ASYMPTOTICALLY MOST POWERFUL RANK ORDER TESTS FOR GROUPED DATA[1]

By Pranab Kumar Sen

University of North Carolina, Chapel Hill, and University of Calcutta

0. Summary. The object of the present investigation is to extend the findings of Hájek [7] on asymptotically most powerful rank order tests (AMPROT) to grouped data where the underlying distributions are essentially continuous but the observable random variables correspond to a finite or countable set of contiguous class intervals. In this context, the two sample problem for grouped data is considered and various efficiency results are also studied.

1. Introduction. Let us consider a sequence of random vectors $\mathbf{X}_\nu = (X_{\nu 1}, \cdots, X_{\nu N_\nu})$ consisting of N_ν independent random variables, where $X_{\nu i}$ has a continuous cumulative distribution function (cdf) $F_{\nu i}(x)$, for $i = 1, \cdots, N_\nu$, $1 \leqq \nu < \infty$. As in Hájek [7], we consider the model

$$(1.1) \qquad F_{\nu i}(x) = F(\sigma^{-1}[x - \alpha - \beta c_{\nu i}]), \qquad i = 1, \cdots, N_\nu, 1 \leqq \nu < \infty,$$

where α, β and $\sigma(>0)$ are real parameters, $(c_{\nu 1}, \cdots, c_{\nu N_\nu})$ are known quantities, concerning which we make the following assumptions:

$$(1.2) \qquad \textstyle\sum_{i=1}^{N_\nu} c_{\nu i} = 0, \qquad \sum_{i=1}^{N_\nu} c_{\nu i}^2 = C_\nu^2, \qquad 0 < \sup_\nu C_\nu^2 < \infty;$$

$$(1.3) \qquad \operatorname{Max}_{1 \leqq i \leqq N_\nu} c_{\nu i}^2 / C_\nu^2 = o(1).$$

Hájek [7] has considered the class of cdf's for which the square root of the probability density function possesses a quadratically integrable derivative i.e.,

$$(1.4) \qquad \textstyle\int_{-\infty}^{\infty} [f'(x)/f(x)]^2 f(x)\, dx = A^2(F) < \infty,$$

where $f(x) = dF(x)/dx$ and $f'(x) = df(x)/dx$. Throughout this paper, we shall also stick to the assumptions (1.1) through (1.4). In Hájek's case, $\mathbf{X}_\nu$ is observable, while in our case, we have a finite or countable set of class intervals

$$(1.5) \qquad I_j : a_j < x \leqq a_{j+1}, \qquad j = 0, 1, \cdots, \infty \quad \text{(without any loss of generality)},$$

[where $-\infty = a_0 < a_1 < a_2 < \cdots < \infty$ is any (finite or countable) set of ordered points on the real line $(-\infty, \infty)$,] and the observable stochastic vector is $\mathbf{X}_\nu^* = (X_{\nu 1}^*, \cdots, X_{\nu N_\nu}^*)$, where

$$(1.6) \qquad X_{\nu i}^* = \textstyle\sum_{j=0}^{\infty} I_j Z_{ij},$$

and

Received 9 September 1966; revised 25 January 1967.

[1] Work supported by the Army Research Office, Durham, Grant DA-31-124-ARO-D-G432.

$$(1.7)\quad Z_{ij} = 1, \quad \text{if} \quad X_{\nu i} \,\varepsilon\, I_j, \\ = 0, \quad \text{otherwise}, \quad \text{for all } i = 1, \cdots, N_\nu, \quad j = 0, \cdots, \infty.$$

Thus, having observed $\mathbf{X}_\nu^*$, we want to test the null hypothesis

$$(1.8)\qquad H_0 : \beta = 0 \quad \text{i.e.,} \quad \text{no regression,}$$

against the set of alternatives that $\beta > 0$.

It may be noted that in actual practice, even if the parent cdf's are continuous, the process of data collection mostly introduces such a set of class intervals on which the data are recorded. This results in so called grouped data, where the usual nonparametric methods (for continuous variables) are not strictly applicable. The object of the present investigation is to consider some permutationally distribution-free tests for regression for grouped data and by a generalization of Hájek's [5] ideas, to show that these are AMPROT for the same problem. In this context, the two sample problem for grouped data is also considered and the allied efficiency results are studied. It may be noted that for the two sample case, the problem of finding AMPROT for grouped data censored by sample percentiles is asymptotically equivalent to the problem of finding AMPROT for the case considered here (cf. Chernoff et al. [2], Gastwirth [4], [5], Sarndal [15] and Kulldorf [11]). In fact, the derivations of the results in [4], [5] are comparatively shorter and they relate particularly to the one sample and two sample location problems. Our results are not only true for the more general regression problem but also the desired asymptotic normality is proved for "nearby" alternatives. Thus, these may also be regarded as generalizations of the earlier works referred to.

