Saddlepoint Approximations of the Two-sample Wilcoxon Statistic

BY RUNGAO JIN AND JOHN ROBINSON

University of Sydney

Froda and van Eeden [1] obtain an approximation for the two-sample Wilcoxon statistic based on the moment generating function due to van Dantzig [8]. A direct saddlepoint approximation based on this moment generating function is obtained and is shown to have uniform relative error. These approximations are compared numerically to those based on the conditional saddlepoint method and the method of [1].

1. Introduction Froda and van Eeden [1] obtain an approximation for the two-sample Wilcoxon statistic, W, the sum of the ranks of the first sample, for samples of size m and n with $m + n = N$, based on an exact moment generating function of van Dantzig [8]. This gives relative errors of order $N^{-3/2}$ for the approximation of the tail probabilities, $P((W - EW)/\sqrt{VarW} \geq w)$, in the compact region, when w is bounded, but it does not give relative errors for a region including the large deviation region, when $w = O(\sqrt{N})$.

We consider approximations for the distribution of the two-sample Wilcoxon statistic by a number of different methods and obtain numerical comparisons of them. The first method is based on the moment generating function used by Froda and van Eeden [1]. We have used the usual saddlepoint method to obtain approximations to probabilities and tail probabilities. We have obtained two distribution approximations, the indirect Edgeworth approximation and the Barndorff-Nielsen approximation, asymptotically equivalent to that of Lugananni-Rice with relative error of order N^{-1}.

The second method is based on a conditional representation of the distribution of the Wilcoxon statistic which has been given in [5] and generalised in [6]. Also the method of [7] may be applied to give a Lugananni-Rice version of the conditional method. We give indirect Edgeworth and Barndorff-Nielsen approximations based on the conditional technique. Froda and van Eeden [1] also suggest that their method should greatly improve on the conditional method, but this is shown here not to be so, particularly in the large deviation region.

Finally, we give numerical examples for $n = m = 5$ and $m = 10$, $n = 6$ comparing approximations to probabilities and to tail probabilities from the saddlepoint approximations using indirect Edgeworth and Barndorff-Nielsen forms, and based on the conditional method and the approximation of [1]. These indicate that there is effectively no difference between the new saddlepoint approximations and the approximations based on the conditional method. The approximation of [1] is very good in the center of the distribution but quite poor in the tails.

2. Direct Saddlepoint Approximation Let W be the two-sample Wilcoxon rank sum statistic. Put $U = W - m(m+1)/2$ and $N = m+n$. Then, from Question 15 of page 126 of [2], we have

LEMMA 1.

$$Et^U = \prod_{j=1}^{m} \frac{j}{n+j} \frac{1-t^{n+j}}{1-t^j}, -\infty < t < \infty. \tag{1}$$

Therefore,

$$M(u) = Ee^{uU} = \prod_{j=1}^{m} \frac{j}{n+j} \frac{1-e^{(n+j)u}}{1-e^{ju}}$$

and

$$K(u) = \frac{1}{N}\log(Ee^{uU}) = \frac{1}{N}\left[-\log\binom{N}{m} + \sum_{j=1}^{m} \log\frac{1-e^{(n+j)u}}{1-e^{ju}}\right].$$

Define the tilted variable U_u by

$$P(U=l) = e^{NK(u)-lu}P(U_u = l),\ l = 0, 1, \ldots, nm. \tag{2}$$

We will choose u as the solution of the saddlepoint equation

$$K'(u) = l/N. \tag{3}$$

We know

$$EU = \frac{nm}{2}, \qquad \sigma^2 = VarU = \frac{mn(N+1)}{12}.$$

We only consider $0 \le l - mn/2 \le \delta N^2$ from now on, where $0 < \delta < mn/(2N^2) = pq/2$. There are analogous results for $l - mn/2 < 0$. The endpoints where l is near 0 or mn are considered in [4].

