1 Introduction

The usual saddlepoint approximations of Daniels (1954) are based on asymptotic normality of tilted means. However, if the means have finite support then the approximations can fail near the endpoints of the support where asymptotic normality of the tilted means does not hold. This paper gives new accurate methods for saddlepoint approximations near the extremes of the support for binomial, hypergeometric, one and two sample Wilcoxon statistics. We will consider only the lower extremes of the support in each case, the upper extreme having analogous results.

For a binomial statistic B, saddlepoint approximations are obtained by considering the asymptotic probabilities for a tilted variable B_t, where t is chosen so that $EB_t = k$. There are 3 cases, namely, k equal to 0, fixed k and k tending to infinity. When k is equal to 0, the probability that B_t is equal to 0 is 1, when k is fixed, B_t converges to a Poisson variable with parameter k as noted by Jensen(1995, section 10.4) and when k tends to infinity, B_t is asymptotically normal. In this last case, we obtain a relative error for the saddlepoint approximation as $1/k$. This extends the known result for $cn < k < n/2$ where $0 < c < 1/2$ in which case the saddlepoint approximation has relative error $1/n$.

For a hypergeometric variable H we obtain similar results to those of the binomial variable, using an approach based on conditioning.

The one-sample Wilcoxon statistic W can be expressed as the sum of independent variables. Saddlepoint approximations are obtained by considering the asymptotic probabilities for a tilted variable W_t, where t is chosen so that $EW_t = k$ (see, for example, Robinson (1982)). Also there are 3 cases, namely, k equal to 0, fixed k and k tending to infinity. When k is equal to 0, W_t is equal to 0 with probability 1, when k is fixed, a new combinatorial result is obtained and when k tends to infinity, W_t is asymptotically normal. In this last case, we obtain a relative error for the saddlepoint approximation as $1/\sqrt{k}$ which extends the known result for $cn(n+1)/2 < k < n(n+1)/4$ where $0 < c < 1/2$ in which case the relative error is $1/n$.

For the two-sample Wilcoxon rank sum statistic, $U_{m,n}$, conditioning is used to obtain results like those of the one-sample Wilcoxon statistic.

In the sequel, for two sequences a_k and b_k, $a_k \sim b_k$ means there exist constants $C > c > 0$ such that $cb_k < a_k < Cb_k$.

2 Binomial Distribution

Assume $X_1, \ldots, X_n$ are i.i.d. and

$$P(X_i = 1) = p, P(X_i = 0) = 1 - p = q, i = 1, \ldots, n.$$

Put

$$B = n\bar{X} = X_1 + \ldots + X_n.$$

Then define tilted random variable $X_{1t}, \ldots, X_{nt}$ as

$$f_t(x) = P(X_{it} = x) = \frac{e^{tx} p^x q^{1-x}}{q + pe^{tx}} = e^{-K(t)+tx} f(x), \tag{2.1}$$

where $f(x) = p^x q^{1-x}$, $x = 0, 1$ and

$$K(t) = q + pe^{tx}.$$

Put

$$B_t = X_{1t} + \ldots + X_{nt}.$$

Then, for $x = k/n$, we have

$$P(B = nx) = e^{-n\Lambda(x)} P(B_{t(x)} = nx),$$

where

$$t(x) = \log \frac{(1-p)x}{p(1-x)}$$

is the solution of $K'(t) = x$ and

$$\Lambda(x) = -K(t(x)) + xt(x) = (1-x)\log\frac{1-x}{1-p} + x\log\frac{x}{p}.$$

Also,

$$B_{t(x)} \sim bin(n, x) \tag{2.2}$$

because

$$P(X_{it} = 1) = \frac{pe^{t(x)}}{q + pe^{t(x)}} = \frac{\frac{(1-p)x}{p(1-x)}p}{q + p\frac{(1-p)x}{p(1-x)}} = x.$$

Consider the following 3 cases.
(a) $k \to \infty$ and $n - k \to \infty$
(b) $k = 0$(or n),
(c) k is fixed and finite integer (or $n - k$ is fixed and finite integer).
In each of these cases we have a different limiting distribution for $B_{t(x)}$.

Theorem 2.1 *For B a binomial (n, p) variable, as $n \to \infty$, we have*
(a) for $k \to \infty$, $k \leq np$,

$$P(B = k) = \frac{e^{-n\Lambda(k/n)}}{\sqrt{2\pi k(1 - k/n)}}(1 + O(\frac{1}{k}));$$

(b) for $k = 0$,

$$P(B = 0) = e^{-n\Lambda(0)} = q^n;$$

(c) for k a fixed integer,

$$P(B = k) = e^{-n\Lambda(k/n)} \frac{k^k e^{-k}}{k!}(1 + O(\frac{1}{n})). \tag{2.3}$$

Proof: (a) when $k \to \infty$, we have

$$\begin{cases} K''(t) = \frac{pqe^t}{(q+pe^t)^2} \sim \frac{k}{n}, \\ K^{(3)}(t) = \frac{pqe^t(q-pe^t)}{(q+pe^t)^3} \sim \frac{k}{n}, \\ K^{(4)}(t) = \frac{pqe^t[(q-2pe^t)(q+pe^t)-3pe^t(q-pe^t)]}{(q+pe^t)^4} \sim \frac{k}{n}, \\ K^{(5)}(t) \sim \frac{k}{n} \end{cases} \tag{2.4}$$

So $nK''(t) \to \infty$ as $k \to \infty$. We know

$$P(B_t = k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \exp\{-i\xi k + nK(t+i\xi) - nK(t)\}d\xi = J_1 + J_2,$$

where

$$J_1 = \frac{1}{2\pi}\int_{-\epsilon}^{\epsilon} \exp\{-i\xi k + nK(t+i\xi) - nK(t)\}d\xi$$

and

$$J_2 = \frac{1}{2\pi}\int_{(-\pi,\pi)\setminus(-\epsilon,\epsilon)} \exp\{-i\xi k + nK(t+i\xi) - nK(t)\}d\xi.$$

Now we look at J_1 first.

$$J_1 = \frac{1}{2\pi}\int_{-\epsilon}^{\epsilon} \exp\{-\frac{1}{2}nK''(t)\xi^2 + \frac{nK^{(3)}(t)}{3!}(i\xi)^3 + \frac{nK^{(4)}(t)}{4!}(i\xi)^4 + \frac{nK^{(5)}(t+\theta i\xi)}{5!}(i\xi)^5\}d\xi.$$

Put $\eta = \sqrt{nK''(t)}\xi$. Then, from

$$\frac{nK^{(3)}(t)}{(nK''(t))^{3/2}} \sim \frac{1}{k^{1/2}},\ \frac{nK^{(4)}(t)}{(nK''(t))^2} \sim \frac{1}{k},\ \frac{K^{(5)}(t)}{(K''(t))^{5/2}} \sim \frac{1}{k^{3/2}},$$

when $\epsilon > 0$ is small enough, we have

$$\begin{aligned} J_1 &= \frac{1}{2\pi\sqrt{nK''(t)}}\int_{-\epsilon\sqrt{nK''(t)}}^{\epsilon\sqrt{nK''(t)}} e^{-\frac{1}{2}\eta^2}(1 + \frac{K^{(3)}(t)}{3!K''(t)^{3/2}\sqrt{n}}(i\eta)^3 + \frac{K^{(4)}(t)}{4!K''(t)^2 n}(i\eta)^4 \\ &\quad + \frac{K^{(3)}(t)^2}{3!K''(t)^3 n}(i\eta)^6 + P_5(\eta)e^{\frac{1}{4}\eta^2})d\eta, \end{aligned}$$

where $P_5(\eta)$ is the polynomial of degree 5. Because

$$\frac{K^{(3)}(t)}{(K''(t))^{3/2}\sqrt{n}} \sim \frac{1}{k^{1/2}},\ \frac{K^{(4)}(t)}{(K''(t))^2 n} \sim \frac{1}{k},$$

we have

$$J_1 = \frac{1}{\sqrt{2\pi nK''(t)}}(1 + O(\frac{1}{k})).$$

Now let us look at J_2. J_2 is rewritten as

$$J_2 = \frac{1}{2\pi}\int_{(-\pi,\pi)\backslash(-\epsilon,\epsilon)} e^{-i\xi k}\left(\frac{q+pe^{t+i\xi}}{q+pe^t}\right)^n d\xi.$$

When $\epsilon \le |\xi| \le \pi$, we have

$$\left|\frac{q+pe^{t+i\xi}}{q+pe^t}\right|^n \le \frac{((q+pe^t)^2 - 2pqe^t(1-cos\epsilon))^{n/2}}{(q+pe^t)^n} = \left(1 - 2(1-\frac{k}{n})(1-cos\epsilon)\frac{k}{n}\right)^{n/2}.$$

We know $x = k/n \le p$. So, when n is big enough and $\epsilon < \pi/3$, we have

$$\left|\frac{q+pe^{t+i\xi}}{q+pe^t}\right|^n \le (1-\frac{\delta k}{n})^{n/2} = e^{-\delta_1 k/2},$$

where $0 < \delta, \delta_1 < 1$. Therefore,

$$J_2 = O(k^{-3/2}).$$

We know

$$P(B = k) = e^{-n\Lambda(x)} P(S_t = k).$$

Therefore,

$$P(B = nx = k) = \frac{e^{-n\Lambda(x)}}{\sqrt{2\pi nx(1-x)}}(1 + O(\frac{1}{k})).$$

(b) For $k = 0$, from (2.2), we have

$$B_{t(x)} = B_{t(0)} \sim bin(n, 0).$$

So $P(B_{t(0)} = 0) = 1$ and the correct saddlepoint approximation is

$$P(B = 0) = e^{-n\Lambda(0)}1 = (1-p)^n = q^n.$$

This is indeed the exact probability.

(c) The tilted variable $B_{t(x)}$ is binomial (n, x). Since $nx = k$ is fixed, $B_{t(x)}$ converges in distribution to a Poisson(k) variable. Indeed

$$P(B_{t(x)} = k) = \frac{k^k e^{-k}}{k!}(1 + O(\frac{1}{n}))$$

and (2.3) follows. □

Remark: In (b), The usual saddlepoint approximation tends to ∞ as $n \to \infty$. In (c), we can use Stirlings formula to obtain

$$P(B = k) = e^{-n\Lambda(k/n)}\frac{1}{\sqrt{2\pi k}}(1 + \frac{1}{12k} + O(\frac{1}{k^2})).$$

In particular, for $k = 1$, the ratio of the approximation of (2.3) to the usual saddlepoint approximation is close to $\sqrt{2\pi}/e = 0.922$. So even for $k = 1$ the usual saddlepoint approximation based on a normal approximation for the tilted mean is quite good.