2. Asymptotically most powerful parametric test. This test is considered in brief, as it will be essentially required in the sequel. Let us define

$$(2.1)\qquad F_j = F([a_j - \alpha]/\sigma), \quad P_j = F_{j+1} - F_j \quad \text{for } j = 0, 1, \cdots, \infty;$$

$$(2.2)\qquad \Delta_j = [f(F^{-1}(F_j)) - f(F^{-1}(F_{j+1}))]/P_j, \qquad j = 0, \cdots, \infty,$$

$$(2.3)\qquad A^2(F, \{I_j\}) = \textstyle\sum_{j=0}^{\infty} \Delta_j^2 P_j.$$

Now, Δ_j can be written as

$$(2.4)\qquad \textstyle\int_{F_j}^{F_{j+1}} \phi(u)\, du / \int_{F_j}^{F_{j+1}} du,$$

where

$$\phi(u) = -f'(F^{-1}(u))/f(F^{-1}(u)), \qquad 0 < u < 1,$$

and hence,

$$(2.5)\qquad A(F, \{I_j\}) = \textstyle\sum_{j=0}^{\infty} \{\int_{F_j}^{F_{j+1}} \phi(u)\, du\}^2 / \int_{F_j}^{F_{j+1}} du \\ \leqq \textstyle\sum_{j=0}^{\infty} \int_{F_j}^{F_{j+1}} \phi^2(u)\, du = A^2(F),$$

uniformly in $\{I_j\}$ i.e., for all possible $-\infty = a_0 < a_1 < a_2 < \cdots < \infty$. Again under (1.2) and (1.3), we have for any real and finite β

$$F_{\nu i}(a_{j+1}) - F_{\nu i}(a_j) = P_j\{1 + (\beta/\sigma)c_{\nu i}\Delta_j + o(\beta)\}, \tag{2.6}$$

uniformly in $j(= 0, \cdots, \infty)$. Thus, for any real and finite β, the likelihood function (under (1.2) and (1.3)) is

$$L(\mathbf{X}_\nu^* \mid \beta) = \prod_{i=1}^{N_\nu} \{ \sum_{j=0}^{\infty} Z_{ij}P_j[1 + (\beta/\sigma)c_{\nu i}\Delta_j + o(\beta)]\}. \tag{2.7}$$

Consequently,

$$L(\mathbf{X}_\nu^* \mid \beta)/L(X_\nu^* \mid \beta = 0) = 1 + (\beta/\sigma)\sum_{i=1}^{N_\nu} c_{\nu i}\sum_{j=0}^{\infty} \Delta_j Z_{ij} + o(\beta). \tag{2.8}$$

Let us denote by T_ν

$$T_\nu = \sum_{i=1}^{N_\nu} c_{\nu i}\sum_{j=0}^{\infty} \Delta_j Z_{ij}. \tag{2.9}$$

Then from Neyman-Pearson's fundamental lemma (cf. Lehmann [12], p. 65), it readily follows that for testing $H_0 : \beta = 0$ against $\beta > 0$, the asymptotically most powerful test function is

$$\psi_1(X_\nu^*) = \begin{cases} 1, & \text{if } T_\nu > T_{\nu,\epsilon}, \\ \gamma_\epsilon, & \text{if } T_\nu = T_{\nu,\epsilon}, \\ 0, & \text{if } T_\nu < T_{\nu,\epsilon}, \end{cases} \tag{2.10}$$

where $T_{\nu,\epsilon}$ and γ_ϵ are so chosen that $E\{\psi_1(\mathbf{X}_\nu^*) \mid H_0\} = \epsilon$, $0 < \epsilon < 1$, ϵ being the desired level of significance of the test.

[It may be noted that in actual practice both α and σ in (1.1) are mostly unknown, and as a result, Δ_j's are also so. However, as in (3.1)–(3.4) of Hájek [7], we may estimate α and σ and work with the estimated Δ_j's.]

Now $\sum_{j=0}^{\infty} Z_{ij}\Delta_j$, $i = 1, \cdots, N_\nu$, are independent random variables, and (1.3) guarantees the condition for the central limit theorem to be satisfied by the coefficients. Consequently, we get on using the classical central limit theorem and avoiding the details of derivation, the following:

THEOREM 2.1. *Under* (1.1) *through* (1.4) *and for any finite* β

$$\mathcal{L}([T_\nu - (\beta/\sigma)C_\nu^2A^2(F, \{I_j\})]/C_\nu A(F, \{I_j\})) \to N(0, 1),$$

where C_ν^2 *and* $A^2(F, \{I_j\})$ *are defined in* (1.2) *and* (2.3), *and* $\mathcal{L}(Z) \to N(0, 1)$ *indicates that* Z *converges in law to a normal distribution with zero mean and unit variance.*

Thus, from (2.10) and Theorem 2.1 we get that the asymptotic power of the test (2.10) is given by

$$1 - \Phi(\tau_\epsilon - (\beta/\sigma)C_\nu A(F, \{I_j\})), \tag{2.11}$$

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

3. Asymptotically most powerful rank order test. Let us define

$$\sum_{i=1}^{N_\nu} Z_{ij} = N_{\nu j}, \quad \text{for } j = 0, \cdots, \infty, \quad \text{so that } N_\nu = \sum_{j=0}^{\infty} N_{\nu j}; \tag{3.1}$$

$$F_{N_\nu,0} = 0, \qquad F_{N_\nu,j+1} = \sum_{l=0}^{j} N_{\nu l}/N_\nu \quad \text{for } j = 0, 1, \cdots, \infty. \tag{3.2}$$

