

Minimum Bias Design for Polynomial Regression

Dae-Heung Jang^a · Youngil Kim^{b,1}

^aDepartment of Statistics, Pukyong National University

^bSchool of Business and Economics, ChungAng University

(Received October 28, 2015; Revised November 27, 2015; Accepted December 12, 2015)

Abstract

Traditional criteria for optimum experimental designs depend on the specifications of the model; however, there will be a dilemma when we do not have perfect knowledge about the model. Box and Draper (1959) suggested one direction to minimize bias that may occur in this situation. We will demonstrate some examples with exact solutions that provide a no-bias design for polynomial regression. The most interesting finding is that a design that requires less bias should allocate design points away from the border of the design space.

Keywords: bias, minimum bias design, Q -optimal design, integrated mean squared error(IMSE)

1. 서론

가정된 모형에 대한 불확실성하에서 실험계획을 다룬 대표적인 초기 논문으로 Box와 Draper (1959)을 들 수 있다. 후속되는 논문들의 일부는 이 효시적인 논문의 정신을 이어 받아 다른 분야로의 접목을 시도하였고 일부는 모형의 불확실성을 감안한 복합실험(composite design)이라든지 참의 모형을 구축하기 위한 모형차별실험(model discrimination design)으로 연구가 진행되었다. 본 연구에서는 먼저 Box와 Draper (1959)의 논문이 왜 아직까지 각광을 받고 있는지에 대한 이해를 다항회귀모형(polynomial regression model)을 중심으로 소개하고 그들이 제안한 최소편의(minimum bias) 실험의 구조적인 성격을 다양한 예제를 통해 알아본다. 또한 주어진 모형을 가정한 전통적인 실험기준과 어떤 다른 특징이 있는지 알아 볼 것이다. 제 2절에서는 이를 위하여 최적실험에 관한 기본적인 내용을 표현하고 왜 최소편의 실험계획이 중요한지에 관한 동기유발 예시를 한 다음 제 3절에서는 다항회귀모형을 중심으로 최소편의 실험의 속성에 대해 알아본다. 제 4절에서는 결론과 향후 과제를 논하여 본다.

2. Q -최적실험과 최소편의 실험계획

앞으로 전개될 논의를 위해 식 (2.1)과 같은 선형모형을 가정한다.

$$y_{ij} = \theta^T f(x_i) + \epsilon_{ij}, \quad i = 1, 2, \dots, k; \quad j = 1, 2, \dots, n_i, \quad \sum_1^k n_i = N, \quad (2.1)$$

¹Corresponding author: School of Business and Economics, ChungAng University, 84 Heukseok-ro, Dongjak-gu, Seoul 06974, Korea. E-mail: yik01@cau.ac.kr

여기서 $\theta \in R^m$ 는 미지의 모수벡터 $\theta^T = (\theta_1, \theta_2, \dots, \theta_m)$ 이고, $f(x)$ 는 주어진 반응변수의 형태에 의존하는 $m \times 1$ 벡터이며 $f^T(x) = (f_1(x), f_2(x), \dots, f_m(x))$ 로 표기한다. 받힘점(support point) x 는 어느 집합 Ω 에서 선택될 수 있는 통제 가능(controllable)한 변수이고 ϵ_{ij} 는 오차벡터로서 기댓값이 0이고, 분산은 σ^2 로 가정한 서로 비상관인 확률오차이다. 제시된 Ω 는 실험점이 위치하는 실험영역으로 통상 $\Omega \in [-1, 1]$ 으로 간주한다.

실험설계는 실험영역 Ω 내의 유한개의 k 개의 점 $x_i \in \Omega, i = 1, 2, \dots, k$ 에 대한 확률 질량함수 ξ 로 기술된다. 식 (2.1)에 의하면 이러한 ξ 은 $\xi = \{n_1/N, n_2/N, \dots, n_k/N\}$ 로 표기된다. 앞으로 각각의 x_i 에서의 질량은 $\xi(x_i)$ 로 표기한다.

$N \times \xi(x_i)$ 가 정수라는 제약조건이 주어지면 정확실험설계(exact design), 제약조건이 없는 경우를 근사실험설계(approximate design) 혹은 연속실험설계(continuous design)이라 한다. 본 연구에서는 후자를 선택한다. 정확실험설계의 중요성은 결론 부분에서 토의한다.

