Ann. Inst. Statist. Math.
Vol. 54, No. 4, 934–944 (2002)

MULTIVARIATE PERCENTILE TESTS FOR INCOMPLETE DATA

HYO-IL PARK

Department of Statistics, Chong-ju University, Chong-ju, Choong-book, 360-764, Korea,
e-mail: hipark@chongju.ac.kr

(Received May 23, 2001; revised October 19, 2001)

Abstract. In this paper, we consider the percentile test procedures for multivariate and right censored data. Because of the involvement of censoring distribution into the distribution of the proposed test statistic, we study the asymptotic normality using the estimated covariance matrix. Finally, we derive the asymptotic relative efficiency and illustrate our procedures with an example.

Key words and phrases: Asymptotic relative efficiency, noncentrality parameter, Pitman translation parameter, two sample problem.

1. Introduction

Median tests as nonparametric procedures for two sample problem are well known and useful for detecting location translations. Basically there are two kinds of median tests in the univariate case. One is the control median test (e.g. Mathisen (1943)) and the other, the combined median test (e.g. Mood (1950)). The distinction between two kinds of median tests is as follows: the control median test uses a median from control sample whereas the combined median test uses a median from combined sample. From now on, we call simply median test for the combined median test. Two kinds of median tests have been modified or extended to the various directions. As a particular modification of the control median test, Gastwirth (1968) proposed the first median test in order to improve its performance as a two-sided test, which permits the experimenter to reach a decision early. Therefore the first median test would be useful in case of the life trial situation. Also Gastwirth discussed the application of the curtailed sampling to the first median test for the early decision in the same paper. For more detailed discussion of the curtailed sampling, we may refer to Alling (1963). Hettmansperger (1973) further considered a conservative test based on the first median test statistic to cover the Behrens-Fisher problem. Chatterjee and Sen (1964), Hettmansperger (1984) and Babu and Rao (1988) considered extensions of the median test to multivariate data. Recently, Park and Desu (1999) extended the control median test to multivariate data. Brookmeyer and Crowley (1982) modified the median test for right censored data. Gastwirth and Wang (1988) proposed the control median test for right censored data. Also Park and Desu (1998) considered an extension of the control median test to multivariate and right censored data. Therefore one may expect the advent of a median test procedure for multivariate and right censored data. However, for the case of right censoring data, it is not rare that one may not obtain a sample median or medians because of heavy censoring for larger observations or early termination of experiments. In this case, it is impossible to compare treatment effects with sample medians. In order to circumvent this stalemate, we consider a percentile test which uses the corresponding quantlie points instead of using medians. Also Gastwirth and Wang (1988) proposed a control percentile test for the consideration of the efficiency in the univariate case. In the next section, we propose a multivariate percentile test for right censored data. We deal with the large sample approximation in Section 3. Finally, we consider the asymptotic relative efficiency and show an example for illustration of our proposed test procedure.

2. Multivariate percentile test for right censored data

Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be two independent q-variate random vectors with continuous distribution functions F and G, respectively. It is of our concern to test the hypothesis $H_0 : F = G$. Since the location translation alternatives are of interest, we assume that in general,

$$G(\boldsymbol{x}) = F(\boldsymbol{x} - \boldsymbol{\Delta}) \text{ for all } \boldsymbol{x} \in \boldsymbol{R}^q \text{ and for some } \boldsymbol{\Delta} \in \boldsymbol{R}^q. \tag{2.1}$$

In view of this assumption, the null hypothesis can be restated as $H_0 : \boldsymbol{\Delta} = \boldsymbol{0}$. Usually, a random sample $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ of $\boldsymbol{X}$ and an independent random sample $\boldsymbol{Y}_1, \dots, \boldsymbol{Y}_m$ of $\boldsymbol{Y}$ are observed and tests are performed based on these samples. However in some experiments, one can only observe $\{(\boldsymbol{V}_i, \boldsymbol{\delta}_i), i = 1, \dots, n\}$ and $\{(\boldsymbol{W}_j, \boldsymbol{\tau}_j), j = 1, \dots, m\}$, where $V_{ki} = \min(X_{ki}, C_{ki})$, $\delta_{ki} = I(V_{ki} = X_{ki})$, $W_{kj} = \min(Y_{kj}, D_{kj})$ and $\tau_{kj} = I(W_{kj} = Y_{kj})$ for $i = 1, \dots, n$, $j = 1, \dots, m$ and $k = 1, \dots, q$. $I(\cdot)$ is the indicator function. It is assumed that $\boldsymbol{C}_1, \dots, \boldsymbol{C}_n$ is a censoring random sample with distribution function H_F and $\boldsymbol{D}_1, \dots, \boldsymbol{D}_m$ is an independent censoring random sample with distribution function H_G. Furthermore, it is assumed that $\boldsymbol{X}$'s, $\boldsymbol{Y}$'s, $\boldsymbol{C}$'s, and $\boldsymbol{D}$'s are all independent each other. For each k, $k = 1, \dots, q$, we denote F_k and G_k as the marginal distribution functions of F and G and $\hat{F}_{kn}$ and $\hat{G}_{km}$, as the corresponding Kaplan-Meier estimates. Also for each k, let $H_{kN} = (n/N)F_k + (m/N)G_k$ and $\hat{H}_{kN} = (n/N)\hat{F}_{kn} + (m/N)\hat{G}_{km}$ with $N = m + n$. Finally, for each p with $0 < p < 1$ and for each k, let $\xi^*_{kN}(p)$ be a p-th quantile of H_{kN} and $\hat{\xi}^*_{kN}(p)$, the corresponding p-th sample quantile of $\hat{H}_{kN}$. Then for any consistent estimate $\hat{\boldsymbol{\Sigma}}_N(p)$ of the limiting null covariance matrix $\boldsymbol{\Sigma}_0(p)$ of