THEOREM 1. *For $\epsilon N^2 < l < mn - \epsilon N^2$, where $0 < \epsilon < pq$, we have*

$$P(U=l) = \frac{e^{NK(u)-lu}}{\sqrt{2\pi N K''(u)}}\left(1 + P_2(u)\frac{1}{N} + O(N^{-3/2})\right), \tag{4}$$

where $K'(u) = l/N$ and

$$P_2(u) = \frac{3K^{(4)}(u)}{4!(K''(u))^2} - \frac{15(K^{(3)}(u))^2}{2!(3!)^2(K''(u))^3}.$$

Proof: We prove this for $0 \le l - mn/2 \le \delta N^2$, where $0 < \delta < mn/(2N^2) = pq/2$. Consider the moment generating function of $U' = (U - mn/2)/N$, given by

$$M'(u') = Ee^{u'(U-mn/2)/N},$$

and let

$$\kappa(u') = \frac{1}{N}\log(Ee^{u'(U-mn/2)/N}) = K(u'/N) - \frac{pq}{2}u'.$$

So, from (3), we have

$$\kappa'(u'(x)) = (l - mn/2)/N^2 = x. \tag{5}$$

We have

$$P(U' = (l - mn/2)/N) = e^{-N\Lambda(x)} P(U'_{u'} = (l - mn/2)/N), \tag{6}$$

where $\Lambda(x) = -\kappa'(u'(x)) + xu'(x)$, and $U'_{u'}$ is the tilted variable defined by (6). We know

$$\begin{aligned} P(U'_{u'} = (l - mn/2)/N) &= \frac{1}{2\pi} \int_{-\pi N}^{\pi N} e^{-iv(l-mn/2)/N} \exp(N\kappa(u' + iv) - N\kappa(u'))dv \\ &= J_1 + J_2, \end{aligned}$$

where

$$J_1 = \frac{1}{2\pi} \int_{-\epsilon}^{\epsilon} e^{-iv(l-mn/2)/N} \exp(N\kappa(u' + iv) - N\kappa(u'))dv$$

and

$$J_2 = \frac{1}{2\pi} \int_{(-\pi N, \pi N)-(-\epsilon,\epsilon)} e^{-iv(l-mn/2)/N} \exp(N\kappa(u' + iv) - N\kappa(u'))dv.$$

Let us look at J_1 first. We can write

$$\begin{aligned} \kappa(u') &= \int_0^p [\log\{y(1 - e^{(q+y)u'})\} - \log\{(q+y)(1 - e^{yu'})\}]dF_N(y) - \frac{pq}{2}u' \\ &= \int_0^p [\log\{y(1 - e^{(q+y)u'})\} - \log\{(q+y)(1 - e^{yu'})\}]dy - \frac{pq}{2}u' + O(1/N), \end{aligned}$$

where

$$F_N(y) = \frac{j}{N}, \qquad \frac{j-1}{N} < y \le \frac{j}{N},\ j = 1, \ldots, m,$$

and the order term is uniform in $|u'| < C$. Further the derivatives of $\kappa(u')$ can be written in a similar way and $\kappa^{(k)}(u')$, $k = 0, 1, \ldots, 5$ are bounded for $|u'| < C$ and $\kappa''(u') > 0$ for $|u'| < C$. So expanding J_1 in the usual way we can get for $|\theta| < 1$,

$$\begin{aligned} J_1 &= \frac{1}{2\pi} \int_{-\epsilon}^{\epsilon} \exp\{-\frac{1}{2}N\kappa''(u')v^2 + \frac{N\kappa^{(3)}(u')}{3!}(iv)^3 + \frac{N\kappa^{(4)}(u')}{4!}(iv)^4 \\ &\quad + \frac{N\kappa^{(5)}(u' + \theta iv)}{5!}(iv)^5\}dv \\ &= \frac{1}{2\pi\sqrt{N\kappa''(u')}} \int_{-\epsilon\sqrt{N\kappa''(u')}}^{\epsilon\sqrt{N\kappa''(u')}} e^{-\frac{1}{2}w^2}[1 + \frac{\kappa^{(3)}(u')}{3!(\kappa''(u'))^{3/2}}(iw)^3 \frac{1}{\sqrt{N}} \\ &\quad + (\frac{(\kappa^{(3)}(u'))^2}{2!(3!)^2(\kappa''(u'))^3}(iw)^6 + \frac{\kappa^{(4)}(u')}{4!(\kappa''(u'))^2}(iw)^4)\frac{1}{N} + |w|^5 e^{\frac{1}{4}w^2} O(N^{-3/2})]dw \\ &= \frac{1}{\sqrt{2\pi N\kappa''(u')}}[1 + P_2(u')\frac{1}{N} + O(N^{-3/2})], \end{aligned}$$

where $w = \sqrt{N\kappa''(u')}v$ and

$$P_2(u) = \frac{3\kappa^{(4)}(u')}{4!(\kappa''(u'))^2} - \frac{15(\kappa^{(3)}(u'))^2}{2!(3!)^2(\kappa''(u'))^3}.$$

