Asymptotic Optimality of Three Stage Design for Estimating Product of Means with Applications in Reliability Estimation and Risk Assessment

Xing Song* and Kamel Rekab

Department of Mathematics and Statistics, University of Missouri-Kansas City

Abstract: The one-parameter exponential family is a practically convenient and widely used unified family of distributions, which contains both discrete and continuous distributions that can be used for practical modelling, such as the Binomial, Beta, Normal, etc. The problem of estimating product of means has been explored for independent populations from one-parameter exponential family in a general sense, with a three-stage sampling design proposed and proven to be first-order efficient. The purpose of this paper is to apply the theoretical results to specific applications and to provide practical guidance on implementing the proposed sequential design. One popular application problem of interest is to estimate the system reliability, for which a Beta-Binomial model will be adopted. The other practical problem, which is often encountered in environmental study, is risk assessment and a Normal-Normal model will be used for the case.

Keywords: One-parameter exponential family, first-order optimality, three-stage sampling scheme, sequential design, applications, Beta-Binomial, Normal-Normal, system reliability, risk assessment, bayesian estimation.

1. INTRODUCTION

As pointed out by Hardwick and Stout, the sequential adaptive sampling scheme is more efficient than fixed design, since the former allows the opportunity to recover from possible mistakes by learning from the accruing data [6]. Fully sequential design, in which the data is monitored one at a time, has been extensively studied and proven to be very powerful. However, there are difficulties with implementation on a fully sequential design and could be practically undesirable due to the cumulating delay in response and set-up costs. A three-stage design was first addressed by Woodroofe for estimating the difference between two normal means with a quasi-Bayesian approach [17]. It was also adopted in the work of Benkamra Z, *et al*. for estimating a product of several Bernoulli proportions in a frequentist framework [1]. Most of the distributions of concern in the previous studies belong to the one-parameter exponential family, which motivates a further investigation for the efficiency of a three-stage sampling scheme to the entire one-parameter exponential family [14].

From a fully Bayesian perspective, the three-stage sampling design has not yet been thoroughly exploited for estimating product of means, a problem which has been of interest in many applications of engineering [2] [12], economics [7] and environmental studies [18], to name a few. A fully Bayesian approach incorporates available prior information both in defining the risk and designing the sampling procedure. Among various types of priors, either being purely subjective assessment of an experienced expert or conforming to some principles, the conjugate priors, if exist, turn out to be an adequate choice for providing analytical tractable solutions.

In this article, two specific models will be carefully examined under fully Bayesian framework with conjugate priors: a Beta-Binomial model used for reliability estimation in the context of software engineering [11]; and a Normal-Normal model which is usually employed for assessing the risk due to exposure to radiation or various pollutants [16]. For both applications, the three-stage sampling design will be adopted and justified to be first-order efficient theoretically as well as through Monte Carlo simulations.

2. APPLICATION I: SYSTEM RELIABILITY ESTIMATION—BETA-BINOMIAL MODEL

The first application is to estimate the reliability of a system by adopting a Beta-Binomial model, where the data collected are discrete values. Reliability is an important factor of any system design since any user of the system would expect some type of guarantee that the system will function to some level of confidence, which is especially true in such critical systems that the tolerance of failure is on the order of 10^{-3} percent or even smaller [9]. The most fundamental system is a series system, which functions if and only if every individual component of the system functions

*Address correspondence to this author at the Department of Mathematics and Statistics, University of Missouri-Kansas City; Tel: 8165174614;
Fax: 816-2355517; E-mail: xsm7f@mail.umkc.edu

E-ISSN: 2409-5761/16

successfully [3]. As depicted in Figure **1**, the simplest series system consists of two components with their individual reliability denoted by θ and ω. The reliability of such a series system, denoted by η, is commonly represented as the product of reliabilities of both components, that is, $\eta = \theta\,\omega$.

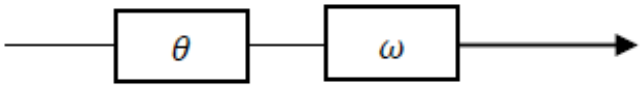

Figure 1: Representation of a simple series system

To obtain an estimate of η in the Bayesian point of view, a fixed number of *t* testing cases will be reasonably distributed between the two components to minimize the Bayes risk in terms of squared error loss. In particular, the outcome of the *i*th test randomly taken from one component is modeled as a Bernoulli trial such that:

$$f_{Xi}(x_i) = \left\{ \begin{array}{ll} \theta & if\ x_i = 1, success \\ 1-\theta & if\ x_i = 0, failure \end{array} \right\}$$

the outcome of the *j*th test randomly taken from the other component is modeled as another Bernoulli trial such that:

$$f_{Yj}(y_j) = \left\{ \begin{array}{ll} \omega & if\ y_j = 1, success \\ 1-\omega & if\ y_j = 0, failure \end{array} \right\}$$

where θ, ω are both unknown and assumed to be independent with conjugate prior distributions as $Beta(a_0, b_0), Beta(c_0, d_0)$ respectively, namely,

$$d\Pi(\theta) = \frac{\theta^{a_0-1}(1-\theta)^{b_0-1}}{B(a_0, b_0)}, \qquad 0 < \theta < 1, a_0 > 0, b_0 > 0$$

where

$$r = a_0 + b_0;\ \mu = \frac{a_0}{a_0 + b_0}$$

represents the fictious sample size and prior mean of its conjugate prior distribution. Similarly,

$$d\Gamma(\omega) = \frac{\omega^{c_0-1}(1-\omega)^{d_0-1}}{B(c_0, d_0)}, \qquad 0 < \omega < 1, c_0 > 0, d_0 > 0$$

where

$$s = c_0 + d_0;\ \ v = \frac{c_0}{c_0 + d_0}$$

By following certain allocation rule or sampling procedure *P*, *t* testing cases are distributed with m_t to one component and n_t to the other, where $m_t + n_t = t$. Let F_{m_t, n_t} be the σ-algebra generated by $X_1, ..., X_{m_t}$ *and* $Y_1, ..., Y_{n_t}$. To estimate the system reliability η, with squared error loss, the terminal Bayes estimator has been shown to be $A_{m_t}\ B_{n_t}$, where A_{mt} is the posterior mean of θ based on $X_1, ..., X_{m_t}$, and B_{n_t} is the posterior mean of ω based on $Y_1, ..., Y_{n_t}$ which can be further specified as:

$$\begin{aligned} A_{m_t} &= E\left(\theta \middle| F_{m_t, n_t}\right) = \frac{a_{m_t}}{a_{m_t} + b_{m_t}}; \\ B_{n_t} &= E\left(\omega \middle| F_{m_t, n_t}\right) = \frac{c_{n_t}}{c_{n_t} + d_{n_t}} \end{aligned} \qquad (2.1.1)$$

with

$$a_{m_t} = a_0 + \sum_{i=1}^{m_t} X_i;\ b_{m_t} = b_0 + m_t - \sum_{i=1}^{m_t} X_i$$

$$c_{n_t} = c_0 + \sum_{i=1}^{n_t} Y_i;\ d_{n_t} = d_0 + n_t - \sum_{i=1}^{n_t} Y_i$$

