

# Understanding Bayesian Experimental Design with Its Applications

Gunhee Lee<sup>a,1</sup>

<sup>a</sup>Graduate School of Business, Sogang University

(Received September 16, 2014; Revised October 25, 2014; Accepted October 27, 2014)

---

## Abstract

Bayesian experimental design is a useful concept in applied statistics for the design of efficient experiments especially if prior knowledge in the experiment is available. However, a theoretical or numerical approach is not simple to implement. We review the concept of a Bayesian experiment approach for linear and nonlinear statistical models. We investigate relationships between prior knowledge and optimal design to identify Bayesian experimental design process characteristics. A balanced design is important if we do not have prior knowledge; however, prior knowledge is important in design and expert opinions should reflect an efficient analysis. Care should be taken if we set a small sample size with a vague improper prior since both Bayesian design and non-Bayesian design provide incorrect solutions.

Keywords: Experimental design, Bayesian method, Bayesian decision theory, Monte-Carlo method.

---

## 1. 베이زي안 실험계획법의 구조

최적화된 실험설계는 자료가 수집되기 이전에 결정되어야 하고, 설령 자료가 존재한다고 하여도 특정한 형태의 실험설계 조건아래서 자료 수집이 가능하기 때문에 연구자 입장에서 실험설계를 선택하는 문제는 대단히 어려운 문제이다. 더군다나 전문가로부터 모수에 대한 사전적 정보가 존재하여 이를 기반으로 최적화된 실험설계를 찾고자 한다면, 기존의 우도함수 기반 최적화 접근방법에는 한계가 존재하며, 따라서 베이زي안적 접근방법 이외에는 해결방법이 없게 된다. 베이زي안 실험계획법은 Raiffa와 Schlaifer (1961)와 Lindley (1972)의 의사결정론적 접근방법에서 시작된다. 베이زي안 실험계획법의 구조를 간단하게 소개하면 다음과 같다.

만일 실험설계  $\eta$ 가 적당한 공간,  $H$ 에서 선택되고 표본공간  $X$ 로부터 표본  $x$ 가 관찰되고, 모수공간인  $\Theta$ 로부터 모수인  $\theta$ 가 설정되었다면, 효용함수  $U(\cdot)$ 에 대하여 실험설계  $\eta$ 에 대한 기대 효용은 다음과 같이 표현된다.

$$U(\eta) = \int_X \int_{\Theta} U(\theta, \eta, x) \cdot p(\theta|x, \eta) \cdot p(x|\eta) d\theta dx. \quad (1.1)$$

결국 베이زي안 측면에서의 최적화된 실험설계법은 식 (1.1)에서 표현된 기대효용함수  $U(\eta)$ 를 최대화 시키는  $\eta^*$ 를 찾는 과정이라고 볼 수 있다.

---

<sup>1</sup>Graduate School of Business, Sogang University, Seoul 121-742, Korea. E-mail : [ghlee@sogang.ac.kr](mailto:ghlee@sogang.ac.kr)

그렇다면 효용함수  $U(\cdot)$ 를 어떻게 선택할 것인가? 다양한 형태의 효용함수가 있겠지만, 기본적인 원칙은 실험목적에 맞는 함수를 선택하여야 한다는 것이다. 베이저안 실험계획에서 가장 많이 사용되는 효용함수는 Shannon 정보량 (Shannon, 1948)으로 이를 최대화 시키는 실험설계는 사전분포와 사후분포간의 차이를 나타내는 Kullback-Leibler 거리를 최대화 시키는 것과 동일한 것으로 다음과 같이 정의된다.

$$\max_{\eta \in \mathcal{H}} \int_X \int_{\Theta} \log \frac{p(\theta|x, \eta)}{p(\theta)} \cdot p(x, \theta|\eta) d\theta dx. \quad (1.2)$$

식 (1.2)에 나타난 사전분포는 실험설계  $\eta$ 와 관계없으므로 기대효용은 모수  $\theta$ 에 대한 사후분포의 Shannon 정보량 기대치로써 다음과 같이 표현된다.

$$U_1(\eta) = \int_X \int_{\Theta} \log[p(\theta|x, \eta)] \cdot p(x, \theta|\eta) d\theta dx = E_{\theta, X}[\log(p(\theta|X, \eta))], \quad (1.3)$$

여기서,  $\theta$ 의 사후분포에 대한 Shannon 정보량 기대값을 효용함수에 적용시켰다는 의미는 실험목적이  $\theta$ 에 대한 추론으로 나타났을 경우 적절하다는 의미를 가지고 있다. 만일 실험의 주요목적이 예측이라면  $\theta$ 보다는 예측에 해당하는  $x_{n+1}$ 에 대한 예측분포(predictive distribution)를 기반으로 효용함수가 정의되는 것이 바람직하며, 효용함수는 예측치  $x_{n+1}$ 에 대한 예측분포의 Shannon 정보량 기대치 또는 사전 예측분포와 사후예측분포간의 차이를 나타내는 Kullback-Leibler 거리가 된다. 이를 정리하여 표현하면 다음과 같다.

$$U_2(\eta) = \int \log[p(x_{n+1}|x, \eta)] \cdot p(x, x_{n+1}|\eta) dx dx_{n+1} = E[\log(p(X_{n+1}|X, \eta))]. \quad (1.4)$$

결국 베이저안 실험계획법은 실험목적에 맞는 효용함수를 정의하고, 효용함수를 최대화시키는 실험설계를 찾는 과정을 의미한다. 본 연구에서는 통계분석모형에서 나타나는 효용함수의 모습을 알아보고 최적화 실험설계를 찾기 위한 방법과 이에 대한 특징을 예를 통하여 알아보기로 한다.