Froda and van Eeden [1] have proved $|J_2| = O(N^{-3/2})$ using a method based on [9]. By the relation between U and U', (4) follows.

In a similar way we can obtain the indirect Edgeworth approximations for the tail probabilities given in the next theorem. These are obtained in the same manner as the result of section 3.2 of [6]. Here we have used a continuity correction since the approximation is the same as one based on continuous distributions yet the true distribution of the standardised statistic is lattice with maximum jump of order $N^{-3/2}$.

THEOREM 2. *The indirect Edgeworth approximation of the tail probabilities of U is*

$$P(U \geq l) = \frac{e^{NK(u)-(l-1/2)u}}{\sqrt{2\pi}} \frac{(1 - \Phi(u\sqrt{NK''(u)}))}{\phi(u\sqrt{NK''(u)})} \times$$

$$(1 + P_{1u}(u\sqrt{NK''(u)})\frac{1}{\sqrt{N}} + P_{2u}(u\sqrt{NK''(u)})\frac{1}{N} + O(N^{-3/2})), \tag{7}$$

where

$$P_{1u}(v) = \frac{K^{(3)}(u)}{6(K''(u))^{3/2}}\left(\frac{(v^2-1)\phi(v)}{1-\Phi(v)} - v^3\right) \tag{8}$$

and

$$P_{2u}(v) = \frac{(K^{4)}(u))}{24(K''(u))^2}\left(\frac{(-v^3+v)\phi(v)}{1-\Phi(v)} + v^4\right) + \frac{(K^{(3)}(u))^2}{72(K''(u))^3}\left(\frac{(-v^5+v^3-3v)\phi(v)}{1-\Phi(v)} + v^6\right), \tag{9}$$

and u is the solution of $K'(u) = l - 1/2$.

Summing the result in Theorem 1 we can obtain the following Barndorff-Nielsen form of the tail probability. The method approximating the sum by an integral is given in Lemma 3.2.4 and Section 3.3 of [3]. We note that this approximation does not use continuity corrections.

THEOREM 3. *The tail probability approximation of U is*

$$P(U \geq \hat{x}) = (1 - \Phi(\sqrt{N}w^*))(1 + O(1/N)), \tag{10}$$

where $w^* = \hat{w} - \frac{\log \psi(\hat{w})}{N\hat{w}}$, $w^2(y) = -2(K(u(y)) - yu(y)/N)$, sign w = sign $\hat{x}$, $\hat{w} = w(\hat{x})$ and

$$\psi(\hat{w}) = \frac{\hat{w}}{(1 - e^{-u(\hat{x})})\sqrt{K''(u(\hat{x}))}}.$$

Remark: The Lugananni-Rice formula can also be obtained by summing in Theorem (1) as

$$P(U \geq \hat{x}) = \left[1 - \Phi(\sqrt{N}\hat{w}) + \frac{\psi(\hat{w}\sqrt{N})}{\sqrt{N}}\left(\frac{\psi(\hat{w})}{\hat{w}} - \frac{1}{\hat{w}}\right)\right](1 + O(1/N)). \tag{11}$$

In Section 5.2 of [3] it is shown that the ratio of the Lugananni-Rice approximation, (11), and the Barndorff-Nielsen approximation, (10), is $1 + O(N^{-1})$ in a large deviation region, that is, a region where $\hat{w}$ is of order one, and is $1 + O(N^{-3/2})$ in a normal region when $\hat{w}$ is $O(N^{-1/2})$.

