Study of Optimal Adaptive Rule in Testing Composite Hypothesis

Subir Kr. Bhandari

Bayesian and Interdisciplinary Research Unit, Indian Statistical Institute, Kolkata, India

Shyamal Krishna De*

Department of Mathematical Sciences, The University of Texas at Dallas, Richardson TX, USA

Siddhartha Mandal

Department of Biostatistics, The University of North Carolina at Chapel Hill, Chapel Hill, USA

Santanu Pradhan

Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, Kanpur, India

Biswajit Ghosh

Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, Kanpur, India

Abstract: In this paper on sequential adaptive testing, we have studied the optimal allocation between two populations for testing a composite hypothesis involving the parameters with the goal of decreasing allocation of one of the treatments to the order of the logarithm of the sample size while decreasing the probability of incorrect selection to zero. We have proved the result for large sample sizes both mathematically and by simulation studies.

Key words: Adaptive allocation; Average sample number; Composite hypothesis; Conditional log-likelihood; Log-likelihood; Probability of incorrect selection.

Subject Classifications: 93C99.

1. INTRODUCTION

In recent years, adaptive allocation has been an important topic of research in sequential testing procedures. Extensive research work has been done in adaptive sequential context [e.g. see Friedman et al. (1981), Ivanova et al. (2000), Rosenberg et al. (2001), Berry et al. (1986) etc]. Compared to fixed sample procedures, adaptive allocation reduces the average sample size required for testing a hypothesis. This procedure is important in clinical trials since it can help reduce the number of applications of the less effective drug so as to mitigate the ethical question of randomly assigning an inferior treatment to volunteers. Moreover, adaptive allocation also increases the probability of correct selection (PCS).

The main idea is to skew the allocation probabilities to alter the response history of patients and, hopefully, increase the chance of a patient to receive the treatment that performs better in the trials by more than 50%.

*Corresponding author: Department of Mathematical Sciences, EC 35, The University of Texas at Dallas, 800 West Campbell Road, Richardson, TX 75080-3021, USA; Fax: (972)883-6622; E-mail: shyamalkd@gmail.com.

In existing literature, many types of allocation rules have been considered. Historically, equal allocation has been used because this invariant fixed sample rule turns out to be computationally easier whereas adaptive allocation turns out better not only from the ethical point of view but may also result in minimal losses or even gains in power for experiments comparing the two treatments.

Adaptive designs for clinical trials have been largely based on urn models. A popular method is the Randomized-Play-The-Winner (RPTW) rule. In this method, an urn contains balls representing two treatments (say, A and B). Firstly, a patient is assigned a treatment A(B) according to the type of ball drawn. A success of a treatment results in the addition of a ball of same type to the urn while failure results in addition of a ball of opposite color. Let $N_{A,n}$ be the number of patients assigned to treatment A after n patients are assigned and $N_{B,n} = n - N_{A,n}$. If p_A =P(success of treatment A) and p_B =P(success of treatment B), $q_A = 1 - p_A$, $q_B = 1 - p_B$, then the limiting allocation is q_A/q_B, a measure of relative risk, as $n \to \infty$.

In Randomized-Play-The-Winner rule and other rules prior to Bhandari et al. (2007), the ratio of allocation to the two treatments tends to a positive constant bounded away from zero as the total sample number n tends to infinity. For the case of testing simple hypothesis in parametric setup, Bhandari et al. (2007) has given Procedure I and Procedure II in which the said ratio tends to zero. Moreover, the number of allocation to the less effective treatment tends to a finite number even if the total sample number tends to infinity.

In this paper, we have considered the testing of composite hypotheses between any two populations with sequential adaptive allocation with a fixed number of patients, and we have successfully shown that we can make the expected sample size of one of the treatments to the order of $\log n$, i.e., $E(N_{A,n})/\log n$ tends to a finite number both mathematically and by simulation studies. All the procedures used here are permutation invariant and the allocation at each stage is based on some function of the existing likelihood ratio or some other estimates arising out of it. At each stage, we determine the optimal decision and allocate to the population which is preferred.

2. PRELIMINARIES : SIMPLE AND COMPOSITE HYPOTHESIS

Two populations with probability density functions f_{θ_0}, f_{θ_1} (with respect to some σ-finite measure) ($f_{\theta_0} \neq f_{\theta_1}$) are considered. The total sample size from the two populations is N which is preassigned. We shall draw samples one by one in such a way that the population to be chosen at some stage depends upon previous sample observations. At each stage n, let $N_{0,n}$ and $N_{1,n}$ be the number of units drawn from f_{θ_0} and f_{θ_1} respectively and $N_{0,n} + N_{1,n} = n$. We stop when $n = N$ and take a decision according to an appropriate rule.