$$\sqrt{n}(\hat{F}_{1n}(\hat{\xi}^*_{1N}(p)), \dots, \hat{F}_{qn}(\hat{\xi}^*_{qN}(p))),$$

assuming that the inverse $\hat{\boldsymbol{\Sigma}}_N^{-1}(p)$ of $\hat{\boldsymbol{\Sigma}}_N(p)$ exists, we propose a q-variate p-th percentile test statistic $\boldsymbol{M}_N$ as follows:

$$\boldsymbol{M}_N = n \begin{pmatrix} \hat{F}_{1n}(\hat{\xi}^*_{1N}(p)) - p \\ \cdots \\ \hat{F}_{qn}(\hat{\xi}^*_{qN}(p)) - p \end{pmatrix}^T \hat{\boldsymbol{\Sigma}}_N^{-1}(p) \begin{pmatrix} \hat{F}_{1n}(\hat{\xi}^*_{1N}(p)) - p \\ \cdots \\ \hat{F}_{qn}(\hat{\xi}^*_{qN}(p)) - p \end{pmatrix},$$

where T means the transpose of a matrix or a vector. We will identify $\boldsymbol{\Sigma}_0(p)$ and $\hat{\boldsymbol{\Sigma}}_N(p)$ later. Then an α-level test of $H_0 : F = G$ against $H_1 : F \neq G$ is to

$$\text{“reject } H_0 \quad \text{if} \quad \boldsymbol{M}_N \geq C(\alpha)\text{”}.$$

The constant $C(\alpha)$ is chosen so that the size of the test is α. Since the exact null distribution of $\boldsymbol{M}_N$ depends on F, G, H_F and H_G in a complicated manner, it is natural to consider the large sample approximation.

3. Limiting distribution of M_N

For the derivation of the limiting null distribution of M_N, we introduce some more notations about distribution and subdistribution functions. In the following, for each k, H_{F_k} and H_{G_k} denote the marginal distribution functions for H_F and H_G, respectively. For each k, $k = 1, \ldots, q$, let

$$F_k^*(u) = P(V_{ki} \le u, \delta_{ki} = 1), \qquad G_k^*(u) = P(W_{kj} \le u, \tau_{kj} = 1),$$
$$S_{F_k}(u) = P(V_{ki} > u) = (1 - F_k(u))(1 - H_{F_k}(u)),$$
$$S_{G_k}(u) = P(W_{kj} > u) = (1 - G_k(u))(1 - H_{G_k}(u)).$$

Also for each $1 \le k \ne l \le q$, let

$$F_{kl}^*(u, v) = P(V_{ki} \le u, V_{li} \le v, \delta_{ki} = \delta_{li} = 1),$$
$$G_{kl}^*(u, v) = P(W_{kj} \le u, W_{lj} \le v, \tau_{kj} = \tau_{lj} = 1),$$
$$S_{F_{kl}}(u, v) = P(V_{ki} > u, V_{li} > v), \qquad S_{G_{kl}}(u, v) = P(W_{kj} > u, W_{lj} > v),$$
$$N_{F_{kl}}(u, v) = P(V_{ki} \le u, V_{li} \ge v, \delta_{ki} = 1), \qquad M_{F_{kl}}(u, v) = P(V_{ki} \ge u, V_{li} \le v, \delta_{li} = 1),$$
$$N_{G_{kl}}(u, v) = P(W_{kj} \le u, W_{lj} \ge v, \tau_{kj} = 1),$$
$$M_{G_{kl}}(u, v) = P(W_{kj} \ge u, W_{lj} \le v, \tau_{lj} = 1).$$

Also we need the following assumptions:

ASSUMPTION 1. As $N \to \infty$, $n/m \to \lambda \in (0, \infty)$.

ASSUMPTION 2. For each k, $k = 1, \ldots, q$, F_k and G_k are continuous and twice differentiable at $\xi_{kN}^*(p)$ with $f_k(\xi_{kN}^*(p)) > 0$ and $g_k(\xi_{kN}^*(p)) > 0$ for each N, where f_k and g_k are the respective densities.

Now we state Bahadur representation of the Kaplan-Meier estimate, which is due to Lo and Singh (1985).

LEMMA 1. *For each k, $k = 1, \ldots, q$ and for each $0 < p < 1$, with Assumptions 1 and 2, we have with probability one (w.p.1), as $N \to \infty$,*

$$\hat{F}_{kn}(\xi_{kN}^*(p)) - F_k(\xi_{kN}^*(p)) = \frac{1}{n}\sum_{i=1}^{n} \beta(V_{ki}, \delta_{ki}, \xi_{kN}^*(p)) + O(N^{-3/4}(\log N)^{3/4}) \quad \textit{and}$$

$$\hat{G}_{km}(\xi_{kN}^*(p)) - G_k(\xi_{kN}^*(p)) = \frac{1}{m}\sum_{j=1}^{m} \gamma(W_{kj}, \tau_{kj}, \xi_{kN}^*(p)) + O(N^{-3/4}(\log N)^{3/4}),$$

where

$$\beta(V_{ki}, \delta_{ki}, t) = (1 - F_k(t))\left\{\frac{I(V_{ki} \le t, \delta_{ki} = 1)}{S_{F_k}(V_{ki})} - \int_0^t \frac{I(V_{ki} \ge u)dF_k^*(u)}{S_{F_k}^2(u)}\right\} \quad \textit{and}$$

$$\gamma(W_{kj}, \tau_{kj}, t) = (1 - G_k(t))\left\{\frac{I(W_{kj} \le t, \tau_{kj} = 1)}{S_{G_k}(W_{kj})} - \int_0^t \frac{I(W_{kj} \ge v)dG_k^*(v)}{S_{G_k}^2(v)}\right\}.$$

