Online Supplement to "Reducing the Conservativeness of Fully Sequential Indifference-Zone Procedures"

Huizhu Wang
Seong-Hee Kim

H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology
Atlanta, GA 30332-0205 U.S.A.

August 24, 2012

1 Validity Proofs of Procedures for Known Variances

Two lemmas are needed to prove the validity of a number of procedures presented in the paper.

Lemma 1 *Jennison et al. [1982] Suppose that a continuation region $H(t)$ is $(-h(t), h(t))$ given by a non-negative function $h(t), t \geq 0$. Consider two processes: a continuous process $\{\mathcal{W}(t, \Delta), t \geq 0\}$ with $\Delta > 0$, and a discrete process obtained by observing $\mathcal{W}(t, \Delta)$ at a random, increasing sequence of times $\{t_i : i = 1, 2, \ldots\}$ taking values in a given countable set. Let $T_C = \inf\{t : \mathcal{W}(t, \Delta) \notin H(t)\}$ and $T_D = \inf\{t_i : \mathcal{W}(t_i, \Delta) \notin H(t_i)\}$, and assume that $T_D < \infty$ almost surely. Note that $T_D \geq T_C$. The error probabilities are*

$$\Pr\{\mathcal{E}_C\} \equiv \Pr\{\mathcal{W}(T_C, \Delta) \leq -h(T_C)\} = \Pr\{\mathcal{W}(T_C, \Delta) < 0\},$$

$$\Pr\{\mathcal{E}_D\} \equiv \Pr\{\mathcal{W}(T_D, \Delta) \leq -h(T_D)\} = \Pr\{\mathcal{W}(T_D, \Delta) < 0\}.$$

Consider an outcome $\{(b(t); t \geq 0), \{t_i\}\}$, where $b(t)$ is the path of a Brownian motion. Assume that the conditional distribution of $\{t_i\}$ given $\mathcal{W}(t, \Delta) = b(t), t \geq 0$, is the same as the conditional distribution of $\{t_i\}$ given $\mathcal{W}(t, \Delta) = -b(t)$, $t \geq 0$. Under these conditions,

$$\Pr\{\mathcal{E}_D\} \leq \Pr\{\mathcal{E}_C\}.$$

Lemma 2 *Fabian [1974] Consider a standard Brownian motion process with drift $\mathcal{W}(t, \Delta)$ with $\Delta > 0$ and $t \geq 0$. Let $h(t) = a - \gamma t$ for some $a > 0$ and $\gamma \geq 0$. Let $H(t)$ denote the interval $(-h(t), h(t))$ (so that $H(t) = \emptyset$ when $-h(t) \geq h(t)$), and let $T = \inf\{t : \mathcal{W}(t, \Delta) \notin H(t)\}$ be the first time $\mathcal{W}(t, \Delta)$ does not fall in the triangular continuation region defined by $(H(t); t \geq 0)$. Finally, let $\mathcal{E}$ be the event $\{\mathcal{W}(T, \Delta) \leq -h(T)$ and $H(T) \neq \emptyset$, or $\mathcal{W}(T, \Delta) \leq 0$ and $H(T) = \emptyset\}$. If $\gamma = \Delta/(2e)$ for any positive integer e, then*

$$\Pr\{\mathcal{E}\} = \sum_{v=1}^{e} (-1)^{v+1} \left(1 - \frac{1}{2}\mathcal{I}(v = e)\right) \exp\{-2a\gamma(2e - v)v\}.$$

The proof of statistical validity of Procedure ($\mathcal{P}$) is as follows.

Let an incorrect selection (ICS) denote the event that the best system (system 1) is eliminated. This event implies that $\sum_{j=1}^{r}(X_{1j} - X_{ij})$ exits through the lower boundary of the continuation region.

$$\Pr\{\text{ICS}\} = \Pr\left\{\sum_{j=1}^{T_i^d}(X_{1j} - X_{ij}) \leq -R(T_i^d; \delta, \eta, \sigma_{1i}^2) \text{ for some } i \neq 1\right\} \quad \text{(S.1)}$$

$$\overset{\text{BI}}{\leq} \sum_{i=2}^{k} \Pr\left\{\sum_{j=1}^{T_i^d}(X_{1j} - X_{ij}) \leq -R(T_i^d; \delta, \eta, \sigma_{1i}^2)\right\} \quad \text{(S.2)}$$

$$= \sum_{i=2}^{k} \Pr\left\{\sum_{j=1}^{T_i^d}\left(\frac{X_{1j} - X_{ij}}{\sigma_{1i}}\right) \leq -\max\left\{0, \left(\frac{\eta\sigma_{1i}}{\delta} - \frac{\delta}{2e\sigma_{1i}}T_i^d\right)\right\}\right\}$$

$$\overset{\text{DTC}}{\leq} \sum_{i=2}^{k} \Pr\left\{\mathcal{W}\left(T_i^c; \frac{\mu_1 - \mu_i}{\sigma_{1i}}\right) \leq -\max\left\{0, \left(\frac{\eta\sigma_{1i}}{\delta} - \frac{\delta}{2e\sigma_{1i}}T_i^c\right)\right\}\right\} \quad \text{(S.3)}$$

