

## Statistics 583, Problem Set 3 Solutions

Wellner; 4/22/2009

1. (a) What is the locally best rank test of  $F = G$  against  $G = (e^{\theta F} - 1)/(e^\theta - 1)$ ,  $\theta > 0$ ?
- (b) What is the locally best rank test of  $F = G$  against  $G = F/(e^\theta(1 - F) + F)$ ?
- (c) What can you say about the power of these tests (other than the fact that they are locally most powerful)?

**Solution:** (a) and (b) By Hoeffding's formula

$$P_\theta(\underline{Q} = \underline{q}) = \frac{1}{\binom{N}{n}} E_{uniform} \left\{ \prod_{j=1}^n \psi'_\theta(U_{(q_j)}) \right\}$$

where

$$\psi_\theta(u) = G_\theta \circ F^{-1}(u) = \frac{e^{\theta u} - 1}{e^\theta - 1}$$

for the first alternatives, and

$$\psi_\theta(u) = G_\theta \circ F^{-1}(u) = \frac{u}{e^\theta(1 - u) + u}$$

in the case of the second type of alternative. In either case the locally most powerful rank test rejects for those values  $\underline{q}$  of  $\underline{Q}$  which make

$$\begin{aligned} \frac{\partial}{\partial \theta} P_\theta(\underline{Q} = \underline{q}) \Big|_{\theta=0} &= \frac{1}{\binom{N}{n}} E_{uniform} \left\{ \sum_{j=1}^n \frac{\partial}{\partial \theta} \psi'_\theta(U_{(q_j)}) \Big|_{\theta=0} \right\} \\ &= \sum_{j=1}^n E_{uniform} \left\{ \frac{\partial}{\partial \theta} \psi'_\theta(U_{(q_j)}) \Big|_{\theta=0} \right\} \end{aligned}$$

as large as possible. Hence it remains only to calculate

$$\phi(u) \equiv \frac{\partial}{\partial \theta} \psi'_\theta(u) \Big|_{\theta=0}$$

and  $E_{uniform} \phi(U_{(i)})$  for the two alternatives in question.

(i) In the first case,

$$\psi'_\theta(u) = \frac{\theta e^{\theta u}}{e^\theta - 1},$$

and straightforward calculation yields

$$\frac{\partial}{\partial \theta} \psi'_\theta(u) = e^{\theta u} \frac{(e^\theta - 1)(1 + \theta u - \theta) - \theta}{(e^\theta - 1)^2}.$$

By applying L'Hopital's rule twice, we find that

$$\frac{\partial}{\partial \theta} \psi'_\theta(u) \Big|_{\theta=0} = u - \frac{1}{2}.$$

Since  $E(U_{(i)}) = i/(N + 1)$ , the locally most powerful rank test of  $H$  versus this alternative  $K$  is the Wilcoxon test "reject  $H$  if  $S_N = \sum_{j=1}^n Q_j > k_\alpha$ ".

(ii) In the second case,

$$\psi'_\theta(u) = \frac{e^\theta}{(e^\theta(1 - u) + u)^2}.$$

Hence

$$\frac{\partial}{\partial \theta} \psi'_\theta(u) \Big|_{\theta=0} = 2u - 1,$$

and again the locally most powerful rank test is the Wilcoxon rank sum test.

As for interpretations of these alternatives, first note that the functions  $\psi_\theta(u)$  are distribution functions on  $[0, 1]$  with densities  $\psi'_\theta(u)$ . (i) This alternative is the simplest exponential family density related to the uniform(0, 1) distribution: the density is of the form  $p_\theta(u) = \psi'_\theta(u) = c(\theta) \exp(\theta u) 1_{[0,1]}(u)$ .

(ii) For this family, note that

$$1 - \psi_\theta(u) = \frac{e^\theta(1 - u)}{e^\theta(1 - u) + u},$$

and hence the *odds ratio* is

$$\frac{1 - \psi_\theta(u)}{\psi_\theta(u)} = e^\theta \frac{1 - u}{u} = e^\theta \cdot \text{the odds ratio for Uniform}(0,1).$$

Thus this family is one with proportional odds ratios.

