

Without loss of generality, suppose that the true means of the systems are indexed so that $\mu_k \geq \mu_{k-1} \geq \dots \geq \mu_1$. Assume $c > 0$.

Assumption 1 There exist finite constants μ_i and v_i^2 such that the probability of distribution of $C_i(t, r) \equiv \frac{\sum_{j=1}^{\text{floor}(rt)} X_{ij} - rt\mu_i}{v_i\sqrt{r}}$ over $D[0, 1]$ converges to that of a standard Brownian motion process, $W(t)$, for t on the unit interval, as r increases; i.e.

$$C_i(\cdot, r) \implies W(\cdot)$$

as $r \rightarrow \infty$.

Note: $D[0, 1]$ is the Skorohod space, i.e. it's the space of real-valued functions on $[0, 1]$ that are right-continuous and have left-hand limits.

Under Assumption 1, μ_i is the mean, and $v_i^2 = \lim_{r \rightarrow \infty} r \text{Var}(\bar{X}_i(r))$ where $\bar{X}_i(r)$ is the sample mean of the first r observations from system i .

Lemma. For $i \neq l$, if \mathbf{X}_i and \mathbf{X}_l satisfy Assumption 1 and are independent, then there exists a constant v_{il}^2 such that

$$\frac{t_l \left(Y_{\text{floor}(rt_i), i} - rt_i \mu_i \right) - t_i \left(Y_{\text{floor}(rt_l), l} - rt_l \mu_l \right)}{v_{il} \sqrt{r}} \implies t_l W_1(t_i) - t_i W_2(t_l)$$

as $r \rightarrow \infty$.

Proof. First, note that

$$v_{il}^2 := \lim_{r, s \rightarrow \infty} \text{Var}(\sqrt{r} \bar{X}_i(r) + \sqrt{s} \bar{X}_l(s)) = v_i^2 + v_l^2$$

because of the independence of \mathbf{X}_i and \mathbf{X}_l . Now,

$$\begin{aligned} & \frac{t_l \left(Y_{\text{floor}(rt_i),i} - rt_i \mu_i \right) - t_i \left(Y_{\text{floor}(rt_l),l} - rt_l \mu_l \right)}{v_{il} \sqrt{r}} = \\ & t_l \frac{\sum_{j=1}^{\text{floor}(rt_i)} X_{ij} - rt_i \mu_i}{v_{il} \sqrt{r}} - t_i \frac{\sum_{j=1}^{\text{floor}(rt_l)} X_{lj} - rt_l \mu_l}{v_{il} \sqrt{r}} = \\ & t_l \left(\frac{v_i}{v_{il}} \right) C_i(t, r) - t_i \left(\frac{v_l}{v_{il}} \right) C_l(t, r). \end{aligned}$$

Because we assume that \mathbf{X}_i and \mathbf{X}_l are independent, so are $C_i(t, r)$ and $C_l(t, r)$. Assumption 1 implies $C_i(\cdot, r) \Rightarrow W_i(\cdot)$ and $C_l(\cdot, r) \Rightarrow W_l(\cdot)$ where W_i and W_l are independent standard Brownian motion processes. By Theorem 3.2 of Billingsley, $(C_i(\cdot, r), C_l(\cdot, r)) \Rightarrow (W_i(\cdot), W_l(\cdot))$. By the Continuous Mapping Theorem,

$$t_l \left(\frac{v_i}{v_{il}} \right) C_i(t, r) - t_i \left(\frac{v_l}{v_{il}} \right) C_l(t, s) \Rightarrow t_l \left(\frac{v_i}{v_{il}} \right) W_i(t) - t_i \left(\frac{v_l}{v_{il}} \right) W_l(t).$$

Theorem. If samples from system $x \in \{1 \dots, k\}$ are normally distributed and independent, over time and across alternatives, then $\lim_{\delta \rightarrow 0} Pr \{ \text{BIZ selects } k \} \geq P^*$ provided $\mu_k \geq \mu_{k-1} + \delta$. We also suppose that the algorithm ends in at most $R(\delta) \in \mathbb{N}$ iterations, and $R(\delta) \rightarrow \infty$ as $\delta \rightarrow 0$ with probability 1. Furthermore, $\sqrt{R}\delta \rightarrow \Delta$ with probability 1. Suppose $B_1 = \dots = B_k = 1$.

Proof. Suppose the variances are known. We begin by considering the case of only two systems, denoted k and i , with $\mu_k \geq \mu_i + \delta^*$. Let $A = \{k, i\}$. Let

$$T(\delta) = \min \{t \leq R, t \in \mathbb{N} : \min_{x \in A} \hat{q}_{tx}(A) \leq c \text{ or } \max_{x \in A} \hat{q}_{tx}(A) \geq P^*\}.$$

Thus $T(\delta)$ is the stage at which the procedure terminates.

Now, let $a_t, b_t \in A$ such that $\exp\left(\delta\beta_t \frac{W_{ta_t}}{n_{ta_t}}\right) \geq \exp\left(\delta\beta_t \frac{W_{tb_t}}{n_{tb_t}}\right)$, so

$$\begin{aligned} \min_{x \in A} \hat{q}_{tx}(A) &\leq c \\ \Leftrightarrow \min_{x \in A} \exp\left(\delta\beta_t \frac{W_{tx}}{n_{tx}}\right) &\leq c \sum_{x' \in A} \exp\left(\delta\beta_t \frac{W_{tx'}}{n_{tx'}}\right) \\ \Leftrightarrow \exp\left(\delta\beta_t \frac{W_{tb_t}}{n_{tb_t}}\right) (1 - c) &\leq c \exp\left(\delta\beta_t \frac{W_{ta_t}}{n_{ta_t}}\right) \\ \Leftrightarrow \exp\left(\delta\beta_t \left(\frac{W_{tb_t}}{n_{tb_t}} - \frac{W_{ta_t}}{n_{ta_t}}\right)\right) &\leq \frac{c}{1 - c} \end{aligned}$$

