ELSEVIER

Journal of Statistical Planning and Inference 102 (2002) 441–466

journal of statistical planning and inference

www.elsevier.com/locate/jspi

A weighted generalization of the Mann–Whitney–Wilcoxon statistic

Jingdong Xie[a], Carey E. Priebe[b,*]

[a] Bristol-Myers Squibb Pharmaceutical Research Institute, Princeton, NJ 08543, USA

[b] Department of Mathematical Sciences, Whiting School of Engineering, Johns Hopkins University, 104 Whitehead Hall, 3400 N. Charles St., Baltimore, MD 21218, USA

Received 30 September 1999; received in revised form 15 January 2001; accepted 15 March 2001

Abstract

Nonparametric statistics, especially the Mann–Whitney–Wilcoxon (MWW) statistic, have gained widespread acceptance, but are by no means the preferred method for statistical analysis in all situations. A main obstacle to their even wider applicability is that the price paid for their distribution-free property is the loss of efficacy. In fact, the MWW statistic, which is an estimate of the functional $\int_{-\infty}^{\infty} F(x)\,\mathrm{d}G(x)$, has an efficacy which varies with the underlying distributions $F(x)$ and $G(x)$. To improve the efficacy, this dissertation generalizes the classical MWW statistic to estimates of the functional $\int_{-\infty}^{\infty} u\{F(x)\}\,\mathrm{d}v\{G(x)\}$, where the functions $u(x)$ and $v(x)$ are strictly increasing on $[0,1]$. Statistical properties of this generalization such as asymptotic normality and admissibility are fully investigated. The optimal choices of functions $u(x)$ and $v(x)$ are studied via the tail binomial polynomials and the Pitman asymptotic efficacy criterion. In the one-sample problem, a similar generalization, based upon the functional $\int_{-\infty}^{\infty} u(1-F(-x))\,\mathrm{d}v(F(x))$, extends the Wilcoxon signed rank statistic.

MSC: 62G10

Keywords: Pitman asymptotic efficacy; Rank sum test; Signed rank test; Tail binomial polynomials; U-statistic

1. Introduction

A fundamental problem in nonparametric statistics is deciding whether a new treatment constitutes an improvement over some standard treatment. The problem of comparing two treatments is divided into two categories: the one-sample problem and the two-sample problem. In the two-sample problem, a random sample is drawn for each of two treatments. In the one-sample case, a random sample of paired comparisons,

* Corresponding author. Tel.: +1-410-516-7002; fax: +1-410-516-7459.

E-mail addresses: jingdong.xie@bms.com (J. Xie), cep@jhu.edu (C.E. Priebe).

0378-3758/02/$ - see front matter
PII: S0378-3758(01)00111-2

some of which may be positive and some negative, is available. Among the classical statistics for testing the difference of means in these two problems are the Mann–Whitney–Wilcoxon (MWW) statistic (Wilcoxon, 1945; Mann and Whitney, 1947) and the Wilcoxon signed rank (WSR) statistic (Wilcoxon, 1945). The establishment of these statistical procedures marks the beginning of modern nonparametric statistics.

These two statistics have gained widespread acceptance due to the weak assumptions required for their validity; however, they are by no means the preferred methods in all situations. For each, we can identify distributions under which the efficacy of the corresponding test is inferior to that of the best parametric tests. (Herein, we use Pitman asymptotic efficacy (*PAE*) to evaluate the performance of test statistics.) For example, if the underlying distribution is normal, the Student t-test is superior to the MWW statistic in the sense of *PAE* and the paired t-test outperforms the WSR statistic. Hence, developing new nonparametric test procedures which improve efficacy with respect to specific distributions is a desirable goal. In this paper, we generalize both the MWW statistic and the WSR statistic towards this end.

The MWW statistic is used to detect stochastic ordering of two populations on the basis of two independent samples. Let $X_1,\dots,X_n$ and $Y_1,\dots,Y_m$ be (jointly) independent random samples from distribution functions $F(x)$ and $G(y)$, respectively. The problem of interest is to test $\mathrm{H}_0: F(x)=G(x)$ for every x versus $\mathrm{H}_\mathrm{A}: F(x)\geqslant G(x)$ for every x, with strict inequality for at least one x (stochastic ordering). It is important to note that $F(x)\geqslant G(x)$ is equivalent to $u\{F(X)\}\geqslant u\{G(X)\}$ where the real-valued function $u(x)$ defined on $[0,1]$ is continuous and strictly increasing in x. The MWW statistic is based on the functional

$$\delta=\int_{-\infty}^{\infty} F(x)\,\mathrm{d}G(x),$$

which is equal to $1/2$ under H_0 and larger than $1/2$ under H_A. Thus large values of δ mean that H_A is true. In fact, the parameter δ is but one member of a class of such parameters; namely, let v be an arbitrary increasing, continuous and real-valued function on $[0,1]$ and define

$$\delta^{(u,v)}=\int_{-\infty}^{\infty} u\{F(x)\}\,\mathrm{d}v\{G(x)\}. \tag{1}$$

Note that under H_0, $\delta^{(u,v)}=\int_0^1 u(x)\,\mathrm{d}v(x)\equiv\delta_0^{(u,v)}$, and $\delta^{(u,v)}>\delta_0^{(u,v)}$ under H_A. This functional $\delta^{(u,v)}$ provides the framework for a generalization of the MWW statistic.

The WSR statistic is developed to detect the center of a symmetric distribution. Let $X_1,X_2,\dots,X_n$ be independent random variables with continuous c.d.f. F. If the distribution F is symmetric about a point θ such that for each x, $F(x+\theta)+F(-x+\theta)=1$, then θ is referred to as the center of the distribution. This is equivalent to saying that $u\{F(x+\theta)\}=u\{\bar{F}(-x+\theta)\}$, where u is an arbitrary strictly increasing, continuous, real-valued function on $[0,1]$ and $\bar{F}(x)=1-F(x)$. Consider the problem of testing that the center of symmetry, θ, is 0 against $\theta\neq 0$. The Wilcoxon signed rank statistic used for this purpose is a nonparametric estimator of the functional $\gamma=\int_{-\infty}^{\infty}\bar{F}(-x)\,\mathrm{d}F(x)$.

It is natural to consider a general functional $\gamma^{(u,v)}$ defined by

$$\gamma^{(u,v)} = \int_{-\infty}^{\infty} u\{\bar{F}(-x)\}\, \mathrm{d}v\{F(x)\}, \tag{2}$$

where functions v and u on $[0,1]$ are strictly increasing and continuous. Without loss of generality, under $\mathrm{H_A}: \theta > 0$,

$$\gamma^{(u,v)} = \int_{-\infty}^{\infty} u\{F(x+2\theta)\}\, \mathrm{d}v\{F(x)\}$$

is larger than under H_0, where we have $\gamma_0^{(u,v)} \equiv \int_{-\infty}^{\infty} u\{F(x)\}\, \mathrm{d}v\{F(x)\} = \int_0^1 u(x)\, \mathrm{d}v(x)$. Thus the associated test rejects for large values of $\gamma^{(u,v)}$. This functional $\gamma^{(u,v)}$ is the cornerstone for our generalization of the WSR statistic.

The following two lemmas provide the framework for our generalization of the MWW statistic based upon (1) and (2). (See the appendix for detailed proofs.)

Lemma 1.1. *If* $u(x)$ *is a strictly increasing continuous function on* $[0,1]$, *then there always exists a positively weighted tail binomial polynomial, i.e.*

$$u_r(x) = u(0) + \{u(1) - u(0)\} \sum_{k=1}^{r} w_k b_{k:r}(x),$$

where the polynomial $b_{k:r}(x) = \sum_{i=k}^{r} \binom{r}{i} x^i (1-x)^{r-i}$ *is called tail binomial polynomial with degree* r, $\sum_{k=1}^{r} w_k = 1$ *and* $w_k \geqslant 0$, $k = 1, \ldots, r$, *such that* $u_r(x)$ *converges to* $u(x)$ *uniformly on* $[0,1]$ *as* $r \to \infty$.

Notice that under the null hypothesis, both (1) and (2) are reduced to

$$J^{(u,v)} = \int_0^1 u(x)\, \mathrm{d}v(x).$$

Lemma 1.1 suggests we consider estimating the functionals

$$J^{(u_r,v_s)} = \int_0^1 u_r(x)\, \mathrm{d}v_s(x),$$

where $u_r(x)$ and $v_s(x)$ are the positively weighted tail binomial polynomials for approximating the functions $u(x)$ and $v(x)$ with maximum degrees r and s, respectively.

One question that arises naturally is whether $J^{(u_r,v_s)}$ converges to $J^{(u,v)}$ as r and s tend to infinity; namely whether the set of functionals $J^{(u_r,v_s)}$ is dense in the set of functionals $J^{(u,v)}$, where u, u_r, v, v_s are defined as above. The following lemma gives rise to a positive answer.

Lemma 1.2. *Suppose that* $u_r(x)$ *and* $v_s(x)$ *as defined in Lemma* 1.1 *are the approximations of strictly increasing continuous functions* $u(x)$ *and* $v(x)$ *on* $[0,1]$, *respectively. The set of functionals* $J^{(u_r,v_s)}$ *is dense in the set of functionals* $J^{(u,v)}$ *if the derivative of* $v(x)$ *exists and is continuous.*

Remarks: It is sensible to consider an empirical estimate based upon the functional approximation, i.e.

$$J^{(u_r,v_s)} = \int_0^1 u_r(x)\,\mathrm{d}v_s(x)$$

or simply

$$\begin{aligned} J^{(\pi,\mu)} &= \int_0^1 \sum_{k=1}^{r} \pi_k b_{k:r}(x)\,\mathrm{d}\left\{\sum_{l=1}^{s} \mu_l b_{l:s}(x)\right\} \\ &= \int_0^1 \sum_{k=1}^{r} \pi_k \sum_{i=k}^{r} \binom{r}{i} x^i(1-x)^{r-i}\,\mathrm{d}\sum_{k=1}^{s}\mu_l\sum_{i=l}^{s}\binom{s}{i}x^i(1-x)^{s-i}, \end{aligned} \tag{3}$$

where $\sum_{k=1}^{r} \pi_k = 1$, $\pi_k \geqslant 0$, $k = 1,\ldots,r$ and $\sum_{l=1}^{s} \mu_l = 1$, $\mu_l \geqslant 0$, $l = 1,\ldots,s$, for the weight vectors $\pi = (\pi_1,\ldots,\pi_r)$ and $\mu = (\mu_1,\ldots,\mu_s)$. In other words, if we consider only the strictly increasing functions $u(x)$ and $v(x)$ such that $u(0) = v(0) = 0$ and $u(1) = v(1) = 1$, $J^{(\pi,\mu)}$ is a functional of great interest. For simplicity, through the remainder of this paper, $u(x)$ and $v(x)$ are defined as above, unless otherwise stated.

