Supporting Information for Distance-Based Analysis of Variance for Brain Connectivity by Shinohara et al.

July 23, 2019

Software

In Section 3 of the manuscript, we analyze the performance of our test in the context of a DTI study of connectomics in autism. Software implementing these analyses in R is available on Github at `https://github.com/rshinohara/distance_statistics_software/`.

Figures

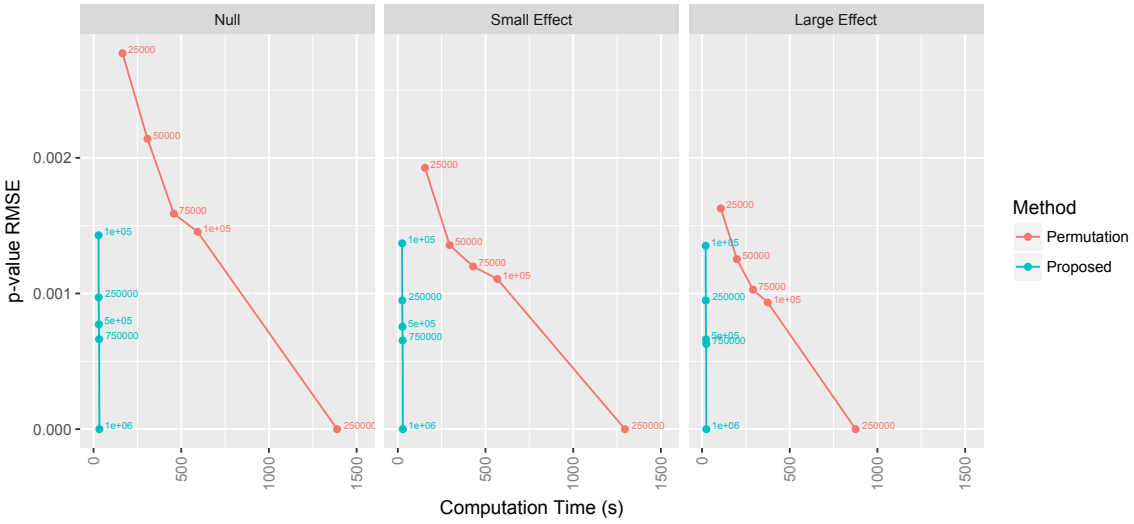

Figure A1: Comparison of p-value estimation and computation time for tuning parameters for the proposed and permutation-based tests. Number in green indicate the number of Monte Carlo samples for the proposed test, and numbers in red indicate the number of permutations used. The three columns indicate three separate simulation scenarios in which 500 graphical simulations from Scenario 2 were conducted using $n = 250$, $\pi_0 = \frac{1}{2}$, $\tau_0 = 0.5$, and $\tau_1 = 0.5, 0.1$, and 0.2 in the left, middle, and right columns respectively.

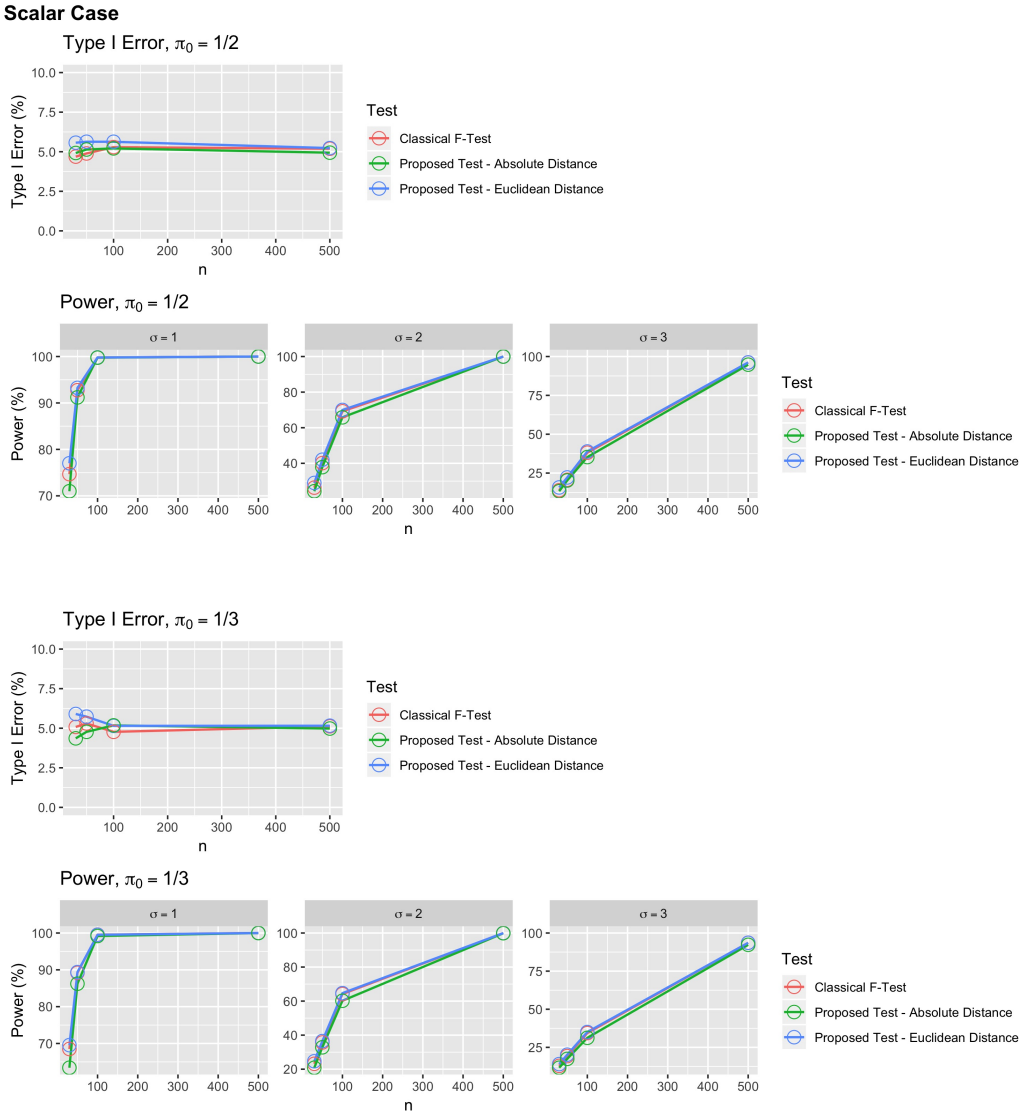

Figure A2: Figures showing the type I error rates and power for various settings with scalar outcomes in Scenario 1. The top two rows show type I error rates and power under several alternatives for the case of $\pi_0 = \frac{1}{2}$, and the bottom two rows show results for the case of imbalanced group sizes.