The simple hypothesis in this case would be H_0: $(f_{\theta_0}, f_{\theta_1})=(f_0, f_1)$ Vs H_1: $(f_{\theta_0}, f_{\theta_1})=(f_1, f_0)$. The composite hypothesis in the above mentioned parametric setup would be H_0: $\theta_0 > \theta_1$ against H_1: $\theta_1 > \theta_0$.

In this case, consider

$$LR_n(0,1) = log[L_n(\underline{x}, H_0)/L_n(\underline{x}, H_1)] \tag{1}$$

$$LR_n(0,1|N_{0,n}, N_{1,n}) = log[L_n(\underline{x}, H_0|N_{0,n}, N_{1,n})/L_n(\underline{x}, H_1|N_{0,n}, N_{1,n})] \tag{2}$$

where under H_i, $L_n(\underline{x}, H_i)$ and $L_n(\underline{x}, H_i|N_{0,n}, N_{1,n})$ are the likelihood and conditional likelihood respectively for the observations $\underline{x}$ and given sample numbers $N_{0,n}$ and $N_{1,n}$ at the stage n. We describe the following procedures for testing the above stated simple and composite hypothesis.

2.1. Procedure I

We consider likelihood ratio $LR_n(0,1)$ as defined in (1) and at each stage n, we adopt the following allocation rule:

if	i) $LR_n(0,1) > 0$	we increase $N_{0,n}$ by 1.
if	ii) $LR_n(0,1) < 0$	we increase $N_{1,n}$ by 1.
if	iii) $LR_n(0,1) = 0$	1 unit observation is allocated to each population with probability $\frac{1}{2}$ each.

Finally, when $n = N$ we accept H_0 with probability 1 if $LR_n(0,1) > 0$ and with probability $\frac{1}{2}$ if $LR_n(0,1) = 0$.

Now, for implementing this procedure we need to know the probability distributions of $N_{0,n}$ and $N_{1,n}$ which may be very cumbersome leading to computational problems. Further, for given values of $N_{0,n}$ and $N_{1,n}$, the conditional distribution of the unit observations in $\underline{x}$ becomes independent. Hence, it is easier to use $LR_n(0,1|N_{0,n}, N_{1,n})$ in place of the usual log-likelihood ratio.

2.2. Procedure II

For the same testing problem as in Procedure I, to define Procedure II we use the similar allocation rule and final rejection/acceptance rule as in Procedure I, except that we use $LR_n(0,1|N_{0,n}, N_{1,n})$ in place of $LR_n(0,1)$. The rationale behind using the conditional likelihoods in the description of Procedure II is that it not only makes computations simple but also gives consistency and results very close to the optimal.

Both the above procedures were extensively studied in Bhandari et al. (2007) in the context of testing the simple hypothesis in sequential adaptive test procedures. Now for testing the composite hypothesis, we state the following procedure which is an extension of Procedure I using LR_n with plug-in values of $\hat{\theta}_{0,n}$ and $\hat{\theta}_{1,n}$.

2.3. Procedure III

To test the composite hypothesis, we need information about the unknown parameters θ_0 and θ_1 in the Likelihood ratio function. So, we consider some suitably chosen efficient estimators $\hat{\theta}_{0,n}$, $\hat{\theta}_{1,n}$ (suitably chosen function of the order statistics in the relevant sample) of θ_0 and θ_1 respectively at each stage n with certain properties:

i) $E(\hat{\theta}_{0,n}) = \theta_0$, $E(\hat{\theta}_{1,n}) = \theta_1$.

ii) $V(\hat{\theta}_{0,n}) \simeq c_0/N_{0,n}$, $V(\hat{\theta}_{1,n}) \simeq c_1/N_{1,n}$.

iii) $\hat{\theta}_{0,n}$, $\hat{\theta}_{1,n}$ are asymptotically normal random variables having same support which contain Θ, the set of all parameters. Here, c_0 and c_1 are positive constants, and $N_{0,n}$ and $N_{1,n}$ are number of observations from f_{θ_0} and f_{θ_1} at stage n respectively. We also assume that θ_0 or θ_1 are not supremum or infimum of the set Θ.

