e., $[T_N/AB - C(F)\lambda AB]$ converges in law to a standard normal distribution.

Consequent on theorem 2.1, we have from (2.5) and (2.9),

$$\lim_{N\to\infty} E\{\psi_1(\underset{\sim}{Z}_1,\ldots,\underset{\sim}{Z}_N)|H_N\} = 1 - \Phi(\tau_\alpha - \lambda C(F)AB) \tag{2.15}$$

where $\Phi(x)$ is the standard normal cdf and $\Phi(\tau_\alpha) = 1-\alpha$.

3. A PROPOSED CLASS OF RANK ORDER TESTS

Let us denote by

$$\Sigma_{i=1}^{N} Z_{ijk} = N_{jk},\ \sum_j N_{jk} = N_{\cdot k},\ \sum_k N_{jk} = N_{j\cdot}\quad, \tag{3.1}$$

$$G_{N,j} = (1/N)\ \Sigma_{\ell=1}^{j} N_{\ell\cdot},\ G_{N,0} = 0,\ H_{N,k} = (1/N)\ \Sigma_{q=1}^{k} N_{\cdot q},\ H_{N,0} = 0, \tag{3.2}$$

for all $j,k=1,2,\ldots$. Let then

$$\mu_{Nj} = \int_{G_{N,j-1}}^{G_{N,j}} \ell_1(G_o^{-1}(u))du/[G_{N,j} - G_{N,j-1}],\ j=1,2,\ldots; \tag{3.3}$$

$$\nu_{Nk} = \int_{H_{N,k-1}}^{H_{N,k}} \ell_2(H^{-1}(u))du/[H_{N,k} - H_{N,k-1}],\ k=1,2,\ldots; \tag{3.4}$$

where $\ell_1(x)$ and $\ell_2(x)$ are defined by (1.9). Our proposed test is then based on the statistic

$$S_N = N^{-\frac{1}{2}} \Sigma_{i=1}^{N} \Sigma_j \Sigma_k \, Z_{ijk} \mu_{Nj} \nu_{Nk} \, . \tag{3.5}$$

To formulate the test function, we let

$$P_{Njo} = G_{N,j} - G_{N,j-1}, \; P_{Nok} = H_{N,k} - H_{N,k-1}, \; j,k=1,2,\ldots; \tag{3.6}$$

$$A_N^2 = \sum_j \mu_{Nj}^2 P_{Njo} \text{ and } B_N^2 = \sum_k \nu_{Nk}^2 P_{Nok} \, . \tag{3.7}$$

It may be noted that as we are dealing with grouped data, the distribution of S_N will depend on the parent c.d.f. $F(\underset{\sim}{x})$ even under the null hypothesis (1.3). However, we may apply the same logic of permutation test theory as in Sen (1967a) to formulate a permutationally distribution free test based on S_N. Avoiding therefore the details of arguments, denoting by $\mathcal{P}_N$ the permutation (conditional) distribution generated by the matching invariant transformations and proceeding as in Sen (1967a), we obtain

$$E(S_N | \mathcal{P}_N) = 0, \; E(S_N^2 | \mathcal{P}_N) = \frac{N}{N-1} A_N^2 B_N^2; \tag{3.8}$$

$$\mathcal{L}_{\mathcal{P}_N}\{S_N / A_N B_N\} \to \mathcal{N}(0,1). \tag{3.9}$$

By analogy with (2.9), we may consider the test function

$$\Psi_2(\underset{\sim}{Z}_1,\ldots,\underset{\sim}{Z}_N) = \begin{cases} 1, & S_N > S_{N,\alpha}(\mathcal{P}_N) \\ b_\alpha(\mathcal{P}_N), & S_N = S_{N,\alpha}(\mathcal{P}_N) \\ 0, & S_N < S_{N,\alpha}(\mathcal{P}_N) \end{cases} \tag{3.10}$$

where $S_{N,\alpha}(\mathcal{P}_N)$ and $b_\alpha(\mathcal{P}_N)$ $(0 < b_\alpha < 1)$ are so chosen that

$$E\{\Psi_2(\underset{\sim}{Z}_1,\ldots,\underset{\sim}{Z}_N) | \mathcal{P}_N\} = \alpha. \tag{3.11}$$

(3.11) implies that the unconditional level of significance is also equal to α.