$$\overset{\text{SC}}{\leq} \sum_{i=2}^{k} \Pr\left\{\mathcal{W}\left(T_i^c; \frac{\delta}{\sigma_{1i}}\right) \leq -\max\left\{0, \left(\frac{\eta\sigma_{1i}}{\delta} - \frac{\delta}{2e\sigma_{1i}}T_i^c\right)\right\}\right\} \quad \text{(S.4)}$$

$$= \sum_{i=2}^{k}\left[\sum_{v=1}^{e}(-1)^{v+1}\left(1 - \frac{1}{2}\mathcal{I}(v = e)\right)\exp\left\{-\frac{\eta(2e-v)v}{e}\right\}\right] \quad \text{(S.5)}$$

$$= \sum_{i=2}^{k} g(\eta) = \alpha,$$

where the first inequality (S.2) is due to the BI; the second inequality (S.3) employs Lemma 1, denoted as DTC; the third inequality (S.4) holds because the smallest mean difference between the best system and any inferior system i is δ under the SC assumption; and Equation (S.5) is due to Lemma 2. Since we choose η such that $g(\eta) = \frac{\alpha}{k-1}$, the last equality holds.

Now consider Procedures ($\mathcal{E}$), ($\mathcal{BI}$) and ($\mathcal{BI} + \mathcal{DTC}$). Procedure ($\mathcal{E}$) uses the right-hand side (RHS) of Equation (S.1) to determine η, and one can see that Procedure ($\mathcal{E}$) is an exact procedure in the sense that it guarantees the actual PCS to be exactly equal to the nominal level. Procedure ($\mathcal{BI}$) employs the RHS of Inequality (S.2), therefore Procedure ($\mathcal{BI}$) is conservative due to the use of the BI. Unfortunately, there is no closed-form solution η for Procedures ($\mathcal{E}$) and ($\mathcal{BI}$), so we empirically search for η. Procedure ($\mathcal{BI} + \mathcal{DTC}$) uses the RHS of Inequality (S.3) to solve η, which implies that Procedure ($\mathcal{BI} + \mathcal{DTC}$) needs the BI and DTC for its validity proof. Note that Lemma 2 is applicable to Procedure ($\mathcal{BI} + \mathcal{DTC}$), which results in the same η as in Procedure ($\mathcal{P}$). Procedure ($\mathcal{P}$) is the most conservative procedure, because it employs all three conservativeness sources, and therefore it can also be considered as Procedure ($\mathcal{BI} + \mathcal{DTC} + \mathcal{SC}$).

2 Impacts of the BI, DTC, and the SC Assumption for Correlated Systems

Percentages of average additional number of observations for correlated systems with $\rho = 0.6$ are given in Figures S1 and S2.

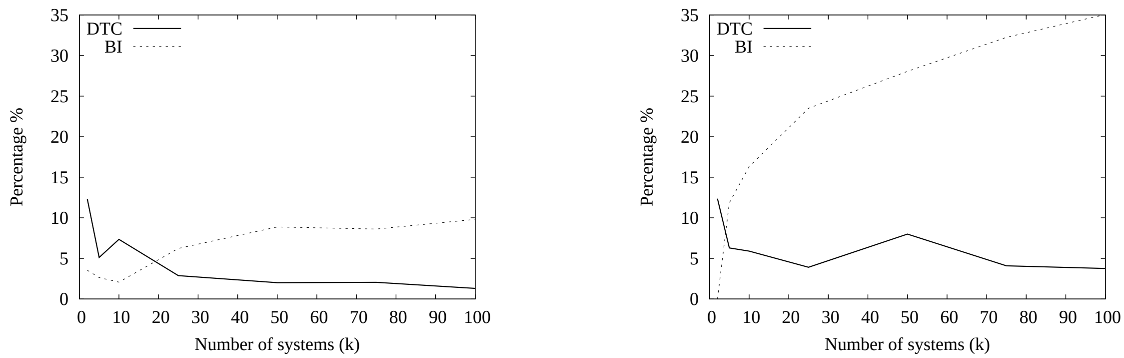

Figure S1: Percentages of average additional number of observations under the SC with INC (left) and DEC (right) variances when $\rho = 0.6$.