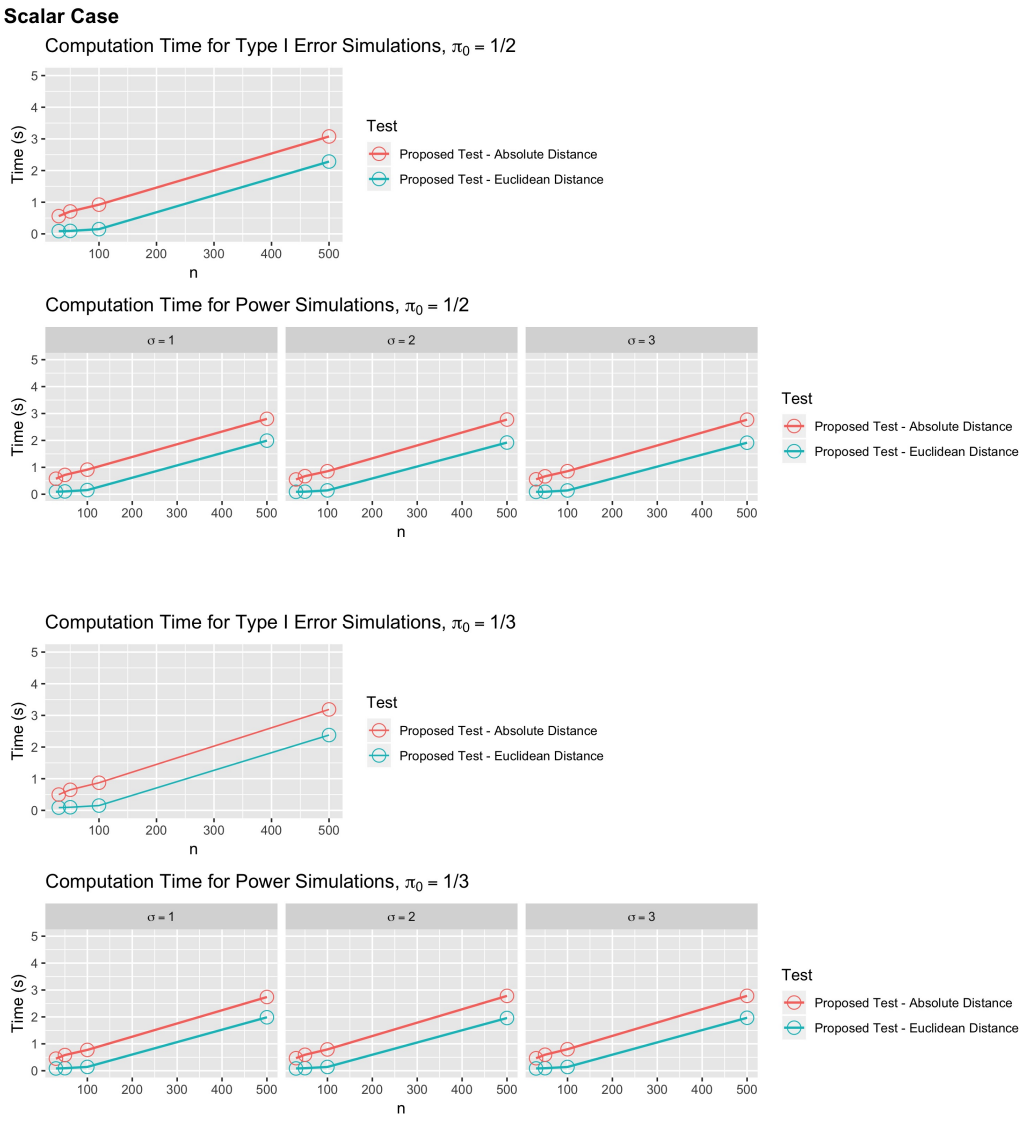

Figure A3: Figures showing the computation time for various settings with scalar outcomes in Scenario 1 in seconds.