3 Hypergeometric Distribution

Put $S_1 = \sum_{i=1}^N a_i I_i$ and $S_2 = \sum_{i=1}^N I_i$ where

$$\begin{cases} a_i = 1 & i = 1, \ldots, m, \\ a_i = 0 & i = m+1, \ldots, N \end{cases}$$

and

$$\begin{cases} P(I_i = 1) = \frac{n}{N} = p, \\ P(I_i = 0) = 1 - p = q \quad i = 1, \ldots, N. \end{cases}$$

$X_1, \ldots, X_n$ are a sample from $a_1, \ldots, a_N$. Then the hypergeometric variable $H = X_1 + \ldots + X_n$ and

$$P(H = k) = P(S_1 = k | S_2 = n).$$

We have $S_1 = B_1$, $S_2 = B_1 + B_2$ where $B_1 = \sum_{i=1}^m I_i$, $B_2 = \sum_{i=m+1}^N I_i$. Then, as binomial situation, define tilted random variable $X_{1t}, \ldots, X_{Nt}$ as

$$f_t(x) = P(X_{1t} = x) = \frac{e^{tx} p^x q^{1-x}}{q + pe^{tx}} = e^{-K(t)+tx} f(x), \tag{3.5}$$

where $f(x) = p^x q^{1-x}$, x=0,1 and

$$K(t) = q + pe^{tx}.$$

Let $B_{1\tau} = \sum_{i=1}^m I_{i\tau}$, $B_{2\eta} = \sum_{i=m+1}^N I_{i\eta}$ and $S_{2t} = \sum_{i=1}^N I_{it}$. Then

$$\begin{aligned} P(H = k) &= \frac{P(B_1 = k, B_1 + B_2 = n)}{P(S_2 = n)} = \frac{P(B_1 = k)P(B_2 = n - k)}{P(S_2 = n)} \\ &= e^{-m\Lambda(k/m) - (N-m)\Lambda((n-k)/(N-m)) + N\Lambda(n/N)} \frac{P(B_{1\tau} = k)P(B_{2\eta} = n - k)}{P(S_{2t} = n)}. \end{aligned}$$

From $K'(\tau) = k/m$, $K'(\eta) = (n-k)/(N-m)$ and $K'(t) = n/N$ we can get $\tau = \tau(x)$, $\eta = \eta(x)$ and $t = t(x)$ respectively where $x = k/n$. As in the binomial situation, we know $B_{1\tau}$ is binomial $(m, k/m)$, $B_{2\eta}$ is binomial $(N-m, (n-k)/(N-m))$ and S_{2t} is binomial $(n, n/N)$. We consider $n < min(m, N-m)$ situation only in the following theorem. Other situations can be handled similarly.

Theorem 3.1 *For H a hypergeometric variable, for $n < min(m, N-m)$ and $N \to \infty$ we have*
(a) for $k \to \infty$, $k \le n/2$,

$$P(H = nx) = e^{-N\Lambda_1(x)} \sqrt{\frac{q}{2\pi nx(1-x)(1-px/p')(1-p(1-x)/q')}} (1 + O(\frac{1}{k'})), \tag{3.6}$$

where $p' = m/N = 1 - q'$, $k' = min(k, N-m-n)$ *and*

$$\Lambda_1(x) = p'\Lambda(px/p') + q'\Lambda(p(1-x)/q') - N\Lambda(p).$$

(b) for $k = 0$,

$$P(H=0) = e^{-N\Lambda_1(0)}\sqrt{\frac{qq'}{q'-p}}(1+O(\frac{1}{n'})) = \frac{q^{N-n+1/2}q'^{N-m+1/2}}{(q'-p)^{N-m-n+1/2}}(1+O(\frac{1}{n'})), \qquad (3.7)$$

where $n' = min(n, N-m-n)$.
(c) for k *a fixed integer,*

$$P(H=nx) = e^{-N\Lambda_1(x)}\frac{k^k e^{-k}}{k!}\sqrt{\frac{qq'}{q'-p}}(1+O(\frac{1}{n'})). \qquad (3.8)$$

Proof: (a) The proof of the result follows from Theorem 2.1.
(b) When $k=0$, we have $P(B_{1\tau}=0)=1$ and

$$\begin{aligned} P(H=0) &= \exp\{-N\lim_{x\to 0}\Lambda_1(x)\}\frac{P(B_{1\tau}=0)P(B_{2\eta}=n)}{P(S_2=n)} \\ &= \exp\{-N\lim_{x\to 0}\Lambda_1(x)\}\sqrt{\frac{n(1-p)}{n(1-\frac{n}{(N-m)})}}(1+O(\frac{1}{k'})) \\ &= \frac{(1-p)^{N-n}(1-p')^{N-m}}{(1-p'-p)^{N-m-n}}\sqrt{\frac{(1-p')(1-p)}{1-p'-p}}(1+O(\frac{1}{k'})) \\ &= \frac{q^{N-n+1/2}q'^{N-m+1/2}}{(q'-p)^{N-m-n+1/2}}(1+O(\frac{1}{k'})), \end{aligned}$$

where $k' = min(n, N-n-m)$.

(c) For k a fixed integer, the tilted variable $B_{1\tau}$ is binomial $(m, k/m)$. Since $m*k/m = k$ is fixed, $B_{1\tau}$ converges in distribution to a Poisson(k) variable. The proof of the result follows from Theorem 2.1. □

4 One-sample Wilcoxon Statistic

Put $W = \sum_{j=1}^n jI_j$ where $P(I_j=0) = P(I_j=1) = \frac{1}{2}$, $j=1,\ldots,n$. So, the moment generating function of W is

$$M(t) = \prod_{j=1}^n(\frac{1}{2}+\frac{1}{2}e^{jt}).$$

Put $K(t) = \frac{1}{n}\sum_{j=1}^n \log(\frac{1}{2}+\frac{1}{2}e^{jt})$. Then the density functions of the tilted variables I_{jt}, $j=1,\ldots,n$ are

$$P(I_{jt}=1) = \frac{e^{jt}}{1+e^{jt}}, \quad P(I_{jt}=0) = \frac{1}{1+e^{jt}}, \quad j=1,\ldots,n. \qquad (4.9)$$

Put $W_t = \sum_{j=1}^n jI_{jt}$. Then,

$$f_{W_t}(nw) = e^{ntw - nK_n(t)} f_W(nw).$$

From

$$K'(t) = \frac{1}{n}\sum_{j=1}^n \frac{je^{jt}}{1+e^{jt}} = w = k/n, \tag{4.10}$$

we can get saddlepoint $t = t_n(k)$. Put $\Lambda(w) = wt(w) - K(t(w))$. Then

$$f_W(nw) = e^{-n\Lambda(w)} f_{W_t}(nw). \tag{4.11}$$

Put $h_n(t) = nK'(t) = \sum_{j=1}^n \frac{je^{jt}}{1+e^{jt}}$. Then $h_n'(t) = nK''(t) = \sum_{j=1}^n \frac{j^2e^{jt}}{(1+e^{jt})^2} > 0$. So, $h_n(t)$ is an increasing function. We can get the following Lemma.

Lemma 4.1 *The solution t_n of the equation*

$$h_n(t) = \sum_{j=1}^n \frac{je^{jt}}{1+e^{jt}} = k$$

satisfies

$$\begin{cases} t_n(k) = -\infty & \text{for } k = 0, \\ t_n(k) < 0 & \text{for } 0 < k < n(n+1)/4, \\ t_n(k) = 0 & \text{for } k = n(n+1)/4, \\ t_n(k) > 0 & \text{for } n(n+1)/4 < k < n(n+1)/2, \\ t_n(k) = \infty & \text{for } k = n(n+1)/2. \end{cases} \tag{4.12}$$

and $t_n(k)$ is an increasing function with respect to k as n is fixed.

Note $t = t(w) = t_n(k)$. Without losing generality, we consider $k \le n(n+1)/4$ only from now on.

Lemma 4.2 *For $1 \le k \le n(n+1)/c$, where $c > 4$, the solution $t_n(k)$ of the equation*

$$h_n(t) = \sum_{j=1}^n \frac{je^{jt}}{1+e^{jt}} = k \tag{4.13}$$

satisfies

$$-t_n(k) \sim \frac{1}{\sqrt{k}}, \ \text{as } n \to \infty,$$

where this means

$$\frac{C_1}{\sqrt{k}} < -t_n(k) < \sqrt{\frac{1+\sqrt{5}}{2}}\frac{1}{\sqrt{k}},$$

where $C_1 = e^{-2/\sqrt{k}}((-c_1-1)e^{-c_1}+1)/2 > 0$, *where* $c_1 = \log(c/4) > 0$. *Also, for* $k < n(n+1)/4$,

$$-t_n(k) = O(\frac{1}{\sqrt{k}}), \text{ as } n \to \infty.$$

Proof: Put $t = t_n(k)$. We know $t < 0$. Therefore,

$$\int_0^1 y e^{nt(y+\frac{1}{n})} dy < \frac{1}{n^2}\sum_{j=1}^n j e^{jt} = \sum_{j=1}^n \frac{j}{n} e^{\frac{j}{n}nt}\frac{1}{n} < \int_0^1 (y+\frac{1}{n}) e^{nty} dy.$$

So,

$$e^t n^2 \int_0^1 y e^{nty} dy < \sum_{j=1}^n j e^{jt} < n^2 \int_0^1 y e^{nty} dy + n\int_0^1 e^{nty} dy.$$