Choosing strictly increasing and continuous functions u and v in both (1) and (2) reduces to choosing the weight vectors $\pi=(\pi_1,\ldots,\pi_r)$ and $\mu=(\mu_1,\ldots,\mu_s)$ in (3) due to Lemmas 1.1 and 1.2. If we choose $\pi=(\underbrace{\overbrace{0,\ldots,0,1}^{k},0,\ldots,0}_{r})$ and $\mu=(\underbrace{\overbrace{0,\ldots,0,1}^{l},0,\ldots,0}_{s})$ in (3), i.e. $u(x)=b_{k:r}(x)$ and $v(x)=b_{l:s}(x)$, then the U-statistic empirical estimates of $\delta^{(u,v)}$ and $\gamma^{(u,v)}$ are order statistic-based subsample generalizations of the Mann–Whitney–Wilcoxon statistic and the Wilcoxon signed rank statistic, respectively (Xie and Priebe, 2000). The generalized MWW statistics (GMWW) include the MWW statistic, the subsample median statistic (Shetty and Govindarajulu, 1988; Kumar, 1997), the subsample maxima statistic (Kochar, 1978; Deshpande and Kochar, 1980; Stephenson and Ghosh, 1985; Ahmad, 1996; Adams et al., 2000) and the subsample minima statistic (Priebe and Cowen, 1999). Moreover, the subsample statistics developed by Mathisen (1943) and Shetty and Bhat (1994) are special cases of the GMWW statistics. On the other hand, the generalized WSR (GWSR) statistic includes the WSR statistic and Maesono–Ahmad subsample statistic. In light of (3), the U-statistic empirical estimates of $\delta^{(u,v)}$ and $\gamma^{(u,v)}$, where $u(x) = \sum_{k=1}^{r} \pi_k b_{k:r}(x)$ and $v(x) = \sum_{k=1}^{s} \mu_l b_{l:s}(x)$, are hereby dubbed the weighted GMWW (WGMWW) and weighted GWSR (WGWSR) statistics, respectively. Apparently, the GMWW and GWSR statistics are special cases of the WGMWW and WGWSR statistics. This further development to a weighted version of GMWW (WGMWW) or GWSR (WGWSR) completes the generalization along the lines of subsample statistics. Thus, it would be of great interest to investigate the admissibility of the WGMWW (WGWSR) statistic with respect to the GMWW (GWSR) statistic in terms of *PAE*.

There are many nonparametric tests available in the literature for both the one-sample and two-sample problems in addition to those presented above. For example, there is a class of locally most powerful linear rank tests (Chapters 9.1 and 10.2 of Randles and Wolfe (1979)). If the optimal score function $\phi(x, f)$ for the two-sample case or $\phi^+(x, f)$ for the one-sample case (f is the density function of the distribution F), i.e.

$$\phi(x, f) = \frac{-f'\{F^{-1}(x)\}}{f\{F^{-1}(x)\}}$$

or

$$\phi^+(x, f) = \frac{-f'\{F^{-1}(x/2 + 1/2)\}}{f\{F^{-1}(x/2 + 1/2)\}},$$

is a strictly increasing function, then the optimal score functional $\int_0^1 \phi(x, f)\,\mathrm{d}x$ or $\int_0^1 \phi^+(x, f)\,\mathrm{d}x$ (the functional associated with the two-sample or one-sample optimal linear rank statistics) is a special case of (1) or (2), indicating that under H_0, Lemmas 1.1 and 1.2 are applicable. Especially, when $\int_0^1 \phi(x, f)\,\mathrm{d}x$ or $\int_0^1 \phi^+(x, f)\,\mathrm{d}u$ is too complicated to be used to construct the optimal linear rank statistic, the empirical estimate based upon the functional approximation (3) is a good alternative.

The asymptotic distribution of the WGMWW statistics is derived in Section 2. Section 3 is devoted to demonstrating the admissibility of the WGMWW statistic with respect to the GMWW statistic in the sense of Pitman efficacy. Precisely, given any continuous underlying distributions F and G, the PAE-optimal weighted generalization of the MWW statistic outperforms the conventional nonparametric subsample statistics, including the MWW statistic itself. An analogous treatment of the WGWSR statistic is given in Section 4.

2. The weighted GMWW statistic

A new class of statistics is introduced for testing stochastic ordering between two independent distributions. This class includes as a special case the GMWW statistic (Xie and Priebe, 2000). This new class is shown to be asymptotically normal under the null and alternative hypotheses. It is distribution-free.

Suppose that $X_1, \ldots, X_r$ $(Y_1, \ldots, Y_s)$ are r (s) independent copies of X (Y) and $X_{1:r}, \ldots, X_{r:r}$ $(Y_{1:s}, \ldots, Y_{s:s})$ are the order statistics obtained arranging the preceding random sample in increasing order of magnitudes. Suppose that $\pi = (\pi_1, \ldots, \pi_r)$ and $\mu = (\mu_1, \ldots, \mu_s)$ are weight vectors, where $\sum_{k=1}^r \pi_k = 1$, $\pi_k \geqslant 0$, $k = 1, \ldots, r$, and $\sum_{l=1}^s \mu_l = 1$, $\mu_l \geqslant 0$, $l = 1, \ldots, s$. Define

$$\begin{aligned}
\delta^{(\pi,\mu)} &= \int_{-\infty}^{\infty} \sum_{k=1}^r \pi_k b_{k:r}\{F(x)\}\,\mathrm{d}\left[\sum_{l=1}^s \mu_l b_{l:s}\{G(x)\}\right] \\
&= \int_{-\infty}^{\infty} \sum_{k=1}^r \pi_k F_{k:r}(x)\,\mathrm{d}\left\{\sum_{l=1}^s \mu_l G_{l:s}(x)\right\}
\end{aligned}$$

$$= \sum_{k=1}^{r} \sum_{l=1}^{s} \pi_k \mu_l \int_{-\infty}^{\infty} F_{k:r}(x)\, \mathrm{d}G_{l:s}(x)$$

$$= \sum_{k=1}^{r} \sum_{l=1}^{s} \pi_k \mu_l \Pr\{X_{k:r} < Y_{l:s}\},$$

where $F_{k:r}(x) = \sum_{i=k}^{r} \binom{r}{i} F^i(x)\{1-F(x)\}^{r-i}$ and $G_{l:s}(x) = \sum_{i=l}^{s} \binom{s}{i} G^i(x)\{1-G(x)\}^{s-i}$. Under H_0,

$$\delta^{(\pi,\mu)} = \sum_{k=1}^{r} \sum_{l=1}^{s} \pi_k \mu_l \sum_{i=k}^{r} \frac{\binom{l+i-1}{i}\binom{r+s-i-l}{r-i}}{\binom{r+s}{r}} \equiv \delta_0^{(\pi,\mu)},$$

while $\delta^{(\pi,\mu)} > \delta_0^{(\pi,\mu)}$ under H_A.

Theorem 2.1. *An empirical estimate of $\delta^{(\pi,\mu)}$ is given by*

$$\delta_{n,m}^{(\pi,\mu)} = \left\{ \binom{n}{r}\binom{m}{s} \right\}^{-1} \sum_{C} \sum_{k=1}^{r} \sum_{l=1}^{s} \pi_k \mu_l I\{X_{k:r}(X_{i_1},\ldots,X_{i_r}) < Y_{l:s}(Y_{j_1},\ldots,Y_{j_s})\}$$

$$= \frac{\sum_{k=1}^{r} \sum_{l=1}^{s} \sum_{i=k}^{n-r+k} \sum_{j=l}^{m-s+l} \pi_k \mu_l \binom{i-1}{k-1}\binom{n-i}{r-k}\binom{j-1}{l-1}\binom{m-j}{s-l} I(X_{i:n} < Y_{j:m})}{\binom{n}{r}\binom{m}{s}}$$

where $\sum_C$ extends over all indices $1 \leqslant i_1 < \cdots < i_r \leqslant n$, $1 \leqslant j_1 < \cdots < j_s \leqslant m$, r and s are fixed, $r \ll n$, $s \ll m$ and where $X_{1:n},\ldots,X_{n:n}$ $(Y_{1:m},\ldots,Y_{m:m})$ denote the order statistics of the random sample $X_1,\ldots,X_n$ $(Y_1,\ldots,Y_m)$ to distinguish these order statistics from the subsample order statistics $X_{k:r}(Y_{l:s})$.

The proof of Theorem 2.1 follows immediately from Theorem 1 of Xie and Priebe (2000).

In this section, we shall study the asymptotic behavior of the weighted GMWW statistic $\delta_{n,m}^{(\pi,\mu)}$. For ease of notation, we define $F_{k:r}(x) = 1$ or 0 as $k = 0$ or $k > r$ and $G_{l:s}(y) = 1$ or 0 as $l = 0$ or $l > s$.