The overall Bayes risk of the system, $R(\mathcal{P})$, is thus the expected loss incurred by estimating η by the Bayes estimator $A_{m_t}\ B_{n_t}$, which can be written as,

$$R(\mathcal{P}) = E\{Var[\eta | \mathcal{F}_{m_t, n_t}]\}$$

and further expanded into the following form,

$$R(\mathcal{P}) = E\left\{\frac{U_{m_t} B_{n_t}^2}{m_t + r} + \frac{V_{n_t} A_{m_t}^2}{n_t + s} + \frac{U_{m_t} V_{n_t}}{(m_t + r)(n_t + s)}\right\} \qquad (2.1.2)$$

where

$$\begin{aligned} U_{m_t} &= E[\theta(1-\theta) | \mathcal{F}_{m_t, n_t}] = \frac{a_{m_t} b_{m_t}}{(a_{m_t} + b_{m_t})(a_{m_t} + b_{m_t} + 1)} \\ V_{n_t} &= E[\omega(1-\omega) | \mathcal{F}_{m_t, n_t}] = \frac{c_n d_n}{(c_{n_t} + d_{n_t})(c_{n_t} + d_{n_t} + 1)} \end{aligned} \qquad (2.1.3)$$

which gives better guidance in making allocation decisions [8, 14, 15].

It will be shown theoretically that the Bayes risk is bounded from below asymptotically and the key problem can be further specified as carefully determining an allocation of *t* test cases, i.e. values of m_t, n_t, such that the overall Bayes risk as in (2.1.2) is minimized.

2.1. Optimal Fixed Design

In a fixed sampling scheme, test cases are allocated before reliability testing even begins. Let $m = m_t$, $n = n_t$ and by assuming m, n being fixed, we can take m, n outside the expectations and have the incurred Bayes risk of (2.1.2) simplified by averaging out the posterior expectations. It gives:

$$R(\mathcal{F}) = \frac{E[\theta(1-\theta)]E(\omega^2)}{m+r} + \frac{E[\omega(1-\omega)]E(\theta^2)}{n+s} - \frac{E[\theta(1-\theta)\omega(1-\omega)]}{(m+r)(n+s)} \tag{2.1.4}$$

where $R(F)$ denotes the Bayes risk incurred by any fixed sampling design. Note that the last term of (2.1.4) can be rewritten as:

$$\frac{1}{t+r+s}\left\{\frac{E[\theta(1-\theta)]}{m+r} + \frac{E[\omega(1-\omega)]}{n+s}\right\}$$

which will converge to zero in first order as $t \to \infty$ for all possible combinations of m, n as long as they are proportional to t asymptotically, i.e. $m/t \to c$ where c is a constant. Then, (2.1.4) can be further written as:

$$R(F) = \frac{\left\{\left[\sqrt{E\left[\theta(1-\theta)\right]E\left(\omega^2\right)} + \sqrt{E\left[\omega(1-\omega)\right]E\left(\theta^2\right)}\right]^2\right\}}{t+r+s} + \frac{\left\{\left[\frac{(n+s)\sqrt{E\left[\theta(1-\theta)\right]E\left(\omega^2\right)} - (m+r)\sqrt{E\left[\omega(1-\omega)\right]E\left(\theta^2\right)}}{(m+r)(n+s)}\right]^2\right\}}{t+r+s} + o\left(\frac{1}{t}\right) \tag{2.1.5}$$

As t getting sufficiently large, it can be easily deduced from (2.1.5) that the Bayes risk incurred by any fixed sampling design is bounded below by:

$$R(\mathcal{F}) \geq \frac{\left(\sqrt{E[\theta(1-\theta)]E(\omega^2)} + \sqrt{E[\omega(1-\omega)]E(\theta^2)}\right)^2}{t+r+s} + o\left(\frac{1}{t}\right) \tag{2.1.6}$$

where the equality can be achieved if the second term of (2.1.5) vanishes by allocating t cases between the two populations as follows:

$$m^* = \frac{\sqrt{E\left[\theta(1-\theta)\right]E\left(\omega^2\right)}}{\sqrt{E\left[\theta(1-\theta)\right]E\left(\omega^2\right)} + \sqrt{E\left[\omega(1-\omega)\right]E\left(\theta^2\right)}}(t+r+s) - r$$

$$n^* = t - m \tag{2.1.7}$$

Or in a more explicit way,

$$R(\mathcal{F}) \geq \frac{(\alpha_0+\beta_0)^2}{t+r+s} + o\left(\frac{1}{t}\right) \tag{2.1.8}$$

with attainment of the equality if t test cases are allocated as:

$$m^* = \frac{\alpha_0}{\alpha_0+\beta_0}(t+r+s) - r; \quad n^* = t - m^* \tag{2.1.9}$$

where

$$\alpha_0 = \sqrt{E\left[\theta(1-\theta)\right]E\left(\omega^2\right)} = \sqrt{\frac{a_0 b_0 c_0 (c_0+1)}{(a_0+b_0)(a_0+b_0+1)(c_0+d_0)(c_0+d_0+1)}}$$
$$\beta_0 = \sqrt{E\left[\omega(1-\omega)\right]E\left(\theta^2\right)} = \sqrt{\frac{c_0 d_0 a_0 (a_0+1)}{(a_0+b_0)(a_0+b_0+1)(c_0+d_0)(c_0+d_0+1)}} \tag{2.1.10}$$

It is interesting to note from (2.1.9) and (2.1.10) that Bayes risk incurred by any fixed sampling scheme are highly mathematically dependent on, and thus very sensitive to, the choice of prior parameters. Inaccuracies in a_0, b_0, c_0, d_0 may greatly mislead the allocation of test cases and never get to be corrected. It is this drawback that motivates an adaptive sampling scheme which will be discussed in the next section.

2.2. First-Order Optimal Sequential Sampling Design

Now, the assumption of fixed allocation is relaxed and to stress on such randomness, M, N (or M_t, N_t) will be adopted in the expression of (2.1.2) and (2.1.3), while their sum t remains fixed. A sequential procedure, P_s, is defined as a sequence of allocation rules $\{(t, A_t)\}_{t\geq 1}$, where an allocation rule is essentially a stochastic process $A_t = \{(M_t, N_t)\}_{t\geq 1}$ *on* N^2 (N is the set of non-negative integers) satisfying:

$$(M_{t+1}, N_{t+1}) = (M_t, N_t) + \begin{cases} (0,1) \text{ or} \\ (1,0) \end{cases}$$

and

$$\{(M_t, N_t) = (m,n), (M_{t+1}, N_{t+1}) = (m+1, n) \cup (m, n+1)\} \in F_t$$

$\forall m,n,t \in N$ *and* $F_t = F_t(A_t) = \{A : A \cap \{(M_t, N_t) = (m,n) \in F_{m,n}, \forall m,n, m+n=t\}\}$. Then, F_t is easily seen to be a σ-algebra generated by A_t for every $t \geq 1$, and $F_t \subset F_{t+1}$. In other words, F_t contains maximum information which can be found out about the

procedure until t test cases are allocated between the two populations, which is essential in defining martingales and carrying out the following proofs.