an

$$\begin{aligned} \max_{x \in A} \hat{q}_{tx}(A) &\geq P^* \\ \Leftrightarrow \max_{x \in A} \exp\left(\delta\beta_t \frac{W_{tx}}{n_{tx}}\right) &\geq P^* \sum_{x' \in A} \exp\left(\delta\beta_t \frac{W_{tx'}}{n_{tx'}}\right) \\ \Leftrightarrow \exp\left(\delta\beta_t \frac{W_{ta_t}}{n_{ta_t}}\right) (1 - P^*) &\geq P^* \exp\left(\delta\beta_t \frac{W_{tb_t}}{n_{tb_t}}\right) \\ \Leftrightarrow \exp\left(\delta\beta_t \left(\frac{W_{ta_t}}{n_{ta_t}} - \frac{W_{tb_t}}{n_{tb_t}}\right)\right) &\geq \frac{P^*}{1 - P^*}, \end{aligned}$$

thus

$$\begin{aligned}
T(\delta) &= \min \left\{ t \leq R, t \in \mathbb{N} : \exp \left(\delta \beta_t \left| \frac{W_{tk}}{n_{tk}} - \frac{W_{ti}}{n_{ti}} \right| \right) \geq \min \left\{ \frac{P^*}{1-P^*}, \frac{1-c}{c} \right\} \right\} \\
&= \min \left\{ t \leq R, t \in \mathbb{N} : \left| \frac{W_{tk}}{n_{tk}} - \frac{W_{ti}}{n_{ti}} \right| \geq \frac{\log \left[\min \left\{ \frac{P^*}{1-P^*}, \frac{1-c}{c} \right\} \right]}{\delta \beta_t} \right\}
\end{aligned}$$

Now, let's prove that $n_{ti} \rightarrow \infty, n_{tk} \rightarrow \infty$ as $t \rightarrow \infty$. For each t , $n_{t+1,x} = n_{tx} + B_x$ for some $x \in A$. If $n_{t+1,i} = n_{ti} + B_i$ for a finite number of t 's, then $n_{tk} \rightarrow \infty$ as $t \rightarrow \infty$. So, there exists l_0 such that $n_{li}/\hat{\lambda}_{li}^2 > n_{lk}/\hat{\lambda}_{lk}^2$ if $l > l_0$. Since $\hat{\lambda}_{ti}^2 \rightarrow \sigma_i^2$ and $\hat{\lambda}_{tk}^2 \rightarrow \sigma_k^2$ as $t \rightarrow \infty$, then $n_{ti} \rightarrow \infty$ as $t \rightarrow \infty$. Similarly, we get the same result in the other cases. Thus, $n_{ti} \rightarrow \infty, n_{tk} \rightarrow \infty$ as $t \rightarrow \infty$.

Let $\lambda_r^2 = \min \{\lambda_i^2, \lambda_k^2\}$ and $\lambda_s^2 = \max \{\lambda_i^2, \lambda_k^2\}$. Then $n_{1,s} = n_0 + 1$, $n_{1,r} = \max \left\{ n_0, \text{ceil} \left(\frac{\lambda_r^2}{\lambda_s^2} (n_0 + 1) \right) \right\}$. Note that $\frac{n_{ts}}{\lambda_s^2} \leq \frac{n_{tr}}{\lambda_r^2}$ for all t , because by induction

$$\frac{n_{t+1,r}}{\lambda_r^2} \geq \frac{\frac{\lambda_r^2}{\lambda_s^2} (n_{ts} + 1)}{\lambda_r^2} = \frac{n_{ts} + 1}{\lambda_s^2}.$$

So, $n_{tr} = \max \left\{ n_{t-1,r}, \text{ceil} \left(\frac{\lambda_r^2}{\lambda_s^2} (n_{t-1,s} + 1) \right) \right\} = \max \left\{ n_{t-1,r}, \text{ceil} \left(\frac{\lambda_r^2}{\lambda_s^2} (n_0 + t) \right) \right\}$.

Thus there exists t_0 such that $n_{t_0,r} = \text{ceil} \left(\frac{\lambda_r^2}{\lambda_s^2} (n_0 + t_0) \right)$ and $n_{tr} = n_0$ if $t < t_0$.

Observe that if $t > t_0$, $n_{tr} = \text{ceil} \left(\frac{\lambda_r^2}{\lambda_s^2} (n_0 + t) \right)$. Furthermore, $n_{ts} = n_0 + t$.

Suppose $i = s$ or $i = r$ and $t_0 = 1$. If $0 \leq t \leq 1$, let

$$\begin{aligned}
D_i(t, R) &= \frac{\sum_{j=1}^{n_{\text{floor}}(tR), i} X_{ij} - (n_0 + (tR)) \mu_i}{v_i \sqrt{R}} \\
&= \frac{\sum_{j=1}^{n_0 + \text{floor}(tR)} X_{ij} - (n_0 + (tR)) \mu_i}{v_i \sqrt{R}} \\
&= \frac{\sum_{j=1}^{n_0} X_{ij} - n_0 \mu_i}{v_i \sqrt{R}} + \frac{\sum_{j=1}^{\text{floor}(tR)} X_{i, n_0+j} - (tR) \mu_i}{v_i \sqrt{R}} \\
&\Rightarrow W_i().
\end{aligned}$$

Suppose $\lambda_i^2 \leq \lambda_k^2$. Assume $t_0 = 1$. Let

$$\begin{aligned}
D_{ik}(t, R) &= \frac{n_{tR, i} \left(\sum_{j=1}^{n_{\text{floor}}(tR), k} X_{kj} - (n_0 + (tR)) \mu_k \right)}{v_{ik} R \sqrt{R}} \\
&\quad - \frac{n_{tR, k} \left(\sum_{j=1}^{n_{\text{floor}}(tR), i} X_{kj} - \left(\frac{\lambda_i^2}{\lambda_k^2} (n_0 + (tR)) \mu_i \right) \right)}{v_{ik} R \sqrt{R}} \\
&= \frac{\text{ceil} \left(\frac{\lambda_i^2}{\lambda_k^2} (n_0 + tR) \right) \left(\sum_{j=1}^{(n_0 + \text{floor}(tR))} X_{kj} - (n_0 + (tR)) \mu_k \right)}{v_{ik} R \sqrt{R}} \\
&\quad - \frac{(n_0 + tR) \left(\sum_{j=1}^{\text{ceil} \left(\frac{\lambda_i^2}{\lambda_k^2} (n_0 + \text{floor}(tR)) \right)} X_{kj} - \text{ceil} \left(\frac{\lambda_i^2}{\lambda_k^2} (n_0 + (tR)) \mu_i \right) \right)}{v_{ik} R \sqrt{R}} \\
&\Rightarrow \frac{v_k \lambda_i^2}{v_{ik} \lambda_k^2} t W_k(t) - \frac{v_i}{v_{ik}} t W_i(t) = t W().
\end{aligned}$$