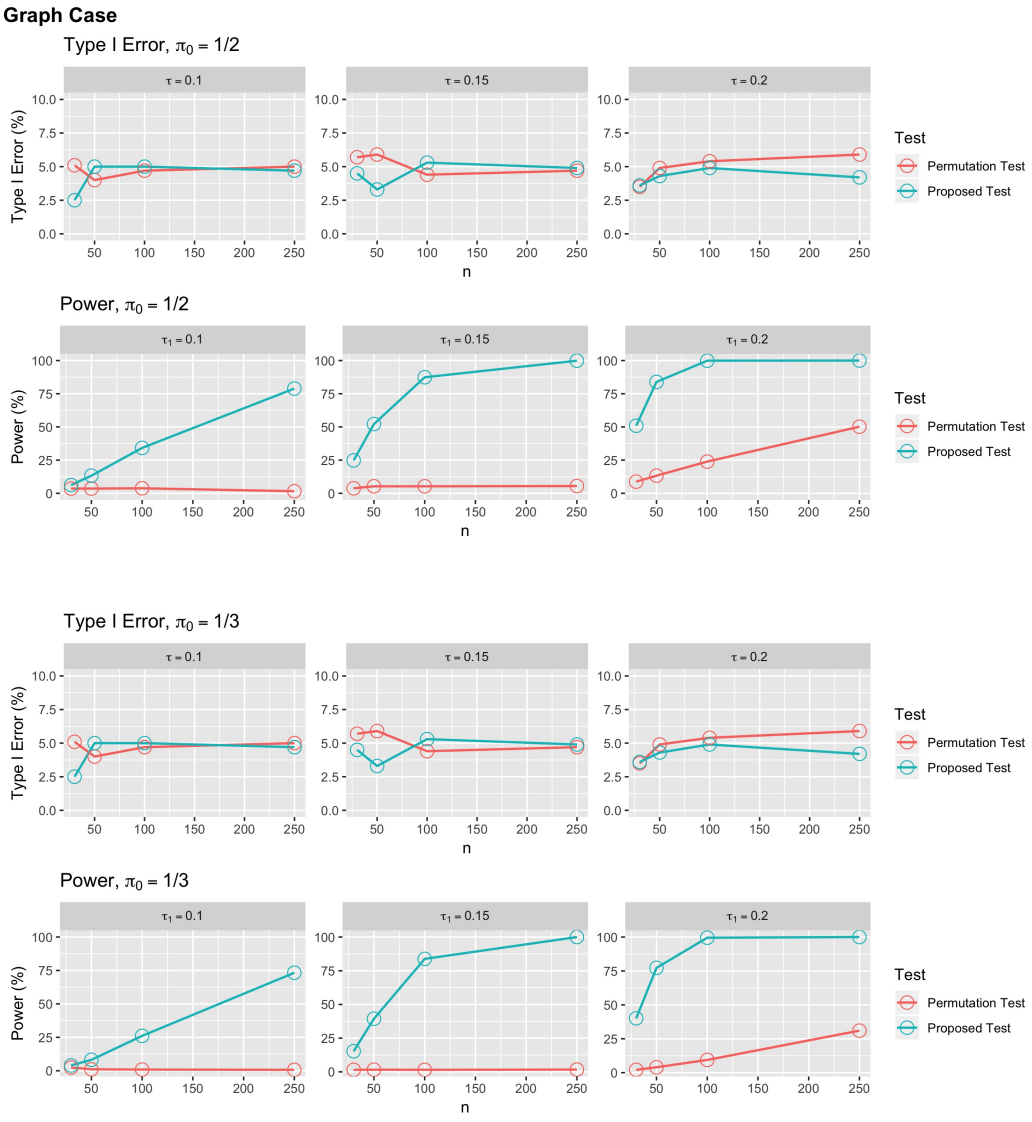

Figure A4: Figures showing the type I error rates and power for various settings with graph outcomes in Scenario 2. The top two rows show type I error rates for several noise levels and power under several alternatives for the case of $\pi_0 = \frac{1}{2}$, and the bottom two rows show results for the case of imbalanced group sizes.

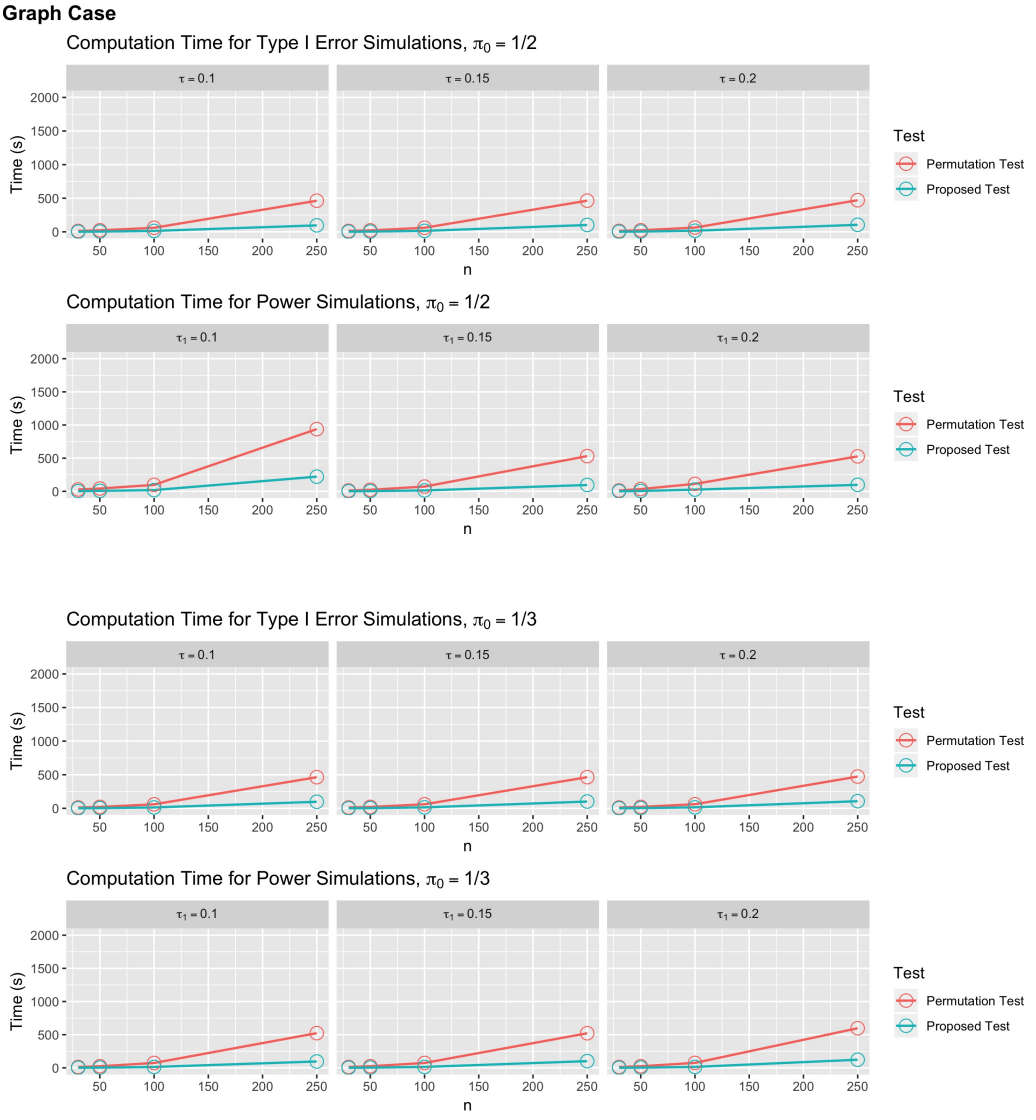

Figure A5: Figures showing the computation time for various settings with graph outcomes in Scenario 2 in seconds. The first and third rows show results from type I error simulations with various noise levels, and the second and fourth rows show power simulations with various effect sizes.