Theorem 2.2. *Let $n \to \infty$ and $m \to \infty$ such that $(n/(n+m)) \to \lambda \in (0,1)$. Then $\sqrt{m+n}(\delta_{n,m}^{(\pi,\mu)} - \delta^{(\pi,\mu)})$ is asymptotically normal with mean 0 and variance $\sigma_{\pi,\mu}^2$, where*

$$\sigma_{\pi,\mu}^2 = \frac{r^2}{\lambda} \operatorname{var}\left\{ \int_{-\infty}^{X_1} \sum_{k=1}^{r} \sum_{l=1}^{s} \pi_k \mu_l (F_{k:r-1} - F_{k-1:r-1})(x)\, \mathrm{d}G_{l:s}(x) \right\}$$

$$+ \frac{s^2}{(1-\lambda)} \operatorname{var}\left\{ \int_{-\infty}^{Y_1} \sum_{k=1}^{r} \sum_{l=1}^{s} \pi_k \mu_l (G_{l:s-1} - G_{l-1:s-1})(x)\, \mathrm{d}F_{k:r}(x) \right\}.$$

Under H_0, *the variance* $\sigma^2_{\pi,\mu}$ *reduces to* $\xi/(\lambda(1-\lambda))$, *where*

$$\xi = r^2 \sum_{k=1}^{r}\sum_{l=1}^{s}\sum_{i=1}^{r}\sum_{j=1}^{s}\pi_k\pi_i\mu_l\mu_j \frac{\binom{k+l-2}{l-1}\binom{r+s-k-l}{r-k}}{\binom{r+s-1}{s}}$$

$$\times \frac{\binom{i+j-2}{i-1}\binom{r+s-i-j}{r-i}}{\binom{r+s-1}{s}} \sum_{p=k+l-1}^{r+s-1}\sum_{q=i+j-1}^{r+s-1}\binom{r+s-1}{p}\binom{r+s-1}{q}$$

$$\times \frac{(p+q)!(2r+2s-p-q-2)!}{(2r+2s-1)!}$$

$$-\left\{\frac{\sum_{k=1}^{r}\sum_{l=1}^{s}\pi_k\mu_l\binom{k+l-2}{k-1}\binom{r+s-k-l+1}{r-k}(r-l+1)}{\binom{r+s}{s}}\right\}^2.$$

For the detailed proof of Theorem 2.2, see the appendix.

Remarks:

1. The GMWW statistic (Xie and Priebe, 2000) is simply $\delta_{n,m}^{(\pi,\mu)}$ with $\pi = (\underbrace{\overbrace{0,\ldots,0,1}^{k},0,\ldots,0}_{r})$ and $\mu = (\underbrace{\overbrace{0,\ldots,0,1}^{l},0,\ldots,0}_{s})$, defined as $\mathrm{GMWW}_{n,m}^{(k:r,l:s)}$. One noteworthy fact is that $\delta^{(\pi',\mu')}$ with the uniform weights $\pi' = (\overbrace{1/r,\ldots,1/r}^{r})$ and $\mu' = (\overbrace{1/s,\ldots,1/s}^{s})$ is identical to the MWW statistic.
2. Theorems 2.1 and 2.2 provide a feasible way to compute $\delta_{n,m}^{(\pi,\mu)}$ and $\sigma^2_{\pi,\mu}$ using modern statistical software. It is easy to calculate the weighted GMWW statistic and its p-value, provided that sample sizes (n,m) and subsample sizes (r,s) are not impractically large.

3. Pitman asymptotic efficacy

In this section the focus is primarily on showing the admissibility of the weighted GMWW statistic relative to the GMWW statistic through theory and examples, both of which shed light onto the relationship between these two classes of statistics. Pitman asymptotic efficacy (PAE) is used for the purpose of comparing two test procedures. The inclusion of the class of the GMWW statistics within the class of the weighted

GMWW statistics as stated in Theorem 3.1 implies that the first class is at most as good as the second in the sense of the maximum *PAE* possible within each class. The efficacy analysis result presented in this investigation indicates nontrivial gain in efficacy using the optimal WGMWW statistics in the sense of *PAE* when the underlying densities are far from unimodal.

Pitman asymptotic efficacy (*PAE*) is defined in our case as

$$\mathrm{PAE}\,(\delta^{(\pi,\mu)}) = \left[\left\{ \frac{\mathrm{d}}{\mathrm{d}\theta} \delta_\theta^{(\pi,\mu)} \right\}^2 \Bigg|_{\theta\to\theta_0} \right] \Bigg/ \sigma^2_{\pi,\mu},$$

where $F = F_{\theta_0}$ and $G = F_\theta$, with $\theta = \theta_0 + c/n^{1/2}$ and where $\sigma^2_{\pi,\mu}$ is the null variance given in Theorem 2.2. Applying this definition we see that for the location problem $G(x) = F(x-\theta)$, under the assumption the density of F exists and is equal to f,

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\delta_\theta^{(\pi,\mu)}\Bigg|_{\theta\to 0} = \sum_{k=1}^{r}\sum_{l=1}^{s} \frac{\pi_k \mu_l r!s!}{(k-1)!(r-k)!(l-1)!(s-l)!} \times \int_{-\infty}^{\infty} F^{k+l-2}(1-F)^{r+s-k-l} f^2(x)\,\mathrm{d}x.$$

Apparently, the explicit expressions for the *PAE* are not easy to derive in general due to the complexity of distribution functions. Hence we study the *PAE*s for various canonical types of distributions and therefore indicate the strengths and weaknesses of the WGMWW statistics. For simplicity, *PAE*s presented here are scaled by the constant $\lambda(1-\lambda)$.

Twenty typical densities have been carefully chosen to investigate the overall performance of the WGMWW statistics in terms of *PAE*. Standard distributions considered here are the Uniform, Exponential, Logistic, Double Exponential and Cauchy distributions. In addition, a test suite of 15 normal mixture densities (Marron and Wand, 1992) are included to represent a wide range of density shape. (Any density can be approximated arbitrarily closely in various senses by a normal mixture.) Density numbers 1–10 represent unimodal densities. The rest of the densities are multimodal. Density numbers 1–6 are symmetric and unimodal, ranging from the heavy tail to the light tail while the skewness of the unimodal densities are increasing from number 7 to number 10. On the other hand, the density numbers 11–15 are *mildly multimodal*. The remaining densities are *strongly multimodal*. We believe these densities effectively model many real data situations.

Let us denote by $C_{r,s}$ the class of two-sample GMWW statistics with subsample sizes r and s and denote by $WC_{r,s}$ the class of two-sample weighted GMWW statistics with subsample sizes r and s. Clearly the fact that $C_{r,s} \subset WC_{r,s}$ leads to the conclusion that the maximum *PAE* of the GMWW statistic in $C_{r,s}$ is not greater than that in $WC_{r,s}$. To prove the admissibility of the weighted GMWW statistic with respect to the GMWW statistic, we need to show that $C_{r,s} \subset WC_{r',s'}$, whenever $r \leqslant r'$ and $s \leqslant s'$ and hence the maximum *PAE* of the WGMWW statistic in $WC_{r',s'}$ is not less than that in $WC_{r,s}$. Moreover, at least one member in $WC_{r',s'}$ has strictly larger *PAE* than does any member in $WC_{r,s}$.

Theorem 3.1. *The class of the two-sample GMWW statistics with subsample sizes r_1 and s_1, i.e. C_{r_1,s_1}, belongs to the class of the two-sample weighted GMWW statistics with subsample sizes r_2 and s_2, i.e. WC_{r_2,s_2}, whenever $r_1 \leqslant r_2$ and $s_1 \leqslant s_2$, that is, $C_{r_1,s_1} \subset WC_{r_2,s_2}$. Furthermore, $WC_{r_1,s_1} \subset WC_{r_2,s_2}$.*

The proof of Theorem 3.1 is provided in the appendix.

Remarks:

1. This theorem implies that the maximum *PAE* for the class $WC_{r,s}$ is monotone in terms of r and s.
2. One interesting identity worth mentioning is the following:

$$\binom{r+1}{k} F_{k:r} = \binom{r}{k} F_{k:r+1} + \binom{r}{k-1} F_{k+1:r+1}. \tag{4}$$

(The proof of identity (4) is given in the appendix.) This key identity allows us to conclude

$$\binom{r+1}{k} \delta^{(k:r,\,l:s)} = \binom{r}{k} \delta^{(k:r+1,\,l:s)} + \binom{r}{k-1} \delta^{(k+1:r+1,\,l:s)}$$

and

$$\binom{s+1}{l} \delta^{(k:r,\,l:s)} = \binom{s}{l} \delta^{(k:r,\,l:s+1)} + \binom{s}{l-1} \delta^{(k:r,\,l+1:s+1)}.$$

Similarly, the analog of Theorem 3.1 is valid for the functionals of the GMWW and WGMWW statistics.

The question arising naturally now is how to determine the optimal weighted GMWW statistic in $WC_{r,s}$ in terms of *PAE*. The *PAE-optimal weighted GMWW statistic* in $WC_{r,s}$ is defined as the weighted GMWW statistic in $WC_{r,s}$ with the maximum *PAE*. Thus it is clear that the problem of finding the optimal weighted GMWW statistic reduces to that of choosing the weights $\pi_i, \mu_j,\ i=1,\ldots,r,\ j=1,\ldots,s$, to maximize the *PAE* in $WC_{r,s}$. The following theorem details this approach, which plays a key role in finding a *PAE*-optimal WGMWW statistic.

Theorem 3.2. *The PAE-optimal weighted GMWW statistic in $WC_{r,s}$ is a weighted GMWW statistic with the weight vectors π and μ, which are a solution of the* *nonlinear programming problem with linear constraints as follows*:

$$\max \frac{\varphi^2(r,s,\pi,\mu)}{\xi(r,s,\pi,\mu)}$$

$$\text{s.t.} \begin{cases} \sum_{i=1}^{r} \pi_i = 1, \\ \sum_{j=1}^{s} \mu_j = 1, \\ \pi_i \geqslant 0, \quad i = 1,\dots r, \\ \mu_j \geqslant 0, \quad j = 1,\dots s, \end{cases} \tag{5}$$

where $\xi(r,s,\pi,\mu)$ *is the null variance given in Theorem* 3.1 *and*

$$\varphi(r,s,\pi,\mu) = \sum_{k=1}^{r} \sum_{l=1}^{s} \frac{r!s!\pi_k\mu_l \int_{-\infty}^{\infty} F^{k+l-2}(1-F)^{r+s-k-l} f^2(x)\,\mathrm{d}x}{(k-1)!(r-k)!(l-1)!(s-l)!}.$$

If $F(x)$ *is a cumulative distribution with finite Fisher information, then there exists a solution to* (5).