LEMMA 2. *Under Assumption 2, for each k, $k = 1, \dots, q$ and for each $0 < p < 1$, w.p.1, as $N \to \infty$,*

$$\hat{\xi}^*_{kN}(p) - \xi^*_{kN}(p) = \frac{p - \hat{H}_{kN}(\xi^*_{kN}(p))}{h_{kN}(\xi^*_{kN}(p))} + O(N^{-3/4}(\log N)^{3/4}),$$

where h_{kN} is the density of H_{kN}.

PROOF. First of all, we note that

$$\hat{H}_{kN}(t) - H_{kN}(t) = \frac{n}{N}[\hat{F}_{kn}(t) - F_k(t)] + \frac{m}{N}[\hat{G}_{km}(t) - G_k(t)]. \tag{3.1}$$

Then from Lemma 3 in Lo and Singh (1985), we see that w.p.1, as $N \to \infty$,

$$\sup_{0<p<1} |\hat{H}^{-1}_{kN}(p) - H^{-1}_{kN}(p)| = O(N^{-1/2}(\log N)^{1/2}). \tag{3.2}$$

Thus from Taylor's expansion around $\xi^*_{kN}(p)$ and (3.2), we have w.p.1, as $N \to \infty$,

$$H_{kN}(\hat{\xi}^*_{kN}(p)) - H_{kN}(\xi^*_{kN}(p)) = h_{kN}(\xi^*_{kN}(p))(\hat{\xi}^*_{kN}(p) - \xi^*_{kN}(p)) + O(N^{-1}\log N). \tag{3.3}$$

Also from Cheng (1984) with (3.1), we have w.p.1, as $N \to \infty$,

$$\begin{aligned} &\hat{H}_{kN}(\hat{\xi}^*_{kN}(p)) - \hat{H}_{kN}(\xi^*_{kN}(p)) - H_{kN}(\hat{\xi}^*_{kN}(p)) + H_{kN}(\xi^*_{kN}(p)) \\ &\quad = O(N^{-3/4}(\log N)^{3/4}). \end{aligned} \tag{3.4}$$

Since $\hat{H}_{kN}(\hat{\xi}^*_{kN}(p)) = p + O(N^{-1})$, we have w.p.1 from (3.4) with (3.3), we obtain the result.

THEOREM 1. *Under Assumption 2, for each k, $k = 1, \dots, q$ and for each $0 < p < 1$, w.p.1, as $N \to \infty$,*

$$\begin{aligned} &\hat{F}_{kn}(\hat{\xi}^*_{kN}(p)) - F_k(\xi^*_{kN}(p)) \\ &\quad = \left\{\frac{1}{n} - \frac{f_k(\xi^*_{kN}(p))}{h_{kN}(\xi^*_{kN}(p))}\frac{1}{N}\right\} \sum_{i=1}^{n} \beta(V_{ki}, \delta_{ki}, \xi^*_{kN}(p)) \\ &\quad\quad - \frac{f_k(\xi^*_{kN}(p))}{h_{kN}(\xi^*_{kN}(p))}\frac{1}{N}\sum_{j=1}^{m} \gamma(W_{kj}, \tau_{kj}, \xi^*_{kN}(p)) + O(N^{-3/4}(\log N)^{3/4}). \end{aligned}$$

PROOF.

$$\begin{aligned} &\hat{F}_{kn}(\hat{\xi}^*_{kN}(p)) - F_k(\xi^*_{kN}(p)) \\ &\quad = \{\hat{F}_{kn}(\hat{\xi}^*_{kN}(p)) - F_k(\hat{\xi}^*_{kN}(p)) - \hat{F}_{kn}(\xi^*_{kN}(p)) + F_k(\xi^*_{kN}(p))\} \\ &\quad\quad + \{F_k(\hat{\xi}^*_{kN}(p)) - F_k(\xi^*_{kN}(p))\} + \{\hat{F}_{kn}(\xi^*_{kN}(p)) - F_k(\xi^*_{kN}(p))\} \\ &\quad = A + B + C, \quad \text{say}. \end{aligned}$$

Then by Cheng (1984), w.p.1, as $N \to \infty$,

$$A = O(N^{-3/4}(\log N)^{3/4}).$$

From Taylor's expansion and Lemmas 1 and 2, w.p.1, as $N \to \infty$,

$$B = -\frac{f_k(\xi_{kN}^*(p))}{h_{kN}(\xi_{kN}^*(p))}\frac{1}{N}\left\{\sum_{i=1}^{n}\beta(V_{ki},\delta_{ki},\xi_{kN}^*(p)) + \sum_{j=1}^{m}\gamma(W_{kj},\tau_{kj},\xi_{kN}^*(p))\right\}$$
$$+O(N^{-3/4}(\log N)^{3/4}).$$

Therefore this theorem is followed by applying Lemma 1 to C.