Theorem 2.1 will first be established to give an asymptotic first-order lower bound for all such sequential procedures, followed by proposing a three-stage sampling design which will then be shown to be first-order efficient and outperform the optimal fixed design.

Theorem 2.1 For any sequential procedure, P_s;

$$R(\mathcal{P}_s) \geq \frac{E\left(\sqrt{\theta(1-\theta)}\omega + \sqrt{\omega(1-\omega)}\theta\right)^2}{t+r+s} + o\left(\frac{1}{t}\right) \quad (2.2.1)$$

Proof: From (2.1.2), we have the overall Bayes risk to be:

$$\begin{aligned} R(\mathcal{P}_s) &= E\left\{\frac{U_M B_N^2}{M+r} + \frac{V_N A_M^2}{N+s} + \frac{U_M V_N}{(M+r)(N+s)}\right\} \geq E\left\{\frac{U_M B_N^2}{M+r} + \frac{V_N A_M^2}{N+s}\right\} \\ &= \frac{1}{t+r+s} E\left(\sqrt{U_M B_N^2} + \sqrt{V_N A_M^2}\right)^2 \\ &\quad + \frac{1}{t+r+s} E\left\{\frac{\left[(N+s)\sqrt{U_M B_N^2} - (M+r)\sqrt{V_N A_M^2}\right]^2}{(M+r)(N+s)}\right\} \\ &\geq \frac{1}{t+r+s} E\left(\sqrt{U_M B_N^2} + \sqrt{V_N A_M^2}\right)^2 \end{aligned}$$

where A_M, U_M, B_N, V_N are specified in (2.1.3). Then, (2.2.1) follows from Fatou's lemma and bounded convergence theorem.

Assume the total sample size t is fixed, the three-stage sampling procedure is delivered as follows, where $\lfloor x \rfloor$ denotes the integer part of x:

Stage 1: Sample $l(t)$ observations from each population and evaluate $\hat{M}, \hat{N}$ by:

$$\hat{M} = \min\left\{\lfloor \kappa(t) - l(t) \rfloor, \max\left\{\kappa(t)\lfloor \hat{C}_{l,l} \rfloor, l(t)\right\}\right\}, \hat{N} = \kappa(t) - \hat{M};$$

where

$$\lim_{t\to\infty} \frac{l(t)}{t} = 0, \qquad \lim_{t} l(t) = \infty \quad (2.2.2)$$

and

$$\lim_{t\to\infty} \frac{l(t)}{\kappa(t)} = 0, \qquad \lim_{t\to\infty} \frac{\kappa(t)}{t} \in (0,1), \qquad \lim_{t\to\infty} \kappa(t) = \infty, \quad (2.2.3)$$

Stage 2: Sample $\hat{M} - l(t), \hat{N} - l(t)$ more cases from each corresponding population and refine the ratio as $\hat{C}_{\hat{M},\hat{N}}$.

Stage 3: Determine *M, N* by the updated $\hat{C}_{\hat{M},\hat{N}}$ such that:

$$M = \min\left\{\left\lfloor t - \hat{N} \right\rfloor, \max\left\{\left\lfloor t\hat{C}_{\hat{M},\hat{N}} \right\rfloor, \hat{M}\right\}\right\}, N = t - M$$

and $\hat{C}_{l,l}, \hat{C}_{\hat{M},\hat{N}}$ are defined as:

$$\hat{C}_{m_t,n_t} = \frac{\sqrt{U_{m_t}} B_{n_t}}{\sqrt{U_{m_t}} B_{n_t} + \sqrt{V_{n_t}} A_{m_t}} \quad (2.2.4)$$

$A_{m_t}, B_{n_t}, U_{m_t}, V_{n_t}$ are defined and evaluated as in (2.1.1) and (2.1.3).

Note that the procedure indicates that $\hat{M}$ should be no less than $l(t)$ but at most $\kappa(t) - l(t)$; while *M* should be at least $\hat{M}$ but no more than $t - \hat{N}$, since there are only $\left(t - \hat{N} - \hat{M}\right)$ test cases left in the last stage. Another feature about this design is that the stage sizes, $l(t)$ and $\kappa(t)$, are allowed to vary freely as long as (2.2.2) and (2.2.3) are satisfied. (2.2.2) was first suggested by Rekab for a two-stage procedure [10], which seems to imply a choice of $l(t)$ of the form $(kt)^{\lambda}$ for some $\lambda \in (0,1)$ and $k \in (0,1)$. Condition (2.2.3) suggests a potential candidate for $\kappa(t)$ to be. One simple assignment is to let $k = \lambda = 1/2$, namely, to select;

$$l(t) = \sqrt{\frac{t}{2}}, \quad \kappa(t) = \frac{t}{2}$$

which appeared in [1]. However, there is still much to be gained by choosing these stage sizes analytically. For example, dividing the $2l(t)$ test cases unevenly in accordance to the optimal fixed scheme, if the prior parameters are quite reliable [13]. More specific guidance with respect to some particular cases has been provided by Hardwick and Stout [5]. Regardless of the specification of $l(t)$ and $\kappa(t)$, first-order optimality of the three-stage sequential design is preserved, which will be shown by the following lemmas and theorem.

Lemma 2.2.1 Let P_{3stgae} be the three-stage sequential sampling scheme and (*M, N*) be the final allocation. Then,

$$\lim_{t\to\infty} \frac{M}{t} = C \ \ w.p.1$$

where;

$$C = \frac{\sqrt{\theta(1-\theta)}\omega}{\sqrt{\theta(1-\theta)}\omega + \sqrt{\omega(1-\omega)}\theta}$$

Proof: From the third stage and (2.2.4), it is obvious that for sufficiently large *t*,

$$\frac{M}{t} = \hat{C}_{\hat{M},\hat{N}} = \frac{\sqrt{U_{\hat{M}}}B_{\hat{N}}}{\sqrt{U_{\hat{M}}}B_{\hat{N}} + \sqrt{V_{\hat{N}}}A_{\hat{M}}}$$

which converges to *C* as a result of martingale convergence theorem. ■

Lemma 2.2.2 Under procedure P_{3stage}, the following inequality holds for all sufficiently large *t*,

$$\frac{t}{\hat{M}} \le \frac{4}{\hat{C}_{l,l}}$$

Proof: From stage 1, on $\{\hat{M} = l\}$, we should have $\lfloor t\hat{C}_{l,l}/2 \rfloor < l$, so that;

$$\frac{t}{2}\hat{C}_{l,l} \le \hat{M} + 1 \le 2\hat{M}$$

Next, on $\{\hat{M} > l\}, t\,\hat{C}_{l,l}/2 \ge \sqrt{t/2} \ge 2, \text{for } t \ge 8$. Then,

$$\frac{t/2}{\hat{M}} = \frac{t/2}{\lfloor \frac{t}{2}\hat{C}_{l,l} \rfloor} \le \frac{t/2}{\frac{t}{2}\hat{C}_{l,l} - 1} \le \frac{2}{\hat{C}_{l,l}}$$

The result follows immediately. ■

Theorem 2.2 Let P_{3stage} be the three-stage sampling scheme, if r>0, s> 0, then;

$$R(\mathcal{P}_{3stage}) = \frac{E\left(\sqrt{\theta(1-\theta)}\omega + \sqrt{\omega(1-\omega)}\theta\right)^2}{t+r+s} + o\left(\frac{1}{t}\right) \tag{2.2.5}$$

as $t \to \infty$.