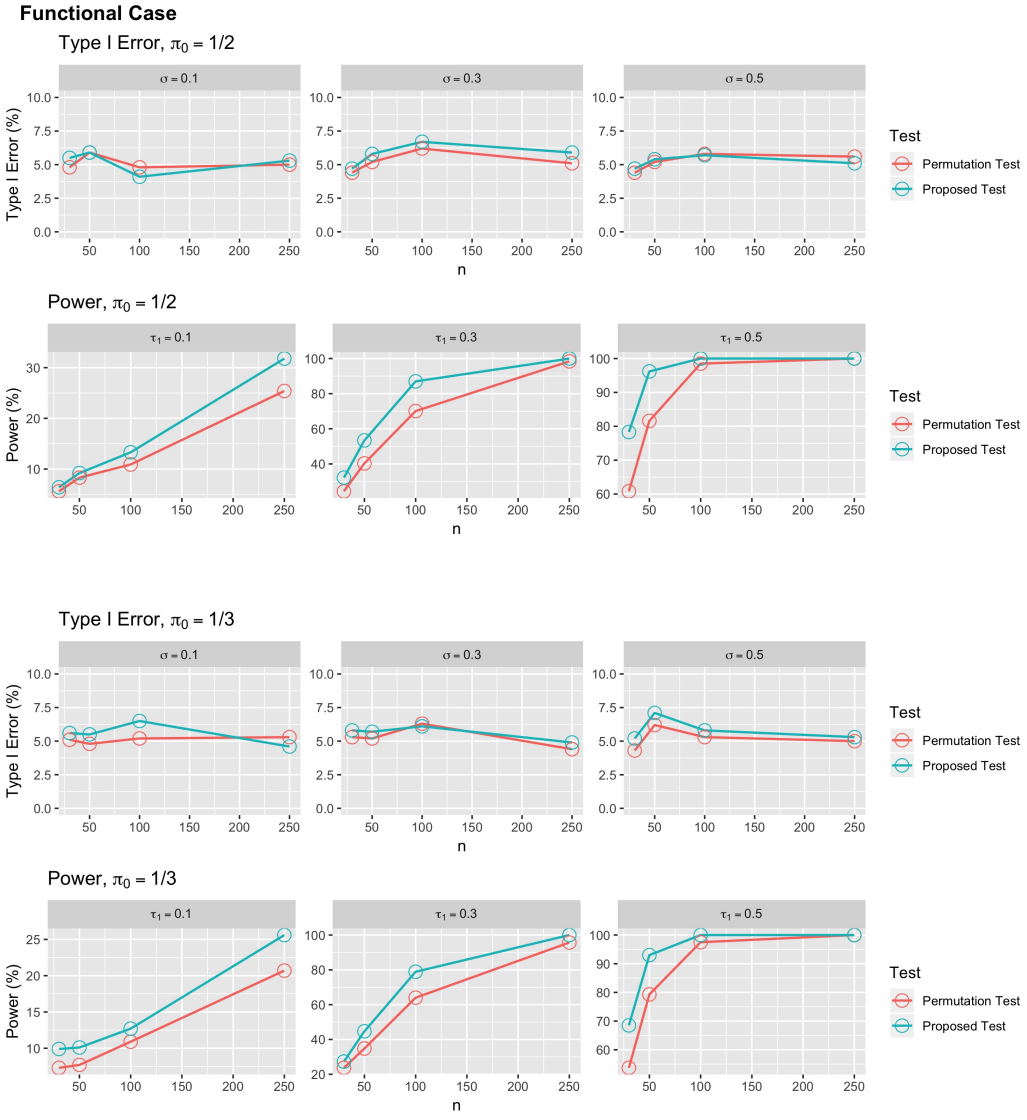

Figure A6: Figures showing the type I error rates and power for various settings with functional outcomes in Scenario 3. The top two rows show type I error rates for several noise levels and power under several alternatives for the case of $\pi_0 = \frac{1}{2}$, and the bottom two rows show results for the case of imbalanced group sizes.