Remarks:

1. The proof follows easily from the fact that given the distribution function F, the *PAE* or the objective function in (5) is bounded by the Fisher information of F.
2. There are many nonlinear programming software packages containing rich maximization procedures which can be easily applied to choose the weights π and μ.
3. Given that the density function f and distribution function F are unknown, an adaptive procedure may be necessary. If large samples are available, a nonparametric estimate for f and an empirical distribution for F may be applicable. This yields an adaptive procedure (Hogg et al., 1975). In fact, one of the major advantages of the weighted GMWW statistic is to take advantage of relatively large samples in order to increase the PAE. Thus it is highly preferable in this scenario. However, application of an adaptive procedure in practice will not be addressed herein.

When the population distributions are assumed to differ only in location, the MWW test is directly comparable with the Student t-test which is known to be optimal with *PAE* of 1 under the assumption of normality. It is well known that if the population distributions are normal, the *PAE* of MWW is quite high at 0.955 (see Table 4 of Xie and Priebe, 2000). In fact, the overall *PAE*-optimal GMWW statistic is indeed the MWW statistic. It is no wonder that many statisticians thus far have considered the MWW test to be an appropriate nonparametric test for the two-sample location problem in the case of F normal. In the following example, we find that the WGMWW dominates the MWW for this case.

Example I. Assume $F(x)$ to be the normal distribution function (our density # 4). Applying Theorem 3.2 and using the computer code in Appendix A of Xie (1999) for solving (5) with $r=s=2$, we obtain the solution

$$\pi_1=0.5, \quad \pi_2=0.5, \quad \mu_1=0.5, \quad \mu_2=0.5,$$

which is in fact the classical MWW statistic. Notice that

$$0.5F_{1:2}(x)+0.5F_{2:2}(x)=F(x)$$

and hence that

$$\begin{aligned}\delta^{(0.5,0.5,0.5,0.5)} &= \int_{-\infty}^{\infty}\{0.5F_{1:2}(x)+0.5F_{2:2}(x)\}\,\mathrm{d}\{0.5G_{1:2}(x)+0.5G_{2:2}(x)\}\\ &= \int_{-\infty}^{\infty}F(x)\,\mathrm{d}G(x)\\ &= \delta^{(1,1)}.\end{aligned}$$

This is an illustrative example of Theorem 3.1, in which $C_{1,1}\subset WC_{2,2}$. Furthermore, it appears that the maximum *PAE* of $WC_{2,2}$ is equal to the maximum *PAE* of $C_{1,1}$, 0.955. On the other hand, the solution of (5) with $r=s=3$ is

$$\pi_1=0.5, \quad \pi_2=0, \quad \pi_3=0.5, \quad \mu_1=0.5, \quad \mu_2=0, \quad \mu_3=0.5$$

and thus the functional associated with this weighted GMWW statistic is

$$\delta^{(0.5,0,0.5,0.5,0,0.5)} = \int_{-\infty}^{\infty}\{0.5F_{1:3}(x)+0.5F_{3:3}(x)\}\,\mathrm{d}\{0.5G_{1:3}(x)+0.5G_{3:3}(x)\},$$

which yields the *PAE* of 0.9891. Clearly the maximum *PAE* of the WGMWW statistic in $WC_{3,3}$ is greater than that in $WC_{2,2}$. It is also greater than the maximum *PAE* in $C_{3,3}$ as shown in Table 1. Moreover, the maximum *PAE* for $WC_{r,s}$, $r=s=4,5,6$ are 0.9932, 0.9936, 0.9958, respectively. Experimentally, the maximum *PAE* for $WC_{r,s}$ increases as r and s grow. This observation not only conforms to Theorem 3.1 as expected, but also demonstrates the admissibility of $WC_{r,s}$ relative to $C_{r,s}$ in terms of *PAE*.

Example II. Let us consider the Bimodal distribution $F(x)$ (our density # 11). By solving (5) with $r=s=2,3,4,5,6$, we obtain the weight vectors as follows:

$$\pi_1=\mu_1=\pi_r=\mu_r=0.5 \quad \text{and} \quad \pi_i=\mu_i=0,\ i=2,\ldots,r-1.$$

Hence the associated functional is

$$\delta^{(\pi,\mu)} = \int_{-\infty}^{\infty}\{0.5F_{1:r}(x)+0.5F_{r:r}(x)\}\,\mathrm{d}\{0.5F_{1:r}(x)+0.5F_{r:r}(x)\}.$$

As displayed in Table 1, the maximum *PAE* for $WC_{r,s}$ is constantly increasing with subsample sizes r and s. Most importantly, the gain in efficacy using the *PAE*-optimal weighted GMWW statistic is substantial.

The *PAE*-optimal weighted GMWW statistic outperforms the *PAE*-optimal GMWW statistic, as shown from Table 1. Regardless of the underlying distribution, the

Table 1
Comparison of the WGMWW and GMWW statistics[a]

ARE(WGMWW, GMWW)							
#	$r=s=1$	$r=s=2$	$r=s=3$	$r=s=4$	$r=s=5$	$r=s=6$	Optimal(r,s)
#1	1	1	1	1	1	1	1
#2	1	1	1	1	1	1	1
#3	1	1.012	1.028	1.044	1.057	1.074	1
#4	1	1.054	1.133	1.168	1.215	1.216	1.042
#5	1	1	1	1	1	1	1
#6	1	1	1.006	1.021	1.036	1.048	1.007
#7	1	1	1	1	1	1	1
#8	1	1	1	1	1	1	1
#9	1	1.016	1.045	1.075	1.081	1.1	1.027
#10	1	1	1	1	1	1	1
#11	1	1.050	1.248	1.476	1.6	1.667	1.615
#12	1	1.007	1.193	1.440	1.585	1.664	1.635
#13	1	1	1	1	1.121	1.145	1.136
#14	1	1.024	1.217	1.451	1.588	1.667	1.657
#15	1	1.079	1.283	1.5	1.554	1.516	1.516
#16	1	1.049	1.249	1.439	1.594	1.674	1.626
#17	1	1.008	1.147	1.310	1.413	1.481	1.481
#18	1	1	1.107	1.297	1.418	1.449	1.425
#19	1	1	1.131	1.263	1.123	1.299	1.299
#20	1	1	1.163	1.389	1.497	1.532	1.532

[a] $\mathrm{ARE(WGMWW, GMWW)} = \frac{\max \mathrm{PAE(WGMWW} \in WC_{r,s})}{\max \mathrm{PAE(GMWW} \in C_{r,s})}$.

PAE-optimal WGMWW statistic always has a nondecreasing *PAE* in terms of subsample sizes r and s. In contrast, this is not generally true for the *PAE*-optimal GMWW statistic. As far as the unimodal densities number 1 to number 10 are concerned, it is clear that under the Gaussian, Outlier, Logistic and Skewed unimodal distributions, the *PAE*-optimal weighted GMWW statistic gains higher efficacy than its counterpart in the class of the GMWW statistics, i.e. the corresponding *PAE*-optimal GMWW statistic (Fig. 1). This indicates the admissibility of the WGMWW statistic relative to the GMWW statistic. In fact, the former dominates the latter; we observe that the *PAE*-optimal weighted GMWW statistic is at least as good as the *PAE*-optimal GMWW statistic for the remaining unimodal densities. As for *PAE* analyses of the multimodal densities displayed in Table 1 (density numbers 11–20), the *PAE*-optimal weighted GMWW statistic shows substantial improvement over the *PAE*-optimal GMWW statistic (Fig. 2). For example, for $r=s=6$, the *PAE*-optimal weighted GMWW statistic gives rise to almost 65.7% increase in efficacy over the *PAE*-optimal GMWW statistic for the density # 14. Overall, the superiority of the former over the latter therefore becomes apparent as anticipated from Theorem 3.1. It is highly recommended that the *PAE*-optimal WGMWW statistic be used, especially when the underlying distribution is multimodal, as opposed to the *PAE*-optimal GMWW statistic.

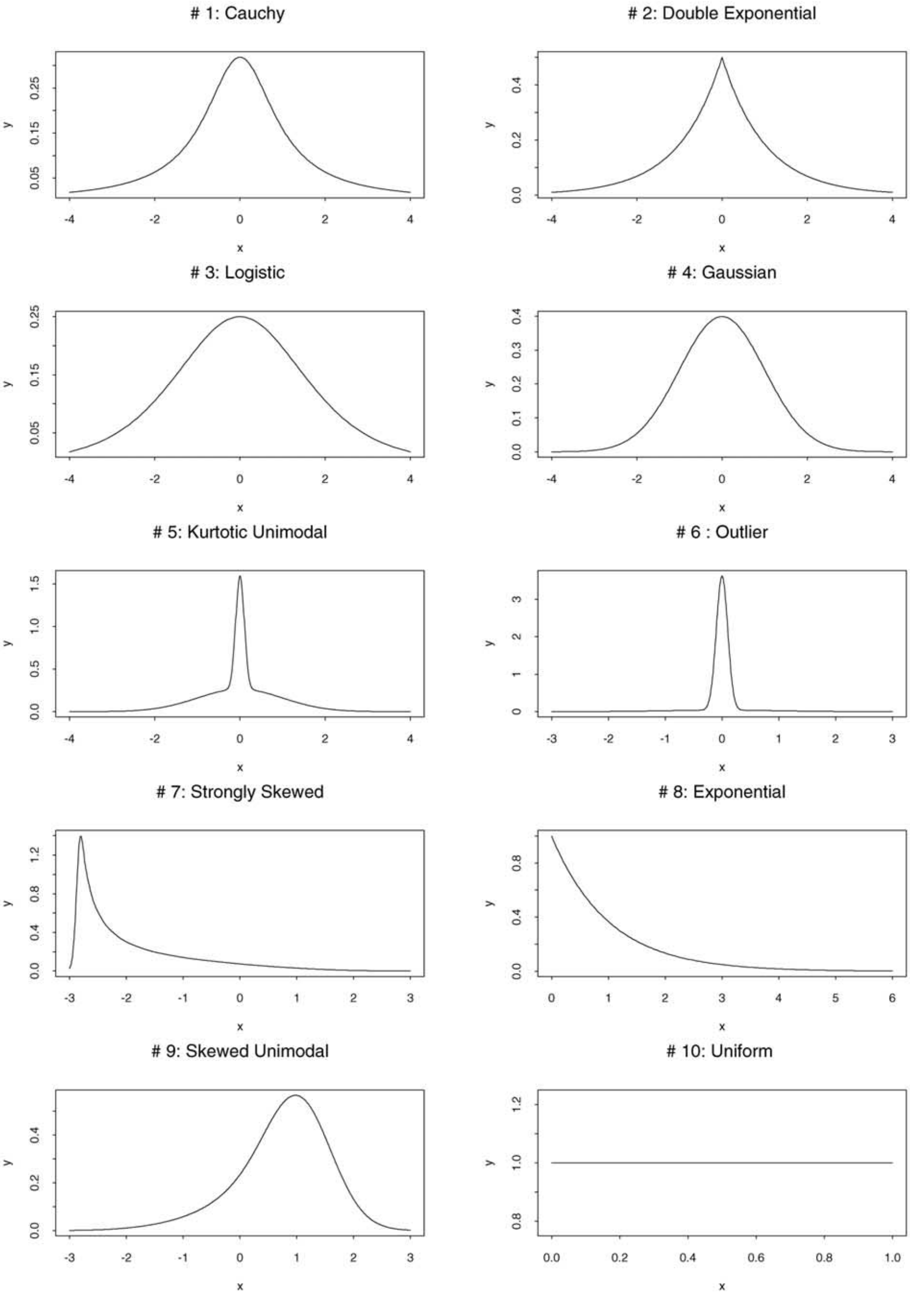

Fig. 1. Testing densities.

