## EXACT AND ASYMPTOTICALLY ROBUST PERMUTATION TESTS

BY EUNYI CHUNG AND JOSEPH P. ROMANO

Stanford University

## APPENDIX A: PROOFS

A.1. Proofs in Section 5.. PROOF OF THEOREM 5.1. The sufficiency part due to Hoeffding [1] is proved in Theorem 15.2.3. of Lehmann and Romano [2]. To prove the necessity part, suppose s and t are continuity points of  $R^{T}(\cdot)$ . Then,

$$P\{T_n(G_nX^n) \le s, T_n(G'_nX^n) \le t\} = E[P\{T_n(G_nX^n) \le s, T_n(G'_nX^n) \le t | X^n\}]$$
$$= E[\hat{R}_n^T(s)\hat{R}_n^T(t)] \to R^T(s)R^T(t) ,$$

since convergence in probability of a bounded sequence of random variables entails convergence of moments. Convergence for a dense set of rectangles in the plane entails weak convergence.

Before proving Slutsky's Theorem for Randomization Distributions (Theorem 5.2), we need three lemmas.

LEMMA A.1. Suppose  $X^n$  has distribution  $P_n$  in  $\mathcal{X}_n$ , and  $\mathbf{G}_n$  is a finite group of transformations g of  $\mathcal{X}_n$  onto itself. Also, let  $G_n$  be a random variable that is uniform on  $\mathbf{G}_n$ . Assume  $X^n$  and  $G_n$  are mutually independent. Let  $\hat{R}_n^A$  denotes the randomization distributions of  $A_n$ , defined by

(S1) 
$$\hat{R}_n^A(t) = \frac{1}{|G_n|} \sum_{g \in G_n} I\{A_n(gX^n) \le t\}.$$

Suppose, under  $P_n$ ,

 $A_n(G_nX^n) \xrightarrow{P} a.$ 

Then, under  $P_n$ ,

(S2) 
$$\hat{R}_n^A(t) = \frac{1}{|G_n|} \sum_{g \in G_n} I\{A_n(gX^n) \le t\} \xrightarrow{P} \delta_a(t) \quad \text{if } t \ne a,$$

where  $\delta_c(\cdot)$  denotes the distribution function corresponding to the point mass function at c.

PROOF OF LEMMA A.1: Let  $G'_n$  have the same distribution as  $G_n$  and be independent from  $G_n$  and  $X_n$ . Since  $A_n(G_nX^n)$  converges in probability to a constant a,  $A_n(G'_nX^n) \xrightarrow{P} a$  and the independence of the limiting distributions is satisfied. Thus, the result follows from Theorem 5.1.

LEMMA A.2. Let  $B_n$  and  $T_n$  be sequences of random variables satisfying the conditions above, i.e.,

$$B_n(G_nX^n) \xrightarrow{P} b,$$

and

(S3) 
$$(T_n(G_nX^n), T_n(G'_nX^n)) \xrightarrow{d} (T, T'),$$

where T and T' are independent, each with common c.d.f.  $R^{T}(\cdot)$ . Let  $\hat{R}_{n}^{T+B}(t)$ denote the randomization distribution of  $T_{n} + B_{n}$ , defined in (S1) with A replaced by T + B. Then,  $\hat{R}_{n}^{T+B}(t)$  converges to T + b in probability. In other words,

$$\hat{R}_n^{T+B}(t) \equiv \frac{1}{|G_n|} \sum_{g \in G_n} I\{T_n(gX^n) + B_n(gX^n) \le t\} \xrightarrow{P} R^{T+b}(t) \quad \text{if } R^{T+b} \text{ is continuous at } t,$$

where  $R^{T+b}(\cdot)$  denotes the corresponding c.d.f. of T+b. (Of course,  $R^{T+b}(t) = R^{T}(t-b)$ .)

PROOF OF LEMMA A.2: Without loss of generality, assume b = 0. For any  $\varepsilon > 0$ ,

$$\frac{1}{|G_n|} \sum \mathrm{I}\{T_n(gX^n) + B_n(gX^n) \le t - \varepsilon\} - \frac{1}{|G_n|} \sum \mathrm{I}\{|B_n(gX^n)| > \varepsilon\}$$
$$\le \frac{1}{|G_n|} \sum \mathrm{I}\{T_n(gX^n) + B_n(gX^n) \le t\}$$

$$\leq \frac{1}{|G_n|} \sum \mathrm{I}\{T_n(gX^n) + B_n(gX^n) \leq t + \varepsilon\} + \frac{1}{|G_n|} \sum \mathrm{I}\{|B_n(gX^n)| > \varepsilon\}.$$

Note that  $\frac{1}{|G_n|} \sum I\{|B_n(gX^n)| > \varepsilon\}$  of the first line and the third line converges in probability to 0 by Lemma A.1. Also, by Theorem 5.1, (S3) implies

(S4) 
$$\hat{R}_n^T(t) = \frac{1}{|G_n|} \sum_{g \in G_n} I\{T_n(gX^n) \le t\} \xrightarrow{P} R^T(t)$$

if  $R^{T}(\cdot)$  is continuous at t. Thus, if both  $t - \varepsilon$  and  $t + \varepsilon$  are continuity points of  $R^{T}(\cdot)$ , the first term of the first line and the first term of the third line converge in probability to  $R^{T}(t - \varepsilon)$  and  $R^{T}(t + \varepsilon)$ , respectively. Therefore,

$$R^{T}(t-\varepsilon) \leq \hat{R}_{n}^{T+b}(t) \leq R^{T}(t+\varepsilon)$$

with probability tending to one, for continuity points  $t - \varepsilon$  and  $t + \varepsilon$  of  $R^T(\cdot)$ .

Now, let  $\varepsilon \downarrow 0$  through continuity points to deduce that

$$\hat{R}_n^{T+B}(t) \xrightarrow{P} R^T(t).$$

LEMMA A.3. Let  $A_n$  and  $T_n$  be sequences of random variables satisfying the conditions above, i.e.,

$$A_n(G_nX^n) \xrightarrow{P} a$$

where a is nonzero, and

$$(T_n(G_nX^n), T_n(G'_nX^n)) \xrightarrow{d} (T, T'),$$

where T and T' are independent, each with common c.d.f.  $R^{T}(\cdot)$ . Then, the randomization distribution of  $A_nT_n$  converges to aT in probability. In other words,

$$\hat{R}_n^{AT}(t) \equiv \frac{1}{|G_n|} \sum_{g \in G_n} \mathbf{I}\{A_n(gX^n)T_n(gX^n) \le t\} \xrightarrow{P} R^{aT}(t),$$

if  $R^{aT}$  is continuous at t, where  $R^{aT}(\cdot)$  denotes the corresponding c.d.f. of aT.

PROOF OF LEMMA A.3: Write

$$A_n T_n = a T_n + (A_n - a) T_n$$

Then, we can apply Lemma A.2 with  $B_n = (A_n - a)T_n$ , if we can verify the condition  $B_n(G_nX^n) \xrightarrow{P} 0$ . But,

$$B_n(G_nX^n) = [A_n(G_nX^n) - a]T_n(G_nX^n) \xrightarrow{P} 0 \cdot T = 0 ,$$

by the usual Slutsky's Theorem. Finally, the behavior of  $aT_n$  follows trivially from that of  $T_n$ .

PROOF OF THEOREM 5.2: The proof follows from Lemma A.2 and Lemma A.3.

REMARK A.1. Under the randomization hypothesis that the distribution of  $X^n$  is the same as that of  $gX^n$  for any  $g \in \mathbf{G}_n$ , the conditions (5.5) and (5.6) are equivalent to the assumptions that  $A_n(X^n) \xrightarrow{P} a$  and  $B_n(X^n) \xrightarrow{P} b$ , i.e. the convergence in probability just based on the original sample  $X^n$ without first transforming by a random  $G_n$ . For more on the randomization hypothesis, see Section 15.2 of Lehmann and Romano [2].