We note that under H_0

$$E(\beta(V_{ki},\delta_{ki},\xi_{kN}^*(p))) = E(\gamma(W_{kj},\tau_{kj},\xi_{kN}^*(p))) = 0$$
$$V(\beta(V_{ki},\delta_{ki},\xi_{kN}^*(p))) = (1-F_k(\xi_{kN}^*(p)))^2\int_0^{\xi_{kN}^*(p)}\frac{dF_k^*(u)}{S_{F_k}^2(u)} \quad \text{and}$$
$$V(\gamma(W_{kj},\tau_{kj},\xi_{kN}^*(p))) = (1-G_k(\xi_{kN}^*(p)))^2\int_0^{\xi_{kN}^*(p)}\frac{dG_k^*(u)}{S_{G_k}^2(u)}.$$

Also we obtain by applying Fubini's theorem that

$$\begin{aligned}
&(3.5)\ \mathrm{Cov}(\beta(V_{ki},\delta_{ki},\xi_{kN}^*(p)),\beta(V_{li},\delta_{li},\xi_{lN}^*(p)))\\
&= (1-F_k(\xi_{kN}^*(p)))(1-F_l(\xi_{lN}^*(p)))\\
&\left[\int_0^{\xi_{kN}^*(p)}\int_0^{\xi_{lN}^*(p)}\frac{d^2F_{kl}^*(u,v)}{S_{F_k}(u)S_{F_l}(v)} + \int_0^{\xi_{kN}^*(p)}\int_0^{\xi_{lN}^*(p)}\frac{S_{F_{kl}}(u,v)dF_k^*(u)dF_l^*(v)}{S_{F_k}^2(u)S_{F_l}^2(v)}\right.\\
&-\int_0^{\xi_{lN}^*(p)}\left\{\int_0^{\xi_{kN}^*(p)}\int_v^{\infty}\frac{d^2N_{F_{kl}}(u,s)}{S_{F_k}(u)}\right\}\frac{dF_l^*(v)}{S_{F_l}^2(v)}\\
&\left.-\int_0^{\xi_{kN}^*(p)}\left\{\int_0^{\xi_{lN}^*(p)}\int_u^{\infty}\frac{d^2M_{F_{kl}}(s,v)}{S_{F_l}(v)}\right\}\frac{dF_k^*(u)}{S_{F_k}^2(u)}\right]\\
&= (1-F_k(\xi_{kN}^*(p)))(1-F_l(\xi_{lN}^*(p)))(C_1(F)+C_2(F)-C_3(F)-C_4(F)), \quad \text{say}
\end{aligned}$$

and

$$\begin{aligned}
&(3.6)\ \mathrm{Cov}(\gamma(W_{kj},\tau_{kj},\xi_{kN}^*(p)),\gamma(W_{lj},\tau_{lj},\xi_{lN}^*(p)))\\
&= (1-G_1(\xi_{kN}^*))(1-G_2(\xi_{lN}^*))\\
&\left[\int_0^{\xi_{kN}^*(p)}\int_0^{\xi_{lN}^*(p)}\frac{d^2G_{kl}^*(u,v)}{S_{G_k}(u)S_{G_l}(v)} + \int_0^{\xi_{kN}^*(p)}\int_0^{\xi_{lN}^*(p)}\frac{S_{G_{kl}}(u,v)dG_k^*(u)dG_l^*(v)}{S_{G_k}^2(u)S_{G_l}^2(v)}\right.\\
&-\int_0^{\xi_{lN}^*(p)}\left\{\int_0^{\xi_{kN}^*(p)}\int_v^{\infty}\frac{d^2N_{G_{kl}}(u,s)}{S_{G_k}(u)}\right\}\frac{dG_l^*(v)}{S_{G_l}^2(v)}\\
&\left.-\int_0^{\xi_{kN}^*(p)}\left\{\int_0^{\xi_{lN}^*(p)}\int_u^{\infty}\frac{d^2M_{G_{kl}}(s,v)}{S_{G_l}(v)}\right\}\frac{dG_k^*(u)}{S_{G_k}^2(u)}\right]\\
&= (1-G_k(\xi_{kN}^*(p)))(1-G_l(\xi_{lN}^*(p)))(C_1(G)+C_2(G)-C_3(G)-C_4(G)), \quad \text{say.}
\end{aligned}$$

We now return to the subject of the limiting null distribution of $\boldsymbol{M}_N$. From Theorem 1, for each k, $k = 1, \ldots, q$, the limiting distributions of $\sqrt{n}(\hat{F}_{kn}(\hat{\xi}^*_{kN}(p)) - p)$ and

$$\frac{1}{n}\sum_{i=1}^{n}\beta(V_{ki}, \delta_{ki}, \xi^*_{kN}(p)) - \frac{1}{N}\left\{\sum_{i=1}^{n}\beta(V_{ki}, \delta_{ki}, \xi^*_{kN}(p)) + \sum_{j=1}^{m}\gamma(W_{kj}, \tau_{kj}, \xi^*_{kN}(p))\right\}$$

are the same under H_0. Therefore from the central limit theorem with Assumptions 1 and 2, under H_0, we see that for each k, $k = 1, \ldots, q$, $\sqrt{n}(\hat{F}_{kn}(\hat{\xi}^*_{kN}(p)) - p)$ converges in distribution to a normal random variable with mean 0 and variance $\sigma_k^2(p)$,

$$\sigma_k^2(p) = \frac{(1-p)^2}{(1+\lambda)^2}\int_0^{\xi^*_k(p)}\frac{dF^*_k(u)}{S^2_{F_k}(u)} + \frac{\lambda(1-p)^2}{(1+\lambda)^2}\int_0^{\xi^*_k(p)}\frac{dG^*_k(u)}{S^2_{G_k}(u)},$$

where $\xi^*_k(p) = \lim_{N\to\infty}\xi^*_{kN}(p)$. Also from (3.5) and (3.6), we see that the limiting null covariance between

$$\sqrt{n}(\hat{F}_{kn}(\hat{\xi}^*_{kN}(p)) - p) \quad \text{and} \quad \sqrt{n}(\hat{F}_{ln}(\hat{\xi}^*_{lN}(p)) - p)$$

is

$$\begin{aligned}\sigma_{kl}(p) &= \frac{(1-p)^2}{(1+\lambda)^2}(C_1(F) + C_2(F) - C_3(F) - C_4(F)) \\ &\quad + \frac{\lambda(1-p)^2}{(1+\lambda)^2}(C_1(G) + C_2(G) - C_3(G) - C_4(G))\end{aligned}$$

with substitution of $\xi^*_k(p)$ for $\xi^*_{kN}(p)$ in (3.5) and (3.6). Then by applying Cramér-Wold device (cf. Billingsley (1986)), we obtain the following result.