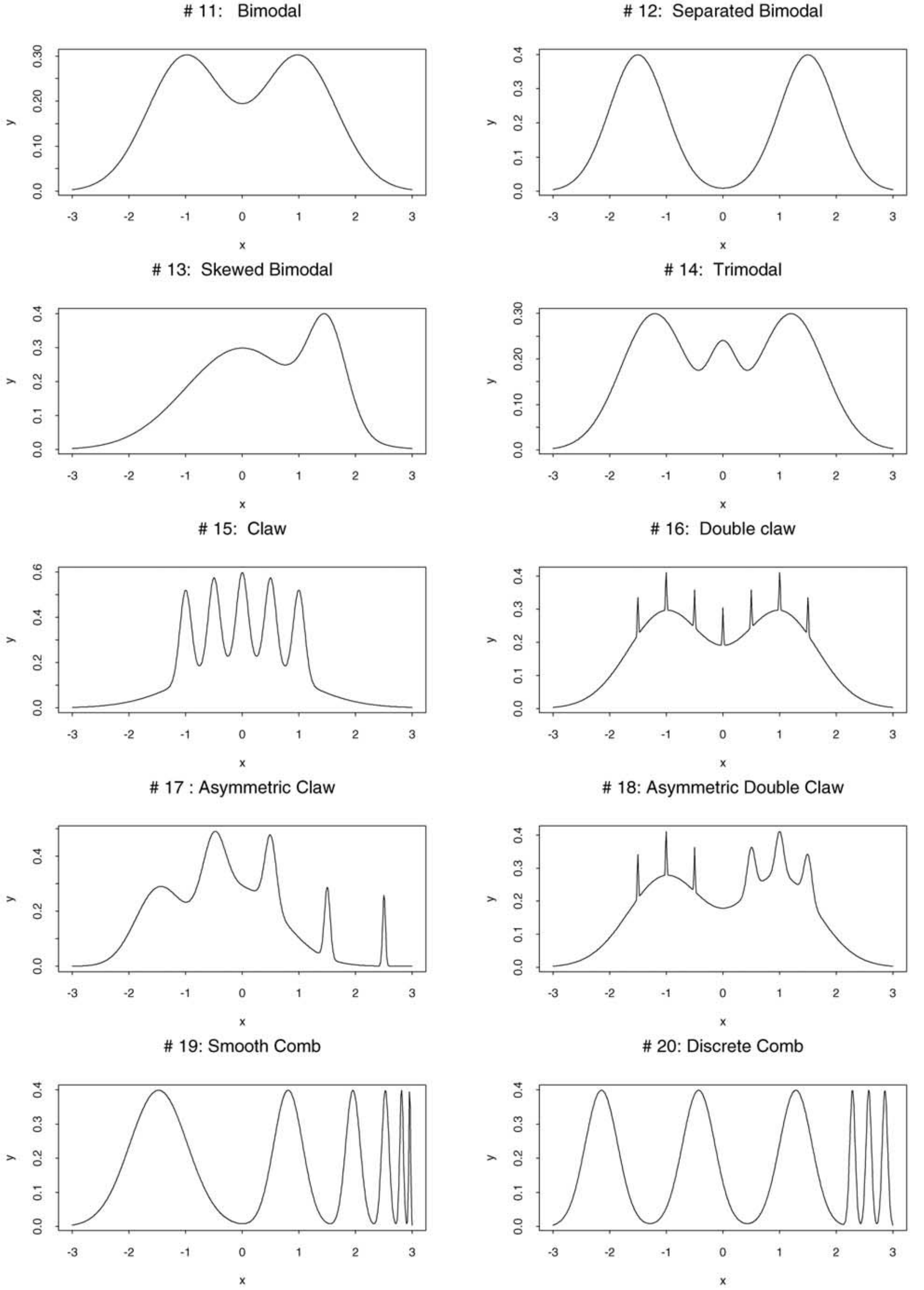

Fig. 2. Testing densities.

4. One-sample analog: a weighted generalization of the Wilcoxon signed rank statistic

Analogous to the two-sample case, a class of weighted generalizations of the Wilcoxon signed rank (WSR) statistics is introduced for testing the center of a symmetric distribution, including as special cases the Wilcoxon Signed Rank, the Maesono–Ahmad (Maesono, 1987; Ahmad, 1996) and the GWSR (Xie and Priebe, 2000) statistics.

As in the two-sample case, Lemmas 1.1 and 1.2 motivate us to estimate the functional $\gamma^{(\pi,\mu)}$ defined as

$$\begin{aligned}\gamma^{(\pi,\mu)} &= \int_{-\infty}^{\infty}\sum_{k=1}^{r}\pi_k b_{k:r}\{1-F(-x)\}\,\mathrm{d}\left[\sum_{l=1}^{s}\mu_l b_{l:s}\{F(x)\}\right]\\ &= \int_{-\infty}^{\infty}\sum_{k=1}^{r}\pi_k \bar{F}_{r-k+1:r}(-x)\,\mathrm{d}\left\{\textstyle\sum_{l=1}^{s}\mu_l F_{l:s}(x)\right\}\\ &= \sum_{k=1}^{r}\sum_{l=1}^{s}\pi_k\mu_l\int_{-\infty}^{\infty}\bar{F}_{r-k+1:r}(-x)\,\mathrm{d}F_{l:s}(x)\\ &= \sum_{k=1}^{r}\sum_{l=1}^{s}\pi_k\mu_l Pr\{X_{k:r}(X_1,\ldots,X_r)+X_{l:s}(X_{r+1},\ldots,X_{r+s})>0\},\end{aligned}$$

where $\bar{F}_{r-k+1:r}(-x)=\sum_{v=r-k+1}^{r}\binom{r}{v}\bar{F}^{v}(-x)(1-\bar{F}(-x))^{r-v}$ and $\pi=(\pi_1,\ldots,\pi_r)$ and $\mu=(\mu_1,\ldots,\mu_s)$ are the weight vectors and where $X_1,\ldots,X_r$ are r independent copies of X and $X_{1:r},\ldots,X_{r:r}$ are the order statistics obtained by arranging the preceding random sample in increasing order of magnitude. Notice that under H_0, $\gamma^{(\pi,\mu)}=\sum_{k=1}^{r}\sum_{l=1}^{s}\pi_k\mu_l\sum_{i=r-k+1}^{r}\binom{l+i-1}{i}\binom{s+r-i-l}{r-i}/\binom{r+s}{r}$.

An empirical estimate of $\gamma^{(\pi,\mu)}$ is given by

$$\begin{aligned}\gamma_n^{(\pi,\mu)} &= \left\{\binom{n}{r+s}\binom{r+s}{s}\right\}^{-1}\sum_{C}\sum_{k=1}^{r}\sum_{l=1}^{s}\\ &\quad \pi_k\mu_l I\{X_{k:r}(X_{i_1},\ldots,X_{i_r})+X_{l:s}(X_{i_{r+1}},\ldots,X_{i_{r+s}})>0\}\\ &= \frac{\sum_{k=1}^{r}\sum_{l=1}^{s}\sum_{i=k}^{n-r+k}\sum_{j=l}^{n-s+l}\pi_k\mu_l w_{ij} I(X_{i:n}+X_{j:n}>0)}{\binom{n}{r+s}\binom{r+s}{r}},\end{aligned}$$

where $\sum_c$ stands for summation over all permutations $(i_1,\ldots,i_{r+s})$ of $r+s$ integers such that $1\leqslant i_1<\cdots<i_r\leqslant n$, $1\leqslant i_{r+1}<\cdots<i_{r+s}\leqslant n$, and $i_e\neq i_f$ if $e\neq f$, $1\leqslant e\leqslant r$ and $1\leqslant f\leqslant s$, and where

$$w_{ij}=$$

$$\begin{cases}\sum_{v=0}^{l-1}\binom{i-1}{k-1}\binom{i-k}{v}\binom{j-i-1}{l-v-1}\binom{n-j}{s-l}\binom{n+v-i-s}{r-k} & i<j,\\ 0 & i=j,\\ \sum_{v=0}^{s-l}\binom{j-1}{l-1}\binom{i-j-l}{v}\binom{n-i-r+k}{s-l-v}\binom{n-i}{r-k}\binom{i-l-v-1}{k-l} & i>j.\end{cases}$$

The GWSR statistic (Xie and Priebe, 2000), including the WSR and Maesono–Ahmad statistics, is simply $\gamma_n^{(\pi,\mu)}$ with $\pi = (\underbrace{\overbrace{0,\ldots,0,1}^{k},0,\ldots,0}_{r})$ and $\mu = (\underbrace{\overbrace{0,\ldots,0,1}^{l},0,\ldots,0}_{s})$. Likewise, $\gamma^{(\pi',\mu')}$ with the uniform weight vectors $\pi' = (\overbrace{1/r,\ldots,1/r}^{r})$ and $\mu' = (\overbrace{1/s,\ldots,1/s}^{s})$ is the WSR statistic because $\frac{1}{r}\sum_{k=1}^{r}\bar{F}_{k:r}(-x) = \bar{F}(-x)$ and $\frac{1}{s}\sum_{l=1}^{s}F_{l:s}(x) = F(x)$.