## 2. 통계분석모형에서 나타나는 베이저안 실험계획법

### 2.1. 선형모형

일반적인 선형모형 형태인  $Y = X\theta + \epsilon$ 에서  $\epsilon \sim N(0, \sigma^2 I)$ 을 가정하고(여기서  $I$ 는 단위행렬),  $k$ 개의 모수로 구성된  $\theta$ 에 대한 사전분포를 공액분포인 정규분포로 가정하면, 다음과 같이 정리할 수 있다.

$$\text{선형 모형의 가정: } [y|\theta, \sigma] \sim N(X\theta, \sigma^2 I).$$

$$\theta \text{에 대한 사전분포: } [\theta|\sigma] \sim N(\theta^0, \sigma^2 R^{-1}).$$

$$\theta \text{에 대한 사후분포: } [\theta|x, y, \sigma] \sim N\left(\left(x^T x + R\right)^{-1} \left(x^T y + R\theta_0\right), \sigma^2 \left(x^T x + R\right)^{-1}\right).$$

$\theta$ 에 대한 사후분포를 보다 간단하게 표현하고, 최적화 실험계획에 적용되는 요소인  $\eta$ 의 의미를 강조하기 위하여 디자인 행렬이라고 부르는  $x$  행렬을 기반으로  $M(\eta) = (x^T x)/n$ 을 정의하면 다음과 같이 재 표현된다.

$$[\theta|x, y, \sigma] \sim N\left((nM(\eta) + R)^{-1} \left(x^T y + R\theta_0\right), \sigma^2 (nM(\eta) + R)^{-1}\right).$$

이제  $\theta$ 에 대한 사후분포가 확정되었으므로 식 (1.3)에서 소개된 Shannon 정보량의 기대치를 계산하면 다음과 같다.

$$U_1(\eta) = -\frac{k}{2} \log(2\pi) - \frac{k}{2} + \frac{1}{2} \log(\det(\sigma^{-2}(nM(\eta) + R))),$$

$$\sim \det(nM(\eta) + R). \tag{2.1}$$

결국 최적화된 실험설계는 행렬  $(nM(\eta) + R)$ 의 행렬식을 최대화시키는 문제로 바뀌게 되며, 이를 베이زي안 D-optimality라고 한다. 만일 사전분포에서 나타난 정보행렬을 0행렬로 설정하면, 비(非) 베이زي안(non-Bayesian) D-optimality가 된다. 따라서 비 베이زي안 D-optimality는 베이زي안 D-optimality의 특별한 경우라고 볼 수 있다. 그렇다면 베이زي안 방법과 비(非) 베이زي안 방법이 일치하는 사전정보에 어떠한 정보도 주지 않는 즉, R행렬에 0 행렬을 할당하는 방법은 객관적 분석결과를 얻기 위하여 바람직한 방법이라고 볼 수 있을까? 베이زي안 방법과 비 베이زي안 방법에서 일치할 결과가 나타났다는 사실 자체로 이 방법은 상당히 매력적인 방법이라고 볼 수 있다. 하지만 이 방법의 우월성을 보여주기 위해서는  $\eta$ 와  $U_1$ 의 분석적 관계를 통하여 사전확률 분포가 무정보적이면서 부적합함에도 불구하고(noninformative improper prior) 안정적이고 우월한 실험설계를 제공하여 준다는 보다 구체적인 증거가 요구된다. 이에 대한 의문은 4장에서 자세하게 논의하기로 하고, 이번에는 실험연구의 목적이 모수추정이 아니라 독립변수  $x_{n+1}$  값에 대한 예측이라면, 효용함수는 예측분포의 기대값으로써 식 (1.4)은 다음과 같이 정리할 수 있다.

$$U_2(\eta) \propto -\frac{1}{2}\sigma^2 \left[ x_{n+1}^T (nM(\eta) + R)^{-1} x_{n+1} + 1 \right]. \tag{2.2}$$

이러한 형태의 효용함수를 통하여 얻은 최적화된 해를 베이زي안 c-optimality 라고 한다. 결국, 예측이 주요 목적이라면 베이زي안 c-optimality를 기반으로 하는  $x_{n+1}^T (nM(\eta) + R)^{-1} x_{n+1}$ 을 최소화 하는  $\eta$ 를 찾아야 한다는 것이다. 만일  $\sigma$ 를 모르는 경우의 선형모형을 고려해 보면,  $(\theta, \sigma)$ 의 공액분포인 정규분포와 역감마분포를 사전분포로 설정하여 정리하면 다음과 같다.