Now, we adopt the following allocation rule which is equivalent to using LR_n with plug-in values $\hat{\theta}_{0,n}$ and $\hat{\theta}_{1,n}$:

if	i) $\hat{\theta}_{0,n} - \hat{\theta}_{1,n} > 0$	we increase $N_{0,n}$ by 1.
if	ii) $\hat{\theta}_{0,n} - \hat{\theta}_{1,n} < 0$	we increase $N_{1,n}$ by 1.
if	iii) $\hat{\theta}_{0,n} - \hat{\theta}_{1,n} = 0$	1 unit observation is allocated to each population with probability $\frac{1}{2}$ each.

Finally, when $n = N$ we accept H_0 with probability 1 if $\hat{\theta}_{0,n}$ - $\hat{\theta}_{1,n} > 0$ and with probability $\frac{1}{2}$ if $\hat{\theta}_{0,n}$ - $\hat{\theta}_{1,n} = 0$. The initial values of $N_{0,n}$ and $N_{1,n}$ are taken to be $M(N)$ depending on N.

With Procedure III, as defined above we prove the following Lemma.

Lemma 2.1. *Under H_0, as $n \to \infty$, both $N_{0,n}$ and $N_{1,n}$ will tend to infinity and $\frac{N_{1,n}}{N_{0,n}} \to 0$ a.e. as $PICS_n \to 0$ where $M = o(logN)$ and $M \to \infty$.*

Proof. Clearly, $N_{0,n}$ and $N_{1,n}$ tends to infinity as $M \to \infty$ and by Lemma 5.2 in the Appendix we have $PICS_n \to 0$. Interpolating E($N_{1,n}$) as smooth increasing function of n we obtain $PICS_n \simeq \frac{dE(N_{1,n})}{dn}$ (see proof of Theorem 3.1). For any $\delta > 0$, a simple application of Markov's inequality and L'Hospital rule yields

$$P(N_{1,n} > n\delta) \leq \frac{E(N_{1,n})}{n\delta} \to 0 \quad \text{as } PICS_n \to 0$$

which implies

$$\frac{N_{1,n}}{n} \to 0 \text{ a.e., i.e., } \frac{N_{1,n}}{n - N_{1,n}} \to 0 \text{ a.e.}$$

Therefore, $\frac{N_{1,n}}{N_{0,n}} \to 0$ a.e. □

3. MAIN RESULT

Our main result is stated in the following theorem.

Theorem 3.1. *With the Procedure III as defined above with two populations having densities f_{θ_0}, f_{θ_1} and having estimators $\hat{\theta}_{0,n}$ and $\hat{\theta}_{1,n}$ of θ_0, θ_1 respectively and satisfying properties (i),(ii),(iii), we have as $N \to \infty$, $M \to \infty$ and $\frac{M}{logN} \to 0$:*

$$\frac{E(N_{1,n})}{logN} \to \quad \textit{a positive constant } c^0 \textit{(say) under } H_0$$

and

$$\frac{E(N_{0,n})}{logN} \to \quad \textit{a positive constant } c^1 \textit{(say) under } H_1.$$

Proof. Under H_0, i.e., $\theta_0 > \theta_1$ and by Lemma 2.1,

$$\begin{aligned} PICS_n &= P(\hat{\theta}_{0,n} - \hat{\theta}_{1,n} < 0) \\ &= E_{N_{0,n},N_{1,n}}[P(\overline{\hat{\theta}_{0,n} - \theta_0} - \overline{\hat{\theta}_{1,n} - \theta_1} < \theta_1 - \theta_0 | N_{0,n}, N_{1,n}] \\ &\simeq E_{N_{0,n},N_{1,n}} \left\{ \Phi \left(\frac{\theta_1 - \theta_0}{\sqrt{\frac{c_0}{N_{0,n}} + \frac{c_1}{N_{1,n}}}} \right) \right\} \end{aligned}$$

since $\left(\overline{\hat{\theta}_{0,n} - \theta_0} - \overline{\hat{\theta}_{1,n} - \theta_1}\right) \sim N\left(0, \frac{c_0}{N_{0,n}} + \frac{c_1}{N_{1,n}}\right)$ asymptotically by C.L.T.