PROOF OF (5.8): Let  $N_j$  denote number of observations in  $\overline{Z}$  which are generated from  $P_j$ . Then,  $(N_1, \ldots, N_k)$  has the multinomial distribution based on N trials and success probabilities  $(p_1, \ldots, p_k)$ . In terms of the  $N_j$ , the number of differing observations in the above coupling construction is

$$D = \sum_{j=1}^k \max(n_j - N_j, 0) \; .$$

If we assume  $p_j > 0$  for all j, then by the usual central limit theorem,

$$N_i - Np_i = O_P(N^{1/2})$$

which together with (5.7) yields

$$N_j - n_j = (N_j - Np_j) + (Np_j - n_j) = O_P(N^{1/2})$$

It follows that  $D = O_P(N^{1/2})$  and so D/N converges to 0 in probability. It also follows that

$$E(D) \le \sum_{j=1}^{k} E|N_j - n_j| \le \sum_{j=1}^{k} E|N_j - p_jN| + |p_jN - n_j|$$

$$\leq \sum_{j=1}^{k} \{ E[(N_j - Np_j)^2] \}^{1/2} + O(N^{1/2}) = \sum_{j=1}^{k} [Np_j(1 - p_j)]^{1/2} + O(N^{1/2}) = O(N^{1/2}) . \quad \blacksquare$$

PROOF OF LEMMA 5.2: First, we can write the likelihood ratio  $L_n(x) = dQ_n(x)/dP_n(x)$  as a product of conditional likelihood ratios, i.e., for  $x = (x_1, \ldots, x_k)$ ,

$$L_n(x) = \frac{P\left(\bigcap_{i=1}^k \{H_{n,i} = x_i\}\right)}{P\left(\bigcap_{i=1}^k \{M_{n,i} = x_i\}\right)}$$

$$=\underbrace{\frac{P(H_{n,1}=x_1)}{P(M_{n,1}=x_1)}}_{=L_{n,1}(x_1)}\cdot\underbrace{\frac{P(H_{n,2}=x_2|H_{n,1}=x_1)}{P(M_{n,2}=x_2|M_{n,1}=x_1)}}_{=L_{n,2}(x_2|x_1)}\cdot\cdot\cdot\underbrace{\frac{P(H_{n,k}=n_k|\bigcap_{i=1}^{k-1}\{H_{n,i}=x_i\})}{P(M_{n,k}=n_k|\bigcap_{i=1}^{k-1}\{M_{n,i}=x_i\})}_{=L_{n,k}(x_k|x_1,\dots,x_{k-1})}$$

Note that conditional on  $\bigcap_{j=1}^{i-1} \{M_{n,j} = x_j\}$ ,  $M_{n,i}$  has the binomial distribution  $B_{n,i}$  based on  $(s - \sum_{j=1}^{i-1} x_j)$  trials and success probability  $\tilde{p}_{n,i}$  given by

(S5) 
$$\tilde{p}_{n,i} = \frac{n_i}{\sum_{j=i}^k n_j} \text{ for } i = 1, \dots, k$$

Likewise, conditional on  $\bigcap_{j=1}^{i-1} \{H_{n,j} = x_j\}$ ,  $H_{n,i}$  has the one-dimensional hypergeometric distribution representing the number of objects from population *i* sampled without replacement. Thus, as long as

(S6) 
$$\max(0, \sum_{j=1}^{i} n_j + s - N - \sum_{j=1}^{i-1} x_j) \le x_i \le n_i \quad \text{for} \quad i = 1, \dots, k ,$$
$$\tilde{L}_n(x) = \frac{\binom{n_1}{x_1}\binom{N-n_1}{s-x_1}}{\binom{N}{s}\binom{s}{x_1}\tilde{p}_{n,1}^{x_1}(1-\tilde{p}_{n,1})^{s-x_1}} \cdot \frac{\binom{n_2}{x_2}\binom{N-(n_1+n_2)}{s-x_1-x_2}}{\binom{N-n_1}{s-x_1}\tilde{p}_{n,2}^{x_1}(1-\tilde{p}_{n,2})^{s-x_1-x_2}}$$
$$\cdots \frac{\binom{n_k}{x_k}\binom{N-\sum_{i=1}^k n_i}{s-\sum_{i=1}^k x_i}}{\binom{N-\sum_{i=1}^k n_i}{s-\sum_{i=1}^k x_i}},$$

for  $s = \sum_{i=1}^{k} x_i$ . Of course, when for x in the support of the multivariate hypergeometric distribution, (S6) holds. Moreover, even when x is an atom

from the the multinomial distribution, we can use the fact that  $x_i = (s - \sum_{j=1}^{i-1} x_j)\tilde{p}_{n,i} + O_P(\sqrt{s})$  to yield

$$P\left(\max(0, \sum_{j=1}^{i} n_j + s - N - \sum_{j=1}^{i-1} x_j) \le x_i \le n_i\right) \to 1 .$$

Thus,  $\tilde{L}_n(x) = L_n(x)$  with probability tending to one. Note that the last kth conditional likelihood ratio  $L_{n,k}(x_k|x_1, \ldots, x_{k-1}) = 1$ , and the *i*th conditional likelihood ratio  $L_{n,i}(x_i|x_1, \ldots, x_{i-1})$  for  $i = 1, \ldots, k-1$  becomes

$$L_{n,i}(x_i|x_1, \dots, x_{i-1}) = \frac{P(H_{n,i} = x_i|\bigcap_{j=1}^{i-1}\{H_{n,j} = x_j\})}{P(M_{n,i} = x_i|\bigcap_{j=1}^{i-1}\{H_{n,j} = x_j\})}$$
$$= \frac{\binom{n_i}{x_i}\binom{N-\sum_{j=1}^{i-1}n_j}{s-\sum_{j=1}^{i-1}x_j}}{\binom{N-\sum_{j=1}^{i-1}n_j}{x_i}\binom{S-\sum_{j=1}^{i-1}x_j}{x_i}}\tilde{p}_{n,i}^{x_i}(1-\tilde{p}_{n,i})^{S-\sum_{j=1}^{i}x_j}}$$
$$= \frac{n_i!(N-\sum_{j=1}^{i}n_j)!(N-\sum_{j=1}^{i-1}n_j-s+\sum_{j=1}^{i-1}x_j)!}{(n_i-x_i)!(N-\sum_{j=1}^{i}n_j-s+\sum_{j=1}^{i}x_j)!(N-\sum_{j=1}^{i-1}n_j)!\tilde{p}_{n,i}^{x_i}(1-\tilde{p}_{n,i})^{S-\sum_{j=1}^{i}x_j}}$$
Since  $x_i|x_1,\dots,x_{i-1} \sim B_{n,i} = (s-\sum_{j=1}^{i-1}x_j)\tilde{p}_{n,i} + O_P(s^{1/2})$ , and thus,

$$\min\left(n_i, N - \sum_{j=1}^{i} n_j, n_i - B_{n,i}, N - \sum_{j=1}^{i} n_j - s + \sum_{j=1}^{i-1} x_j + B_{n,i}\right) \xrightarrow{P} \infty ,$$

we can apply Stirling's approximation

$$n! \sim \sqrt{2\pi n} (n/e)^n (1+O(\frac{1}{n}))$$
 as  $n \to \infty$ ,

which yields  $L_{n,i}(x_i|x_1,...,x_{i-1}) \sim L'_{n,i}(x_i|x_1,...,x_{i-1})$ , where  $L'_{n,i}(x_i|x_1,...,x_{i-1}) =$ 

$$\frac{n_{2}^{n_{2}+\frac{1}{2}}(N-\sum_{j=1}^{i}n_{j})^{N-\sum_{i=1}^{j}n_{j}+\frac{1}{2}}(N-\sum_{j=1}^{i-1}n_{j}-s+\sum_{j=1}^{i-1}x_{j})^{N-\sum_{j=1}^{i-1}n_{j}-s+\sum_{j=1}^{i-1}x_{j}+\frac{1}{2}}}{(N-\sum_{j=1}^{i-1}n_{j})^{N-\sum_{j=1}^{i-1}n_{j}+\frac{1}{2}}(n_{i}-x_{i})^{n_{i}-x_{i}+\frac{1}{2}}(N-\sum_{j=1}^{i}-s+\sum_{j=1}^{i}x_{j})^{N-\sum_{j=1}^{i-1}-s+\sum_{j=1}^{i}x_{j}+\frac{1}{2}}\tilde{p}_{n,i}^{x_{i}}(1-\tilde{p}_{n,i})^{s-\sum_{j=1}^{i}x_{j}}}$$
  
Therefore,  $L_{n,i}(B_{n,i}|B_{n,1}=x_{1},\ldots,B_{n,i-1}=x_{i-1})$  has the same limiting

distribution as  $L'_{n,i}(B_{n,i}|B_{n,1}=x_1,\ldots,B_{n,i-1}=x_{i-1})$ . Write  $L'_{n,i}=a \cdot b \cdot c$ and  $\tilde{q}_{n,i}=1-\tilde{p}_{n,i}$ , where

$$a = \frac{n_2^{n_2 + \frac{1}{2}} (N - \sum_{j=1}^{i} n_j)^{N - \sum_{j=1}^{i} n_j + \frac{1}{2}}}{(N - \sum_{j=1}^{i-1} n_j)^{N - \sum_{j=1}^{i-1} n_j + \frac{1}{2}}},$$

$$b = \frac{(N - \sum_{j=1}^{i-1} n_j - s + \sum_{j=1}^{i-1} x_j)^{N - \sum_{j=1}^{i-1} n_j - s + \sum_{j=1}^{i-1} x_j + \frac{1}{2}}{(n_i - x_i)^{n_i - x_i + \frac{1}{2}} (N - \sum_{j=1}^{i} - s + \sum_{j=1}^{i} x_j)^{N - \sum_{j=1}^{i} - s + \sum_{j=1}^{i} x_j + \frac{1}{2}}}$$

and

$$c = \frac{1}{\tilde{p}_{n,i}^{x_i} \tilde{q}_{n,i}^{s-\sum_{j=1}^i x_j}} \; .$$