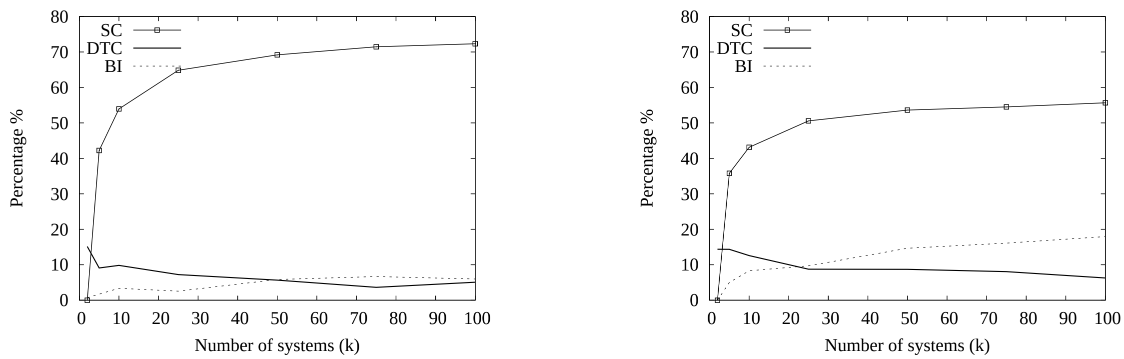

Figure S2: Percentages of average additional number of observations under the DC with INC (left) and DEC (right) variances when $\rho = 0.6$.

3 Derivation of Asymptotic Ratio $\mathcal{S}$

We employ a similar heuristic argument from Perng [1969] to approximate the number of observations for Procedures $(\mathcal{P})$ and $(\mathcal{BI} + \mathcal{DTC})$. The following happens for Procedures $(\mathcal{P})$ and $(\mathcal{BI} + \mathcal{DTC})$ as $\alpha \to 0$: (i) procedures stop making mistakes and system 1 is always found superior to other systems; and (ii) a larger number of observations is needed to make a decision, and hence sample mean $\bar{X}_i(r)$ behaves like the true mean μ_i for system i.

Let N_i^P and N_i^B be the number of observations that system i takes until a decision is reached in Procedures $(\mathcal{P})$ and $(\mathcal{BI} + \mathcal{DTC})$, respectively. For Procedure $(\mathcal{P})$, system i is eliminated when

$$\mu_1 - \mu_i \geq \frac{1}{r} R(r; \delta, \eta, \sigma_{1i}^2) = \frac{\eta \sigma_{1i}^2}{r\delta} - \frac{\delta}{2} \quad \text{and} \quad N_i^P \approx \frac{\eta \sigma_{1i}^2}{\delta(\mu_1 - \mu_i + \delta/2)}, \text{ for } i = 2, \ldots, k.$$

System 1 keeps receiving additional observations until all inferior systems are eliminated and thus takes

$$N_1^P = \max(N_2^P, \ldots, N_k^P).$$

Similarly, for Procedure $(\mathcal{BI} + \mathcal{DTC})$,

$$\mu_1 - \mu_i \geq \frac{1}{r} R(r; \delta_{1i}, \eta, \sigma_{1i}^2) = \frac{\eta \sigma_{1i}^2}{r\delta_{1i}} - \frac{\delta_{1i}}{2} \quad \text{and} \quad N_i^B \approx \frac{\eta \sigma_{1i}^2}{\delta_{1i}(\mu_1 - \mu_i + \delta_{1i}/2)}, \text{ for } i = 2, \ldots, k.$$

Also let $N_1^B = \max(N_2^B, \ldots, N_k^B)$.

Then, the asymptotic ratio of savings between the total number of observations for Procedures $(\mathcal{BI}+\mathcal{DTC})$ and $(\mathcal{P})$ is

$$\mathcal{S} \equiv 1 - \frac{\sum_{i=1}^{k} N_i^B}{\sum_{i=1}^{k} N_i^P} = 1 - \frac{\sum_{i=2}^{k} \frac{\eta\sigma_{1i}^2}{3(\mu_1-\mu_i)^2/2} + \max_{i\in\{2,\ldots,k\}} N_i^B}{\sum_{i=2}^{k} \frac{\eta\sigma_{1i}^2}{\delta(\mu_1-\mu_i+\delta/2)} + \max_{i\in\{2,\ldots,k\}} N_i^P}$$

The following table shows the asymptotic ratios $\mathcal{S}$ and empirical ratios with $\alpha = 0.01$ after 500 macro replications when the mean configuration follows the DC and $\rho = 0$. The table demonstrates that our theoretical asymptotic ratios match well empirical ratios. The table shows that Procedure $(\mathcal{BI}+\mathcal{DTC})$ spends 23% to 41% fewer observations than Procedure $(\mathcal{P})$.

Variances	Ratio	$k=5$	$k=10$	$k=25$	$k=50$	$k=75$	$k=100$
INC	$\mathcal{S}$	0.277	0.345	0.391	0.406	0.412	0.415
	Empirical	0.273	0.334	0.375	0.393	0.396	0.406
Equal	$\mathcal{S}$	0.259	0.324	0.368	0.384	0.390	0.392
	Empirical	0.256	0.311	0.351	0.365	0.369	0.375
DEC	$\mathcal{S}$	0.247	0.310	0.354	0.370	0.375	0.384
	Empirical	0.233	0.276	0.319	0.323	0.327	0.378

4 Proof of Theorem 1

We need the following lemmas for the proof of Theorem 1.