We intend to prove that

$$\lim_{N\to\infty} E\{\Psi_2(\underset{\sim}{Z}_1,\ldots,\underset{\sim}{Z}_N)|H_N\} = 1 - \Phi(\tau_\alpha - \lambda C(F)AB) \tag{3.12}$$

where $\{H_N\}$ is specified by (2.5) and (Φ,τ_α) by (2.15). The proof of this statement rests on the following

<u>THEOREM 3.1</u>. <u>Under (1.6) through (1.10) and (2.5)</u>

$$\mathcal{L}(S_N/AB-\lambda C(F)AB) \to \mathcal{N}(0,1); \tag{3.13}$$

$$A_N^2 \overset{p}{\to} A^2,\ B_N^2 \overset{p}{\to} B^2. \tag{3.14}$$

<u>PROOF</u>. The proof that $A_N^2 \overset{p}{\to} A^2$ follows precisely on the same line as in lemma 3.1 of Sen (1967b) and the same proof also applies to the stochastic convergence of B_N^2 to B^2; the details are therefore omitted. The proof of (3.13) rests on the following.

<u>LEMMA 3.2</u>. <u>Under the conditions (1.6) through (1.10)</u>

$$E[(S_N-T_N)^2|H_o] \to 0 \text{ as } N\to\infty .$$

<u>PROOF</u>. It follows from (2.8) and (3.5) that

$$E[(S_N-T_N)^2|H_o] = E_N^o\{E_{\mathcal{P}_N}[\tfrac{1}{N}\{\sum_j\sum_k(\mu_j\nu_k-\mu_{Nj}\nu_{Nk})\sum_{i=1}^N Z_{ijk}\}^2]\}, \tag{3.15}$$

where $E_{\mathcal{P}_N}$ and E_N^o denote the expectation over the permutation distribution $\mathcal{P}_N$ (of Z_{ijk}'s given $N_{jo},N_{ok},j,k=1,2,\ldots,$) and the distribution of $N_{jo},N_{ok},j,k=1,2,\ldots,$ respectively. We note that

$$E(Z_{ijk}|\mathcal{P}_N) = E(Z_{ijk}^2|\mathcal{P}_N) = N_{jo}\ N_{ok}/N^2\ , \tag{3.16}$$

$$E(Z_{ijk}\ Z_{i'\ell q}|\mathcal{P}_N) = 0 \text{ if } i=i' \text{ but } (j,k)\neq(\ell,q)) \tag{3.17}$$

$$= N_{jo}N_{ok}(N_{\ell o}-\delta_{j\ell})(N_{oq}-\delta_{kq})/N^2(N-1)^2 \tag{3.18}$$

if $i\neq i'$,

for j, ,k,q=1,2,...; where $\delta_{j\ell}$ and δ_{kq} are the Kronecker deltas. Also, it is easy to verify that under E_N^o, $\{N_{jo}, j=1,2,...\}$ and $\{N_{ok}, k=1,2,...\}$ are stochastically independent, and

$$E_N^o\{N_{jo}\} = P_{jo},\ E_N^o\{N_{ok}\} = P_{ok} \text{ for all } j,k=1,2,\ldots\ . \tag{3.19}$$

Hence, from (3.15) through (3.19), we obtain after some simplifications that

$$E[(S_N-T_N)^2|H_o] = \frac{N}{N-1}[A^2B^2-2E_N^o\{(\sum_j \mu_j\mu_{Nj}\ N_{jo}/N)(\sum_k \nu_k\nu_{Nk}N_{ok}/N)\} + E_N^o(A_N^2\cdot B_N^2)]. \tag{3.20}$$

Now

$$E_N^o(A_N^2B_N^2) = E_N^o(A_N^2)\cdot E_N^o(B_N^2), \tag{3.21}$$

and hence, proceeding as in Sen (1967b, (3.33)), we get that (3.21) converges to A^2B^2 as $N\to\infty$. Also

$$E_N^o(\sum_j \mu_j\mu_{Nj}\frac{N_{jo}}{N})(\sum_k \nu_k\nu_{Nk}\frac{N_{ok}}{N}) = E_N^o(\sum_j \mu_j\mu_{Nj}\frac{N_{jo}}{N})\ E_N^o(\sum_k \nu_k\nu_{Nk}\frac{N_{ok}}{N}), \tag{3.22}$$

and hence, proceeding as in Sen (1967b, (3.34)), we get that (3.22) converges to A^2B^2 as $N\to\infty$. Thus, (3.20) converges to zero as $N\to\infty$. Hence the lemma.

An immediate consequence of lemma 3.2 and theorem 2.1 is the following.

<u>LEMMA 3.3</u>. <u>Under</u> H_o <u>in (1.3)</u>, (T_N, S_N) <u>converges in law to a bivariate normal distribution concentrated on the line</u> $T_N=S_N$.