Now, let $N = \max\{n_{Rk}, n_{Ri}\}$, then

$$\begin{aligned} \frac{W_{T(\delta)k}}{n_{T(\delta)k}} &< \frac{W_{T(\delta)i}}{n_{T(\delta)i}} \\ \Leftrightarrow \frac{\mu_k - \mu_i}{R^{3/2}v_{ik}} (n_{T(\delta)k}n_{T(\delta)i}) &+ \frac{n_{T(\delta)i} \left(\sum_{j=1}^{n_{T(\delta)k}} X_{kj} - n_{T(\delta)k}\mu_k \right) - n_{T(\delta)k} \left(\sum_{j=1}^{n_{T(\delta)i}} X_{ij} - n_{T(\delta)i}\mu_i \right)}{R^{3/2}v_{ik}} < 0. \end{aligned}$$

Then

$$\begin{aligned} \mathbb{P} \left\{ \frac{W_{T(\delta)k}}{n_{T(\delta)k}} < \frac{W_{T(\delta)i}}{n_{T(\delta)i}} \right\} &\leq \mathbb{P} \left\{ \frac{n_{T(\delta)i} \left(\sum_{j=1}^{n_{T(\delta)k}} X_{kj} - n_{T(\delta)k}\mu_k \right) - n_{T(\delta)k} \left(\sum_{j=1}^{n_{T(\delta)i}} X_{ij} - n_{T(\delta)i}\mu_i \right)}{R^{3/2}v_{ik}} \right. \\ &\quad \left. + \frac{\delta}{R^{3/2}v_{ik}} (n_{T(\delta)k}n_{T(\delta)i}) < 0 \right\} \\ &= E \left[\mathbb{P} \left\{ \frac{n_{T(\delta)i} \left(\sum_{j=1}^{n_{T(\delta)k}} X_{kj} - n_{T(\delta)k}\mu_k \right) - n_{T(\delta)k} \left(\sum_{j=1}^{n_{T(\delta)i}} X_{ij} - n_{T(\delta)i}\mu_i \right)}{R^{3/2}v_{ik}} \right. \right. \\ &\quad \left. \left. + \frac{\delta}{R^{3/2}v_{ik}} (n_{T(\delta)k}n_{T(\delta)i}) < 0 \right\} \mid \Delta \right] \end{aligned}$$

Define

$$\begin{aligned} \hat{T}(\delta) &= \min \left\{ t \in \left\{ \frac{1}{R}, \dots, 1 \right\} : \left| D_{ik}(t, \delta) + \frac{\text{ceil} \left(\frac{\lambda_i^2}{\lambda_k^2} (n_0 + tR) \right) (n_0 + tR) \delta}{R^{3/2}v_{ik}} \right| \right. \\ &\quad \left. \geq \frac{\text{ceil} \left(\frac{\lambda_i^2}{\lambda_k^2} (n_0 + tR) \right) (n_0 + tR) \log \left[\min \left\{ \frac{P^*}{1-P^*}, \frac{1-c}{c} \right\} \right]}{R^{3/2}v_{ik}\beta_{tR}\delta} \right\} \end{aligned}$$

where $\beta_{tR} = \frac{n_0 + tR + \text{ceil} \left(\frac{\lambda_i^2}{\lambda_k^2} (n_0 + tR) \right)}{\lambda_i^2 + \lambda_k^2}$.

Clearly, $\hat{T}(\delta) = T(\delta)/R$. Also, define the stopping time of the corresponding continuous-time process as

$$\begin{aligned}\tilde{T}(\delta) &= \min \left\{ 1 \geq t \geq \frac{1}{R} : \left| D_{ik}(t, \delta) + \frac{\text{ceil} \left(\frac{\lambda_i^2}{\lambda_k^2} (n_0 + tR) \right) (n_0 + tR) \delta}{R^{3/2} v_{ik}} \right| \right. \\ &\quad \left. \geq \frac{\text{ceil} \left(\frac{\lambda_i^2}{\lambda_k^2} (n_0 + tR) \right) (n_0 + tR) \log \left[\min \left\{ \frac{P^*}{1-P^*}, \frac{1-c}{c} \right\} \right]}{R^{3/2} v_{ik} \beta_{tR} \delta} \right\}\end{aligned}$$

Note that for fixed δ , $D_{ik}(\hat{T}(\delta), \delta)$ corresponds to the right-hand limit of a point of discontinuity of $D_{ik}(\cdot, \delta)$. We can show that $\hat{T}(\delta) \rightarrow \tilde{T}(\delta)$ with probability 1 as $\delta \rightarrow 0$, making use of the fact that $1/R \rightarrow 0$ with probability 1. Thus, in the limit, we can focus on $D_{ik}(\tilde{T}(\delta), \delta)$.

Now, condition on Δ . By Assumption 1, Lemma 1, and the CMT we have that

$$\begin{aligned}D_{ik}(t, \delta) + \frac{\text{ceil} \left(\frac{\lambda_i^2}{\lambda_k^2} (n_0 + tR) \right) (n_0 + tR) \delta}{R^{3/2} v_{ik}} &\Rightarrow tW(t) + t^2 \Delta. \\ &= t(W(t, \Delta))\end{aligned}$$