$$\begin{aligned} \theta \text{에 대한 조건부 사전분포: } & [\theta|\sigma] \sim N(\theta_0, \sigma^2 R^{-1}). \\ \sigma \text{에 대한 사전분포: } & [\sigma^{-2}|\alpha, \beta] \sim \text{Ga}(\alpha, \beta) \propto (\sigma^{-2})^{\alpha+1} \exp(-\beta\sigma^{-2}). \\ \Rightarrow \theta \text{에 대한 조건부 사전분포: } & [\theta] \sim \text{MVT} \left( k, 2\sigma, \theta_0, \frac{\beta}{\alpha} R^{-1} \right). \end{aligned}$$

여기서 MVT는 K변량 t-분포를 의미하는 것으로 자유도는  $2\alpha$ , 중심모수는  $\theta_0$ , 척도모수는  $(\beta/\alpha)R^{-1}$ 임을 의미한다. 이와 같이 공액사전분포를 설정하게 되면, 최종적인  $\theta$ 에 대한 사후분포는 다음과 같이 요약될 수 있다 (Box와 Tiao, 1973).

$$[\theta|x, y] \sim \text{MVT} \left( k, 2\alpha + n, \theta^*, h(x, y)(nM(\eta) + R)^{-1} \right), \tag{2.3}$$

여기서  $\theta^* = (nM(\eta) + R)^{-1}(x^T y + R\theta_0)$ 이고,  $h(x, y) = (2\alpha + n)^{-1}[(y - s\theta_0)^T(I - x(nM(\eta) + R)^{-1})(y - x\theta_0) + 2\beta]$ 이다.

또한  $x_{n+1}$ 에 대한 예측치  $y_{n+1}$ 에 대한 예측분포는 다음과 같이 정리된다 (Chaloner과 Verdinelli, 1995).

$$[y_{n+1}|x, y] \sim \text{MVT} \left( k, 2\alpha + n, (x_{n+1}\theta^*), h(x, y) (x_{n+1}(nM(\eta) + R)^{-1} x_{n+1} + 1) \right). \tag{2.4}$$

식 (1.2)와 식 (1.3)에 소개된 효용함수를 얻기 위하여 계산된 중간과정에서 나타나는 사후분포인 식 (2.3)이나 식 (2.4)에 대한 기대값을 계산하는 문제는 대수적으로 해결할 수 없는 문제로 적분의 수치 해석적인 방법, 특별히 MCMC(Markov Chain Monte Carlo)를 이용한 접근방법에 대하여 소개하고자 한다.

## 2.2. 비선형모형

통계분석모형이 비선형일 경우에는 식 (1.3)에 나타난 효용함수의 기대치를 최대화시키는 문제는 훨씬 복잡하게 나타나므로 사후분포의 정규근사를 이용하여 접근하는 것이 일반적이다. 이 경우 피셔의 기대 정보량,  $nI(\theta, \eta)$ 을 사용하거나 우도함수 또는 사후분포의 2차 미분을 분산으로 사용하여 접근하게 된다. 이 경우  $\theta$ 에 대한 사후분포는 다음과 같이 나타난다 (Berger, 1985).

$$[\theta|x, \eta] \sim N\left(\hat{\theta}, [nI(\hat{\theta}, \eta) + \ddot{R}]^{-1}\right), \quad (2.5)$$

여기서  $\hat{\theta}$ 은 일반화된 최대우도 추정치라 불리는 사후분포의 최빈값이 되며,  $\ddot{R}$ 은 사전확률밀도함수의 2차 미분 행렬을 의미한다. 피셔 정보량을 피셔정보량의 기대치(expected Fisher information matrix)을 사용할지 아니면 관찰된 피셔정보량(observed Fisher information matrix)을 사용할 지에 대한 논란은 끊임없이 제기되고 있지만 아직까지 어느 것이 우월하다는 결론은 내릴 수 없는 상태이다. 피셔정보량의 기대치를 기반으로 식 (1.3)에서 나타나는 기대 효용함수를 정리하면 다음과 같다.

$$\begin{aligned} U_1(\eta) &= -\frac{k}{2} \log(2\pi) - \frac{k}{2} + \frac{1}{2} \int \log \det[nI(\theta, \eta) + R] p(\theta) d\theta, \\ &\propto \int \log \det[nI(\theta, \eta) + R] p(\theta) d\theta. \end{aligned} \quad (2.6)$$

식 (2.6)에 나타난 구체적인 식을 확인하기 위하여 베르누이 반응을 하는 와 이에 대응되는 독립변수 간에 다음과 같은 성공확률을 갖는 로지스틱 회귀모형을 고려하여 보자 (Chaloner과 Larntz, 1989).