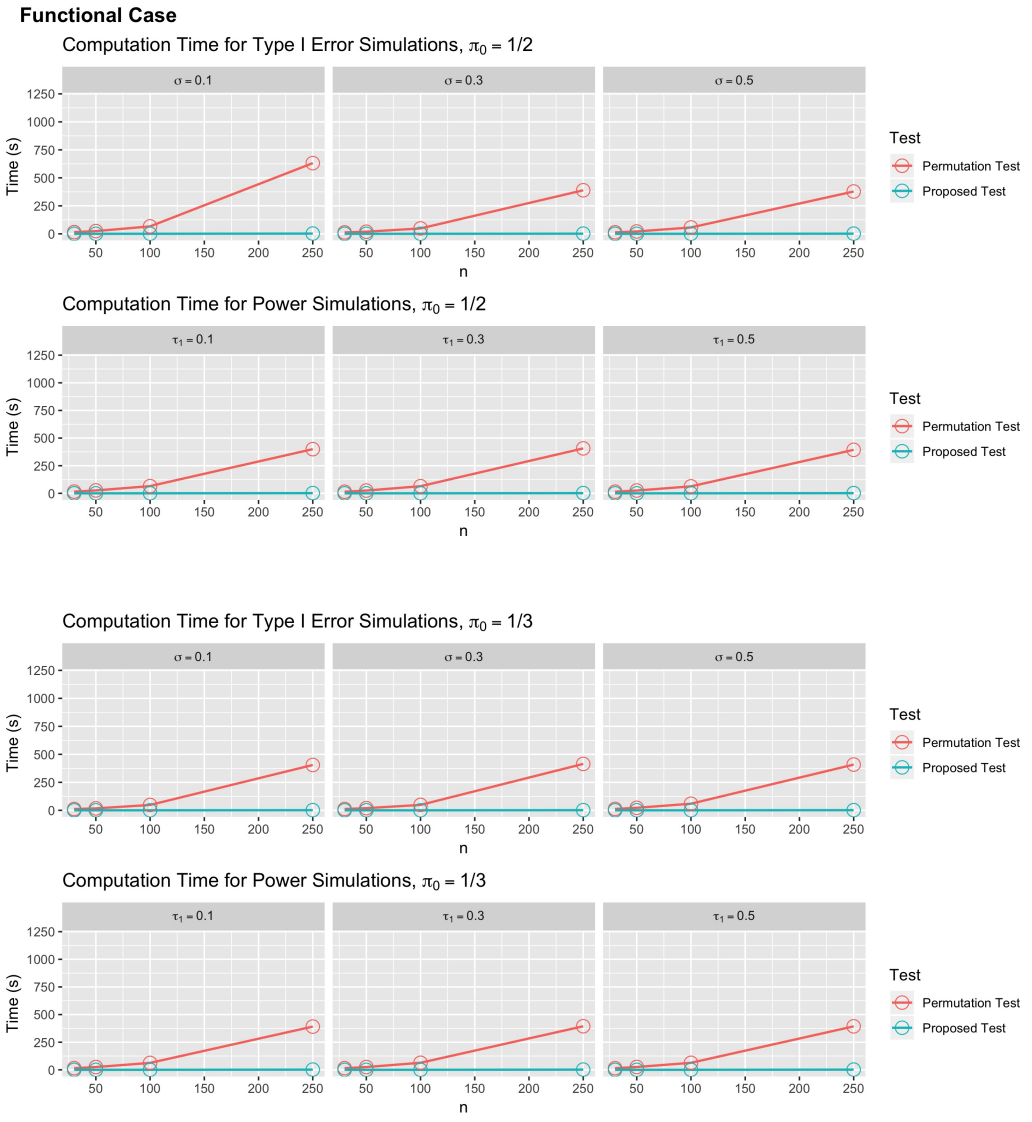

Figure A7: Figures showing the computation time for various settings with functional outcomes in Scenario 3 in seconds. The first and third rows show results from type I error simulations with various noise levels, and the second and fourth rows show power simulations with various effect sizes.

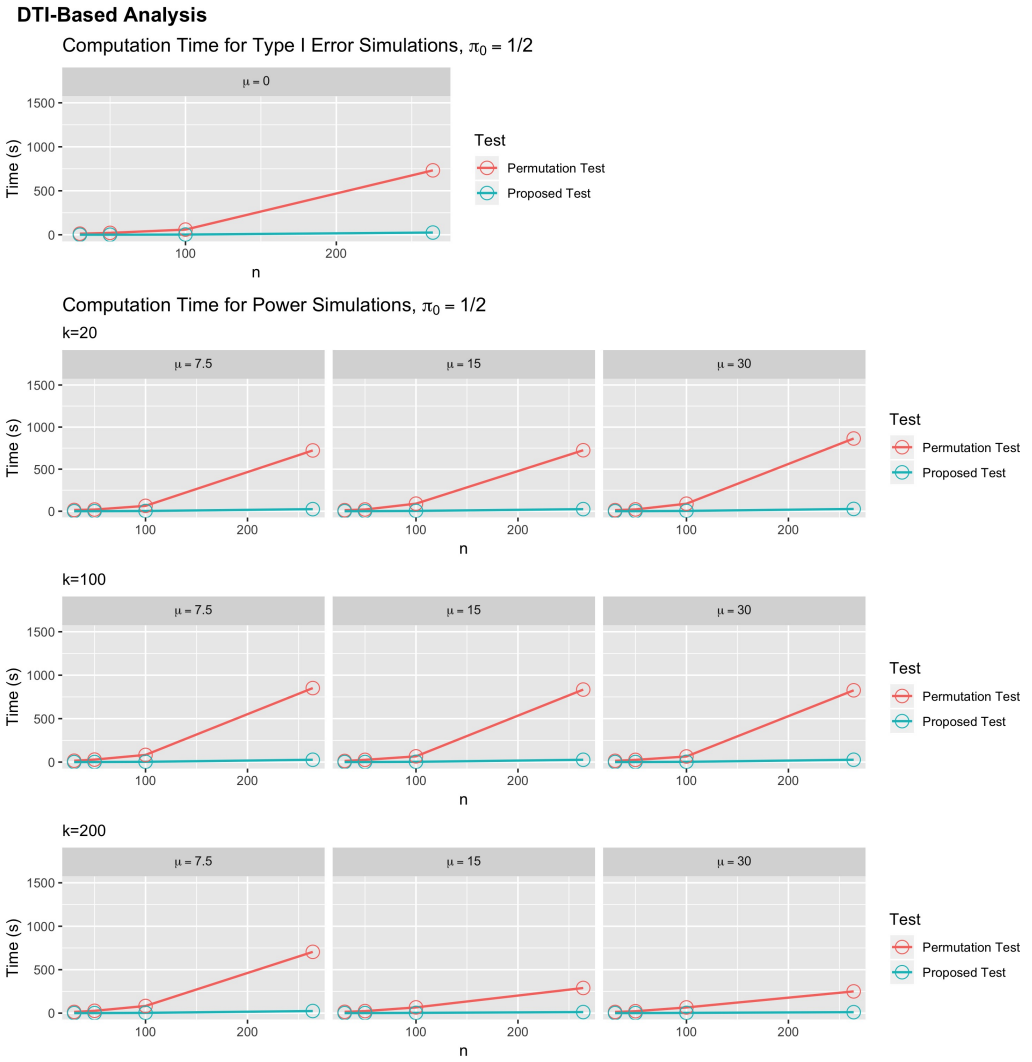

Figure A8: Figures showing the computation time for various settings with the DTI-based network outcomes in Scenario 4 in seconds. The first and third rows show results from type I error simulations with various noise levels, and the second and fourth rows show power simulations with various effect sizes.