Then,

$$a = \tilde{p}_{n,i}^{n_2 + \frac{1}{2}} \tilde{q}_{n,i}^{N - \sum_{j=1}^{i} n_j + \frac{1}{2}} (N - \sum_{j=1}^{i-1} n_j)^{\frac{1}{2}},$$

and so

$$a \cdot c = p_{n,i}^{C_i + \frac{1}{2}} q_{n,i}^{D_i + \frac{1}{2}} (N - \sum_{j=1}^{i-1} n_j)^{\frac{1}{2}},$$

where  $C_i = n_i - x_i$  and  $D_i = N - \sum_{j=1}^{i} n_j - s + \sum_{j=1}^{i} x_j$ . Also,

$$b = \frac{(C_i + D_i)^{C_i + D_i + \frac{1}{2}}}{C_i^{C_i + \frac{1}{2}} D_i^{D_i + \frac{1}{2}}} = \left(\frac{C_i + D_i}{C_i}\right)^{C_i + \frac{1}{2}} \left(\frac{C_i + D_i}{D_i}\right)^{D_i + \frac{1}{2}} (C_i + D_i)^{-\frac{1}{2}} .$$

Therefore,  $L'_{n,i} = a \cdot b \cdot c$  equals

$$L'_{n,i} = \left(\frac{C_i}{\tilde{p}_{n,i}(C_i + D_i)}\right)^{-(C_i + \frac{1}{2})} \left(\frac{D_i}{\tilde{q}_{n,i}(C_i + D_i)}\right)^{-(D_i + \frac{1}{2})} \left(\frac{C_i + D_i}{N - \sum_{j=1}^{i-1} n_j}\right)^{-\frac{1}{2}}$$

We will evaluate  $L_{n,i}$  and  $L'_{n,i}$  not at a generic  $x_i$ , but at the binomial variable  $B_{n,i}$  conditioning on  $B_{n,1} = x_1, \ldots, B_{n,i-1} = x_{i-1}$ , which satisfies

$$B_{n,i} = \left(s - \sum_{j=1}^{i-1} x_j\right) \tilde{p}_{n,i} + O_P(s^{1/2}) ,$$

in which case  $C_{n,i} = C_i(B_{n,i}) = n_i - B_{n,i}$  and  $D_{n,i} = D_i(B_{n,i}) = N - \sum_{j=1}^i n_j - s + \sum_{j=1}^{i-1} x_j + B_{n,i}$  satisfy

$$\frac{C_{n,i}}{\tilde{p}_{n,i}(C_{n,i}+D_{n,i})} = \frac{n_i - (s - \sum_{j=1}^{i-1} x_j)\tilde{p}_{n,i}}{\tilde{p}_{n,i}(N - \sum_{j=1}^{i-1} n_j - s + \sum_{j=1}^{i-1} x_j)} + \frac{O_P(s^{1/2})}{\tilde{p}_{n,i}(N - \sum_{j=1}^{i-1} n_j - s + \sum_{j=1}^{i-1} x_j)}$$
$$= 1 + \frac{O_P(s^{1/2})}{\tilde{p}_{n,i}(N - \sum_{j=1}^{i-1} n_j - s + \sum_{j=1}^{i-1} x_j)}$$

and

$$\begin{aligned} \frac{D_{n,i}}{\tilde{q}_{n,i}(C_{n,i}+D_{n,i})} &= \frac{N - \sum_{j=1}^{i} n_j - (s - \sum_{j=1}^{i-1} x_j) \tilde{q}_{n,i}}{\tilde{q}_{n,i}(N - \sum_{j=1}^{i-1} n_j - s + \sum_{j=1}^{i-1} x_j)} + \frac{O_P(s^{1/2})}{\tilde{q}_{n,i}(N - \sum_{j=1}^{i-1} n_j - s + \sum_{j=1}^{i-1} x_j)} \\ &= 1 + \frac{O_P(s^{1/2})}{\tilde{q}_{n,i}(N - \sum_{j=1}^{i-1} n_j - s + \sum_{j=1}^{i-1} x_j)} \ . \end{aligned}$$

Also, since

(S7) 
$$\frac{B_{n,i}}{(s-\sum_{j=1}^{i-1} x_j)} \xrightarrow{P} \tilde{p}_i, \text{ where } \tilde{p}_i = \frac{p_{n,i}}{\sum_{j=i}^k p_{n,j}},$$

we also have

(S8) 
$$\frac{C_{n,i}}{\tilde{p}_{n,i}(C_{n,i}+D_{n,i})} \xrightarrow{P} 1$$
 and  $\frac{D_{n,i}}{\tilde{q}_{n,i}(C_{n,i}+D_{n,i})} \xrightarrow{P} 1$ .

Therefore, we can expand the logarithm of  $L_{n,i}^\prime$  as long as we keep both the linear and quadratic terms,

$$\log(t) = (t-1) - \frac{1}{2}(t-1)^2 + o(|t-1|^2)$$
 as  $t \to 1$ .

Hence,

$$-\log\left[\left(\frac{N-\sum_{j=1}^{i-1}n_j-s+\sum_{j=1}^{i-1}x_j}{N-\sum_{j=1}^{i-1}n_j}\right)^{\frac{1}{2}}L'_{n,i}(B_{n,i}|B_{n,1}=x_1,\dots,B_{n,i-1}=x_{i-1})\right]$$

$$=(C_{n,i}+\frac{1}{2})\log\left(\frac{C_{n,i}}{\tilde{p}_{n,i}(C_{n,i}+D_{n,i})}\right)+(D_{n,i}+\frac{1}{2})\log\left(\frac{D_{n,i}}{\tilde{q}_{n,i}(C_{n,i}+D_{n,i})}\right)$$

$$=\underbrace{C_{n,i}\left(\frac{C_{n,i}}{\tilde{p}_{n,i}(C_{n,i}+D_{n,i})}-1\right)+D_{n,i}\left(\frac{D_{n,i}}{\tilde{q}_{n,i}(C_{n,i}+D_{n,i})}-1\right)}_{A_i}$$

$$-\underbrace{\frac{1}{2}C_{n,i}\left(\frac{C_{n,i}}{\tilde{p}_{n,i}(C_{n,i}+D_{n,i})}-1\right)^2-\frac{1}{2}D_{n,i}\left(\frac{D_{n,i}}{\tilde{q}_{n,i}(C_{n,i}+D_{n,i})}-1\right)^2}_{B_i}+o_P(1) \ .$$

Noting that

(S9) 
$$\frac{D_{n,i}}{\tilde{q}_{n,i}(C_{n,i}+D_{n,i})} - 1 = -\frac{\tilde{p}_{n,i}}{\tilde{q}_{n,i}} \left(\frac{C_{n,i}}{\tilde{p}_{n,i}(C_{n,i}+D_{n,i})} - 1\right) ,$$

 $\mathbf{S8}$ 

we have that

$$\begin{split} A_{i} &= \left(\frac{C_{n,i} - \tilde{p}_{n,i}(C_{n,i} + D_{n,i})}{\tilde{p}_{n,i}(C_{n,i} + D_{n,i})}\right) \cdot \left[C_{n,i} - \frac{\tilde{p}_{n,i}}{\tilde{q}_{n,i}}D_{n,i}\right] = \left(\frac{\tilde{q}_{n,i}C_{n,i} - \tilde{p}_{n,i}D_{n,i}}{\tilde{p}_{n,i}(C_{n,i} + D_{n,i})}\right) \cdot \frac{\tilde{q}_{n,i}C_{n,i} - \tilde{p}_{n,i}D_{n,i}}{\tilde{q}_{n,i}} \\ &= \frac{(\tilde{q}_{n,i}C_{n,i} - \tilde{p}_{n,i}D_{n,i})^{2}}{\tilde{p}_{n,i}\tilde{q}_{n,i}(C_{n,i} + D_{n,i})} = Z_{n,i}^{2} \cdot \frac{s - \sum_{j=1}^{i-1} x_{j}}{N - \sum_{j=1}^{i-1} n_{j} - s + \sum_{j=1}^{i-1} x_{j}} \,, \end{split}$$

where

$$Z_{n,i} = \frac{\tilde{q}_{n,i}C_{n,i} - \tilde{p}_{n,i}D_{n,i}}{\sqrt{(s - \sum_{j=1}^{i-1} x_j)\tilde{p}_{n,i}\tilde{q}_{n,i}}} = -\frac{B_{n,i} - \tilde{p}_{n,i}(s - \sum_{j=1}^{i-1} x_j)}{\sqrt{(s - \sum_{j=1}^{i-1} x_j)\tilde{p}_{n,i}\tilde{q}_{n,i}}} \xrightarrow{L} Z \sim N(0,1) \,.$$