With $x > 0$, we have the inequality

$$\left(\frac{1}{x} - \frac{1}{x^3}\right)\phi(x) \leq \Phi(-x) \leq \frac{1}{x}\phi(x) \tag{3}$$

Now,

$$\sqrt{\frac{c_1}{N_{1,n}}} \leq \sqrt{\frac{c_0}{N_{0,n}} + \frac{c_1}{N_{1,n}}} \leq max\left\{ \sqrt{\frac{2c_1}{N_{1,n}}}, \sqrt{\frac{2c_0}{N_{0,n}}} \right\}.$$

Thus,

$$\frac{(\theta_1-\theta_0)}{\sqrt{\frac{c_1}{N_{1,n}}}} \leq \frac{(\theta_1-\theta_0)}{\sqrt{\frac{c_0}{N_{0,n}}+\frac{c_1}{N_{1,n}}}} \leq max\left\{\frac{(\theta_1-\theta_0)}{\sqrt{\frac{2c_1}{N_{1,n}}}}, \frac{(\theta_1-\theta_0)}{\sqrt{\frac{2c_0}{N_{0,n}}}}\right\} \quad \text{as } \theta_0 > \theta_1.$$

With the above inequality and (3), we have

$$\begin{aligned}
&E_{N_{0,n},N_{1,n}}\left[\left\{\frac{\sqrt{\frac{c_1}{N_{1,n}}}}{(\theta_0-\theta_1)}-\left(\frac{\sqrt{\frac{c_1}{N_{1,n}}}}{(\theta_0-\theta_1)}\right)^3\right\}\phi\left(\frac{(\theta_1-\theta_0)}{\sqrt{\frac{c_1}{N_{1,n}}}}\right)\right] \\
&\quad \leq PICS_n \\
&\quad \leq E_{N_{0,n},N_{1,n}}\left[\frac{\sqrt{\frac{2c_1}{N_{1,n}}}}{(\theta_0-\theta_1)}\phi\left(\frac{(\theta_1-\theta_0)}{\sqrt{\frac{2c_1}{N_{1,n}}}}\right)\right]+E_{N_{0,n},N_{1,n}}\left[\frac{\sqrt{\frac{2c_0}{N_{0,n}}}}{(\theta_0-\theta_1)}\phi\left(\frac{(\theta_1-\theta_0)}{\sqrt{\frac{2c_0}{N_{0,n}}}}\right)\right] \\
&\quad \leq 2E_{N_{0,n},N_{1,n}}\left[\frac{\sqrt{\frac{2c_1}{N_{1,n}}}}{(\theta_0-\theta_1)}\phi\left(\frac{(\theta_1-\theta_0)}{\sqrt{\frac{2c_1}{N_{1,n}}}}\right)\right]
\end{aligned}$$

since $\frac{N_{1,n}}{N_{0,n}} \to 0$ a.e. and both $N_{0,n}$ and $N_{1,n} \to \infty$ by Lemma 2.1.

This implies,

$$E_{N_{0,n},N_{1,n}}\Big[O(e^{-k_1 N_{1,n}})\Big] \leq PICS_n \leq E_{N_{0,n},N_{1,n}}\Big[O(e^{-k_2 N_{1,n}})\Big],$$

where $k_1 = \frac{(\theta_1-\theta_0)^2}{2c_1} > 0$ and $k_2 = \frac{(\theta_1-\theta_0)^2}{4c_1} > 0$.

Therefore, applying Jensen's inequality

$$PICS_n = E_{N_{0,n},N_{1,n}}\Big[e^{-k_0 N_{1,n}}\Big] \geq e^{-k_0 E(N_{1,n})}, \tag{4}$$

where k_0 is some suitable constant.

We have $E(N_{1,n}) = K + \sum_{i=n_0}^{n} PICS_i$; n_0 is a small integer and K is a constant.
Approximating $PICS_i$ as a continuous decreasing function of i, we can write

$$E(N_{1,n}) \simeq K + \int_{n_0}^{n} PICS_m dm. \tag{5}$$

Using (4) and (5), we obtain

$$PICS_n \simeq \frac{dE(N_{1,n})}{dn} \geq e^{-k_0 E(N_{1,n})}$$

which implies

$$\int_{n_0}^{n} e^{k_0 E(N_{1,m})} dE(N_{1,m}) \geq \int_{n_0}^{n} dm$$

i.e.,

$$e^{k_0 E(N_{1,n})} - e^{k_0 E(N_{1,n_0})} \geq k_0(n-n_0).$$

It follows,

$$k_3 e^{k_0 E(N_{1,n})} \geq k_4 n,$$

where k_3 and k_4 are suitable constants. Hence,

$$E(N_{1,n}) \geq \alpha log(\beta n) \text{ for some constants } \alpha \text{ and } \beta.$$

Let $\lambda_n = E(N_{1,n})$. So, $\lambda_n \to \infty$ as $n \to \infty$. Also, λ_n is increasing in n.

Using Lemma 5.1 in appendix, we have $(1+\delta)\lambda_n = (1+\delta)E(N_{1,n}) \geq V(N_{1,n})$.

