

Some Criteria for Optimal Experimental Design at Multiple Extrapolation Points

YoungIl Kim^{a,1} · Dae-Heung Jang^b

^aSchool of Business and Economics, Chung-Ang University

^bDepartment of Statistics, Pukyong National University

(Received June 23, 2014; Revised August 2, 2014; Accepted August 15, 2014)

Abstract

When setting up an experiment for extrapolation at multiple points outside the design space, we often face a difficulty in which point we should emphasize even if the polynomial model under consideration is given. In this paper we propose various methods under two possible scenarios that deal with extrapolations. One considered in this paper is the situation when the model assumed can be extended beyond the design space. In this setting, the many classical methods(including various approaches the authors proposed before) were revisited in the context of extrapolation. But the real problem arises when there is an uncertainty concerning the validity of the assumed model. Therefore, the second scenario is to develop an appropriate procedure when we have limited information about model. Consequently, a hybrid approach is suggested to deal with this issue of how to handle the multiple extrapolating under model uncertainty. A search algorithm was implemented because the classical exchange algorithm was found difficult to handle the complexity of the problem.

Keywords: Extrapolation design, model uncertainty, constrained design, compound design, maximin design, hybrid approach, genetic algorithm.

1. 서론

앞으로 전개될 논의를 위해 식 (1.1)과 같은 m 차 다항회귀모형을 가정한다.

$$y(x) = \theta_0 + \sum_{i=1}^m \theta_i x^i + \epsilon = f_m^T(x)\theta + \epsilon, \quad (x \in \Omega), \quad (1.1)$$

여기서 $y(x)$ 는 반응변수, $f_m^T(x) = (1, x, \dots, x^m)$ 이며 $\theta^T = (\theta_0, \theta_1, \dots, \theta_m)$ 는 모수벡터이며 ϵ 은 오차 벡터로서 기댓값 0 그리고, 분산은 일반성의 손실 없이 $\sigma^2 = 1$ 로 가정한다. 명시된 Ω 는 실험점이 위치하는 실험영역으로 통상 $\Omega \in [-1, 1]$ 으로 간주한다. 실험설계는 실험영역 Ω 내의 유한개의 s 개의 점 $x_i \in \Omega$, $i = 1, 2, \dots, s$ 에 대한 확률 질량함수 ξ 로 기술되는데 전체 실험에 쓰이는 관측 값의 개수 n 을 곱한 $n\xi(x_i)$ 이 정수라는 제약조건이 주어지면 정확실험설계(exact design), 제약조건이 없는 경우를

This research was supported by the Chung-Ang University's research grant in 2014.

¹Corresponding author: School of Business and Economics, Chung-Ang University, 84 Heukseok-Ro, Dongjak-Gu, Seoul 156-756, Korea. E-mail: yik01@cau.ac.kr

근사실험설계(approximate design)라 한다. 본 연구에서는 후자를 택하였는데 이는 이론적인 최적실험 분야의 연구에서는 이 방법을 통해 실험설계의 근본적인 성질을 연구하기 때문이다.

1960년대 초창기 최적실험의 논문들은 실험설계에 쓰이는 실험기준들의 수학적 성질에 많은 할애를 한 반면 Box와 Draper (1975)의 논문 이후로는 모형에 관한 불확실성이 내재되어 있거나 모형에 대한 가정이 부적합한 경우 실험자가 융통성 있게 사용할 수 있는 실험들이 연구되었다. 최근에 Jones와 Nachtsheim (2011)은 기존의 전통적인 실험설계의 틀을 부정할 만한 논문을 발표하였는데 모형의 부적합 시 발생하는 별명(alias)의 놈(norm)을 없애면서도 3-수준 실험을 완성한 것이다. 이와 같은 결과물은 전형적인 최적실험설계의 부산물이라 할 수 있다.