The rest of the proof of theorem 3.1 follows from lemma 4.2 of Hájek (1962) and our theorem 2.1, lemmas 3.2 and 3.3; for brevity the details are omitted.

Comparing now (2.15) and (3.12) we observe that the permutation test in (3.10) based on the rank order statistic S_N is asymptotically most powerful for H_o: $\beta=0$ against $\beta>0$; a similar result also holds for H_o: $\beta=0$ against $\beta<0$.

4. ASYMPTOTIC EFFICIENCY CONSIDERATIONS

Under the conditions in (1.6) through (1.10), for ungrouped data, the asymptotically most powerful (parametric) test (based on the likelihood ratio test criterion) can be shown to have the asymptotic power (for $\{H_N\}$ specified by (2.5))

$$1 - \Phi(\tau_\alpha - C(F)\gamma), \tag{4.1}$$

where $\Phi(x)$ and τ_α are defined by (2.15) and γ by (1.10). Thus it follows from (2.15) and (3.12) that the loss in asymptotic efficiency due to grouping of data is equal to

$$L(\{I_{jk}\}) = 1 - A^2B^2/\gamma^2, \tag{4.2}$$

where A^2 and B^2 are defined by (2.3). Let us now define

$$L(\{a_o,a_1,\ldots\}) = \sum_j \int_{a_{j-1}}^{a_j} [\ell_1(x)-\mu_j]^2 dG_o(x) / \int_{-\infty}^{\infty} \ell_1^2(x) dG_o(x); \tag{4.3}$$

$$L(\{b_o,b_1,\ldots\}) = \sum_k \int_{b_{k-1}}^{b_k} [\ell_2(x)-\nu_k]^2 dH(x) / \int_{-\infty}^{\infty} \ell_2^2(x) dH(x), \tag{4.4}$$

where $\ell_1(x)$, $\ell_2(x)$ are defined by (1.9) and $\mu_j, \nu_k, j,k=1,2,\ldots,$ by (2.2). Using (1.10), (2.3), (4.2), (4.3) and (4.4), we obtain

$$[1-L(\{I_{jk}\})] = [1-L(\{a_o,a_1,\ldots\})][1-L(\{b_o,b_1,\ldots\})] \tag{4.5}$$

i.e.

$$\begin{aligned} L(\{I_{jk}\}) &= L(\{a_o,a_1,\ldots\})+L(\{b_o,b_1,\ldots\})-L(\{a_o,a_1,\ldots\})L(\{b_o,b_1,\ldots\}) \\ &\le L(\{a_o,a_1,\ldots\}) + L(\{b_o,b_1,\ldots\}) \end{aligned} \tag{4.6}$$

If P_{jo} and $P_{ok}, j,k=1,2,\ldots,$ are all sufficiently small, it follows from (4.3) through (4.6) that the loss in efficiency due to grouping can be made arbitrarily small.

Now, instead of the rank order statistic S_N, we may consider another rank order statistic S_N^* which leads to an asymptotically most powerful test for H_o: $\beta=0$ against $\beta>0$ when the true c.d.f. is $F^*(\underset{\sim}{x})$, $\underset{\sim}{x}\varepsilon R$, i.e.,

$$S_N^* = N^{-\frac{1}{2}} \sum_{i=1}^{N} \sum_{j} \sum_{k} Z_{ijk} \mu_{Nj}^* \nu_{Nk}^* \tag{4.7}$$

where μ_{Nj}^*, ν_{Nk}^*, μ_j^*, ν_k^*, $j,k=1,\ldots$, are defined as in μ_{Nj}, ν_{Nk}, μ_j and ν_k respectively, with the c.d.f. $F(\underset{\sim}{x})$ being replaced by $F^*(\underset{\sim}{x})$. Thus, replacing $F(\underset{\sim}{x})$ by $F^*(\underset{\sim}{x})$ in the definition of $\ell_1(x)$, $\ell_2(x)$, and P_{jk}'s, we obtain analogous expressions which are denoted by $\ell_1^*(x)$, $\ell_2^*(x)$ and P_{jk}^* respectively. We want to study the asymptotic relative efficiency of the test based on S_N^* with respect to the one based on S_N when $F(\underset{\sim}{x})$ is the true c.d.f. For this, let us define

$$\rho_1(\{a_o,a_1,\ldots\}) = \sum_j \mu_j \mu_j^{**} P_{jo} / \{[\sum_j \mu_j^2 P_{jo}][\sum_j (\mu_j^{**})^2 P_{jo}]\}^{\frac{1}{2}}, \tag{4.8}$$

where

$$\mu_j^{**} = \int_{a_{j-1}}^{a_j} \ell_1^*(x)\, dG_o(x)/P_{jo}, \quad j=1,2,\ldots\ ; \tag{4.9}$$

$$\rho_2(\{b_o,b_1,\ldots\}) = \sum_k \nu_k \nu_k^{**} P_{ok} / \{[\sum_k \nu_k^2 P_{ok}][\sum_k (\nu_k^{**})^2 P_{ok}]\}^{\frac{1}{2}} \tag{4.10}$$

where

$$\nu_k^{**} = \int_{b_{k-1}}^{b_k} \ell_2^*(x)\, dH(x)/P_{ok} \text{ for } k=1,2,\ldots\ . \tag{4.11}$$