THEOREM 2. *For any consistent estimate $\hat{\mathbf{\Sigma}}_N(p)$ of $\mathbf{\Sigma}_0(p)$, under H_0, $\boldsymbol{M}_N$ converges in distribution to a χ^2 random variable with q degrees of freedom, where*

$$\mathbf{\Sigma}_0(p) = \begin{pmatrix}\sigma_1^2(p)\cdots\sigma_1 q(p)\\ \cdots \\ \sigma_1 q(p)\cdots\sigma_q^2(p)\end{pmatrix}.$$

We note that under H_0, the first and the second parts of each variance and covariance term are the same except for λ. Therefore we could have reduced the expression of $\mathbf{\Sigma}_0(p)$ to a more concise form. However since we have to obtain $\hat{\mathbf{\Sigma}}_N(p)$ from two samples, we do not reduce them in this manner. A consistent estimate $\hat{\mathbf{\Sigma}}_N(p)$ for $\mathbf{\Sigma}_0(p)$ can be obtained by substituting empirical ones for the quantities, which were introduced at the beginning of this section. Then one can show the consistency of $\hat{\mathbf{\Sigma}}_N(p)$ by proving the consistency of each component of $\hat{\mathbf{\Sigma}}_N(p)$. For more detailed discussion, we may refer to Park and Desu (1998).

4. Asymptotic relative efficiency and an example

In this section, we study the asymptotic relative efficiency (ARE). For this matter, we only consider comparing two types of the median tests. Let $\boldsymbol{L}_N$ be the control median test statistic which was proposed by Park and Desu (1998). We begin this section by stating the definition of ARE for the multivariate version (cf. Puri and Sen (1985)). For two sequences of test statistics, say $\{\boldsymbol{Q}_N\}$ and $\{\boldsymbol{Q}^*_N\}$, having asymptotically (under a sequence $\{H_{1N}\}$ of alternative hypotheses) noncentral chi-square distributions with q degrees of freedom and noncentrality parameters $\boldsymbol{\Psi}$ and $\boldsymbol{\Psi}^*$, respectively, the ARE of $\{\boldsymbol{Q}_N\}$ relative to $\{\boldsymbol{Q}^*_N\}$ is defined by

$$\text{ARE}(\boldsymbol{Q},\boldsymbol{Q}^*) = \frac{\boldsymbol{\Psi}}{\boldsymbol{\Psi}^*}.$$

Noncentrality parameters of $\{\boldsymbol{M}_N\}$ and $\{\boldsymbol{L}_N\}$ under alternatives depend on the censoring distributions in a complicated manner. Therefore we will assume, in this section, that censoring distributions for two samples are equal. We consider ARE under the Pitman translation alternatives: for each k, $k = 1,\ldots,q$ and for each N,

$$H_{1N} : \boldsymbol{\Delta}_N = (\Delta_{1N},\ldots,\Delta_{qN})^T = (\theta_1/\sqrt{N},\ldots,\theta_q/\sqrt{N})^T,$$

where for each k, θ_k is some nonzero constant. Before we derive the ARE$(\boldsymbol{L},\boldsymbol{M})$, we review a useful relation between the noncentrality parameter and the efficacies of components of test statistics under the Pitman translation alternatives. For this purpose, let $\{\boldsymbol{Z}_N = (Z_{1N},\ldots,Z_{qN})^T\}$ be a sequence of q-variate test statistics such that for each N, $\boldsymbol{Z}_N$ is arbitrarily distributed with mean vector, $\boldsymbol{\mu}_N(\boldsymbol{\Delta}_N)$ and covariance matrix, $\boldsymbol{\Sigma}_N(\boldsymbol{\Delta}_N)$, where $\boldsymbol{\mu}_N(\boldsymbol{\Delta}_N) = E(\boldsymbol{Z}_N \mid \boldsymbol{\Delta}_N)$ and $\boldsymbol{\Sigma}_N(\boldsymbol{\Delta}_N) = V(\boldsymbol{Z}_N \mid \boldsymbol{\Delta}_N)$. We assume that $\boldsymbol{Z}_N$ converges in distribution to $\boldsymbol{Z}$, where $\boldsymbol{Z}$ is normally distributed with mean vector, $\boldsymbol{\mu}$ and covariance matrix, $\boldsymbol{\Sigma}$. Then we note that $(\boldsymbol{Z}_N - \boldsymbol{\mu}_N)^T\boldsymbol{\Sigma}_N^{-1}(\boldsymbol{Z}_N - \boldsymbol{\mu}_N)$ converges in distribution to a chi-square random variable with q degrees of freedom. With those notations and assumptions, we state the following result.