식 (1.1)의 모형 $f_m(x)$ 하에서 특정한 점 z 에서의 분산은 식 (1.2)와 같다.

$$v_m(\xi, z) = f_m^T(z)M_m^{-1}(\xi)f_m(z), \quad (1.2)$$

여기서 $M_m(\xi)$ 는 $\int_{\Omega} f_m(x)f_m^T(x)\xi d(x)$ 로 정보행렬(information matrix)이다. 이 정보행렬로 실험설계, ξ 의 정보의 양을 측정하는데 비특이(non-singular) 행렬로 가정한다. 최적실험설계분야에서는 이러한 $M_m(\xi)$ 의 행렬식을 최대화시키는 실험설계기준으로 D -최적 실험설계기준이 있고 $z \in \Omega$ 인 경우 해당하는 예측분산(prediction variance) $v_m(\xi)$ 의 최대값을 최소화하는 G -최적 실험설계기준이 존재한다. 이러한 D -와 G -최적 사이에는 동격성(equivalence)이 존재한다. 그리고 최적성은 $v_m(\xi)$ 의 최대값이 모형의 모수의 개수와 일치하는지 여부를 통해 확인할 수 있다. 물론 식 (1.1)에서의 다항모형의 경우 모수의 개수는 $m + 1$ 이다. 그리고 또 다른 실험설계기준으로서 $\text{tr}(M^{-1}(\xi))$ 를 최소화하는 A -최적 실험설계기준이 있다.

본 연구에서 다루는 실험설계기준은 외삽-최적이다. $z \notin \Omega$ 인 경우

$$v_m(\xi, z) = f_m^T(z)M_m^{-1}(\xi)f_m(z), \quad z \notin \Omega \quad (1.3)$$

를 최소화하는 실험설계가 외삽-최적 ξ_z^* 이다. 즉, 외삽-최적은 다항회귀모형 가정 하에서 식 (1.3)의 값이 최소가 되는 실험설계를 찾는 문제가 된다. Hoel과 Levine (1964)에 의해 다항회귀모형 $f_m(x)$ 의 차수 m 을 알고 있다는 가정에서 실험영역을 벗어나는 점, $z \notin \Omega$ 에 대한 외삽-최적실험설계 ξ_z^* 의 해가 제공된 바 있다. 그러나 이러한 해는 $f_m(x)$ 이 실험영역을 벗어난다 하여도 여전히 모형이 타당한 경우여야 한다. 그러나 $z \notin \Omega$ 가 실험영역을 벗어날수록 이러한 가정은 현실적으로 불안정하다. 이러한 점에 유의하여 Dette와 Wong (1996)이 Läuter (1974)의 실험설계기준을 이용하여 하나의 점 $z \notin \Omega$ 를 대상으로 모형의 불확실성이 존재할 경우 최적실험설계를 제안하였다. 그러나 이들의 논문들은 $z \notin \Omega$ 에서의 분산의 선형결합만을 고려 할 뿐 모형의 불확실성을 고려한 실험설계기준은 고려하지 않아 Kim과 Jang (2012)은 실험영역 바깥에 있는 점 $z \notin \Omega = [-1, 1]$ 에서의 분산을 최소화하는 외삽의 문제점을 모형이 불확실한 상황을 고려해서 다루었다.

본 연구에서는 $z \notin \Omega$ 를 하나에서 여러 개로 확장하려 한다. 즉 실험영역을 벗어나는 여러 개의 점 $z_i \notin \Omega, i = 1, 2, \dots, k$ 로 확장될 때 생기는 문제를 다루려고 한다. 이는 Kim과 Jang (2012)의 토의 부분에서 언급된 바 있다. 본 연구에서 다루고자 하는 모형들은 일변 다항회귀 모형이다. 물론 다변 다항회귀모형으로 전개가 이루어 질 수도 있으나 연구에서 제시하는 방법들은 하이브리드형 접근방법들이므로 이들 방법의 타당성 검증을 위해서 되도록 차수가 낮은 일변 다항회귀모형을 고려하였다.

제 2절에서는 모형 불확실성하에서 일반적으로 쓰이는 복합실험설계기준을 설명하고 이들 방법론들이 본 연구에서 어떻게 적용될 수 있는지 알아보고, 새로운 실험설계 기준을 정의하여 본다, 모형 불확실성이 없는 경우를 일단 정의하고 이를 모형 불확실성하에서의 문제로 확대하여 본다. 그리고 제 3절에서는 이들 정의에 따른 예제를 그리고 제 4절에서는 결론과 토의를 다룰 것이다.