In this section, we shall study the asymptotic behavior of the statistic $\gamma_n^{(\pi,\mu)}$. For ease of illustration, we define $F_{k:r}(x) = 1$ or 0 as $k = 0$ or $k > r$.

Theorem 4.1. *Let $n \to \infty$. Then $\sqrt{n}\{\gamma_n^{(\pi,\mu)} - \gamma^{(\pi,\mu)}\}$ is asymptotically normal with mean* 0 *and variance $\alpha^2_{\pi,\mu}$, where*

$$\begin{aligned}\alpha^2_{\pi,\mu} = \operatorname{var}\Bigg[\sum_{k=1}^{r}\sum_{l=1}^{s}\pi_k\mu_l\Bigg\{ & r\int_{-X_1}^{\infty}\bar{F}_{r-k:r-1}(-x)\,\mathrm{d}F_{l:s}(x)\\ & + r\int_{-\infty}^{-X_1}\bar{F}_{r-k+1:r-1}(-x)\,\mathrm{d}F_{l:s}(x)\\ & + s\int_{-\infty}^{X_1}F_{l-1:s-1}(x)\,\mathrm{d}\bar{F}_{r-k+1:r}(-x)\\ & - s\int_{-\infty}^{X_1}F_{l:s-1}(x)\,\mathrm{d}\bar{F}_{r-k+1:r}(-x)\Bigg\}\Bigg]\end{aligned}$$

for $1 \leqslant k \leqslant r$, $1 \leqslant l \leqslant s$. Under H_0, *the variance $\alpha^2_{\pi,\mu}$ reduces to*

$$\begin{aligned}&\sum_{k=1}^{r}\sum_{l=1}^{s}\sum_{i=1}^{r}\sum_{j=1}^{s}\pi_k\mu_l\pi_i\mu_j\frac{r!s!(r+l-k-1)!(s+k-l-1)!}{(k-1)!(r+s-1)!(r-k)!(l-1)!(s-l)!}\\ &\quad\times\frac{r!s!(r+j-i-1)!(s+i-j-1)!}{(i-1)!(r+s-1)!(r-i)!(j-1)!(s-j)!}\\ &\quad\times\Bigg[1-\frac{(s+r-1)!}{(s-l+k-1)!(r+l-k-1)!}\sum_{i=r+j-i}^{s+r-1}\frac{\binom{s+r-1}{i}}{(2r+2s-1)!}\\ &\quad\times\{(2i+2l-2k-2s+1)(r+l+i-k-1)!(2s+r+k-l-i-2)!\\ &\quad+(2i+2k-2r-2l+1)(s-l+k+i-1)!(s+2r+l-i-k-2)!\}\\ &\quad-\frac{(s+r-1)!}{(s-j+i-1)!(r+j-i-1)!}\sum_{n=r+l-k}^{s+r-1}\frac{\binom{s+r-1}{n}}{(2r+2s-1)!}\end{aligned}$$

$$\times\{(2n+2j-2i-2s+1)(r+j+n-i-1)!(2s+r+i-j-n-2)!$$

$$\left.+(2n+2i-2r-2j+1)(s-j+i+n-1)!(s+2r+j-n-i-2)!\}\right].$$

Thus the proof of Theorem 4.1 is similar to that of Theorem 3.1.

Let us denote by $V_{r,s}$ the class of the one-sample GWSR statistics with subsample sizes r and s and denote by $WV_{r,s}$ the class of the one-sample weighted GWSR statistics with subsample sizes r and s. Apparently, $V_{r,s} \subset WV_{r,s}$. As in the two-sample case, we have the following theorem characterizing the relationship between the classes V and WV.

Theorem 4.2. *The class of the one-sample GWSR statistics with subsample sizes r_1 and s_1, i.e. V_{r_1,s_1}, is included in the class of the one-sample weighted GWSR statistics with subsample sizes r_2 and s_2, i.e. WV_{r_2,s_2}, whenever $r_1 \leqslant r_2$ and $s_1 \leqslant s_2$, that is, $V_{r_1,s_1} \subset WV_{r_2,s_2}$. Furthermore, $WV_{r_1,s_1} \subset WV_{r_2,s_2}$.*

The *PAE-optimal weighted GWSR statistic* in $WV_{r,s}$ is defined as the weighted GWSR statistic with the maximum efficacy in $WV_{r,s}$. The problem of finding the optimal weighted GWSR statistic reduces to that of choosing the weights π_i, μ_j, $i=1,\dots,r$, $j=1,\dots,s$ to maximize the *PAE* of the WGWSR statistic in $WV_{r,s}$.

Theorem 4.3. *Suppose that the cumulative distribution $F(x)$ is known. A PAE-optimal weighted GWSR statistic in $WV_{r,s}$ is the weighted GWSR statistic with weight vectors π and μ, which are a solution of the nonlinear programming problem with linear constraints as follows:*

$$\max \frac{\phi^2(r,s,\pi,\mu)}{\psi(r,s,\pi,\mu)}$$

$$\text{s.t.} \begin{cases} \sum_{i=1}^{r} \pi_i = 1, \\ \sum_{j=1}^{s} \mu_j = 1, \\ \pi_i \geqslant 0, \qquad i=1,\dots r, \\ \mu_j \geqslant 0, \qquad j=1,\dots s, \end{cases} \tag{6}$$

where $\psi(r,s,\pi,\mu)$ is the null variance given by Theorem 4.1 *and*

$$\phi(r,s,\pi,\mu) = \sum_{k=1}^{r}\sum_{l=1}^{s} \frac{2r!s!\pi_k\mu_l \int_{-\infty}^{\infty} F^{r-k+l-1}(1-F)^{s+k-l-1} f^2(x)\,\mathrm{d}x}{(k-1)!(r-k)!(l-1)!(s-l)!}.$$

If the Fisher information of $F(\cdot)$ is bounded, then there exists a solution to (6.15).

Proof. The proof follows from the similar argument in the proof of Theorem 3.2. □

Example I. Suppose that $F(x)$ is the normal distribution function. Applying Theorem 4.3 and using the computer code in Appendix B of Xie (1999) for solving (6), we obtain the maximum *PAEs* for $WV_{r,s}$, $r = s = 2, 3, 4, 5, 6$, which are 0.9760, 0.9896, 0.9932, 0.9937 and 0.9959, respectively. As subsample sizes r, s get large, the optimal *PAE* increases. This agrees with Theorem 4.2. Notice that for the normal distribution, the maximum achievable *PAE* is 1. Since the one-sample t-test with $PAE = 1$ is optimal for the normal distribution, the *PAE*-optimal WGWSR with $r = s = 6$ is a strong nonparametric competitor.

Appendix

Proof of Lemma 1.1. It is sufficient to consider the case in which $u(0)=0$ and $u(1)=1$. Suppose that $u(x)$ is a strictly increasing continuous function on $[0,1]$ and that $u(0)=0$ and $u(1) = 1$. In light of the Weierstrass approximation theorem (Theorem 1.1.1 of Lorentz (1986)), this implies that the Bernstein polynomials $u_r(x)$ of the function $u(x)$ converges to $u(x)$ uniformly on $[0,1]$. Moreover, $u_r(x)$ can be expressed in terms of the tail binomial polynomials $b_{k:r}(x)$, $k = 1, \ldots, r$. To see this, let us take a closer look at $u_r(x)$.

$$
\begin{aligned}
u_r(x) &= \sum_{k=0}^{r} u\left(\frac{k}{r}\right)\binom{r}{k} x^k (1-x)^{r-k} \\
&= u(1) b_{r:r}(x) + \sum_{k=0}^{r-1} u\left(\frac{k}{r}\right)\{b_{k:r}(x) - b_{k+1:r}(x)\} \\
&= u(0) b_{0:r}(x) + \sum_{k=1}^{r}\left\{u\left(\frac{k}{r}\right) - u\left(\frac{k-1}{r}\right)\right\} b_{k:r}(x) \\
&= \sum_{k=1}^{r}\left\{u\left(\frac{k}{r}\right) - u\left(\frac{k-1}{r}\right)\right\} b_{k:r}(x) \\
&= \sum_{k=1}^{r} w_k b_{k:r}(x),
\end{aligned}
$$

where $w_k = u(k/r) - u((k-1)/r) > 0$, $k = 1, \ldots, r$ and $\sum_{k=1}^{r} w_k = 1$. The proof is completed as desired. $\square$

Proof of Lemma 1.2. It is sufficient to show that

$$
\int_0^1 u_r(x)\,\mathrm{d}v_s(x) \to \int_0^1 u(x)\,\mathrm{d}v(x) \quad \text{as } r, s \to \infty \text{ and } r/s \to \lambda \in (0,1).
$$

Without loss of generality, we are assuming that $u(0)=0$, and hence $0 < u(1) < \infty$. The continuity of v', taken together with the Weierstrass approximation theorem, implies that the Bernstein polynomials of $v'(x)$, i.e. $v'_s(x)$, converge to $v'(x)$ uniformly on $[0,1]$. Thus, for any $\varepsilon > 0$, and all $x \in [0,1]$, there exists M' such that $|v'_s(x) - v'(x)| < \varepsilon/2u(1)$, whenever $s > M'$. Once again, it follows from the Weierstrass approximation theorem that $u_r(x)$ and $v_s(x)$ converge to $u(x)$ and $v(x)$ uniformly on $[0,1]$, respectively. This implies that for any $0<\varepsilon<1$ and all $x \in [0,1]$, there exist $N1$ and $M1$ such that $|u_r(x)-u(x)|<\varepsilon/2(2+|v(1)-v(0)|)$ and $|v_s(x)-v(x)|<\varepsilon$, whenever $r>N1$ and $s>M1$. All in all, for any $\varepsilon>0$, there exists $M=\max\{N1,M1,M'\}$ such that

$$\begin{aligned}
&\left|\int_0^1 u_r(x)\,\mathrm{d}v_s(x)-\int_0^1 u(x)\,\mathrm{d}v(x)\right|\\
&\quad=\left|\int_0^1 \{u_r(x)-u(x)\}\,\mathrm{d}v_s(x)+\int_0^1 u(x)\,\mathrm{d}\{v_s(x)-v(x)\}\right|\\
&\quad\leqslant\left|\int_0^1 \{u_r(x)-u(x)\}\,\mathrm{d}v_s(x)\right|+\left|\int_0^1 u(x)\,\mathrm{d}\{v_r(x)-v(x)\}\right|\\
&\quad\leqslant\int_0^1 |u_r(x)-u(x)|\,\mathrm{d}v_s(x)+\int_0^1 |u(x)\{v_s'(x)-v'(x)\}|\,\mathrm{d}x\\
&\quad\leqslant\frac{|v_s(1)-v_s(0)|\varepsilon}{2(2+|v(1)-v(0)|)}+|u(1)|\int_0^1 |v_s'(x)-v'(x)|\,\mathrm{d}x.
\end{aligned}$$

Hence the lemma follows as desired.