Define an event $A_{\lambda_n} = \left\{ \lambda_n - (\lambda_n)^{\frac{1}{4}}\sqrt{\lambda_n} \leq N_{1,n} \leq \lambda_n + (\lambda_n)^{\frac{1}{4}}\sqrt{\lambda_n} \right\}$.

Therefore, $A^c_{\lambda_n} = \left\{ |N_{1,n} - \lambda_n| \geq (\lambda_n)^{\frac{1}{4}}\sqrt{\lambda_n} \right\}$.

Applying Chebyshev's inequality, we obtain

$$P(A^c_{\lambda_n}) \leq \frac{V(N_{1,n})}{(\lambda_n)^{\frac{3}{2}}} \leq (1+\delta)\frac{\lambda_n}{(\lambda_n)^{\frac{3}{2}}} = \frac{1+\delta}{\sqrt{\lambda_n}}.$$

Now, define A_x as $A_{\lambda_{n_x}}$ where $\lambda_{n_x} \geq x$ with n_x being minimum integer , x $\in \mathbb{R}$.

Therefore, $P\left(\bigcup_{t=[\lambda_{n^*}]}^{\infty} A^c_{t^4} \right) \leq \sum_{t=[\lambda_{n^*}]}^{\infty} \frac{(1+\delta)^4}{t^2} \to 0$ as $n^* \to \infty$.

So, with suitable m_i as $i \to \infty, \lambda_{m_i} = \lambda_{n_i{}^4} \simeq i^4 \uparrow \infty$ and $m_i \to \infty$ implies $\frac{N_{1,m_i}}{E(N_{1,m_i})} \to 1$ a.s.
Let $m_i \leq k \leq m_{i+1}$. So,

$$\frac{N_{1,m_i}}{E(N_{1,m_{i+1}})} \leq \frac{N_{1,k}}{E(N_{1,k})} \leq \frac{N_{1,m_{i+1}}}{E(N_{1,m_i})}.$$

Now,

$$\frac{N_{1,m_{i+1}}}{E(N_{1,m_i})} = \frac{E(N_{1,m_{i+1}})}{E(N_{1,m_i})}\frac{N_{1,m_{i+1}}}{E(N_{1,m_{i+1}})} \to 1 \quad \text{a.s.}$$

since $\frac{E(N_{1,m_{i+1}})}{E(N_{1,m_i})} \simeq \frac{(i+1)^4}{i^4} \to 1$ as $i \to \infty$.
Similarly, $\frac{N_{1,m_i}}{E(N_{1,m_{i+1}})} \to 1$ as $i \to \infty$ almost surely. It follows, $\frac{N_{1,k}}{E(N_{1,k})} \to 1$ as $k \to \infty$ a.s.

Hence, $N_{1,n} = O(\lambda_n)$ a.s.

From (4), we have $PICS_n = E\left(e^{-k_0 N_{1,n}}\right) = E\left(e^{-\delta\lambda_n}\right) = e^{-\delta E(N_{1,n})}$ for some constant $\delta > 0$.

Again, approximating $PICS_n$ as a smooth continuous decreasing function of n, we write

$$PICS_n \simeq \frac{d\lambda_n}{dn} = e^{-\delta\lambda_n}$$

which implies

$$\int_{n_0}^{n} e^{\delta\lambda_m} d\lambda_m = \int_{n_0}^{n} dm$$

i.e.,

$$e^{\delta\lambda_n} - e^{\delta\lambda_{n_0}} = \delta(n - n_0).$$

Hence, for some δ_1 and δ_2 suitable constants, we obtain

$$\delta_1 e^{\delta\lambda_n} = \delta_2 n$$

Therefore,

$$\lambda_n = E(N_{1,n}) = c^0 log(\delta_3 n),$$

where $c^0 = \frac{1}{\delta}$ and $\delta_3 = \frac{\delta_2}{\delta_1}$. This leads to our main result $\frac{E(N_{1,n})}{logn} \to c^0$ as $n \to \infty$.

Similarly, it can be shown that $\frac{E(N_{0,n})}{logn} \to c^1$, a positive constant under H_1. □

4. TABLES OF SIMULATION

In this section, we provide simulation results to observe the limiting values of $\frac{E(N_{1,n})}{logn}$ for large n and to support our main result in section 3. We implement Procedure III in testing composite hypothesis of our interest in section 2 for different pairs of (p_0, p_1) in the case of Bernoulli trials and for different pairs of (μ_0, μ_1) in the context of $N(\mu, 1)$ populations.