(c) From Example 4.4 on pages 37-38 and the equivalence of the test based on  $S_N = \sum_{j=1}^n Q_j$  to the test based on

$$U_{m,n} \equiv \int \mathbb{F}_m d\mathbb{G}_n = m^{-1} n^{-1} \sum_{i=1}^m \sum_{j=1}^n 1\{X_i \leq Y_j\},$$

we see that under  $F = G$

$$\frac{U_{m,n} - 1/2}{\sqrt{(N+1)/12mn}} \rightarrow_d Z \sim N(0, 1),$$

while under local alternatives of the form  $G_N \equiv \psi_{c/\sqrt{N}}(F)$

$$\frac{U_{m,n} - 1/2}{\sqrt{(N+1)/12mn}} \rightarrow_d Z + \Delta \sim N(\Delta, 1)$$

where, assuming that  $\lambda_N \equiv m/N \rightarrow \lambda$ , and using  $\int F dG_N = P(X \leq Y_N) = 1 - P(Y_N < X) = 1 - \int G_N dF$

$$\begin{aligned} \Delta &\equiv \lim_{N \rightarrow \infty} \sqrt{\frac{12mn}{N(N+1)}} \sqrt{N} \left( \int F dG_N - 1/2 \right) \\ &= \sqrt{12\lambda(1-\lambda)} \lim_{N \rightarrow \infty} \sqrt{N} \left( \int F dG_N - 1/2 \right) \\ &= -\sqrt{12\lambda(1-\lambda)} \lim_{N \rightarrow \infty} \sqrt{N} \left( \int G_N dF - 1/2 \right) \\ &= -\sqrt{12\lambda(1-\lambda)} c \lim_{N \rightarrow \infty} \int_0^1 \frac{\psi_{c/\sqrt{N}}(u) - u}{c/\sqrt{N}} du \\ &= -\sqrt{12\lambda(1-\lambda)} c \int_0^1 \frac{d}{d\theta} \psi_\theta(u) |_{\theta=0} du \end{aligned}$$

if we can justify the interchange of limit and integral. For the two functions  $\psi_\theta$  in (a) and (b) we calculate

$$\frac{d}{d\theta} \psi_\theta(u) |_{\theta=0} = -\frac{1}{2} u(1-u)$$

in case (a), and

$$\frac{d}{d\theta} \psi_\theta(u) |_{\theta=0} = -u(1-u)$$

with  $\int_0^1 u(1-u) du = 1/6$ . Thus

$$\Delta = \begin{cases} \sqrt{12\lambda(1-\lambda)} c / 12, & \text{in case (a)} \\ \sqrt{12\lambda(1-\lambda)} c / 6, & \text{in case (b)}. \end{cases} \quad (1)$$

Thus the asymptotic power of the Mann-Whitney-Wilcoxon test under these two local alternatives is given by

$$\lim_{N \rightarrow \infty} P_{F, G_N}((U_{m,n} - 1/2) / \sqrt{(N+1)/12mn} > z_\alpha) = P(Z + \Delta > z_\alpha) = P(Z > z_\alpha - \Delta)$$

where  $\Delta$  is given in (1) respectively.

2. Suppose that an urn contains  $N$  balls with the numbers  $z_i = -\log(1 - i/(N+1))$ ,  $i = 1, \dots, N$  and we sample  $n < N$  balls from this urn. Let  $\bar{Y}_n = n^{-1} \sum_1^n Y_i$  denote the sample mean of the sampled balls.

(a) Calculate the mean  $\mu_N = E(\bar{Y}_n)$  and variance  $\sigma_N^2 = \text{Var}(\bar{Y}_n)$  of  $\bar{Y}_n$ .

Find the limits of  $\bar{z}_N$  and  $\sigma_z^2$  as  $N \rightarrow \infty$ .

(b) Use the Wald-Wolfowitz-Noether-Hajek finite-sampling CLT to prove that  $(\bar{Y}_n - \mu_N)/\sigma_N \rightarrow_d N(0, 1)$ .

