ASYMPTOTIC OPTIMALITY OF EXPERIMENTAL DESIGNS IN ESTIMATING A PRODUCT OF MEANS*

Kamel Rekab

Department of Applied Mathematics
Florida Institute of Technology
Melbourne, FL 32905

ABSTRACT

In nonlinear estimation problems with linear models, one difficulty in obtaining optimal designs is their dependence on the true value of the unknown parameters. A Bayesian approach is adopted with the assumption the means are independent apriori and have conjuguate prior distributions. The problem of designing an experiment to estimate the product of the means of two normal populations is considered. The main results determine an asymptotic lower bound for the Bayes risk, and a necessary and sufficient condition for any sequential procedure to achieve the bound.

1. INTRODUCTION

There are many statistical problems in which a good choice for the design depends on the true value of the unknown parameters; For such problems, the

idea of designing the experiment sequentially is very appropriate. For example, in nonlinear estimation problems with linear models, one difficulty in obtaining optimal designs is their dependence on the value of the parameters. For another illustrative example, consider the problem of estimating the difference of the means of two normal populations with unknown means and unknown variances. It is optimal to design the experiment sequentially in order to have the ratio of the two sample sizes equal the ratio of the standard deviations, which are unknown apriori.

The problem of finding efficient experimental designs in nonlinear problems is of considerable practical importance, since efficient designs use the available resources more effectively.

Robbins, Simons, and Starr (1966) considered this design problem for estimating the parameter $\Delta = \mu_1 - \mu_2$, in the normal model; that is, suppose that

$$X_i \sim N(\mu_1, \sigma_1{}^2) \quad \text{and} \quad Y_i \sim N(\mu_2, \sigma_2{}^2) \quad \text{for all} \quad i = 1, \ldots.$$

The four parameters μ_1, μ_2, σ_1, σ_2 are assumed unknown. Their procedure can be described as follows. Let

$$u_i{}^2 = \frac{\sum_{k=1}^{i}(x_k - \overline{x}_i)^2}{(i-1)} \quad v_j{}^2 = \frac{\sum_{k=1}^{j}(y_k - \overline{y}_j)^2}{(j-1)}.$$

Then choose the next observation on x or on y according as i/j is less or greater than u_i/v_j. This sequential procedure was shown to be asymptotically efficient.

Consider the general linear model $y = x^t\beta + \epsilon$, where β is $k \times 1$, ϵ is an error which is $N(0,1)$, and the design variable x can be chosen within a bounded region $\mathcal{X}$. The $p \times 1$ vector parameter of interest is $\Psi = g(\beta)$, a nonlinear smooth function of β.

In nonlinear regression problems, the performance of a design depends on the unknown parameters. For instance, the Fisher information for Ψ depends on the x's chosen and β. In these circumstances, efficient designs must be constructed sequentially. The choice of the next design point should be chosen using the information about the parameters from previous observations.

Ford and Silvey (1980) considered the design problem for estimating the ratio $\rho = -\theta_1/2\theta_2$, which is the turning point of the regression function $\theta_1 x + \theta_2 x^2$ in the

linear model

$$y = \theta_1 x + \theta_2 x^2 + \epsilon, \quad \epsilon \sim N(0, \sigma^2), \quad \mid x \mid \leq 1.$$

They proposed a sequential procedure which effectively selects x to maximize the estimated Fisher Information at each stage. The properties of this procedure were then studied by Wu (1985). It is very important to note that neither one of these authors concentrates toward the optimality; their work is restricted entirely to the ad hoc design. The goal of this article is to refine the above analysis, searching for procedures with higher order efficiency.

To obtain a higher order efficiency, the Bayesian approach is adopted with squared error loss. It assumes that the means are independent apriori and have conjuguate prior distributions before the observations are known. These are modified through Bayes theorem, to posterior distributions given the observations. By adopting the Bayesian approach, we find a necessary and sufficient condition for the first order optimality of any sequential design. Throughout this article, we assume that the total number of observations from the two populations is fixed, so that the problem is one of selecting the sample sizes.

The investigation presented here concentrates on two normal populations with unit variance. The parameter of interest is the product of the means. An asymptotic lower bound for the Bayes risk and a necessary and sufficient condition for any sequential procedure to achieve the bound are derived.