If now $N_{\nu j} > 0$, we define

$$(3.3)\qquad \hat{\Delta}_{\nu j} = [f(F^{-1}(F_{N_\nu,j})) - f(F^{-1}(F_{N_\nu,j+1}))]/[F_{N_\nu,j+1} - F_{N_\nu,j}] = \int_{F_{N_\nu,j}}^{F_{N_\nu,j+1}} \phi(u)\, du \Big/ \int_{F_{N_\nu,j}}^{F_{N_\nu,j+1}} du,$$

where F^{-1} is the inverse of $F(x)$ and $\phi(u)$ is defined by (2.4). If $N_{\nu,j} = 0$, we conventionally let

$$(3.4)\qquad \hat{\Delta}_{\nu j} = \phi(F_{N_\nu,j}).$$

Our proposed test-statistic is then

$$(3.5)\qquad S_\nu = \sum_{i=1}^{N_\nu} c_{\nu i} \sum_{j=0}^{\infty} \hat{\Delta}_{\nu j} Z_{ij},$$

and we shall see later on that S_ν provides an AMPROT for the hypothesis $\beta = 0$ against $\beta > 0$.

3.1. *Null distribution of* S_ν. Since, we are dealing with grouped data, even under H_0 in (1.8), the distribution of S_ν will depend on the unknown Δ_j ($j = 0, \cdots, \infty$). However, under a very simple permutation model, S_ν will provide a distribution-free test. Now, under H_0 in (1.8), $X_{\nu i}^*$ ($i = 1, \cdots, N_\nu$) are independent and identically distributed random variables (iidrv), and hence $\mathbf{X}_\nu^*$ has a joint distribution which remains invariant under any permutation of its N_ν arguments. Thus, in the N_ν-dimensional real space R^{N_ν}, we have a set of $N_\nu!$ permutationally equiprobable points. The permutational (conditional) probability measure defined on this set is denoted by $\mathcal{P}_\nu$. Hence, under $\mathcal{P}_\nu$, all the $N_\nu!$ equally likely realizations have the common probability $1/N_\nu!$. Now, $N_{\nu j}$ in (3.1) and $\hat{\Delta}_{\nu j}$ in (3.3) are unaffected by the permutations of the coordinates of $\mathbf{X}_\nu^*$ i.e., they are permutation-invariant. Hence, by some simple reasonings it follows that

$$(3.6)\qquad E_{\mathcal{P}_\nu}\{Z_{ij}\} = N_{\nu j}/N_\nu \quad \text{for all } i = 1, \cdots, N_\nu, j = 0, \cdots, \infty;$$

$$(3.7)\qquad E_{\mathcal{P}_\nu}\{Z_{ij} \cdot Z_{ij'}\} = 0 \quad \text{for all } i = 1, \cdots, N_\nu, j \neq j' = 0, \cdots, \infty;$$

$$(3.8)\qquad E_{\mathcal{P}_\nu}\{Z_{ij} \cdot Z_{i'j'}\} = N_{\nu j}(N_{\nu j'} - \delta_{jj'})/N_\nu(N_\nu - 1)$$

for all $i \neq i' = 1, \cdots, N_\nu$ and $j, j' = 0, \cdots, \infty$; where $\delta_{jj'}$ is the usual Kronecker delta. By analogy with (2.3), we define

$$(3.9)\qquad A^2(F_{N_\nu}, \{I_j\}) = \sum_{j=0}^{\infty} \hat{\Delta}_{\nu j} N_{\nu j}^2 / N_\nu,$$

and it is easy to see that (3.9) also satisfies (2.5), uniformly in $\{I_j\}$ and for all $\{F_{N_\nu}\}$. From (3.5) through (3.9), we have

$$(3.10)\qquad E_{\mathcal{P}_\nu}\{S_\nu\} = 0 \quad \text{and} \quad E_{\mathcal{P}_\nu}\{S_\nu^2\} = [N_\nu/(N_\nu - 1)]C_\nu^2 A^2(F_{N_\nu}, \{I_j\}),$$

where C_ν^2 is defined by (1.2). Now, under $\mathcal{P}_\nu$, S_ν can only assume $N_\nu!$ possible values (actually, there are $N_\nu!/\prod_{j=0}^{\infty} N_{\nu j}!$ distinct equally likely permutations of $\mathbf{X}_\nu^*$), and hence, for small values of N_ν, the upper tail of the permutation distribution of S_ν, can be evaluated to formulate the test function:

$$\begin{aligned} \psi_2(\mathbf{X}_\nu^*) &= 1, &&\text{if } S_\nu > S_{\nu,\epsilon}, \\ &= \delta_\epsilon, &&\text{if } S_\nu = S_{\nu,\epsilon}, \\ &= 0, &&\text{if } S_\nu < S_{\nu,\epsilon}, \end{aligned} \tag{3.11}$$

where $S_{\nu,\epsilon}$ and δ_ϵ are so chosen that $E\{\psi_2(\mathbf{X}_\nu^*) \mid \mathcal{P}_\nu\} = \epsilon$, the level of significance. This implies that $E\{\psi_2(\mathbf{X}_\nu^*) \mid H_0\} = \epsilon$, i.e., $\psi_2(\mathbf{X}_\nu^*)$ is a similar size ϵ test.

