Large Sample Theory
Ferguson

Exercises, Section 12, Some Rank Statistics.

1. Suppose $z_j = j$ and $a(j) = 1/\sqrt{j}$ for all $j = 1, 2, \ldots$. Find the asymptotic distribution of $S_N = \sum_1^N z_j a(R_j)$, where $(R_1, \ldots, R_N)$ is a random permutation of $(1, \ldots, N)$ with each permutation having probability $1/N!$.

2. The van der Waerden Test is a competitor of the Rank-Sum Test, in which the value an observation of rank r is replaced by $\psi(r/(N+1))$ where $\psi = \Phi^{-1}$ is the inverse of the standard normal distribution function. Thus, the van der Waerden Statistic is of the form $S_N = \sum_{j=1}^N z_j a(R_j)$ with $z_j = \psi(j/(N+1))$ and $a(j)$ as in Example 3.

(a) Note $\bar{z} = 0$. Show that $\sigma_z^2 \to 1$ as $N \to \infty$.

(b) Suppose that $n/N \to r$ as $N \to \infty$, with $0 < r < 1$. Is it true that $\sqrt{N} S_N \xrightarrow{\mathcal{L}} \mathcal{N}(0, r(1-r))$?

3. Show for general $S_N = \sum_1^N z_j a(R_j)$,

$$\mathrm{E}(S_N - \mathrm{E}S_N)^3 = \frac{N^3}{(N-1)(N-2)} \mu_3(z) \mu_3(a),$$

where

$$\mu_3(z) = (1/N) \sum_1^N (z_j - \bar{z}_N)^3 \qquad \text{and} \qquad \mu_3(a) = (1/N) \sum_1^N (a(j) - \bar{a}_N)^3.$$

(The third central moment of S_N may be useful in improving the normal approximation through the Edgeworth expansion.)

4.(a) Let $Z = \sum_1^N z_j a(R_j)$ and $T = \sum_1^N t_j b(R_j)$. Generalize Lemma 1 by showing that $\mathrm{Cov}(Z, T) = (N^2/(N-1)) \sigma_{zt} \sigma_{ab}$ where

$$\sigma_{zt} = \frac{1}{N} \sum_1^N (z_j - \bar{z}_N)(t_j - \bar{t}_N) \qquad \text{and} \qquad \sigma_{ab} = \frac{1}{N} \sum_1^N (a(j) - \bar{a}_N)(b(j) - \bar{b}_N).$$

(b) If $z_j = t_j = b(j) = j$ and $a(j) = \mathrm{I}(1 \le j \le m)$, then Z is the rank-sum test statistic and T is the statistic of Example 5, related to Spearman's rho. Assume $\sqrt{N}((m/N) - r) \to 0$ as $N \to \infty$, $r \in (0, 1)$, and show

$$\sqrt{N} \left(\begin{pmatrix} Z/N^2 \\ T/N^3 \end{pmatrix} - \begin{pmatrix} r/2 \\ 1/4 \end{pmatrix} \right) \xrightarrow{\mathcal{L}} \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{r(1-r)}{12} & -\frac{r(1-r)}{24} \\ -\frac{r(1-r)}{24} & \frac{1}{144} \end{pmatrix} \right).$$

5. In sampling from a population of N objects having values $z_1, z_2, \ldots, z_N$, first a sample of size $n < N/2$ is taken without replacement. Later a second sample of size n is

taken from the remaining $N-n$ objects without replacement. The difference of the means of the two samples is used to compare the samples. This leads to a rank statistic of the form $S_N = \sum_1^N z_j a(R_j)$, where $a(i) = 1$ for $i = 1, \ldots, n$, $a(i) = -1$ for $i = n+1, \ldots, 2n$, and $a(i) = 0$ for $i = 2n+1, \ldots, N$.
(a) What is the mean and the variance of S_N?
(b) Assume that $n \to \infty$ as $N \to \infty$. Under what condition on the z_i is it true that $(S_N - \mathrm{E}S_N)/\sqrt{\mathrm{Var}(S_N)} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$?

6. Let a sample of size n be taken from each of three distributions, and let T_N, respectively V_N, denote the sum of the ranks of the observations from the first, respectively second, distribution when all $N = 3n$ observations are ranked in order from 1 to N. Let $S_N = b_1 T_N + b_2 V_N$, for arbitrary real numbers b_1 and b_2. Let H_0 be the hypothesis that the three distributions are identical.

(a) Show that S_N is a linear rank statistic under H_0 of the form $S_N = \sum_{j=1}^N z_j a(R_j)$ where $z_j = j$; that is, find $a(i)$ for $i = 1, \ldots, N$.

(b) We have $\bar{z}_N = (N+1)/2$ and $\sigma_z^2 = (N^2-1)/12$. Find the asymptotic distribution of S_N.

(c) Find the asymptotic joint distribution of T_N and V_N. (Use the Cramér-Wold device of Exercise 3.2, p. 18.)