LEMMA 3. *For the sequence $\{\boldsymbol{Z}_N\}$ of test statistics, suppose that*

(1) *for each k and for each N, $\frac{d}{d\Delta}\mu_{kN}(\Delta) = \mu'_{kN}(\Delta)$ is assumed to exist and be continuous in some neighborhood of 0 with $\mu'_N(0) \neq 0$,*

(2) $\lim_{N\to\infty} \mu'_{kN}(\Delta_{kN})/\mu'_{kN}(0) = 1$ *and*

(3) $\lim_{N\to\infty} \boldsymbol{\Sigma}_N(\boldsymbol{\Delta}_N) = \boldsymbol{\Sigma}$.

Then under the Pitman translation alternatives, the limiting distribution of $\boldsymbol{Z}_N\boldsymbol{\Sigma}_N^{-1}(\boldsymbol{\Delta}_N)$ $\boldsymbol{Z}_N$ is a noncentral chi-square distribution with q degrees of freedom and the noncentrality parameter

$$\boldsymbol{\Psi} = \begin{pmatrix}\theta_1 e_1\\ \cdots\\ \theta_p e_p\end{pmatrix}^T \boldsymbol{P}^{-1} \begin{pmatrix}\theta_1 e_1\\ \cdots\\ \theta_p e_p\end{pmatrix},$$

where for each k, e_k is the efficacy of the k-th component, Z_{kN} of the test statistic, $\boldsymbol{Z}_N$ and $\boldsymbol{P}$ is the limiting correlation matrix.

PROOF. See Park and Desu (1999).

We note that the conditions (1) and (2) in Lemma 3 with the condition that the sequence $\{\boldsymbol{Z}_N\}$ of test statistics has a limiting distribution, are exactly the same as those for the derivation of the efficacy of $\{\boldsymbol{Z}_N\}$ in Theorem 5.2.7 of Randles and Wolfe (1979) except the existence of the efficacy itself. Assumption 1 in Section 3 implies that $\xi^*_{kN} \to \xi^*_k$, where ξ^*_k is a median of $H_k = (1/(1+\lambda))F_k + (\lambda/(1+\lambda))G_k$. Since the noncentrality parameter contains the expressions of the limiting distribution of the sequence of test statistics, without loss of generality, we use ξ_k^* instead of ξ_{kN}^* in the sequel to obtain the noncentrality parameter. Also we use Δ_k instead of Δ_{kN} for each N when there is no confusion. We note that under H_0 and the Pitman translation alternatives, ξ_k^* becomes also a median of F_k and G_k.

Therefore in view of Lemma 3, it is enough to derive the efficacies of q components and limiting correlation matrix for the noncentrality parameter. For each k, $k = 1, \dots, q$ and $1 \le k \ne l \le q$, define

$$\mu_{kN}(\Delta_k) = \sqrt{n}(G_k(\xi_k^* + \Delta_k) - 1/2),$$

$$\sigma_{kN}^2(\Delta_k) = \left\{1 - \frac{n}{N}\frac{f_k(\xi_k^* + \Delta_k)}{h_{kN}(\xi_k^* + \Delta_k)}\right\}^2 (1 - F_k(\xi_k^* + \Delta_k))^2 \int_0^{\xi_k^*+\Delta_k} \frac{dF_k^*(u)}{S_{F_k}^2(u)} + \frac{mn}{N^2}\left\{\frac{f_k(\xi_k^* + \Delta_k)}{h_{kN}(\xi_k^* + \Delta_k)}\right\}^2 (1 - G_k(\xi_k^* + \Delta_k))^2 \int_0^{\xi_k^*+\Delta_k} \frac{dG_k^*(u)}{S_{G_k}^2(u)}$$

and

$$\sigma_{klN}(\mathbf{\Delta}) = \left\{1 - \frac{n}{N}\frac{f_k(\xi_k^* + \Delta_k)}{h_{kN}(\xi_k^* + \Delta_k)}\right\}\left\{1 - \frac{n}{N}\frac{f_l(\xi_l^* + \Delta_l)}{h_{lN}(\xi_l^* + \Delta_l)}\right\} \mathrm{Cov}\{\beta(V_{ki}, \delta_{ki}, \xi_k^* + \Delta_k), \beta(V_{li}, \delta_{li}, \xi_l^* + \Delta_l)\} + \frac{mn}{N^2}\frac{f_k(\xi_k^* + \Delta_k)f_l(\xi_l^* + \Delta_l)}{h_{kN}(\xi_k^* + \Delta_k)h_{lN}(\xi_l^* + \Delta_l)} \mathrm{Cov}\{\gamma(W_{kj}, \tau_{kj}, \xi_k^* + \Delta_k), \gamma(W_{lj}, \tau_{lj}, \xi_l^* + \Delta_l)\}.$$

Then under the Pitman translation alternatives,

$$n \begin{pmatrix} \hat{F}_{1n}(\hat{\xi}_1^*) - 1/2 - \mu_{1N}(\Delta_1) \\ \cdots \\ \hat{F}_{qn}(\hat{\xi}_q^*) - 1/2 - \mu_{qN}(\Delta_q) \end{pmatrix}^T \mathbf{\Sigma}_N^{-1}(\mathbf{\Delta}) \begin{pmatrix} \hat{F}_{1n}(\hat{\xi}_1^*) - 1/2 - \mu_{1N}(\Delta_1) \\ \cdots \\ \hat{F}_{qn}(\hat{\xi}_q^*) - 1/2 - \mu_{qN}(\Delta_q) \end{pmatrix}$$

converges in distribution to a chi-square random variable with q degrees of freedom. Therefore we can use the Lemma 3 to derive the noncentrality parameter by checking the three conditions. Assumption 2 in Section 3 guarantees the condition (1). Thus we have

$$\frac{d\mu_{kN}(\Delta_k)}{d\Delta_k} = \sqrt{n} f_k(\xi_k^* + \Delta_k) \quad \text{and} \quad \left.\frac{d\mu_{kN}(\Delta_k)}{d\Delta_k}\right|_{\Delta_k=0} = \sqrt{n} f_k(\xi_k^*).$$

With the fact that $\Delta_k \to 0$ as $N \to \infty$, we see that

$$\lim_{N\to\infty} \frac{d\mu_{kN}(\Delta_k)/d\Delta_k}{d\mu_{kN}(\Delta_k)/d\Delta_k|_{\Delta_k=0}} = 1,$$

which confirms the condition (2).