(c) What classical two-sample rank statistic is  $\bar{Y}_n$  equivalent to under the null hypothesis (of all  $X_1, \dots, X_m, Y_1, \dots, Y_n$  equal in distribution with a common continuous distribution function  $F$ )?

**Solution:** (a) The mean is

$$\begin{aligned} \mu_N &= E(\bar{Y}_n) = \bar{z}_N \\ &= \frac{1}{N} \sum_{i=1}^N \left\{ -\log \left( 1 - \frac{i}{N+1} \right) \right\} = \int_0^1 F^{-1}(t) d\mu_N(t) \\ &\rightarrow \int_0^1 \{-\log(1-t)\} dt = 1 \end{aligned} \quad (2)$$

upon noticing that  $F^{-1}(t) = -\log(1-t)$  for the standard exponential distribution  $F(x) = 1 - e^{-x}$ ,  $x \geq 0$ , so that  $F^{-1}(U) =_d Y \sim \text{Exponential}(1)$ . The convergence in (2) is justified by noting that the measure  $\mu_N$  corresponds to the uniform measure on the set  $\{1/(N+1), \dots, N/(N+1)\}$ , and hence these measures converge weakly to the uniform measure on  $(0, 1)$ , namely Lebesgue measure with df  $t$  on  $[0, 1]$ . Furthermore, for each  $\delta > 0$  there is a constant  $M = M_\delta$  such that  $-\log(1-t) \leq M(1-t)^{-\delta}$  for  $0 \leq t < 1$ , and hence the function  $-\log(1-t)$  is uniformly integrable with respect to the sequence  $\mu_N$ : choosing  $\delta = 1/2$ ,

$$\begin{aligned} &\int_{-\log(1-t) > \lambda} (-\log(1-t)) d\mu_N(t) \\ &\leq \int_{M(1-t)^{-1/2} > \lambda} M(1-t)^{-1/2} d\mu_N(t) \\ &= \frac{1}{N} \sum_{i=1}^N M \left( 1 - \frac{i}{N+1} \right)^{-1/2} \mathbf{1}\{i/(N+1) > (\lambda/M)^{-2}\} \\ &\leq \frac{N+1}{N} M \int_{t > 1 - (\lambda/M)^{-2}} (1-t)^{-1/2} dt = \frac{N+1}{N} M \int_0^{(\lambda/M)^{-2}} s^{-1/2} ds \\ &\leq 4\lambda^{-1}, \end{aligned}$$

so

$$\lim_{N \rightarrow \infty} \int_{-\log(1-t) > \lambda} (-\log(1-t)) d\mu_N(t) \leq 4\lambda^{-1} \rightarrow 0$$

as  $\lambda \rightarrow \infty$ . Similarly, the variance is

$$\sigma_N^2 = \text{Var}(\bar{Y}_n) = \frac{\sigma_z^2}{n} \left(1 - \frac{n-1}{N-1}\right),$$

where

$$\begin{aligned} \sigma_z^2 &= \frac{1}{N} \sum_{i=1}^N (z_i - \bar{z}_N)^2 \\ &\rightarrow \int_0^1 \{-\log(1-t) - 1\}^2 dt \\ &= \text{Var}(Y) = 1. \end{aligned}$$

The convergence can again be justified by a uniform integrability argument. (b) The Wald-Wolfowitz-Noether-Hájek finite-sampling CLT yields  $(\bar{Y}_n - \mu_N)/\sigma_N \rightarrow_d N(0, 1)$  as long as  $0 < \liminf(n/N) \leq \limsup(n/N) < 1$  if we show that the Noether condition holds. But the Noether condition is

$$\eta_N \equiv \frac{\max_{1 \leq i \leq N} |z_i - \bar{z}_N|}{\sum_{i=1}^N (z_i - \bar{z}_N)^2} \rightarrow 0.$$