Let

$$\begin{aligned}A(\delta) &= \frac{n_0 \text{ceil} \left(\frac{\lambda_i^2}{\lambda_k^2} (n_0 + tR) \right) \log \left[\min \left\{ \frac{P^*}{1-P^*}, \frac{1-c}{c} \right\} \right]}{R^{3/2} v_{ik} \beta_{tR} \delta} \\ &= \frac{(\lambda_i^2 + \lambda_k^2) n_0 \text{ceil} \left(\frac{\lambda_i^2}{\lambda_k^2} (n_0 + tR) \right) \log \left[\min \left\{ \frac{P^*}{1-P^*}, \frac{1-c}{c} \right\} \right]}{\left(n_0 + tR + \text{ceil} \left(\frac{\lambda_i^2}{\lambda_k^2} (n_0 + tR) \right) \right) R^{3/2} v_{ik} \delta} \xrightarrow{\delta \rightarrow 0} 0\end{aligned}$$

and

$$\begin{aligned}
tB(\delta) &= \frac{(\lambda_i^2 + \lambda_k^2) \text{ceil} \left(\frac{\lambda_i^2}{\lambda_k^2} (n_0 + tR) \right) (tR) \log \left[\min \left\{ \frac{P^*}{1-P^*}, \frac{1-c}{c} \right\} \right]}{\left(n_0 + tR + \text{ceil} \left(\frac{\lambda_i^2}{\lambda_k^2} (n_0 + tR) \right) \right) R^{3/2} v_{ik} \delta} \\
&= \frac{(\lambda_i^2 + \lambda_k^2) \text{ceil} \left(\frac{\lambda_i^2}{\lambda_k^2} (n_0 + tR) \right) (t) \log \left[\min \left\{ \frac{P^*}{1-P^*}, \frac{1-c}{c} \right\} \right]}{\left(n_0 + tR + \text{ceil} \left(\frac{\lambda_i^2}{\lambda_k^2} (n_0 + tR) \right) \right) R^{1/2} v_{ik} \delta} \\
&\xrightarrow{\delta \rightarrow 0} \frac{(\lambda_i^2 + \lambda_k^2) \frac{\lambda_i^2}{\lambda_k^2} (t) \log \left[\min \left\{ \frac{P^*}{1-P^*}, \frac{1-c}{c} \right\} \right]}{\left(1 + \frac{\lambda_i^2}{\lambda_k^2} \right) v_{ik} \Delta} = Bt.
\end{aligned}$$

Note that the stopping time $\tilde{T}(\delta)$ is the first time t at which the event

$$\left\{ \left| D_{ik}(t, \delta) + \frac{\text{ceil} \left(\frac{\lambda_i^2}{\lambda_k^2} (n_0 + tR) \right) (n_0 + tR) \delta}{R^{3/2} v_{ik}} \right| - A(\delta) - tB(\delta) \geq 0 \right\}$$

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

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

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

$$T_Y = \inf \{ t : |Y(t)| - Bt \geq 0 \}$$

for every $Y \in D[0, 1]$. Note that

$$s_\delta \left(D_{ik}(t, \delta) + \frac{\text{ceil} \left(\frac{\lambda_i^2}{\lambda_k^2} (n_0 + tR) \right) (n_0 + tR) \delta}{R^{3/2} v_{ik}} \right) = D_{ik}(\tilde{T}(\delta), \delta) \\ + \frac{\text{ceil} \left(\frac{\lambda_i^2}{\lambda_k^2} (n_0 + \tilde{T}(\delta) R) \right) (n_0 + \tilde{T}(\delta) R) \delta}{R^{3/2} v_{ik}},$$

$$s(tW(t, \Delta)) = T_{\mathfrak{W}(\cdot)} W(T_{\mathfrak{W}(\cdot)}, \Delta) \text{ where } \mathfrak{W}(t) = tW(t, \Delta).$$

We need to show that

$$s_\delta(G_{ik}(t, \delta)) \Rightarrow s(tW(t, \Delta))$$

as $\delta \rightarrow 0$, where

$$G_{ik}(t, \delta) \equiv D_{ik}(t, \delta) + \frac{\text{ceil} \left(\frac{\lambda_i^2}{\lambda_k^2} (n_0 + tR) \right) (n_0 + tR) \delta}{R^{3/2} v_{ik}}$$

for $t \in [0, 1]$ and $\delta > 0$. If $\mathbb{P}(tW(t, \Delta) \in D[0, 1] - D_s) = 1$ where $D_s \equiv \{x \in D[0, 1] : \text{for some sequence } \{x_n\} \subset D[0, 1] \text{ with } \lim_n d(x_n, x) = 0, \text{ the sequence } \{s_{\delta_n}(x_n)\} \text{ does not converge to } s(x)\}$, and $d(X, Y)$ is the infimum of those positive w for which there exists $\lambda \in \Lambda$ such that $\sup_{t \in [0, 1]} |X(t) - Y(\lambda(t))| \leq w$ and $\sup_{t \in [0, 1]} |\lambda(t) - t| \leq w$ (Λ is the class of strictly increasing, continuous mappings of $[0, 1]$ onto itself such that for every $\lambda \in \Lambda$, we have $\lambda(0) = 0$ and $\lambda(1) = 1$), then by Theorem 5.5 of Billingsley 1968 we conclude that this is true. By Kim et al. (2005), we know that $\mathbb{P}(W(t, \Delta) \in D[0, 1] - D_s) = 1$, thus

it follows this is true (to prove).

Now, unconditioning on Δ gives

$$\limsup_{\delta \rightarrow 0} \mathbb{P}(ICS) \leq E \left[\mathbb{P} \left[\tilde{T}(\delta) W(\tilde{T}(\delta), \Delta) < 0 \mid \Delta \right] \right]$$

Observe that,

$$\mu_k - \mu_i + \frac{v_k}{\sqrt{n_{T(\delta)k}}} C_k(1, n_{T(\delta)k}) - \frac{v_i}{\sqrt{n_{T(\delta)i}}} C_i(1, n_{T(\delta)i}) \Rightarrow \mu_k - \mu_i$$

when $\delta \rightarrow 0$ if $n_{T(\delta)j}$ is independent of $C_j(1, n)$ for n sufficiently large and $j = k, i$.