Lemma 3 *Define the standardized partial sum for system i as*

$$C_i(t,r) \equiv \frac{\sum_{j=1}^{\lfloor rt\rfloor} X_{ij} - rt\mu_i}{\sigma_i\sqrt{r}}, \quad 0 \le t \le 1 \text{ and } r \in \mathbb{Z}^+, \tag{S.6}$$

where $\lfloor\cdot\rfloor$ indicates truncation of any fractional part. The probability distribution of $C_i(t,r)$ over $D[0,1]$, the space of functions that are right-continuous and have left-hand limits, converges to that of a standard Brownian motion process, $\mathcal{W}(t)$, as r increases; i.e., $C_i(\cdot,r) \Rightarrow \mathcal{W}(\cdot)$ as $r \to \infty$, where $\Rightarrow$ denotes convergence in distribution. Further, we assume that for any $t \in [0,1]$, the family of random variables $\{C_i^2(t,r) : r = 1,2,\ldots\}$ is uniformly integrable.

Lemma 3 is basically Functional Central Limit Theorem (Billingsley [1968]). It is straightforward that Lemma 3 holds under Assumption 1.

Lemma 4 *If Assumptions 1, 2 and 3 hold, then there exists $N \in \mathbb{Z}^+$ such that $\delta \le \hat{\delta}_{1i}(r) < (c_i+\kappa)\delta$ with probability 1 (w.p.1), for any $r > N$ and some $\kappa \in \mathbb{R}^+$.*

Proof: Under Assumptions 1, 2 and 3, $|\bar{X}_1(r) - \bar{X}_i(r)| - A_{1i}(r) \to c_i\delta$ w.p.1 as $r \to \infty$ for $i \in \{2,\ldots,k\}$. This follows by the strong law of large numbers and continuous mapping theorem (CMT, Billingsley [1968]). Set $\epsilon = \kappa\delta$ for some $\kappa \in \mathbb{R}^+$. Then there exists some $N \in \mathbb{Z}^+$, $(c_i-\kappa)\delta < |\bar{X}_1(r) - \bar{X}_i(r)| - A_{1i}(r) < (c_i+\kappa)\delta$ and therefore $\delta \le \hat{\delta}_{1i}(r) < (c_i+\kappa)\delta$ for $r > N$ w.p.1. □

Proof of Theorem 1:

Let

$$T_{1i} \equiv T_{1i}(\delta) = \min\left\{ r : r \in \{n_0, n_0+1, \ldots\} \text{ and } \left|\sum_{j=1}^{r}(X_{1j} - X_{ij})\right| \geq R(r; \hat{\delta}_{1i}(r), h^2, S_{1i}^2(n_0)) \right\}$$

and

$$N_{1i}(r) = \left\lfloor \frac{2eh^2 S_{1i}^2(n_0)}{\hat{\delta}_{1i}^2(r)} \right\rfloor.$$

First of all, we will show that $T_{1i} \to \infty$ as $\delta \to 0$ w.p.1, which is a fundamental requirement for the convergence of estimators. We first represent the output process $\{X_{1j} - X_{ij}\}$ as $\{Z_j + c_i\delta\}$, where $Z_j \overset{\text{IID}}{\sim} N(0, \sigma_{1i}^2)$, $i = 2, \ldots, k$ and $j = 1, 2, \ldots$. Consider a sample path going up for which $\limsup_{\delta \to 0} T_{1i} = T^* \leq \infty$. (When the path goes down, the proof is analogous.) There must exist $\delta^* > 0$ such that for all $\delta \leq \delta^*$,

$$\sum_{j=1}^{T^*} Z_j + T^* c_i \delta \geq \frac{h^2 S_{1i}^2(n_0)}{\hat{\delta}_{1i}(T^*)} - \frac{\hat{\delta}_{1i}(T^*)}{2e} T^*. \tag{S.7}$$

By Lemma 4, there must exist some $\kappa \in \mathbb{R}^+$ and $n < T^*$ that ensure $\delta \leq \hat{\delta}_{1i}(r) < (c_i + \kappa)\delta$ for any $r > n$ w.p.1. This implies that $\delta \leq \hat{\delta}_{1i}(T^*) < (c_i + \kappa)\delta$ and we obtain

$$\sum_{j=1}^{T^*} Z_j + T^* c_i \delta \geq \frac{h^2 S_{1i}^2(n_0)}{(c_i + \kappa)\delta} - \frac{(c_i + \kappa)\delta}{2e} T^*. \tag{S.8}$$

Equation (S.8) can be rewritten as follows:

$$\sum_{j=1}^{T^*} Z_j + T^* \delta (c_i + \frac{c_i + \kappa}{2e}) \geq \frac{h^2 S_{1i}^2(n_0)}{(c_i + \kappa)\delta}.$$

Note that when T^* is finite, $\sum_{j=1}^{T^*} Z_j$ is finite w.p.1, but $T^*\delta$ goes to zero while $\frac{h^2 S_{1i}^2(n_0)}{(c_i+\kappa)\delta}$ goes to infinity as $\delta \to 0$. Thus (S.7) occurs with probability zero and then $T_{1i} \to \infty$ w.p.1 as $\delta \to 0$.