Upon dividing the numerator and denominator by  $N$ , we know from part A that the denominator (divided by  $N$ ) converges to 1. Hence it suffices to show that

$$N^{-1} \max_{1 \leq i \leq N} |z_i - \bar{z}_N|^2 \rightarrow 0.$$

Now since  $z_i$  increases with  $i$ ,

$$\begin{aligned} \max_{1 \leq i \leq N} |z_i - \bar{z}_N| &\leq \max_{1 \leq i \leq N} (\bar{z}_N - z_i) \vee \max_{1 \leq i \leq N} (z_i - \bar{z}_N) \\ &\leq \bar{z} \vee (z_N - \bar{z}_N) \end{aligned}$$

where  $z_N = -\log(1 - N/(N+1)) = -\log(1/(N+1)) = \log(N+1)$ . Thus we have

$$\begin{aligned} N^{-1} \max_{1 \leq i \leq N} |z_i - \bar{z}_N|^2 &\leq N^{-1} \bar{z}_N^2 \vee N^{-1} (\log(N+1) - \bar{z}_N)^2 \\ &\rightarrow 0 \vee 0 = 0. \end{aligned}$$

(c) Under the null hypothesis  $\bar{Y}_n$  is equivalent to the “log-rank” statistic

$$T_N \equiv \frac{1}{n} \sum_{i=1}^n \left\{ -\log \left( 1 - \frac{R_i}{N+1} \right) \right\}$$

where  $R_i$  is the rank of  $Y_i$ ,  $i = 1, \dots, n$  in the combined sample,  $X_1, \dots, X_m, Y_1, \dots, Y_n$ .

3. Suppose that  $X_1, \dots, X_n$  are independent Exponential(1) random variables. Let  $Y_i \equiv X_{(i)}$ , for  $i = 1, \dots, n$ , denote the *order statistics* corresponding to  $X_1, \dots, X_n$ .  
 (a) Show that the vector  $(Y_1, \dots, Y_n)$  has the same joint distribution as  $(W_1, \dots, W_n)$  where  $W_i \equiv \sum_{j=1}^i Z_j / (n - j + 1)$  and  $Z_1, \dots, Z_n$  are i.i.d. Exponential(1).  
 (b) Use the result of (a) to compute  $E(Y_i)$ ,  $Var(Y_i)$ , and  $Cov(Y_i, Y_j)$  for any fixed  $i, j$ .

**Solution:** (a) Note that  $0 \leq W_1 \leq \dots \leq W_n$  and

$$Z_i = (n - i + 1)(W_i - W_{i-1}), \quad i = 1, \dots, n \quad (3)$$

(with  $W_0 \equiv 0$ ). Let  $g(\underline{Z}) \equiv \underline{W}$  be the map defined in the problem statement so that  $g^{-1}(\underline{W}) = \underline{Z}$  is given in (3). Then the Jacobian of  $g^{-1}$  has entries  $n, (n-1), \dots, 1$  on the diagonal, entries  $-(n-1), \dots, -2, -1$  below the diagonal, and zero elsewhere. Hence  $\det(J_{g^{-1}}) = \text{tr}(J_{g^{-1}}) = n!$  and the density of  $\underline{W}$  is given by

$$\begin{aligned} f_{\underline{W}}(\underline{w}) &= f_{\underline{Z}}(g^{-1}(\underline{w})) \det(J_{g^{-1}}) \\ &= n! \prod_{i=1}^n \exp(-(n-i+1)(w_i - w_{i-1})) \\ &= n! \exp\left(-\sum_{i=1}^n (n-i+1)(w_i - w_{i-1})\right) \\ &= n! \exp\left(-\sum_{i=1}^n \left(\sum_{j=i}^n 1\right)(w_i - w_{i-1})\right) \\ &= n! \exp\left(-\sum_{j=1}^n \sum_{i \leq j} (w_i - w_{i-1})\right) \\ &= n! \exp\left(-\sum_{j=1}^n w_j\right) = n! f(w_1) \cdots f(w_n) \end{aligned}$$

on the set  $0 \leq w_1 \leq \dots \leq w_n < \infty$  where  $f(x) = \exp(-x)1_{[0, \infty)}(x)$  is the standard exponential density. Hence  $\underline{Y} \stackrel{d}{=} \underline{W} \equiv \underline{X}_{(\cdot)}$  where  $X_1, \dots, X_n$  are i.i.d. exponential(1).