Suppose $i = r$ and $t_0 > 1$. Let $s_0 \leq t \leq 1$ where $\text{floor}(s_0 R) \geq t_0$ and $\text{floor}(sR) < t_0$ if $s < s_0$. Let

$$\begin{aligned} D_i(t, R) &= \frac{\sum_{j=1}^{n_{\text{floor}(tR), i}} X_{ij} - \frac{\lambda_i^2}{\lambda_k^2} (n_0 + (tR)) \mu_i}{v_i \sqrt{R}} \\ &= \frac{\sum_{j=1}^{\text{ceil}\left(\frac{\lambda_i^2}{\lambda_k^2} (n_0 + \text{floor}(tR))\right)} X_{ij} - (n_0 + (tR)) \mu_i}{v_i \sqrt{R}} \end{aligned}$$

and if $t < s_0$

$$\begin{aligned} D_i(t, R) &= \frac{\sum_{j=1}^{n_{floor(tR), i}} X_{ij} - n_0 \mu_i}{v_i \sqrt{R}} \\ &= \frac{\sum_{j=1}^{n_0} X_{ij} - n_0 \mu_i}{v_i \sqrt{R}}, \end{aligned}$$

Suppose $D_i(t, R) \Rightarrow W()$ if $t \geq s_0$.

$$D_{ik}(t, R) = \frac{\sum_{j=1}^n}{v_{ik} \sqrt{R}}$$

By Assumption 1, Lemma 1, and the CMT we have that

$$C_{ik} \left(\frac{n_{t,i}}{N}, \frac{n_{t,k}}{N}, \delta \right) + \frac{n_{t,k} n_{t,i} \delta}{N^{1/2} v_{ik}} \Rightarrow n_{ti} W_1(n_{ti}) - n_{tk} W_2(n_{tk}) + \frac{n_{t,k} n_{t,i} \Delta}{v_{ik}}.$$

Let

$$A(\delta) = \frac{\log \left[\min \left\{ \frac{P^*}{1-P^*}, \frac{1-c}{c} \right\} \right]}{N^{1/2} v_{ik} \delta} \rightarrow \frac{\log \left[\min \left\{ \frac{P^*}{1-P^*}, \frac{1-c}{c} \right\} \right]}{N^{1/2} v_{ik} \delta}$$

$$\beta_t = \sum$$

Observe that,

$$\mu_k - \mu_i + \frac{v_k}{\sqrt{n_{T(\delta)k}}} C_k(1, n_{T(\delta)k}) - \frac{v_i}{\sqrt{n_{T(\delta)i}}} C_i(1, n_{T(\delta)i}) \Rightarrow \mu_k - \mu_i$$

when $\delta \rightarrow 0$ if $n_{T(\delta)j}$ is independent of $C_j(1, n)$ for n sufficiently large and $j = k, i$.

Let ICS denote the event that an incorrect selection is made. Then,

$$\begin{aligned} \liminf_{\delta \rightarrow 0} \mathbb{P}\{ICS\} &= \liminf_{\delta \rightarrow 0} \mathbb{P}\left\{ \frac{W_{T(\delta)k}}{n_{T(\delta)k}} < \frac{W_{T(\delta)i}}{n_{T(\delta)i}} \right\} \\ &= \mathbb{P}\{\mu_k < \mu_i\} = 0. \end{aligned}$$

However, it's very likely $n_{T(\delta)k}$ is not independent of $C_k(1, n)$. Thus, let's give other proof. First, we'll show that n_{ti} is independent of $C_i(1, n)$ if $n \geq 1$ and $t \geq 0$. Note that

$$\begin{aligned} C_i(1, n) &= \frac{\sqrt{n} \sum_{j=1}^n X_{ij}}{v_i n} - \frac{\sqrt{n}}{v_i} \mu_i \\ &= \frac{\sqrt{n} \bar{X}_i(n)}{v_i} - \frac{\sqrt{n}}{v_i} \mu_i. \end{aligned}$$

Since $n_{0i} = n_0$, then $C_i(1, n)$ is independent of n_{0i} . Note that

$$n_{1,i} = \text{ceil} \left(\hat{\lambda}_{0i}^2 (n_0 + B_z) / \hat{\lambda}_{0z}^2 \right),$$

if $z = i$, then $n_{1,i}$ is independent of $C_i(1, n)$. Suppose $z = k$. Note that

$$\begin{aligned}
\hat{\lambda}_{1i}^2 &= \frac{1}{n-1} \sum_{j=1}^{n_{1i}} (X_{ij} - \bar{X}_i(n_{1i}))^2 \\
&= \frac{1}{n-1} \left((X_{i1} - \bar{X}_i(n_{1i}))^2 + \sum_{j=2}^{n_{1i}} (X_{ij} - \bar{X}_i(n_{1i}))^2 \right) \\
&= \frac{1}{n-1} \left(\left[\sum_{j=2}^{n_{1i}} (X_{ij} - \bar{X}_i(n_{1i}))^2 \right] + (X_{i1} - \bar{X}_i(n_{1i}))^2 \right).
\end{aligned}$$

Then $\hat{\lambda}_{1i}^2$ can be written as a function only of $(X_{i2} - \bar{X}_i(n_{0i}), \dots, X_{in_{ti}} - \bar{X}_i(n_{ti}), n_{ti})$.

Suppose $C_i(1, n)$ is independent of n_{ti} . Observe that

$$n_{t+1,i} = \text{ceil} \left(\hat{\lambda}_{ti}^2 (n_{tz} + B_z) / \hat{\lambda}_{tz}^2 \right),$$

if $z = i$, then $n_{t+1,i} = \text{ceil}((n_{tz} + B_z))$ and $C_i(1, n)$ are independent random

variables. Now suppose $z = k$. Note that

$$\begin{aligned}
\hat{\lambda}_{ti}^2 &= \frac{1}{n-1} \sum_{j=1}^{n_{ti}} (X_{ij} - \bar{X}_i(n_{ti}))^2 \\
&= \frac{1}{n-1} \left((X_{i1} - \bar{X}_i(n_{ti}))^2 + \sum_{j=2}^{n_{ti}} (X_{ij} - \bar{X}_i(n_{ti}))^2 \right) \\
&= \frac{1}{n-1} \left(\left[\sum_{j=2}^{n_{ti}} (X_{ij} - \bar{X}_i(n_{ti})) \right]^2 + \sum_{j=2}^{n_{ti}} (X_{ij} - \bar{X}_i(n_{ti}))^2 \right).
\end{aligned}$$

Then $\hat{\lambda}_{ti}^2$ can be written as a function only of $(X_{i2} - \bar{X}_i(n_{ti}), \dots, X_{in_{ti}} - \bar{X}_i(n_{ti}), n_{ti})$.