In order to check the condition (3), we take $\mathbf{\Sigma}_0$ as $\mathbf{\Sigma}$. $\mathbf{\Sigma}_0$ was defined in Section 3. Since $h_{kN} = (n/N)f_k + (m/N)g_k$ with the fact that ξ_k^* is a common median of F_k and G_k under the Pitman translation alternatives, we have

$$\lim_{N\to\infty} \frac{f_k(\xi_k^* + \Delta_k)}{h_{kN}(\xi_k^* + \Delta_k)} = 1, \quad \text{and}$$

$$\lim_{N\to\infty} F_k(\xi_k^* + \Delta_k) = \lim_{N\to\infty} G_k(\xi_k^* + \Delta_k) = G_k(\xi_k^*) = 1/2.$$

Thus with the assumption that censoring distributions for the control and the treatment are equal, we can conclude with Assumption 1 that for each k,

$$\lim_{N\to\infty} \sigma_{kN}^2(\Delta_k) = \sigma_k^2.$$

Also with the same arguments used for σ_k^2, we can show that

$$\lim_{N\to\infty} \sigma_{klN}(\boldsymbol{\Delta}) = \sigma_{kl}.$$

Therefore we have shown that all the three conditions in Lemma 3 are satisfied. This means that, in view of Lemma 3, it is enough to consider the efficacies of two components with limiting null correlation matrix to obtain the noncentrality parameter for the median test $\boldsymbol{M}_N$.

The conditions and method for the derivation of the efficacy for tests statistics are well summarized in Randles and Wolfe (1979). Already we have noticed that all the conditions in Theorem 5.2.7 in Randles and Wolfe are satisfied except the existence of the efficacy. Therefore it is enough to check that condition. Then some simple considerations for the efficacy e_k of the k-th component of $\boldsymbol{M}_N$ leads as follows:

$$e_k = 2g_k(\xi_k^*)\left\{\frac{(1+\lambda)^2}{\lambda}\int_0^{\xi_k^*}\frac{dG_k^*(u)}{S_{G_k}^2(u)}\right\}^{-1/2}.$$

Also straightforward calculations produce the limiting null correlation matrix $\boldsymbol{P}$ with

$$P_{11} = \cdots = P_{qq} = 1 \quad \text{and}$$

$$P_{kl} = P_{lk} = (C_1(G) + C_2(G) - C_3(G) - C_4(G))\left\{\int_0^{\xi_k^*}\frac{dG_k(u)}{S_{G_k}^2(u)}\int_0^{\xi_l^*}\frac{dG_l(u)}{S_{G_l}^2(u)}\right\}^{-1/2}.$$

In the following, we denote ξ_k as a median of G_k for each k. Then we note that under H_0 and Pitman translation alternatives, $\xi_k = \xi_k^*$. In order to derive the noncentrality parameter for the control median tests statistics $\{\boldsymbol{L}_N\}$ (cf. Park and Desu (1998)), define for each k, $k = 1, \ldots, q$ and $1 \le k \ne l \le q$,

$$\begin{aligned}
\mu_{kN}(\Delta_k) &= \sqrt{n}(G_k(\xi_k + \Delta_k) - 1/2),\\
\sigma_{kN}^2(\Delta_k) &= (1 - F_k(\xi_k + \Delta_k))^2\int_0^{\xi_k+\Delta_k}\frac{dF_k^*(u)}{S_{F_k}^2(u)}\\
&\quad + \frac{n}{m}(1 - G_k(\xi_k+\Delta_k))^2\frac{f_k^2(\xi_k+\Delta_k)}{g_k^2(\xi_k+\Delta_k)}\int_0^{\xi_k+\Delta_k}\frac{dG_k^*(u)}{S_{G_k}^2(u)}
\end{aligned}$$

and

$$\begin{aligned}
\sigma_{klN}(\boldsymbol{\Delta}) &= \mathrm{Cov}\{\beta(V_{ki}, \delta_{ki}, \xi_k + \Delta_k), \beta(V_{li}, \delta_{li}, \xi_l + \Delta_l)\}\\
&\quad + \frac{n}{m}\frac{f_k(\xi_k+\Delta_k)f_l(\xi_l+\Delta_l)}{g_k(\xi_k+\Delta_k)g_l(\xi_l+\Delta_l)}\\
&\quad \mathrm{Cov}\{\gamma(W_{kj}, \tau_{kj}, \xi_k+\Delta_k), \gamma(W_{lj}, \tau_{lj}, \xi_l+\Delta_l)\}.
\end{aligned}$$