2. 외삽-최적실험설계와 모형 불확실성

제 1절에서도 언급이 되었지만 다항회귀모형에 대한 외삽-최적은 오랜 역사를 가지고 있다. Kiefer와 Wolfowitz (1964)와 Hoel과 Levine (1964)은 일변다항회귀모형에 대해 Studden (1971)은 다변다항회귀모형에 대해서도 외삽최적의 해를 연구하였다.

본 연구의 목적상 식 (1.1)을 다음과 같이 분해하여 본다.

$$y(x) = f_m^T(x)\theta + \epsilon = f_j^T(x)\theta_j + \sum_{i=j+1}^m \theta_i x^i + \epsilon, \quad (2.1)$$

여기서 모수벡터 $\theta_j^T = (\theta_0, \theta_1, \dots, \theta_j)$ 는 $j+1$ 차원을 가진다. 식 (2.1)은 편의상 모형의 불확실성을 염두에 둔 모형 표현식이다. 경우에 따라 참의 모형은 식 (1.1)일 수도 있고 식 (2.1)의 부분 모형인

$$y(x) = f_j^T(x)\theta_j + \epsilon \quad (2.2)$$

이 참일 수도 있다. 본 연구에서는 외삽실험설계에 관한 기준과 모형의 모수에 관한 기준을 복합적으로 다루어야 하므로 간단히 외삽-효율과 D -효율을 정의한다.

외삽-효율: 외삽-최적에 대해 실험설계 ξ 의 외삽-효율 ϕ_ξ^z 은 다음과 같이 정의된다.

$$\phi_\xi^z = \frac{v_m(\xi_z^*, z)}{v_m(\xi, z)},$$

여기서 ξ_z^* 는 $z \notin \Omega$ 에서 $f_m(x)$ 에 대한 외삽최적실험설계(optimal extrapolation design)이다. 이와 비슷하게 실험설계 ξ 의 D -효율 및 G -효율도 정의된다.

D -효율: D -최적에 대해 실험설계 ξ 의 D -효율 ϕ_ξ^D 은 다음과 같이 정의된다.

$$\phi_\xi^D = \left(\frac{\det(M(\xi))}{\det(M(\xi_m^D))} \right)^{\frac{1}{m+1}},$$

여기서 ξ_m^D 는 $f_m(x)$ 에 대한 D -최적실험설계이며 \det 는 행렬식을 의미한다.

G -효율: G -최적에 대해 실험설계 ξ 의 G -효율 ϕ_ξ^G 은 다음과 같이 정의된다.

$$\phi_\xi^G = \frac{m+1}{\max_{z \in \Omega} f_m^T(z) M_m^{-1}(\xi) f_m(z)},$$

여기서 G -최적인 경우 분자의 값은 동격성에 의해 $m+1$ 이 된다.

위와 같은 효율은 어느 실험설계기준 ψ 이든지 정의가 될 수 있다. 간단히 표현하면 실험설계 ξ 의 ψ -효율이란 실험설계기준 ψ 에 대해 실험설계 ξ 가 ψ -최적실험설계 ξ_ψ^* 에 대해 가지는 효율 ϕ_ξ^ψ 이라 정의가 된다.

이러한 D -, G - 그리고 외삽-최적과 같은 실험설계기준을 이용하여 최적 실험설계를 구하는 문제는 주어진 모형이 참이라고 가정하여야 한다. 본 연구에서는 언급된 다항회귀 모형의 차수가 사전에 주어져야 한다는 문제이다. 이는 현실적으로 가정하기 힘든 상황이며 또한 실험설계 기준 역시 실험자의 선호도에 따라 복수의 실험설계기준을 설정하여야 하는 문제가 발생된다. 이를 위해 문헌에서는 다음과 같은 고전적인 방법이 존재한다.