Using first (S9) and then (S8), we find that

$$-2 \cdot B_{i} = \left(\frac{C_{n,i}}{\tilde{p}_{n,i}(C_{n,i} + D_{n,i})} - 1\right)^{2} \cdot \left(C_{n,i} + D_{n,i}\frac{\tilde{p}_{n,i}^{2}}{\tilde{q}_{n,i}^{2}}\right)$$
$$= Z_{n,i}^{2} \cdot \frac{s - \sum_{j=1}^{i-1} x_{j}}{N - \sum_{j=1}^{i-1} n_{j} - s + \sum_{j=1}^{i-1} x_{j}} + o_{P}(1)$$

. Therefore,

$$A_i + B_i = \frac{1}{2} Z_{n,i}^2 \cdot \frac{s - \sum_{j=1}^{i-1} x_j}{N - \sum_{j=1}^{i-1} n_j - s + \sum_{j=1}^{i-1} x_j} + o_P(1) .$$

Since

$$\frac{s - \sum_{j=1}^{i-1} x_j}{N} \xrightarrow{P} \theta \prod_{j=1}^{i-1} (1 - \tilde{p}_j) = \theta (1 - \sum_{j=1}^{i-1} p_j) ,$$

we conclude that

$$L_{n,i}(B_{n,i}|B_{n,1},\dots,B_{n,i-1}) \xrightarrow{L} (1-\theta)^{-\frac{1}{2}} \exp\left(-\frac{\theta}{2(1-\theta)}Z^2\right)$$
 for  $i = 1,\dots,k-1$ ,

where  $Z^2 \sim \chi_1^2$ , and therefore,

$$L_n(M_n) = L_{n,1}(B_{n,1}) \cdot L_{n,1}(B_{n,2}|B_{n,1}) \cdots L_{n,k}(B_{n,k}|B_{n,1}, \dots, L_{n,k}(B_{n,k-1})$$
$$\xrightarrow{L} (1-\theta)^{-\frac{k-1}{2}} \exp\left\{-\frac{\theta}{2(1-\theta)}\chi_{k-1}^2\right\} .$$

To prove (ii), note that

$$E\left[(1-\theta)^{-\frac{k-1}{2}}\exp\left\{-\frac{\theta}{2(1-\theta)}\chi_{k-1}^{2}\right\}\right] = 1 ,$$

since  $\chi^2_{k-1}$  is the Chi-squared distribution with k-1 degrees of freedom and moment generating function  $\psi(t) = (1-2t)^{-\frac{k-1}{2}}$ . Since the mean of the limiting distribution has mean 1, by Theorem 12.3.2 (iii) of Lehmann and Romano [2]  $Q_n$  is contiguous with respect to  $P_n$ . Since the limiting distribution has no mass at 0, by Problem 12.23 of Lehmann and Romano [2], it also follows that  $P_n$  is contiguous to  $Q_n$ .

PROOF OF LEMMA 5.3 Imagine  $V_1, \ldots, V_s$  are sampled in a two-stage process. First  $\overline{M}_n = (\overline{M}_{n,1}, \ldots, \overline{M}_{n,k})$  is drawn from the multinomial distribution with parameters s and  $p = (p_1, \ldots, p_k)$ , where  $s = \sum_{i=1}^k \overline{M}_{n,i}$ . Then,  $V_1, \ldots, V_s$  are obtained by drawing  $\overline{M}_{n,i}$  i.i.d. observations from  $P_i$ for  $i = 1, \ldots, k$ . Similarly, let  $H_{n,i}$  denote the number of observations among  $Z_{\pi(1)}, \ldots, Z_{\pi(s)}$  which were among the  $X_{i,1}, \ldots, X_{i,n_i}$ , so that  $H_n = (H_{n,1}, \ldots, H_{n,k})$ has the hypergeometric distribution based on sampling s objects from N = $\sum_{i=1}^k n_i, n_i$  of which are  $X_{i,1}, \ldots, X_{i,n_i}$ . By Lemma 5.2, Remark 5.1 and (5.14),  $\overline{M}_n$  and  $H_n$  are contiguous. Importantly, conditional on  $\overline{M}_n = H_n =$  $b = (b_1, \ldots, b_k)$ , the conditional probabilities

(S10) 
$$P\{W_n(V_1, ..., V_s) - t | > \epsilon | M_n = b\}$$

and

(S11) 
$$P\{|W_n(Z_{\pi(1)},\ldots,Z_{\pi(s)})-t| > \epsilon | H_n = b\}$$

are the same, because  $W_n$  is evaluated at a random sample of  $b_i$  observations from  $P_i$  for i = 1, ..., k, in both cases. Let  $f_n(\bar{M}_n)$  be defined by

(S12) 
$$f_n(\bar{M}_n) \equiv P\{|W_n(V_1, \dots, V_s) - t| > \epsilon | \bar{M}_n \}$$

By assumption (5.13),

$$E[f_n(\bar{M}_n)] \to 0$$
,

and hence

$$f_n(\bar{M}_n) \stackrel{P}{\to} 0$$
,

by Markov's inequality. But then, by contiguity,

$$f_n(H_n) \xrightarrow{P} 0$$
,

and so

(S13) 
$$E[f_n(H_n)] \to 0$$

since  $f_n$  is uniformly bounded. But (S13) is exactly same as showing, for any  $\epsilon > 0$ ,  $P\{|W_n(Z_{\pi(1)}, \ldots, Z_{\pi(s)}) - t| > \epsilon\} \to 0$  as  $s = s(n) \to \infty$ .

REMARK A.2. The assumption (5.14) is stronger than the more basic assumption  $n_i/N \to p_i$ , where no rate is required between the difference  $n_i/N$ and  $p_i$ . Alternatively, we can replace (5.14) with the more basic assumption  $n_i/N \to p_i$  as long as we slightly strengthen the requirement (5.13) to

$$W_n(Z_{n,1},\ldots,Z_{n,s}) \xrightarrow{P} t$$

when  $Z_{n,1}, \ldots, Z_{n,s}$  are i.i.d. according to the mixture distribution  $\sum_{i=1}^{k} \frac{n_i}{N} P_i$ (rather than  $\sum_{i=1}^{k} p_i P_i$ ), so that the data distribution at time *n* depends on *n*. We prefer to assume the convergence hypothesis based on an i.i.d. sequence, though it is really a matter of choice. Usually, we can appeal to some basic convergence in probability results with ease, but if convergence in probability results are available (or can be derived) which are "uniform" in the underlying probability distribution, then such results can be used instead with the weaker hypothesis  $n_i/N \to p_i$  for  $i = 1, \ldots, k$ .

**A.2.** Proofs of Theorems in Section 2.. PROOF OF THEOREM 2.1 First, argue in the case  $\theta(P) = \mu(P) = \int x dP(x)$ , so  $f_P(x) = x - \theta(P)$  for all x, P and

$$T_{m,n} = N^{1/2} (\bar{X}_m - \bar{Y}_n).$$

Independent of Zs, let  $(\pi(1), \ldots, \pi(N))$  and  $(\pi'(1), \ldots, \pi'(N))$  be independent random permutations of  $\{1, \ldots, N\}$ . We will show

$$\left(T_{m,n}(Z_{\pi(i)}), T_{m,n}(Z_{\pi'(i)})\right)$$

converges in distribution to a bivariate normal distribution with independent, identically distributed marginals having mean 0 and variance

$$\tau^2(\bar{P}) = \frac{1}{p(1-p)}\sigma^2(\bar{P}),$$

where  $\sigma^2(P)$  denotes the variance of P. To do this, we will use the coupling argument of Lemma 5.1, and so we must verify (5.9) and (5.10). When all observations are from the mixture distribution  $\bar{P}$ , the result (5.9) holds by the arguments in Example 15.2.6 of Lehmann and Romano [2]. Moreover, (5.10) is verified in the case of difference of means in Example 5.1. Thus, the result is true.