Proof: As in the proof of Theorem 2.1,

$$(t+r+s)\left\{R(\mathcal{P}_{3stage}) - \left[E\left(\sqrt{\theta(1-\theta)}\omega + \sqrt{\omega(1-\omega)}\theta\right)^2\right]\right\}$$

is equal to the sum of the following three terms:

$$E\left(\sqrt{U_M}|B_N| + \sqrt{V_N}|A_M|\right)^2 - E\left(\sqrt{\theta(1-\theta)}\omega + \sqrt{\omega(1-\omega)}\theta\right)^2 \tag{2.2.6a}$$

$$E\left\{\frac{\left[(N+s)\sqrt{U_M}|B_N| - (M+r)\sqrt{V_N}|A_M|\right]^2}{(M+r)(N+s)}\right\} \tag{2.2.6b}$$

$$E\left[\frac{U_M V_N}{(M+r)} + \frac{U_M V_N}{(N+s)}\right] \tag{2.2.6c}$$

Since *M, N* > 0 and *r, s* > 0, both terms in (2.2.6c) are bounded by a uniformly integrable martingale, $U_M V_N$, from above. Also, the procedure indicates that $M > l(t)$ and $l(t) \to \infty$ along with t as required by (2.2.2), so *m* must also approach infinity as $t \to \infty$. As a result, (2.2.6c) will vanish by the dominated convergence theorem. By virtue of the following inequality:

$$\left(\sqrt{U_M}|B_N| + \sqrt{V_N}|A_M|\right)^2 \le 2(U_M B_N^2 + V_N A_M^2), \qquad w.p.1$$

where $\sqrt{U_M}|B_N|$ and $\sqrt{V_N}|A_M|$ are clearly uniformly integrable, it follows that (2.2.6a) approaches zero as $t \to \infty$. Finally, it directly follows from Lemma 2.2.1 that the random quantity within the expectation of (2.2.6b) converges to zero with probability one, that is,

$$\frac{\left[(N+s)\sqrt{U_M}|B_N| - (M+r)\sqrt{V_N}|A_M|\right]^2}{(M+r)(N+s)} \to 0, \qquad w.p.1$$

$$as\ t \to \infty$$

So as to claim that the expectation itself, i.e. (2.2.6b) also converges to zero, we still need to show that both terms:

$$U_M B_N^2 \frac{t}{M} \quad and \quad V_N A_M^2 \frac{t}{N}$$

are uniformly integrable. To this end, we first establish the following inequality:

$$U_M B_N^2 \frac{t}{M} \le 1 + \frac{1}{2}\left(\frac{1}{U_l} + \frac{1}{4B_l^2}\right)$$

Note that $0 \le \theta \le 1 \text{ and } 0 \le \omega \le 1, \text{then,}$

$$U_M B_N^2 = E[\theta(1-\theta)|\mathcal{F}_t]E(\omega|\mathcal{F}_t) \le \frac{1}{4}$$

It then follows from Lemma 2.2.2 that for sufficiently large *t*, we have:

$$U_M B_N^2 \frac{t}{M} \le \frac{1}{\hat{C}_{l,l}} = \frac{\sqrt{U_l}B_l + \sqrt{V_l}A_l}{\sqrt{U_l}B_l}$$

$$= 1 + \sqrt{\frac{A_l^2}{U_l}}\sqrt{\frac{V_l}{B_l^2}} \le 1 + \frac{1}{2}\left(\frac{A_l^2}{U_l} + \frac{V_l}{B_l^2}\right)$$

$$\leq 1+\frac{1}{2}\left(\frac{1}{U_l}+\frac{1}{4B_l^2}\right)$$

where the term in the second line is a non-negative sub-martingale, since $1/U_l$ and $1/B_l^2$ are both non-negative sub-martingales. Then, the uniform integrability of $U_M B_N^2 t/M$ follows from Doob's inequality and dominated convergence theorem. The other term $V_N A_M^2 t/N$ can be shown to be uniformly integrable in a similar fashion. ■

2.3. Sampling Scheme Comparison

In this section, the three-stage sequential sampling design will be shown to perform better than the optimal fixed design in the sense that the former incurred less Bayes risk than the latter, especially when t is large.

Theorem 2.3 Let P_{3stage} be the three-stage sequential sampling design and F_0 be the optimal fixed design, then;

$$R(\mathcal{F}_0) \geq R\left(\mathcal{P}_{3stage}\right)+o\left(\frac{1}{t}\right) \qquad (2.2.7)$$

Proof: It follows easily from (2.1.6) and (2.2.5) that,

$$\lim_{t\to\infty}(t+r+s)\left[R(\mathcal{F}_0)-R\left(\mathcal{P}_{3stage}\right)\right]$$

$$=\left(\sqrt{E[\theta(1-\theta)]E(\omega^2)}+\sqrt{E[\omega(1-\omega)]E(\theta^2)}\right)^2$$

$$-E\left(\sqrt{\theta(1-\theta)}\omega+\sqrt{\omega(1-\omega)}\theta\right)^2$$

and the result immediately follows from the Cauchy-Schwarz inequality. A similar proof can be found in Rehab and Li [15], whose work also implies that the gain of efficiency of the three-stage design over the best fixed design should be no more than 50%. ■

2.4. Monte Carlo Simulation

The results of experimental comparisons between the three-stage sampling design and the optimal fixed sampling scheme as well as the first-order efficiency of three-stage sampling design are presented in Table **1** and **2**. We consider the case where the system consists of two components, or two sub-systems connected in series, with reliability θ and ω respectively. Prior knowledge about the individual reliabilities are quantified by beta distribution, such that $\theta \sim Beta(a_0,b_0), \omega \sim Beta(c_0,d_0)$.

Table 1: The Bayes Risk Ratios of Three-Stage vs. Optimal Fixed (Various Prior Parameters, 10000 Replications)

(a_0,b_0) / (c_0,d_0) Size (t)	(1,1) (1,1)	(0.5,0.01) (0.5,0.01)	(0.5,0.01) (1,1)	(1,0.05) (0.1,0.005)
t = 60	0.935848	0.781562	0.977919	0.932582
t = 80	0.908820	0.715508	0.967004	0.867259
t = 100	0.888615	0.674085	0.967431	0.815990
t = 200	0.857535	0.578334	0.941608	0.725076
t = 400	0.849106	0.542225	0.935700	0.679830
t = 600	0.848983	0.530122	0.934358	0.664879
t = 800	0.847717	0.521657	0.931591	0.652924
t = 1000	0.847176	0.517955	0.931344	0.652541
$t\to\infty$	0.849783	0.512973	0.924828	0.649841

Table **1** shows the ratio of the Bayes risk of the three-stage sampling design to the Bayes risk of the optimal fixed scheme. The table presents results for various total samples sizes, t, and different prior parameters. The last row indicates the limiting value of $R(P_{3stage})$ when t becomes very large. Among all different settings, even with relative small sample sizes, the three-stage sampling scheme outperformed the optimal fixed scheme, just as what we would expect since the former procedure learns and refines the estimate during the testing process. This result can be confirmed by Theorem 2.3.