Let us now define

$$W_{\nu i} = \textstyle\sum_{j=0}^{\infty} \hat{\Delta}_{\nu j} Z_{ij} \qquad \text{for } i = 1, \cdots, N_\nu. \tag{3.12}$$

Under $\mathcal{P}_\nu$, $\hat{\Delta}_{\nu j}$'s are all invariant while Z_{ij}'s are stochastic. Thus, it follows that $N_{\nu j}$ of $W_{\nu i}$'s are equal to $\hat{\Delta}_{\nu j}$, for $j = 0, \cdots, \infty$. We now impose the nondegeneracy condition on $F(x)$ as

$$\sup_j \,[F_{j+1} - F_j] < 1 \qquad \text{with probability one.} \tag{3.13}$$

Then, writing S_ν equivalently as $\sum_{i=1}^{N_\nu} c_{\nu i} W_{\nu i}$, it follows from the well-known permutational central limit theorem by Wald-Wolfowitz-Noether-Hoeffding-Hájek (cf. [4]) that under (1.2), (1.3), (3.9) and (3.13)

$$\mathcal{L}_{\mathcal{P}_\nu}(S_\nu / C_\nu A(F_{N_\nu}, \{I_j\})) \to N(0, 1), \qquad \text{in probability.} \tag{3.14}$$

Consequently, from (3.11) and (3.14), we have

$$S_{\nu,\epsilon} \to \tau_\epsilon C_\nu A(F_{N_\nu}, \{I_j\}), \qquad \delta_\epsilon \to 0, \qquad \text{in probability,} \tag{3.15}$$

where τ_ϵ is defined by (2.11). [It may be noted that $\mathcal{P}_\nu$ being a conditional probability measure, (given $(N_{\nu j}, j = 0, \cdots, \infty)$,) (3.14) and (3.15) hold in probability i.e., for almost all $(N_{\nu j}, j = 0, \cdots, \infty)$.]

3.2 *Asymptotic optimality of* $\psi_2(\mathbf{X}_\nu^*)$. The main contention of this paper is to establish the asymptotic equivalence of $\psi_1(\mathbf{X}_\nu^*)$ and $\psi_2(\mathbf{X}_\nu^*)$, in (2.10) and (3.11), respectively. For this, let us first consider the following lemmas:

LEMMA 3.1. *Under* (1.1) *through* (1.4) *and for any real and finite* β, $A^2(F_{N_\nu}, \{I_j\})$ *converges in probability to* $A^2(F, \{I_j\})$, *uniformily in* $\{I_j\}$.

PROOF. We shall prove the lemma only for $\beta = 0$ as the rest of the proof will follow by the contiguity argument of Hájek [7]. Let us select a sequence of real and positive numbers $\{\eta_\nu\}$ in such a manner that

$$\lim_{\nu=\infty} \eta_\nu = 0 \quad \text{but} \quad \lim_{\nu=\infty} N_\nu^{\frac{1}{2}} \eta_\nu = \infty. \tag{3.16}$$

For any given N_ν, the set of class intervals $\{I_j\}$ is divided into two subsets $G_\nu^{(1)}$ and $G_\nu^{(2)}$, where

$$G_\nu^{(1)} = \{I_j : P_j \geqq \eta_\nu\}, \qquad G_\nu^{(2)} = \{I_j : P_j < \eta_\nu\}. \tag{3.17}$$

Let $F_{N_\nu}(x) = N_\nu^{-1}$ [Number of $X_{\nu i} \leqq x$] be the empirical cdf of $\mathbf{X}_\nu$. Then by making use of the fact that $\sup_j |F_{N_\nu, j} - F_j| \leqq \sup_x |F_{N_\nu}(x) - F(x)|$ and the well-known result regarding Kolmogorov-Smirnoff statistic, viz., $\sup_x \{N_\nu^{\frac{1}{2}} |F_{N_\nu}(x) - F(x)|\}$ is bounded in probability, we obtain that

$$\sup_j \{N_\nu^{\frac{1}{2}} |F_{N_\nu, j} - F_j|\} \quad \text{is bounded in probability.} \tag{3.18}$$

Further, it is easily seen that

$$(3.19)\qquad \begin{aligned} \sup_j |f(F^{-1}(F_{N_\nu,j})) - f(F^{-1}(F_j))| &= \sup_j |\textstyle\int_{F_{N_\nu,j}}^{F_j} \phi(u)\,du| \\ &\leqq \sup_j |\textstyle\int_{F_{N_\nu,j}}^{F_j} \phi^2(u)\,du|^{\frac{1}{2}} |F_{N_\nu,j} - F_j|^{\frac{1}{2}} \\ &\leqq A^2(F) \sup_j |F_{N_\nu,j} - F_j|^{\frac{1}{2}}. \end{aligned}$$

Hence, from (2.2), (3.3), (3.18) and (3.19), we have for all $I_j \varepsilon G_\nu^{(1)}$,

$$(3.20)\qquad \hat{\Delta}_{\nu j}^2 N_{\nu j}/N_\nu = P_j[\Delta_j + R_{\nu j}^{(1)}]^2[1 + R_{\nu j}^{(2)}],$$

where

$$(3.21)\qquad \sup_j |R_{\nu j}^{(2)}| = O_p([N_\nu^{\frac{1}{2}}\eta_\nu]^{-1}), \qquad \sup_j |R_{\nu j}^{(1)}| = O_p(N_\nu^{-\frac{1}{4}}).$$

