Biometrical Journal **42** (2000) 1, 17–25

The Nonparametric Behrens-Fisher Problem: Asymptotic Theory and a Small-Sample Approximation

EDGAR BRUNNER and ULLRICH MUNZEL

Abt. Med. Statistik
Universität Göttingen
Germany

Summary

A generalization of the Behrens-Fisher problem for two samples is examined in a nonparametric model. It is not assumed that the underlying distribution functions are continuous so that data with arbitrary ties can be handled. A rank test is considered where the asymptotic variance is estimated consistently by using the ranks over all observations as well as the ranks within each sample. The consistency of the estimator is derived in the appendix. For small samples $(n_1,\, n_2 \geq 10)$, a simple approximation by a central t-distribution is suggested where the degrees of freedom are taken from the Satterthwaite-Smith-Welch approximation in the parametric Behrens-Fisher problem. It is demonstrated by means of a simulation study that the Wilcoxon-Mann-Whitney-test may be conservative or liberal depending on the ratio of the sample sizes and the variances of the underlying distribution functions. For the suggested approximation, however, it turns out that the nominal level is maintained rather accurately. The suggested nonparametric procedure is applied to a data set from a clinical trial. Moreover, a confidence interval for the nonparametric treatment effect is given.

Key words: Rank Test; Heteroscedastic Model; Satterthwaite-Smith-Welch Approximation; Ties; Ordered Categorical Data.

1. Introduction

The problem of analyzing two independent samples with possibly heteroscedastic normally distributed errors has been considered extensively in the literature. For a recent discussion, see, for example, MOSER and STEVENS (1992) or ROTH (1983). In this context, an accurate and simple approximation is of particular interest. The so-called *Satterthwaite-Smith-Welch (SSW)* approximation is used by many statistical software packages because this method provides a good approximation of the pre-assigned level, even in the case of extreme heteroscedasticity and unequal sample sizes.

In nonparametric models, asymptotic procedures for the heteroscedastic case have been considered by FLIGNER and POLICELLO (1981) and by BRUNNER and NEUMANN (1982, 1986) under the assumption of continuous distribution functions.

In the present paper, this assumption is relaxed and arbitrary distribution functions are admitted (with the exception of the trivial case of one-point-distributions). This includes the cases where ties occur by rounding off observations from continuous distributions as well as by observing count data or ordered categorical data. Based on some recent results of AKRITAS and BRUNNER (1997) an asymptotically distribution free rank test is derived and a consistent estimator for the unknown asymptotic variance is given. Moreover, a simple approximation for small samples is considered and its accuracy is examined by means of a simulation study.

Models and hypotheses are considered in Section 2 while the main results are given in Section 3 and 4. Finally, the application of the procedure is demonstrated by an example in Section 6. Some technical details are provided in the Appendix.

2. Nonparametric Model and Hypothesis

We consider a general nonparametric model where $N = n_1 + n_2$ independent random variables

$$X_{ik} \sim F_i(x)\,, \qquad k = 1, \ldots, n_i\,, \tag{2.1}$$

are observed under $i = 1, 2$ treatments. The distribution functions $F_i(x)$ may be arbitrary (with the exception of the trivial case of one-point-distributions). In a nonparametric context, the hypothesis of no treatment effect is commonly formulated as $H_0^F : F_1 = F_2$ which implies homoscedasticity under the hypothesis.

To formulate a nonparametric hypothesis of no treatment effect, which entails the parametric Behrens-Fisher problem as a special case, we consider the relative treatment effect

$$p = P(X_{11} < X_{21}) + \tfrac{1}{2}P(X_{11} = X_{21})\,. \tag{2.2}$$

The random variable X_{11} is called *to tend to smaller (larger) values* than the random variable X_{21} if $p > \frac{1}{2}$ $(p < \frac{1}{2})$ and the two random variables are called *tendentiously equal* if $p = \frac{1}{2}$. To illustrate the meaning of $p = \frac{1}{2}$, consider two normal distributions F_i, $i = 1, 2$ with expectations and variances μ_i and σ_i^2, respectively. Then it is easily seen that $\mu_1 = \mu_2 \Leftrightarrow p = \frac{1}{2}$, where the variances σ_1^2 and σ_2^2 may be different. Thus, a reasonable hypothesis of no treatment effect in the general nonparametric model (2.1) can be expressed as $H_0^p : p = \frac{1}{2}$. This hypothesis has been called *generalized* or *nonparametric Behrens-Fisher problem* (FLIGNER and POLICELLO, 1981; BRUNNER and NEUMANN, 1986) since the parametric Behrens-Fisher problem is contained as a special case.