Column 1 in the table shows scenarios with uniform or non-informative priors where $a_0=b_0=c_0=d_0=1$. From a practical point of view, it indicates that very little is known about the reliability of each component or sub-system. In this case, the three-stage sampling design still outperforms the best fixed scheme with their ratios of Bayes risk being less than 1. As t becomes large, the limits converge to values in the last row, which are all above 50%.

Column 2 gives the case with informative and identical priors where $a_0=c_0=0.5,\ b_0=d_0=0.01$. This choice represents an expected reliability of 0.998 for each component or sub-system. It is also a practical scenario, such as for highly reliable one-shot systems like missiles and rockets [4], or for software applications which are already in place. Apparently, we gain more by using the three stage sampling scheme over the best fixed scheme. As t goes to infinity, the limiting value gets very close to the best efficiency gain, 50%.

Column 3 resembles the case when a new component, which we know very little about, is introduced to an existing system, where $a_0 = 0.5,\ \ b_0 = 0.01,\ \ c_0 = d_0 = 1$. Under these conditions, the three-stage sampling scheme still shows an improvement over the best fixed design. It is very interesting to see that the efficiency gain is not as good as the case with both uniform priors. One possible explanation is that the allocation scheme for a series system is collectively affected by information of both components, the uncertainty of one component may in turn reduce the faith about the other component and introduce extra risk to the estimation of the whole system.

The last parameter configuration in column 4 represents a choice of prior parameters where the expected reliability in both populations are believed to be equally high but we are more certain about the first component then the second, which is reflected by the difference of the prior variances. In this case, the three stage sampling scheme still performs much better than the best fixed scheme manifested by ratios listed in the last column.

Table 2: The Excessive First-Order Bayes Risk Incurred by Three-Stage Design (Various Prior Parameters, 10000 Replications)

(a_0, b_0) / (c_0, d_0) Size (t)	(1,1) (1,1)	(0.5,0.01) (0.5,0.01)	(0.5,0.01) (1,1)	(1,0.05) (0.1,0.005)
t = 60	0.003525	0.000369	0.003484	0.000609
t = 80	0.003550	0.000283	0.002873	0.000469
t = 100	0.003301	0.000230	0.002343	0.000381
t = 200	0.002389	0.000119	0.001298	0.000199
t = 400	0.001233	0.000065	0.000688	0.000103
t = 500	0.001051	0.000049	0.000560	0.000083
t = 600	0.000859	0.000041	0.000470	0.000072
t = 800	0.000658	0.000031	0.000357	0.000054
t = 1000	0.000534	0.000025	0.000289	0.000042

Table **2** and Figure **2** both demonstrate the first-order optimality of the three-stage sampling scheme by calculating and plotting the excess of first-order Bayes risk, which is defined as:

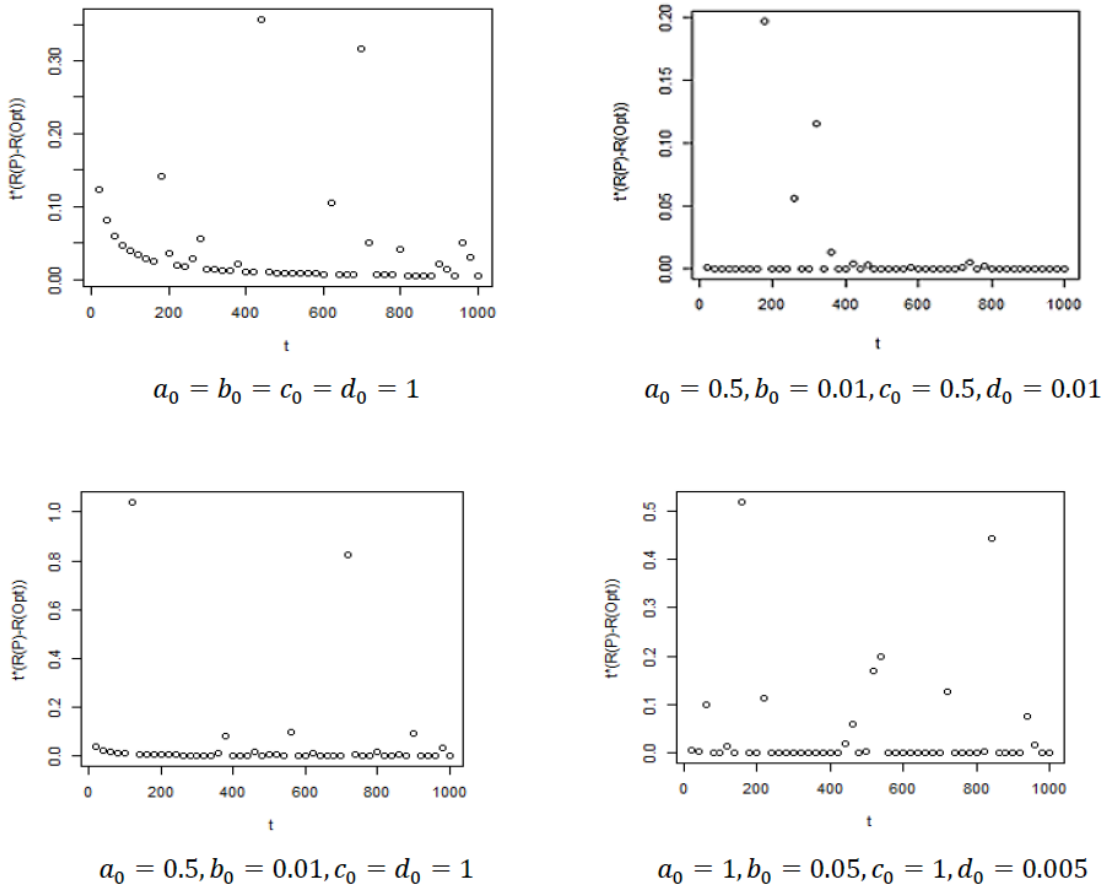

Figure 2: The excess of bayes risk incurred by the three-stage sampling scheme (various cases of a_0, b_0, c_0, d_0, 10000 replications).

$$(t+r+s)\left(R(\mathcal{P}_{3stage})-\frac{E\left(\sqrt{\theta(1-\theta)}\omega+\sqrt{\omega(1-\omega)}\theta\right)^2}{t+r+s}\right)$$

with increasing total sample size, t.

Four practical scenarios of prior parameters are considered separately as before and the corresponding excessive first-order Bayes risks all exhibit a trend of convergence towards zero, which agrees with Theorem 3.2. With smaller increment of *t*, Figure **2** demonstrates such converging tendency more evidently.

In addition, by comparing between column 1 and 3 with column 2 and 4, one may also discover that the Bayes risk incurred by the three-stage sampling scheme converges faster with informative priors than without or only with partial information.