3. Conditional Saddlepoint Approximation Put $S_1 = \sum_{i=1}^{N} jI_j$, $S_2 = \sum_{j=1}^{N} I_j$ where

$$P(I_j = 1) = \frac{m}{n+m} = \frac{m}{N} = p, P(I_j = 0) = \frac{n}{N} = 1 - p = q, \quad j = 1, \ldots, N.$$

Then, for the two-sample Wilcoxon rank sum statistic W, we have

$$P(W = x) = P(S_1 = x | S_2 = m).$$

Let $U = W - m(m+1)/2$. We have $P(U = x) = P(S_1 = \frac{m(m+1)}{2} + x | S_2 = m)$. Define

$$K(s,t) = \frac{1}{N}\log Ee^{sS_1 + tS_2} = \frac{1}{N}\sum_{j=1}^{N}\log(q + pe^{js+t}).$$

Then, as in [6], we have the following theorems.

THEOREM 4. *The approximation of $P(U = l)$ is*

$$P(U = l) = \frac{\sqrt{pq}\exp(NK(s,t) - (l + m(m+1)/2)s - tn)}{\sqrt{2\pi NK_{02}}\sigma_{st}} \times$$

$$(1 + P_{2st}(\sqrt{N}s\sigma_{st})\frac{1}{N} + O(N^{-3/2})), \tag{12}$$

where s, t is the solution of

$$NK_{10}(s,t) = x + m(m+1)/2, \qquad NK_{01}(s,t) = n,$$

and

$$P_{2st}(v) = -\frac{k_2}{2\sigma_{st}^2} + \frac{3k_4}{24\sigma_{st}^4} - 15\frac{k_6}{72\sigma_{st}^6} + \frac{3K_{04}}{24K_{02}^2}H_4(0) - \frac{15K_{03}^2}{72K_{02}^3}H_6(0) + \frac{1-pq}{12pq},$$

where

$$\sigma_{st}^2 = \frac{N}{N-1}\left(K_{20} - \frac{K_{11}^2}{K_{02}}\right),$$

$$k_1 = \frac{1}{2K_{02}}\left(K_{03}\frac{K_{11}}{K_{02}} - K_{12}\right),$$

$$k_2 = -\frac{1}{2K_{02}}\left(K_{04}\frac{K_{11}^2}{K_{02}^2} - 2\frac{K_{11}}{K_{02}}K_{13} + K_{22}\right) + \frac{3}{4K_{02}^2}\left(K_{03}\frac{K_{11}}{K_{02}} - K_{12}\right)^2$$
$$+\frac{1}{2}\frac{K_{03}}{K_{02}^2}\left(K_{21} - 2\frac{K_{11}K_{12}}{K_{02}}\right) - \sigma_{st}^2,$$
$$k_3 = K_{30} - 3\frac{K_{11}}{K_{02}}K_{21} + 3\frac{K_{11}^2}{K_{02}^2}K_{12} - \frac{K_{11}^3}{K_{02}^3}K_{03},$$
$$k_4 = \left(K_{40} - 4\frac{K_{11}}{K_{02}}K_{31} + 6\frac{K_{11}^2}{K_{02}^2}K_{22} - 4\frac{K_{11}^3}{K_{02}^3}K_{13} + K_{04}\frac{K_{11}^4}{K_{02}^4}\right)$$
$$-\frac{3}{K_{02}}\left(K_{21} - 2\frac{K_{11}K_{12}}{K_{02}} + \frac{K_{03}K_{11}^2}{K_{02}^2}\right)^2$$
$$+\frac{2}{K_{02}}\left(K_{03}\frac{K_{11}}{K_{02}} - K_{12}\right)\left(K_{30} - 3\frac{K_{11}}{K_{02}}K_{21} + 3\frac{K_{11}^2}{K_{02}^2}K_{12} - \frac{K_{11}^3}{K_{02}^3}K_{03}\right),$$
$$k_6 = \left(K_{30} - 3\frac{K_{11}}{K_{02}}K_{21} + 3\frac{K_{11}^2}{K_{02}^2}K_{12} - \frac{K_{11}^3}{K_{02}^3}K_{03}\right)^2,$$

where $K_{ij} = K_{ij}(s,t) = \dfrac{\partial^{i+j}K(s,t)}{\partial s^i \partial t^j}$.