To estimate the relative treatment effect p and to derive its asymptotic distribution it is more convenient to express p in terms of the distribution functions. To this end, we use the so-called *normalized version* $F_i(x) = \frac{1}{2}[F_i^-(x) + F_i^+(x)]$ of the distribution function (RUYMGAART, 1980) where $F_i^-(x) = P(X_{i1} < x)$ is the left-continuous version and $F_i^+(x) = P(X_{i1} \leq x)$ is the right-continuous version of the distribution function. Then, the relative teatment effect p can be written as $p = \int F_1 \, dF_2$ and the hypothesis of no treatment effect is written as $H_0^p : p = \int F_1 \, dF_2 = \frac{1}{2}$. We note that $H_0^F : F_1 = F_2 = F$ implies $H_0^p : p = \frac{1}{2}$, because $\int F \, dF = \frac{1}{2}$, which follows from integration by parts.

3. The Test and its Asymptotic Distribution

To estimate the relative treatment effect p, the distribution functions F_1 and F_2 are replaced by their empirical counterparts $\hat{F}_i(x) = \frac{1}{2}[\hat{F}_i^-(x) + \hat{F}_i^+(x)]$ where $\hat{F}_i^-(x)$ is the left-continuous version and $\hat{F}_i^+(x)$ is the right-continuous version of the empirical distribution function. Let $H(x) = \sum_{i=1}^{2} \frac{n_i}{N} F_i(x)$ denote the combined distribution function and let $\hat{H}(x) = \sum_{i=1}^{2} \frac{n_i}{N} \hat{F}_i(x)$ denote the normalized version of the combined empirical distribution function. Note that $R_{ij} = N \cdot \hat{H}(X_{ik}) + \frac{1}{2}$ is the rank of X_{ik} among all $N = n_1 + n_2$ observations $X_{11}, \ldots, X_{2n_2}$. In case of ties, R_{ij} is the mid-rank, which comes out in a natural way by using the normalized version $\hat{F}_i(x)$ of the empirical distribution function, as defined above. Let $\bar{R}_{i\cdot} = n_i^{-1} \sum_{k=1}^{n_i} R_{ik}$, $i = 1, 2$, denote the mean of the ranks R_{ik} in the ith sample. Then, it follows that

$$\hat{p} = \int \hat{F}_1 \, d\hat{F}_2 = \frac{1}{n_1} \left(\bar{R}_{2\cdot} - \frac{n_2 + 1}{2} \right) \tag{3.3}$$

is an unbiased and consistent estimator for the relative treatment effect p. The unbiasedness of $\hat{p}$ follows immediately from $E[c(X_{21} - X_{11})] = \int F_1 \, dF_2$, where $c(u) = 0, \frac{1}{2}, 1$ according as $u <, =, >$, denotes the normalized version of the count function. The L_2-consistency follows from (7.13) in the Appendix as a special case by letting $g(u) = u$.

Finally, the asymptotic normality of the statistic $\sqrt{N} \left(\hat{p} - \frac{1}{2} \right) = \frac{1}{\sqrt{N}} (\bar{R}_{2\cdot} - \bar{R}_{1\cdot})$ is based on the following decomposition

$$\begin{aligned}
\sqrt{N}(\hat{p} - p) &= \sqrt{N} \left(\int \hat{F}_1 \, d\hat{F}_2 - \int F_1 \, dF_2 \right) \\
&= \sqrt{N} \left[\int (\hat{F}_1 - F_1) \, dF_2 + \int F_1 \, d(\hat{F}_2 - F_2) + \int (\hat{F}_1 - F_1) \, d(\hat{F}_2 - F_2) \right] \\
&= \sqrt{N} \left(\int F_1 \, d\hat{F}_2 - \int F_2 \, d\hat{F}_1 + 1 - 2 \int F_1 \, dF_2 \right) + C_N = U_N + C_N \,,
\end{aligned}$$

where $C_N = \sqrt{N} \int (\hat{F}_1 - F_1) \, d(\hat{F}_2 - F_2)$ and

$$U_N = \sqrt{N} \left(\frac{1}{n_2} \sum_{j=1}^{n_2} F_1(X_{2j}) - \frac{1}{n_1} \sum_{j=1}^{n_1} F_2(X_{1j}) + 1 - 2p \right).$$