3. APPLICATION II: RISK ASSESSMENT —NORMAL-NORMAL MODEL

Suppose that we observe $X \sim N(\theta, 1), \mathrm{Y} \sim \mathrm{N}(\omega,1)$ independently and are concerned with inference about the product of means, $\theta\omega$. The problem was first recognized as the determination of an area of a rectangle based on measurements of length and width and was then introduced to some environmental applications, such as exposure assessment and risk modeling. For example, to assess the risk due to exposure to radiation or various pollutants, *X* can be defined to measure the dose per unit time, *Y* the number of time units during which an individual is exposed and are independent of *X,* and thus the total exposure is $\theta\omega$ [16].

In the Bayesian framework with assumption of conjugate priors, the two parameters $\theta\omega$ are assigned with independent normal prior distributions such as;

$$\theta \sim N(\mu, 1/r) \ \ and \ \ \omega \sim N(v, 1/s)$$

respectively, where $\mu, v \in R, \ \ \text{and} \ \ r, s > 0$.

Let the study proceed until a fixed total number of samples, t, have been observed. According to a sampling procedure, P, m_t cases are randomly sampled from population X and n_t sampled from Y, where $m_t+n_t=t$. The resulting Bayes risk now becomes:

$$R(P)=E\left\{\frac{v_{n_t}^2}{m_t+r}+\frac{\mu_{m_t}^2}{n_t+s}+\frac{1}{(m_t+r)(n_t+s)}\right\} \qquad (3.1.1)$$

where μ_{m_t} denotes the posterior means of θ given $X_1, X_2, ..., X_{m_t}$, that is,

$$\mu_0=\mu; \ \ \mu_{m_t}=E\{\mu|\mathcal{F}_{m_t,n_t}\}=\frac{r\mu+\sum_{i=1}^{m_t}X_i}{m_t+r} \qquad (3.1.2)$$

Similarly, let v_{n_t} be the posterior means of ω given $Y_1, ..., Y_{n_t}$ then,

$$v_0=v; \ \ v_{n_t}=E\{v|\mathcal{F}_{m_t,n_t}\}=\frac{sv+\sum_{i=1}^{n_t}Y_i}{n_t+s} \qquad (3.1.3)$$

The observed value of μ_{m_t} v_{n_t} is thus the estimate for the parameter of interest, $\theta\omega$.

3.1. Optimal Fixed Design

Let $m=m_t, n=n_t$ and in any fixed sampling scheme, the allocation m,n are determined before any observation being collected, leading to the corresponding Bayes risk to be:

$$R(\mathcal{F}) =\frac{E^2(\omega)}{m+r}+\frac{E^2(\theta)}{n+s}+\frac{1}{(m+r)(n+s)} \qquad (3.1.4)$$

Similarly as in section 2.1, the Bayes risk, *R*(*F*), incurred by any fixed sampling scheme has an asymptotical lower bound specified as:

$$\begin{aligned} R(\mathcal{F}) &\geq \frac{\left(\sqrt{E(\omega^2)}+\sqrt{E(\theta^2)}\right)^2}{t+r+s}+o\left(\frac{1}{t}\right) \\ &= \frac{\left(\sqrt{\mu^2+\frac{1}{r}}+\sqrt{v^2+\frac{1}{s}}\right)^2}{t+r+s}+o\left(\frac{1}{t}\right) \end{aligned} \qquad (3.1.5)$$

which, with sufficiently large *t*, can be attained by the following allocation:

$$m^*=\frac{|E(\omega)|}{|E(\omega)|+|E(\theta)|}=\frac{|\mu|}{|\mu|+|v|}, \qquad n^*=t-m^*$$

3.2. First-Order Optimal Sequential Sampling Design

In this section, an asymptotic first-order lower bound for any sequential procedure P_s will be given first with a sketch of proof and the three-stage sequential sampling scheme will later be adopted again and justified to be first-order efficient both theoretically and through Monte Carlo simulations. Again, let $M = m_t$ and $N = n_t$ to indicate the allocation randomness.

Theorem 3.1 For any sequential procedure, P_s,

$$R(\mathcal{P}_s) \geq \frac{E(|\omega| + |\theta|)^2}{t + r + s} + o\left(\frac{1}{t}\right) \tag{3.2.1}$$

Proof: The proof of Theorem 3.2 is analogous to the proof of Theorem 2.2 by replacing A_M by μ_M, B_N by v_N, V_M and U_N both by 1, and follows directly from the key identity:

$$E\left(\frac{v_N^2}{M+r} + \frac{\mu_N^2}{N+s}\right) = \frac{E(|v_N| + |\mu_M|)^2}{t+r+s} + \frac{E\left\{\frac{[(N+s)|v_N| - (M+r)|\mu_M|]^2}{(M+r)(N+s)}\right\}}{t+r+s}$$

A similar proof can also be found in Rekab [10]. ■

The three-stage sampling procedure *P3stage* follows a very similar structure that has been elaborated in section 2.2, regardless of distribution specifications. Details are not to be repeated here, except that the key ratio involved to determine the allocation at each stage is adjusted to:

$$\hat{C}_{m_t, n_t} = \frac{|v_{n_t}|}{|\mu_{m_t}| + |v_{n_t}|} \tag{3.2.2}$$

with μ_{mt}, v_{nt} defined and evaluated by (3.1.2) and (3.1.3). For the sake of practicality, we will simply choose $l(t) = \sqrt{t/2}$, and $\kappa(t) = t/2$ as sizes for the first two stages.

Lemma 3.2.1 Let (*M, N*) be the final allocation given by *P3stage*, then,

$$\lim_{t\to\infty} \frac{M}{t} = C = \frac{|\theta|}{|\theta| + |\omega|} \quad w.p.1$$

Proof: The proof is very similar to the proof of Lemma 2.2.1. ■

Theorem 3.2 Let *P3stage* be three-stage sequential sampling design defined as above, if r > 0, s > 0 and *μ* and *v* are both finite, then;

$$R(\mathcal{P}_{3stage}) = \frac{E(|\theta| + |\omega|)^2}{t + r + s} + o\left(\frac{1}{t}\right) \tag{3.2.3}$$

as $t \to \infty$.