식 (2.1)의 확률오차 ϵ 에 대하여 정규분포를 가정한다면 실험설계 ξ 의 정보행렬(information matrix)은 식 (2.2)와 같이 주어진다. 이 정보행렬로 실험설계의 정보의 양을 측정하는데 통상적으로 비특이(non-singular) 행렬을 가정한다.

$$M(\xi) = \frac{1}{N} \sum_{i=1}^k n_i f(x_i) f^T(x_i). \quad (2.2)$$

이러한 정보행렬은 $1/N$ 을 제외하면 통상적인 $X'X$ 이다. 따라서 식 (2.2)는 정규화(normalized)된 정보행렬로 불린다. 또한 이를 연속적인 표기방법으로 구현하면 식 (2.3)으로 표기가 가능하다.

$$M(\xi) = \int_{\Omega} f(x) f^T(x) d\xi(x). \quad (2.3)$$

최적실험설계분야에서는 대표적인 실험설계기준으로 $M(\xi)$ 의 행렬식을 최대화시키는 D -최적 실험설계기준이 있고, 다른 하나는 $x_0 \in \Omega$ 인 경우 예측분산(prediction variance), $\sigma^2 f^T(x_0) M^{-1}(\xi) f(x_0)$ 의 최대값을 최소화하는 G -최적 실험설계기준이 존재한다. 이러한 D -와 G -최적 사이에는 동격성(equivalence)이 존재한다. 최적성은 예측분산의 최대값이 모형의 모수의 개수와 일치하는지 여부를 통해 확인할 수 있다. 그리고 본 연구에서 주로 쓰일 선형최적의 한 형태인 Q -최적실험이 있다. G -최적은 예측분산의 최대값의 최소화하는 실험기준이나 Q -최적은 식 (2.4)에서와 같이 실험영역에서 예측분산의 (가중) 평균을 최소화하는 실험기준이다.

$$M(\xi) = \int_{\Omega} \sigma^2 f^T(x) M^{-1}(\xi) f(x) d\lambda(x), \quad (2.4)$$

여기서 λ 는 가중치를 반영하기 위한 가중함수로 이해한다. 그리고 보통 식 (2.4)을 계산할시 λ 를 실험영역에서 균일분포로 가정하는데 반드시 그럴 이유는 없다. 실험영역, $\Omega \in [-1, 1]$ 을 벗어나 확장될 수 있으며 균일분포가 아닌, 가중치의 유연성을 반영한 베타(beta) 분포를 적용하는 경우가 있다. 예외적인 경우는 Kim (1995)과 같은 논문을 참고하기 바란다. 2차 회귀모형을 예를 들어 G -최적과 Q -최적의 차이점을 간단하게 설명한다. 단순회귀 모형인 경우는 두 최적은 동일한 실험계획을 제공한다.

예제 2.1: 2차 회귀모형 $f^T(x) = (1, x, x^2)$ 인 경우 $D(G)$ -최적은 $\xi(\pm 1) = \xi(0) = 1/3$ 이고 Q -최적인 경우는 $\xi(\pm 1) = 0.25, \xi(0) = 0.5$ 이다. 분산함수 $f^T(x) M^{-1}(\xi) f(x)$ 를 실험영역에 그려보면 Figure 2.1과 같은데 실선이 G -최적 그리고 점선이 Q -최적의 분산함수이다. Figure 2.1에서 보듯이 $D(G)$ -최적의 최대값은 2차 회귀모형의 모수의 개수인 3과 정확히 일치한다. 반면 Q -최적은 실험영역의 양극