This also ensures that as $\delta \to 0$,

$$\hat{\delta}_{1i}(T_{1i})/\delta_{1i} \to 1 \quad w.p.1. \tag{S.9}$$

Due to (S.9) and CMT, it is also true that $N_{1i}(T_{1i}) \to \infty$ as $\delta \to 0$. Let ICS_i denote the event that the best system gets eliminated by an inferior system i, $i = 2, \ldots, k$. Then

$$\begin{aligned}
\Pr\{\text{ICS}_i\} &= \Pr\left\{ \sum_{j=1}^{T_{1i}} (X_{1j} - X_{ij}) \leq -\max\left\{0, \frac{h^2 S_{1i}^2(n_0)}{\hat{\delta}_{1i}(T_{1i})} - \frac{\hat{\delta}_{1i}(T_{1i})}{2e} T_{1i} \right\} \right\} \\
&= \Pr\left\{ \frac{\sum_{j=1}^{T_{1i}} (X_{1j} - X_{ij}) - \delta_{1i} T_{1i}}{\sigma_{1i}\sqrt{N_{1i}(T_{1i}) + 1}} + \frac{\delta_{1i} T_{1i}}{\sigma_{1i}\sqrt{N_{1i}(T_{1i}) + 1}} \leq \right. \\
&\quad \left. -\max\left\{0, \frac{h^2 S_{1i}^2(n_0)}{\hat{\delta}_{1i}(T_{1i})\sigma_{1i}\sqrt{N_{1i}(T_{1i}) + 1}} - \frac{\hat{\delta}_{1i}(T_{1i})}{2e\sigma_{1i}\sqrt{N_{1i}(T_{1i}) + 1}} T_{1i} \right\} \right\}
\end{aligned} \tag{S.10}$$

Let

$$C_{1i}(t,\delta) = \frac{\sum_{j=1}^{\lfloor (N_{1i}(T_{1i})+1)t \rfloor}(X_{1j} - X_{ij}) - (N_{1i}(T_{1i})+1)\delta_{1i}t}{\sigma_{1i}\sqrt{N_{1i}(T_{1i})+1}}$$

for $0 \le t \le 1$. Further, define

$$\begin{aligned}\widehat{T}_{1i} = \min\Bigg\{ t \in \left\{ \frac{n_0}{N_{1i}(T_{1i})+1}, \frac{n_0+1}{N_{1i}(T_{1i})+1}, \ldots, 1 \right\} : \left| C_{1i}(t,\delta) + \frac{(N_{1i}(T_{1i})+1)\delta_{1i}t}{\sigma_{1i}\sqrt{N_{1i}(T_{1i})+1}} \right| \\ \ge \frac{h^2 S_{1i}^2(n_0)}{\hat{\delta}_{1i}(T_{1i})\sigma_{1i}\sqrt{N_{1i}(T_{1i})+1}} - \frac{\hat{\delta}_{1i}(T_{1i})(N_{1i}(T_{1i})+1)}{2e\sigma_{1i}\sqrt{N_{1i}(T_{1i})+1}} t \Bigg\}.\end{aligned}$$

Clearly, $\widehat{T}_{1i} = T_{1i}/(N_{1i}(T_{1i})+1)$. Also define the stopping time of the corresponding continuous-time process as

$$\begin{aligned}\tilde{T}_{1i} = \min\Bigg\{ t \ge \frac{n_0}{N_{1i}(T_{1i})+1} : \left| C_{1i}(t,\delta) + \frac{(N_{1i}(T_{1i})+1)\delta_{1i}t}{\sigma_{1i}\sqrt{N_{1i}(T_{1i})+1}} \right| \\ \ge \frac{h^2 S_{1i}^2(n_0)}{\hat{\delta}_{1i}(T_{1i})\sigma_{1i}\sqrt{N_{1i}(T_{1i})+1}} - \frac{\hat{\delta}_{1i}(T_{1i})(N_{1i}(T_{1i})+1)}{2e\sigma_{1i}\sqrt{N_{1i}(T_{1i})+1}} t \Bigg\}.\end{aligned}$$

Notice that for fixed δ, $C_{1i}(\widehat{T}_{1i},\delta)$ corresponds to the right-hand limit of a point of discontinuity of $C_{1i}(\cdot,\delta)$. We can show that $\widehat{T}_{1i} \to \tilde{T}_{1i}$ w.p.1 as $\delta \to 0$, making use of the fact that $1/(N_{1i}(T_{1i})+1) \to 0$ w.p.1. Thus, in the limit, we can focus on $C_{1i}(\tilde{T}_{1i},\delta)$.