Consequently, from (3.16), (3.20), (3.21) and some simple algebraic manipulations, we have

$$(3.22)\qquad \textstyle\sum_{G_\nu^{(1)}} \hat{\Delta}_{\nu j}^2 N_{\nu j}/N_\nu = \sum_{G_\nu^{(1)}} \Delta_j^2 P_j + o_p(1).$$

Again, using (2.4) and (3.3), we have

$$(3.23)\qquad \Delta_j^2 P_j = \textstyle\int_{F_j}^{F_{j+1}} \phi^2(u)\,du - \int_{F_j}^{F_{j+1}} [\phi(u) - \Delta_j]^2\,du;$$

$$(3.24)\qquad \hat{\Delta}_{\nu j}^2 N_{\nu j}/N_\nu = \textstyle\int_{F_{N_\nu,j}}^{F_{N_\nu,j+1}} \phi^2(u)\,du - \int_{F_{N_\nu,j}}^{F_{N_\nu,j+1}} [\phi(u) - \hat{\Delta}_{\nu j}]^2\,du,$$

where Δ_j and $\hat{\Delta}_{\nu j}$ are also the conditional means of $\phi(u)$ on $[F_j, F_{j+1}]$ and $[F_{N_\nu,j}, F_{N_\nu,j+1}]$, respectively. Now, using (3.18) and some routine analysis, it is easily seen that

$$(3.25)\qquad \textstyle\sum_{G^{(2)}} \int_{F_{N_\nu,j}}^{F_{N_\nu,j+1}} \phi^2(u)\,du = \sum_{G_\nu^{(2)}} \int_{F_j}^{F_{j+1}} \phi^2(u)\,du + o_p(1),$$

and we shall show that the second integral on the right hand side of (3.23) converges to zero (as $\nu \to \infty$) and the second one on the right hand side of (3.24) converges to zero, in probability. For an unessential simplification of this proof we shall assume (as in Lemma 2.2 of [6]) that $\phi(u)$ is $\uparrow$ in u. It follows from the existence of (1.4) that if $P_j < \eta_\nu$ for all $I_j \varepsilon G_\nu^{(2)}$, then

$$(3.26)\qquad \lim_{\nu=\infty} \textstyle\sum_{G_\nu^{(2)}} \phi^2(F_j)P_j = \lim_{\nu=\infty} \sum_{G_\nu^{(2)}} \phi^2(F_{j+1})P_j = \int_{G_\nu^{(2)}} \phi^2(u)\,du.$$

Now for all $u \varepsilon [F_j, F_{j+1}]$

$$(3.27)\qquad [\phi(u) - \Delta_j]^2 \leqq [\phi(F_{j+1}) - \phi(F_j)]^2 \leqq \phi^2(F_{j+1}) - \phi^2(F_j).$$

Therefore

$$(3.28)\qquad \textstyle\sum_{G_\nu^{(2)}} \int_{F_j}^{F_{j+1}} [\phi(u) - \Delta_j]^2\,du \leqq \sum_{G_\nu^{(2)}} [\phi^2(F_{j+1})P_j - \phi^2(F_j)P_j] \to 0$$

as $\nu \to \infty$ (by (3.26)). Similarly, from (3.16), (3.18) we have for all $I_j \varepsilon G_{\nu(2)}$, $F_{N_\nu,j+1} - F_{N_\nu,j} < \eta_\nu + O_p(N_\nu^{-\frac{1}{2}}) < 2\eta_\nu$ for adequately large N_ν. Consequently, as in (3.26) and (3.27), we get that $\sum \int_{F_{N_\nu,j}}^{F_{N_\nu,j+1}} [\phi(u) - \hat{\Delta}_{\nu j}]^2\,du \to^P 0$. Hence,

$$(3.29)\qquad \textstyle\sum_{G_\nu^{(2)}} \hat{\Delta}_{\nu j}^2 N_{\nu j}/N_\nu = \sum_{G_\nu^{(1)}} \Delta_j^2 P_j + o_p(1).$$

Hence, the lemma follows from (3.22) and (3.29).

LEMMA 3.2 *Under* H_0 *in* (1.8), $E\{S_\nu - T_\nu\}^2 \to 0$ *as* $\nu \to \infty$, *where* S_ν *and* T_ν *are defined in* (3.7) and (2.9), *respectively.*

PROOF.

