Mixture Bayesian Robust Design

Han Son Seo*†

* Konkuk University, Department of Applied Statistics

Key Words : Bayesian Design, Optimal Design, Robustness

Abstract

Applying Bayesian optimal design principles is not easy when a prior distribution is not certain. We present a optimal design criterion which possibly yield a reasonably good design and also robust with respect to misspecification of the prior distributions. The criterion is applied to the problem of estimating the turning point of a quadratic regression. Exact mathematical results are presented under certain conditions on prior distributions. Computational results are given for some cases not satisfying our conditions.

1. Introduction

Suppose x is a design variable with experimental region X and y is a random variable with density function $p(y|\theta,x)$ which depends on parameters $\theta^T=(\theta_1,\cdots,\theta_P)$, and design variable x. A design may be represented by a probability measure η on X with finite support. Though an exact design of N design points requires that $\eta_i=N\,\eta(x_i)$ should be an integer for all i. we consider any probability measure η on X as a design. Assume that we are interested in estimating a function of parameters, $g(\theta)$. In Bayesian analysis a usual optimal design criterion is the approximate expected posterior variance of $g(\theta)$ by taking the usual squared error loss $L(\tilde{g}, g(\theta)) = (\tilde{g}-g(\theta))^2$ as a loss function. Posterior expected loss can be approximated using an approximate distribution of θ (Berger 1985). For estimating several functions of θ, $g(\theta)=(g_1(\theta),\cdots,g_k(\theta))^T$ a general criterion is suggested by Chaloner and Larntz (1989). The criterion is the expected weighted trace of the product of a symmetric matrix and the inverse of the information matrix,

$$\phi^*(\eta)=E_\theta(trB(\theta)M(\theta,\eta)^{-1}) \qquad (1)$$

where $B(\theta)$ is a nonnegative definite symmetric p by p matrix and M is the Fisher information matrix, defined as

$$[M(\tilde{\theta},\eta)]_{ij}=-\int\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log(p(y|\theta,x))\right)_{\theta=\tilde{\theta}}\eta(dx).$$

The Bayesian optimal design theory outlined so far assumed that there is only one prior distribution. But in reality there may be several plausible prior distributions. DasGupta and Studden (1991) suggested several optimal robust criteria in normal linear models by formulating uncertainty of the prior in terms of having a family of priors induced by a metric on the space of nonnegative measures. Dette (1990) suggested Bayesian optimal designs which are robust to models in polynomial regression. Seo (2001) deals with robustness aspects by specifying design assumptions in several prior distributions, one of which is assumed to be more favored than others. In this paper we present a new criterion which is designed to be robust with respect to the choice of prior distribution

† 교신저자 hsseo@konkuk.ac.kr

※ This paper was supported by Konkuk University in 2005.

among several plausible prior distributions. The criterion is applied to the problem of estimating turning point of a quadratic regression. Many optimal design criteria were applied to the turning point problem (Chaloner, 1989). Both analytic and numerical results are given demonstrating the robustness of the design criterion.

2. Robust optimal design

2.1 Mixture Bayesian design

Denote the prior distribution for θ as $\Delta(\theta)$. Bayesian design problem with Δ is finding a design which minimize $\Phi(\eta,\Delta)$ over η .

Definition 1.
We call η^* a B-optimal design (for the prior distribution Δ) if η^* minimizes $\Phi(\eta,\Delta)$ among all designs η.

We consider the situation that there are several (n, say) possible prior distributions for θ. Denote these by Δ_i, $i=1,\cdots,n$. For each prior distribution, let $\Phi(\eta,\Delta_i)$ denote a certain function of interest for the design η evaluated for prior distribution Δ_i.

We suggest a new criterion for this situation. Suppose $\{\alpha_1, \alpha_2, \cdots, \alpha_n\}$ are the weights assigned to prior distributions $\{\Delta_1, \Delta_2, \cdots, \Delta_n\}$ respectively. More weight would be assigned to the more plausible prior. The $\alpha-$ mixture of the prior distributions is defined as $\alpha_1\Delta_1+\alpha_2\Delta_2\cdots+\alpha_n\Delta_n$ for fixed$\{\alpha_1, \alpha_2, \cdots, \alpha_n\}$, $\alpha_i \in (0, 1)$, $\sum_{i=1}^{n}\alpha_i=1$.