The so-called *Asymptotic Equivalence Theorem* states that $C_N \xrightarrow{p} 0$, which means that the two statistics $\sqrt{N}(\hat{p} - p)$ and U_N have, asymptotically, the same distribution. This result follows from Theorem 2.2 of AKRITAS and BRUNNER (1997) as a special case by letting $m_{ijk} = 1$, $c = 1$ and $r = 2$ in this theorem. Further, note that U_N is a linear combination of independent and identically distributed random variables $F_1(X_{2j})$, $j = 1, \ldots, n_2$, and $F_2(X_{1j})$, $j = 1, \ldots, n_1$, which are uniformly bounded by 1 and the asymptotic normality of U_N follows immediately from the Central Limit Theorem, if $\sigma_1^2 = \mathrm{Var}\,(F_2(X_{11})) > 0$ and $\sigma_2^2 = \mathrm{Var}\,(F_1(X_{21})) > 0$. Moreover, it follows that $E(U_N) = 0$ and that

$$\sigma_N^2 = \mathrm{Var}\,(U_N) = N[\sigma_1^2/n_1 + \sigma_2^2/n_2]\,. \tag{3.4}$$

Thus, $\sqrt{N}\,(\hat{p} - \frac{1}{2})/\sigma_N = (\bar{R}_{2\cdot} - \bar{R}_{1\cdot})/\sqrt{N\sigma_N^2}$ has, asymptotically, a standard normal distribution under $H_0^p : p = \frac{1}{2}$.

4. Estimation of the Variance

Note that, even under H_0^p, the variances σ_1^2 and σ_2^2 in (3.4) are unknown and must be estimated from the data. To this end, let $Y_{1k} = F_2(X_{1k})$, $k = 1, \ldots, n_1$, and $Y_{2k} = F_1(X_{2k})$, $k = 1, \ldots, n_2$, and note that the random variables Y_{1k} and Y_{2k} are independent by assumption. Then, the quantities $\tilde{\sigma}_i^2 = (n_i - 1)^{-1} \sum_{k=1}^{n_i} (Y_{ik} - \bar{Y}_{i\cdot})^2$ are unbiased and consistent for σ_i^2, $i = 1, 2$. However, the random variables Y_{ik} are unobservable and, for the computation of an estimator, they must be replaced by observable random variables which are "close enough" to the unobservable random variables Y_{ik}. Therefore, we replace the distribution functions $F_i(x)$ by their empirical counterparts $\hat{F}_i(x)$, $i = 1, 2$. Then, by definition,

$$n_1\hat{F}_1(X_{2k}) = N\hat{H}(X_{2k}) - n_2\hat{F}_2(X_{2k}) = R_{2k} - R_{2k}^{(2)}\,,$$

$$n_2\hat{F}_2(X_{1k}) = N\hat{H}(X_{1k}) - n_1\hat{F}_1(X_{1k}) = R_{1k} - R_{1k}^{(1)}\,,$$

where $R_{ik}^{(i)} = n_i\hat{F}_i(X_{ik}) + \frac{1}{2}$ denotes the (within) rank of X_{ik} among the n_i observations within the ith sample $X_{i1}, \ldots, X_{in_i}$, $i = 1, 2$. In the case of ties, the mid-ranks come out automatically as already noted for the (overall) ranks $R_{ik} = N\hat{H}(X_{ik}) + \frac{1}{2}$. Then, the variances σ_i^2 are estimated by

$$\hat{\sigma}_i^2 = S_i^2/(N - n_i)^2\,, \tag{4.5}$$

where

$$S_i^2 = \frac{1}{n_i - 1} \sum_{k=1}^{n_i} \left(R_{ik} - R_{ik}^{(i)} - \bar{R}_{i\cdot} + \frac{n_i + 1}{2} \right)^2 . \tag{4.6}$$

is the empirical variance of $R_{ik} - R_{ik}^{(i)}$, $k = 1, \ldots, n_i$, $i = 1, 2$.