Now

$$n^2 \int_0^1 y e^{nty} dy = \int_0^{-nt} z e^{-z} dz \frac{1}{t^2} = ((nt-1)e^{nt}+1)\frac{1}{t^2},$$

$$n\int_0^1 e^{nty} dy = \int_0^{-nt} e^{-z} d(-z)\frac{1}{t} = (e^{nt}-1)\frac{1}{t},$$

and $1 < 1 + e^{jt} < 2$. So,

$$\begin{aligned} e^t((nt-1)e^{nt}+1)\frac{1}{2t^2} &< k = \sum_{j=1}^n \frac{je^{jt}}{1+e^{jt}} \\ &< ((nt-1)e^{nt}+1+t(e^{nt}-1))\frac{1}{t^2}. \end{aligned}$$

We know

$$\frac{n(n+1)}{c} > k = \sum_{j=1}^n \frac{je^{jt}}{1+e^{jt}} > \sum_{j=1}^n \frac{je^{nt}}{1+e^{jt}} > \frac{1}{2}e^{nt}\sum_{j=1}^n j = \frac{n(n+1)}{4}e^{nt}.$$

That is $e^{nt} < \frac{4}{c} = e^{-c_1} < 1$ where $c_1 > 0$. So,

$$nt < -c_1 < 0.$$

Also $f(x) = (x-1)e^x + 1$ is a decreasing function when $x \le 0$. So,

$$0 = f(0) < (-c_1-1)e^{-c_1}+1 < (x-1)e^x+1 \le 1$$

when $x < -c_1$. Therefore we have

$$e^t((-c_1-1)e^{-c_1}+1)\frac{1}{2k} < t^2 < (1-t)\frac{1}{k}. \tag{4.14}$$

When $k = 1$ we have $t^2 + t - 1 < 0$. So, $-(1+\sqrt{5})/2 < t_n(1) < 0$. From Lemma 4.1, we know $t_n(k)$ is increasing function with respect to k. So,

$$t_n(1) < t_n(k) \text{ for } k > 1.$$

Therefore, from (4.14),

$$\frac{C_1}{\sqrt{k}} < -t_n(k) < \sqrt{\frac{1+\sqrt{5}}{2}}\frac{1}{\sqrt{k}} \quad \text{as } n \to \infty,$$

where $C_1 = \sqrt{e^{-2/\sqrt{k}}((-c_1-1)e^{-c_1}+1)/2}$. Also because $t_n(k)$ is an increasing function with respect to k, we have

$$-t_n(k) = O(\frac{1}{\sqrt{k}}), \text{ as } n \to \infty$$

when $1 \leq k \leq n(n+1)/4$. $\square$

Lemma 4.3 *If k is fixed, we have*

$$t_n(k) \to t(k) \qquad \textit{as } N \to \infty,$$

where $t(k) < 0$ is dependent on k only.

Proof: We have

$$h_n(t_n(k)) = \sum_{j=1}^{n} \frac{je^{jt_n(k)}}{1+e^{jt_n(k)}} = k$$

and

$$\sum_{j=1}^{n} \frac{je^{jt_{n+1}(k)}}{1+e^{jt_{n+1}(k)}} = k - \frac{(n+1)e^{(n+1)t_{n+1}(k)}}{1+e^{(n+1)t_{n+1}(k)}} < k.$$

We know $h_n(t)$ is an increasing function. Therefore,

$$t_{n+1}(k) \leq t_n(k) < 0.$$

From Lemma 4.2, we know $t_n(k)$ is absolutely bounded for any n. Therefore,

$$t_n(k) \to t(k) \qquad \text{as } N \to \infty. \qquad \square$$

Consider the following 3 cases.
(a) $k \to \infty$ and $n(n+1)/2 - k \to \infty$
(b) $k = 0$(or n),
(c) k is fixed and finite integer (or $n(n+1)/2 - k$ is fixed and finite integer).
In each of these cases we have a different limiting distribution for $W_{t(w)}$.

Theorem 4.1 *Assume $W = \sum_{j=1}^n jI_j$ where $P(I_j = 0) = P(I_j = 1) = \frac{1}{2}$, $j = 1, \ldots, n$. Then, as $n \to \infty$ we have*
(a) for $k \to \infty$, $k \leq n(n+1)/4$,

$$P(W = k) = e^{-n\Lambda(k/n)} \frac{1}{\sqrt{2\pi K''(t_n(k))}}(1 + O(k^{-1/2})); \tag{4.15}$$

(b) for $k = 0$,

$$P(W = 0) = e^{-n\Lambda(0)} = 2^{-n};$$

(c) for k *a fixed number and* $k \leq n$,

$$P(W = k) = e^{-n\Lambda(k/n)} f_{W_{t(k)}}(k)(1 + O(ne^{nt(k)})), \tag{4.16}$$

where $t(k) = \lim_{n\to\infty} t_n(k)$, $f_{W_{t(k)}}(k) = B(k)e^{kt(k)}/(\prod_{j=1}^{\infty}(1 + e^{jt(k)}))$ *and* $B(k)$ *is the number of sets of distinct positive integers whose sum is* k.

Proof: (a) When $k \to \infty$ and $k \leq n(n+1)/4$, note $-t = O(k^{-1/2})$ from Lemma 4.2, we have

$$\begin{cases} K''(t) = n^2 \frac{1}{n} \sum_{j=1}^{n} \frac{(\frac{j}{n})^2 e^{\frac{j}{n}nt}}{(1+e^{\frac{j}{n}nt})^2} \sim n^2 \sum_{j=1}^{n} (\frac{j}{n})^2 e^{\frac{j}{n}nt} \frac{1}{n} \sim n^2 \int_0^1 y^2 e^{ynt} dy \sim n^{-1}k^{3/2}, \\ K^{(3)}(t) = n^3 \frac{1}{n} \sum_{j=1}^{n} \frac{(\frac{j}{n})^3 e^{\frac{j}{n}nt}(1-e^{\frac{j}{n}nt})}{(1+e^{\frac{j}{n}nt})^3} \sim n^3 \sum_{j=1}^{n} (\frac{j}{n})^3 e^{\frac{j}{n}nt} \frac{1}{n} \sim n^3 \int_0^1 y^2 e^{ynt} dy \sim n^{-1}k^2, \\ K^{(4)}(t) \sim n^{-1}k^{5/2}, \\ K^{(5)}(t) \sim n^{-1}k^3 \end{cases} \tag{4.17}$$

So $nK''(t) \sim k^{3/2} \to \infty$. We know

$$P(W_t = k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp\{-i\xi k + nK(t + i\xi) - nK(t)\} d\xi = J_1 + J_2,$$

where

$$J_1 = \frac{1}{2\pi} \int_{-\epsilon k^{-1/2}}^{\epsilon k^{-1/2}} \exp\{-i\xi k + nK(t + i\xi) - nK(t)\} d\xi$$

and

$$J_2 = \frac{1}{2\pi} \int_{(-\pi,\pi)\setminus(-\epsilon k^{-1/2},\epsilon k^{-1/2})} \exp\{-i\xi k + nK(t + i\xi) - nK(t)\} d\xi.$$

Now we look at J_1 first.

$$J_1 = \frac{1}{2\pi} \int_{-\epsilon k^{-1/2}}^{\epsilon k^{-1/2}} \exp\{-\frac{1}{2}nK''(t)\xi^2 + \frac{nK^{(3)}(t)}{3!}(i\xi)^3 + \frac{nK^{(4)}(t)}{4!}(i\xi)^4 + \frac{nK^{(5)}(t + \theta i\xi)}{5!}(i\xi)^5\} d\xi.$$

Put $\eta = \sqrt{nK''(t)}\xi$. Then, from

$$\frac{nK^{(3)}(t)}{(nK''(t))^{3/2}} \sim \frac{1}{k^{1/4}}, \quad \frac{nK^{(4)}(t)}{(nK''(t))^2} \sim \frac{1}{k^{1/2}}, \quad \frac{K^{(5)}(t)}{(K''(t))^{5/2}} \sim \frac{1}{k^{3/4}},$$

when $\epsilon > 0$ is small enough, we have

$$\begin{aligned} J_1 \quad &= \quad \frac{1}{2\pi\sqrt{nK''(t)}} \int_{-\epsilon k^{-1/2}\sqrt{nK''(t)}}^{\epsilon k^{-1/2}\sqrt{nK''(t)}} e^{-\frac{1}{2}\eta^2} (1 + \frac{K^{(3)}(t)}{3!K''(t)^{3/2}\sqrt{n}}(i\eta)^3 + \frac{K^{(4)}(t)}{4!K''(t)^2 n}(i\eta)^4 \\ &\quad + \frac{K^{(3)}(t)^2}{2!(3!)^2 K''(t)^3 n}(i\eta)^6 + P_5(\eta)e^{\frac{1}{4}\eta^2}) d\eta, \end{aligned}$$

where $P_5(\eta)$ is the polynomial of degree 5. Because

$$\frac{K^{(3)}(t)}{(K''(t))^{3/2}\sqrt{n}} \sim \frac{1}{k^{1/4}}, \ \frac{K^{(4)}(t)}{(K''(t))^2 n} \sim \frac{1}{k^{1/2}},$$

we have

$$J_1 = \frac{1}{\sqrt{2\pi n K''(t)}}(1 + O(\frac{1}{k^{1/2}})).$$

Now let us look at J_2. We use a proof related to one used in Albers, W., Bickel, P. J. and van Zwet, W. R. (1976). J_2 is rewritten as

$$J_2 = \frac{1}{2\pi}\int_{(-\pi,\pi)\setminus(-\epsilon k^{-1/2},\epsilon k^{-1/2})} e^{-i\xi l}\prod_{j=1}^{n}\frac{1+e^{jt+ij\xi}}{1+e^{jt}}d\xi. \tag{4.18}$$

When $\epsilon k^{-1/2} \leq |\xi| \leq \pi$ where ϵ is small enough, note $t \sim -k^{-1/2}$, we have

$$\begin{aligned} \left|\prod_{j=1}^{n}\frac{1+e^{jt+ij\xi}}{1+e^{jt}}\right| &= \prod_{j=1}^{n}\frac{((1+e^{jt})^2 - 2e^{jt}(1-cos(j\xi)))^{1/2}}{(1+e^{t})} \\ &= \prod_{j=1}^{n}\left(1 - 4\frac{e^{jt}}{(1+e^{jt})^2}sin^2(j\xi/2)\right)^{1/2} \leq \exp(-c_1\sum_{j=1}^{[k^{1/2}]} sin^2(j\xi/2)), \end{aligned}$$

where $c_1 > 0$. For $\epsilon k^{-1/2} \leq |\xi| \leq \epsilon$, then $|\xi| = \epsilon k_1^{-1/2}$ where $1 \leq k_1 \leq k$. Separating the integers $1,\ldots,[k^{1/2}]$ to $1,\ldots,[k_1^{1/2}]$; $[k_1^{1/2}]+1,\ldots,2[k_1^{1/2}]$ and so on, we can get $[[k^{1/2}]/[k_1^{1/2}]]$ groups. In at least half of the groups at least one third of the j's satisfy

$$|j\xi/2 - 2K\pi| > \epsilon/4 \text{ for all } K.$$

So,

$$\exp(-c_1\sum_{j=1}^{[k^{1/2}]} sin^2(j\xi/2)) \leq \exp(-\frac{c_1 k^{1/2}}{6}sin^2(\epsilon/4)).$$