$$p(y = 1|x, \mu, \beta) = \frac{1}{1 + \exp(-\beta(x - \mu))}.$$

이 경우 모수  $\mu$ 는 반응확률 50%에 해당하는  $x$ 값을 의미하고,  $\beta$ 는 확률의 로짓 스케일에 대한 기울기를 의미한다. 만일  $k$ 개의 서로 다른  $x_i$  값들이 있어, 이를 기반으로 실험설계,  $\eta$ 가 이루어진다면( $i = 1, 2, \dots, k$ ), 관찰된 피셔정보량은 다음과 같이 정리된다.

$$I(\theta, \eta) = \begin{pmatrix} \beta^2 t & -\beta t(\bar{x} - \mu) \\ -\beta t(\bar{x} - \mu) & s + t(\bar{x} - \mu)^2 \end{pmatrix}. \quad (2.7)$$

여기서  $\bar{x}$ 는 가중평균을 나타내는,  $t^{-1} \sum_{i=1}^k n_i w_i x_i$ 이고,  $s$ 는 가중 제곱합  $\sum_{i=1}^k n_i w_i (x_i - \bar{x})^2$ 을 의미하며, 여기서  $n_i$ 는 실험설계  $\eta$ 를 구체적으로 표현한  $x_i$  값에 해당하는 관찰값의 수를 나타내고, 가중치  $w_i$ 는 베르누이 분포의 분산을 나타내는  $p(y = 1|x, \mu, \beta)(1 - p(y = 1|x, \mu, \beta))$ 의 의미이며,  $t$ 는 가중 총합을 나타내는  $\sum_{i=1}^k n_i w_i$ 을 의미한다.

결국 식 (2.6)에 나타난 효용함수를 로지스틱 회귀모형의 관찰된 피셔정보량에 대입하여 정리하면 다음과 같은 식이 된다.

$$U_1(\eta) \propto E_{\mu, \beta} [\log(\beta^2 t s)].$$

효용함수를 최대화시키는 최적의 실험계획을 찾는 방법은 Nelder와 Mead (1965)가 제시한 심플렉스 방법이나 수치해석적인 적분방법이 주로 사용되어 왔다. 하지만 1990년 이후 소개된 몬테칼로 시뮬레이션 기반을 하는 접근방법이 보편화되면서 (Tanner, 1998) 문제를 쉽게 해결할 수 있게 되었다. 다음 장에서는 이러한 몬테칼로 시뮬레이션을 통한 효용함수 최대화방법에 대하여 설명하고자 한다.

### 3. 몬테칼로 적분을 통한 베이저안 실험계획법

일반적으로 식 (1.2)에 나타난 기대 효용치는 간단한 형태가 아니므로 수치해석적이나 대수적으로 문제를 해결하는 것은 매우 어렵다. 문제 접근을 쉽게 하기 위하여 몬테칼로 시뮬레이션으로 접근하기 쉬운 형태로 기대 효용치를 변환시켜 보기로 한다. 먼저 기대 효용치를 계산하기 위하여 사용되는 확률,  $p(\theta, x|\eta)$ 는 조건부 확률의 곱셈법칙에 의하여  $p(x|\theta, \eta) \times p(\theta|\eta)$ 로 표현될 수 있고, 사후확률을 나타내는  $p(\theta|x, \eta)$ 는 베이즈 정리에 의하여 우도함수와 사전분포의 곱의 형태로 표현될 수 있으며, 식 (1.2)에서 언급한 바와 같이  $\theta$ 에 대한 사전확률,  $p(\theta|\eta)$ 는 실험설계를 결정하는  $\eta$ 와 무관하므로  $p(\theta)$ 으로 표현되어 다음과 같은 식이 유도된다.

$$p(\theta|x, \eta) = \frac{p(x|\theta, \eta) \times p(\theta|\eta)}{\int_{\Theta} p(x|\theta, \eta) \times p(\theta|\eta) d\theta} = \frac{p(x|\theta, \eta) \times p(\theta)}{p(x|\eta)}. \quad (3.1)$$

Kullback-Leibler 거리를 기반으로 하는 기대 효용치를 나타내는 식 (1.2)에  $\theta$ 에 대한 사후확률을 나타내는 식 (3.1)을 대입하면 다음과 같은 형태의 식이 유도된다 (Huan과 Marzouk, 2013).

$$\begin{aligned} U_1(\eta) &= \int_X \int_{\Theta} \log \frac{p(\theta|x, \eta)}{p(\theta)} \cdot p(x, \theta|\eta) d\theta dx \\ &= \int_X \int_{\Theta} \log \frac{p(x|\theta, \eta) \times p(\theta)}{p(x|\eta)} \cdot p(x, \theta|\eta) d\theta dx \\ &= \int_X \int_{\Theta} \log \frac{p(x|\theta, \eta)}{p(x|\eta)} \cdot p(x|\theta, \eta) p(\theta) d\theta dx \\ &= \int_X \int_{\Theta} [\log p(x|\theta, \eta) - \log p(x|\eta)] \cdot p(x|\theta, \eta) p(\theta) d\theta dx. \end{aligned} \quad (3.2)$$