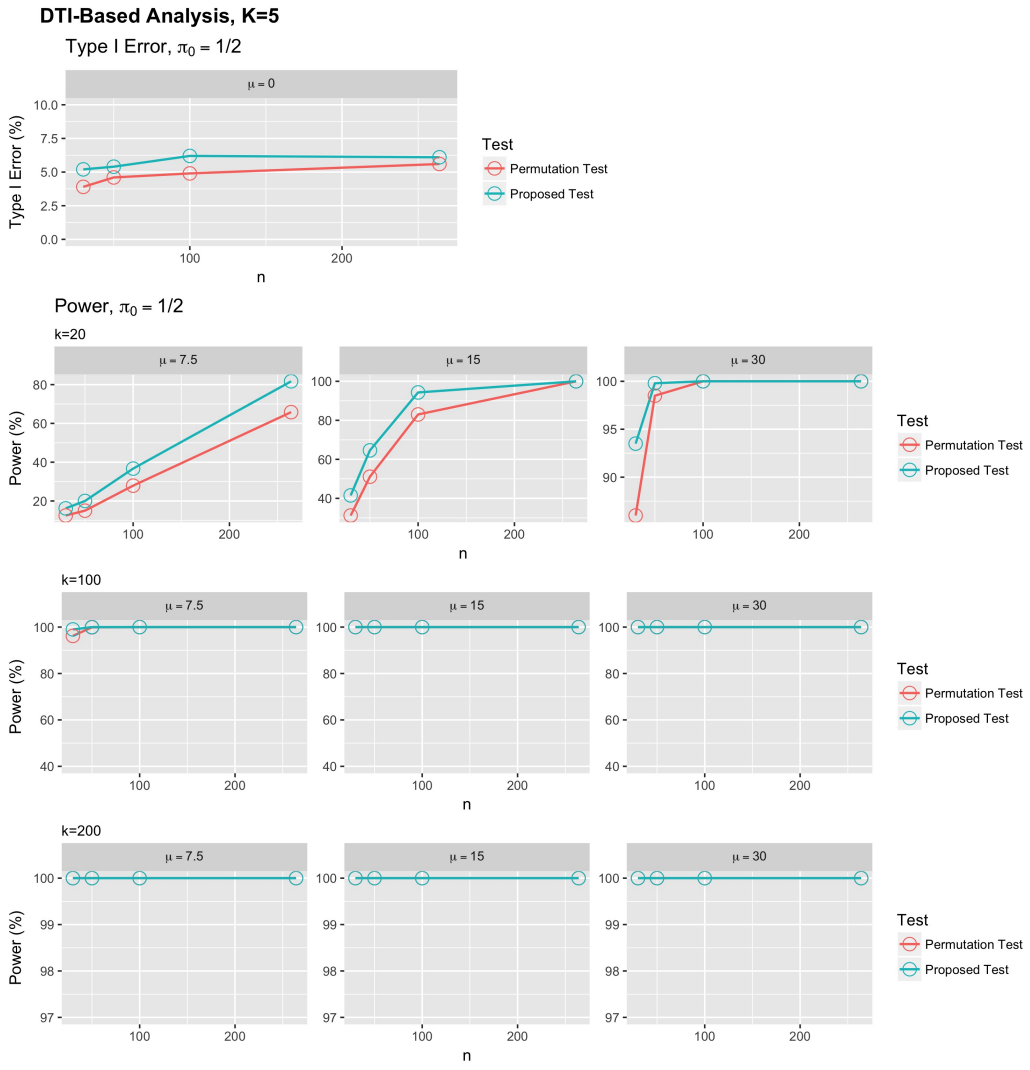

Figure A9: Figures showing the type I error rates and power for various settings with network outcomes in Scenario 4 for $K = 5$. The top row shows type I error rates for several noise levels and power under several alternatives for the case of $\pi_0 = \frac{1}{2}$, and the bottom rows show results for the case of imbalanced group sizes.

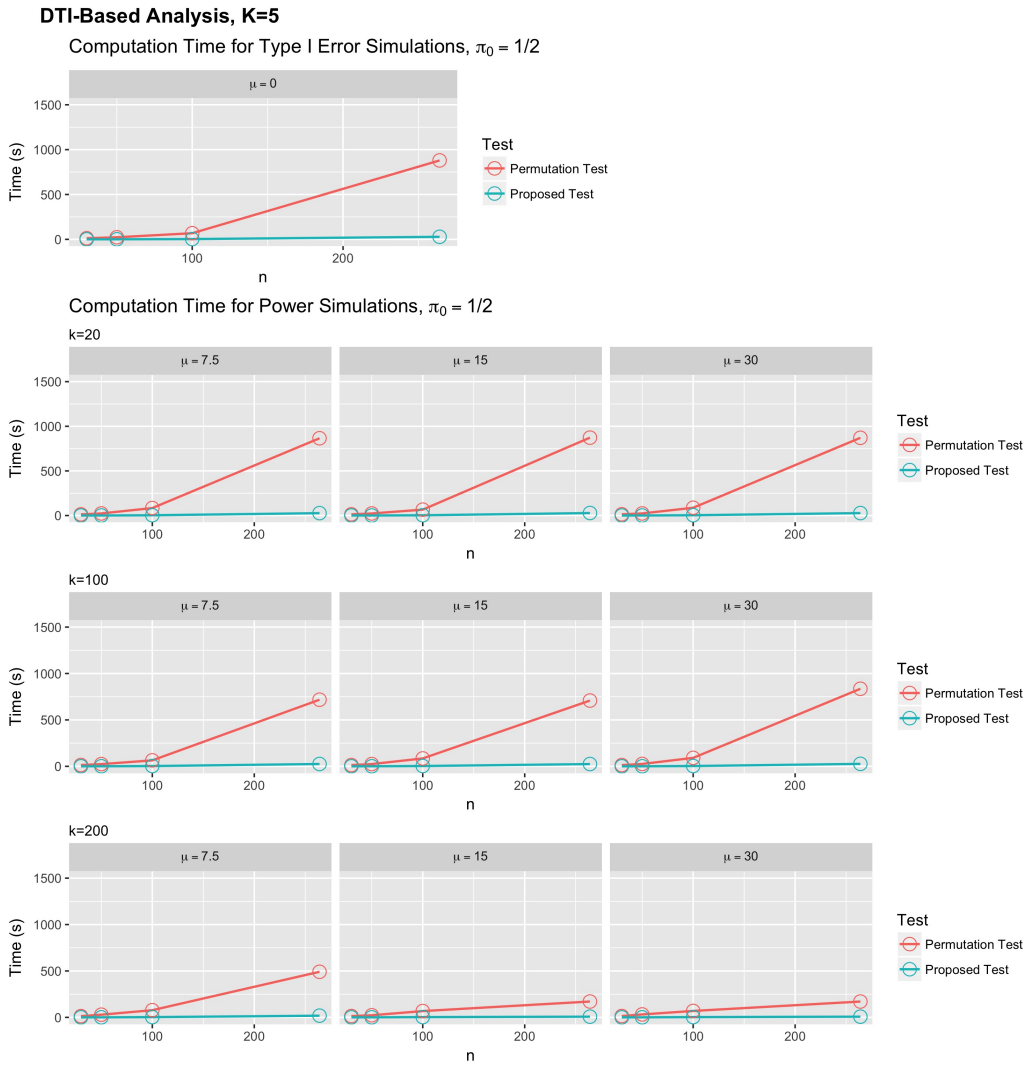

Figure A10: Figures showing the computation time for various settings with the DTI-based network outcomes in Scenario 4 for $K = 5$ in seconds. The first and third rows show results from type I error simulations with various noise levels, and the second and fourth rows show power simulations with various effect sizes.