Next, consider the case  $\theta(P) = \int f(x)dP(x)$ , so  $f_P(x) = f(x) - \theta(P)$ . However, this problem is the same as the mean case. Instead of observing  $(Z_1, \ldots, Z_N) = (X_1, \ldots, X_m, Y_1, \ldots, Y_n)$ , we now observe  $(\tilde{Z}_1, \ldots, \tilde{Z}_N) = (f(X_1), \ldots, f(X_m), f(Y_1), \ldots, f(Y_n))$  and we are interested in means of  $\tilde{Z}$ s. Thus, the proof for this case would be the same as above except we replace  $\sigma^2(P) = E_P(X_i - \mu(P))^2$  with  $E_P(f_P(X_i) - \theta(P))^2$ .

Finally, we consider the general case. Let  $\pi$  be a random permutation of  $\{1, \ldots, N\}$ , so that

$$T_{m,n}(Z_{\pi(1)},\ldots,Z_{\pi(N)}) = N^{1/2}[\hat{\theta}_m(Z_{\pi(1)},\ldots,Z_{\pi(m)}) - \hat{\theta}_n(Z_{\pi(m+1)},\ldots,Z_{\pi(N)})].$$

Let  $V_1, V_2, \ldots$  be i.i.d.  $\overline{P}$ . By assumption,

(S14) 
$$N^{1/2}[\hat{\theta}_m(V_1,\ldots,V_m) - \theta(\bar{P})] - \frac{N^{1/2}}{m} \sum_{i=1}^m f_{\bar{P}}(V_i) \xrightarrow{P} 0.$$

By Lemma 5.3 and (S14),

$$\epsilon_m(Z_{\pi(1)},\ldots,Z_{\pi(m)}) \equiv N^{1/2}[\hat{\theta}_m(Z_{\pi(1)},\ldots,Z_{\pi(m)}) - \theta(\bar{P})] - \frac{N^{1/2}}{m} \sum_{i=1}^m f_{\bar{P}}(Z_{\pi(i)}) \xrightarrow{P} 0$$

Similarly,

$$\epsilon_n(Z_{\pi(m+1)},\ldots,Z_{\pi(N)}) \equiv N^{1/2}[\hat{\theta}_n(Z_{\pi(m+1)},\ldots,Z_{\pi(N)}) - \theta(\bar{P})] - \frac{N^{1/2}}{n} \sum_{j=m+1}^N f_{\bar{P}}(Z_{\pi(j)}) \xrightarrow{P} 0$$

Hence, we can write

$$T_{m,n}(Z_{\pi(1)},\ldots,Z_{\pi(N)}) =$$

$$N^{1/2}\left[\frac{1}{m}\sum_{i=1}^{m}f_{\bar{P}}(Z_{\pi(i)})-\frac{1}{n}\sum_{j=m+1}^{N}f_{\bar{P}}(Z_{\pi(j)})\right]+\epsilon_{m}(Z_{\pi(1)},\ldots,Z_{\pi(m)})-\epsilon_{n}(Z_{\pi(m+1)},\ldots,Z_{\pi(N)})$$

and each of the last two terms goes to zero in probability. Therefore, we can apply Theorem 5.2; that is, it suffices to determine the limit behavior of just

$$N^{1/2}\left[\frac{1}{m}\sum_{i=1}^{m}f_{\bar{P}}(Z_{\pi(i)})-\frac{1}{n}\sum_{j=m+1}^{N}f_{\bar{P}}(Z_{\pi(j)})\right],$$

which reduces the problem to the previous case considered.

REMARK A.3. Using similar arguments, one can deduce the behavior of the permutation distribution even if p = 0 (or p = 1). Of course,  $\tau^2(\bar{P})$  in (2.5) is not properly defined now. But, the scaling factor  $\sqrt{N}$  in the definition of  $T_{m,n}$  plays a minor role and can be replaced by  $\sqrt{\min(m,n)}$  in order to get a nondegenerate limiting distribution.

REMARK A.4. As mentioned in Remark A.2, the assumption (2.4) is of course a little stronger than the more basic assumption  $m/N \to p$ , where no rate is required between the difference m/N and p. Of course, we are free to choose p as m/N in any situation, and the assumption is rather innocuous. (Indeed, for any  $m_0$  and  $N_0$  with  $m_0/N_0 = p$ , we can always let m and N tend to infinity with  $m = km_0$  and  $N = kN_0$  and let  $k \to \infty$ .) Alternatively, we can replace (2.4) with the more basic assumption  $m/N \to p$ as long as we slightly strengthen the basic assumption that the statistic has a linear expansion under  $\bar{P} = pP + qQ$  to also have a linear expansion under sequences

$$\bar{P}_{m,n} = \frac{m}{N}P + \frac{n}{N}Q \; ,$$

which is a rather weak form of local uniform triangular array type of convergence. We prefer to assume the convergence hypothesis based on an i.i.d. sequence from a fixed  $\bar{P}$ , though it is really a matter of choice. Usually, we can appeal to some basic convergence in distributions results with ease, but if linear expansions are available (or can be derived) which are "uniform" in the underlying probability distribution near  $\bar{P}$ , then such results can be used instead with the weaker hypothesis  $p_m \to p$ .

PROOF OF THEOREM 2.2: Write  $V_{m,n} = V_{m,n}(Z_1, \ldots, Z_N)$ , where the  $Z_i$  are defined in (1.1). Let  $(\pi(1), \ldots, \pi(N))$  denote a random permutation of

 $\{1, \ldots, N\}$  (and independent of all other variables). We first will show that

(S15) 
$$V_{m,n}^2(Z_{\pi(1)},\ldots,Z_{\pi(N)}) \xrightarrow{P} \tau^2(\bar{P})$$
.

To do this, it suffices to show that

(S16) 
$$\hat{\sigma}_m^2(Z_{\pi(1)},\ldots,Z_{\pi(m)}) \xrightarrow{P} \sigma^2(\bar{P})$$

and

(S17) 
$$\hat{\sigma}_n^2(Z_{\pi(m+1)},\ldots,Z_{\pi(N)}) \xrightarrow{P} \sigma^2(\bar{P}) .$$

But (S16) and (S17) both follow from Lemma 5.3. Now let  $R_{m,n}^V(\cdot)$  denote the permutation distribution corresponding to the statistic  $V_{m,n}$ , as defined in (1.2) with T replaced by V. By Lemma A.1,  $\hat{R}_{m,n}^V(t)$  converges to  $\delta_{\tau^2(\bar{P})}(t)$ for all  $t \neq \tau^2(\bar{P})$ , where  $\delta_c(\cdot)$  denotes the c.d.f. of the distribution placing mass one at the constant c. Using this fact together with Theorem 2.1, we can apply Lemma A.3 to conclude that the permutation distribution of the ratio of statistics  $S_{m,n}$  satisfies (2.9).

**A.3.** Proofs in Section 3.. PROOF OF LEMMA 3.1 First, consider the case where  $\theta(P_i) = \int f_{P_i}(x) dP_i(x)$ . Without loss of generality, assume  $\theta(P_i) = 0$  for all *i*. Let  $Z_n$  be the column vector with *i*th component  $n_i^{1/2} \hat{f}_{P_i}/\sigma_i$ , where  $\hat{f}_{P_i} \equiv \frac{1}{n_i} \sum_{j=1}^{n_i} f_{P_i}(X_{i,j})$ . Also, let *I* denote the  $k \times k$ identity matrix, let **1** denote the  $k \times 1$  vector of ones, and let  $D_n$  denote the diagonal matrix with (i, i) entry  $N\sigma_i^2/n_i$ . Then, we can write

$$T_{n,0} = Z'_n P_n Z_n \; .$$

where

$$P_n \equiv (I - \frac{D_n^{-1/2} \mathbf{11}' D_n^{-1/2}}{\mathbf{1}' D_n^{-1} \mathbf{1}}) \ .$$

Of course,  $Z_n$  converges in distribution to Z, where Z has the multivariate normal distribution with mean 0 and covariance matrix I. If we let D denote the diagonal matrix with (i, i) entry  $\sigma_i^2/p_i$ , then the convergence of  $D_n$  to D (for each entry) as well as the convergence of  $D_n^{-1}$  to  $D^{-1}$  implies (using the continuous mapping theorem) that

$$T_{n,0} \stackrel{d}{\to} Z'PZ$$

where  $\boldsymbol{P}$  is the matrix

(S18) 
$$P \equiv \left(I - \frac{D^{-1/2} \mathbf{11}' D^{-1/2}}{\mathbf{1}' D^{-1} \mathbf{1}}\right) \,.$$

The matrix P is a symmetric idempotent or projection matrix, and its rank therefore is its trace, which is then easily checked to be k - 1. Indeed, Prepresents the projection orthogonal to the unit vector  $D^{-1/2}\mathbf{1}/\mathbf{1}'D^{-1}\mathbf{1}$ . It follows that  $Z'PZ \sim \chi^2_{k-1}$ , as required.