For $\epsilon \leq |\xi| \leq \pi$, we can find $k_2 \geq [k^{1/2}/2]$ values of j such that for each of these

$$|j\xi/2 - 2K\pi| > \delta > 0 \text{ for all } K$$

since in this case two consecutive j's cannot both be within δ of multiples of 2π. So,

$$\sum_{j=1}^{[k^{1/2}]} c_1 sin^2(j\xi/2) \geq c_0 k^{1/2} \text{ for all } \epsilon k^{-1/2} \leq |\xi| \leq \pi$$

for some $c_0 > 0$. Therefore,

$$\left|\prod_{j=1}^{n}\frac{1+e^{jt+ij\xi}}{1+e^{jt}}\right| \leq e^{-c_0 k^{1/2}}.$$

So, from (4.18), we know

$$J_2 = O(e^{-c_0 k^{1/2}}) = O(\frac{1}{k^{3/2}}).$$

Therefore,

$$P(W=k) = e^{-n\Lambda(k/n)} \frac{1}{\sqrt{2\pi K''(t_n(k))}} (1 + O(k^{-1/2}))$$

for $k \to \infty$ and $k \leq n(n+1)/4$.

(b) for $k=0$, we know $t = -\infty$. So,

$$\lim_{x\to 0}(2^n \exp(-n\Lambda(x))) = \lim_{t\to-\infty} \prod_{j=1}^{n} (1+e^{jt}) = 1$$

and ,from (4.9),

$$f_{W_t}(0) = \lim_{t\to-\infty} \prod_{j=1}^{n} \frac{1}{1+e^{jt}} = 1.$$

From (4.11), we have

$$f_W(0) = \exp(-n \lim_{x\to 0} \Lambda(x)) f_{W_t}(0) = 2^{-n}.$$

(c) When k is fixed and $k \leq n$, we have

$$e^{-n\Lambda(k/n)} = 2^{-n} e^{-kt_n(k)} \prod_{j=1}^{n} (1+e^{jt_n(k)}). \tag{4.19}$$

From Lemma 4.2, we have $t_n(k) \to t(k) < 0$. We also can prove $t_n(k) = t(k)(1+O(ne^{nt(k)}))$ for some $c > 0$. Let $B_n(k)$ and $B(k)$ be the number of sets of distinct positive integers from $\{1,\ldots,n\}$ and $\{1,2,\ldots\}$ respectively whose sums are k. From $W_t = \sum_{j=1}^n jI_{jt}$, where $P(I_{jt}=1) = e^{jt}/(1+e^{jt})$, when $k \leq n$, we have

$$f_{W_{t_n(k)}}(k) = \frac{B_n(k)e^{kt_n(k)}}{\prod_{j=1}^n (1+e^{jt_n(k)})} = \frac{B(k)e^{kt(k)}}{\prod_{j=1}^\infty (1+e^{jt(k)})}(1+O(ne^{nt(k)})) = f_{W_{t(k)}}(k)(1+O(ne^{nt(k)})), \tag{4.20}$$

where $t(k)$ and $f_{W_{t(k)}}(k)$ depend only on k, and (4.16) follows. $\square$

Remark: Table 1 gives values of $B(k)$, $t(k)$, exact values of $P(W=k)$ and approximations of $P(W=k)$ based on (4.16) and (4.15) for $k=10$. From these we see that the error $O(ne^{nt(k)})$ is not small unless n is large and k is small. So, for $k=8,9,10$, the new approximations begin to fail, however , it is greatly superior to the normal approximations for $k \leq 5$.

Table 1

Approximations for one sample Wilcoxon for $n = 10$

k	1	2	3	4	5	6	7	8	9	10
$B_{10}(k) = B(k)$	1	1	2	2	3	4	5	6	8	10
t(k)	-.889	-.635	-.520	-.451	-.404	-.369	-.342	-.320	-.302	-.286
exact	.0010	.0010	.0020	.0020	.0029	.0039	.0049	.0059	.0078	.0098
$P_{new}(W = k)$	.0010	.0010	.0019	.0019	.0028	.0037	.0044	.0051	.0065	.0077
$P_{nor}(W = k)$	.0012	.0015	.0020	.0026	.0034	.0043	.0054	.0067	.0082	.0098

4.1 Two-sample Wilcoxon statistic

Put $S_1 = \sum_{i=1}^{N} jI_j$, $S_2 = \sum_{j=1}^{N} I_j$ where

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

Put $U_{m,n}$ is the two-sample Wilcoxon statistic. Then

$$P(U_{m,n} = k) = P(S_1 = k|S_2 = m) = \frac{P(S_1 = k, S_2 = m)}{P(S_2 = m)}. \tag{4.21}$$

The moment generating function of S_1 and S_2 is

$$M(s,t) = Ee^{sS_1 + tS_2} = \sum_{j=1}^{N} Ee^{(js+t)I_j} = \prod_{j=1}^{N}(q + pe^{js+t}).$$

Put $K(s,t) = \frac{1}{N}\sum_{j=1}^{N} \log(q + pe^{js+t})$. Then the density functions of the tilted variables I_{jst}, $j = 1, \ldots, N$ are

$$P(I_{jst} = 1) = \frac{pe^{js+t}}{q + pe^{js+t}}, \quad P(I_{jst} = 0) = \frac{q}{q + pe^{js+t}}, \quad j = 1, \ldots, N. \tag{4.22}$$

Put $S_{1st} = \sum_{j=1}^{N} jI_{jst}$ and $S_{2st} = \sum_{j=1}^{N} I_{jst}$ Then,

$$P(S_1 = k, S_2 = m) = e^{NK(s,t) - sk - tm} P(S_{1st} = k, S_{2st} = m).$$

From

$$\begin{cases} NK_{10}(s,t) = \sum_{j=1}^{N} \frac{jpe^{js+t}}{q+pe^{js+t}} = k \\ NK_{01}(s,t) = \sum_{j=1}^{N} \frac{pe^{js+t}}{q+pe^{js+t}} = m, \end{cases} \tag{4.23}$$

where $K_{ij}(s,t) = \partial^{i+j}K(s,t)/(\partial s^i \partial t^j)$, we can get saddlepoint $s = s(k,m)$, $t = t(k,m)$. Put $\Lambda(k/N, m/N) = \frac{1}{N}(s(k,m)k + t(k,m)m - K(s(k,m), t(k,m)))$. Then

$$P(S_1 = k, S_2 = m) = e^{-N\Lambda(k/N, m/N)} P(S_{1st} = k, S_{2st} = m). \tag{4.24}$$

Partially differentiating (4.23) with respect to k, we can get

$$\begin{cases} K_{20}(s,t)s_{10}(k,m) + K_{11}(s,t)t_{10}(k,m) = \frac{1}{N} \\ K_{11}(s,t)s_{10}(k,m) + K_{02}(s,t)t_{10}(k,m) = 0. \end{cases}$$

So,

$$\begin{cases} s_{10}(k,m) = \frac{1}{N}\frac{K_{20}(s,t)}{K_{02}(s,t)K_{20}(s,t)-K_{11}^2(s,t)} > 0 \\ t_{10}(k,m) = \frac{1}{N}\frac{-K_{11}(s,t)}{K_{02}(s,t)K_{20}(s,t)-K_{11}^2(s,t)} < 0. \end{cases} \tag{4.25}$$

Therefore we can get the following Lemma.

Lemma 4.4 *Put $k = m(m+1)/2 + l$ where $0 \le l \le \frac{m(2N-m+1)}{2} - \frac{m(m+1)}{2} = mn$. The solution s and t of equation (4.23) satisfies*

$$\begin{cases} s \to -\infty, ms + t \to \infty, (m+1)s + t \to -\infty, & as\ l \to 0, \\ s < 0, t > 0 & as\ 0 < l < \frac{1}{2}mn, \\ s = t = 0 & as\ l = \frac{1}{2}mn,\ when\ mn\ is\ even, \\ s > 0, t < 0 & as\ \frac{1}{2}mn < l < mn, \\ s \to \infty, (N-m)s + t \to -\infty, (N-m+1)s + t \to \infty, & as\ l \to mn \end{cases} \tag{4.26}$$

If N, p are fixed, $s(k,m)$ is an increasing function with respect to k and $t(k,m)$ is an decreasing function with respect to k.

We consider $0 < l < \frac{1}{2}mn$ first. In this situation $s < 0$ and $t > 0$ from Lemma 4.4. So, from (4.23),

$$N\int_0^1 \frac{e^{(y+1/N)Ns+t}}{q + pe^{(y+1/N)Ns+t}}dy < \sum_{j=1}^{N} \frac{e^{\frac{j}{N}s+t}}{q + pe^{\frac{j}{N}s+t}}\frac{1}{N}N = N < N\int_0^1 \frac{e^{yNs+t}}{q+pe^{yNs+t}}dy.$$

That is

$$\frac{1}{ps}ln\frac{q+pe^{Ns+s+t}}{q+pe^{s+t}} < N < \frac{1}{ps}ln\frac{q+pe^{Ns+t}}{q+pe^{t}}.$$

By doing some derivation we can get

$$\frac{q}{p}\frac{1-e^{ms}}{1-e^{ns}}e^{-ms} < e^t < \frac{q}{p}\frac{1-e^{ms}}{1-e^{(n-1)s}}e^{-(m+1)s}. \tag{4.27}$$

From (4.23), we have

$$\begin{cases} f_{1N}(s,t) = \sum_{j=1}^N \frac{jpe^{js+t}}{q+pe^{js+t}} - \frac{m(m+1)}{2} = -\sum_{j=1}^m \frac{jq}{q+pe^{js+t}} + \sum_{j=m+1}^N \frac{jpe^{js+t}}{q+pe^{js+t}} = l \\ f_{2N}(s,t) = \sum_{j=1}^N \frac{pe^{js+t}}{q+pe^{js+t}} - m = -\sum_{j=1}^m \frac{q}{q+pe^{js+t}} + \sum_{j=m+1}^N \frac{pe^{js+t}}{q+pe^{js+t}} = 0 \end{cases} \tag{4.28}$$

If $Ns \to 0$ we can get $-ms \le t \le -(m+1)s$ from (4.27). So,

$$\sum_{j=1}^N \frac{jpe^{js+t}}{q+pe^{js+t}} - \frac{m(m+1)}{2} = \frac{mn}{2}(1+\epsilon(N))$$

where $\epsilon(N) \to 0$ as $N \to \infty$. Therefore we have

Lemma 4.5 *If $0 < l < mn/c$ where $c > 2$, then $-Ns > c_1 > 0$.*

Without loss of generality, we assume $m \geq n$.