Technical proofs

We begin by establishing a result on the asymptotic behavior of the product of the estimation error of two asymptotically linear estimators. Specifically, let θ_1 and θ_2 be two parameters, and denote by θ_{10} and θ_{20} their respective true values. Suppose that $\hat{\theta}_{1n}$ and $\hat{\theta}_{2n}$ are asymptotically linear estimators of θ_{10} and θ_{20}, respectively, with influence functions ϕ_{10} and ϕ_{20}.

Lemma. *The mapping $h : (u, v) \mapsto \frac{1}{2}\{\phi_{10}(u)\phi_{20}(v) + \phi_{10}(v)\phi_{20}(u)\}$ is a first-order degenerate kernel, and furthermore,*

$$n(\hat{\theta}_{1n} - \theta_{10})(\hat{\theta}_{2n} - \theta_{20}) = E_0\{\phi_{10}(X)\phi_{20}(X)\} + nU_n + o_P(1) \ ,$$

where U_n is a U-statistic with kernel h. Additionally, it holds that

$$\iiint f(x_1, x_2)h(x_1, x_3)dP_0(x_1)dP_0(x_2)dP_0(x_3) = 0$$

for any function $(x_1, x_2) \mapsto f(x_1, x_2)$ such that $\iint f^2(x_1, x_2)dP_0(x_1)dP_0(x_2) < \infty$.

Proof. In view of asymptotic linearity, we may write

$$\begin{aligned}
(\hat{\theta}_{1n} - \theta_{10})(\hat{\theta}_{2n} - \theta_{20}) &= \left\{\frac{1}{n}\sum_{i=1}^{n}\phi_{10}(X_i) + o_P(n^{-1/2})\right\}\left\{\frac{1}{n}\sum_{i=1}^{n}\phi_{20}(X_i) + o_P(n^{-1/2})\right\} \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\phi_{10}(X_i)\phi_{20}(X_j) + \left\{\frac{1}{n}\sum_{i=1}^{n}\phi_{10}(X_i)\right\}o_P(n^{-1/2}) \\
&\qquad + \left\{\frac{1}{n}\sum_{i=1}^{n}\phi_{20}(X_i)\right\}o_P(n^{-1/2}) + o_P(n^{-1})
\end{aligned}$$

and so, using the fact that both $\frac{1}{n}\sum_{i=1}^{n}\phi_{10}(X_i)$ and $\frac{1}{n}\sum_{i=1}^{n}\phi_{20}(X_i)$ are $O_P(n^{-1/2})$ by the Central Limit Theorem,

$$n(\hat{\theta}_{1n} - \theta_{10})(\hat{\theta}_{2n} - \theta_{20}) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\phi_{10}(X_i)\phi_{20}(X_j) + o_P(1) \ .$$

Furthermore, we can write

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\phi_{10}(X_i)\phi_{20}(X_j) \\
&\qquad = n\left(\frac{n-1}{n}\right)\binom{n}{2}^{-1}\sum_{i<j}h(X_i, X_j) + \frac{1}{n}\sum_{i=1}^{n}\phi_{10}(X_i)\phi_{20}(X_i) + o_P(1) \\
&\qquad = nU_n + E_0\{\phi_{10}(X)\phi_{20}(X)\} + o_P(1) \ .
\end{aligned}$$

Since $E_0\{\phi_{10}(X)\} = E_0\{\phi_{20}(X)\} = 0$, it follows that $E_0\{h(X_i, X_j)\} = 0$ for $i \neq j$ and furthermore that $E_0\{h(X_1, X_2)h(X_1, X_3)\} = 0$.

□

We now establish the validity of our main theorea and their corollary.

Proof of Theorem 1. We prove the theorem for the general case of $K \geq 2$. *(a)* The weak consistency of $SSE_n/(n-K)$ to $\sigma^2(P_{0*})$ is a consequence of the weak law of large numbers for U-statistics. In particular,

$$SSE_n := \sum_{s=0}^{K-1} (n_s - 1) \binom{n_s}{2}^{-1} \sum_{i<j} r(M_i, M_j) I(D_i = D_j = s)$$

and $(n_s - 1)\binom{n_s}{2}^{-1} \sum_{i<j} r(M_i, M_j) I(D_i = D_j = s)$ is a U-statistic with kernel $(x_1, x_2) \mapsto r(m_1, m_2) I(d_1 = d_2 = s)$ where $x_1 := (m_1, d_1)$ and $x_2 := (m_2, d_2)$ represent two arbitrary realizations of X, and $E_0\{r(M_i, M_j) I(D_i = D_j = s)\} = \sigma^2(P_{0s})\pi_s^2$. Thus, since $(n_s - 1)/(n-K) \xrightarrow{P} \pi_s$ as $n \to \infty$, we have that

$$SSE_n/(n-K) \xrightarrow{P} \sum_{d=0}^{K-1} \pi_s \sigma^2(P_{0s}) = \sigma^2(P_{0*}),$$

under the null hypothesis. *(b)* Using the fact that $SST_n = SS_n - SSE_n$, we observe that we may represent SST_n as

$$\begin{aligned}
&(n-1)\binom{n}{2}^{-1} \sum_{j<k} r(M_j, M_k) - \sum_{s=0}^{K-1} (n_s - 1)\binom{n_s}{2}^{-1} \sum_{j<k} r(M_j, M_k) I(D_j = D_k = s) \\
&= \frac{2}{n} \sum_{j<k} r(M_j, M_k) \left\{ 1 - \sum_{s=0}^{K-1} \frac{I(D_j = D_k = s)}{\pi_s} \right\} \\
&\quad - \frac{2}{n} \sum_{j<k} r(M_j, M_k) \left[\sum_{s=0}^{K-1} \left(\frac{n}{n_s} - \frac{1}{\pi_s} \right) I(D_j = D_k = s) \right] \\
&= A_{1n} - A_{2n} - A_{3n} + o_P(1) \ ,
\end{aligned}$$

where we have defined the summands

$$\begin{aligned}
A_{1n} &:= n \cdot \binom{n}{2}^{-1} \sum_{j<k} r(M_j, M_k) \left\{ 1 - \sum_{s=0}^{K-1} \frac{I(D_j = D_k = s)}{\pi_s} \right\} \\
A_{2n} &:= n \cdot \sum_{s=0}^{K-1} \left(\frac{n}{n_s} - \frac{1}{\pi_s} \right) \binom{n}{2}^{-1} \sum_{j<k} \left\{ r(M_j, M_k) I(D_j = D_k = s) - \pi_s^2 \sigma^2(P_{0*}) \right\} \\
A_{3n} &:= n \cdot \sum_{s=0}^{K-1} \left(\frac{n}{n_s} - \frac{1}{\pi_s} \right) \pi_s^2 \sigma^2(P_{0*}) \ .
\end{aligned}$$