The mean square error caused by replacing the unobservable random variables Y_{ik} by the observable random variables $(R_{ik} - R_{ik}^{(i)})/(N - n_i)$ is of the order

$1/(N - n_i)$. This follows from $E(\hat{F}_2(X_{1k}) - F_2(X_{1k}))^2 \leq 1/n_2$ and $E(\hat{F}_1(X_{2k}) - F_1(X_{2k}))^2 \leq 1/n_1$ (see Appendix 7.1). The estimator $\hat{\sigma}_i^2$ is consistent for σ_i^2, since $E(\hat{\sigma}_i^2/\sigma_i^2 - 1)^2 \to 0$, $i = 1, 2$ (the technical details are outlined in the Appendix 7.2). Thus, under H_0^p,

$$\hat{\sigma}_N^2 = N \cdot [\hat{\sigma}_1^2/n_1 + \hat{\sigma}_2^2/n_2] \tag{4.7}$$

is consistent for σ_N^2 and it follows that the statistic

$$W_N^{BF} = \frac{\sqrt{N}\,(\hat{p} - \frac{1}{2})}{\hat{\sigma}_N} = \frac{1}{\sqrt{N}} \cdot \frac{\bar{R}_{2\cdot} - \bar{R}_{1\cdot}}{\hat{\sigma}_N} \tag{4.8}$$

has, asymptotically, a standard normal distribution under the hypothesis $H_0^p : p = \frac{1}{2}$.

5. Approximation for Small Samples

In a comprehensive simulation study, the behaviour of the Wilcoxon-Mann-Whitney-test (WMW) and of the test based on the statistic W_N^{BF} in (4.8) was examined under heteroscedasticity, where several types of distributions were compared: (1) two normal distributions with unequal variances, (2) one unimodal and one bimodal distribution, (3) two symmetric discrete distributions with unequal variances, and (4) two distributions with different variances and with different skewnesses. The ratio of the variances of the two distributions F_1 and F_2 ranged from 0.1 to 10. It turned out that the WMW-test led to conservative decisions if the larger sample size was taken from the population with the larger variance while it led to rather liberal decisions in the opposite case. This fact did not depend on the total sample size N but on the ratio of the variances σ_1^2 and σ_2^2 and on the ratio n_1/n_2 of the two sample sizes, i.e. the liberal or the conservative behaviour did not vanish asymptotically. For samples sizes $n_1, n_2 \geq 10$, the simulated type-I error rates for the WMW test ranged from 7% to 19.9% (nominal level 10%), from 2.8% to 12.2% (nominal level 5%) and from 0.2% to 4% (nominal level 1%). The test based on the statistic W_N^{BF}, however, was more or less liberal for medium or small sample sizes (smaller than about 50) and was quite accurate for larger sample sizes. These results gave rise to develop a more accurate approximation for small samples.

Note that the distribution of $\hat{\sigma}_N^2$ in (4.7) becomes degenerate rather quickly (at the rate of $1/N$), because $\hat{\sigma}_N^2$ is consistent for σ_N^2 and thus, the small sample distribution of W_N^{BF} may be approximated by a distribution which converges to the standard normal disribution with increasing sample sizes. A simulation study showed that the quality of the approximation was mainly affected by the ratio of the variances, the sample sizes and, of course, by the total sample size. These findings gave rise to use an approximation by a t-distribution where the degrees of freedom were taken from the parametric SSW-approximaion, i.e. for small sample sizes, the null distribution of W_N^{BF} is approximated by a central t-distribution with

$$\hat{f} = \frac{\left(\sum_{i=1}^{2} \hat{\sigma}_i^2/n_i\right)^2}{\sum_{i=1}^{2} (\hat{\sigma}_i^2/n_i)^2/(n_i-1)} = \frac{\left(\sum_{i=1}^{2} S_i^2/(N-n_i)\right)^2}{\sum_{i=1}^{2} [S_i^2/(N-n_i)]^2/(n_i-1)} \tag{5.9}$$

degrees of freedom, where S_i^2 is given in (4.6). It is easily seen that $\hat{f} \to \infty$ and thus, the $t_{\hat{f}}$-distribution converges to the standard normal distribution which means that the approximation is asymptotically correct.