Proof of Theorem 2.2. Using Theorem 3.4.13 of Randles and Wolfe (1979), we need only to establish the variance. Let

$$\varphi^{(\pi,\mu)}(X_1,\ldots,X_r;Y_1,\ldots,Y_s)=\sum_{k=1}^{r}\sum_{l=1}^{s}\pi_k\mu_l I\{X_{k:r}(X_1,\ldots,X_r)<Y_{l:s}(Y_1,\ldots,Y_s)\}.$$

Thus

$$\begin{aligned}
\varphi_{10}^{(\pi,\mu)}(X_1)&=E[\varphi^{(\pi,\mu)}(X_1,\ldots,X_r;Y_1,\ldots,Y_s)|X_1]\\
&=\sum_{k=1}^{r}\sum_{l=1}^{s}\pi_k\mu_l\,\Pr\{X_{k:r}(X_1,\ldots,X_r)\leqslant Y_{l:s}(Y_1,\ldots,Y_s)|X_1\}\\
&=\sum_{k=1}^{r}\sum_{l=1}^{s}\pi_k\mu_l\Pr\{X_{k:r}\leqslant Y_{l:s}|X_1\}\\
&=\sum_{k=1}^{r}\sum_{l=1}^{s}\pi_k\mu_l\varphi_{10}^{(k:r,l:s)}(X_1).
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
\varphi_{01}^{(\pi,\mu)}(Y_1)&=E\{\varphi^{(\pi,\mu)}(X_1,\ldots,X_r;Y_1,\ldots,Y_s)|Y_1\}\\
&=\sum_{k=1}^{r}\sum_{l=1}^{s}\pi_k\mu_l\,\Pr\{X_{k:r}\leqslant Y_{l:s}|Y_1\}\\
&=\sum_{k=1}^{r}\sum_{l=1}^{s}\pi_k\mu_l\varphi_{01}^{(k:r,l:s)}(Y_1).
\end{aligned}$$

Under H_0,

$$
\begin{aligned}
E\varphi_{10}^{(\pi,\mu)}(X_1) &= E\varphi_{01}^{(\pi,\mu)}(Y_1)\\
&= \sum_{k=1}^{r}\sum_{l=1}^{s}\pi_k\mu_l E\varphi_{01}^{(k:r,l:s)}(Y_1)\\
&= \delta^{(\pi,\mu)}\\
&= \binom{r+s}{r}^{-1}\sum_{k=1}^{r}\sum_{l=1}^{s}\pi_k\mu_l\sum_{i=k}^{r}\binom{l+i-1}{i}\binom{r+s-i-l}{r-i}.
\end{aligned}
$$

$$
\begin{aligned}
\xi_{1,0} &= \mathrm{var}\{\varphi_{10}^{(\pi,\mu)}(X_1)\}\\
&= \mathrm{var}\left\{\sum_{k=1}^{r}\sum_{l=1}^{s}\pi_k\mu_l\int_{-\infty}^{X_1}(F_{k:r-1}-F_{k-1:r-1})(x)\,\mathrm{d}F_{l:s}(x)\right\}\\
&= E\left\{\sum_{k=1}^{r}\sum_{l=1}^{s}\pi_k\mu_l\int_{-\infty}^{X_1}(F_{k:r-1}-F_{k-1:r-1})(x)\,\mathrm{d}F_{l:s}(x)\right\}^2\\
&\quad - E^2\left\{\sum_{k=1}^{r}\sum_{l=1}^{s}\pi_k\mu_l\int_{-\infty}^{X_1}(F_{k:r-1}-F_{k-1:r-1})(x)\,\mathrm{d}F_{l:s}(x)\right\}\\
&= A-B^2,
\end{aligned}
$$

where

$$
\begin{aligned}
B &= E\left\{\sum_{k=1}^{r}\sum_{l=1}^{s}\pi_k\mu_l\int_{-\infty}^{X_1}(F_{k:r-1}-F_{k-1:r-1})(x)\,\mathrm{d}F_{l:s}(x)\right\}\\
&= \sum_{k=1}^{r}\sum_{l=1}^{s}\pi_k\mu_l E\left\{\int_{-\infty}^{X_1}(F_{k:r-1}-F_{k-1:r-1})(x)\,\mathrm{d}F_{l:s}(x)\right\}\\
&= \frac{\sum_{k=1}^{r}\sum_{l=1}^{s}\pi_k\mu_l\binom{k+l-2}{k-1}\binom{r+s-k-l+1}{r-k}(r-l+1)}{r\binom{r+s}{s}}
\end{aligned}
$$

and

$$
\begin{aligned}
A &= E\left\{\sum_{k=1}^{r}\sum_{l=1}^{s}\pi_k\mu_l\int_{-\infty}^{X_1}(F_{k:r-1}-F_{k-1:r-1})(x)\,\mathrm{d}F_{l:s}(x)\right\}^2\\
&= \sum_{k=1}^{r}\sum_{l=1}^{s}\sum_{i=1}^{r}\sum_{j=1}^{s}\pi_k\pi_i\mu_l\mu_j E\left\{\int_{-\infty}^{X_1}(F_{k:r-1}-F_{k-1:r-1})(x)\,\mathrm{d}F_{l:s}(x)\right.\\
&\quad\left.\times\int_{-\infty}^{X_1}(F_{i:r-1}-F_{i-1:r-1})(x)\,\mathrm{d}F_{j:s}(x)\right\}
\end{aligned}
$$

$$= \sum_{k=1}^{r}\sum_{l=1}^{s}\sum_{i=1}^{r}\sum_{j=1}^{s}\pi_k\pi_i\mu_l\mu_j \frac{\binom{k+l-2}{l-1}\binom{r+s-k-l}{r-k}}{\binom{r+s-1}{s}}$$

$$\times \frac{\binom{i+j-2}{i-1}\binom{r+s-i-j}{r-i}}{\binom{r+s-1}{s}}$$

$$\times \sum_{p=k+l-1}^{r+s-1}\sum_{q=i+j-1}^{r+s-1}\binom{r+s-1}{p}\binom{r+s-1}{q} E\{F^{p+q}(1-F)^{2r+2s-p-q-2}\}$$

$$= \sum_{k=1}^{r}\sum_{l=1}^{s}\sum_{i=1}^{r}\sum_{j=1}^{s}\pi_k\pi_i\mu_l\mu_j \frac{\binom{k+l-2}{l-1}\binom{r+s-k-l}{r-k}}{\binom{r+s-1}{s}}$$

$$\times \frac{\binom{i+j-2}{i-1}\binom{r+s-i-j}{r-i}}{\binom{r+s-1}{s}}$$

$$\times \sum_{p=k+l-1}^{r+s-1}\sum_{q=i+j-1}^{r+s-1}\binom{r+s-1}{p}\binom{r+s-1}{q} \frac{(p+q)!(2r+2s-p-q-2)!}{(2r+2s-1)!}.$$

Similarly

$$\begin{aligned}
\xi_{0,1} &= \mathrm{var}\{\varphi_{0,1}^{(\pi,\mu)}(Y_1)\} \\
&= \mathrm{var}\left\{\sum_{k=1}^{r}\sum_{l=1}^{s}\pi_k\mu_l \int_{-\infty}^{X_1}(F_{l:s-1}-F_{l-1:s-1})(x)\,\mathrm{d}F_{k:r}(x)\right\} \\
&= E\left\{\sum_{k=1}^{r}\sum_{l=1}^{s}\pi_k\mu_l \int_{-\infty}^{X_1}(F_{l:s-1}-F_{l-1:s-1})(x)\,\mathrm{d}F_{k:r}(x)\right\}^2 \\
&\quad - E^2\left\{\sum_{k=1}^{r}\sum_{l=1}^{s}\pi_k\mu_l \int_{-\infty}^{X_1}(F_{l:s-1}-F_{l-1:s-1})\,\mathrm{d}F_{k:r}(x)\right\} \\
&= C - D^2,
\end{aligned}$$

where

$$D = E\left\{\sum_{k=1}^{r}\sum_{l=1}^{s}\pi_k\mu_l\int_{-\infty}^{X_1}(F_{l:s-1}-F_{l-1:s-1})\,\mathrm{d}F_{k:r}(x)\right\}$$

$$= \sum_{k=1}^{r}\sum_{l=1}^{s}\pi_k\mu_l E\left\{\int_{-\infty}^{X_1}(F_{l:s-1}-F_{l-1:s-1})\,\mathrm{d}F_{k:r}(x)\right\}$$

$$= \frac{\sum_{k=1}^{r}\sum_{l=1}^{s}\pi_k\mu_l\binom{k+l-2}{l-1}\binom{r+s-k-l+1}{s-l}(s-k+1)}{s\binom{r+s}{r}}$$

and

$$C = \sum_{k=1}^{r}\sum_{l=1}^{s}\sum_{i=1}^{r}\sum_{j=1}^{s}\pi_k\pi_i\mu_l\mu_j\frac{\binom{k+l-2}{k-1}\binom{r+s-k-l}{s-l}}{\binom{r+s-1}{r}}$$

$$\times\frac{\binom{i+j-2}{j-1}\binom{r+s-i-j}{s-j}}{\binom{r+s-1}{r}}$$

$$\times\sum_{p=k+l-1}^{r+s-1}\sum_{q=i+j-1}^{r+s-1}\binom{r+s-1}{p}\binom{r+s-1}{q}\frac{(p+q)!(2r+2s-p-q-2)!}{(2r+2s-1)!}.$$

Thus $r^2\xi_{1,0}/\lambda+s^2\xi_{0,1}/(1-\lambda)$ is identical to the variance under H_0, i.e. $\xi/(\lambda(1-\lambda))$.