Suppose n_{ti} is given. The joint pdf of the sample $X_{i1}, \dots, X_{in_{ti}}$ is given by

$$f(x_1, \dots, x_{n_{ti}}) = \frac{1}{(2\pi)^{n_{ti}/2}} e^{-(1/2) \sum_{i=1}^{n_{ti}} x_i^2}.$$

Make the transformation $y_1 = \bar{x}(n_{ti}), y_2 = x_2 - \bar{x}(n_{ti}), \dots, y_{n_{ti}} = x_{n_{ti}} - \bar{x}(n_{ti})$. This is a linear transformation with a Jacobian equal to $1/n_{ti}$. We have

$$\begin{aligned}
f(y_1, \dots, y_{n_{ti}}) &= \frac{n_{ti}}{(2\pi)^{n_{ti}/2}} e^{-(1/2)(y_1 - \sum_{i=2}^{n_{ti}} y_i)^2} e^{-(1/2) \sum_{i=2}^{n_{ti}} (y_i + y_1)^2} \\
&= \left[\left(\frac{n_{ti}}{2\pi} \right)^{1/2} e^{(-n_{ti} y_1^2)/2} \right] \left[\frac{n_{ti}^{1/2}}{(2\pi)^{(n_{ti}-1)/2}} e^{-(1/2) [\sum_{i=2}^{n_{ti}} y_i^2 + (\sum_{i=2}^{n_{ti}} y_i)^2]} \right].
\end{aligned}$$

Then y_1 is independent of $\hat{\lambda}_{ti}^2$ given n_{ti} . If $n_{ti} \geq n$, then $\bar{X}_i(n)$ is independent

of $\hat{\lambda}_{ti}^2$ because X_{ij} is independent of $X_{ij'}$ for $j' < j$. If $n_{ti} < n$,

$$\begin{aligned}\bar{X}_i(n) &= \frac{\sum_{j=1}^{n_{ti}} X_{ij}}{n_{ti}} \frac{n_{ti}}{n} + \frac{\sum_{j=n_{ti}+1}^n X_{ij}}{n} \\ &= Y_1 \frac{n_{ti}}{n} + \frac{\sum_{j=n_{ti}+1}^n X_{ij}}{n}\end{aligned}$$

and then $\bar{X}_i(n)$ is independent of $\hat{\lambda}_{ti}^2$. Thus, $C_i(1, n)$ is independent of $\hat{\lambda}_{ti}^2$ given n_{ti} . So, by induction,

$$\begin{aligned}f\left(C_i(1, n), \hat{\lambda}_{ti}^2\right) &= \int f\left(C_i(1, n), \hat{\lambda}_{ti}^2 \mid n_{ti}\right) f(n_{ti}) dn_{ti} \\ &= f\left(C_i(1, n)\right) f\left(\hat{\lambda}_{ti}^2\right),\end{aligned}$$

then $C_i(1, n)$ is independent of $\hat{\lambda}_{ti}^2$. Consequently, $C_i(1, n)$ is independent of $n_{t,i+1}$. Thus

$$\mu_k - \mu_i + \frac{v_k}{\sqrt{n_{tk}}} C_k(1, n_{tk}) - \frac{v_i}{\sqrt{n_{ti}}} C_i(1, n_{ti}) \Rightarrow \mu_k - \mu_i$$

as $t \rightarrow \infty$. Let $Y_t = C_k(1, n_{T(1/t)k})$ and $X_t = C_k(1, n_{tk})$, if $d(Y_t, X_t) \rightarrow 0$ in probability, then $Y_t \Rightarrow W()$. Thus the result would be proved. Let's show that $d(Y_t, X_t) \rightarrow 0$. ???Let $\epsilon > 0$.

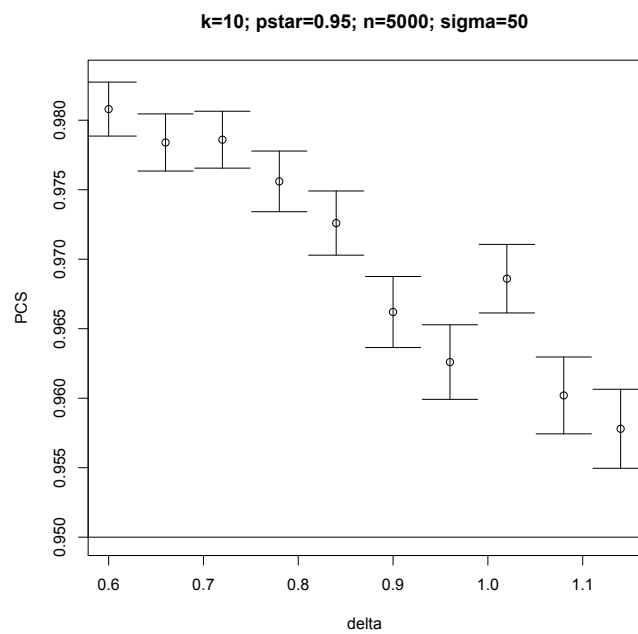
$$\mu_k - \mu_i + \frac{v_k}{\sqrt{n_{T(\delta)k}}} C_k(1, n_{tk}) - \frac{v_i}{\sqrt{n_{T(\delta)i}}} C_i(1, n_{T(\delta)i}) \Rightarrow \mu_k - \mu_i$$

$$\mathbb{P} \left\{ \frac{W_{T(\delta)k}}{n_{T(\delta)k}} < \frac{W_{T(\delta)i}}{n_{T(\delta)i}} \right\} = \mathbb{P} \left\{ \frac{W_{T(\delta)k}}{n_{T(\delta)k}} < \frac{W_{T(\delta)i}}{n_{T(\delta)i}} \right\}$$

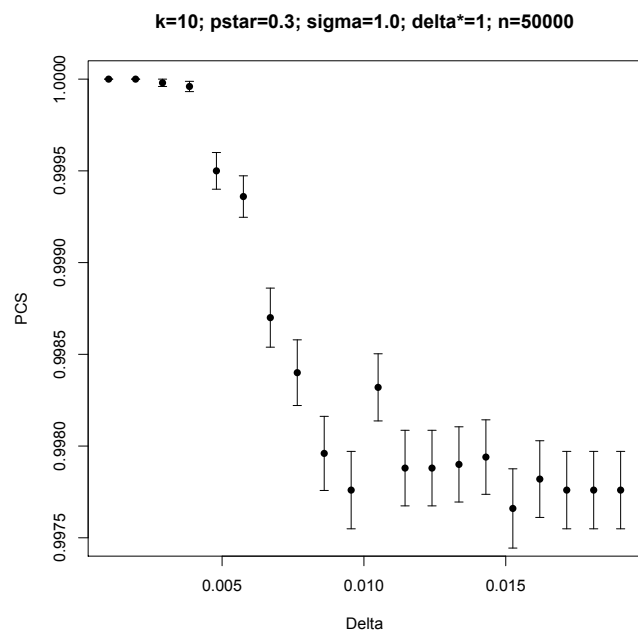
Now consider $k \geq 2$ systems and let CS be the event that k is selected and let ICS_i be the event that an incorrect selection is made when systems k and i are considered in isolation.