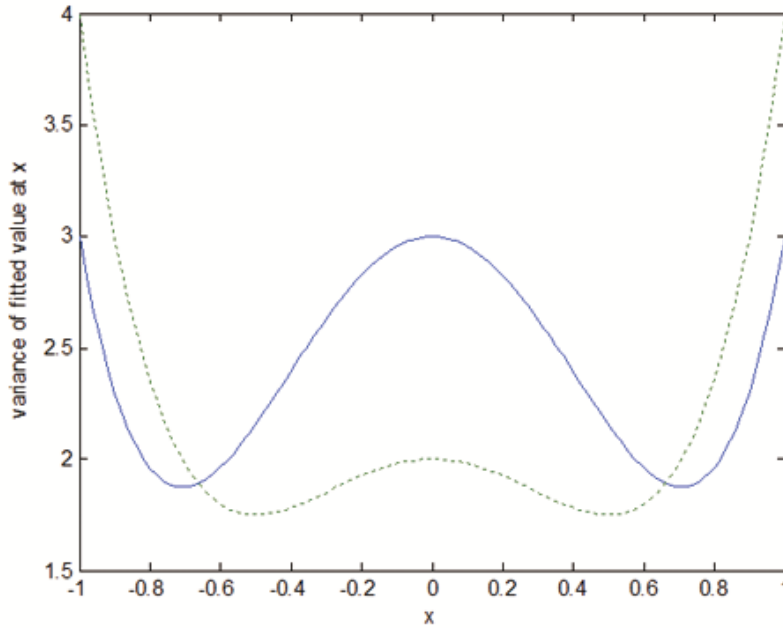


Figure 2.1. Variance function of fitted value at Q -optimality and D -optimality (dotted line: Q -optimality, real line: D -optimality).

점에서는 분산의 값이 4로 G -최적의 분산의 최대값인 3보다 높지만 실험영역 대부분에서는 $D(G)$ -최적 보다는 분산의 값이 낮게 나온다는 사실을 확인할 수 있다.

그러나 이와는 달리 단순 선형모형인 경우는 G -최적이거나 Q -최적이거나 다 같이 $\xi(\pm 1) = 0.5$ 로 나온다. 기본적으로 모형이 단순회귀 모형이라면 양극점에 등질의 질량을 부여하는 것이 최상의 실험이다. 그러나 Figure 2.2에서 짐작하듯이 참의 모형이 2차 회귀모형이라면 주어진 모형의 분산에 기초한 실험 $\xi(\pm 1) = 0.5$ 보다는 대안실험 $\xi(\pm 3/4) = 0.5$ 에서 $E[y(x)], x \in \Omega$ 의 추정치의 기대제곱오차 값(expected squared error; MSE)이 대부분의 실험영역에서 더 작다. 따라서 참의 모형이 2차 회귀모형이라면 $D(G)$ -최적과 같은 실험기준보다는 대안실험이 MSE 관점에서 보면 더 나올 수 있다.

따라서 가정된 모형과 참의 모형사이의 간극을 좁히기 위한 방법이 Box와 Draper (1959)의 논문의 배경이다. 이를 위하여 식 (2.1)을 다음과 같이 분해하여 본다.

$$y_{ij} = \theta_1^T f_1(x_i) + \theta_2^T f_2(x_i) + \epsilon_{ij}, \tag{2.5}$$

여기서 모수벡터 θ_1 과 θ_2 의 크기는 각각 m_1 과 m_2 이며 f_1 과 f_2 역시 해당되는 $f(x)$ 의 부분집합이다. 그러나 $\hat{y}_{ij} = \hat{\theta}_1^T f_1(x)$ 을 적합 시켰다고 가정하자. 그러면 어떤 점 x 에서의 \hat{y} 의 제곱오차(squared error), $\text{err}(x)^2$ 는 다음과 같고

$$\text{err}(x)^2 = \left(\theta_1^T f_1(x_i) + \theta_2^T f_2(x_i) - \hat{\theta}_1^T f_1(x) \right)^2$$

이고 이에 대한 기댓값인 평균제곱오차(MSE)는 $E[\text{err}(x)^2] = (E[\text{err}(x)])^2 + \text{Var}[\text{err}(x)]$ 로 나타난다. 이를 Box와 Draper (1959)는 λ 에 대해 적분을 취한 IMSE(integrated mean squared error)는

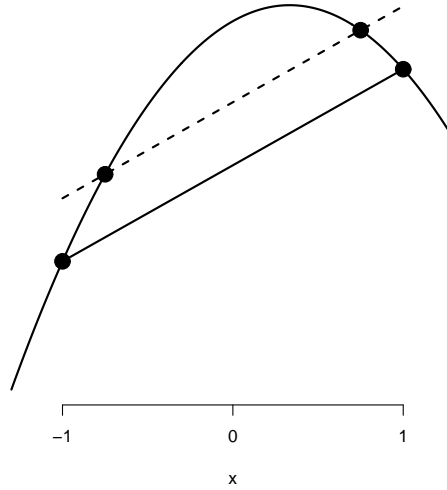


Figure 2.2. Alternative design of $D(G-)$ optimal design (dotted line).