Thus, proceeding as in theorem 4.1 of Sen (1967b) and generalizing it along the lines of theorem 3.1 of this paper, we obtain after some simplifications that the asymptotic relative efficiency of the test based on S_N^* with respect to the one based on S_N, when $F(\underset{\sim}{x})$ is the true c.d.f., is given by

$$e_{\{S_N^*\}\{S_N\}}(F) = [\rho_1(\{a_o,a_1,\ldots\})\rho_2(\{b_o,b_1,\ldots\})]^2 \ . \tag{4.12}$$

If $F(\underset{\sim}{x})$ and $F^*(\underset{\sim}{x})$ are specified, (4.12) may be evaluated for specified $\{a_o,a_1,\ldots\}$ and $\{b_o,b_1,\ldots\}$.

5. SOME ILLUSTRATIVE EXAMPLES

Let us first consider the case when $F(\underset{\sim}{x})$ is a bivariate normal cdf. In this case, from (1.9), we have on taking ρ, the correlation coefficient as a measure of association (i.e., $\rho=\beta$)

$$c(F) = 1,\ \ell_1(x) = \frac{x_1-E(X_1)}{\sqrt{V(X_1)}} \text{ and } \ell_2(x) = \frac{X_2-E(X_2)}{\sqrt{V(X_2)}} \quad , \tag{5.1}$$

Thus, as in section 2, on denoting by $\Phi(x)$ the standard normal c.d.f., we have from (3.3) and (3.4)

$$\mu_{Nj} = \int_{\Phi^{-1}(G_{N,j-1})}^{\Phi^{-1}(G_{N,j})} x d\Phi(x)/[G_{N,j}-G_{N,j-1}],\ j=1,\ldots; \tag{5.2}$$

$$\nu_{Nk} = \int_{\Phi^{-1}(H_{N,k-1})}^{\Phi^{-1}(H_{N,k})} x d\Phi(x)/[H_{N,k}-H_{N,k-1}],\ k=1,2,\ldots\ . \tag{5.3}$$

On denoting by $\phi(x)$ the standard normal density function, we may write

$$\mu_{Nj} = [\phi(\Phi^{-1}(G_{N,j-1})) - \phi(\Phi^{-1}(G_{N,j}))]/[G_{N,j}-G_{N,j-1}], \tag{5.4}$$

$$\nu_{Nk} = [\phi(\Phi^{-1}(H_{N,k-1})) - \phi(\Phi^{-1}(H_{N,k}))]/[H_{N,k}-H_{N,k-1}] \tag{5.5}$$

for all $j,k=1,2,\ldots$. The corresponding S_N, defined by (4.7), may be termed the <u>normal scores statistic for doubly-grouped data</u>. The test based on this statistic is therefore asymptotically most efficient for testing H_o: $\rho=0$ against $\rho>0$ (or <0)

when $F(\underset{\sim}{x})$ is a bivariate normal c.d.f.

A second example may be as follows. Suppose that the second variate X_{2i} has marginally a normal distribution while the conditional distribution of X_{1i} given X_{2i} is logistic with a location parameter $\alpha+\beta X_{2i}$. In this case, it is easily seen that for the optimal S_N, ν_{Nk}, k=1,2,..., are defined by (5.3) and (5.5), while

$$\mu_{Nj} = (1/2)(G_{N,j-1} + G_{N,j}) \text{ for } j=1,2,\ldots.. \tag{5.6}$$

Many other examples may be considered.

REFERENCES

HÁJEK, J. (1962): Asymptotically most powerful rank order tests. Ann. Math. Statist. 33, 1124-1147.

RAO, C. R. (1965): Linear Statistical Inference and its Applications. John Wiley: New York.

SEN, P. K. (1967a): On a class of permutation tests for stochastic independence, I. Sankhya, Ser. A. 29, in press.

SEN, P. K. (1967b): Asymptotically most powerful rank order tests for grouped data. Ann. Math. Statist. 38, No. 4 (in press).