$$
\begin{aligned}
E\{S_\nu - T_\nu\}^2 &= E_\nu\{E_{\mathcal{P}_\nu}[S_\nu - T_\nu]^2\} \\
&= E_\nu\{E_{\mathcal{P}_\nu}[\textstyle\sum_{j=0}^{\infty} \sum_{i=1}^{N_\nu} c_{\nu i} Z_{ij}(\hat{\Delta}_{\nu j} - \Delta_j)]^2\} \\
&= [N_\nu/(N_\nu - 1)] \textstyle\sum_{i=1}^{N_\nu} c_{\nu i}^2 E_\nu[\sum_{j=0}^{\infty} (\hat{\Delta}_{\nu j} - \Delta_j)^2 N_{\nu j}/N_\nu \\
&\qquad - \{\textstyle\sum_{j=0}^{\infty} (\hat{\Delta}_{\nu j} - \Delta_j) N_{\nu j}/N_\nu\}^2],
\end{aligned} \tag{3.30}
$$

where $E_{\mathcal{P}_\nu}$ and E_ν stand for the expectation over the permutation distribution and the distribution of the order statistic associated with $\mathbf{X}_\nu^*$, respectively. Thus, it follows from (3.30), that we are only to show that

$$E_\nu\{\textstyle\sum_{j=0}^{\infty}(\hat{\Delta}_{\nu j} - \Delta_j)^2 N_{\nu j}/N_\nu\} \to 0 \quad \text{as} \quad \nu \to \infty. \tag{3.31}$$

We rewrite the expression in (3.31) as

$$E_\nu\{\textstyle\sum_{j=0}^{\infty} \hat{\Delta}_{\nu j}^2 N_{\nu j}/N_\nu\} - 2\sum_{j=0}^{\infty} \Delta_j E_\nu\{\hat{\Delta}_{\nu j} N_{\nu j}/N_\nu\} + \sum_{j=0}^{\infty} \Delta_j^2 P_j. \tag{3.32}$$

Now, $\sum_{j=0}^{\infty} \hat{\Delta}_{\nu j}^2 N_{\nu j}/N_\nu$ is just $A^2(F_{N_\nu}, \{I_j\})$, defined by (3.9), and hence, it is bounded by $A^2(F)$, defined by (1.4), uniformly in all $(N_{\nu j}, j = 0, 1, \cdots, \infty)$. Further, by Lemma 3.1, it converges to $\sum_{j=0}^{\infty} \Delta_j^2 P_j$, in probability. Since, for bounded valued random variables convergence in probability to a constant implies convergence of the expectation to the same constant, we readily obtain

$$E_\nu\{\textstyle\sum_{j=0}^{\infty} \hat{\Delta}_{\nu j}^2 N_{\nu j}/N_\nu\} \to \sum_{j=0}^{\infty} \Delta_j^2 P_j \quad \text{as} \quad \nu \to \infty. \tag{3.33}$$

Similarly, it follows that

$$E_\nu\{\textstyle\sum_{j=0}^{\infty} \Delta_j \hat{\Delta}_{\nu j} N_{\nu j}/N_\nu\} \to \sum_{j=0}^{\infty} \Delta_j^2 P_j \quad \text{as} \quad \nu \to \infty. \tag{3.34}$$

(3.32), (3.33) and (3.34) imply (3.31), which in conjunction with (3.30) proves the lemma.

LEMMA 3.3. *Under* H_0 *in* (1.8), (T_ν, S_ν) *converges in law to a bivariate normal distribution which degenerates on the line* $T_\nu = S_\nu$.

PROOF. By virtue of Lemma 3.2, any linear function $aS_\nu + bT_\nu$ converges in mean square to $(a + b)T_\nu$, and hence, from Theorem 2.1, has (under H_0) an asymptotic normal distribution with mean zero and variance $(a + b)^2 C_\nu^2 A^2(F, \{I_j\})$. The rest of the proof follows from Theorem 2.1, Lemma 3.1 and Lemma 3.2, and hence is omitted.

THEOREM 3.4. *Under the sequence of alternatives in* (1.1) *through* (1.4) *with a real and finite* β

$$\mathcal{L}([S_\nu - (\beta/\sigma)C_\nu^2 A^2(F, \{I_j\})]/C_\nu A(F, \{I_j\})) \to N(0, 1).$$

The proof is an immediate consequence of Lemma 4.2 of Hájek [7] and our Lemma 3.3, and hence is not reproduced again.

From Theorem 3.4, we get that the test $\psi_2(\mathbf{X}_\nu^*)$ in (3.11) has asymptotically the power function

$$(3.35) \qquad 1 - \Phi(\tau_\epsilon - (\beta/\sigma)C_\nu A(F, \{I_j\})),$$

which agrees with (2.11). Thus, S_ν provides the AMPROT for $H_0 : \beta = 0$ against $\beta > 0$.

This result also applies in particular to the two sample location problem, where

$$(3.36) \qquad c_{\nu i} = \delta_\nu/m_\nu^{\frac{1}{2}} \qquad \text{for } i = 1, \cdots, m_\nu$$
$$= -m_\nu^{\frac{1}{2}}\delta_\nu/n_\nu \qquad \text{for } i = m_\nu + 1, \cdots, N_\nu, m_\nu + n_\nu = N_\nu,$$

where δ_ν is real. Further, the results derived here can also be extended to the problems of symmetry and scalar alternatives. For that one will have to work with Capon's [1] technique and use his $\phi(u)$ (defined by (iv) on p. 89 of [1]) instead of the $\phi(u)$ in (1.4). The rest of the procedure will be very similar to the one considered here, and hence is omitted. Finally, the impact of these findings on AMPROT for truncated/censored two sample problem will be considered in the next section.