방법 1 복합실험설계: Läuter (1974)가 제안한 실험설계기준으로 k 개의 실험설계기준 ψ_i , $i = 1, 2, \dots, k$ 이 존재하는 경우 다음과 같은 선형 결합을 최대화하는 실험설계이다.

$$\max_{\xi} \alpha_1 \phi_\xi^{\psi_1} + \alpha_2 \phi_\xi^{\psi_2} + \dots + \alpha_k \phi_\xi^{\psi_k}, \quad 0 \leq \alpha_i \leq 1, \quad \sum_{i=1}^k \alpha_i = 1.$$

방법 2 제약조건실험설계: Stigler (1971) 등이 제안한 실험설계기준으로 실험설계기준 $\psi_i, i = 1, 2, \dots, k$ 개 존재 하는 경우, 이 중 목적함수에 들어갈 실험설계기준 ψ_j 을 제외한 나머지를 제약조건 식으로 두고 목적함수에 위치한 실험설계기준 ψ_j 에 대한 효율을 최대화하는 실험설계이다.

$$\max_{\xi} \phi_{\xi}^{\psi_j}, \text{ subject to } \phi_{\xi}^{\psi_i} \geq c_i, \quad i \neq j.$$

그러나 $k > 2$ 인 경우에 제약조건식을 모두 만족하지 못하는 비타당(infeasible) 해의 가능성 때문에 이러한 방법은 특수한 경우를 제외하고는 $k = 2$ 에 국한 하는 경우가 대부분이다. 이를 순차적이거나 우회적으로 접근하는 논문들이 다수 존재하는데 Wong (1995)이나 Kahng과 Kim (2002)이나 Kim과 Lim (2007) 등이 그것이다.

그리고 방법 1과 방법 2에도 동격성 이론이 존재하는데 Cook과 Wong (1994)에 의하면 이는 방법 1에 주어진 α_i 값들에 해당하는 제약조건 실험설계가 존재 한다는 의미이다. 따라서 제약조건 실험설계에 대한 알고리즘은 방법 1에 쓰이는 알고리즘을 이용하면 된다.

방법 3 최소-최대(maxi-min) 실험설계: k 개의 실험설계기준 $\psi_i, i = 1, 2, \dots, k$ 이 존재하는 경우 실험설계가 개개의 실험설계기준에 대해 가지는 효율을 $\phi_{\xi}^{\psi_i}, i = 1, 2, \dots, k$ 라 한다면 최소-최대 실험설계는 다음과 같은 실험설계를 의미한다.

$$\max_{\xi} \min_i \phi_{\xi}^{\psi_i}, \quad i = 1, 2, \dots, k.$$

이 실험설계는 매우 보수적인 방법으로 Imhof와 Wong (2000)에 의하면 $k = 2$ 인 경우, $\phi_{\xi}^{\psi_1} = \phi_{\xi}^{\psi_2}$ 가 되어 버린다. 물론 $k > 2$ 인 경우는 이를 보장하지 못한다. 이 방법은 방법 1에 의해 가중치를 부여하지 못하고 방법 2의 제약조건에서 최저효율성을 보장하지 못하는 단점이 있으나 이러한 실험자의 주관적인 정보를 배려하지 못하는 경우 매우 유용하다고 본다.

본 연구에서는 이러한 3가지 방법에 빗대어 주안점을 두고 있는 외삽 문제를 고려하여 본다. 주어진 실험영역에서 외삽점이 벗어나면 날수록 그 모형의 타당성은 의심을 받는데 이러한 문제 해결을 위한 시도로 Huang과 Studden (1988)에 의해 강건실험설계가 이루어 졌다. 주어진 모형에서 이탈되는 모형들을 하나의 집합체의 구성원으로 보고 이탈되는 모형의 평균제곱오차(MSE)의 최대값을 최소화하는 이론적인 논문을 발표하였다. 그러나 이 논문은 보편적으로 많이 사용하는 실험설계의 효율을 기준으로 작성된 것이 아니다. 한편으로 Dette와 Wong (1996)은 주어진 다항회귀모형의 차수가 조금만 이탈하여도 $z \notin \Omega$ 가 실험영역에서 멀리 벗어나면 날수록 외삽-최적의 실험설계는 그 효율이 급속히 나빠짐을 보인 바 있다.