Proof of Theorem 3.1. Firstly, we claim that

$$\binom{r+1}{k}\mathrm{GMWW}_{n,m}^{(k:r,\,l:s)} = \binom{r}{k}\mathrm{GMWW}_{n,m}^{(k:r+1,\,l:s)}$$

$$+\binom{r}{k-1}\mathrm{GMWW}_{n,m}^{(k+1:r+1,\,l:s)}. \tag{7}$$

Since

$$\binom{r+1}{k}\mathrm{GMWW}_{n.m}^{(k:r,\,l:s)}$$

$$=\frac{\binom{r+1}{k}\sum_{i=1}^{n}\sum_{j=1}^{m}\binom{i-1}{k-1}\binom{n-i}{r-k}\binom{j-1}{l-1}\binom{m-j}{s-l}I(X_{i:n}<Y_{j:m})}{\binom{n}{r}\binom{m}{s}},$$

$$\binom{r}{k}\mathrm{GMWW}_{n,m}^{(k:r+1,\,l:s)}$$
$$=\frac{\binom{r}{k}\sum_{i=1}^{n}\sum_{j=1}^{m}\binom{i-1}{k-1}\binom{n-i}{r-k+1}\binom{j-1}{l-1}\binom{m-j}{s-l}I(X_{i:n}<Y_{j:m})}{\binom{n}{r+1}\binom{m}{s}}$$

and

$$\binom{r}{k-1}\mathrm{GMWW}_{n,m}^{(k+1:r+1,\,l:s)}$$
$$=\frac{\binom{r}{k-1}\sum_{i=1}^{n}\sum_{j=1}^{m}\binom{i-1}{k}\binom{n-i}{r-k}\binom{j-1}{l-1}\binom{m-j}{s-l}I(X_{i:n}<Y_{j:m})}{\binom{n}{r+1}\binom{m}{s}},$$

it is sufficient to show that

$$\frac{\binom{r+1}{k}\binom{i-1}{k-1}\binom{n-i}{r-k}}{\binom{n}{r}}=\frac{\binom{r}{k-1}\binom{i-1}{k}\binom{n-i}{r-k}}{\binom{n}{r+1}}$$
$$+\frac{\binom{r}{k}\binom{i-1}{k-1}\binom{n-i}{r-k+1}}{\binom{n}{r+1}}.$$

By expanding and simplifying both sides, this identity follows immediately. Likewise, we obtain

$$\binom{s+1}{l}\mathrm{GMWW}_{n,m}^{(k:r,\,l:s)}=\binom{s}{l}\mathrm{GMWW}_{n.m}^{(k:r,\,l:s+1)}+\binom{s}{l-1}\mathrm{GMWW}_{n,m}^{(k:r,\,l+1:s+1)}. \tag{8}$$

Secondly, combining (7) and (8) yields

$$\mathrm{GMWW}_{n,m}^{(k:r,\,l:s)}$$

$$= \frac{\binom{r}{k}\left\{\binom{s}{l}\mathrm{GMWW}_{n,m}^{(k:r+1,\,l:s+1)} + \binom{s}{l-1}\mathrm{GMWW}_{n,m}^{(k:r+1,\,l+1:s+1)}\right\}}{\binom{r+1}{k}\binom{s+1}{l}}$$

$$+ \frac{\binom{r}{k-1}\left\{\binom{s}{l}\mathrm{GMWW}_{n,m}^{(k+1:r+1,\,l:s+1)} + \binom{s}{l-1}\mathrm{GMWW}_{n,m}^{(k+1:r+1,\,l+1:s+1)}\right\}}{\binom{r+1}{k}\binom{s+1}{l}}.$$

Hence (7) implies that

$$C_{r,s} \subset WC_{r+1,s} \quad \Rightarrow \quad WC_{r,s} \subset WC_{r+1,s}$$

and (8) implies that

$$C_{r,s} \subset WC_{r,s+1} \quad \Rightarrow \quad WC_{r,s} \subset WC_{r+1,s+1}.$$

Finally, using the above recursive relations yields

$$C_{r_1,s_1} \subset WC_{r_1,s_1} \subset WC_{r_2,s_2},$$

whenever $r_1 \leqslant r_2$ and $s_1 \leqslant s_2$.

The proof of theorem is completed as desired. □

Proof of Identity (4).

$$F_{k:r} = \sum_{i=k}^{r} \binom{r}{i} F^i (1-F)^{r-i}$$

$$= \sum_{i=k}^{r} \binom{r}{i} \{F^i(1-F)^{r-i+1} + F^{i+1}(1-F)^{r-i}\}$$

$$= \sum_{i=k}^{r} \binom{r}{i} \left(\frac{F_{i:r+1} - F_{i+1:r+1}}{\binom{r+1}{i}} + \frac{F_{i+1:r+1} - F_{i+2:r+1}}{\binom{r+1}{i+1}} \right)$$

$$= \sum_{i=k}^{r} \left[\frac{\binom{r}{i}}{\binom{r+1}{i}} F_{i:r+1} + \frac{\left(\binom{r+1}{i} - \binom{r+1}{i+1}\right)\binom{r}{i}}{\binom{r+1}{i}\binom{r+1}{i+1}} F_{i+1:r+1} \right.$$

$$\left. - \frac{\binom{r}{i}}{\binom{r+1}{i+1}} F_{i+2:r+1} \right]$$

$$= \frac{\binom{r}{k}}{\binom{r+1}{k}} F_{k:r+1}$$

$$+ \frac{\left\{\binom{r+1}{k} - \binom{r+1}{k+1}\right\}\binom{r}{k} + \binom{r+1}{k}\binom{r}{k+1}}{\binom{r+1}{k}\binom{r+1}{k+1}} F_{k+1:r+1}$$

$$+ \sum_{i=k}^{r-2} \left[-\frac{\binom{r}{i}}{\binom{r+1}{i+1}} + \frac{\left\{\binom{r+1}{i+1} - \binom{r+1}{i+2}\right\}\binom{r}{i+1}}{\binom{r+1}{i+1}\binom{r+1}{i+2}} \right.$$

$$\left. + \frac{\binom{r}{i+2}}{\binom{r+1}{i+2}} \right] F_{i+2:r+1}$$

$$+ \left[\frac{\left\{\binom{r+1}{r} - \binom{r+1}{r+1}\right\}\binom{r}{r}}{\binom{r+1}{r}\binom{r+1}{r+1}} - \frac{\binom{r}{r-1}}{\binom{r+1}{r}} \right] F_{r+1:r+1}$$

$$= \frac{\binom{r}{k}}{\binom{r+1}{k}} F_{k:r+1} + \frac{\binom{r}{k-1}}{\binom{r+1}{k}} F_{k+1:r+1}.$$

The proof is completed as desired. □

References

Adams et al., 2000. Ahmad, I.A. (1996). A class of Mann–Whitney–Wilcoxon type statistics, The American Statistician, 50, 324–327: Comment by Adams, Adams, Chang, Etzel, Kuo, Montemayor, and Schucany; and Reply. Amer. Statist. 54, 160.

Ahmad, I.A., 1996. A class of Mann–Whitney–Wilcoxon type statistics. Amer. Statist. 50, 324–327.

Deshpande, J.V., Kochar, S.C., 1980. Some competitors of tests based on powers of ranks for the two-sample problem. Sankhya Ser. B 42, 236–241.

Hogg, R.V., Fisher, D.M., Randles, R.H., 1975. A two-sample adaptive distribution-free test. J. Amer. Statist. Assoc. 70, 656–661.

Kochar, S.C., 1978. A class of distribution-free tests for the two-sample slippage problem. Comm. Statist. Theory Methods A 7 (13), 1243–1252.

Kumar, N., 1997. A class of two-sample tests for location based on sub sample medians. Comm. Statist. Theory Methods 26, 943–951.

Lorentz, G.G., 1986. Bernstein Polynomials, 2nd Edition. Chelsea, New York.
Maesono, Y., 1987. Competitors of the Wilcoxon signed rank test. Ann. Inst. Statist. Math. 39, 363–375.
Mann, H.B., Whitney, D.R., 1947. On a test of whether one of two random variables is stochastically larger than the other. Ann. Math. Statist. 18, 50–60.
Marron, J.S., Wand, M.P., 1992. Exact mean integrated squared error. Ann. Statist. 20, 712–736.
Mathisen, H.C., 1943. A method of testing the hypothesis that two samples are from the same population. Ann. Math. Statist. 14, 188–194.
Priebe, C.E., Cowen, L.J., 1999. A generalized Wilcoxon–Mann–Whitney statistic. Comm. Statist. Theory Methods 28 (12), 2871–2878.
Randles, R.H., Wolfe, D.A., 1979. Introduction to the Theory of Nonparametric Statistics, Wiley Series in Probability and Mathematical Statistics. Wiley, New York.
Shetty, I.D., Bhat, S.V., 1994. A note on the generalization of Mathisen's median test. Stat. Probab. Lett. 19, 199–204.
Shetty, I.D., Govindarajulu, Z., 1988. A two-sample test for location. Comm. Statist. Theory Methods 17, 2389–2401.
Stephenson, R.W., Ghosh, M., 1985. Two sample nonparametric tests based on subsamples. Comm. Statist. Theory Methods 14 (7), 1669–1684.
Wilcoxon, F., 1945. Individual comparisons by ranking methods. Biometrics 1, 80–83.
Xie, J., 1999. Generalizing the Mann–Whitney–Wilcoxon statistic. Dissertation, The Johns Hopkins University.
Xie, J., Priebe, C.E., 2000. Generalizing the Mann–Whitney–Wilcoxon Statistic. J. Nonparametric Statist. 12, 661–682.