식 (2.6)과 같다.

$$\text{IMSE} = \int_{\Omega} \text{MSE} d\lambda(x) = V + B, \quad (2.6)$$

여기서 V 는 integrated variance의 약자이고 B 는 integrated bias의 약자이다. 그리고 V 와 B 는 다음과 같이 각각 표기된다.

$$V = \int_{\Omega} \sigma^2 \text{Var}[\text{err}(x)] d\lambda(x) = \frac{\sigma^2}{N} \text{tr} [M_{11}(\xi)^{-1} \mu_{11}],$$

$$B = \int_{\Omega} E[\text{err}(x)]^2 d\lambda(x) = \theta_2^T (\mu_{22} - M_{21} M_{11}^{-1} \mu_{12} - \mu_{21} M_{11}^{-1} M_{12} + M_{21} M_{11}^{-1} \mu_{11} M_{11}^{-1} M_{12}) \theta_2,$$

여기서

$$M_{11}(\xi) = \frac{1}{N} \sum_{i=1}^k n_i f_1(x_i) f_1^T(x_i), \quad M_{12}(\xi) = \frac{1}{N} \sum_{i=1}^k n_i f_1(x_i) f_2^T(x_i)$$

이고

$$\mu_{11} = \int_{\Omega} f_1(x) f_1^T(x) d\lambda(x), \quad \mu_{12} = \int_{\Omega} f_1(x) f_2^T(x) d\lambda(x)$$

이다.

V 는 $f_1(x)$ 에 대하여 Q -최적의 대상이고 더 나아가 B 는 $B = B_1 + B_2$ 로 분해된다. 여기서 B_1 과 B_2 는 아래와 같다.

$$B_1 = \theta_2^T (\mu_{22} - \mu_{21} \mu_{11}^{-1} \mu_{12}) \theta_2,$$

$$B_2 = \theta_2^T ((M_{21} M_{11}^{-1} - \mu_{21} \mu_{11}^{-1}) \mu_{11} (M_{11}^{-1} M_{12} - \mu_{11}^{-1} \mu_{12})) \theta_2. \quad (2.7)$$

그러나 식 (2.7)에서 B_1 은 실험과는 무관한 항목이기 때문에 실험하는 단계에서는 무시한다. 그리고 θ_2 에 상관없이 $B_2 = 0$ 이 되는 필요충분조건은 식 (2.8)과 같다.

$$M_{11}^{-1}M_{12} = \mu_{11}^{-1}\mu_{12}, \quad (2.8)$$

여기서 $M_{11}^{-1}M_{12}$ 는 별명(alias) 행렬임을 주지하기 바란다. 이러한 $B_2 = 0$ 을 만족하는 조건의 실험의 존재 여부 및 그 형태에 Box와 Draper (1959)에 의하여 언급이 되었고 Karson (1970), Hunga와 Fan (2004), Draper와 Guttman (1992), Montepiedra와 Fedorov (1997), Abdelbasit와 Butler (2006) 여러 연구자에 의하여 연구가 되었다. 그러나 많은 논문들은 이론적인 틀만 제시하고 그 해가 구체적이지 못하였다. 그 이유로는 아무리 모형이 단순하다고 하더라도 그 해가 복잡한 수식으로 전개되기 때문이라고 본다. 또한 대다수의 논문은 B 에만 국한하지 않고 V 와 B 를 동시에 고려하였는데 이는 σ^2 과 θ_2 를 모르고 실험을 하는 문제가 되므로 단순한 직관적인 해를 구하기 힘들다. 따라서 본 연구에서는 $B_2 = 0$ 를 만족하는 연속적인 실험에 국한한다. 이러한 실험을 본 연구에서는 최소편의실험(minimum bias design)이라 부른다.