Lemma 4.6 *If $0 < l < mn/2$ and $p \to p_0$, then*

$$t = -ms - \delta s + \log(q/p) + \log c_3$$

where $0 \leq \delta \leq 1$ and $1 \leq c_3 \leq m/(n-1)$.

Proof: Put $a = m/(n-1) > 1$ and $x = e^{(n-1)s}$. We know $s < 0$ for $0 < l < mn/2$. So, $0 < x < 1$. Put

$$f(x) = \frac{1 - x^a}{1 - x} = \frac{1 - e^{ms}}{1 - e^{(n-1)s}}.$$

Then,

$$f'(x) = \frac{(a-1)x^a - ax^{a-1} + 1}{(1-x)^2} = \frac{1}{(1-x)^2} g(x), \tag{4.29}$$

where $g(x) = (a-1)x^a - ax^{a-1} + 1$. We have

$$g'(x) = a(a-1)x^{a-2}(1-x) > 0$$

and $\lim_{x\to 0} g(x) = 1$. So, $f'(x) > 0$. From $\lim_{x\to 0} f(x) = 1$ and $\lim_{x\to 1} f(x) = a$, we have $1 \leq f(x) \leq a$. In the same way, we can prove $1 \leq (1 - e^{ms})/(1 - e^{ns}) \leq m/n$. From (4.27), the proof of the Lemma is now complete. $\square$

Lemma 4.7 *If $0 < l < mn/c$ where $c > 2$ and $p \to p_0$, then, there exist constants $c_5 > 0$ and $c_6 < 3 + \sqrt{13}$ such that*

$$\frac{c_5}{\sqrt{l}} < -s < \frac{c_6}{\sqrt{l}}, \text{ as } N \to \infty$$

for large enough N. Also, for $mn/c \leq l < mn/2$,

$$-s = O(\frac{1}{\sqrt{l}}), \text{ as } N \to \infty.$$

Proof: From (4.28), we have

$$\begin{cases} f_{1N}(s,t) = -\sum_{j=1}^{m} \frac{j}{1+c_3 e^{js-ms-\delta s}} + \sum_{j=m+1}^{N} \frac{jc_3 e^{js-ms-\delta s}}{1+c_3 e^{js-ms-\delta s}} = l, \\ f_{2N}(s,t) = -\sum_{j=1}^{m} \frac{1}{1+c_3 e^{js-ms-\delta s}} + \sum_{j=m+1}^{N} \frac{c_3 e^{js-ms-\delta s}}{1+c_3 e^{js-ms-\delta s}} = 0. \end{cases} \tag{4.30}$$

That is

$$\begin{cases} f_{1N}(s,t) = -\sum_{j=1}^{m} \frac{je^{(m-j)s+\delta s}}{c_3+e^{(m-j)s+\delta s}} + \sum_{j=1}^{n} \frac{(m+j)c_3 e^{js-\delta s}}{1+c_3 e^{js-\delta s}} = l, \\ f_{2N}(s,t) = -\sum_{j=1}^{m} \frac{e^{(m-j)s+\delta s}}{c_3+e^{(m-j)s+\delta s}} + \sum_{j=1}^{n} \frac{c_3 e^{js-\delta s}}{1+c_3 e^{js-\delta s}} = 0. \end{cases} \tag{4.31}$$

Therefore

$$f_N(s,t) = f_{1N}(s,t) - m f_{2N}(s,t) = \sum_{j=1}^{m-1} \frac{j e^{js+\delta s}}{c_3 + e^{js+\delta s}} + \sum_{j=1}^{n-1} \frac{(j+1)c_3 e^{js+(1-\delta)s}}{1 + c_3 e^{js+(1-\delta)s}} + \frac{c_3 e^{(1-\delta)s}}{1 + c_3 e^{(1-\delta)s}} = l. \tag{4.32}$$

We know $s < 0$ and $f_1(x) = x/(1+x)$ is an increasing function. So

$$\int_0^{1-1/m} y e^{ms(y+\frac{1+\delta}{m})} dy < \frac{1}{m^2} \sum_{j=1}^{m-1} j e^{js+\delta s} = \sum_{j=1}^{m-1} \frac{j}{m} e^{\frac{j+\delta}{m} ms} \frac{1}{m} < \int_{\delta/m}^{(m-1+\delta)/m} (y + \frac{1-\delta}{m}) e^{msy} dy.$$

That is

$$m^2 e^{(1+\delta)s} \int_0^{1-1/m} y e^{msy} dy < \sum_{j=1}^{m-1} j e^{js+\delta s} < m^2 \int_{\delta/m}^{(m-1+\delta)/m} y e^{msy} dy + m \int_{\delta/m}^{(m-1+\delta)/m} (1-\delta) e^{msy} dy$$

Now

$$\begin{cases} m^2 \int_0^{1-1/m} y e^{msy} dy = \{((m-1)s - 1)e^{(m-1)s} + 1\} \frac{1}{s^2}, \\ m^2 \int_{\delta/m}^{(m-1+\delta)/m} y e^{msy} dy = \{((m-1+\delta)s - 1)e^{(m-1+\delta)s} + (1-\delta s)e^{\delta s}\} \frac{1}{s^2}, \\ m \int_{\delta/m}^{(m-1+\delta)/m} (1-\delta) e^{msy} dy = (1-\delta)(e^{(m-1+\delta)s} - e^{\delta s}) \frac{1}{s} \end{cases} \tag{4.33}$$

and $1 < c_3 + e^{js+\delta s} < 1 + c_3$ as $j = 1, \ldots, m-1$. So,

$$e^{(1+\delta)s} \{((m-1)s - 1)e^{(m-1)s} + 1\} \frac{1}{(1+c_3)s^2} < \sum_{j=1}^{m-1} \frac{j e^{js+\delta s}}{c_3 + e^{js+\delta s}} < (mse^{(m-1)s} + (1-s)) \frac{e^{\delta s}}{s^2}.$$

We know $-1 \le mse^{(m-1)s} < 0$ and $-1 \le (m-1)s - 1)e^{(m-1)s} < 0$. Therefore,

$$e^{(1+\delta)s} \{((m-1)s - 1)e^{(m-1)s} + 1\} \frac{1}{(1+c_3)s^2} < \sum_{j=1}^{m-1} \frac{j e^{js+\delta s}}{c_3 + e^{js+\delta s}} < (1-s) \frac{e^{\delta s}}{s^2}. \tag{4.34}$$

In the same way we can get

$$\int_0^{1-1/n} (y + 1/n) e^{ns(y+\frac{2-\delta}{n})} dy < \frac{1}{n^2} \sum_{j=1}^{n-1} (j+1) e^{js+(1-\delta)s} < \int_{(1-\delta)/n}^{(n-\delta)/n} (y + \frac{1+\delta}{n}) e^{nsy} dy.$$

That is

$$\begin{aligned} n^2 e^{(2-\delta)s} \int_0^{1-1/n} y e^{nsy} dy \quad &+ \quad n e^{(2-\delta)s} \int_0^{1-1/n} e^{msy} dy < \sum_{j=1}^{n-1} j e^{js+\delta s} \\ &< \quad n^2 \int_{(1-\delta)/n}^{(n-\delta)/n} y e^{nsy} dy + n \int_{(1-\delta)/n}^{(n-\delta)/n} (1+\delta) e^{nsy} dy. \end{aligned}$$

Now

$$\begin{cases} n^2\int_0^{1-1/n} ye^{nsy}dy = \{((n-1)s-1)e^{(n-1)s}+1\}\frac{1}{s^2}, \\ n\int_0^{1-1/n} e^{nsy}dy = (e^{(n-1)s}-1)\frac{1}{s}, \\ n^2\int_{(1-\delta)/n}^{(n-\delta)/n} ye^{nsy}dy = \{((n-\delta)s-1)e^{(n-\delta)s}+(1-(1-\delta)s)e^{(1-\delta)s}\}\frac{1}{s^2}, \\ n\int_{(1-\delta)/n}^{(n-\delta)/n}(1+\delta)e^{nsy}dy = (1+\delta)(e^{(n-\delta)s}-e^{(1-\delta)s})\frac{1}{s} \end{cases} \tag{4.35}$$

and $1 < 1+c_3e^{js-\delta s} < 1+c_3$ as $j = 1,\ldots,n$. So,

$$\begin{aligned} e^{(2-\delta)s}((ns-1)e^{(n-1)s}-s+1)\frac{c_3}{(1+c_3)s^2} &< \sum_{j=1}^{n-1}\frac{(j+1)c_3e^{js+(1-\delta)s}}{1+c_3e^{js+(1-\delta)s}} \\ &< c_3((n+1)se^{(n-1)s}+(1-2s))\frac{e^{(1-\delta)s}}{s^2}. \end{aligned}$$

We know $-1 \le (n+1)se^{(n-1)s} < 0$ and $-1 \le (ns-1)e^{(n-1)s} < 0$. Therefore,

$$e^{(2-\delta)s}((ns-1)e^{(n-1)s}-s+1)\frac{c_3}{(1+c_3)s^2} < \sum_{j=1}^{n-1}\frac{(j+1)c_3e^{js+(1-\delta)s}}{1+c_3e^{js+(1-\delta)s}} < c_3(1-2s)\frac{e^{(1-\delta)s}}{s^2}. \tag{4.36}$$