(b) It follows immediately from (a) that

$$\begin{aligned} E(Y_i) &= E\left(\sum_{j=1}^i \frac{Z_j}{n-j+1}\right) = \sum_{j=1}^i \frac{1}{n-j+1}, \\ Var(Y_i) &= Var\left(\sum_{j=1}^i \frac{Z_j}{n-j+1}\right) = \sum_{j=1}^i \frac{1}{(n-j+1)^2}, \end{aligned}$$

and

$$\begin{aligned} \text{Cov}(Y_i, Y_j) &= \text{Cov}\left(\sum_{k=1}^i \frac{Z_k}{n-k+1}, \sum_{k'=1}^i \frac{Z_{k'}}{n-k'+1}\right) \\ &= \sum_{k=1}^{i \wedge j} \frac{1}{(n-k+1)^2}. \end{aligned}$$

for any fixed  $i, j$ .

4. Suppose that, in Example 6.3.17, page 28,  $1 - F_i = (1 - F)^{\Delta_i}$  where  $\Delta_i = \exp(\theta z_i)$  and  $z_1, \dots, z_N$  are given real numbers and  $\theta \in \mathbb{R}$ . Then the distribution of the ranks of  $X_1, \dots, X_N$  (independent with respective d.f.'s  $F_1, \dots, F_N$ ) is

$$P_\theta(\underline{R} = \underline{r}) = \prod_{i=1}^N \frac{e^{\theta z_{d_i}}}{\sum_{j=i}^N e^{\theta z_{d_j}}}.$$

- (a) Find the locally most powerful rank test of  $H : \theta = 0$  versus  $K : \theta > 0$ . (Call the statistic  $S_N$  and express it explicitly in terms of some scores  $a_N(j)$ ,  $j = 1, \dots, N$ , the ranks  $\underline{R}$ , and the  $z_j$ 's.)  
 (b) Compute  $E(S_N)$  and  $\text{Var}(S_N)$  under the null hypothesis  $\theta = 0$ ? How would you carry out the test you found in (a)?  
 (c) Show that when  $z_1 = \dots = z_m = 0$  and  $z_{m+1} = \dots = z_N = 1$ , the test reduces to the test "reject when  $S_N = \sum_{j=1}^n a_N(Q_j) > c_{N,\alpha}$ " found in Example 3.20 with

$$a_N(i) = \sum_{j=1}^i \frac{1}{N-j+1}.$$

- (d) Let  $S_{N,1}(x) \equiv N^{-1} \sum_{i=1}^N z_i 1_{[X_i \geq x]}$  and  $S_{N,0}(x) \equiv N^{-1} \sum_{i=1}^N 1_{[X_i \geq x]}$ . Show that the statistic  $S_N$  can be rewritten as

$$\begin{aligned} S_N &= N \left( \bar{z} - \int \frac{S_{N,1}(x)}{S_{N,0}(x)} d\mathbb{F}_N(x) \right) \\ &= N \int \left( z - \frac{S_{N,1}(x)}{S_{N,0}(x)} \right) d\mathbb{P}_N(x, z) \end{aligned}$$

where

$$\mathbb{F}_N(x) \equiv N^{-1} \sum_{i=1}^N 1_{(-\infty, x]}(X_i), \quad \mathbb{P}_N \equiv N^{-1} \sum_{i=1}^N \delta_{(X_i, z_i)}.$$