Suppose we restrict our attention to allocation problems where the design space has two points, $\mathcal{X} = \{(0,1)^t, (1,0)^t\}$ say. Shortly after the beginning stages of an experiment, a sequential adaptive procedure will have learned the general location of β. Since any smooth function $g(\beta)$ can be expanded into a Taylor series, a thorough understanding of problems where Ψ is polynomial should have value in a general setting.

Definitions and Notation

Suppose that θ and ω are independent random variables which have (prior) normal distributions, say

$$\theta \sim N(\mu, 1/r) \quad \text{and} \quad \omega \sim N(\nu, 1/s)$$

where $\mu, \nu \in \Re$ and $r, s > 0$. Given θ, ω, let $X_1, X_2, \ldots Y_1, Y_2, \ldots$ be independent, with

$$X_i \sim N(\theta, 1) \quad \text{and} \quad Y_i \sim N(\omega, 1) \quad \text{for all} \quad i = 1, \ldots.$$

Allocation rules are algorithms for determining the number of $X's$ and $Y's$ to be sampled. These must satisfy measurability conditions prohibiting clairvoyance. Let $\mathcal{F}_{jl} = \sigma(X_1, \ldots, X_j; Y_1, \ldots, Y_l)$, the sigma algebra generated by $X_1, \ldots, X_j$ and $Y_1, \ldots, Y_l$. Formally, an allocation rule will be a stochastic process $\mathcal{A} = \{(n_k, m_k)\}_{k \geq 1}$ on $\mathcal{N}^2$ ($\mathcal{N}$ is the set of non-negative integers) satisfying

$$(n_{k+1}, m_{k+1}) = (n_k, m_k) + \begin{cases} (0,1), & \text{or} \\ (1,0) \end{cases}$$

and

$$\{(n_k, m_k) = (j, l), (n_{k+1}, m_{k+1}) = (j+1, l)\} \in \mathcal{F}_{jl},$$

for all $k \in \mathcal{N}, j \in \mathcal{N}, l \in \mathcal{N}$. Here, let n_k, and m_k denote respectively the number of X's and Y's sampled up to stage k. Let $\mathcal{F}_k = \mathcal{F}_k(\mathcal{A}) = \{A : A \cap \{(n_k, m_k) = (j, l)\} \in \mathcal{F}_{j,l}, \forall j, l\}$. Then $\mathcal{F}_k$ is easily seen to be a sigma algebra for every $k \geq 1$, and $\mathcal{F}_k \subset \mathcal{F}_{k+1}$. A procedure will be a sequence of allocation rules $\{(N, \mathcal{A}_N)\}_{N \geq 1}$.

Let μ_j be the posterior mean of θ given $X_1, \ldots, X_j$; that is

$$\mu_0 = \mu$$

and

$$\mu_j = \frac{r\mu + \sum_{i=1}^j X_i}{j + r}$$

for $j = 1, \ldots$. Similarly let ν_j be the posterior mean of ω given $Y_1, \ldots, Y_j$; that is

$$\nu_0 = \nu$$

and

$$\nu_j = \frac{s\nu + \sum_{i=1}^j Y_i}{j + s}$$

for $j = 1, \ldots$. Observe $n_k + r$ and $m_k + s$ are the precisions of θ and ω respectively at stage k. So,

$$\theta, \omega | \mathcal{F}_k \sim N(\mu_{n_k}, \frac{1}{n_k + r}) \times N(\nu_{m_k}, \frac{1}{m_k + s}).$$

The study proceeds until stage N (fixed). To simplify the notation, let

$$n = n_N \quad \text{and} \quad m = m_N.$$

Now, consider the problem of estimating the product $\theta\omega$ with squared error loss. It is well known that the Bayes risk is minimized by taking the posterior means as estimates. Then

$$\mathcal{R}(\mathcal{P}) = E\operatorname{Var}(\theta\omega|\mathcal{F}_N)$$

$$(1) \qquad = E\left(\frac{\mu_n^2}{m+s} + \frac{\nu_m^2}{n+r} + \frac{1}{(n+r)(m+s)}\right),$$

where E denotes the expectation with respect to the joint distribution of θ, ω, $X_1, X_2, \ldots$ and $Y_1, Y_2, \ldots$

2. First Order Lower Bound

In this section an asymptotic lower bound for $\mathcal{R}(\mathcal{P})$ as N goes to infinity and a necessary and sufficient condition for any procedure to achieve the bound are derived.