그리고 본 연구에서는 문제의 단순성과 그러한 실험의 속성을 구체화하기 위하여 다항회귀모형만을 중심으로 수치적인 해를 제공하고자 한다. 과연 위의 동기유발예제에서 보였던 대안실험의 성질이 최소편의실험에서도 존재 하는지 알아보하고자 한다. 그리고 본 연구에서는 Figure 2.2에서 언급한 이슈를 더 잘 설명할 수 있도록 연속실험의 관점에서 다루기로 한다. 수치적인 해로서는 필요충분조건의 만족 여부에 따라 이산형의 실험인 경우는 Myer와 Nachtsheim (1995)의 교환알고리즘(exchange algorithm)을 이용할 수 있으나 본 연구에서는 GRG2(generalized reduced gradient 2)란 비선형계획법 알고리즘을 이용하여 그 필요충분조건을 만족하는 연속적인 실험계획의 해를 찾았다. GRG2 알고리즘은 Leon Lasdon 교수가 개발하여 Excel의 해찾기(solver)기능에 내장되어 있다. 이전에는 모형이 복잡한 경우 구체적인 해를 얻지 못하였으나 이러한 알고리즘의 도움으로 수치 해석적이거나 해의 구조를 좀 더 정확하게 알아 볼 수 있게 되었다. 제 3절에서는 다항회귀모형을 중심으로 $B_2 = 0$ 의 조건을 가지는 실험의 형태를 알아본다.

3. 예제

예제 3.1: 주어진 $f^T(x)$ 를 $f_1(x) = (1, x)$, $f_2(x) = (x^2)$ 로 분해하여 보자. 선형모형을 가정하고 실험을 하려고 하나 참의 모형이 $f_2(x) = (x^2)$ 을 포함한 2차 회귀모형인 경우 나타날 편의를 최소화하는 실험을 하여 본다. $f^T(x) = (1, x, x^2)$ 인 경우 잘 알려져 있듯이 $D(G)$ -최적은

$$\xi(\pm 1) = \xi(0) = \frac{1}{3}$$

이다. 각 반힘점에 부여된 질량은 $1/3$ 로 고정하고 $B_2 = 0$ 의 필요충분조건 $M_{11}^{-1}M_{12} = \mu_{11}^{-1}\mu_{12}$ 을 만족하는 $0, \pm x$ 의 구조가 어떻게 나타나는지 살펴 보았다.

$$\xi(\pm 0.707) = \xi(0) = \frac{1}{3}$$

참고로 위의 실험 하에서는 $M_{11} = \begin{pmatrix} 1 & 0 \\ 0 & 2/3 \end{pmatrix}$, $M_{12} = \begin{pmatrix} 2/3 \\ 0 \end{pmatrix}$, $\mu_{11} = \begin{pmatrix} 1 & 0 \\ 0 & 1/3 \end{pmatrix}$, $\mu_{12} = \begin{pmatrix} 1/3 \\ 0 \end{pmatrix}$ 으로 나타나 필요충분조건을 만족시킴을 알 수 있다.

만약 반힘점의 구조를 $0, \pm 1$ 로 그대로 두고 질량을 재배치한다면 $\xi(\pm 1) = 1/6$, $\xi(0) = 2/3$ 으로 나온다. 이러한 실험은 식 (3.1)에서 언급한 실험과 비교하여 본다면 양극점이 안쪽으로 축소가 되지 않고 가운

데 반함점인 0에 배정되는 질량을 $1/3$ 에서 $2/3$ 로 2배가 증가함을 알 수 있다. 참고로 Q -최적 실험은 $\xi(0) = 1/2$ 을 배정한다는 것을 감안한다면 오히려 $B_2 = 0$ 을 위한 실험은 Q -최적보다도 많은 질량을 가운데 반함점에 배정한다는 사실을 알 수 있다. 이 두 실험은 모두 다 $B_2 = 0$ 을 만족시키는 실험이다.

$N = 3$ 이라면 당연히 첫 번째 실험이 이상적이다. N 이 3의 배수가 되지 않는 한 각 반함점에 배정되는 실험의 질량에 따라 $\pm x$ 의 값은 달라 질 수밖에 없다. 예를 들어 $\xi(\pm x) = 1/5$, $\xi(0) = 3/5$ 로 질량이 정해져 있다면 더 이상 $B_2 = 0$ 으로 하는 실험은 $\pm x = \pm 1$ 이 아니다. 그러나 가운데 질량이 $3/5$ 만큼 부여가 되어 있으므로 ± 1 에서 이탈되는 양이 많지 않을 것이다. 이는 수치적으로 증명이 가능하다. $\pm x = \pm 0.80$ 으로 밝혀지는데 $\pm x = \pm 0.707$ 보다 ± 1 에서 이탈되는 양이 크지 않음을 알 수 있다. 결론적으로 편의를 최소화 하는 실험은 예제 3.1인 경우는 기본적으로 양극단에 반함점을 배치하기 보다는 양극단에서 이탈된 지점에서 반함점을 재배치하거나 혹은 가운데 반함점에 좀 더 많은 질량을 배정하는 것이 타당하다고 보여진다.