THEOREM 5. *The indirect Edgeworth approximation of the tail probabilities of* U *is*

$$P(U \geq l) = \frac{\sqrt{pq}e^{NK(s,t)-(l-1/2+m(m+1)/2)s-tn}}{\sqrt{2\pi K_{02}}} \frac{(1-\Phi(\sqrt{N}s\sigma_{st}))}{\phi(\sqrt{N}s\sigma_{st})}$$
$$(1 + P_{1st}(\sqrt{N}s\sigma_{st})\frac{1}{\sqrt{N}} + P_{2st}(\sqrt{N}s\sigma_{st})\frac{1}{N} + O(N^{-3/2})), \tag{13}$$

where s, t *is the solution of*

$$NK_{10}(s,t) = l - 1/2 + m(m+1)/2, \qquad NK_{01}(s,t) = n,$$

$$P_{1st}(v) = \frac{k_1}{\sigma_{st}}\left(\frac{\phi(v)}{1-\Phi(v)} - v\right) + \frac{k_3}{6\sigma_{st}^{3/2}}\left(\frac{(v^2-1)\phi(v)}{1-\Phi(v)} - v^3\right),$$

and

$$P_{2st}(v) = \frac{k_2}{2\sigma_{st}^2}\left(\frac{-v^2\phi(v)}{1-\Phi(v)} + v^2\right) + \frac{k_4}{24\sigma_{st}^4}\left(\frac{(-v^3+v)\phi(v)}{1-\Phi(v)} + v^4\right)$$
$$+\frac{k_6}{72\sigma_{st}^6}\left(\frac{(-v^5+v^3-3v)\phi(v)}{1-\Phi(v)} + v^6\right) + \frac{3K_{04}}{24K_{02}^2} - \frac{15K_{03}^2}{72K_{02}^3} + \frac{1-pq}{12pq},$$

where k_j, $j = 1,2,3,4,6$, *are the same as in Theorem 4.*

Robinson [5] gives the coefficient of the relative order $n^{-1/2}$ for the indirect Edgeworth approximation as expression (7). And [6] has the general formula for the coefficient of the relative orders $N^{-1/2}$ and N^{-1}.

Again we can obtain the Barndorff-Nielsen approximation, and so, as in (11), the Lugananni-Rice version, as follows:

THEOREM 6. *The tail probability approximation of U is*

$$P(U \geq \hat{x}) = (1 - \Phi(\sqrt{N}w^*))(1 + O(1/N)), \tag{14}$$

where $w^* = \hat{w} - \frac{\log \psi(\hat{w})}{N\hat{w}}$, $w^2(y) = -2(K(s(y), t(y)) - yu(y)/N - pt)$, *sign* w =*sign* $\hat{x}$, $\hat{w} = w(\hat{x})$ *and*

$$\psi(\hat{w}) = \frac{\sqrt{pq}}{\sigma'_{st}\sqrt{K_{02}(s,t)}} \frac{\hat{w}}{1 - e^{-s(\hat{x})}},$$

where $\sigma'_{s,t} = \sqrt{K_{20} - K_{11}^2/K_{02}}$.

4. The Method of Froda and van Eeden The following definition and theorem are simply restatements of Lemma 4.5 and Theorem 3.2 of [1]. Let $F(x)$ be the distribution of the standardized Wilcoxon-Mann-Whitney statistic, $T = (U - EU)/Var(U)$, let $a(u)$ and $b(u)$ respectively be the mean and variance of the distribution function

$$V(t : u) = \int_{-\infty}^{x} e^{ut} dF(t)/Q_F(u), \quad -\infty < x < \infty,$$

where $Q_F(u) = Ee^{uT}$ and let $a(u) = x$.

Definition: The saddlepoint approximation $1 - F_S(x)$ of $1 - F(x)$, up to and including terms of $O(1/N)$ is, for $x > 0$, given by

$$Q_F(u)e^{-ua(u)+w^2/2}\{1 - \Phi(w)\}\{1 + \frac{3}{N}\frac{c_{20}}{4!}\frac{u}{b(u)^{3/2}}W_3(w)\}, \tag{15}$$

where $w = ub(u)^{1/2}$, $W_3(v)$ is given by

$$W_3(v) = \frac{(v^2 - 1)\phi(v)}{1 - \Phi(v)} - v^3,$$

and u is such that $a(u) = x$.