The accuracy of this approximation was examined by means of a simulation study, where the same sample sizes and heteroscedastic distributions were used as in the simulation study for the WMW test. For the approximation by the t-distribution with $\hat{f}$ degrees of freedom given in (5.9), the simulated type-I error rates ranged from 9.5% to 10.7% (nominal level 10%), from 4.6% to 5.7% (nominal level 5%) and from 0.5% to 1.5% (nominal level 1%). The quality of this approximation is comparable to that of the SSW-approximation in the parametric case. For extremely small sample sizes $(n_i < 10)$, simple and accurate approximations in a general nonparametric model cannot be expected.

6. An Example

In this section, we apply the statistic W_N^{BF} and the approximation by the t-distribution with $\hat{f}$ degrees of freedom given in (5.9) to the shoulder tip pain trial as reported by LUMLEY (1996). In this clinical trial, a pain score was observed for every patient after a laparoscopic surgery. The pain score ranged from 1 (low) to 5 (high). Two treatments (Y and N) were randomly assigned to 25 eligible female patients where 14 patients received the active treatment Y and 11 patients the control treatment N. (Here we consider only the subset of the pain scores on the third day after the surgery for the 25 female patients). The observed pain scores are listed in table 6.1.

For the physician, it was mainly of interest to know whether the pain scores after the active treatment Y "tended to be lower" than after the control treatment N. In this trial, the "difference" of the two treatments can be described by the relative treatment effect $p = \int F_1 \, dF_2$, where F_1 denotes the distribution of the pain scores under treat-

Table 6.1
Pains scores on the third day after surgery for $n_1 = 14$ patients under the treatment Y and $n_2 = 11$ patients under the treatment N

Treatment	Pain Score
Y	1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 2, 4, 1, 1
N	3, 3, 4, 3, 1, 2, 3, 1, 1, 5, 4

ment Y and F_2 denotes the distribution of the pain scores under treatment N. It is of interest to detect the one-sided alternative $H_1^p : p > \frac{1}{2}$ and thus, it appears to be natural to test the hypothesis $H_0^p : p = \frac{1}{2}$ against this one-sided alternative. Moreover, the possible benefit of the active treatment can be estimated by the relative treatment effect $\hat{p} = (\bar{R}_{2\cdot} - (n_2 + 1)/2)/n_1$, where $\bar{R}_{2\cdot}$ is the mean of the (mid-)ranks R_{ik}, $i = 1, 2;\ k = 1, \ldots, n_i$, of the observed pain scores under the control treatment. We note that it is not reasonable to apply a parametric procedure (e.g. the t-test) for the analysis of this trial because the pain scores are ordered categorical data.

The two rank means are $\bar{R}_{1\cdot} = 9.79$ (treatment Y), $\bar{R}_{2\cdot} = 17.09$ (treatment N) and the estimated relative treatment effect is $\hat{p} = 0.792$. The hypothesis $H_0^p : p = \frac{1}{2}$ is rejected at the 1%-level ($W_N^{BF} = 3.24$, one sided p-value: 0.002), where the approximation by the t-distribution for small samples is used ($\hat{f} = 21.01$).

Moreover, from the derivation of the statistic in section 3, it follows that $\sqrt{N}\,(\hat{p} - p)/\hat{\sigma}_N$ has, asymptotically, a standard normal distribution and, for large sample sizes, a one-sided $(1 - \alpha)$-confidence interval for the relative treatment effect p is given by $P(p \geq p_L) \doteq 1 - \alpha$, where $p_L = \hat{p} - u_{1-\alpha} \cdot \hat{\sigma}_N/\sqrt{N}$. For small samples, the suggested approximation by the t-distribution may be used and the critical value $u_{1-\alpha}$ is replaced by the critical value $t_{\hat{f};1-\alpha}$ of the t-distribution with $\hat{f}$ degrees of freedom given in (5.9). For the present example, the lower bound of a one-sided 95%-confidence interval for the relative treatment effect is $p_L = 0.792 - 1.721 \cdot 0.451/\sqrt{25} = 0.64$, which is (at a confidence level of 95%) the minimal probability of observing a lower pain score under treatment Y than under treatment N.