**Solution:** (a) The locally most powerful rank test reject for large values of

$$\frac{\partial}{\partial \theta} P_{\theta}(\underline{R} = \underline{r})|_{\theta=0},$$

or, equivalently, for large values of

$$\begin{aligned} & \frac{\partial}{\partial \theta} \log P_{\theta}(\underline{R} = \underline{r})|_{\theta=0} \\ &= \frac{\partial}{\partial \theta} \sum_{i=1}^N \left\{ \theta z_{d_i} - \log \left( \sum_{j=i}^N \exp(\theta z_{d_j}) \right) \right\} |_{\theta=0} \\ &= \sum_{i=1}^N \left\{ z_{d_i} - \frac{\sum_{j=i}^N z_{d_j}}{N-i+1} \right\} \\ &= N\bar{z} - \sum_{i=1}^N \frac{\sum_{j=i}^N z_{d_j}}{N-i+1} \\ &= N\bar{z} - \sum_{j=1}^N \sum_{i=1}^N 1_{[j \geq i]} \frac{z_{d_j}}{N-i+1} \\ &= N\bar{z} - \sum_{j=1}^N a_N(j) z_{d_j} \\ & \quad \text{where } a_N(j) \equiv \sum_{i=1}^j \frac{1}{N-i+1} \\ &= N\bar{z} - \sum_{j=1}^N a_N(r_j) z_j \end{aligned}$$

since  $d$  is the inverse permutation of  $r$ . Thus the locally most powerful rank test of  $H : \theta = 0$  versus  $K : \theta > 0$  is of the form “reject  $H$  if  $S_N \equiv N\bar{z} - \sum_{j=1}^N a_N(R_j) z_j > k_N(\alpha)$ ” where  $k_N(\alpha)$  satisfies  $P_{\theta=0}(S_N > k_N(\alpha)) \approx \alpha$ .

(b) Now

$$\begin{aligned} E(S_N) &= N\bar{z} - \sum_{j=1}^N E a_N(R_j) z_j = N\bar{z} - \sum_{j=1}^N \left( \sum_{i=1}^N N^{-1} a_N(i) \right) z_j \\ &= N\bar{z} (1 - \bar{a}_N) \\ &= 0 \end{aligned}$$

since

$$\begin{aligned}
\bar{a}_N &= N^{-1} \sum_{i=1}^N a_N(i) = N^{-1} \sum_{i=1}^N \left( \sum_{j=1}^i \frac{1}{N-j+1} \right) \\
&= N^{-1} \sum_{j=1}^N \frac{1}{N-j+1} \sum_{i=1}^N 1_{\{j \leq i\}} \\
&= N^{-1} \sum_{j=1}^N \frac{1}{N-j+1} (N-j+1) = N^{-1} \sum_{j=1}^N 1 = 1.
\end{aligned}$$

To calculate  $Var(S_N)$ , first note that  $\sum_{j=1}^N a_N(R_j) = \sum_{i=1}^N a_N(i)$ , so, by symmetry

$$\begin{aligned}
0 &= Var\left(\sum_{i=1}^N a_N(i)\right) = Var\left(\sum_{j=1}^N a_N(R_j)\right) \\
&= \sum_{j=1}^N Var(a_N(R_j)) + \sum_{j,j'=1, j \neq j'}^N Cov(a_N(R_j), a_N(R_{j'})) \\
&= NVar(a_N(R_j)) + N(N-1)Cov(a_N(R_j), a_N(R_{j'})) \\
&= N\sigma_a^2 + N(N-1)Cov(a_N(R_j), a_N(R_{j'}))
\end{aligned}$$

where  $\sigma_a^2 \equiv N^{-1} \sum_{j=1}^N (a_N(j) - \bar{a}_N)^2$ . Thus we find