THEOREM 7. *(Saddlepoint expansion) There exist $\eta > 0$ and $M_0 > 1$ such that, for $M_0 < u < \eta N^{1/2}$, and $N \to \infty$,*

$$1 - F(x) = \{1 - F_S(x)\}\{1 + O(u^3/N^{3/2})\}.$$

In particular, if $u = o(N^{1/2})$, and thus x is $o(N^{1/2})$,

$$1 - F(x) = \{1 - F_S(x)\}\{1 + o(1)\},$$

if $u = O(N^{1/6})$, and thus x is $o(N^{1/6})$,

$$1 - F(x) = \{1 - F_S(x)\}\{1 + O(1/N)\},$$

while, if u (and thus x) stay bounded,

$$1 - F(x) = \{1 - F_S(x)\}\{1 + O(1/N^{3/2})\}.$$

Remark: This result does not have the relative error properties of the saddlepoint method and cannot be expected to give good approximations in the large deviation region. The result is related to (7) but uses only the term of order $N^{-1/2}$ and a first order Taylor approximation of the coefficient in $P_{1u}(v)$ defined in (8).

5. Numerical Comparison Table 1 gives numerical comparison of the approximations of probabilities from the method based on the moment generating function and conditional methods for $n = m = 5$. Row (4) is obtained by equation (4) (new method) including the $O(1/N)$ terms and row (4)* is the saddlepoint approximation from (4) excluding the $O(1/N)$ terms. So Row (4) has the relative error $N^{-3/2}$ and row (4)* has the relative error N^{-1}. In the same way row (12) and (12)* are based on equation (12) (conditional method). We can see the new method is a little better, although not at all values, than the conditional method under the same order of relative error. Table 2 gives numerical comparisons of the

Table 1: Approximations of $P(U = l)$ for the two-sample Wilcoxon for $n = m = 5$

l	1	2	3	4	5
exact	.00397	.00794	.01190	.01984	.02778
(4)	.00463	.00807	.01293	.01925	.02696
(12)	.00556	.00894	.01389	.02033	.02817
(4)*	.00583	.00928	.01433	.02092	.02894
(12)*	.00689	.01011	.01510	.02165	.02961
l	6	7	8	9	10
exact	.03571	.04365	.05556	.06349	.07143
(4)	.03577	.04523	.05472	.06353	.07095
(12)	.03710	.04667	.05626	.06516	.07263
(4)*	.03810	.04791	.05774	.06686	.07453
(12)*	.03868	.04838	.05809	.06710	.07465

approximations of tail probabilities based on the new and conditional methods for $n = m = 5$. Row (7) is obtained by equation (7) (indirect Edgeworth) including the $O(1/N)$ terms and row (7)* includes neither the $O(N^{-1/2})$ nor the $O(1/N)$ terms. So Row (7) has the relative error of $O(N^{-3/2})$ and row (7)* has the relative error of $O(N^{-1/2})$. Row (10) is obtained by equation (10) (Barndorff-Nielsen approximation based on the new method) which has the relative error of $O(N^{-1})$. Row (13) and (13)* are based on equation (13) (conditional method). Row (14) is obtained by equation (14) (Barndorff-Nielsen approximation based on conditional method) which also has the relative error of $O(N^{-1})$. Row (15) is obtained by equation (15) (Froda and van Eeden). In general, we see the new method is a little better than the conditional method under the same order of relative error, but again not at all points. The method of [1] is very good in the center of the distribution but quite poor in the tails. Table 3 and Table 4 are for $m = 10$, $n = 6$. The results are similar

Table 2: Approximations of $P(U \leq l)$ for the two-sample Wilcoxon for $n = m = 5$

l	1	2	3	4	5
exact	.00794	.01587	.02778	.04762	.07540
(7)	.00803	.01610	.02905	.04833	.07528
(13)	.00828	.01632	.02920	.04834	.07511
(10)	.00845	.01639	.02933	.04865	.07570
(14)	.00912	.01702	.03004	.04948	.07667
(7)*	.00667	.01401	.02597	.04402	.06962
(13)*	.00644	.01330	.02446	.04136	.06542
(15)	.02690	.02701	.03713	.05469	.08023
l	6	7	8	9	10
exact	.1111	.1548	.2103	.2738	.3452
(7)	.1110	.1561	.2107	.2741	.3449
(13)	.1106	.1555	.2099	.2732	.3441
(10)	.1116	.1569	.2117	.2754	.3464
(14)	.1127	.1582	.2133	.2771	.3483
(7)*	.1040	.1481	.2023	.2661	.3384
(13)*	.0980	.1401	.1925	.2554	.3283
(15)	.1146	.1586	.2121	.2746	.3449

to those with $n = m = 5$.