To handle  $T_{n,1}$ , let  $t_n$  be the column vector with *i*th component  $n_i^{1/2} \hat{f}_{P_i} / \hat{\sigma}_{n,i}$ and let  $\hat{D}_n$  be the diagonal matrix with (i, i) entry  $N\hat{\sigma}_{n,i}^2/n_i$ . Then, let  $\hat{P}_n$ be the projection matrix where D is replaced by  $\hat{D}_n$  in the definition (S18) of P. Of course by Slutsky's Theorem,  $Z_n - t_n$  converges in probability to 0. Also,  $\hat{D}_n$  converges in probability to D (as well as its inverse),  $\hat{P}_n$  converges in probability to P, and so  $\hat{P}_n - P_n$  converges in probability to 0. Since

$$T_{n,0} - T_{n,1} = Z'_n P_n Z_n - t'_n \hat{P}_n t_n$$
$$= (Z_n - t_n)' P_n (Z_n - t_n) + 2(Z_n - t_n)' P_n t_n + t'_n (P_n - \hat{P}_n) t_n \xrightarrow{P} 0,$$

then  $T_{n,1}$  must have the same limiting distribution as that of  $T_{n,0}$ .

Finally, consider the general case when the estimators are asymptotically linear as in (2.2). Let  $W_n$  and  $V_n$  be the column vector with *i*th component  $n_i^{1/2}\hat{\theta}_{n,i}/\sigma_i$  and  $n_i^{1/2}\hat{\theta}_{n,i}/\hat{\sigma}_i$ , respectively. Then, the test statistics  $T_{n,0}$  and  $T_{n,1}$  become

$$T_{n,0} = W'_n P_n W_n$$
 and  $T_{n,1} = V'_n P_n V_n$ .

Using the fact that  $W_n = Z_n + o_P(1)$  and  $V_n = t_n + o_P(1)$ , we can apply the continuous mapping theorem to conclude that both  $T_{n,0}$  and  $T_{n,1}$  converge in distribution to the same limiting distribution as the previous case considered.

PROOF OF THEOREM 3.1 Put all the  $N = \sum_{i=1}^{k} n_i$  observations in one vector

$$Z^{N} \equiv (Z_{1}, \dots, Z_{N}) = (X_{1,1}, \dots, X_{1,n_{1}}, \dots, X_{k,1}, \dots, X_{k,n_{k}}) ,$$

where  $X_{i,1}, \ldots, X_{i,n_i}$  are i.i.d.  $P_i$  for  $i = 1, \ldots, k$ . Also, let  $\overline{Z}^N$  be a sample of N i.i.d. observations

$$\bar{Z}^N \equiv (\bar{Z}_1, \dots, \bar{Z}_N) \equiv (\bar{Z}_{1,1}, \dots, \bar{Z}_{1,n_1}, \dots, \bar{Z}_{k,1}, \dots, \bar{Z}_{k,n_k})$$

from the mixture distribution

$$\bar{P} \equiv p_1 P_1 + \dots + p_k P_k$$
.

Without loss of generality, we can assume  $\theta(\bar{P}) = 0$  and we write  $\sigma^2 = \sigma^2(f_{\bar{P}}) = \mathbb{E}_{\bar{P}} f_{\bar{P}}^2(\bar{Z}_{i,j}).$ 

For now, we consider the case where all the N observations are i.i.d., so that  $P_i = \overline{P}$  for i = 1, ..., k. First, we will show the randomization distribution based on  $T_{n,0}$ , say  $\hat{R}_{n,0}(\cdot)$ , behaves the same as  $T_{n,1}$ . (Of course, we can't use  $T_{n,0}$  as  $\sigma$  is unknown, but we treat it now in essence as if it is known.) Let  $\pi = (\pi(1), ..., \pi(N))$  denote a random permutation of  $\{1, ..., N\}$  (and independent of the observations). From Theorem 5.1, we must verify

(S19) 
$$(T_{n,0}(\overline{Z}), T_{n,0}(\overline{Z}_{\pi})) \xrightarrow{d} (T, T')$$
,

where T and T' are independent and each distributed as the Chi-squared distribution with degrees of freedom k - 1. (Note that we do not need to consider the joint behavior of  $T_{n,0}$  at  $\bar{Z}_{\pi}$  and at  $\bar{Z}_{\pi'}$ , where  $\pi'$  is another independent random permutation, because since the  $\bar{Z}_i$  are i.i.d.,  $\bar{Z}_{\pi'}$  and  $\bar{Z}$ have the same distribution.) When all the N observations are i.i.d.  $\bar{P}$ ,  $T_{n,0}$ simplifies to

$$T_{n,0} = \frac{1}{\sigma^2} \sum_{i=1}^k \left[ n_i^{1/2} \hat{\theta}_{n,i} - \sum_{j=1}^k n_j^{1/2} \hat{\theta}_{n,j} n_i^{1/2} n_j^{1/2} / N \right]^2$$
$$= \frac{1}{\sigma^2} \sum_{i=1}^k n_i \left( \bar{f}_{\bar{P}}(\bar{Z}_{n,i}) - \bar{f}_{\bar{P}}(\bar{Z}_N) \right)^2 + o_P(1),$$

where  $\bar{f}_{\bar{P}}(\bar{Z}_{n,i}) = \frac{1}{n_i} \sum_{j=1}^{n_i} f_{\bar{P}}(\bar{Z}_{i,j})$  and  $\bar{f}_{\bar{P}}(\bar{Z}_N) = \frac{1}{N} \sum_{l=1}^{N} f_{\bar{P}}(\bar{Z}_l)$  and the  $o_P(1)$  term on the right side is derived from the condition (2.2) under  $\bar{P}$ .

Therefore, we can apply both Theorem 5.1 and Slutsky's Theorem for randomization distributions (Theorem 5.2); that is, it suffices to show that the joint limit behavior

$$\left(\frac{1}{\sigma^2}\sum_{i=1}^k n_i \left(\bar{f}_{\bar{P}}(\bar{Z}_{n,i}) - \bar{f}_{\bar{P}}(\bar{Z}_N)\right)^2, \ \frac{1}{\sigma^2}\sum_{i=1}^k n_i \left(\bar{f}_{\bar{P}}(\bar{Z}_{n,i}) - \bar{f}_{\bar{P}}(\bar{Z}_N)\right)^2\right) \xrightarrow{d} (T,T')$$

To do this, define

$$V_{n,i} = n_i^{1/2} \bar{f}_{\bar{P}}(\bar{Z}_{n,i}) = n_i^{-1/2} \sum_{l=1}^N f_{\bar{P}}(\bar{Z}_l) I\{l \in I_i\}$$

and

$$V'_{n,j} = n_j^{-1/2} \sum_{l=1}^N f_{\bar{P}}(\bar{Z}_l) I\{\pi(l) \in I_j\} ,$$

where  $I_i$  is the set of indices corresponding to the *i*th sample; that is,  $I_1 = \{1, \ldots, n_1\}$ .  $I_2 = \{n_1 + 1, \ldots, n_1 + n_2\}$ , and ultimately  $I_k = \{N - n_k + 1, \ldots, N\}$ . We claim the joint asymptotic normality of

$$(V_{n,1},\ldots,V_{n,k},V'_{n,1},\ldots,V'_{n,k})$$
.