제 1절에서 언급하였듯이 본 연구에서는 외삽에 해당하는 $z \notin \Omega$ 를 하나에서 여러 개로 확장하려 한다. 즉, 실험영역을 벗어나는 여러 개의 점 $z_i \notin \Omega, i = 1, 2, \dots, k$ 으로 확장될 때 생기는 문제를 다루려고 한다. 실험자는 하나의 점에서의 단 하나의 외삽문제 뿐 아니라 여러 개의 외삽에 관한 관심을 두는 경우가 종종 있다. 그러한 $z_i \notin \Omega, i = 1, 2, \dots, k_1$ 가 실험영역에서 조금 벗어나면 모형에 관한 불확실성의 문제가 대두 되지 않겠으나 $z_i \notin \Omega, i = k_1 + 1, \dots, k$ 가 실험영역에서 많이 벗어나면 모형 불확실성의 문제가 많이 발생된다. 따라서 본 연구에서는 이러한 문제를 다룰 것이다. 실험영역을 벗어나는 점들의 선택은 편의상 Dette와 Wong (1996)의 논문의 예를 따르도록 한다. 그들은 논문에서 세 외삽점, $z_1 = 1.1, z_2 = 2, z_3 = 4.0$ 을 선택하여 각각의 z_i 에 대한 복합 실험설계를 모형 불확실성에 비추어 보았다. $z_1 = 1.1$ 은 실험영역에서 아주 인접한 점 그리고 $z_3 = 4.0$ 은 아주 멀리 떨어진 점을 고려한 것이다.

본 연구에서는 이러한 점들에 대한 실험자의 외삽문제가 동시에 발생하는 경우를 다룰 것이다. 실험자의 실험설계기준이 3개의 외삽 실험설계기준으로 확장되어도 위에서 언급된 3가지 방법은 자연스럽게 적용된다. 그러나 방법 1과 방법 2는 본질적으로 동격이므로 본 연구는 방법 2와 방법 3을 선택한다.

정의 2.1 다음과 같은 실험설계 ξ 를 k 개의 외삽점 $z_i, i = 1, 2, \dots, k$ 이 존재하는 경우 식 (2.1)에서의 차수가 m 인 다항회귀모형 f_m 에 대한 최소-최대 외삽실험설계(MAXI-MIN EXTRAPOLATION OPTIMAL DESIGN)이라 한다.

$$\max_{\xi} \min_i \phi_{\xi}^{z_i}, \quad i = 1, \dots, k,$$

여기서 $\phi_{\xi}^{z_i}$ 는 i 번째 외삽점 z_i 에 대해 실험설계 ξ 가 가지는 외삽-효율이다. 이 정의는 모형에 관한 불확실성이 없는 경우에 국한하는 실험설계이다. 그러나 실험영역을 벗어나는 외삽인 경우 이를 보장하지 못하기 때문에 이를 고려한 실험설계 기준을 정의 2.2와 정의 2.3에서 제안한다.

정의 2.2 다음과 같은 실험설계 ξ 를 k 개의 외삽점 $z_i, i = 1, 2, \dots, k$ 가 존재하는 경우 두 다항회귀모형 $f_h, h = j, m$ 에 대한 최소-최대 외삽실험설계(MAXI-MIN MODEL-DEPENDENT EXTRAPOLATION DESIGN)이라 한다.

$$\max_{\xi} \min_h \min_i \phi_{\xi}^{z_i}(f_h), \quad i = 1, \dots, k,$$

여기서 f_m 는 식 (2.1)에서의 차수가 m 인 다항회귀 모형이고 f_j 는 차수가 j 인 다항회귀모형으로 부분모형이다. 그리고 $\phi_{\xi}^{z_i}(f_h)$ 는 모형 f_h 와 i 번째 외삽점 z_i 에 대해 실험설계 ξ 가 가지는 외삽-효율이다.