Now condition on $S_{1i}^2(n_0)$. Then Lemma 3, (S.9) and CMT imply that

$$C_{1i}(t,\delta) + \frac{(N_{1i}(T_{1i})+1)\delta_{1i}t}{\sigma_{1i}\sqrt{N_{1i}(T_{1i})+1}} \Rightarrow \mathcal{W}(t,\Delta)$$

as $\delta \to 0$, where

$$\Delta = \lim_{\delta\to 0} \frac{(N_{1i}(T_{1i})+1)\delta_{1i}}{\sigma_{1i}\sqrt{N_{1i}(T_{1i})+1}} = \frac{\sqrt{2eh^2 S_{1i}^2(n_0)}}{\sigma_{1i}}.$$

Still conditional on $S_{1i}^2(n_0)$, let

$$\mathcal{A}(\delta) = \frac{h^2 S_{1i}^2(n_0)}{\hat{\delta}_{1i}(T_{1i})\sigma_{1i}\sqrt{N_{1i}(T_{1i})+1}} \stackrel{\delta\to 0}{\longrightarrow} \frac{\sqrt{2eh^2\, S_{1i}^2(n_0)}}{2e\sigma_{1i}} \equiv \mathcal{A}$$

$$\mathcal{B}(\delta) = \frac{\hat{\delta}_{1i}(T_{1i})(N_{1i}(T_{1i})+1)}{2e\sigma_{1i}\sqrt{N_{1i}(T_{1i})+1}} \stackrel{\delta\to 0}{\longrightarrow} \frac{\sqrt{2eh^2\, S_{1i}^2(n_0)}}{2e\sigma_{1i}} \equiv \mathcal{B}.$$

Notice that the stopping time $\tilde{T}_{1i}$ is the first time t at which the event

$$\left\{ \left| C_{1i}(t,\delta) + \frac{(N_{1i}(T_{1i})+1)\delta_{1i}t}{\sigma_{1i}\sqrt{N_{1i}(T_{1i})+1}} \right| - \mathcal{A}(\delta) + \mathcal{B}(\delta)t \ge 0 \right\}$$

occurs. Define the mapping $s_\delta : D[0,1] \to \mathbb{R}$ such that $s_\delta(Y) = Y(T_{Y,\delta})$ where

$$T_{Y,\delta} = \inf\{t : |Y(t)| - \mathcal{A}(\delta) + \mathcal{B}(\delta)t \geq 0\}$$

for every $Y \in D[0,1]$ and $\delta > 0$. Similarly, define $s(Y) = Y(T_Y)$ where

$$T_Y = \inf\{t : |Y(t)| - \mathcal{A} + \mathcal{B}t \geq 0\}$$

for every $Y \in D[0,1]$ and $\delta > 0$. Notice that

$$s_\delta\left(C_{1i}(t,\delta) + \frac{(N_{1i}(T_{1i})+1)\delta_{1i}t}{\sigma_{1i}\sqrt{N_{1i}(T_{1i})+1}}\right) = C_{1i}(\tilde{T}_{1i},\delta) + \frac{(N_{1i}(T_{1i})+1)\delta_{1i}\tilde{T}_{1i}}{\sigma_{1i}\sqrt{N_{1i}(T_{1i})+1}}$$

$$s\left(\mathcal{W}(\cdot,\Delta)\right) = \mathcal{W}(T_{\mathcal{W}(\cdot,\Delta)},\Delta).$$

We need to show that

$$s_\delta\left(\mathcal{G}_{1i}(\cdot,\delta)\right) \Rightarrow s\left(\mathcal{W}(\cdot,\Delta)\right) \tag{S.11}$$

as $\delta \to 0$, where

$$\mathcal{G}_{1i}(t,\delta) \equiv C_{1i}(t,\delta) + \frac{(N_{1i}(T_{1i})+1)\delta_{1i}t}{\sigma_{1i}\sqrt{N_{1i}(T_{1i})+1}} \tag{S.12}$$

for $t \in [0,1]$ and $\delta > 0$. This follows Proposition 2 from Kim et al. [2005] which is established on the extended CMT (Billingsley [1968]).

Unconditioning on S_{1i}^2, (S.11) gives

$$\begin{aligned}
&\limsup_{\delta\to 0} \Pr\{\mathrm{ICS}_i\} \\
&= \mathrm{E}\left[\Pr\left\{\mathcal{W}(t,\Delta) \text{ exits continuation region through the lower boundary} | S_{1i}^2(n_0)\right\}\right] \\
&= \mathrm{E}\left[\sum_{v=1}^{e}(-1)^{v+1}\left(1-\frac{1}{2}\mathcal{I}(v=e)\right)\exp\left\{-\frac{h^2 S_{1i}^2(n_0)}{e\sigma_{1i}^2}(2e-v)v\right\}\middle| S_{1i}^2(n_0)\right] &\text{(S.13)} \\
&= \mathrm{E}\left[\sum_{v=1}^{e}(-1)^{v+1}\left(1-\frac{1}{2}\mathcal{I}(v=e)\right)\exp\left\{-\frac{\eta(n_0-1) S_{1i}^2(n_0)}{e\sigma_{1i}^2}(2e-v)v\right\}\middle| S_{1i}^2(n_0)\right] &\text{(S.14)} \\
&= \sum_{v=1}^{e}(-1)^{v+1}\left(1-\frac{1}{2}\mathcal{I}(v=e)\right)\mathrm{E}\left[\exp\left\{-\frac{\eta\chi_{n_0-1}^2}{e}(2e-v)v\right\}\right] \\
&= \sum_{v=1}^{e}(-1)^{v+1}\left(1-\frac{1}{2}\mathcal{I}(v=e)\right)\left(1+\frac{2\eta(2e-v)v}{e}\right)^{-(n_0-1)/2} &\text{(S.15)} \\
&= \frac{\alpha}{k-1},
\end{aligned}$$

where (S.14) holds due to Lemma 2, χ_f^2 is a chi-squared random variable with degrees of freedom f, and (S.15) is from the moment generating function of χ_f^2. Finally the probability of incorrect selection is less than the nominal level as below.