4. Asymptotic efficiency. Suppose now for the AMPROT we work with the assumed density function $f(x)$ instead of the true density function $g(x)$ $(= G'(x))$, where

$$(4.1) \qquad A^2(G) = \int_{-\infty}^{\infty} [g'(x)/g(x)]^2 \, dG(x) < \infty.$$

We define G_j and P_j^* as in (2.1) with F replaced by $G(x)$, and $F_{N_\nu,j}$ as in (3.2). Also, we let

$$(4.2) \qquad \phi(u) = f'(F^{-1}(u))/f(F^{-1}(u)),$$
$$\phi^*(u) = g'(G^{-1}(u))/g(G^{-1}(u)); \qquad 0 < u < 1;$$
$$(4.3) \qquad \Delta_j^* = \int_{G_j}^{G_{j+1}} \phi^*(u) \, du/P_j^*,$$
$$\Delta_j^{**} = \int_{G_j}^{G_{j+1}} \phi(u) \, du/P_j^*;$$
$$(4.4) \qquad A^2(G, \{I_j\}) = \sum_{j=0}^{\infty} \Delta_j^{*2} P_j^*,$$
$$B^2(F, \{I_j\}) = \sum_{j=0}^{\infty} (\Delta_j^{**})^2 P_j^*;$$
$$(4.5) \qquad C(F, G, \{I_j\}) = \sum_{j=0}^{\infty} \Delta_j^* \Delta_j^{**} P_j^*;$$
$$(4.6) \qquad \rho(\{I_j\}) = C(F, G, \{I_j\})/[A(G, \{I_j\}) \cdot B(F, \{I_j\})].$$

THEOREM 4.1. *Under* (1.1)–(1.4) *and* (4.1), *the asymptotic power of the test* (3.11) *is equal to*

$$(4.7) \qquad 1 - \Phi(\tau_\epsilon - \rho(\{I_j\})(\beta/\sigma)C_\nu A(G, \{I_j\})).$$

PROOF. Defining

$$(4.8) \qquad T_\nu^* = \sum_{i=1}^{N_\nu} c_{\nu i} \sum_{j=0}^{\infty} \Delta_j^* Z_{ij}, \qquad T_\nu^{**} = \sum_{i=1}^{N_\nu} c_{\nu i} \sum_{j=0}^{\infty} \Delta_j^{**} Z_{ij},$$

and following the same approach as in Sections 2 and 3, we get that

(4.9) (i) $E\{[S_\nu - T_\nu^{**}]^2 \mid H_0\} \to 0$ as $\nu \to \infty$,

(4.10) (ii) $\mathcal{L}([T_\nu^* - (\beta/\sigma)C_\nu^2 A^2(G, \{I_j\})]/C_\nu A(G, \{I_j\})) \to N(0, 1)$,

(4.11) (iii) $\mathcal{L}([T_\nu^{**}/C_\nu B(F, \{I_j\})] \mid H_0) \to N(0, 1)$,

(4.12) (iv) $\mathcal{L}(T_\nu^*, T_\nu^{**} \mid H_0)$ tends to a bivariate normal distribution with a correlation coefficient $\rho(\{I_j\})$, given by (4.6).

The rest of the proof follows directly from Lemma 4.2 of Hájek [7] and (4.9). Hence, the theorem.

As in Hájek [7], we can interpret $[\rho(\{I_j\})]^2$ as the efficiency factor of (3.11) with respect to the asymptotically most powerful parametric test; the interpretation for the two sample problem again being the ratio of the sample sizes needed to attain the same power.

REMARK 1. The loss in efficiency due to grouping of data, as obtained from our Theorem 3.4 and Theorem 1.1 of Hájek [7], is equal to

(4.13) $$1 - A^2(F, \{I_j\})/A^2(F) = \sum_{j=0}^{\infty} \int_{F_j}^{F_{j+1}} [\phi(u) - \Delta_j]^2 \, du \Big/ \int_0^1 \phi^2(u) \, du,$$

where Δ_j and $\phi(u)$ are defined by (2.2) and (2.4), respectively. (4.13) can be made arbitrarily small, provided the Lebesgue measures of all the class intervals $\{I_j\}$ are also arbitrarily small. Again, from our Theorem 4.1 and Theorem 6.1 of Hájek [7], it follows that the loss in efficiency in the case where the assumed density differs from the true density, is given by

(4.14) $$1 - [\rho(\{I_j\})/\rho]\{1 - \sum_{j=0}^{\infty} \int_{F_j}^{F_{j+1}} [\phi(u) - \Delta_j]^2 \, du \Big/ \int_0^1 \phi^2(u) \, du\},$$

where $\rho(\{I_j\})$ is defined by (4.6) and ρ by (6.3) of [7]. (4.14) may be greater than, equal to or less than (4.13), depending on ρ and $\rho(\{I_j\})$.