정의 2.3 다음과 같은 실험설계 ξ 를 k 개의 외삽점 $z_i, i = 1, 2, \dots, k$ 이 존재하는 경우 다항회귀 두 모형 $f_h, h = j, m$ 에 대한 제약 최소-최대 외삽실험설계(MAXI-MIN MODEL-DEPENDENT CONSTRAINED EXTRAPOLATION DESIGN)라 한다.

$$\max_{\xi} \min_i \phi_{\xi}^{z_i}(f_m), \quad i \neq l,$$

$$\text{SUBJECT TO } \phi_{\xi}^{z_l}(f_j) \geq c_l,$$

여기서 f_m 는 식 (2.1)에서의 차수가 m 인 다항회귀 모형이고 f_j 는 차수가 j 인 다항회귀모형으로 부분모형이다. 그리고 제약조건에 명시된 l 은 $i = 1, 2, \dots, k$ 의 하나이다.

정의 2.3은 부분모형에 대한 한 외삽점의 외삽-효율이 어느 정도 보장되는 상태에서 다른 외삽점들에서 전체 모형에 대한 최소효율을 최대화하는 실험설계기준이다. 정의 2.3에서 제안된 기준은 위에서 설명된 방법 2를 응용하여 Kim과 Lim (2007)이 제안한 방법을 따른 것이다. 즉, 3개 이상의 실험설계기준이 존재하는 경우 방법 2는 위에서 언급된 해의 비타당성(infeasibility) 문제가 대두되어 이중 하나만 제약 조건식으로 돌리고 나머지를 maxi-min 방법을 적용한 것이다. 물론 $k = 2$ 인 경우는 통상적인 방법 2와 일치 한다.

정의 2.2에 제시된 실험설계기준은 정의 2.3과 달리 모든 모형에 관한 최소효율을 최대화하는 실험설계 기준으로 정의 2.3에 비해 단순하나 실험영역을 벗어나는 모형에 관한 사전 정보가 없는 경우에 유용하게 쓰일 수 있다. 정의 2.3은 특정모형에 대한 최소효율을 보장한 상태에서 실험기준을 찾는 것이 다른 점이다.

그러나 정의 2.2와 정의 2.3의 실험설계기준은 모형에 따른 외삽-효율만 다를 뿐 모형에 관한 불확실성을 감안하지 않아 정의 2.4와 같은 종합적인 실험설계기준을 설정할 수 있다.

정의 2.4 다음과 같은 실험설계 ξ 를 k 개의 외삽점 $z_i, i = 1, 2, \dots, k$ 이 존재하는 경우 두 다항회귀모형 $f_h, h = j, m$ 에 대한 모형-강건 제약 최소-최대 외삽실험설계(MODEL-ROBUST MAXI-MIN CONSTRAINED EXTRAPOLATION DESIGN)이라 한다.

$$\max_{\xi} \min_h \min_i \phi_{\xi}^{z_i}(f_h), \quad i = 1, \dots, k,$$

$$\text{SUBJECT TO } \max_{\xi} \min_h \phi_{\xi}^D(f_h) \geq c,$$

Table 3.1. D - and extrapolation optimal design for simple linear regression model.

$D(G)$ -최적	$\xi(\pm 1) = 0.5$
외삽최적, $z = 1.1$	$\xi(-1) = 0.0455, \xi(1) = 0.9545$
외삽최적, $z = 2.0$	$\xi(-1) = 0.2500, \xi(1) = 0.7500$
외삽최적, $z = 4.0$	$\xi(-1) = 0.3750, \xi(1) = 0.6250$

Table 3.2. D - and extrapolation optimal experimental design for quadratic regression model.

$D(G)$ -최적	$\xi(\pm 1) = \xi(0) = 0.3333$
외삽최적, $z = 1.1$	$\xi(-1) = 0.0387, \xi(0) = 0.1479, \xi(1) = 0.8134$
외삽최적, $z = 2.0$	$\xi(-1) = 0.1429, \xi(0) = \xi(1) = 0.4286$
외삽최적, $z = 4.0$	$\xi(-1) = 0.1935, \xi(0) = 0.4839, \xi(1) = 0.3226$

여기서 $\phi_{\xi}^{z_i}(f_h)$ 는 모형 f_h 과 i 번째 외삽점 z_i 에 대해 실험설계가 ξ 가 가지는 외삽-효율이며, $\phi_{\xi}^D(f_h)$ 는 모형 f_h 에 대해 실험설계 ξ 가 가지는 D -효율이다. 정의 2.4는 위에서 언급된 LÄUTER (1974)의 방법 1을 따라 더 발전시킬 수 있으나 본 연구에서는 가중치를 부여하는 문제가 객관적이지 못한 점이 발생하여 이는 생략하고 방법 2와 방법 3을 혼합한 방법만을 고려하였다.