$$\limsup_{\delta\to 0}\Pr\{\mathrm{ICS}\} \leq \limsup_{\delta\to 0}\sum_{i=2}^{k}\Pr\{\mathrm{ICS}_i\} \leq \alpha.$$

□

5 Proof of Theorem 2

For the proof of $\mathcal{WK}$++ we need the following lemma:

Lemma 5 *Billingsley [1968]*
If $Y_n \Rightarrow Y$ and $Z_n \xrightarrow{P} a$ where a is a constant vector, then $(Y_n,\, Z_n) \Rightarrow (Y,\, a)$.

Proof of Theorem 2: First set T_{1i} and $N_{1i}(r)$ in the same way as the proof of Theorem 1, except replacing $S^2_{1i}(n_0)$ with $S^2_{1i}(r)$. Sample variance $S^2_{1i}(n_0)$ in (S.7) and (S.8) is also replaced with $S^2_{1i}(T^*)$. Then using a similar argument we can show $T_{1i} \to \infty$ as $\delta \to 0$ w.p.1 to guarantee the convergence of estimators. This ensures that as $\delta \to 0$,

$$\hat{\delta}_{1i}(T_{1i})/\delta_{1i} \to 1 \quad \text{and} \quad S^2_{1i}(T_{1i}) \to \sigma^2_{1i}, \text{ w.p.1.} \tag{S.16}$$

Also, $N_{1i}(r) \to \infty$ as $\delta \to 0$ w.p.1 due to (S.16) and CMT. The definition of $C_{1i}(t,\delta)$ is the same as in the proof of Theorem 1. Redefine $\hat{T}_{1i}$ and $\tilde{T}_{1i}$ as follows:

$$\begin{aligned}\widehat{T}_{1i} = \min\Bigg\{ t \in \left\{\frac{n_0}{N_{1i}(T_{1i})+1}, \frac{n_0+1}{N_{1i}(T_{1i})+1}, \ldots, 1\right\} : \left| C_{1i}(t,\delta) + \frac{(N_{1i}(T_{1i})+1)\delta_{1i}t}{\sigma_{1i}\sqrt{N_{1i}(T_{1i})+1}}\right| \\ \geq \frac{h^2 S^2_{1i}(T_{1i})}{\hat{\delta}_{1i}(T_{1i})\sigma_{1i}\sqrt{N_{1i}(T_{1i})+1}} - \frac{\hat{\delta}_{1i}(T_{1i})(N_{1i}(T_{1i})+1)}{2e\sigma_{1i}\sqrt{N_{1i}(T_{1i})+1}}t \Bigg\}. \\ \tilde{T}_{1i} = \min\Bigg\{ t \geq \frac{n_0}{N_{1i}(T_{1i})+1} : \left| C_{1i}(t,\delta) + \frac{(N_{1i}(T_{1i})+1)\delta_{1i}t}{\sigma_{1i}\sqrt{N_{1i}(T_{1i})+1}}\right| \\ \geq \frac{h^2 S^2_{1i}(T_{1i})}{\hat{\delta}_{1i}(T_{1i})\sigma_{1i}\sqrt{N_{1i}(T_{1i})+1}} - \frac{\hat{\delta}_{1i}(T_{1i})(N_{1i}(T_{1i})+1)}{2e\sigma_{1i}\sqrt{N_{1i}(T_{1i})+1}}t \Bigg\}.\end{aligned}$$

We can show that $\widehat{T}_{1i} \to \tilde{T}_{1i}$ w.p.1 as $\delta \to 0$ making use of the fact that $1/(N_{1i}(T_{1i})+1) \to 0$ w.p.1. Thus, in the limit, we can focus on $C_{1i}(\tilde{T}_{1i},\delta)$.