$$Cov(a_N(R_j), a_N(R_{j'})) = -\frac{1}{N-1}\sigma_a^2.$$

Now we calculate

$$\begin{aligned}
Var(S_N) &= Var\left(\sum_{j=1}^N a_N(R_j)z_j\right) \\
&= \sum_{j=1}^N z_j^2 Var(a_N(R_j)) + \sum_{j=1}^N \sum_{j'=1, j' \neq j}^N z_j z_{j'} Cov(a_N(R_j), R_{j'}) \\
&= \sum_{j=1}^N z_j^2 \sigma_a^2 - \frac{1}{N-1} \sum_{j=1}^N \sum_{j'=1, j' \neq j}^N z_j z_{j'} \sigma_a^2 \\
&= \sigma_a^2 \left\{ \sum_{j=1}^N z_j^2 - \frac{1}{N-1} \sum_{j'=1, j' \neq j}^N z_j z_{j'} \right\} \\
&= \sigma_a^2 \left\{ \sum_{j=1}^N z_j^2 - \frac{1}{N-1} \left( \left( \sum_{j=1}^N z_j \right)^2 - \sum_{j=1}^N z_j^2 \right) \right\} \\
&= \frac{N^2}{N-1} \sigma_a^2 \cdot \sigma_z^2
\end{aligned} \tag{4}$$

where  $\sigma_z^2 \equiv \frac{1}{N} \sum_{i=1}^N (z_i - \bar{z})^2$ . Theorem 4.1 of Hájek (1961), page 513, says that if

$$\max_{1 \leq i \leq N} \frac{|a_N(i) - \bar{a}_N|}{N\sigma_a^2} \rightarrow 0, \quad \text{and} \quad \max_{1 \leq i \leq N} \frac{|z_i - \bar{z}_N|}{N\sigma_z^2} \rightarrow 0, \tag{5}$$

then  $(S_N - E(S_N))/\sqrt{Var(S_N)} \rightarrow_d N(0, 1)$  if and only if

$$\sum \sum_{\{(i,j): \sqrt{N}|a_i - \bar{a}| |z_j - \bar{z}| > \epsilon N^2 \sigma_a^2 \sigma_z^2\}} \frac{|a_i - \bar{a}|^2 |z_j - \bar{z}|^2}{N^2 \sigma_a^2 \sigma_z^2} \rightarrow 0 \tag{6}$$

for every  $\epsilon > 0$ . It is easily verified that  $\{a_N(i) : 1 \leq i \leq N\}$  satisfies the first part of (5), and we will assume that the second part of (5) holds for  $\{z_i : 1 \leq i \leq N\}$ . We will also assume that (6) holds. Then it follows from our computation of the mean and variance of  $S_N$  that we can carry out our test approximately for large  $N$  by rejecting  $H$  if  $S_N > z_\alpha \sqrt{Var(S_N)} = z_\alpha N \sigma_a \sigma_z / \sqrt{N-1}$  where  $P(Z > z_\alpha) = \alpha$  for  $Z \sim N(0, 1)$ .

(c) When  $z_1 = \dots = z_m = 0$  and  $z_{m+1} = \dots = z_N = 1$ , the test reduces to the test “reject when  $S_N = N\{n/N - \sum_{j=1}^n a_N(R_{m+j})\} > k_{N,\alpha}$ ”, or, equivalently, “reject when  $\tilde{S}_N \equiv \sum_{j=1}^n a_N(Q_j) < \tilde{k}_{N,\alpha}$ ”; this is closely related to the Savage test of Example 3.20, but with the direction of the test reversed because  $\theta > 0$  in our current setting corresponds to  $\Delta > 1$ , and this corresponds to the parameter  $\theta$  of Example 3.20 being less than 1.

(d) To see that  $S_N$  can be rewritten as claimed, note that

$$\begin{aligned}
\sum_{i=1}^N \frac{\sum_{j=i}^N z_{d_j}}{N-i+1} &= \sum_{i=1}^N \frac{\sum_{j=1}^N z_{d_j} 1\{X_{(j)} \geq X_{(i)}\}}{\sum_{j=1}^N 1\{X_{(j)} \geq X_{(i)}\}} \\
&= N \int_{-\infty}^{\infty} \frac{N^{-1} \sum_{j=1}^N z_j 1\{X_j \geq x\}}{N^{-1} \sum_{j=1}^N 1\{X_j \geq x\}} d\mathbb{F}_N(x) \\
&= N \int_{-\infty}^{\infty} \frac{S_{N,1}(x)}{S_{N,0}(x)} d\mathbb{F}_N(x)
\end{aligned}$$

where  $S_{N,1}$  and  $S_{N,0}$  are as defined in the problem statement.