Table 3: Approximations of $P(U = l)$ for the two-sample Wilcoxon for $n = 6$, $m = 10$

l	1	2	3	4	5
exact	.000125	.000250	.000375	.000624	.000874
(4)	.000145	.000246	.000398	.000614	.000911
(12)	.000140	.000244	.000396	.000612	.000911
(4)*	.000180	.000284	.000444	.000673	.000986
(12)*	.000228	.000331	.000500	.000743	.001080
l	6	7	8	9	10
exact	.00137	.00175	.00250	.00325	.00437
(4)	.00131	.00183	.00248	.00330	.00429
(12)	.00131	.00183	.00249	.00332	.00432
(4)*	.00140	.00195	.00263	.00348	.00452
(12)*	.00152	.00208	.00280	.00369	.00477

REFERENCES

[1] S. Froda and C. van Eeden. A uniform saddlepoint expansion for the null-distribution of the Wilcoxon-Mann-Whitney statistics. *Canadian Journal of Statistics*, 28:137–149, 2000.

Table 4: Approximations of $P(U \leq l)$ for the two-sample Wilcoxon for $n = 6$, $m = 10$

l	1	2	3	4	5
exact	.000250	.000450	.000874	.00150	.00237
(7)	.000252	.000497	.000894	.00151	.00242
(13)	.000265	.000514	.000918	.00154	.00247
(10)	.000264	.000503	.000897	.00151	.00242
(14)	.000288	.000526	.000925	.00154	.00246
(7)*	.000203	.000419	.000774	.00133	.00216
(13)*	.000208	.000422	.000774	.00132	.00214
(15)	.00537	.00342	.00319	.00356	.00439
l	6	7	8	9	10
exact	.00375	.00550	.00780	.0112	.0156
(7)	.00373	.00556	.00805	.0114	.0157
(13)	.00379	.00564	.00815	.0115	.0158
(10)	.00373	.00556	.00804	.0113	.0156
(14)	.00378	.00562	.00812	.0114	.0158
(7)*	.00337	.00506	.00738	.0105	.0145
(13)*	.00333	.00499	.00726	.0103	.0143
(15)	.00753	.0101	.0134	.0177	.0232

[2] J. Hájek and Z. Šidák. *Theory of Rank Tests.* Academic Press, New York, 1967.

[3] J.L. Jensen. *Saddlepoint Approximations.* Number 16 in Oxford Statistical Science Series. Clarenden Press, Oxford, 1995.

[4] R. Jin and J. Robinson. Saddlepoint approximation near the endpoints of the support. *Statistics and Probability Letters*, 45:295–303, 1999.

[5] J. Robinson. Saddlepoint approximations for permutation tests. *Journal of the Royal Statistical Society Series B*, 44:91–101, 1982.

[6] J. Robinson, T. Höglund, L. Holst, and M. P. Quine. On approximating probabilities for small and large deviations in r^{d1}. *Annals of Probability*, 18:727–753, 1990.

[7] I. M. Skovgaard. Saddlepoint expansions for conditional distributions. *Journal of Applied Probability*, 24:875–887, 1987.

[8] D. van Dantzig. *Kader cursus Mathematische Statistiek*, pages 301–304. Mathematisch Centrum Amsterdam, 1947-1950.

[9] W. R. van Zwet. On the Edgeworth expansion for the simple linear rank statistic. In B. V. Gnedenko, M. L. Puri, and I. Vincze, editors, *Colloquia Mathematica Societatis Janos Bolyai*, pages 889–909. North-Holland, Amsterdam, 1980.

School of Mathematics and Statistics
University of Sydney NSW 2006 AUSTRALIA
johnr@maths.usyd.edu.au