Proof: From the identity used for proving Theorem 3.1, the quantity;

$$(t + r + s)\{R(\mathcal{P}_{3stage}) - [E(|\theta| + |\omega|)^2]\}$$

can be rewritten as the sum of the following three terms:

$$E(|v_N| + |\mu_M|)^2 - E(|\theta| + |\omega|)^2 \tag{3.2.4a}$$

$$E\left\{\frac{[(N+s)|v_N| - (M+r)|\mu_M|]^2}{(M+r)(N+s)}\right\} \tag{3.2.4b}$$

$$E\left[\frac{t+r+s}{(M+r)(N+s)}\right] \tag{3.2.4c}$$

Since r > 0, s > 0, and

$$\frac{t+r+s}{(M+r)(N+s)} \leq 2\max\left\{\frac{1}{M+r}, \frac{1}{N+s}\right\}$$

It follows that (3.2.4c) will vanish. Also note that,

$$E(|v_N| + |\mu_M|)^2 \leq 2E(v_N^2 + \mu_M^2)$$

$$\leq 2E\left(\sup_N v_N^2 + \sup_M \mu_N^2\right) \leq 8[E(\omega^2) + E(\theta^2)]$$

where the last inequality follows from Doob's inequality and the right-most bound is finite. Hence, (3.2.4a) also vanishes. Turning to the term (3.2.4b), by using Lemma 3.2.1, it follows that:

$$\frac{[(N+s)|v_N| - (M+r)|\mu_M|]^2}{(M+r)(N+s)} \to 0, \qquad w.p.1,$$

$as\ t \to \infty$. Then, the only thing left to be checked is the uniform integrability of $v_N^2 t/M$ and $\mu_M^2 t/N$, which follows in the similar way as the proof for Theorem 2.2, as long as we can validate the following two relations:

$$v_N^2 \frac{t}{M} \leq 2\sup_{k\geq 1}(v_k^2 + \mu_k^2) + 4v_N^2$$

$$\mu_M^2 \frac{t}{N} \leq 2\sup_{k\geq 1}(v_k^2 + \mu_k^2) + 4\mu_M^2$$

From Lemma 2.2.2, for sufficiently large *t* and under procedure *P3stage*, it can be easily established that:

$$v_N^2 \frac{t}{M} \leq 4v_N^2 \max_{k\leq l}\left|\frac{\mu_l}{v_l}\right| + 4v_N^2 \leq 4\sup_{k\geq 1}|v_k|\,|\mu_k| + 4v_N^2$$
$$\leq 2\sup_{k\geq 1}(v_k^2 + \mu_k^2) + 4v_N^2$$

The other relation can be shown in the same fashion. A similar proof appears in Rekab [10] for a two-stage procedure. ■

3.3. Sampling Scheme Comparison

In this section, the three-stage sequential sampling scheme will be shown to outperform the optimal fixed design in terms of the Bayes risk, especially when t is large.

Theorem 3.3 Let P_{3stage} be the three-stage design and F_0 be the optimal fixed design, then;

$$R(\mathcal{F}_0) \geq R\left(\mathcal{P}_{3stage}\right) + o\left(\frac{1}{t}\right) \tag{3.2.5}$$

Proof: It follows easily from (3.1.5) and (3.2.3) that,

$$\lim_{t\to\infty}(t+r+s)\left[R(\mathcal{F}_0) - R\left(\mathcal{P}_{3stage}\right)\right]$$

$$=\left(\sqrt{E(\omega^2)} + \sqrt{E(\theta^2)}\right)^2 - E(|\theta| + |\omega|)^2$$

$$=2\left\{\sqrt{E(\omega^2\theta^2)} - E\left[\sqrt{(\omega^2\theta^2)}\right]\right\}$$

and the result follows easily from the Jensen's Inequality for concave functions. ■

3.4. Monte Carlo Simulation

The results of experimental comparisons between the three-stage sampling design and the optimal fixed sampling scheme as well as the first-order efficiency of three-stage sampling design will be presented in Table **3** and **4**. We consider the case of assessing the risk due to exposure to certain type of pollutant. Let X_i be the dose of the pollutant per unit time for the ith subject, and Y_i be the number of time units during which the ith subject is exposed.

Table 3: The Bayes Risk Ratios of Three-Stage vs. Optimal Fixed (Various Prior Parameters, 10000 Replications)

Prior / Size	$\mu = v$ $r = s$	$\mu \geq 10v$ $r = s$	$\mu = v$ $\mu \geq 10s$	$\mu \leq 0.1v$ $r \geq 10s$
t = 60	0.895004	0.999035	0.951571	0.931777
t = 80	0.899987	0.913864	0.905650	0.968722
t = 100	0.967188	0.979956	0.905229	0.890654
t = 200	0.992475	0.946160	0.901752	0.885033
t = 400	0.986781	0.966606	0.902478	0.853707
t = 600	0.895689	0.930659	0.927050	0.874488
t = 800	0.972394	0.973464	0.905229	0.861526
t = 1000	0.966911	0.961724	0.897031	0.861428
$t \to \infty$	0.866811	0.964907	0.885927	0.844912

Table 4: The Excessive First-Order Bayes Risk Incurred by Three-Stage Design (Various Initial Conditions, 10000 Replications)

Prior / Size	$\mu = v$ $r = s$	$\mu = 10v$ $r = s$	$\mu = v$ $r = 10s$	$\mu = 0.1v$ $r = 10s$
t = 60	0.192910	1.676767	0.313537	0.822316
t = 80	0.053572	1.334689	0.012233	0.263540
t = 100	0.071821	0.025763	0.012406	0.075005
t = 200	0.069218	0.019987	0.009390	0.018621
t = 400	0.022048	0.008023	0.004869	0.008148
t = 600	0.006256	0.005701	0.002442	0.002258
t = 800	0.003215	0.005148	0.001369	0.003702
t = 1000	0.001381	0.001233	0.000150	0.000142

Without loss of generality, we assume $X_i \sim N(\theta,1)$, $Y_i \sim N(\omega,1)$ and they are independent random variables. Prior knowledge about the expected dose of pollutant and exposing time are quantified by normal distributions, such that,

$$\theta \sim Normal(\mu, 1/r),\ \omega \sim Normal(v, 1/s).$$

where r, s are the precisions of θ and ω, respectively.

Table **3** shows the ratio of Bayes risk of the three-stage design to that of the optimal fixed scheme. The table presents results for various prior parameters and different total samples sizes, t, including the limiting case as t approaches infinity.

The first column of Table **3** is a case of equal priors, such that $\mu = v, r = s$. The second column simulates the case where μ/v is much larger than 1, while r/s = 1. In a practical sense, the first two scenarios indicate that we equally believe in how the two factors, pollutant concentration and contact time, will affect overall exposure, but blame more on the pollutant concentration in the second scenario. In the third column, with r/s being large, we seem to be surer about the effect from the pollutant concentration. The last scenario, where μ/v gets very small but r/s is relatively large, indicates that we are not very concerned about the pollutant concentration to be the key factor which changes the level of exposure.