Theorem : Let $\mathcal{R}(\mathcal{P})$ be defined as in (1). Then

$$(2) \qquad \mathcal{R}(\mathcal{P}) \geq \frac{E(|\theta| + |\omega|)^2}{N + r + s} + o(1/N)$$

as $N \to +\infty$ with equality if and only if the following three conditions are satisfied:

(i) $m, n \to +\infty$ in probability as $N \to +\infty$

(ii) $\dfrac{m}{N} \to \dfrac{|\theta|}{|\theta| + |\omega|}$ in probability as $N \to +\infty$

(iii) $\mu_n^2 \dfrac{N}{m}$ and $\nu_m^2 \dfrac{N}{m}$ are uniformly integrable.

Proof of Theorem 1

The proof of the theorem requires the following five lemmas.

Lemma 1: Let $T_1, T_2, \ldots$ and T be random variables on a probability space $(\Omega, \mathcal{A}, P)$. Then

(a)   If $T_k \to T$ almost surely, given T, as $k \to +\infty$, then $T_k \to T$ almost surely as $k \to +\infty$.

(b)   If $T_k \to T$ in probability, given T, as $k \to +\infty$, then $T_k \to T$ in probability as $k \to +\infty$.

Proof: To establish (a), observe that

$$\Pr\{T_k \to T\} = E\left(\Pr\{T_k \to T | T\}\right)$$
$$= E(1) = 1.$$

To establish (b), observe that

$$\forall \epsilon > 0, \quad \Pr\{| T_k - T | > \epsilon\} = E\left(\Pr\{| T_k - T | > \epsilon | T\}\right).$$

The proof follows by the dominated convergence theorem.

Lemma 2: If n , $m \to +\infty$ in probabilty as $N \to +\infty$, then

$$\mu_n \to \theta \quad \text{and} \quad \nu_m \to \omega \quad \text{in probability as} \quad N \to +\infty.$$

Proof: The following result is crucial for the proof of the lemma:

Let Z_k, $k \geq 1$ be a sequence of random variables, let Z be a random variable, and let τ_a be a sequence of integer valued random variables such that:

$$Z_k \to Z \quad \text{almost surely as} \quad k \to +\infty$$

and

$$\tau_a \to +\infty \quad \text{in probability as} \quad a \to +\infty.$$

Then

$$Z_{\tau_a} \to Z \quad \text{in probability as} \quad a \to +\infty.$$

The lemma will follow if we show $\mu_k \to \theta$ almost surely as $k \to +\infty$ and $\nu_k \to \omega$ almost surely as $k \to +\infty$. By the strong law of large numbers, $\mu_k \to \theta$ almost

surely, given θ, as $k \to \infty$. Hence $\mu_k \to \theta$ almost surely as $k \to +\infty$, by Lemma (1) and similarly $\nu_k \to \omega$ almost surely as $k \to +\infty$.

Lemma 3: For any sequential procedure $\{(\mid \mu_n \mid + \mid \nu_m \mid)^2 ; N \geq 0\}$ is dominated by an integrable random variable .

Proof:

$$(\mid \mu_n \mid + \mid \nu_m \mid)^2 \leq 2(\mid \mu_n \mid^2 + \mid \nu_m \mid^2)$$

$$\leq 2(\sup_{k \geq r} \mid \mu_k \mid^2 + \sup_{k \geq s} \mid \nu_k \mid^2).$$

So in order to establish the proof we need to show that

$$E(\sup_{k \geq r} \mid \mu_k \mid^2) < +\infty$$

and

(3)
$$E(\sup_{k \geq s} \mid \nu_k \mid^2) < +\infty.$$

Now μ_k and ν_k, $k \geq 1$, are martingales, since they are formed by taking succesive conditional expectations. So domination follows from Doob's inequality. In fact

$$E(\sup_{k \geq r} \mid \mu_k \mid^2) \leq 4 \sup_k E(\mu_k{}^2)$$

$$= 4 \left(\frac{1}{r} + \mu^2 \right).$$

Lemma 4: If $\mathcal{R}(\mathcal{P}) \to 0$ as $N \to +\infty$, then $m \to +\infty$, and $n \to +\infty$ in probability.