식 (3.2)의 형태로 나타난 기대 효용치를 몬테칼로 적분으로 계산하려면, 먼저  $\theta$ 의 사전분포,  $p(\theta)$ 로부터  $\theta^{(i)}$ 를 추출하고,  $x$ 를 우도함수,  $p(x|\theta^{(i)}, \eta)$ 로부터  $x^{(i)}$ 를 추출하는 과정을  $n_1$ 번 반복하여 평균값을 계산하면 얻을 수 있다. 하지만, 식 (3.2)에서 나타난  $p(x|\eta)$ 는  $p(x|\theta, \eta)$ 의 기대치 또는 적분값으로 표현되는 또 다른 형태의 몬테칼로 적분을 요구하게 되며, 이를 위하여  $p(\theta)$ 로부터  $\theta$ 를 다시 한 번 추출하고, 이를  $\theta^{(i,j)}$ 라 표현하여  $n_2$ 번 반복하고 평균을 계산하면  $x$ 에 대한 주변확률분포를 얻을 수 있다. 이를 수식으로 표현하면 다음과 같다.

$$p(x^{(i)}|\eta) = \int_{\Theta} p(x^{(i)}|\theta, \eta) p(\theta|\eta) d\theta = \frac{1}{n_2} \sum_{j=1}^{n_2} p(x^{(i)}|\theta^{(i,j)}, \eta). \quad (3.3)$$

결국 식 (3.2)에 나타난 기대효용치는 식 (3.3)과 함께 다음과 같이 표현된다.

$$U_1(\eta) \approx \frac{1}{n_1} \sum_{i=1}^{n_1} \log p(x^{(i)}|\theta^{(i)}, \eta) - \log p(x^{(i)}|\eta). \quad (3.4)$$

실험설계를 나타내는 각각의  $\eta$ 에 대하여 식 (3.4)에서 나타난 기대효용치를 계산하는 것은  $n_1$ 의 반복을 통한  $\theta^{(i)}$  및  $x^{(i)}$ 의 생성과  $\log p(x^{(i)}|\eta)$ 를 계산하기 위한  $n_2$ 번의 몬테칼로 적분을 적용하기 때문에 간단한 계산은 아니다. 예를 들어,  $\eta$ 에 대한 실험설계 유형이 1,000가지가 있고  $n_1$ 과  $n_2$ 를 10만번씩 반복한다면, 최적 설계를 위하여  $10^{13}$ 번의 반복이 요구되는 과정을 거쳐야 하지만 요즘의 컴퓨팅 기술로는 그리 어려운 계산은 아닐 것이다.

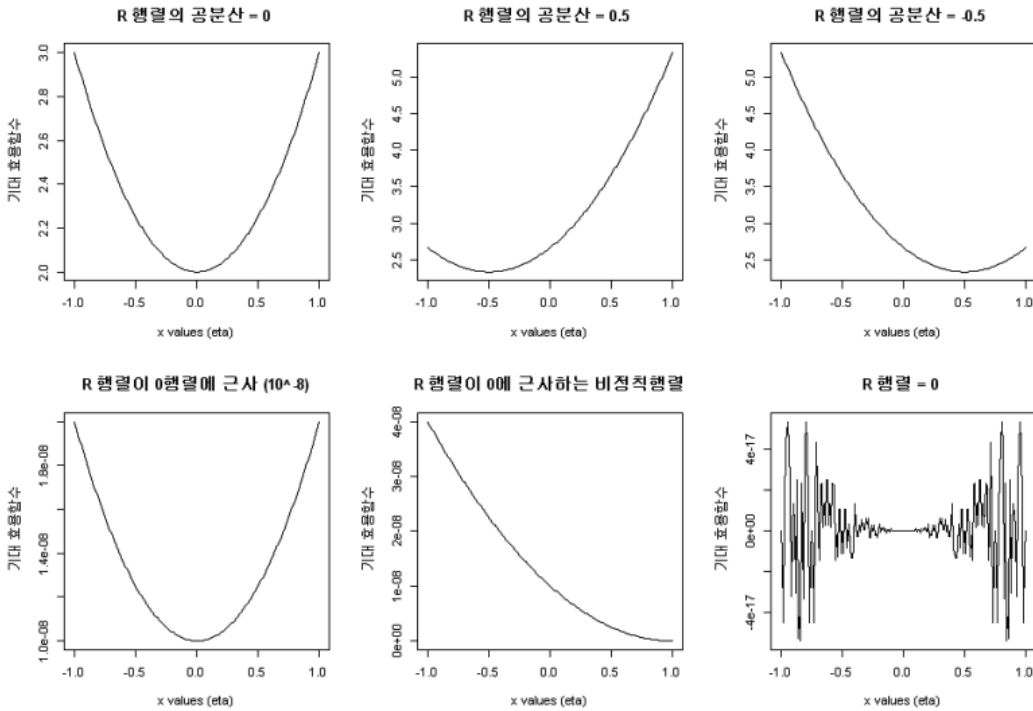


Figure 4.1. Expected Utility with different priors, R matrix ( $k = 1$ )