To do this we use the Cramér-Wold device, i.e., we must show that

$$V_n \equiv V_N(a,b) \equiv \sum_{i=1}^k (a_i V_{n,i} + b_i V'_{n,i})$$

is asymptotically normal for any choices of constants  $a_i$  and  $b_i$ . We can write

$$V_n = \sum_{l=1}^N C_{n,l} f_{\bar{P}}(\bar{Z}_l) ,$$

where

$$C_{n,l} = \sum_{i=1}^{k} \left[ \frac{a_i I\{l \in I_i\}}{n_i^{1/2}} + \frac{b_i I\{\pi(l) \in I_i\}}{n_i^{1/2}} \right] .$$

Note that the  $C_{n,l}$  are random (as they depend on the random permutation  $\pi$ ), but are independent of the  $Z_l$ . By Lemma 11.3.3 in Lehmann and Romano [2], a sufficient condition for

(S20) 
$$\sum_{l=1}^{N} C_{n,l} f_{\bar{P}}(\bar{Z}_l) / \sqrt{\sum_{l=1}^{N} C_{n,l}^2} \xrightarrow{d} N(0,\sigma^2)$$

is

(S21) 
$$\frac{\max_{l=1,\dots,N} C_{n,l}^2}{\sum_{l=1}^N C_{n,l}^2} \xrightarrow{P} 0$$

as  $N \to \infty$ . Note that

$$\begin{split} C_{n,l}^2 &= \sum_{i=1}^k \left[ \frac{a_i I\{l \in I_i\}}{n_i^{1/2}} + \frac{b_i I\{\pi(l) \in I_i\}}{n_i^{1/2}} \right] \cdot \sum_{j=1}^k \left[ \frac{a_j I\{l \in I_j\}}{n_j^{1/2}} + \frac{b_j I\{\pi(l) \in I_j\}}{n_j^{1/2}} \right] \\ &= \sum_{i=1}^k \frac{a_i^2}{n_i} I\{l \in I_i\} + \sum_{i=1}^k \sum_{j=1}^k \frac{a_i}{n_i^{1/2}} \frac{b_j}{n_j^{1/2}} I\{l \in I_i, \ \pi(l) \in I_j\} \\ &+ \sum_{i=1}^k \sum_{j=1}^k \frac{b_i a_j}{n_i^{1/2} n_j^{1/2}} I\{\pi(l) \in I_i, \ l \in I_j\} + \sum_{i=1}^k \frac{b_i^2}{n_i} I\{\pi(l) \in I_i\} \ . \end{split}$$

Certainly,

$$\max_{l=1,\dots,N} C_{n,l}^2 = O_P(1/N) \to 0 \; .$$

Furthermore,

$$\sum_{l=1}^{N} C_{n,l}^{2} = \sum_{i=1}^{k} \frac{a_{i}^{2}}{n_{i}} \sum_{l=1}^{N} I\{l \in I_{i}\} + \sum_{i=1}^{k} \sum_{j=1}^{k} \frac{a_{i}}{n_{i}^{1/2}} \frac{b_{j}}{n_{j}^{1/2}} \sum_{l=1}^{N} I\{l \in I_{i}, \ \pi(l) \in I_{j}\}$$
$$+ \sum_{i=1}^{k} \sum_{j=1}^{k} \frac{b_{i}a_{j}}{n_{i}^{1/2}n_{j}^{1/2}} \sum_{l=1}^{N} I\{\pi(l) \in I_{i}, \ l \in I_{j}\} + \sum_{i=1}^{k} \frac{b_{i}^{2}}{n_{i}} \sum_{l=1}^{N} I\{\pi(l) \in I_{i}\}$$
$$= \sum_{i=1}^{k} (a_{i}^{2} + b_{i}^{2}) + \sum_{i=1}^{k} \sum_{j=1}^{k} \frac{a_{i}}{n_{i}^{1/2}} \frac{b_{j}}{n_{j}^{1/2}} \sum_{l=1}^{N} [I\{l \in I_{i}, \ \pi(l) \in I_{j}\} + I\{\pi(l) \in I_{i}, l \in I_{j}\}]$$

Now, the term

(S22) 
$$A_n(i,j) \equiv \sum_{l=1}^N I\{l \in I_i, \pi(l) \in I_j\}$$

represents the  $n_i$  indices in  $I_i$  such that, after permuted by  $\pi$ , are in  $I_j$ ; hence, its distribution is that of the hypergeometric distribution when sampling  $n_i$  observations from N, of which  $n_j$  are "special". The expectation of (S22) is then  $n_i n_j/N$ . Hence,

$$E[A_n(i,j)/n_i] \to p_j$$

and  $Var[A_n(i, j)/n_i] = O(1/n_i)$ , implying

$$A_n(i,j)/n_i \xrightarrow{P} p_j$$
.

It follows that

(S23) 
$$\sum_{l=1}^{N} C_{n,l}^{2} \xrightarrow{P} \sum_{i=1}^{k} (a_{i}^{2} + b_{i}^{2}) + 2 \sum_{i=1}^{k} \sum_{j=1}^{k} a_{i} b_{j} p_{i}^{1/2} p_{j}^{1/2}$$

Of course, the right side of (S23) is nonnegative. By the Cauchy-Schwarz inequality,

$$|\sum_{i=1}^{k} a_i p_i^{1/2}| \le [\sum_{i=1}^{k} a_i^2]^{1/2}$$

with equality if and only if  $a_i = cp_i^{1/2}$  for some constant c. It follows that the right side of (S23) is greater than or equal to  $(A^{1/2} - B^{1/2})^2$ , where  $A = \sum_i a_i^2$  and  $B = \sum_i b_i^2$ , and is equal to 0 if and only if A = B, i.e.  $a_i = cp_i^{1/2}$  and  $b_i = -cp_i^{1/2}$ .

When the right side of (S23) is positive, we have that the condition (S21) holds, and so

(S24) 
$$\sum_{l=1}^{N} C_{n,l} f_{\bar{P}}(\bar{Z}_l) \xrightarrow{d} N(0, \sigma^2 [\sum_{i=1}^{k} (a_i^2 + b_i^2) + 2\sum_{i=1}^{k} \sum_{j=1}^{k} a_i b_j p_i^{1/2} p_j^{1/2}]) .$$

But, even if the right side of (S23) is zero, we can still claim  $\sum_{l} C_{n,l} f_{\bar{P}}(\bar{Z}_{l})$ converges in distribution to N(0,0), i.e., it converges in probability to 0. To see why,  $\sum_{l} C_{n,l} f_{\bar{P}}(\bar{Z}_{l})$  has mean 0 and variance  $\sigma^{2} \sum_{l} E(C_{n,l}^{2})$ . But the above argument showing  $\sum_{l} C_{n,l}^{2}$  converges to 0 in probability (in this case only) shows that its expectation does as well.

In general, we can now conclude that

$$(V_{n,1},\ldots,V_{n,k},V'_{n,1},\ldots,V'_{n,k}) \stackrel{d}{\to} (V,V')$$

is asymptotically multivariate normal with mean 0 (and each of V and V' are k-vectors). Moreover, by appropriate choices of constants  $a_i$  and  $b_i$ , we can read off the covariance matrix from the limiting variance in (S24). In particular, by taking  $a_i = 1$  and  $a_j = 0$  if  $j \neq i$  and taking  $b_j = 0$  for all j, yields  $Var(V_{n,i}) = \sigma^2$ . Also,  $Cov(V_{n,i}, V_{n,j}) = 0$  if  $i \neq j$ . Similarly,  $Var(V'_{n,j}) = \sigma^2$ , and for  $i \neq j$ , (by taking  $a_i = 1 = b_j$  and the rest of the constants 0),

(S25) 
$$Cov(V_{n,i}, V'_{n,j}) = \sigma^2 (p_i p_j)^{1/2}$$

Of course, the statistic  $T_{n,0}$  that is of current interest is indeed a function of the  $V_{n,i}$ ; however, the fact that the covariances in (S25) are nonzero would not allow us to conclude the asymptotic independence of  $T_{n,0}(Z)$ and  $T_{n,0}(Z_{\pi})$ . So we first need to consider a simple transformation of the  $V_{n,i}$  and  $V'_{n,j}$ . For  $i = 1, \ldots, k$ , define

$$W_{n,i} \equiv n_i^{1/2} (\bar{f}_{\bar{P}}(\bar{Z}_{n,i}) - \bar{f}_{\bar{P}}(\bar{Z}_N))$$
$$= V_{n,i} - n_i^{1/2} \bar{f}_{\bar{P}}(\bar{Z}_N) = V_{n,i} - (n_i/N)^{1/2} \sum_{m=1}^k p_m^{1/2} V_{n,m} .$$

Similarly,

$$W'_{n,j} = V'_{n,j} - (n_j/N)^{1/2} \sum_{m=1}^k p_m^{1/2} V'_{n,m} .$$

The joint asymptotic multivariate normality of the  $V_{n,i}$  together with the  $V'_{n,j}$  implies the joint asymptotic multivariate normality of the  $W_{n,i}$  together with the  $W'_{n,j}$ . Indeed,

$$(W_{n,1},\ldots,W_{n,k},W'_{n,1},\ldots,W'_{n,k}) \xrightarrow{d} (W_1,\ldots,W_k,W'_1,\ldots,W'_k)$$
,

where

$$W_i = V_i - p_i^{1/2} \sum_{m=1}^k p_m^{1/2} V_m$$

and

$$W'_j = V'_j - p_j^{1/2} \sum_{m=1}^k p_m^{1/2} V'_m$$

Importantly,

$$Cov(W_i, W'_j) = Cov(V_i - p_i^{1/2} \sum_{m=1}^k p_m^{1/2} V_m, V'_j - p_j^{1/2} \sum_{m=1}^k p_m^{1/2} V'_m)$$
  