Proof: Recalling equation (1), $\mathcal{R}(\mathcal{P}) \to 0$ as $N \to +\infty$ implies $E\left(\frac{\mu_n{}^2}{m+s} \right) \to 0$ as $N \to +\infty$. So $\frac{\mu_n^2}{m+s} \to 0$ in probability as $N \to +\infty$. Observe that

$$\frac{\mu_n^2}{m+s} \geq \frac{1}{m+s} \inf_k \mu_k^2$$

But $\inf_k \mu_k^2 > 0$ with probability one since $\mu_k \to \theta \neq 0$ with probability one and $\Pr(\bigcap_{k\geq 1}^{\infty}\{\mu_k \neq 0\}) = 1$. So

$$\frac{1}{m+s} \leq (\sup_k \frac{1}{\mu_k^2})\frac{\mu_n^2}{m+s}$$

$$\to 0.$$

By the same argument involving $\frac{\nu_m^2}{n+r}$ it follows that $n \to +\infty$ in probability, as $N \to +\infty$.

Lemma 5: Let

$$f(a,b,x) = \frac{a^2}{x} + \frac{b^2}{1-x} - (a+b)^2$$

for $a, b > 0$ and $0 < x < 1$. Then

(a) $f(a,b,x) \geq 0$ with equality if and only if $x = \dfrac{a}{a+b}$.

(b) If $f(\mu_n, \nu_m, \frac{m}{N}) \to 0$, $n \to +\infty$, and $m \to +\infty$ in probability as $N \to +\infty$, then

$$\frac{m}{N} \to \frac{|\theta|}{|\theta| + |\omega|} \quad \text{in probability as} N \to +\infty.$$

Proof: (a) follows directly from the identity

(4) $$\frac{a^2}{x} + \frac{b^2}{1-x} - (a+b)^2 = \frac{[a-(a+b)x]^2}{x(1-x)}.$$

To establish (b), first observe that

$$(|\mu_n| - \frac{m+s}{N+r+s}(|\mu_n| + |\nu_m|))^2 \leq \frac{1}{4}f(|\mu_n|, |\nu_m|, \frac{m+s}{N+r+s})$$

by (4), since $x(1-x) \leq 1/4$ for $0 < x < 1$. So, if $f(\mu_n, \nu_m, m/N) \to 0$ in probability then

$$(\frac{m+s}{N+r+s} - \frac{|\mu_n|}{|\mu_n| + |\nu_m|})^2(|\mu_n| + |\nu_m|)^2 \to 0 \quad \text{in probability as} \quad N \to +\infty.$$

By the remarks at the beginning of the proof of Lemma 2, $\mu_n \to \theta$ and $\nu_m \to \omega$ in probability as $N \to +\infty$, so

$$\left(\frac{m+s}{N+r+s} - \frac{|\mu_n|}{|\mu_n| + |\nu_m|}\right)^2 \to 0 \quad \text{in probability as} \quad N \to +\infty,$$

and

$$\frac{m}{N} \to \frac{|\theta|}{|\theta| + |\omega|} \quad \text{in probability as } N \to +\infty.$$

Proof of the first order lower bound

To establish the lower bound, let $\mathcal{P} = \mathcal{P}_N$ denote any procedure for which $\mathcal{R}(\mathcal{P}) \leq \inf_{\mathcal{P}'} \mathcal{R}(\mathcal{P}') + o(\frac{1}{N})$. Then $\mathcal{R}(\mathcal{P}) \to 0$ as $N \to +\infty$, since the risk approaches zero for equal allocation. To show that $\mathcal{R}(\mathcal{P}) \geq \frac{E(|\theta| + |\omega|)^2}{N+r+s} + o(\frac{1}{N})$, write

$$\mathcal{R}(\mathcal{P}) = E\left\{\frac{\mu_n^2}{m+s} + \frac{\nu_m^2}{n+r} + \frac{1}{(n+r)(m+s)}\right\}$$

$$\geq E\left\{\frac{\mu_n^2}{m+s} + \frac{\nu_m^2}{n+r}\right\}$$

$$\geq \frac{1}{N+r+s}E\left\{(|\mu_n| + |\nu_m|)^2\right\}.$$

The last inequality follows from Lemma (5) part (a). So

$$\mathcal{R}(\mathcal{P}) - \frac{E(|\theta| + |\omega|)^2}{N+r+s} \geq \frac{1}{N+r+s}E\left\{(|\mu_n| + |\nu_m|)^2 - (|\theta| + |\omega|)^2\right\}.$$

Combining Lemmas (2) and (3) it follows that

(5) $\qquad E\left\{(|\mu_n| + |\nu_m|)^2 - (|\theta| + |\omega|)^2\right\} \to 0 \quad \text{as} \quad N \to +\infty.$