#### 4. 모의실험 연구

실험설계와 효용함수의 특징을 알아보기 위하여 먼저 분산이 알려진 선형모형의 경우를 고려하여 보자. 문제를 간단하게 만들기 위하여  $x$ 가 취하는 값이  $-1$ 에서  $1$  사이에 있다고 가정하고,  $x$ 값이 변하면서, 그리고 사전정보 행렬인 R행렬이 변하면서 기대 효용치 함수가 어떻게 변하는지를 살펴보았다. 회귀모형의 절편과 기울기를 나타내는  $\theta = (\alpha, \beta)$ 에 대한 사전분포의 분산이 각각 1이라고 가정하고, 공분산이 0일 경우, 0.5일 경우,  $-0.5$ 일 경우, 그리고 R행렬이 0에 가까운 경우와 0에 가까운 비정칙 행렬일 경우, 마지막으로 사전적 정보가 없는 형태인 부적합 분포로 만들어진 0행렬을 가정하였을 경우, 식 (2.1)의 형태로 나타난 기대 효용함수가 어떠한 변화를 나타내는 지 살펴보았다. 적용된 R행렬을 정리하면 다음과 같다.

$$R_1^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad R_2^{-1} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}, \quad R_3^{-1} = \begin{pmatrix} 1 & -0.5 \\ -0.5 & 1 \end{pmatrix},$$

$$R_4 = \begin{pmatrix} 10^{-8} & 0 \\ 0 & 10^{-8} \end{pmatrix}, \quad R_5 = \begin{pmatrix} 10^{-8} & 10^{-8} \\ 10^{-8} & 10^{-8} \end{pmatrix}, \quad R_6 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

베이저안 최적 실험설계를 위한 기대 효용함수의 형태를 살펴보면 Figure 4.1과 같다. Figure 4.1을 살펴보면 사전정보를 나타내는 R행렬의 역할이 매우 중요하다는 사실을 확인할 수 있는데, 사전분포의 중심인  $\theta_0$ 를 0행렬로 가정하고 공분산이 0이라고 가정하는 경우는 ( $R_1$  유형)  $\eta$ 에 대한  $x$ 값의 설계가 0을 중심으로 멀리 떨어져 있을수록 큰 값을 갖는 특징을 가지고 있음을 확인할 수 있다.

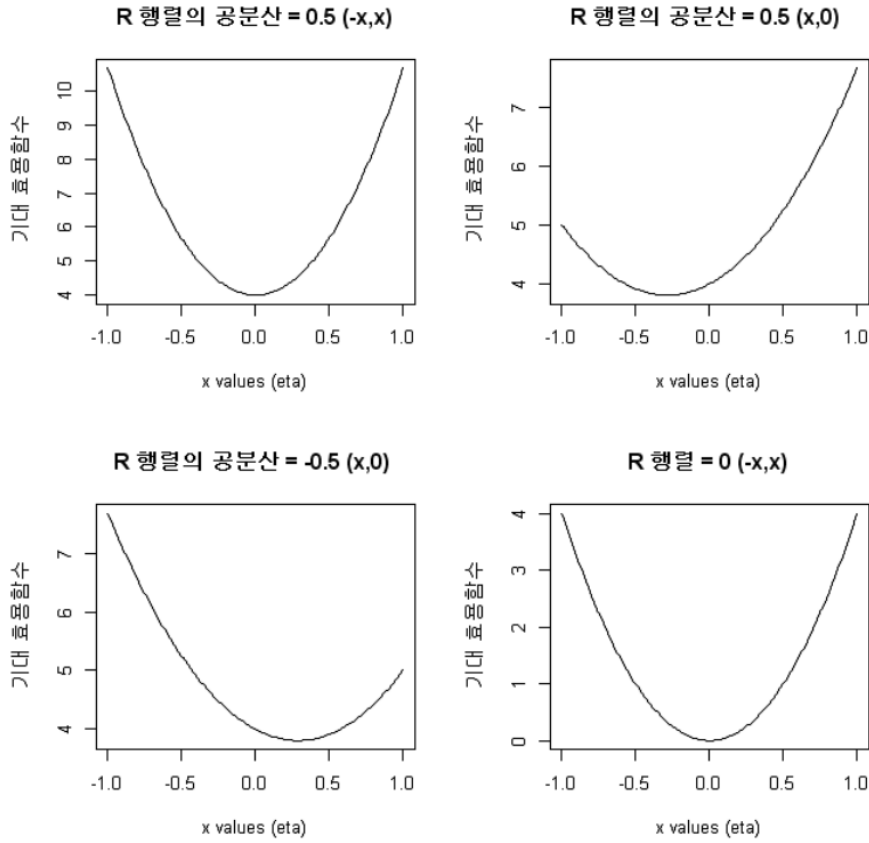


Figure 4.2. Expected Utility with different priors, R matrix ( $k = 2$ )

만일 R의 역행렬, 즉  $\theta$ 에 대한 사전적 공분산행렬이 양의 상관관계나 음의 상관관계를 나타내는 경우에는 ( $R_2$ 와  $R_3$ 유형)  $\eta$ 에 대한 기대 효용함수의 형태가 매우 비대칭적으로 나타나고 있음을 확인할 수 있다. 즉, 두 모수  $\theta = (\alpha, \beta)$ 에 대한 사전적 상관관계가 양으로 설정되는 경우에는  $x$ 의 값이 0으로부터 큰 값으로 설계될수록 기대 효용함수가 크게 나타나며, 반대로  $\theta = (\alpha, \beta)$ 가 사전적으로 음의 관계를 가지고 있을 경우에는  $x$ 의 값이 0으로부터 작은 값으로 설계될수록 기대 효용함수가 크게 나타나고 있음을 확인할 수 있다.