=  $Cov(V_i, V'_j) - p_j^{1/2} \sum_{m=1}^k p_m^{1/2} Cov(V_i, V'_m) - p_i^{1/2} \sum_{m=1}^k p_m^{1/2} Cov(V_m, V'_j)$   
+ $(p_i p_j)^{1/2} \sum_{l=1}^k \sum_{m=1}^k (p_l p_m)^{1/2} Cov(V_l, V'_m)$ 

$$= \sigma^{2}[(p_{i}p_{j})^{1/2} - p_{j}^{1/2}\sum_{m=1}^{k} p_{i}^{1/2}p_{m} - p_{i}^{1/2}\sum_{m=1}^{k} p_{j}^{1/2}p_{m} + (p_{i}p_{j})^{1/2}\sum_{l=1}^{k}\sum_{m=1}^{k} p_{l}p_{m}]$$
$$= \sigma^{2}[(p_{i}p_{j})^{1/2} - (p_{i}p_{j})^{1/2} - (p_{i}p_{j})^{1/2} + (p_{i}p_{j})^{1/2}] = 0.$$

It follows that  $(W_1, \ldots, W_k)$  and  $(W'_1, \ldots, W'_k)$  are independent. But since

$$T_{n,0}(\bar{Z}) = \frac{1}{\sigma^2} \sum_{i=1}^{k} W_{n,i}^2 \xrightarrow{d} \frac{1}{\sigma^2} \sum_{i=1}^{k} W_i^2$$

and

$$T_{n,0}(\bar{Z}_{\pi}) = \frac{1}{\sigma^2} \sum_{i=1}^k (W'_{n,i})^2 \xrightarrow{d} \frac{1}{\sigma^2} \sum_{i=1}^k (W'_i)^2 ,$$

it now follows that  $T_{n,0}(\bar{Z})$  and  $T_{n,0}(\bar{Z}_{\pi})$  are asymptotically independent. Moreover, by Lemma 3.1,  $T_{n,0}(Z)$  is asymptotically Chi-squared with k-1 degrees of freedom. Since,  $T_{n,0}(Z_{\pi})$  has the same distribution as  $T_{n,0}(Z)$ , it has the same limiting distribution as well.

Next, we show the same result with  $T_{n,1}$  replaced by  $T_{n,0}$ . However, by the fact that Z and  $Z_{\pi}$  have the same distribution,

$$T_{n,1}(\bar{Z}_{\pi}) - T_{n,0}(\bar{Z}_{\pi}) \stackrel{d}{=} T_{n,1}(\bar{Z}) - T_{n,0}(\bar{Z}) ,$$

and so by the proof of Lemma 3.1,

$$T_{n,1}(\bar{Z}_{\pi}) - T_{n,0}(\bar{Z}_{\pi}) \xrightarrow{P} 0$$

Writing  $T_{n,1} = T_{n,0} + [T_{n,1} - T_{n,0}]$ , we can then apply Slutsky's Theorem for Randomization distributions to conclude that  $\hat{R}_{n,1}(\cdot)$  has the same limiting behavior as  $\hat{R}_{n,0}(\cdot)$ .

The proof is now complete under the assumption that all N observations are i.i.d. We now argue, using the coupling argument in Section section:coupling, that the behavior of the permutation distribution under general  $P_1, \ldots, P_k$  is the same as when all observations are i.i.d. with distribution given by the mixture distribution  $\bar{P}$ . So, construct Z,  $\bar{Z}$  and  $\bar{Z}_{\pi_0}$  as in the coupling construction. By Lemma 5.1, it now suffices to show that, for a random permutation  $\pi$ ,

(S26) 
$$T_{n,1}(Z_{\pi}) - T_{n,1}(\bar{Z}_{\pi\pi_0}) \xrightarrow{P} 0$$
.

Write (S27)

$$T_{n,1}(Z) = \sum_{i=1}^{k} \frac{1}{\hat{\sigma}_{n,i}^2} \left[ n_i^{1/2} \bar{f}_{P_i}(X_{n,i}) - \frac{\sum_{j=1}^{k} n_j^{1/2} \bar{f}_{P_i}(X_{n,i}) (n_i^{1/2} n_j^{1/2} / N) / \hat{\sigma}_{n,j}^2}{\sum_{j=1}^{k} (n_j / N) / \hat{\sigma}_{n,j}^2} \right]^2 + o_P(1)$$

Then,  $T_{n,1}(Z_{\pi})$  is computed by replacing

$$n_i^{1/2}\bar{f}_{P_i}(X_{n,i}) = n_i^{1/2}\bar{f}_{P_i}(Z) = n_i^{-1/2}\sum_{l=1}^N f_{P_i}(Z_l I\{l \in I_i\})$$

with

$$n_i^{1/2} \bar{f}_{P_i}(Z_{\pi}) = n_i^{-1/2} \sum_{l=1}^N f_{P_i}(Z_l I\{\pi(l) \in I_i\})$$

and  $\hat{\sigma}_{n,i}^2(Z) = \hat{\sigma}_{n,i}^2(Z_l I\{l \in I_i\})$  gets replaced by

$$\hat{\sigma}_{n,i}^2(Z_\pi) \equiv \hat{\sigma}_{n,i}^2(Z_l I\{\pi(l) \in I_i\}) .$$

Note that the last  $o_P(1)$  term in (S27) can be "negligible" by the Slutusky's Theorem for randomization distributions.

From (S27), it now suffices to show that, for each i,

(S28) 
$$n_i^{1/2} \bar{f}_{P_i}(Z_{\pi}) - n_i^{1/2} \bar{f}_{P_i}(\bar{Z}_{\pi\pi_0}) \xrightarrow{P} 0$$

and

(S29) 
$$\hat{\sigma}_{n,i}^2(Z_{\pi}) - \hat{\sigma}_{n,i}^2(\bar{Z}_{\pi\pi_0}) \xrightarrow{P} 0$$

To show (S28), first note that the left side has mean 0; so, it suffices to show its variance tends to 0. Now, remember that  $Z_{\pi}$  and  $\bar{Z}_{\pi\pi_0}$  differ in at most  $D = O_P(N^{1/2})$  entries. But, conditional on  $\pi$ ,  $\pi_0$  and the multinomial variables  $(N_1, \ldots, N_k)$  in the coupling construction, for indices l where  $Z_l \neq \bar{Z}_{\pi_0(l)}$ ,

$$Var\left(f_{P_i}(Z_l) - f_{P_i}(\bar{Z}_{\pi_0(l)}) | \pi, \pi_0, N_1, \dots, N_k\right) \le 2V$$
.

where  $V = \max(\sigma_1^2(f_{P_1}), \ldots, \sigma_k^2(f_{P_k}))$ . But the left side of (S28) is

$$n_i^{-1/2} \sum_{l=1}^N [f_{P_i}(Z_l) - f_{P_i}(\bar{Z}_{\pi_0(l)})] I\{\pi(l) \in I_i\} ,$$

and the sum here is conditionally a sum of at most D independent variables with variance  $\leq 2V$ . Hence, the variance of the left side of (S28) is conditionally at most  $2VD/n_i$ , and hence the unconditional variance is at most  $2VE(D)/n_i \rightarrow 0$ .

To show (S29), note that for  $i = 1, \ldots, k$ ,

$$\hat{\sigma}_{n,i}^2\left(\bar{Z}_1,\ldots,\bar{Z}_{n_i}\right) \xrightarrow{P} \sigma^2(f_{\bar{P}}) ,$$

which implies  $\hat{\sigma}_{n,i}^2(\bar{Z}_{\pi\cdot\pi_0})) \xrightarrow{P} \sigma^2$ . It also follows by Lemma 5.3 that

$$\hat{\sigma}_{n,i}^2 \left( Z_{\pi(l)} I\{\pi(l) \in I_i\} \right) \xrightarrow{P} \sigma^2(f_{\bar{P}}) \ .$$

Thus, the result (S29) now follows.

## REFERENCES

- HOEFFDING, W. (1952). The large-sample power of tests based on permutations of observations. Ann. Math. Statistics 23, 169–192. MR0057521 MR2571714
- [2] LEHMANN, E. L. and ROMANO, J. (2005). Testing Statistical Hypotheses. 3rd edition, Springer-Verlag, New York.

Department of Economics Stanford University Stanford, California 94305-6072 USA E-MAIL: eunyi@stanford.edu DEPARTMENT OF STATISTICS AND ECONOMICS STANFORD UNIVERSITY STANFORD, CALIFORNIA 94305-4068 USA E-MAIL: romano@stanford.edu