Since strong consistency implies convergence in probability, (S.16) and Lemma 5 can be applied to $\left(C_{1i}(T_{1i},\delta), \begin{bmatrix} \hat{\delta}_{1i}(T_{1i})/\delta_{1i} \\ S^2_{1i}(T_{1i}) \end{bmatrix}\right)$. By the random-change-of-time theorem (Billingsley [1968]), the followings hold.

$$C_{1i}(t,\delta) + \frac{(N_{1i}(T_{1i})+1)\delta_{1i}t}{\sigma_{1i}\sqrt{N_{1i}(T_{1i})+1}} \Rightarrow \mathcal{W}(t,\Delta)$$

as $\delta \to 0$, where

$$\Delta = \lim_{\delta\to 0} \frac{(N_{1i}(T_{1i})+1)\delta_{1i}}{\sigma_{1i}\sqrt{N_{1i}(T_{1i})+1}} = \sqrt{2eh^2}.$$

$$\begin{aligned}\mathcal{A}(\delta) &= \frac{h^2 S^2_{1i}(T_{1i})}{\hat{\delta}_{1i}(T_{1i})\sigma_{1i}\sqrt{N_{1i}(T_{1i})+1}} \xrightarrow{\delta\to 0} \sqrt{\frac{h^2}{2e}} \equiv \mathcal{A} \\ \mathcal{B}(\delta) &= \frac{\hat{\delta}_{1i}(T_{1i})(N_{1i}(T_{1i})+1)}{2e\sigma_{1i}\sqrt{N_{1i}(T_{1i})+1}} \xrightarrow{\delta\to 0} \sqrt{\frac{h^2}{2e}} \equiv \mathcal{B}\end{aligned}$$

The definitions of mappings s_δ, s, $T_{Y,\delta}$, and T_Y follow those in the proof of Theorem 1 . Then (S.11) holds again. Therefore,

$$\begin{aligned}\limsup_{\delta\to 0} \Pr\{\mathrm{ICS_i}\} &= \Pr\{\mathcal{W}(t,\Delta) \text{ exits continuation region through the lower boundary}\} \\ &= \sum_{v=1}^{e}(-1)^{v+1}\left(1-\frac{1}{2}\mathcal{I}(v=e)\right)\exp\left\{-2\frac{h^2}{2e}(2e-v)v\right\} \text{ (by Lemma 2)} \\ &= \sum_{v=1}^{e}(-1)^{v+1}\left(1-\frac{1}{2}\mathcal{I}(v=e)\right)\exp\left\{-\frac{\eta}{e}(2e-v)v\right\} = \alpha/(k-1) \qquad \text{(S.17)}\end{aligned}$$

where the equality follows from the way we choose η. Finally,

$$\limsup_{\delta\to 0}\Pr\{\mathrm{ICS}\} \le \limsup_{\delta\to 0}\sum_{i=2}^{k}\Pr\{\mathrm{ICS}_i\} \le \alpha.$$

□

6 Estimated PCS for $\mathcal{WK}$ and $\mathcal{WK}$++

Figure S3 shows estimated PCS for $\mathcal{WK}$ and $\mathcal{WK}$++ when $\rho = 0$ and $\rho = 0.6$.

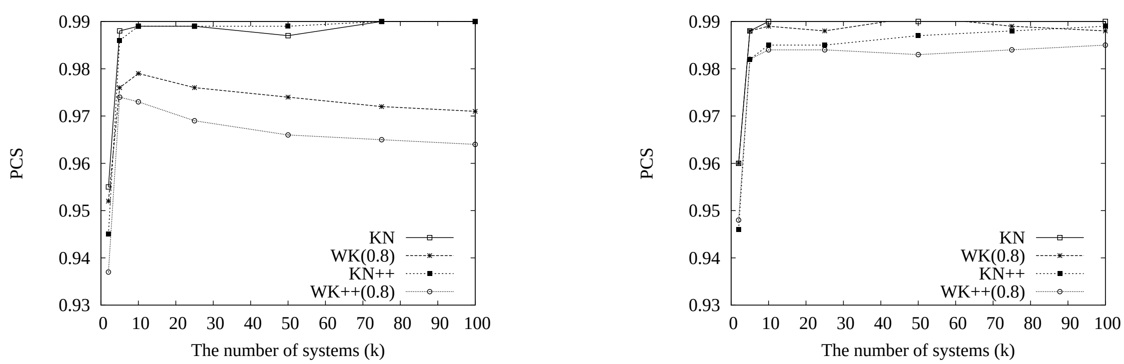

Figure S3: Estimated PCS under the DC-INC with $\rho = 0$ (left) and $\rho = 0.6$ (right).

References

P. Billingsley. *Convergence of Probability Measures*. John Wiley & Sons, Inc., first edition, 1968.

V. Fabian. Note on anderson's sequential procedures with triangular boundary. *Annals of Statistics*, 2:170–176, 1974.

C. Jennison, I. M. Johnstone, and B. W. Turnbull. Asymptotically optimal procedures for sequential adaptive selection of the best of several normal means. In S. S. Gupta and J. O. Berger, editors, *Statistical Decision Theory and Related Topics, iii*, volume 2, pages 55–86. Academic Press, 1982.

S.-H. Kim, B. L. Nelson, and J. R. Wilson. Some almost-sure convergence properties useful in sequential analysis. *Sequential Analysis*, 24:411–419, 2005.

S. K. Perng. A comparison of the asymptotic expected sample sizes of two sequential procedures for ranking problem. *Annals of Mathematical Statistics*, 40:2198–2202, 1969.