Put $\delta_1 = c_3e^{(1-\delta)s}/(1+c_3e^{(1-\delta)s})$. We know $0 < \delta_1 \le c_3/(1+c_3)$. So, substituting (4.34), (4.36) and δ_1 in (4.32), we get

$$e^{2s}\{(ns-1)e^{(n-1)s}+((m-1)s-1)e^{(m-1)s}-s+2\}\frac{1}{(1+c_3)s^2} < l-\delta_1 < c_3(2-3s)\frac{1}{s^2}.$$

That is

$$e^{2s}\{(ns-1)e^{(n-1)s}+((m-1)s-1)e^{(m-1)s}-s+2\}\frac{1}{(1+c_3)(l-\delta_1)} < s^2 < \frac{c_3(2-3s)}{l-\delta_1}.$$

From Lemma 4.5, we can prove

$$(ns-1)e^{(n-1)s}+((m-1)s-1)e^{(m-1)s}-s+2 > c_4 > 0.$$

So,

$$\frac{c_4e^{2s}}{(1+c_3)(l-\delta_1)} < s^2 < \frac{c_3(2-3s)}{l-\delta_1}. \tag{4.37}$$

When $l = 1$, from the right inequality of (4.37), we have $0 < -s < c_7$. Therefore, from the left inequality of (4.37), we have $-s > c_8 > 0$. That is

$$0 < c_8 < -s < c_7.$$

So, from (4.27) and last inequality, we can get $c_3 = 1+\epsilon$ where $\epsilon \to 0$, as $N \to \infty$. Therefore, we can have $0 < \delta_1 < \frac{1}{2}$ as N is big enough. So, when $l = 1$, from the right inequality of (4.37), we have

$$s^2+6s-4 < 0.$$

From Lemma 4.4, we know s is an increasing function with respect to l. Therefore,

$$0 < -s < 3 + \sqrt{13}, \text{ for any } l < mn/2 \ ,$$

when N is big enough. So, from (4.37), we have

$$\frac{c_5}{\sqrt{l}} < -s < \frac{c_6}{\sqrt{l}}, \text{ as } N \to \infty.$$

Because s is an increasing function with respect to k, we have

$$-s = O(\frac{1}{\sqrt{l}}), \text{ as } N \to \infty. \qquad \Box$$

If l is fixed we have the following Lemma.

Lemma 4.8 *If l is fixed and $p \to p_0$ as $N \to \infty$, we have*

$$s \to s(l) \qquad as\ N \to \infty,$$

where $s(l) < 0$ is only dependent on l.

Proof: From Lemma 4.7 we know $-s > c > 0$. Also,

$$\frac{1 - e^{ms}}{1 - e^{ns}} \longrightarrow 1 \qquad \text{as } N \to \infty.$$

So, from Lemma 4.6, we have

$$t = -ms - \delta_0 s + \log(q/p),$$

where $0 \leq \delta_0 \leq 1$ as $N \to \infty$. Therefore, from the second formula of (4.31), we have

$$\sum_{j=1}^{n}(\frac{e^{js-\delta_0 s}}{1 + e^{js-\delta_0 s}} - \frac{e^{js-(1-\delta_0)s}}{1 + e^{js-(1-\delta_0)s}}) = \sum_{j=n+1}^{m} \frac{e^{js-(1-\delta_0)s}}{1 + e^{js-(1-\delta_0)s}} > 0.$$

Therefore,

$$\frac{e^{js-\delta_0 s}}{1 + e^{js-\delta_0 s}} - \frac{e^{js-(1-\delta_0)s}}{1 + e^{js-(1-\delta_0)s}} > 0 \qquad \text{for all } j.$$

So $\delta_0 > 1/2$. Because $-s > c > 0$ we have

$$\sum_{j=1}^{n}(\frac{e^{js-\delta_0 s}}{1 + e^{js-\delta_0 s}} - \frac{e^{js-(1-\delta_0)s}}{1 + e^{(js-(1-\delta_0)s}}) = \sum_{j=n+1}^{m} \frac{e^{js-(1-\delta_0)s}}{1 + e^{js-(1-\delta_0)s}} = e^{ns} \sum_{j=1}^{m-n} \frac{e^{js-(1-\delta_0)s}}{1 + e^{js-(1-\delta_0)s}} = d_0 e^{ns}, \tag{4.38}$$

where $d_0 > 0$. Because

$$\frac{e^{js-\delta_0 s}}{1 + e^{js-\delta_0 s}} - \frac{e^{js-(1-\delta_0)s}}{1 + e^{js-(1-\delta_0)s}} = \frac{e^{(j-\delta_0)s} - e^{js-(1-\delta_0)s}}{(1 + e^{(1-\delta_0)s})(1 + e^{\delta_0 s})}$$

and

$$\begin{aligned} e^{(j-\delta_0)s} - e^{js-(1-\delta_0)s} &= e^{js}\sum_{u=1}^{\infty}((-\delta_0)^u - (-(1-\delta_0))^u)\frac{s^u}{u!} \\ &= e^{js}(1-2\delta_0)\sum_{u=2}^{\infty}((-\delta_0)^{u-1} + \ldots + (-(1-\delta_0))^{u-1})\frac{s^u}{u!} \\ &= -c_1(1-2\delta_0)e^{js}, \end{aligned}$$

where $c_1 > 0$. Therefore, from (4.38), we have

$$0 < -(1-2\delta_0) < d_1 e^{ns},$$

where $d_1 > 0$. So

$$\delta_0 = 1/2 + d_2 e^{ns}, \tag{4.39}$$

where $d_2 > 0$. From (4.32), we have

$$\sum_{j=1}^{m-1}\frac{je^{js+\delta_0 s}}{1+e^{js+\delta_0 s}} + \sum_{j=1}^{n}\frac{je^{js-\delta_0 s}}{1+e^{js-\delta_0 s}} = l.$$

Now,

$$\begin{aligned} \sum_{j=1}^{m-1}\frac{je^{js+\delta_0 s}}{1+e^{js+\delta_0 s}} - \sum_{j=1}^{m-1}\frac{je^{js+\frac{1}{2}s}}{1+e^{js+\frac{1}{2}s}} &= \sum_{j=1}^{m-1}\frac{je^{js+\frac{1}{2}s}(e^{d_2e^{ns}}-1)}{(1+e^{js+\delta_0 s})(1+e^{js+\frac{1}{2}s})} \\ &= d_2 e^{ns+\frac{1}{2}s}\sum_{j=1}^{m-1}\frac{je^{js}}{(1+e^{js+\delta_0 s})(1+e^{js+\frac{1}{2}s})}(1+O(e^{ns})) \end{aligned} \tag{4.40}$$

In the same way we have

$$\begin{aligned} &\sum_{j=1}^{n}\frac{je^{js-\delta_0 s}}{1+e^{js-\delta_0 s}} - \sum_{j=1}^{n}\frac{je^{js-\frac{1}{2}s}}{1+e^{js-\frac{1}{2}s}} = \sum_{j=1}^{n}\frac{je^{js-\frac{1}{2}s}(e^{-d_2e^{ns}}-1)}{(1+e^{js-\delta_0 s})(1+e^{js-\frac{1}{2}s})} \\ &= -d_2 e^{ns+\frac{1}{2}s}\sum_{j=1}^{m-1}\left(\frac{je^{js}}{(1+e^{js+\delta_0 s})(1+e^{js+\frac{1}{2}s})}\frac{(1+e^{js+\delta_0 s})(1+e^{js+\frac{1}{2}s})}{(1+e^{js-\delta_0 s})(e^s+e^{js+\frac{1}{2}s})}\right)(1+O(e^{ns})). \end{aligned} \tag{4.41}$$

We know

$$\begin{aligned} 1 - \frac{(1+e^{js+\delta_0 s})(1+e^{js+\frac{1}{2}s})}{(1+e^{js-\delta_0 s})(e^s+e^{js+\frac{1}{2}s})} &= \frac{(e^s-1) + e^{2js+\frac{1}{2}s+\delta_0 s}(e^{-2\delta_0 s}-1) + e^{js}(e^{(1-\delta_0)s}-e^{\delta_0 s})}{(1+e^{js-\delta_0 s})(e^s+e^{js+\frac{1}{2}s})} \\ &= \frac{(e^s-1) + e^{2js+s}(e^{-s}-1) + c_{11}e^{js}e^{ns}}{(1+e^{js-\delta_0 s})(e^s+e^{js+\frac{1}{2}s})} \\ &= \frac{(1-e^s)(e^{2js}-1)}{(1+e^{js-\delta_0 s})(e^s+e^{js+\frac{1}{2}s})}(1+e^{ns}) < 0, \end{aligned}$$

where $c_{11} > 0$. Therefore, from (4.40) + (4.41), we get

$$\sum_{j=1}^{m-1} \frac{je^{js+\frac{1}{2}s}}{1+e^{js+\frac{1}{2}s}} + \sum_{j=1}^{n} \frac{je^{js-\frac{1}{2}s}}{1+e^{js-\frac{1}{2}s}} = l + c_{12}d_2e^{ns}(1+O(e^{ns})), \tag{4.42}$$

where $c_{12} > 0$. When $m_1 > m$ and $n_1 > n$, we have

$$\sum_{j=1}^{m_1-1} \frac{je^{js_1+\frac{1}{2}s_1}}{1+e^{js_1+\frac{1}{2}s_1}} + \sum_{j=1}^{n_1} \frac{je^{js_1-\frac{1}{2}s_1}}{1+e^{js_1-\frac{1}{2}s_1}} = l + c_{13}d_2e^{n_1s_1}(1+O(e^{n_1s})), \tag{4.43}$$

where $c_{13} > 0$. Subtracting (4.42) from (4.43), we get

$$\begin{aligned} & \sum_{j=1}^{m-1} \left(\frac{je^{js_1+\frac{1}{2}s_1}}{1+e^{js_1+\frac{1}{2}s_1}} - \frac{je^{js+\frac{1}{2}s}}{1+e^{js+\frac{1}{2}s}}\right) + \sum_{j=1}^{n} \left(\frac{je^{js_1-\frac{1}{2}s_1}}{1+e^{js_1-\frac{1}{2}s_1}} - \frac{je^{js-\frac{1}{2}s}}{1+e^{js-\frac{1}{2}s}}\right) \\ = & -\sum_{j=m}^{m_1-1} \frac{je^{js_1+\frac{1}{2}s_1}}{1+e^{js_1+\frac{1}{2}s_1}} - \sum_{j=n+1}^{n_1} \frac{je^{js_1-\frac{1}{2}s_1}}{1+e^{js_1-\frac{1}{2}s_1}} + c_{13}d_2e^{n_1s_1}(1+O(e^{n_1s_1})) - c_{12}d_2e^{ns}(1+O(e^{ns})) \\ = & -d_{13}me^{ms_1} - d_{14}ne^{ns_1} + c_{13}d_2e^{n_1s_1}(1+O(e^{n_1s_1})) - c_{12}d_2e^{ns}(1+O(e^{ns})) < 0, \end{aligned}$$

where $d_{13} > 0$ and $d_{14} > 0$. Therefore,

$$0 > c > s > s_1 \qquad \text{as } N \text{ is big enough.}$$

Because $p \to p_0$ and s is absolutely bounded when l is fixed, the proof of the Lemma is now completed. □