Thus

$$n\begin{pmatrix}\hat{F}_{1n}(\hat{G}_{1m}^{-1}(1/2)) - 1/2 - \mu_{1N}(\Delta_1)\\ \cdots \\ \hat{F}_{qn}(\hat{G}_{qm}^{-1}(1/2)) - 1/2 - \mu_{qN}(\Delta_q)\end{pmatrix}^T \boldsymbol{\Sigma}_N^{-1}(\boldsymbol{\Delta})\begin{pmatrix}\hat{F}_{1n}(\hat{G}_{1m}^{-1}(1/2)) - 1/2 - \mu_{1N}(\Delta_1)\\ \cdots \\ \hat{F}_{qn}(\hat{G}_{qm}^{-1}(1/2)) - 1/2 - \mu_{qN}(\Delta_q)\end{pmatrix}$$

converges in distribution to a chi-square random variable with q degrees of freedom. Therefore by the same arguments used for $\{\boldsymbol{M}_N\}$, we can show that the conditions (1), (2) and (3) in Lemma 3, are all satisfied. Then by checking all the conditions of Theorem 5.4.7 in Randles and Wolfe (1979), we obtain the same efficacies as those of $\{\boldsymbol{L}_N\}$. Also it is easy to show that the limiting correlation matrix for $\boldsymbol{L}_N$ is the same as that of $\boldsymbol{M}_N$. Therefore we conclude that with the fact that $\xi_k^* = \xi_k$ under Pitman translation alternatives, $\text{ARE}(\boldsymbol{L}, \boldsymbol{M}) = 1$.

Finally we illustrate our procedure with the NCGS data considered by Wei and Lachin (1984). The patients are allocated into two groups, i.e. control (placebo) and treatment (high dose) groups with sample sizes $n = 48$ and $m = 65$. The Kaplan-Meier estimate for the second component of high dose group (X_{12}) shows that a sample median cannot be obtained because of the heavy censoring of higher observations. Therefore one can not apply any median test procedure. Since the lower (or first) sample quartile point (25%) can be achieved for all components, we consider applying the 25 percentile test to this example. The necessary statistics for obtaining the 25 percentile test statistic are as follows:

$$\hat{\xi}^*_{1,113}(.25) = 249.23 \quad \text{and} \quad \hat{\xi}_{2,113}(.25) = 640.26$$
$$\hat{F}_{1,48}(249.23) - 0.25 = 0.44 - 0.25 = 0.19$$
$$\hat{F}_{2,48}(640.26) - 0.25 = 0.27 - 0.25 = 0.02$$
$$\Sigma_{113} = \begin{pmatrix} 0.2842884 & 0.1733398 \\ 0.1733398 & 0.1834684 \end{pmatrix} \quad \text{and} \quad \Sigma_{113}^{-1} = \begin{pmatrix} 8.2975131 & -7.839439 \\ -7.839439 & 12.857183 \end{pmatrix}.$$

Then we obtain that $M_{113} = 11.765$, whose p-value is less than 0.005 from the chi-square distribution with 2 degrees of freedom. Therefore we may conclude that the two groups of patients are significantly different for the disease progression.

Acknowledgements

The author would like to express his sincere appreciations to an anonymous referee for suggesting constructive opinions, introducing the necessary materials and pointing out errors to improve the original version of this paper.

REFERENCES

Alling, D. W. (1963). Early decision in the Wilcoxon two-sample test, *J. Amer. Statist. Assoc.*, **58**, 713–720.

Babu, G. J. and Rao, C. R. (1988). Joint asymptotic distribution of marginal quantile functions in samples from a multivariate population, *J. Multivariate Anal.*, **27**, 15–23.

Billingsley, P. (1986). *Probability and Measure*, 2nd ed., Wiley, New York.

Brookmeyer, R. and Crowley, J. (1982). A k-sample median test for censored data, *J. Amer. Statist. Assoc.*, **77**, 433–440.

Chatterjee, S. K. and Sen, P. K. (1964). Non-parametric tests for the bivariate two-sample location problem, *Calcutta Statist. Assoc. Bull.*, **13**, 18–58.

Cheng, K. F. (1984). On almost sure representations for quantiles of the product-limit estimator with applications, *Sankhyā, Ser. A*, **46**, 426–443.

Gastwirth, J. L. (1968). The first-median test: A two-sided version of the control median test, *J. Amer. Statist. Assoc.*, **63**, 692–706.

Gastwirth, J. L. and Wang, J. L. (1988). Control percentile test procedures for censored data, *J. Statist. Plann. Inference*, **18**, 267–276.

Hettmansperger, T. P. (1973). A large sample conservative test for location with unknown scale parameters, *J. Amer. Statist. Assoc.*, **68**, 466–468.

Hettmansperger, T. P. (1984). *Statistical Inference Based on Ranks*, Wiley, New York.

Lo, S. H. and Singh, K. (1985). The product-limit estimator and the bootstrap: Some asymptotic representations, *Probab. Theory Related Fields*, **71**, 455–465.

Mathisen, H. C. (1943). A method of testing the hypothesis that two samples are from the same population, *Ann. Math. Statist.*, **14**, 188–194.

Mood, A. M. (1950). *Introduction to the Theory of Statistics*, McGraw-Hill, New York.

Park, H. I. and Desu, M. M. (1998). Multivariate control median test for right censored data, *Comm. Statist. Theory Methods*, **27**(8), 1923–1935.

Park, H. I. and Desu, M. M. (1999). A multivariate control median test, *J. Statist. Plann. Inference*, **79**, 123–139.

Puri, M. L. and Sen, P. K. (1985). *Nonparametric Methods in General Linear Models*, Wiley, New York.

Randles, H. R. and Wolfe, D. A. (1979). *Introduction to the Theory of Nonparametric Statistics*, Wiley, New York.

Wei, L. J. and Lachin, J. M. (1984). Two-sample asymptotically distribution-free tests for incomplete multivariate observations, *J. Amer. Statist. Assoc.*, **79**, 653–661.