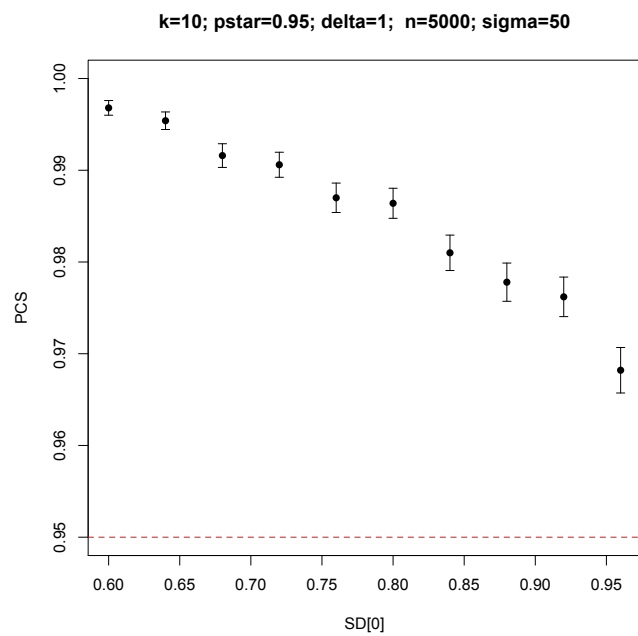
Theorem. If samples from system $x \in \{1 \dots, k\}$ are normally distributed and independent, over time and across alternatives, then $\lim_{\min_x \sigma_x^2 \rightarrow \infty} Pr \{ \text{BIZ selects } k \} \geq P^*$ provided $\mu_k \geq \mu_{k-1} + \delta^*$ for some $\delta^* > 0$.

Graphs Delta goes to 0

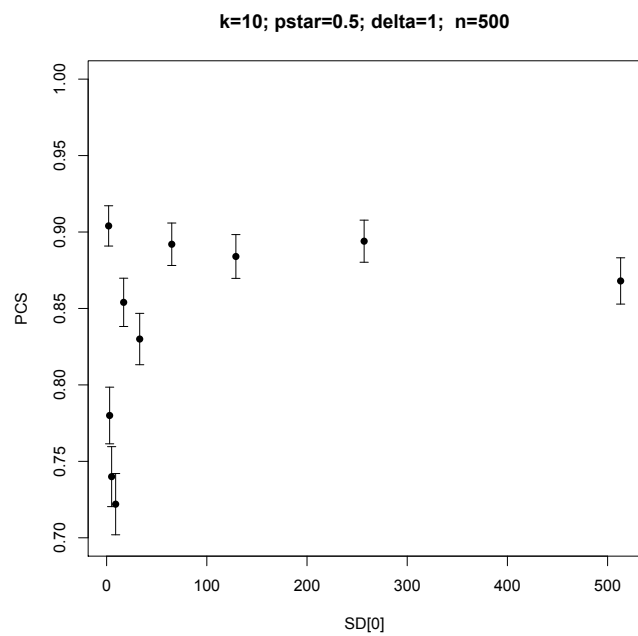


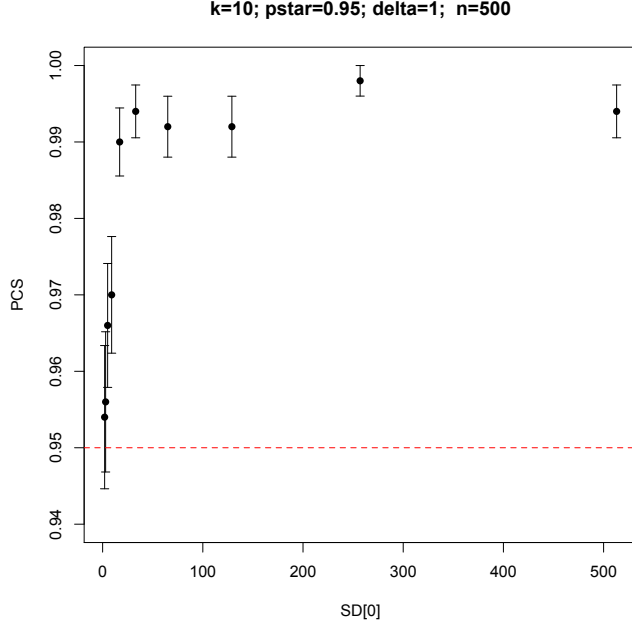
Delta goes to 0.





Variances go to infinity.





Theorem. If samples from system $x \in \{1 \dots, k\}$ are normally distributed and independent, over time and across alternatives, then $\lim_{\delta \rightarrow 0} Pr \{ \text{BIZ selects } k \} \geq P^*$ (or $=1??$) provided $\mu_k \geq \mu_{k-1} + \delta^*$.

Proof. We begin by considering the case of only two systems, denoted k and i , with $\mu_k \geq \mu_i + \delta^*$. Let $A = \{k, i\}$. For $\delta \leq \delta^*$, let

$$T(\delta) = \min \{ t \in \mathbb{N} : \min_{x \in A} \hat{q}_{tx}(A) \leq c \text{ or } \max_{x \in A} \hat{q}_{tx}(A) \geq P^* \}.$$

Thus $T(\delta)$ is the stage at which the procedure terminates.