Definition 2.
For given $\alpha=\{\alpha_1,\alpha_2,\cdots,\alpha_n\}$, where $\alpha_i \in (0, 1)$ and $\sum_{i=1}^{n}\alpha_i=1$ and $\Delta_\alpha=\alpha_1\Delta_1+\alpha_2\Delta_2\cdots+\alpha_n\Delta_n$, $\tilde{\eta}$ is (α-mixture) B_M-optimal design (for the set of prior distributions $\Delta=(\Delta_1,\Delta_2,\cdots,\Delta_n)$ if $\tilde{\eta}$ minimizes $\Phi(\eta,\Delta_\alpha)$ among all designs.

If we let $\Phi(\eta,\Delta)=E_\theta(tr\,B(\theta)M(\theta,\eta)^{-1})$, then we have the following results.

(a) Design criteria $\Phi(\eta,\Delta_\alpha)$ for B_M-optimal designs are convex function of η.

(b) The directional derivative for the B_M-optimal design criterion at η_1 in the direction of η_2 is

$$F(\eta_1, \eta_2) = \Phi(\eta_1,\Delta_\alpha) - E_\theta(tr\,B(\theta)M(\theta,\eta_1)^{-1}M(\theta,\eta_2)M(\theta,\eta_1)^{-1}) \quad (2)$$

the expectation being taken over the distribution for θ, Δ_α.

2.2 Example : Turning point problem

The following quadratic model is considered with observation y_i and design points x_i

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + e_i \quad (3)$$

where $\beta=(\beta_0,\beta_1,\beta_2)$ are unknown coefficients and the errors e_i are independent, normally distributed with mean zero and variance σ^2. Under the model (3) the expected value of y is a maximum if β_2 is negative, or is a minimum if β_2 is positive at the value of $x=-\beta_0/2\beta_2$, denoted by γ, which is called the turning point.

2.2.1 Case A : Analytic results

The B-optimal design for estimating the turing point γ is presented by Mandal (1978). Without loss of generality design region X is restricted to be in the interval [-1, 1]. Since the B-optimal design depends on β_0, β_2 and γ only through the first two moments of distribution of γ, the prior for γ can be summarized as the vector, $\Delta=(m,\nu)$ where $m=E(\gamma)$ and $\nu=\mathrm{var}(\gamma)$.

For a large sample size, if σ^2 and β_2 are assumed to be independent of γ in the prior distribution the expected posterior variance of γ

is proportional to

$$\Phi(\eta, \Delta) = E_{\gamma}(\mathrm{tr}B(\gamma)I(\theta, \eta)^{-1})$$

where $B(\gamma) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 2\gamma \\ 0 & 2\gamma & 4\gamma^2 \end{pmatrix}$

and

$$\Phi(\eta, \Delta) = \frac{1}{d^2}\frac{1}{\mu} + \frac{\frac{4\nu}{d^2} + \left\{ \frac{2(m-c)}{d} - \frac{(\mu'_3 - \mu'_2\mu'_1)}{\mu_2} \right\}^2}{d^2\left\{ \mu'_4 - \mu'^2_2 - \frac{1}{\mu_2}(\mu'_3 - \mu'_2\mu'_1)^2 \right\}}$$

where

$x_{(1)}$ = minimum value of supporting points of a design η,

$x_{(N)}$ = maximum value of supporting points of a design η,

$$z = \frac{2x - x_{(1)} - x_{(N)}}{x_{(N)} - x_{(1)}}, \ \mu'_r = \int z^r \eta(dz),$$

$$\mu_r = \int (z - \mu_1)^r \eta(dz)$$

$$c = \frac{x_{(1)} + x_{(N)}}{2} \text{ and } d = \frac{x_{(N)} - x_{(1)}}{2}.$$

A B-optimal design for the special case of the prior centered at 0, i.e. $\Delta = (0, \nu)$ is

$$\eta^*(-1) = \eta^*(1) = \mu_2'^*/2,$$
$$\eta^*(0) = 1 - \mu_2'^* \tag{4}$$

where $\mu_2'^* = \{1 + 2(\nu^{-1} + 4)^{-1/2}\}^{-1}$.

We now apply a B_M-optimal design criterion to the turning point problem of a quadratic regression. Under certain conditions on the prior distributions, we establish analytic results.

Theorem 1.