7. Appendix

7.1 *Some inequalities*

To prove the consistency of the variance estimators $\hat{\sigma}_i^2$, $i = 1, 2$, given in (4.5), we first need some inequalities. Consider the notation of sections 3 and 4. Then,

$$E[\hat{F}_1(X_{2k}) - F_1[X_{2k})]^2 \leq 1/n_1\,. \tag{7.10}$$

Proof: Using the independence of X_{1j} and $X_{1j'}$ for $j \neq j'$, the independence of X_{1j} and X_{2k}, $j = 1, \ldots, n_1$, $k = 1, \ldots, n_2$, and Fubini's theorem, it follows that

$$\begin{aligned}
&E[\hat{F}_1(X_{2k}) - F_1(X_{2k})]^2 \\
&= \frac{1}{n_1^2} \sum_{j=1}^{n_1} \sum_{j'=1}^{n_1} E([c(X_{2k} - X_{1j}) - F_1(X_{2k})]\,[c(X_{2k} - X_{1j'}) - F_1(X_{2k})]) \\
&= \frac{1}{n_1^2} \sum_{j=1}^{n_1} E[c(X_{2k} - X_{1j}) - F_1(X_{2k})]^2 \leq \frac{1}{n_1}\,,
\end{aligned}$$

since $E[c(X_{2k} - X_{1j}) - F_1(X_{2k})] = 0$ and $|c(X_{2k} - X_{1j}) - F_1(X_{2k})| \leq 1$. □

Next we consider a function $g(u)$ on $[0, 1]$ with the properties $\sup_{0 \le u \le 1} |g(u)| = C_0 < \infty$ and $\sup_{0 \le u \le 1} |g'(u)| = C_1 < \infty$. Then,

$$(i) \quad \text{Var}\,(g[F_1(X_{2k})]) \le C_0^2\,, \tag{7.11}$$

$$(ii) \quad E[g(\hat{p}) - g(p)]^2 \le C_1^2 E(\hat{p} - p)^2\,. \tag{7.12}$$

Proof: Inequality (7.11) is obvious, since $|g(\cdot)| \le C_0$, by assumption. The second inequality (7.12) follows by applying the mean value theorem. □

7.2 *Consistency of the variance estimator*

Next, we consider the ratio of the estimator $\hat{\sigma}_2^2$ in (4.5) and $\sigma_2^2 = \int F_1^2 \, \mathrm{d}F_2 - (\int F_1 \, \mathrm{d}F_2)^2$ and we show that $E(\hat{\sigma}_2^2/\sigma_2^2 - 1)^2 \to 0$. Note that it suffices to show that $E(\hat{\sigma}_2^2 - \sigma_2^2)^2 \to 0$, since $\sigma_2^2 > 0$ by assumption. The result for $\hat{\sigma}_1^2$ will follow in the same way.

To apply the inequalities (7.11) and (7.12), let $g(u) = u^2$. Then, $C_0 = 1$ in (7.11) and $C_1 = 2$ in (7.12). Furthermore, let $p(g) = \int g(F_1) \, \mathrm{d}F_2$ and $\hat{p}(g) = \int g(\hat{F}_1) \, \mathrm{d}\hat{F}_2$. Then, σ_2^2 and $\hat{\sigma}_2^2$ can be written as $\sigma_2^2 = p(g) - g(p)$ and

$$\hat{\sigma}_2^2 = \frac{S_2^2}{n_1^2} = \frac{n_2}{n_2 - 1} \left[\int \hat{F}_1^2 \, \mathrm{d}\hat{F}_2 - \hat{p}^2 \right] = \frac{n_2}{n_2 - 1} \, [\hat{p}(g) - g(\hat{p})]\,.$$

Thus,

$$\hat{\sigma}_2^2 - \sigma_2^2 = \frac{n_2}{n_2 - 1} \, [\hat{p}(g) - g(\hat{p}) - p(g) + g(p)] + \frac{1}{n_2 - 1} \, \sigma_2^2\,.$$