We now analyze each of the above three terms under the null hypothesis. The term A_{1n} is an n-scaled U-statistic with first-order degenerate kernel h_1 given by

$$(x_1, x_2) \mapsto r(m_1, m_2) \left\{ 1 - \sum_{s=0}^{K-1} \frac{I(d_1 = d_2 = s)}{\pi_s} \right\}.$$

To study the term A_{2n}, we first note that

$$\binom{n}{2}^{-1} \sum_{j<k} \left\{ r(M_j, M_k) I(D_j = D_k = s) - \pi_s^2 \sigma^2(P_{0*}) \right\}$$

is a non-degenerate U-statistic with kernel $(x_1, x_2) \mapsto r(m_1, m_2) I(d_1 = d_2 = s) - \pi_s^2 \sigma^2(P_{0*})$, and by the Hájek projection, it is asymptotically linear with influence function

$$\begin{aligned} x \;\mapsto\; & 2 \left[E_0 \left\{ r(M, m) I(D = d = s) \right\} - \pi_s^2 \sigma^2(P_{0*}) \right] \\ = \; & 2\pi_s \left[I(d = s) E_0 \left\{ r(M, m) \right\} - \pi_s \sigma^2(P_{0*}) \right] \;=:\; \phi_{21s}(x) \;, \end{aligned}$$

where $E_0 \{ r(M, m) \}$ is the average distance from m to a random draw from the distribution of M. By the delta method, we have that

$$\frac{n}{n_s} - \frac{1}{\pi_s} = \pi_s^{-2} \left\{ \frac{n - n_s}{n} - (1 - \pi_s) \right\} + o_P(n^{-1/2}) = \frac{1}{n} \sum_{i=1}^{n} \phi_{22s}(X_i) + o_P(n^{-1/2})$$

with $\phi_{22s}(x) := \{ I(d \neq s) - (1 - \pi_s) \} / \pi_s^2$. By the Lemma, we can then write $A_{2n} = \sum_{s=0}^{K-1} B_{2s} + n U_{2ns} + o_P(1)$, where $B_{2s} := E_0 \{ \phi_{21s}(X) \phi_{22s}(X) \} = -2(1 - \pi_s) \sigma^2(P_{0*})$ and U_{2ns} is a first-order degenerate U-statistic with kernel

$$h_2 : (x_1, x_2) \mapsto \sum_{s=0}^{K-1} \left\{ \phi_{21s}(x_1) \phi_{22s}(x_2) + \phi_{21s}(x_2) \phi_{22s}(x_1) \right\} / 2 \;.$$

To study the term A_{3n}, we note that by a second-order approximation we have

$$\frac{n}{n_s} - \frac{1}{\pi_s} = \pi_s^{-2} \left\{ \frac{n - n_s}{n} - (1 - \pi_s) \right\} + \pi_s^{-3} \left(\frac{n_s}{n} - \pi_s \right)^2 + o_P(n^{-1}) \;.$$

Thus, we readily find that

$$A_{3n} = n \sum_{s=0}^{K-1} \left(\frac{n_s}{n} - \pi_s \right)^2 \pi_s^{-1} \sigma^2(P_{0*}) + o_P(1)$$

for all $s = 1, \ldots, K - 1$, and hence $A_{3n} = B_3 + n U_{3n} + o_P(1)$, where

$$B_3 := (K - 1) \cdot \sigma^2(P_{0*})$$

and U_{3n} is a first-order degenerate U-statistic with kernel

$$h_3 : (x_1, x_2) \mapsto \sum_{s=0}^{K-1} [\sigma^2(P_{0*}) / \pi_s] \{ I(d_1 = s) - \pi_s \} \{ I(d_2 = s) - \pi_s \} \;.$$

In view of the above facts, we obtain that

$$SST_n = (K - 1) \cdot \sigma^2(P_{0*}) + n \cdot \binom{n}{2}^{-1} \sum_{j<k} u(X_j, X_k) + o_P(1) \;,$$

where u is the first-order degenerate kernel $(x_1, x_2) \mapsto h_1(x_1, x_2) - h_2(x_1, x_2) - h_3(x_1, x_2)$. It is easy to verify that $Cov\{f(X_1, X_2), h_j(X_1, X_3)\} = 0$ for any $f : \mathbb{R}^2 \to \mathbb{R}$, and so, in particular, we find that $Cov\{u(X_1, X_2), u(X_1, X_3)\} = 0$. It follows then that SST_n tends in distribution to $(K-1) \cdot \sigma^2(P_{0*}) + \sum_{j=1}^{\infty} \lambda_j (Z_j^2 - 1)$ as $n \to \infty$, where $Z_1, Z_2, \ldots$ are independent standard normal random variables and $\lambda_1, \lambda_2, \ldots$ are eigenvalues of the operator $g \mapsto \Gamma(g) : m \mapsto \int u(m, m_2) g(m_2) dP_{00}(m_2)$ (Lee 1990).

Proof of Corollary. Follows from the theorem by an application of Slutsky's theorem.

Proof of Theorem 2. Under the alternative, by the law of large numbers for U-statistics, we see that $T(P_0) > 0$. Since $T_n = T(P_0) + o_P(1)$, the power of the test, $P_0(Q_n > q_\alpha) = P_0(T_n > q_\alpha/n)$, tends to one.

References

Lee, J. (1990), 'U-statistics: Theory and practice'.