For each (p_0, p_1) pair and each total sample number n, we apply Procedure III using some simulation technique and iterate 500 times. We compute probability of correct selection (PCS) as proportion of correct decision among 500 trials. Tables 1, 2, and 3 include simulated results for different pairs of (p_0, p_1) and n in Bernoulli set up. Simulated results for different (μ_0, μ_1) pairs and n in normal set up are shown in tables 4 and 5. In the 4^{th} and 8^{th} column of each table, we include estimated limiting values of $\frac{E(N_{1,n})}{logn}$ for large n with some consideration of standard error in simulation. However, these estimates are obtained by averaging 10 individual estimates to further increase the precision. In this simulation study, we note that the samples assigned to the medicine with lower value of p or μ tends to a positive constant when divided by logarithm of the total sample number.

At the beginning of each of the 500 loops, we start with fixed, constant, equal values of $N_{0,n}$ and $N_{1,n}$ depending on the total sample number so that the small constant values do not affect the asymptotic properties. At each stage of application of Procedure III, we take the usual maximum likelihood estimate of p and μ.

Remark. From Lemma 5.1 in the appendix, an approximate upper bound of $V(N_{1,n})$ is $E(N_{1,n})(1+\delta)$ for small $\delta > 0$. In the tables of section 4, the estimates of $\frac{E(N_{1,n})}{logn}$, denoted as $\frac{\widehat{E(N_{1,n})}}{logn}$, are obtained from 5000 iterations. The standard deviation of $\frac{\widehat{E(N_{1,n})}}{logn}$ is given in percentage as

$$\left(\frac{\sqrt{V\left[\frac{\widehat{E(N_{1,n})}}{logn}\right]}}{\frac{E(N_{1,n})}{logn}} \right) 100\% = \left(\frac{\sqrt{\frac{V(N_{1,n})}{5000}}}{E(N_{1,n})} \right) 100\%$$

Hence, an approximate upper bound of the standard deviation of the estimates in percentage would be $\sqrt{\frac{(1+\delta)}{5000E(N_{1,n})}}100\%$. This implies that an approximate upper bound of the standard error of estimates of $\frac{E(N_{1,n})}{logn}$ is less than 1%. Such an upper bound helps to get an idea about the limiting values of the quantities as n tends to ∞.

4.1. Tables of Binomial Simulation

Table 1: Simulated results for Bernoulli probability pairs (p_0, p_1)=(0.7, 0.4) and (p_0, p_1)=(0.65, 0.45)

pair	n	PCS	$E(N_{1,n})/logn$	pair	n	PCS	$E(N_{1,n})/logn$
(0.7, 0.4)	25	0.98	3.10834	(0.65, 0.45)	30	0.83	3.25786
	30	0.97	2.97910		55	0.92	3.40195
	35	0.96	2.89738		90	0.95	3.42370
	40	0.98	2.80564		110	0.97	3.51814
	50	0.99	2.66729		130	0.98	3.52644
	55	0.99	2.62035		150	0.98	3.39577
	65	0.99	2.50487		170	0.99	3.41673
	75	0.99	2.41565		190	0.99	3.33941
	80	0.99	2.39041		220	0.99	3.27872
	90	0.99	2.33294		250	0.99	3.33849
	100	0.99	2.26462		300	0.99	3.35803
	110	0.99	2.21267		400	0.99	3.26250
	120	0.99	2.17687		500	0.99	3.35967

Table 2: Simulated results for Bernoulli probability pairs (p_0, p_1)=(0.7, 0.5) and (p_0, p_1)=(0.65,0.5)

pair	n	PCS	$E(N_{1,n})/logn$	pair	n	PCS	$E(N_{1,n})/logn$
(0.7, 0.5)	30	0.84	3.23979	(0.65, 0.5)	30	0.84	3.24419
	40	0.87	3.37006		40	0.88	3.33386
	55	0.91	3.48199		75	0.95	3.44253
	90	0.95	3.47010		90	0.95	3.50518
	100	0.99	3.54686		100	0.96	3.50990
	110	0.96	3.55521		110	0.97	3.42913
	130	0.98	3.49544		130	0.98	3.52382
	140	0.97	3.53382		140	0.98	3.46321
	150	0.98	3.54348		160	0.98	3.52186
	160	0.98	3.47132		170	0.99	3.39444
	170	0.99	3.42871		180	0.98	3.30910
	180	0.99	3.44402		200	0.98	3.46892