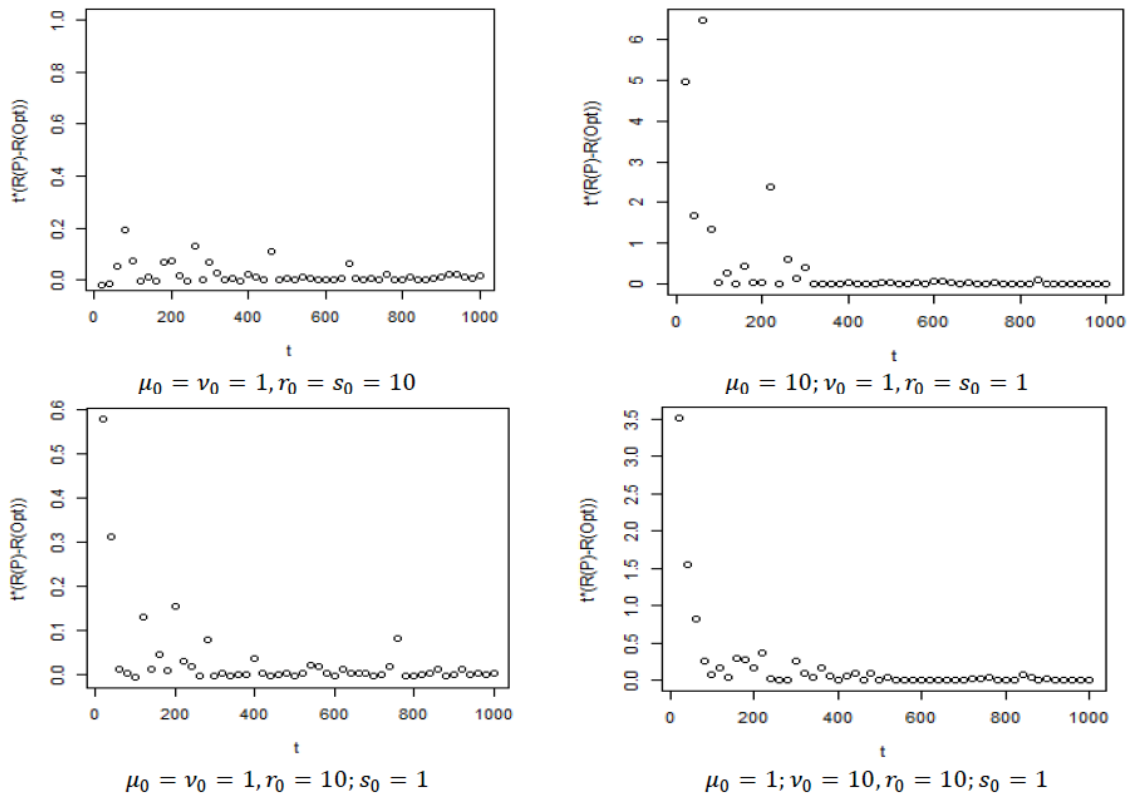

Figure 3: The excess of bayes risk incurred by the three-stage sampling scheme (various cases of μ_0, r_0, ν_0, s_0, 10000 replications).

Among all different scenarios listed in Table **3**, even with relatively small sample sizes, the three-stage design outperformed the optimal fixed scheme. Asymptotically, there is also gain of efficiency of the three-stage design over the optimal fixed scheme, which is consistent with the result of Theorem 3.3.

Table **4** and Figure **3** both demonstrate the first-order optimality of the three-stage sampling scheme by calculating and plotting the excess of first-order Bayes risk, which is defined as:

$$(t+r+s)\left(R(\mathcal{P}_{3stage}) - \frac{E(|\omega|+|\theta|)^2}{t+r+s}\right)$$

with growing total sample size, t.

Four practical scenarios of prior parameters are considered separately as before and the corresponding excessive first-order Bayes risks all exhibit a nice trend of convergence towards zero, which is even more evident in Figure **3**.

CONCLUSION

The Beta-Binomial model, a typical case used for discrete data; and the Normal-Normal model, a popular case for continuous data, have been discussed in the paper in the context of two application problems. The three-stage sequential sampling scheme, which is shown to be first-order efficient both theoretically and through Monte Carlo simulations, also turns out to be practically implementable. The three-stage sampling design has the potential to be adapted to solve more problems which involves distributions from one-parameter exponential family. Furthermore, to achieve better accuracy and higher efficiency, sequential sampling scheme of second-order optimality will be of interest for the future study.

REFERENCE

[1] Benkamra Z, *et al*. Nearly second order three-stage design for estimating a product of several Bernoulli proportions. J Stat Plan Inference 2015; 167: 90-101. http://dx.doi.org/10.1016/j.jspi.2015.05.006

[2] Benkamra Z, Terbeche M and Tlemcani M. Bayesian sequential estimation of the reliability of a parallel-series system. Appl Math Comput 2013; 219(23): 10842-10852. http://dx.doi.org/10.1016/j.amc.2013.05.010

[3] Billinton R, *et al*. Reliability Evaluation of Engineering System, Concepts and Techniques. Springer: New York 1983. http://dx.doi.org/10.1007/978-1-4615-7728-7

[4] Guo H, Honecker S, Mettas A and Ogden D. Reliability Estimation for On-Shot System with Zero Components test Failures. Proceedings to the Reliability and Maintainability

Symposium (RAMS) – San Jose 2010. IEEE: pp 1-7. http://dx.doi.org/10.1109/rams.2010.5448016

[5] Hardwick J and Stout Q. Determining optimal few-stage allocation procedures. Comput Sci Stat 1995; 27: 342-346.

[6] Harkwick J and Stout QF. Optimal few-stage designs. J Stat Plan Inference 2002; 104: 121-145. http://dx.doi.org/10.1016/S0378-3758(01)00242-7

[7] Kenneth VD, John SD and Kenneth JS. Incorporating a geometric mean formula into CPI. Mon Labor Rev 1998; October: 2-7.

[8] Lehmann EL. Theory of Point Estimation. Wiley: New York 1983. http://dx.doi.org/10.1007/978-1-4757-2769-2

[9] Littlewood B and Wright D. Some conservative stopping rules for the operational testing of safety critical software. IEEE Trans Softw Eng 1997; 23(11): 673-683. http://dx.doi.org/10.1109/32.637384

[10] Rekab K. A Nearly Optimal Two Stage Procedure. Comm Stat – Theory Meth 1992; 12(1): 197-201. http://dx.doi.org/10.1080/03610929208830772

[11] Rekab K. A Sampling Scheme for Estimating the Reliability of Series System. IEEE. Trans Rel 1992; 42: 287-291. http://dx.doi.org/10.1109/24.229502

[12] Rekab K and Cheng Y. An accelerated sequential sampling scheme for estimating the reliability of a n-parallel network system. Int J Rel Appl 2013; 14(2): 71-78.

[13] Rekab K and Thompson H and Wu W. A multistage sequential test allocation for software reliability estimation. IEEE Trans Rel 2013; 62(2): 424-433. http://dx.doi.org/10.1109/TR.2013.2259195

[14] Rekab K and Song X. First-Order Asymptotic Efficiency in Sequential Designs for Estimating product of means in the Exponential Family case. Manuscript submitted for publication, 2016.

[15] Rekab K and Li Y. Bayesian Estimation of the Product of Two Proportions, Stoch. Anal Appl 1994; 12(3): 369-377. http://dx.doi.org/10.1080/07362999408809357

[16] Sun D and Ye K. Reference Prior Bayesian Analysis for Normal Mean Products. J Am Stat Assoc Theory Meth 1995; 90(430). http://dx.doi.org/10.1080/01621459.1995.10476551

[17] Woodroofe M and Hardwick J. Sequential allocation for an estimation problem with ethical cost. Ann Stat 1991; 18: 1358-1367. http://dx.doi.org/10.1214/aos/1176347754

[18] Yfantis E and Flatman GT. On Confidence Interval for the Product of Three Means of Three Normally Distributed Populations. J Chemom1991; 5: 309-319. http://dx.doi.org/10.1002/cem.1180050317

Received on 24-05-2016 Accepted on 25-05-2016 Published on 16-06-2016

DOI: http://dx.doi.org/10.15377/2409-5761.2016.03.01.3