Applying the c_r-inequality for $r = 2$ and taking the expectation, it follows that

$$E(\hat{\sigma}_2^2 - \sigma_2^2)^2 \le \frac{4n_2}{n_2 - 1} \, (E[\hat{p}(g) - p(g)]^2 + E[g(\hat{p}) - g(p)]^2) + \frac{2}{(n_2 - 1)^2}\,,$$

since $0 < \sigma_2^2 \le 1$. First, we show that

$$E[\hat{p}(g) - p(g)]^2 \to 0\,. \tag{7.13}$$

Applying the c_r-inequality for $r = 2$ and Jensen's inequality it follows that

$$\begin{aligned}
[\hat{p}(g) - p(g)]^2 &= \left[\int g(\hat{F}_1) \, \mathrm{d}\hat{F}_2 - \int g(F_1) \, \mathrm{d}\hat{F}_2 + \int g(F_1) \, \mathrm{d}\hat{F}_2 - p(g) \right]^2 \\
&\le 2 \int [g(\hat{F}_1) - g(F_1)]^2 \, \mathrm{d}\hat{F}_2 + 2 \left(\frac{1}{n_2} \sum_{k=1}^{n_2} [g[F_1(X_{2k})] - p(g)] \right)^2.
\end{aligned}$$

Taking the expectation, it follows for the first term $\int [g(\hat{F}_1) - g(F_1)]^2 \,\mathrm{d}\hat{F}_2$ that

$$
\begin{aligned}
&E\left(\int [g[\hat{F}_1] - g[F_1]]^2 \,\mathrm{d}\hat{F}_2\right) \\
&= \frac{1}{n_2} \sum_{k=1}^{n_2} E[g[\hat{F}_1(X_{2k})] - g[F_1(X_{2k})]]^2 \\
&\le C_1^2 \cdot \frac{1}{n_2} \sum_{k=1}^{n_2} E[\hat{F}_1(X_{2k}) - F_1(X_{2k})]^2 \le 4/n_1 \to 0\,, \quad \text{for} \quad n_1 \to \infty
\end{aligned}
$$

by applying the mean value theorem, inequality (7.10) and noting that $C_1 = 2$. For the second term, we note that $p(g) = E(g[F_1(X_{2k})])$ and thus, by independence, it follows that

$$
\begin{aligned}
&E\left(\frac{1}{n_2} \sum_{k=1}^{n_2} [g[F_1(X_{2k})] - p(g)]\right)^2 \\
&= \frac{1}{n_2^2} \sum_{k=1}^{n_2} E[g(F_1(X_{2k})) - p(g)]^2 = \frac{1}{n_2^2} \sum_{k=1}^{n_2} \mathrm{Var}\,(g[F_1(X_{2k})]) \\
&\le C_0^2/n_2 = 1/n_2 \to 0\,, \quad \text{for} \quad n_2 \to \infty
\end{aligned}
$$

using inequality (7.11). Finally, it follows that $E[g(\hat{p}) - g(p)]^2 \to 0$ by inequality (7.12) and from (7.13) by letting $g(u) = u$. This completes the proof. □

References

AKRITAS, M. G. and BRUNNER, E., 1997: A unified approach to ranks tests in mixed models. *J. Statist. Plann. Inference* **61**, 249–277.

BRUNNER, E. and NEUMANN, N., 1982: Rank Tests for Correlated Random Variables. *Biom. J.* **24**, 373–389.

BRUNNER, E. and NEUMANN, N., 1986: Two-sample rank tests in general models. *Biom. J.* **28**, 395–402.

FLIGNER, M. A. and POLICELLO II, G. E., 1981: Robust rank procedures for the Behrens-Fisher problem. *J. Amer. Statist. Assoc.* **76**, 162–168.

LUMLEY, T., 1996: Generalized estimating equations for ordinal data: A note on working correlation structures. *Biometrics* **52**, 354–361.

MOSER, B. K. and STEVENS, G. R., 1992: Homogeneity of variance in the two sample mean test. *Amer. Statist.* **46**, 19–21.

ROTH, A. J., 1983: Robust Trend Tests Derived and Simulated: Analogs of the Welch and Brown-Forsythe Tests. *J. Amer. Statist. Assoc.* **78**, 972–980.

RUYMGAART, F. H., 1980: A unified approach to the asymptotic distribution theory of certain midrank statistics. In: *Statistique non Parametrique Asymptotique*, 1–18, J. P. Raoult (Ed.), Lecture Notes on Mathematics, No. 821, Springer, Berlin.

EDGAR BRUNNER and ULLRICH MUNZEL
Abt. Med. Statistik
Universität Göttingen
Humboldt Allee 32
D-37073 Göttingen
Germany
E-mail: brunner@ams.med.uni-goettingen.de

Received, June 1999
Accepted, October 1999