만일 R행렬이 0에 근사한 정칙행렬인 경우는 ( $R_4$ 유형) 대칭의 형태로 안정적으로 나타나고 있으며, R행렬이 0에 근사한 비정칙행렬로 설정하였을 경우에도 ( $R_5$ 유형) 공분산의 부호에 민감한 특징을 확인할 수 있었다. 하지만, R행렬을 0행렬로 설정하여 무정보적 부적합한 사전분포로 설정하게 되면 ( $R_6$ 유형), 기대 효용함수는 우리가 예상하지 못한 이상한 형태의 모습을 가지게 됨을 확인할 수 있게 된다.  $R_6$ 유형은 사실상 우도함수를 기반으로 하는 비 베이저안 방법으로 표본의 크기가 1인 경우에는 적용해서는 안 된다는 사실을 확인할 수 있다.

만일 표본의 크기가 어느 정도 크다면, 우도함수의 영향력이 커져서 사후분포는 사전분포에 영향을 받지 않는 것으로 알려져 있다. 하지만, 표본의 수가 작은 경우에는, 특별히 실험설계를 위하여 표본의 수가

아직 결정되지 않은 상황이라면 사전정보행렬인 R행렬의 역할은 매우 커지게 되며, 특별히 베이지안 접근법에서 잘 알려진 비 베이지안 접근방법과 일치하는 무정보적 부적합 사전분포를 사용하는 것에는 세심한 주의가 필요하다는 사실을 간과해서는 안 될 것으로 판단된다.

만일 표본크기를 증가시키면서 실험설계법을 적용하면, 이러한 동일한 현상이 나타날까? 이러한 사실을 확인하기 위하여 동일한 조건에서 다양한 형태의  $x$ 를 설정하고 이에 대한 기대효용함수의 변화를 관찰하여 보면 Figure 4.2와 같이 나타난다.

Figure 4.2는 두 모수의 공분산에 대한 사전적 관계가 양의 상관관계를 가지고 있는 경우,  $k = 2$  경우에서  $(x_1, x_2)$ 로 설계된  $\eta$ 의 형태에 따라 기대효용함수가 어떻게 나타나고 있는지를 보여주고 있다. Figure 4.2에서 나타난 사실을 요약하면, 최적 실험설계  $\eta$ 는 사전분포의 중심점  $\theta_0 = (0, 0)$ 로부터 멀리 떨어진 설계 점들을 찾고 대칭적 관계를 유지하는 것이 바람직하다는 사실을 확인시켜 주고 있다. Figure 4.2에서 나타난 또 다른 특징적인 사실은 무정보적 부적합 사전확률을 적용하였을 경우의 변화이다. 이 경우는,  $k = 1$ 의 경우와 다르게, 안정적인 형태의 기대효용함수 모습이 나타나고 있음을 확인할 수 있다. 본 연구에서는 간단한 선형모형을 적용하였다는 사실을 감안할 때, 선형모형의 복잡도가 증가되면 심각한 문제가 발생할 수 있다는 사실을 확인할 수 있으며, 표본의 크기가 어느 정도 커지면 효용함수는 급격하게 안정적으로 변한다는 사실을 알 수 있다.

## 5. 결론 및 한계점

본 연구에서는 실험계획법 분야에서 매우 중요하게 활용될 수 있는, 하지만 아직 학문적으로나 실용적인 측면에서 한계점이 많이 지적되고 있는 베이지안 실험계획법에 대하여 논의하고 간단한 모의실험을 통하여 최적화된 베이지안 실험계획법이 어떠한 특징을 가지고 있는지 설명하였다. 어떠한 실험을 설계하는 경우, 모수 또는 예측값에 대한 사전적 정보가 존재하지 않는다면 무정보적 부적합 사전확률을 이용하여 실험을 설계할 수 있으며, 이는 비 베이지안적 접근방법과 일치하게 된다. 이 경우, 모형의 복잡도에 따라 다르지만, 표본의 크기를 적게 설정하고 접근하는 경우 베이지안 방법이나 비베이지안 방법 모두 불안정한 설계가 나타날 수 있으므로 세심한 주의가 필요하다는 사실을 확인할 수 있었다. 만일 모수나 예측값에 대한 사전적 정보가 존재하는 경우에는 베이지안 실험계획법이 유일한 해결 방법이지만, 사전적 정보를 어떠한 방법으로 부여하고, 이러한 사전적 정보가 최종 결과에 어떠한 형태로 영향을 미치는 지에 대한 분석은 아직까지도 많은 연구가 필요한 부분이다. 본 연구에서는 간단한 선형모형을 가정하고 사전분포가 최적화된 실험설계에 미치는 영향을 살펴본 결과 사전분포의 중심모수와 척도모수의 설정은 실험설계에 결정적으로 영향을 미치는 매우 중요한 설정이며, 중심모수를 중심으로 대칭으로 만들어진 실험설계가 최적화된 실험설계에 가깝다는 사실을 확인하였다.