Consider the following 3 cases.
(a) $l \to \infty$, and $mn - l \to \infty$
(b) $l = 0$ or $l = mn$,
(c) l is fixed and finite integer or $mn - l$ is fixed and finite integer.
In each of these cases we have a different limiting distribution for $P_{U_{m,n}=k}$. Put $K_{ij} = K_{ij}(s,t)$. We have the following theorem.

Theorem 4.2 *Assume* $S_1 = \sum_{i=1}^{N} jI_j$, $S_2 = \sum_{j=1}^{N} I_j$ *where*

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

Put $k = m(m+1)/2 + l$. *Then, for the two-sample Wilcoxon rank sum statistic* $U_{m,n}$, *as* $N \to \infty$, *we have*
(a) for $l \to \infty$ *and* $l \le nm/2$, *we have*

$$P(U_{m,n} = k) = P(S_1 = k|S_2 = m) = e^{-N\Lambda(k/N,m/N)} \frac{\sqrt{pq}}{\sqrt{2\pi N(K_{20}K_{02} - K_{11}^2)}}(1 + O(l^{-1/2})); \tag{4.44}$$

(b) for $l=0$,

$$P(U_{m,n}) = e^{-N\Lambda(m(m+1)/(2N),m/N)}\frac{1}{P(S_2=m)} = p^m q^n\sqrt{2\pi Npq}(1+O(\frac{1}{N})); \qquad (4.45)$$

(c) for l a fixed and finite integer,

$$P(U_{m,n} = \frac{m(m+1)}{2}+l) = e^{-N\Lambda(k/N,m/N)}f_{s(l)}(l)\sqrt{2\pi Npq}(1+O(N^{-1})), \qquad (4.46)$$

where

$$f_{s(l)}(l) = \frac{B_1(l)e^{ls(l)}}{\prod_{j=1}^{\infty}(1+e^{(j-\frac{1}{2})s(l)})\prod_{j=1}^{\infty}(1+e^{(j-\frac{1}{2})s(l)})},$$

where

$$B_1(l) = \sum_{j=1}^{\infty}\sharp\{i_1+i_1'+\ldots+i_j+i_j'-j=l,\ \ i_{j_1}\neq i_{j_2}\ \textit{and}\ i'_{j_1}\neq i'_{j_2}\ \textit{when}\ j_1\neq j_2\ \textit{and}$$

i_{j_1}, i'_{j_2} *are positive integers* $\}$.

Proof: (a) We have

$$\begin{aligned} P(S_1=k|S_2=m) &= \frac{P(S_1=k,S_2=m)}{P(S_2=m)} = e^{NK(s,t)-sk-tm}\frac{P(S_{1st}=k,S_{2st}=m)}{P(S_2=m)} \\ &= e^{NK(s,t)-sk-tm}\frac{P(S_{1st}=k|S_{2st}=m)P(S_{2st}=m)}{P(S_2=m)}, \end{aligned} \qquad (4.47)$$

where $k=\frac{m(m+1)}{2}+l$ and

$$K(s,t) = \frac{1}{N}\log Ee^{sS_1+tS_2} = \frac{1}{N}\sum_{j=1}^{N}\log(q+pe^{js+t}). \qquad (4.48)$$

When $l\to\infty$ and $l\le nm/2$, from Lemma 4.6 and 4.7, we know $t=-ms-\delta s+\log(q/p)+\log c_3$ and $-s=O(l^{-1/2})$. Therefore,

$$\begin{cases} K_{10}\sim N^{-1}l \\ K_{20} = \frac{1}{N}\sum_{j=1}^{N}\frac{pqj^2c_3e^{js-ms-\delta s}}{(q+qc_3e^{js-ms-\delta s})^2} = N^2\sum_{j=0}^{m-1}\frac{c_3(\frac{j}{N})^2e^{\frac{j}{N}Ns+\delta s}}{(c_3+e^{\frac{j}{N}Ns+\delta s})^2}\frac{1}{N} + N^2\sum_{j=1}^{m}\frac{c_3(\frac{j}{N})^2e^{\frac{j}{N}Ns-\delta s}}{(1+c_3e^{\frac{j}{N}Ns-\delta s})^2}\frac{1}{N} \\ \quad\sim N^2\int_0^p y^2e^{yNs}dy + N^2\int_0^q y^2e^{yNs}dy \sim N^{-1}l^{3/2}, \\ K_{30}\sim N^{-1}l^2, \\ K_{40}\sim N^{-1}l^{5/2}, \\ K_{50}\sim N^{-1}l^3 \end{cases} \qquad (4.49)$$

Note $K_{rv} \sim K_{r0}$, $r, v = 1, \dots, 5$. Therefore,

$$\sigma_{st}^2 = K_{20} - \frac{K_{11}^2}{K_{02}} \sim K_{20} \sim N^{-1} l^{3/2}.$$

So, $N\sigma_{st}^2 \sim l^{3/2} \to \infty$. We know

$$\begin{aligned} P(S_{1st} = k, S_{2st} = m) &= \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \exp(-i\xi k - i\eta m + NK(s + i\xi, t + i\eta) - NK(s,t)) d\xi d\eta \\ &= J_1 + J_2, \end{aligned}$$

where $|\theta_1| < 1$, $|\theta_2| < 1$ and $P_3(i\xi, i\eta)$, $P_4(i\xi, i\eta)$, $P_5(i\xi, i\eta)$, $P_6(i\xi, i\eta)$ are polynomials of degree 3, 4, 5, 6.

$$J_1 = \int_{-\epsilon_2}^{\epsilon_2} \int_{-\epsilon_1 l^{-1/2}}^{\epsilon_1 l^{-1/2}} \exp(-i\xi k - i\eta m + NK(s + i\xi, t + i\eta) - NK(s,t)) d\xi d\eta$$

and

$$J_2 = \int_{(-\pi,\pi)^2 \setminus (-\epsilon_1 l^{-1/2}, \epsilon_1 l^{-1/2}) \times (-\epsilon_2, \epsilon_2)} \exp(-i\xi k - i\eta m + NK(s + i\xi, t + i\eta) - NK(s,t)) d\xi d\eta.$$

Now we look at J_1 first.

$$\begin{aligned} J_1 &= \int_{-\epsilon_2}^{\epsilon_2} \int_{-\epsilon_1 l^{-1/2}}^{\epsilon_1 l^{-1/2}} \exp(-i\xi k - i\eta m + NK(s + i\xi, t + i\eta) - NK(s,t)) d\xi d\eta \\ &= \int_{-\epsilon_2}^{\epsilon_2} \int_{-\epsilon_1 l^{-1/2}}^{\epsilon_1 l^{-1/2}} \exp\{-\frac{1}{2} N\sigma_{st}^2 \xi^2 - \frac{1}{2} NK_{02}(\eta + \frac{K_{11}}{K_{02}}\xi)^2 + \\ &\quad \frac{N}{3!} P_3(i\xi, i\eta) + \frac{N}{4!} P_4(i\xi\theta_1, i\eta\theta_2)\} d\xi d\eta, \end{aligned}$$

where $|\theta_1| < 1$, $|\theta_2| < 1$ and $P_3(i\xi, i\eta)$ and $P_4(i\xi, i\eta)$ are polynomials of degree 3 and 4. Put $\xi_1 = \sqrt{N}\sigma_{st}\xi$ and $\eta_1 = \sqrt{NK_{02}}(\eta + \frac{K_{11}}{K_{02}}\xi)$. We know $K_{rv} \sim K_{r0} \sim N^{-1} l^{(r+1)/2}$, $r, v = 1, 2, 3, 4$, $l = O(N^2)$ and

$$\frac{NK_{ij}}{(N\sigma_{st}^2)^{i/2}} \sim l^{-(i-2)/4}, \; i = 1, 2, 3, 4. \tag{4.50}$$

Therefore,

$$\begin{aligned} J_1 &= \int_{-\epsilon_3 N^{1/2}}^{\epsilon_3 N^{1/2}} \int_{-\epsilon_4 l^{1/4}}^{\epsilon_5 l^{1/4}} e^{-\frac{1}{2}\xi_1 - \frac{1}{2}\eta_1} \\ &\quad (1 + P_3'(\xi_1, \eta_1) l^{-1/4} + (P_4'(\xi_1, \eta_1) + P_6'(\xi_1, \eta_1)) l^{-1/2} e^{\frac{1}{4}\xi_1 + \frac{1}{4}\eta_1}) d\xi_1 d\eta_1 + O(e^{-cN^{1/2}}), \end{aligned}$$

where $P_3'(i\xi, i\eta)$, $P_4'(i\xi, i\eta)$ and $P_6'(i\xi, i\eta)$ are polynomials of degree 3, 4 and 6, $\epsilon_3 > 0$, $\epsilon_4 > 0$ and $\epsilon_5 > 0$ can be small enough by chosing ϵ_1 and ϵ_2. Therefore,

$$J_1 = \frac{1}{2\pi N\sqrt{K_{02}}\sigma_{st}}(1 + O(\frac{1}{l^{1/2}})) = \frac{1}{2\pi N\sqrt{K_{20}K_{02} - K_{11}^2}}(1 + O(\frac{1}{l^{1/2}})).$$