REMARK 2. If we take $I_0 : x < x_0$, while $I_1, \cdots, I_\infty$ all have sufficiently small Lebesgue measure, the results will relate to the AMPROT for truncated case ($x \leqq x_0$ being truncated). More than one truncation can be dealt in a similar way. Again, in the two sample problem, Gastwirth [4] has considered the censored case where only $N_\nu^*(<N_\nu)$ of the ordered variables of the combined sample are observable, and N_ν^*/N_ν approaches $p(0 < p < 1)$ as $\nu \to \infty$. In his case, N_ν^* is given but the corresponding truncation point is random, while in our case, the truncation points are given and $N_{\nu j}$ ($j = 0, \cdots, \infty$) are random. In spite of this basic difference, the power properties can be studied by the same formulae. In this connection, the reader is also referred to Gastwirth [5], Chernoff et al. [2], Sarndal [15] and Kulldorff [11] for some interesting results related to the percentile censored case for the one sample and two sample location problems.

REMARK 3. In the particular case of $f(x)$ being assumed to be normal, the corresponding S_ν will be termed the *grouped normal store statistics*. The corresponding test in (3.11) will be AMPROT for normal alternatives. The *grouped Wilcoxon's test* also belongs to the class of tests considered here (namely when we work with logistic distribution). The test-statistic may be written as

$$W_\nu = \sum_{i=1}^{N_\nu} c_{\nu i} \sum_{j=0}^{\infty} \hat{\gamma}_{\nu j} Z_{ij}, \tag{4.15}$$

where Z_{ij}'s are defined by (1.7) and

$$\hat{\gamma}_{\nu j} = F_{N_\nu, j} + \tfrac{1}{2} N_{\nu, j}/N_\nu = \tfrac{1}{2}(F_{N_\nu, j} + F_{N_\nu, j+1}), \qquad j = 0, \cdots, \infty. \tag{4.16}$$

For the two sample problem, W_ν defined by (4.15) and (4.16) may be simplified further. On defining $N_{\nu j}$, $j = 0, \cdots, \infty$, as in (3.1) and $m_{\nu j}$ as the number of observations of the first sample (of size m_ν) belonging to I_j, $j = 0, \cdots, \infty$, it can be shown that W_ν is a linear function of

$$U_\nu = (1/n_\nu) \sum_{j=0}^{\infty} (m_{\nu j}/m_\nu)[\tfrac{1}{2} N_{\nu j} + \sum_{k<j} N_{\nu k}], \tag{4.17}$$

where $n_\nu (= N_\nu - m_\nu)$ is the size of the second sample. For a set of finite number of class intervals, statistics of the type (4.17) have been considered by Natrella [13], Halperin [8], Sugiura [17], Klotz [10], among others, while the present author ([16]) considered the same when the number of class intervals need not be finite (but be countable). For details, references may be given to [8], [10] and [16]. For a comparison of the grouped normal score test and the group Wilcoxon's test one may proceed as in [9] with the modifications suggested earlier in this section.

5. Acknowledgment. Thanks are due to the referee for his helpful suggestions.

REFERENCES

[1] Capon, J. (1961). Asymptotic efficiency of certain locally most powerful rank tests. *Ann. Math. Statist.* **32** 88–100.

[2] Chernoff, H., Gastwirth, J. L. and John, M. V., Jr. (1967). Asymptotic distributions of linear combinations of functions of order statistics. *Ann. Math. Statist.* **38** 52–72.

[3] Chernoff, H. and Savage, I. R. (1958). Asymptotic normality and efficiency of certain nonparametric test statistics. *Ann. Math. Statist.* **29** 972–994.

[4] Gastwirth, J. (1965). Asymptotically most powerful rank order tests for the two-sample problem with censored data. *Ann. Math. Statist.* **36** 1243–1247.

[5] Gastwirth, J. (1966). On robust procedures. *J. Amer. Statist. Assoc.* **61** 929–948.

[6] Hájek, J. (1961). Some extensions of the Wald-Wolfowitz-Noether theorem. *Ann. Math. Statist.* **32** 506–523.

[7] Hájek, J. (1962). Asymptotically most powerful rank order tests. *Ann. Math. Statist.* **33** 1124–1147.

[8] Halperin, M. (1960). Extension of the Wilcoxon-Mann-Whitney test to samples censored at the same fixed point. *J. Amer. Statist. Assoc.* **55** 125–138.

[9] Hodges, J. L., Jr. and Lehmann, E. L. (1961). Comparison of the normal score and Wilcoxon tests. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* **1** 307–317. Univ. of California Press.

[10] Klotz, J. (1966). The Wilcoxon, ties and the computer. *J. Amer. Statist. Assoc.* **61** 772–787.

[11] Kulldorff, G. (1961). *Estimation From Grouped and Partially Grouped Samples.* Almqvist and Wicksel, Stockholm.

[12] Lehmann, E. L. (1959). *Testing Statistical Hypotheses.* Wiley, New York.

[13] Natrella, M. G. (1954). Critical values for the Wilcoxon rank-sum test. SEL Note

54-14, Statistical Engineering Laboratory, National Bureau of Standards, Washington, D. C.

[14] Noether, G. E. (1955). On a theorem of Pitman. *Ann. Math. Statist.* **26** 64–68.

[15] Sarndäll, Carl E. (1962). *Information from censored samples.* Almqvist and Wicksel, Stockholm.

[16] Sen, P. K. (1965). On some permutation tests based on U-statistics. *Calcutta Statist. Assoc. Bull.* **14** 106–125.

[17] Sugiura, N. (1963). On a generalization of the Wilcoxon test for censored data. *Osaka Math. J.* **15** 257–268.