For the turning point problem, consider prior distributions, $\Delta_i = (m_i, \nu_i)$ with $\nu_i > 0$ and $m_i \in R$ $i=1, \cdots, n$.

For given $\alpha = \{\alpha_1, \alpha_2, \cdots, \alpha_n\}$, where $\alpha_i \in (0, 1)$ and $\sum_{i=1}^{n} \alpha_i = 1$, if $\sum_{i=1}^{n} \alpha_i m_i = 0$ then a α-mixture B_M-optimal design (for the set of prior distributions $\Delta = (\Delta_1, \Delta_2, \cdots, \Delta_n)$ is given by

$$\bar{\eta}(-1) = \bar{\eta}(1) = \frac{\mu_2'^*}{2} \text{ and } \bar{\eta}(0) = 1 - \mu_2'^*$$

where

$$\mu_2'^* = \left\{ 1 + 2\left(\frac{1}{\sum_{i=1}^{n} \alpha_i \{\nu_i + m_i^2\}} + 4 \right)^{-\frac{1}{2}} \right\}^{-1}.$$

Proof.

It is easily verified that for $\{\Delta_1, \Delta_2, \cdots, \Delta_n\}$ with $\Delta_i = (m_i, \nu_i)$, $\alpha_i \in (0, 1)$ and $\sum_{i=1}^{n} \alpha_i = 1$,

$$\Delta_\alpha = \sum_{i=1}^{n} \alpha_i \Delta_1$$
$$= (\sum_{i=1}^{n} \alpha_i m_i, \ \sum_{i=1}^{n} \alpha_i(\nu_i + m_i^2) - \{\sum_{i=1}^{n} \alpha_i m_i\}^2).$$

So

$$\Delta_\alpha = \sum_{i=1}^{n} \alpha_i \Delta_1$$
$$= (\sum_{i=1}^{n} \alpha_i m_i = 0, \ \sum_{i=1}^{n} \alpha_i(\nu_i + m_i^2)).$$

The B-optimal design only depends on the first two moments of the prior distribution. So we apply the equation (4) and have the required result.

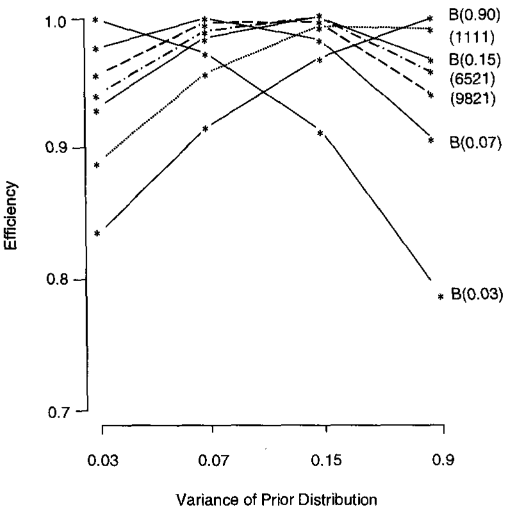
B_M-optimal design is found analytically for any value of $\alpha = \{\alpha_1, \alpha_2, \cdots, \alpha_n\}$ for the turning point problem that results in the weighted average of the mean being at the center of the design region.

Based on the analytic results we demonstrate the robustness of the mixture designs by presenting efficiencies of mixture designs for several examples of prior distributions.

We use the efficiency as a measure to assess the relative worth of a design η for prior distribution Δ. The efficiency of a design η with respect to the prior distribution Δ is defined by

$$EFF(\eta, \Delta) = \inf_X \Phi(X, \Delta) / \Phi(\eta, \Delta)$$
$$= \Phi(\eta^*, \Delta) / \Phi(\eta, \Delta)$$

where η^* is a B-optimal design for prior distribution Δ.



<Figure 1> Efficiencies of B_M-optimal designs compared to B-optimal designs based on analytic results for prior distributions Δ_1=(0, 0.03), Δ_2=(0, 0.07), Δ_3=(0, 0.15), Δ_4=(0, 0.90). For example, B(0.03) stands for the B-optimal design for Δ=(0, 0.03) and (1,1,1,1) stands for the B_M-optimal design with α=(1/4, 1/4. 1/4, 1/4).