Table 3: Continuation of table 2 for pairs (p_0, p_1)=(0.7, 0.5) and (p_0, p_1)=(0.65, 0.5)

pair	n	PCS	$E(N_{1,n})/logn$	pair	n	PCS	$E(N_{1,n})/logn$
(0.7, 0.5)	190	0.99	3.41115	(0.65, 0.5)	220	0.99	3.47242
	200	0.99	3.40578		250	0.99	3.23987
	220	0.99	3.33686		275	0.99	3.30053
	250	0.98	3.52462		300	0.99	3.22519
	275	0.99	3.35387		350	0.99	3.15934
	300	0.99	3.22278		400	0.99	3.26372
	400	0.99	3.27865		500	0.99	3.25120

4.2. Tables of Normal Simulation

Table 4: Simulated results for normal mean pairs (μ_0, μ_1)=(0.7,0.3) and (μ_0, μ_1)=(0.8,0.2)

pair	n	PCS	$E(N_{1,n})/logn$	pair	n	PCS	$E(N_{1,n})/logn$
(0.7, 0.3)	25	0.96	3.12819	(0.8, 0.2)	25	0.99	3.10122
	30	0.92	3.04883		30	0.98	2.94865
	40	0.91	3.07554		35	0.98	2.83686
	55	0.94	3.08183		40	0.98	2.76555
	75	0.97	3.08618		45	0.99	2.70059
	90	0.95	3.16842		50	0.96	2.66852
	100	0.97	3.15280		55	0.97	2.65765
	110	0.97	3.16917		60	0.97	2.61195
	120	0.97	3.20783		65	0.98	2.57846
	130	0.97	3.20319		70	0.99	2.53912
	140	0.97	3.20208		75	0.99	2.51435
	150	0.97	3.22822		80	0.97	2.49531
	160	0.97	3.24329		85	0.97	2.48463
	170	0.97	3.31301		90	0.98	2.47081
	180	0.98	3.30345		95	0.99	2.44051
	190	0.98	3.11181		100	0.99	2.41974
	200	0.97	3.31959		110	0.99	2.39976
	220	0.97	3.28628		120	0.99	2.34502
	250	0.97	3.47673		150	0.99	2.24720

Table 5: Simulated results for normal mean pairs (μ_0, μ_1)=(1.5,1.0) and (μ_0, μ_1)=(1.35,0.9)

pair	n	PCS	$E(N_{1,n})/logn$	pair	n	PCS	$E(N_{1,n})/logn$
(1.5, 1.0)	25	0.99	3.03449	(1.35, 0.9)	30	0.94	2.99894
	30	0.95	2.97259		35	0.93	2.96149
	35	0.94	2.88975		40	0.94	2.92653
	40	0.97	2.83314		45	0.96	2.92784
	45	0.98	2.78061		50	0.93	2.91960
	50	0.95	2.76179		55	0.96	2.87731
	55	0.96	2.76310		60	0.95	2.89460
	60	0.97	2.73026		65	0.96	2.92048
	65	0.98	2.71349		70	0.98	2.89782
	70	0.99	2.71426		75	0.98	2.85128
	75	0.99	2.62937		80	0.96	2.86837
	80	0.97	2.66634		85	0.96	2.85865
	85	0.98	2.59304		90	0.97	2.87939
	90	0.98	2.64287		95	0.98	2.90915
	95	0.99	2.63393		100	0.98	2.87150
	100	0.99	2.55961		110	0.98	2.84377
	110	0.99	2.57603		120	0.98	2.81239
	120	0.99	2.56636		150	0.99	2.80481
	150	0.99	2.50198		200	0.99	2.79036

5. APPENDIX

We have used the following results in previous sections.

Lemma 5.1. *Under H_0, with the previously stated Procedure III for composite hypotheses for large values of n, $E(N_{1,n})(1+\delta) \geq V(N_{1,n})$ with small $\delta > 0$.*

Proof. Define,

$$X_i = \begin{cases} 1 & \text{if } \hat{\theta}_{0,i} - \hat{\theta}_{1,i} < 0 \text{ at } i^{th} \text{ trial} \\ 0 & o.w. \end{cases}$$

Therefore, $N_{1,n} = X_1 + X_2 + \cdots + X_n$. For $j > i$, consider the distribution of $(X_i, X_j) \mid X_1, \cdots, X_{i-1}, X_{i+1}, \cdots, X_{j-1}$. Let us assume $(X_1, \cdots, X_{i-1}, X_{i+1}, \cdots, X_{j-1}) \sim f$ and consider n is large and also i, j are large with $i, j \geq M$.
Following is the 2×2 table for the joint distribution of (X_i, X_j) :

$x_i \diagdown x_j$	0	1	total
0	a	b	$1-(c+d)$
1	c	d	$c+d$
total	$1-(b+d)$	$b+d$	1

where $a = P(X_i = 0, X_j = 0)$, $b = P(X_i = 0, X_j = 1)$, $c = P(X_i = 1, X_j = 0)$, and $d = P(X_i = 1, X_j = 1)$.