예제 3.2: 이번 예제는 예제 3.1을 확장하여 본다. 3차 회귀모형을 참의 모형으로 가정하고 같은 증상이 나타나는지 수치적으로 확인하여 본다. $f^T(x) = (1, x, x^2, x^3)$, $f_1(x) = (1, x, x^2)$, $f_2(x) = (x^3)$ 으로 설정되는 경우이다. 기존 문헌에서는 필요충분조건을 맞추기에는 모멘트의 함수가 복잡하므로 그 조건에 대한 구체적인 해가 나와 있지 않다. 참고로 $D(G)$ -최적실험은 $\xi(\pm 1) = \xi(\pm 0.4472) = 1/4$ 로 나타난다. $\xi(\pm x_1), \xi(\pm x_2)$ 에 부여되는 질량을 등질량인 $1/4$ 로 하고 구하여 보면 $\pm x_1 = \pm 0.8425$, $\pm x_2 = \pm 0.43611$ 로 나타난다. 역시 예제 3.1과 마찬가지로 양극단에서 바깥쪽 반함점들의 위치가 ± 1 에서 안으로 줄어드는 현상이 나타난다. 그러나 안쪽 반함점의 위치는 거의 변화가 없다. 그리고 반함점의 구조를 $\pm 1, \pm x$ 로 하고 질량을 변화하였을 때는

$$\xi(\pm 1) = 0.08994, \quad \xi(\pm 0.50249) = 0.4101$$

로 나타난다. 역시 안쪽 반함점에 질량이 많이 배정되는 결과를 보여준다. 역시 $B_2 = 0$ 의 실험을 잘 특징 지어준다고 본다.

예제 3.3: $f^T(x) = (1, x, x^2, x^3, x^4)$, $f_1(x) = (1, x)$, $f_2(x) = (x^2, x^3, x^4)$ 인 경우를 살펴보자. 4차 회귀모형에서 $D(G)$ -최적인 경우 반함점이 5개로 나오는데 본 연구에서는 $0, \xi(\pm x_1), \xi(\pm x_2)$ 에 주어지는 질량을 0.2 로 두고 반함점의 위치가 어떻게 조정되는지 보았다.

$$\pm x_1 = 0.8325, \quad \pm x_2 = 0.3745.$$

바깥 반함점과 안쪽 반함점의 위치가 D -최적에 비해 양극단에서 안쪽으로 이탈됨을 알 수 있어 예제 3.1과 예제 3.2와 마찬가지로의 현상이 나타난다. 다만 안쪽 반함점의 위치가 조정됨이 달라지는 점이라 할 수 있다. 예제 3.2 보다는 안쪽으로의 이탈이 조금 더 심하게 일어나는데 이는 모형의 설정상 당연한 결과로 보인다. 또한 $\pm 1, \pm x, 0$ 을 반함점으로 하고 x 및 질량을 구하면 $\xi(\pm 1) = 0.0667$, $\xi(\pm 0.5774) = 0.2999$, $\xi(0) = 0.2667$ 로 나타나 역시 안쪽 반함점에 많은 질량을 배정함을 알 수 있다.

몇 가지 예제를 보게 되면 차수가 높은 모형을 염두에 둔 $B_2 = 0$ 실험은 우리가 동기부여 예제에서 보았듯이 V 만 염두에 둔 실험보다는 향후 발생할 편의를 염두에 둔 실험으로 모형의 불확실성에 대한 상황을 유연하게 대처 할 수 있다고 본다. 즉 양 극단에서 반함점을 구하지 않고 반함점의 위치를 실험영역 안쪽으로 정하면 된다.