Consider the following four prior distributions, Δ_1=(0, 0.03), Δ_2=(0, 0.07), Δ_3=(0, 0.15), Δ_4=(0, 0.90). All four prior distributions have mean at the center of the design region, so the theorem 1 applies. To illustrate, we use integer numbers from 1 to 9 as relative weights for each prior distribution. For example (6551) represents the mixture giving weight 6/17 to Δ_1, 5/17 to Δ_2, 5/17 to Δ_3, 1/17 to Δ_4. <Figure 1> gives efficiencies for B_M-optimal designs for several sets of weights. Not all sets of weights result in robust designs for all prior distributions. In essence, robust design moderate the extreme efficiencies without losing much to the B-optimal design for each prior alone. (1111) is robust compared to the B-optimal designs for Δ_1 and Δ_4 but not for Δ_2 and Δ_3. But some sets of weights, for example (6551) and (9821), yield designs which are robust for all prior distributions.

2.2.2 Case B : Numerical results

We now use numerical methods to find mixture Bayesian optimal designs. To verify the optimality of a design we use Whittle's (1973) general equivalence theorem.

Let Ξ to be the set of all probability measures on X. Then the design problem can be viewed as finding a measure in Ξ that minimizes $\phi(\eta)$. We only consider convex functions ϕ on Ξ. For non-linear design, we characterize optimal designs using directional derivatives of the criterion with respect to design measures.

For two measures η_1 and η_2 in Ξ, the directional derivative at η_1 in the direction of η_2 is denoted by $F(\eta_1, \eta_2)$ and is defined, when the limit exists, by

$$F(\eta_1, \eta_2) = \lim_{\varepsilon \downarrow 0} \varepsilon^{-1}[\phi((1-\varepsilon)\eta_1 + \varepsilon\eta_2) - \phi(\eta_1)]$$

The extreme points of Ξ are the measures which put point mass at a single point x in X. We denote such a measure by η_x and use $d(\eta, x)$ to denote $F(\eta, \eta_x)$. With this notation, we give Whittle's general equivalence theorem which characterizes optimal designs.

Theorem 2. (Whittle, 1973)
Assume that X is compact, that directional derivatives exist and are continuous in x, that there is at least one measure in Ξ for which ϕ is finite, and that ϕ is such that if $\eta_i \rightarrow \eta$ in weak convergence then $\phi(\eta_i) \rightarrow \phi(\eta)$.

If ϕ is convex, then a ϕ-optimal design η_0 can be equivalently characterized by any of the three conditions

(i) η_0 minimizes $\phi(\eta)$.

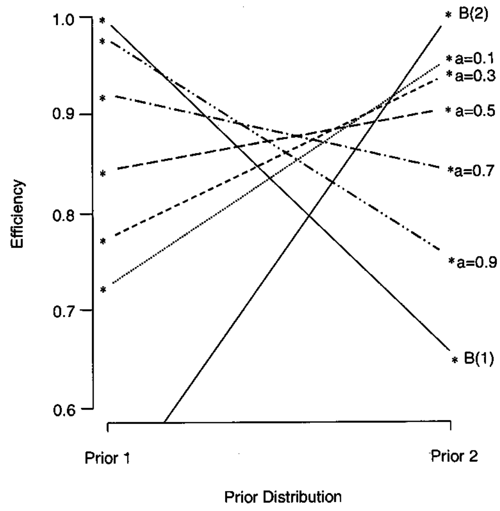
(ii) η_0 maximizes $\inf_{x \in X} d(\eta, x)$

(iii) $\inf_{x \in X} d(\eta_0, x) = 0$

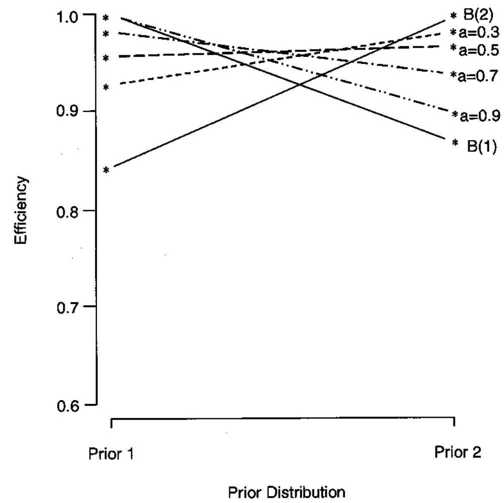
Since the criterion ϕ^* in the equation (1) is convex function of η in Ξ we can apply theorem 2 to our problem. From equation (1) and equation (2), directional derivatives in the direction of η_x, $d(\eta, x)$, of B- and B_M-optimal designs are fourth degree polynomials. By theorem 2 it is clear that the number of supporting points of optimal design is three. We use the Nelder and Mead (1965) version of the simplex algorithm finding the best three points design. For simplicity, we only work with two prior distributions. With two prior distributions, we consider two cases. : First, both two prior distributions are certain, i.e., have small variances. Second, neither prior distribution 1 nor the minor prior distribution 2 is certain. Corresponding to the two cases, we take prior distributions as follows in our examples : a) $\Delta_1 = (-0.2, 0.07)$ $\Delta_2 = (0.5, 0.07)$, b) $\Delta_1 = (-0.2, 0.30)$ $\Delta_2 = (0.5, 0.30)$.