Now let us look at J_2. We use a proof related to one used in Albers, Bickel and van Zwet (1976). J_2 is rewritten as

$$J_2 = \int_{(-\pi,\pi)^2\setminus(-\epsilon_1 l^{-1/2},\epsilon_1 l^{-1/2})\times -\epsilon_2,\epsilon_2} \exp(-i\xi k - i\eta m)\sum_{j=1}^{N}\frac{q + pe^{js+t+i(j\xi+\eta)}}{q + pe^{js+t}}d\xi d\eta.$$

When $\epsilon_1 l^{-1/2} \le |\xi| \le \pi$ and $\epsilon_2 \le |\eta| \le \pi$, note $-s = O(l^{-1/2})$ and $t = -ms - \delta s$, where $0 \le \delta \le 1$, we have

$$\begin{aligned}
\left|\prod_{j=1}^{N}\frac{q + pe^{js+t+i(j\xi+\eta)}}{q + pe^{js+t}}\right| &= \prod_{j=1}^{N}\frac{((1 + e^{js-ms-\delta s})^2 - 2e^{js-ms-\delta s}(1 - cos(j\xi + \eta)))^{1/2}}{(1 + e^{js-ms-\delta s})} \\
&= \prod_{j=1}^{N}\left(1 - 4\frac{e^{js-ms-\delta s}}{(1 + e^{js-ms-\delta s})^2}sin^2((j\xi + \eta)/2)\right)^{1/2} \\
&\le \exp(-c_1\sum_{j=m}^{m+[cl^{1/2}]} sin^2((j\xi + \eta)/2)),
\end{aligned}$$

where $c, c_1 > 0$ and $cl^{1/2} \le n$. For $\epsilon_1 l^{-1/2} \le |\xi| \le \epsilon_1$, we can have $|\xi| = \epsilon_1 l_1^{-1/2}$ where $1 \le l_1 \le l$. Separating the integers $m, \ldots, m + [cl^{1/2}]$ to $m, \ldots, m + [cl_1^{1/2}]$; $m + [cl_1^{1/2}] + 1, \ldots, m + 2[cl_1^{1/2}]$ and so on, we can get $[[cl^{1/2}]/[cl_1^{1/2}]]$ groups. In at least half of the groups at least one third of the j's satisfy

$$|j\xi + \eta - 2k\pi| > \delta > 0 \text{ for all } k$$

Therefore,

$$\exp(-c_1\sum_{j=m}^{m+[cl^{1/2}]} sin^2((j\xi + \eta)/2)) \le \exp(-\frac{c_1[cl^{1/2}]}{6}sin^2(\epsilon/4)).$$

For $\epsilon_1 \le |\xi| \le \pi$ and $\epsilon_2 \le |\eta| \le \pi$, we can find $l_3 \ge [l^{1/2}/2]$ values of j such that for each of these

$$|j\xi + \eta - 2k\pi| > \delta \text{ for all } k$$

since in this case two consecutive j's cannot both be within δ of multiples of 2π. So,

$$\sum_{j=m}^{m+[cl^{1/2}]} c_1 sin^2((j\xi + \eta)/2) \ge l^{1/2}\delta_1 \text{ for all } \epsilon_1 l^{-1/2} \le |\xi| \le \pi \text{ and } \epsilon_2 \le |\eta| \le \pi,$$

where $\delta_1 > 0$. Therefore,

$$\left|\prod_{j=1}^{N} \frac{1+e^{js+t+i(j\xi+\eta)}}{1+e^{js+t}}\right| \le e^{-\delta_1 l^{1/2}}.$$

From (4.18), we know

$$J_2 = O(\frac{1}{l^{3/2}}).$$

Therefore,

$$P(S_{1st}=k, S_{2st}=m) = \frac{1}{2\pi N\sqrt{K_{20}K_{02}-K_{11}^2}}(1+O(\frac{1}{l^{1/2}})).$$

We know

$$P(S=m) = \frac{1}{\sqrt{2\pi Npq}}(1+O(\frac{1}{N})). \tag{4.51}$$

So,

$$P(U_{m,n}=k) = P(S_1=k|S_2=m) = e^{-N\Lambda(s,t)}\frac{\sqrt{pq}}{\sqrt{2\pi N(K_{20}K_{02}-K_{11}^2)}}(1+O(l^{-1/2})).$$

(b) We have $s \to -\infty$, $ms+t \to \infty$ and $(m+1)s+t \to -\infty$. So, from (4.48), we have

$$\begin{aligned}
\exp(-N\lim_{x\to 0}\Lambda(k/N, m/N)) &= \lim_{s\to-\infty}\prod_{j=1}^{N}(q+pe^{js+t})e^{-mt-\frac{m(m+1)}{2}s} \\
&= \lim_{s\to-\infty}\prod_{j=1}^{m}(qe^{-js-t}+p)\prod_{j=m+1}^{N}(q+pe^{js+t}) = p^m q^n.
\end{aligned}$$

$$\begin{aligned}
P_{st}(S_1=m(m+1)/2, S_2=m) &= \lim_{s\to-\infty}\frac{p^m q^n e^{mt+\frac{m(m+1)}{2}s}}{\prod_{j=1}^{N}(q+pe^{js+t})} \\
&= \lim_{s\to-\infty}\frac{p^m q^n e^{mt+\frac{m(m+1)}{2}s}}{p^m q^n e^{mt+\frac{m(m+1)}{2}s}} = 1.
\end{aligned}$$

Therefore, (4.45) follows from (4.51), (4.21) and (4.24).

(c) When l is fixed, note $t = -(m+\delta)s + \log(q/p)$, where $\delta = 1/2 + O(e^{ns})$, we have

$$\begin{aligned}
e^{-N\Lambda(k/N,m/N)} &= \prod_{j=1}^{N}(q+pe^{js+t})e^{-mt-(m(m+1)/2+l)s} = \prod_{j=1}^{m}(qe^{-t-js}+p)\prod_{j=m+1}^{N}(q+pe^{js+t})e^{-ls} \\
&= p^m q^n \prod_{j=0}^{m-1}(1+e^{(j+\delta)s})\prod_{j=1}^{n}(1+e^{(j-\delta)s})e^{-ls} \\
&= p^m q^n \prod_{j=1}^{m}(1+e^{(j-\frac{1}{2})s})\prod_{j=1}^{n}(1+e^{(j-\frac{1}{2})s})e^{-ls}(1+O(e^{ns})).
\end{aligned} \tag{4.52}$$

From Lemma 4.1, $s \to s(l) < 0$ as $N \to \infty$. We can prove $s = s(l)(1 + O(ne^{ns(l)}))$ for some $c > 0$. Let $B_{N1}(l) = \sum_{r=1}^{l} \sharp\{i_1 + i'_1 + \ldots + i_r + i'_r - r = l,\ \ i_{j_1} \neq i_{j_2}$ and $i'_{j_1} \neq i'_{j_2}$ when $j_1 \neq j_2$ and $\ i_{j_1} \leq m, i'_{j_2} \leq n$ are positive integers $\}$ and $B_1(l) = \sum_{r=1}^{\infty} \sharp\{i_1 + i'_1 + \ldots + i_r + i'_r - r = l,\ \ i_{j_1} \neq i_{j_2}$ and $i'_{j_1} \neq i'_{j_2}$ when $j_1 \neq j_2$ and $\ i_{j_1}, i'_{j_2}$ are positive integers $\}$. From $P(I_{jst} = 1) = pe^{js+t}/(q + pe^{js+t})$, we have

$$P_{st}(S_1 = m(m+1)/2 + l, S_2 = m) = \frac{p^m q^n B_{N1}(l) e^{mt+(m(m+1)/2+l)s}}{\prod_{j=1}^{N}(q + pe^{js+t})}$$
$$= \frac{B_{N1}(l)e^{ls}}{\prod_{j=0}^{m-1}(1 + e^{(j+\delta)s})\prod_{j=1}^{n}(1 + e^{(j-\delta)s})}$$
$$= \frac{B_1(l)e^{ls(l)}}{\prod_{j=1}^{\infty}(1 + e^{(j-\frac{1}{2})s(l)})\prod_{j=1}^{\infty}(1 + e^{(j-\frac{1}{2})s(l)})}(1 + O(ne^{ns(l)})) = f_{s(l)}(l)(1 + O(ne^{ns(l)})), \tag{4.53}$$

where $s(l)$ and $f_{s(l)}(l)$ are only dependent on l. So, (4.46) follows from (4.51), (4.21) and (4.24). $\square$

Remark: Table 2 gives values of $B_1(l)$, $s(l)$, exact values of $P(U_{m,n} = m(m + 1)/2 + l)$ and approximations of $P(U_{m,n} = m(m+1)/2 + l)$ based on (4.46) and (4.44) for $m = n = 8$. From these we see that the new approximation even for $l = 8$ is good, however, the normal approximations for $k \leq 4$ are very poor.

Table 2

Approximations for two sample Wilcoxon for $n = m = 8$

k	1	2	3	4	5	6	7	8	9	10
$B_{10}(k) = B(k)$	1	1	2	2	3	4	5	6	8	10
t(k)	-.889	-.635	-.520	-.451	-.404	-.369	-.342	-.320	-.302	-.286
exact	.0010	.0010	.0020	.0020	.0029	.0039	.0049	.0059	.0078	.0098
$P_{new}(W = k)$	.0010	.0010	.0019	.0019	.0028	.0037	.0044	.0051	.0065	.0077
$P_{nor}(W = k)$	.0012	.0015	.0020	.0026	.0034	.0043	.0054	.0067	.0082	.0098

References

Albers, W., Bickel, P. J. and van Zwet, W. R. (1976), Asymptotic expansions for the Power of Distribution Free Tests in the One-sample Problem, *Ann. Statist.* **4** 108–156.

Daniels, H. E. (1954), Saddlepoint Approximations in statistics, *Ann. Math. Statist.* **25** 631–649.

Jensen, J. L. (1995), Saddlepoint Approximations, Oxford Uni. Press Inc.

Robinson, J. (1982), Saddlepoint Approximations for Permutation Tests , *J. Roy. Statist. Soc. Ser. B* **44** 91–101.