4. 결론 및 토의

본 연구에서는 Box와 Draper (1959)가 제안한 IMSE 중에서 B_2 만을 최소화하는 실험을 다항회귀모형을 중심으로 알아보았다. 다른 연구와 달리 IMSE의 V, B 의 관계를 감안한 실험을 하지 않고 B_2 만을 감안하여 실험한다는 것은 B 가 V 에 비해 압도적으로 크다고 가정할 경우가 된다. 따라서 본 연구의 목적을 B_2 만의 실험이 V 만의 실험과 비교하여 어떤 특징을 가지는지를 알아보려고 설정하였다. 편의가 발생하는 상황이라면 실험영역의 극단에 반힘점을 배정하는 것과 달리 영역가운데 쪽으로 재배치되는 것이 더 바람직하다는 사실을 다항회귀모형에서는 수치적으로 확인하였다. 향후과제로서 Jones와 Nachtsheim (2011a, 2011b)이 시도하였던 별명(alias)의 놈(norm)을 최소화 하는 과정에서 밝혀진 확정적인 선별실험설계(definitive screening experiment)와 본 연구에서 제안한 $B_2 = 0$ 을 추구하는 실험간에 어떤 연관성이 있는 지 혹은 모형을 다차원 다항회귀모형으로 확장을 하였을 때 어떤 규칙성이 존재하는지를 탐색하는 것이다. 본 연구의 결과로 추정컨대 분명히 유사한 관계가 존재한다고 보여진다. 또한 V 와 B 의 복합실험에도 다양한 예제를 수치적으로 밝힐 필요가 있다. 또한 이의 내용을 정확실험설계로 확장하는 것도 실제 실험자에게는 매우 중요한 연구주제가 될 것이다.

References

- Abdelbasit, K. M. and Butler, N. A. (2006). Minimum bias design for generalized linear models, *The Indian Journal of Statistics*, **68**, 587–599.
- Box, G. E. P. and Draper, N. R. (1959). A bias for the selection of a response surface design, *Journal of the Statistical Association*, **54**, 622–654.
- Draper, N. R. and Guttman, I. (1992). Treating bias as variance for experimental design purposes, *Annals of the Institute of Statistical Mathematics*, **44**, 659–671.
- Hunga, K. and Fan, S. (2004). A note on minimum bias estimation in response surfaces, *Statistics and Probability Letters*, **70**, 71–85.
- Jones, B. and Nachtsheim, C. J. (2011a). A class of three-level designs for definitive screening in the presence of second-order effects, *Journal of Quality Technology*, **43**, 1–15.
- Jones, B. and Nachtsheim, C. J. (2011b). Efficient designs with minimal aliasing, *Technometrics*, **53**, 62–71.
- Karson, M. J. (1970). Design criterion for minimum bias estimation of response surfaces, *Journal of the Statistical Association*, **65**, 1565–1572.
- Kim, Y. (1995). A Study for additional characteristics of I_λ -optimal experimental designs, *Communications for Statistical Applications and Methods*, **2**, 55–63.
- Montepiedra, G. and Fedorov, V. V. (1997). Minimum bias design with constraints, *Journal of Statistical Planning and Inference*, **63**, 97–111.
- Myer, R. K. and Nachtsheim, C. J. (1995). The coordinate exchange algorithm for constructing exact experimental designs, *Technometrics*, **37**, 60–69.

다항회귀모형에 대한 최소편의 실험계획

장대흥^a · 김영일^{b,1}

^a부경대학교 통계학과, ^b중앙대학교 경영학부

(2015년 10월 28일 접수, 2015년 11월 27일 수정, 2015년 12월 12일 채택)

요약

전통적으로 최적실험을 위한 실험기준들은 기본적으로 가정된 모형에 의존한다. 따라서 모형에 대한 완벽한 정보를 가지지 않는 경우 실험자는 곤란에 빠질 수 밖에 없다. Box와 Draper (1959) 이런 상황에 대비해 적분된 평균제곱오차의 편이부분에 해당하는 적분된 편의를 최소화하는 실험기준을 제안하고 필요충분조건을 명시하였다. 그러나 간단한 예제를 제외하고는 문헌에서는 이러한 필요충분조건을 만족하는 실험에 대한 구체적인 예제는 계산상의 문제로 예외로 많이 연구가 되어 있지 않다. 비록 수치적인 해이긴 하지만 다항회귀모형을 중심으로 최소편의를 만족하는 실험의 성격을 파악하였는데 결론적으로 양극단에서 안쪽 방향으로 이탈되는 위치에서 반힘점이 형성되는 것을 알 수 있었다.

주요용어: 편의, 최소편의 실험, Q -최적 실험, 적분평균제곱오차(IMSE)

¹교신저자: (06974) 서울특별시 동작구 흑석동 84, 중앙대학교 경영경제대학 경영학부.
E-mail: yik01@cau.ac.kr