Proof of the sufficient condition

At stage N,

$$(N+r+s)\left\{\mathcal{R}(\mathcal{P}) - \frac{E(|\theta| + |\omega|)^2}{N+r+s}\right\} = E\left\{\frac{N+r+s}{(m+s)(n+r)}\right\}$$

$$+ E\left\{ \frac{\mu_n^2}{\frac{m+s}{N+r+s}} + \frac{\nu_m^2}{\frac{n+r}{N+r+s}} - (|\,\mu_n\,| + |\,\nu_m\,|)^2 \right\}$$

$$+ E\left\{ \left(|\,\mu_n\,| + |\,\nu_m\,|\right)^2 - \left(|\,\theta\,| + |\,\omega\,|\right)^2 \right\}.$$

Combining Lemmas (2), (3) it follows that the last line approaches zero as $N \to +\infty$. Since $r > 0$, $s > 0$, and

$$(N + r + s)/(m + s)(n + r) \le 2 \max\{1/(m + s), 1/(n + r)\},$$

it follows that

$$E\left(\frac{N + r + s}{(m + s)(n + r)} \right) \to 0 \quad \text{as} \quad N \to +\infty$$

by the bounded convergence theorem. Combining Lemmas (2), (3), and conditions (ii) and (iii) it follows that the middle line approaches zero as $N \to +\infty$. This concludes the proof of the sufficient condition.

Proof of the necessary condition

Let $\mathcal{P}$ be any procedure for which there is equality in (1). Then Conditions (i) and (ii) follow easily by using lemma(4) and (5). For condition (iii), first observe that

$$(N + r + s)\left\{ \mathcal{R}(\mathcal{P}) - \frac{E(|\,\theta\,| + |\,\omega\,|)^2}{N + r + s} \right\} = E\left\{ \frac{N + r + s}{(m + s)(n + r)} \right\}$$

$$+ E\left\{ \frac{\mu_n^2}{\frac{m+s}{N+r+s}} + \frac{\nu_m^2}{\frac{n+r}{N+r+s}} - (|\,\mu_n\,| + |\,\nu_m\,|)^2 \right\}$$

$$+ E\left\{ \left(|\,\mu_n\,| + |\,\nu_m\,|\right)^2 - \left(|\,\theta\,| + |\,\omega\,|\right)^2 \right\}.$$

Here the left hand side approaches zero as $N \to \infty$, by assumption; the first term on the right approaches zero by the dominated convergence theorem; the second term on the right is non-negative; and the last term on the right approaches zero as $N \to \infty$, by Doob's inequality and the dominated convergence theorem. So

$$E\left\{ \frac{\mu_n^2}{\frac{m+s}{N+r+s}} + \frac{\nu_m^2}{\frac{n+r}{N+r+s}} - (|\,\mu_n\,| + |\,\nu_m\,|)^2 \right\}$$

approaches zero as $N \to \infty$. Moreover since condition(ii) is satisfied, then

$$\frac{\mu_n^2}{\frac{m+s}{N+r+s}} + \frac{\nu_m^2}{\frac{n+r}{N+r+s}} + \frac{N+r+s}{(m+s)(n+r)} - (|\mu_n| + |\nu_m|)^2 \to 0$$

in probability as $N \to \infty$ and since it is non negative, then it is uniformly integrable. See Woodroofe (1982). Therefore condition (iii) is satisfied since $\mu_n^2 N/m$ and $\nu_m^2 N/n$ are bounded above by a uniformly integrable quantity.

ACKNOWLEDGEMENT

BIBLIOGRAPHY

[1] Doob, J. (1953). **Stochastic Processes**. John Wiley & Sons, Inc. New York.

[2] Ford, I. and Silvey, S. D. (1980). A sequentially constructed design for estimating a nonlinear parametric function. *Biometrika.* **67**, 381-388.

[3] Robbins, H., Simons, G., and Starr, N. (1967). A sequential analogue of the Behrens-Fisher problem. *Ann. Math. Stat.* **38**, 1384-1388.

[4] Woodroofe, M. (1982). **Nonlinear Renewal Theory in Sequential Analysis**. SIAM, Philadelphia.

[5] Wu, C. F. J. (1985). Asymptotic inference from a sequential design in a nonlinear situation. *Biometrika* **72**, 553-558.