<Figure 2> B_M-optimal designs based on numerical results for prior distributions and weights $\Delta_1 = (-0.2, 0.07)$ $\Delta_2 = (0.5, 0.07)$ and $\alpha = (.5, .5)$.

B_M-optimal designs were found for various values of α for each case. We could check the optimality of the designs by plotting the derivative function, which is positive almost everywhere and has exactly three roots at the design points. <Figure 2> show B_M-optimal designs with $\Delta_1 = (-0.2, 0.07)$ $\Delta_2 = (0.5, 0.07)$ and $\alpha = (.5, .5)$. <Figure 3> indicates that the B_M-optimal design criterion yields robust designs for two cases with various values of the weights. For more than 3 priors cases B_M-optimal designs were found by a similar procedure using numerical method of optimality.



(a)



(b)

<Figure 3> Efficiencies of B_M-optimal designs compared B-optimal designs based on numerical results for prior distributions (a) $\Delta_1 = (-0.2, 0.07)$ $\Delta_2 = (0.5, 0.07)$ (b) $\Delta_1 = (-0.2, 0.30)$ $\Delta_2 = (0.5, 0.30)$, a = 0.1 stands for B_M-optimal designs with $\alpha = (0.1, 0.9)$.

3. Conclusion

We have addressed the problem of handling uncertainty of the prior distribution in Bayesian optimal design and suggested a robust Bayesian optimal design criterion. The criterion is applied to the turning point problem in the quadratic regression, and analytic results for mean-centered prior distributions and numerical results for non mean-centered prior distributions were presented. In our example with two prior distributions, we found reasonably good robust designs for all prior distributions using the B_M-optimal design criterion. If we have several competing prior distributions, it is generally more difficult to find a design which may be considered robust for all prior distributions. B_M-optimal design criterion can be applied to any problems.

References

[1] Berger, J. O.(1985), *Statistical Decision Theory and Bayesian Analysis*, Springer-Verlag, New York.

[2] Chaloner, K.(1989), "Bayesian Design for Estimating the Turning Point of a Quadratic Regression", *Communication in Statistics-Simulation and Computation*, Vol. 13, pp. 723-731.

[3] Chaloner, K. and Larntz, K.(1989), "Optimal Bayesian Design Applied to Logistic Regression Experiments", *Journal of Statistical Planning and Inference*, Vol. 21, pp. 191-208.

[4] DasGupta, A. and Studden, W. J.(1991), "Robust Bayesian Experimental Designs in Normal Linear Models", *The Analysis of Statistics*, Vol. 19, pp. 1244-1256.

[5] Dette, H.(1990), "A Generalization of D- and D_l-optimal designs in polynomial regression", *The Analysis of Statistics*, Vol 18, pp. 1784-1804.

[6] Mandal, N. K.(1978), "On Estimation of the Maximal Point of a Single Factor Quadratic Response Function", *Calcutta Statistical Association Bulletin*, Vol. 27, pp. 119-125.

[7] Nelder, J. A. and Mead, R.(1965), "A Simplex Method for Function Minimization", *Computer Journal*, Vol. 7, pp. 308-313.

[8] Seo, H. S.(2001), "Restricted Bayesian Optimal Designs in Turning Point Problem", *JKSS*, Vol. 30, pp. 163-178.

[9] Whittle, P.(1973), "Some General Points in the Theory of Optimal Experimental design", *Journal of Royal Statistical Society*, Series B, Vol. 35, pp. 123-130.