Now,

$$\begin{aligned}
\min_{x \in A} \hat{q}_{tx}(A) &\leq c \\
\Leftrightarrow \min_{x \in A} \exp\left(\delta\beta_t \frac{W_{tx}}{n_{tx}}\right) &\leq c \sum_{x' \in A} \exp\left(\delta\beta_t \frac{W_{tx'}}{n_{tx'}}\right) \\
\Leftrightarrow \min_{x \in A} \frac{W_{tx}}{n_{tx}} &\leq \log \left[c \sum_{x' \in A} \exp\left(\delta\beta_t \frac{W_{tx'}}{n_{tx'}}\right) \right] / (\delta\beta_t) \\
\Leftrightarrow \left(\frac{W_{tk}}{n_{tk}} + \frac{W_{ti}}{n_{ti}} - \left| \frac{W_{tk}}{n_{tk}} - \frac{W_{ti}}{n_{ti}} \right| \right) &\leq 2 \log \left[c \sum_{x' \in A} \exp\left(\delta\beta_t \frac{W_{tx'}}{n_{tx'}}\right) \right] / (\delta\beta_t)
\end{aligned}$$

and

$$\begin{aligned}
\max_{x \in A} \hat{q}_{tx}(A) &\geq P * \\
\Leftrightarrow \max_{x \in A} \exp\left(\delta\beta_t \frac{W_{tx}}{n_{tx}}\right) &\geq P * \sum_{x' \in A} \exp\left(\delta\beta_t \frac{W_{tx'}}{n_{tx'}}\right) \\
\Leftrightarrow \max_{x \in A} \frac{W_{tx}}{n_{tx}} &\geq \log \left[P * \sum_{x' \in A} \exp\left(\delta\beta_t \frac{W_{tx'}}{n_{tx'}}\right) \right] / (\delta\beta_t) \\
\Leftrightarrow \left(\frac{W_{tk}}{n_{tk}} + \frac{W_{ti}}{n_{ti}} + \left| \frac{W_{tk}}{n_{tk}} - \frac{W_{ti}}{n_{ti}} \right| \right) &\geq 2 \log \left[P * \sum_{x' \in A} \exp\left(\delta\beta_t \frac{W_{tx'}}{n_{tx'}}\right) \right] / (\delta\beta_t).
\end{aligned}$$

Let $d(\delta) := 2 \log \left(\sum_{x' \in A} \exp\left(\delta\beta_t \frac{W_{tx'}}{n_{tx'}}\right) \right) / (\delta\beta_t)$, then

$$\begin{aligned}
T(\delta) = \min \left\{ t \in \mathbb{N} : \left(\frac{W_{tk}}{n_{tk}} + \frac{W_{ti}}{n_{ti}} + \left| \frac{W_{tk}}{n_{tk}} - \frac{W_{ti}}{n_{ti}} \right| \right) \geq 2 \frac{\log(P*)}{\delta\beta_t} + d(\delta) \text{ or } \right. \\
\left. \left(\frac{W_{tk}}{n_{tk}} + \frac{W_{ti}}{n_{ti}} - \left| \frac{W_{tk}}{n_{tk}} - \frac{W_{ti}}{n_{ti}} \right| \right) \leq 2 \frac{\log c}{\delta\beta_t} + d(\delta) \right\}.
\end{aligned}$$

Now, let's prove that $n_{ti} \rightarrow \infty, n_{tk} \rightarrow \infty$ as $t \rightarrow \infty$. For each t , $n_{t+1,x} = n_{tx} + B_x$ for some $x \in A$. If $n_{t+1,i} = n_{ti} + B_i$ for a finite number of t 's, then $n_{tk} \rightarrow \infty$ as $t \rightarrow \infty$. So, there exists l_0 such that $n_{li}/\hat{\lambda}_{li}^2 > n_{lk}/\hat{\lambda}_{lk}^2$ if $l > l_0$. Since $\hat{\lambda}_{ti}^2 \rightarrow \sigma_i^2$ and $\hat{\lambda}_{tk}^2 \rightarrow \sigma_k^2$ as $t \rightarrow \infty$, then $n_{ti} \rightarrow \infty$ as $t \rightarrow \infty$. Similarly, we get the same result in the other cases. Thus, $n_{ti} \rightarrow \infty, n_{tk} \rightarrow \infty$ as $t \rightarrow \infty$. Furthermore, it's easy to see that $T(\delta) \rightarrow \infty$ as $\delta \rightarrow 0$.

Now, we'll show that $T(\delta)$ is finite. Observe that there exists t_0 such that if $t > t_0$

$$W_{ti}/n_{ti} < W_{tk}/n_{tk},$$

otherwise there would exists t such that

$$\mu_k - \delta/2 < W_{tk}/n_{tk} \leq W_{ti}/n_{ti} < \delta/2 + \mu_i$$

which is a contradiction. Moreover, there exists $t_1 > t_0$ such that if $t > t_1$

$$\beta_t (W_{tk}/n_{tk} - W_{ti}/n_{ti}) > \log(P^* / (1 - P^*)) / \delta$$

since $\beta_t \rightarrow \infty$ as $t \rightarrow \infty$ and $P^* > \frac{1}{2}$. Consequently, if $t > t_1$

$$\begin{aligned} \exp\left(\delta\beta_t \frac{W_{tk}}{n_{tk}}\right) / \exp\left(\delta\beta_t \frac{W_{ti}}{n_{ti}}\right) &> \frac{P^*}{1 - P^*} \\ \Rightarrow \exp\left(\delta\beta_t \frac{W_{tk}}{n_{tk}}\right) &> P^* \exp\left(\delta\beta_t \frac{W_{ti}}{n_{ti}}\right) \\ &+ P^* \exp\left(\delta\beta_t \frac{W_{tk}}{n_{tk}}\right) \end{aligned}$$

and so $T(\delta) \leq t_1(\delta) < \infty$.

By the strong law of large numbers, almost surely

$$\frac{W_{T(\delta)k}}{n_{T(\delta)k}} - \frac{W_{T(\delta)i}}{n_{T(\delta)i}} \rightarrow \mu_k - \mu_i$$

as $\delta \rightarrow 0$. Thus,

$$\frac{W_{T(\delta)k}}{n_{T(\delta)k}} - \frac{W_{T(\delta)i}}{n_{T(\delta)i}} \Rightarrow \mu_k - \mu_i$$

as $\delta \rightarrow 0$.

Let ICS denote the event that an incorrect selection is made. Then,

$$\begin{aligned} \liminf_{\delta \rightarrow 0} \mathbb{P}\{ICS\} &= \liminf_{\delta \rightarrow 0} \mathbb{P}\left\{ \frac{W_{T(\delta)k}}{n_{T(\delta)k}} < \frac{W_{T(\delta)i}}{n_{T(\delta)i}} \right\} \\ &= \mathbb{P}\{\mu_k < \mu_i\} = 0. \end{aligned}$$