비선형모형의 경우 사전정보를 이용하여 최적화된 실험설계를 찾는 베이지안 실험계획법 문제는 더욱 복잡하게 나타난다. 하지만 Tsutakawa (1972)의 선구적인 연구 이후 Chaloner과 Larntz (1989) 연구와 Sun 등 (1996)의 연구를 통하여 제한적이지만 비선형모형, 특별히 로지스틱 회귀모형에 대한 다양한 형태의 해결방법이 제안되었다. 또한 1990년 이후 소개된 몬테칼로 적분법 및 MCMC(Monte-Carlo Markov Chain) 기법의 등장으로 최적화에 대한 문제는 더욱 쉽게 접근할 수 있게 되었다. 그러나 이러한 노력에도 불구하고 비선형모형의 복잡도가 증가하는 경우에 발생할 수 있는 우도함수의 불안정성, 실험설계에 대한 사전정보를 어떠한 형태로 설정하여야 하는지에 대한 이슈, 근사적인 사후분포를 기반으로 최적화된 실험설계를 찾을 경우 실제 최적화된 실험설계와 어느 정도 유사한 지에 대한 문제는 응용 단계에서 극복해야 하는 문제점으로 지적되고 있다. 이러한 한계점에 대해서는 많은 논의가 있지만 2000년 이후 획기적인 연구 성과가 나타나고 있지는 않은 형편이다.



## References

- Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*, Springer, New York.
- Box, G. E. P. and Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis*, Addison-Wesley, Reading, MA.
- Chaloner, K. and Larntz, K. (1989). Optimal Bayesian design applied to logistic regression experiments, *Journal of Statistical Planning and Inference*, **21**, 191–208.
- Chaloner, K. and Verdinelli, I. (1995). Bayesian experimental design: A review, *Statistical Science*, **10**, 273–304.
- Huan, X. and Marzouk, Y. M. (2013). Simulation-based optimal Bayesian experimental design for nonlinear systems, *Journal of Computational Physics*, **232**, 288–317.
- Lindley, D. V. (1972). *Bayesian Statistics-A Review*, SIAM, Philadelphia.
- Nelder, J. A. and Mead, R. (1965). A Simplex method for function minimization, *Computer Journal*, **7**, 308–313.
- Raiffa, H. and Schlaifer, R. (1961). *Applied Statistical Decision Theory*, Harvard Business School, Boston.
- Shannon, C. E. (1948). A mathematical theory and communication, *Bell System Technology Journal*, **27**, 379–423, 623–656.
- Sun, D., Tsutakawa, R. K. and Lu, W. S. (1996). Bayesian design of experiment for quantal response: What is promised versus what is delivered, *Journal of Statistical Planning and Inference*, **52**, 289–306.
- Tanner, M. A. (1998). *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*, Springer, New York.
- Tsutakawa, R. K. (1972). Design of experiment for bioassay, *Journal of the American Statistical Association*, **67**, 584–590.

# 베이지안 실험계획법의 이해와 응용

이군희<sup>a,1</sup>

<sup>a</sup>서강대학교 경영전문대학원

(2014년 9월 16일 접수, 2014년 10월 25일 수정, 2014년 10월 27일 채택)

---

## 요약

본 연구에서는 베이지안 실험계획법에 대하여 논의하고 간단한 모의실험을 통하여 최적화된 베이지안 실험계획법이 어떠한 특징을 가지고 있는지 설명하였다. 실험을 설계하는 경우 연구자는 관심있는 주제가 모수추정인지 아니면 예측인지를 결정하고 사전확률과 우도함수를 기반으로 이에 맞는 사후확률을 찾아 효용함수와 결합하여 최적의 실험설계를 찾는 것이 베이지안 실험계획법의 기본 원리이다. 만일 사전적 정보가 존재하지 않는다면 무정보적 부적합 사전확률을 이용하여 실험을 설계할 수 있으며, 이는 비 베이지안적 접근방법과 일치하게 된다. 만일 모수나 예측값에 대한 사전적 정보가 존재하는 경우에는 베이지안 실험계획법이 유일한 해결 방법이다. 하지만 모형의 복잡도가 증가하게 되면, 최적해를 찾는 과정이 매우 복잡해져서 극복해야 하는 많은 문제점들이 존재하므로 향후 많은 연구가 필요한 분야이다.

주요용어: 실험계획법, 베이지안 분석, 베이지안 의사결정론, 몬테칼로 방법.

---

<sup>1</sup>(121-741) 서울특별시 마포구 신수동 1번지, 서강대학교 경영전문대학원. E-mail: ghlee@sogang.ac.kr