We have $PICS_n = P((\hat{\theta}_{0,n} - \theta_0) - (\hat{\theta}_{1,n} - \theta_1) < \theta_1 - \theta_0)$. Clearly, as $N_{1,n}$ decreases $Var((\hat{\theta}_{0,n} - \theta_0) - (\hat{\theta}_{1,n} - \theta_1))$ increases which implies increase in $PICS_n$ as well. This statement is justified by property (iii) of $\hat{\theta}$ and Lemma 2.1. $\hat{\theta}_{0,n}$ and $\hat{\theta}_{1,n}$ are functions of the order statistics of the relevant samples. Now, if X_i changes from 1 to 0, i.e., upto $(j-1)^{th}$ trial $N_{1,n}$ decreases, then $X_j = 1$ is more likely to occur in the next trial than $X_j = 0$. So, it is obvious that $P(X_i = 1, X_j = 0) \geq P(X_i = 0, X_j = 0)$, i.e., $c \geq a$. Similarly, $b \geq d$.

Now, from the contingency table we have, $E(X_i X_j) = d$, $E(X_i) = c + d$, and $E(X_j) = b + d$. Therefore,

$$E(X_i)E(X_j) - E(X_i X_j) = (c+d)(b+d) - d = bc - ad \geq 0$$

i.e., $Cov(X_i, X_j) \leq 0$. Hence,

$$\begin{aligned} V(N_{1,n}) &\leq (1+\delta)\sum_{k=1}^{n} V(X_k) \\ &\leq (1+\delta)\sum_{k=1}^{n} E({X_k}^2) \\ &= (1+\delta)\sum_{k=1}^{n} E(X_k) \\ &= (1+\delta)E(N_{1,n}). \end{aligned}$$

This holds for large values of n as we assumed in the beginning of the lemma. □

Lemma 5.2. *Under H_0 for large n, $N_{0,n}$ and $N_{1,n}$ both tend to infinity implies $PICS_n \to 0$.*

Proof. Under H_0, $PICS_n \simeq E_{N_{0,n},N_{1,n}} \left\{ \Phi \left(\frac{(\theta_1 - \theta_0)}{\sqrt{\frac{c_0}{N_{0,n}} + \frac{c_1}{N_{1,n}}}} \right) \right\}$ (see proof of Theorem 3.1).

Now, under H_0, $\frac{(\theta_1 - \theta_0)}{\sqrt{\frac{c_0}{N_{0,n}} + \frac{c_1}{N_{1,n}}}} \to -\infty$ as $n \to \infty$, which implies

$$E_{N_{0,n},N_{1,n}} \left\{ \Phi \left(\frac{(\theta_1 - \theta_0)}{\sqrt{\frac{c_0}{N_{0,n}} + \frac{c_1}{N_{1,n}}}} \mid N_{0,n}, N_{1,n} \right) \right\} \to 0 \quad \text{as } n \to \infty$$

i.e., $PICS_n \to 0$ as $n \to \infty$. □

ACKNOWLEDGEMENT

We appreciate the comments made by the Associate Editor and the referee.

REFERENCES

Berry, D. A. and Fristedt, B. (1986). *Bandit Problems: Sequential Allocation of Experiments*, London: Chapman and Hall.

Bhandari, S. K., Dutta, R., and Guha Niyogi, R. (2007). Study of Optimal Adaptive Rule in Testing Problem, In *ISI Platinum Jubilee Conference on Multivariate Analysis*, Singapore: World Scientific.

Friedman, L. M., Furgberg, C. D., and DeMets, D. L. (1981). *Fundamentals of Clinical Trials*, third edition, New York: Springer.

Ivanova, A. and Rosenberger, W. F. (2000). A Comparison of Urn Designs for Randomized Clinical Trials of $K > 2$ Treatments, *Journal of Biopharmaceutical Statistics* 10: 93-107.

Rosenberger, William F., Stallard, Nigel, Ivanova, Anastasia, Harper, Cherice N., Ricks, and Michelle L. (2001). Optimal Adaptive Designs for Binary Response Trials, *Biometrics* 57: No.3, 909-913.