정의 2.4는 혼합적인 방법으로 다중점에서의 외삽문제 뿐 아니라 모형에 관한 불확실성을 고려하여 제공된 기준이다. 제 3절에서는 제 2절에서 밝힌 세 가지 정의에 의한 실험설계를 간단한 예제를 통해 그 의미를 알아보하고자 한다.

3. 예제

본 연구에서는 고려되는 모형의 개수는 단순회귀모형과 2차회귀모형 두 개로만 설정한다. 실험자는 이 두 모형 중 어느 모형이 참의 모형인지 모르는 상태에서 실험설계를 진행하여야 하고 외삽에 관한 점은 3개로 국한한다. 이미 명시 되었지만 $z_i \notin \Omega, i = 1, 2, 3$ 이며 $z_1 = 1.1, z_2 = 2.0, z_3 = 4.0$ 으로 한다. 제 2절에서 정의한 새로운 실험설계기준을 구하기 전에 이해에 도움을 주는 몇 가지 실험설계를 예제 3.1부터 예제 3.3까지 표기하여 놓았다. 이 내용은 Kim과 Jang (2012)의 논문에서 발췌하였다.

예제 3.1: 단순회귀모형에 대한 $D(G)$ -최적과 외삽-최적 실험설계

실험영역 $\Omega = [-1, 1]$ 에 단순회귀모형에 대한 $D(G)$ -최적과 외삽-최적 실험설계인데 특히 외삽-최적에서 $z \notin \Omega$ 가 1에 가까우면 $\xi(1)$ 의 값이 커지고 $z \notin \Omega$ 가 1에서 멀리 치우칠수록 $D(G)$ -최적, $\xi(\pm 1) = 0.5$ 에 가까워진다는 사실을 알 수 있다. 외삽점이 실험영역에서 벗어나가면 외삽점에 어떤 의미도 부여하지 못한다는 의미로 해석된다.

예제 3.2: 2차 회귀모형에 대한 $D(G)$ -최적과 외삽-최적 실험설계

실험영역 $\Omega = [-1, 1]$ 에서 2차 회귀모형에 대한 $D(G)$ -최적과 외삽-최적 실험설계는 Table 3.2와 같다. 역시 예제 3.1과 마찬가지로 $z \notin \Omega$ 가 1에서 멀리 벗어날수록 $\xi(1)$ 의 질량이 상대적으로 줄어든다.

참고로 2차 회귀 모형을 가정하고 구한 외삽-최적 실험설계는 참의 모형이 단순회귀모형인 경우 Table 3.3과 같은 효율이 나온다. Table 3.3에 의하면 최저효율이 37.83%까지 떨어짐을 알 수 있다. 즉, 외삽점 $z = 1.1$ 에서의 이차 회귀모형의 외삽-최적 실험설계가 참의 모형이 단순회귀모형이고 외삽점이 4.0에 대해서는 효율이 37.83%로 떨어진다는 의미이다. 또한 대각선의 값을 보면 같은 외삽점이라도 외삽점이 1에서 이탈될수록 효율이 떨어진다.

Table 3.3. Extrapolation efficiency of extrapolation-optimal design of quadratic regression model with respect to simple regression model.

단순회귀	2차 모형		
	$z_1 = 1.1$	$z_2 = 2.0$	$z_3 = 4.0$
$z_1 = 1.1$	0.8521	0.5142	0.4190
$z_2 = 2.0$	0.5748	0.5715	0.4999
$z_3 = 4.0$	0.3783	0.5487	0.5161

Table 3.4. Maxi-min $D(G)$ - optimal design for simple and quadratic regression model.

ξ	$\phi_{\xi}^D(\xi_1^*)$	$\phi_{\xi}^D(\xi_2^*)$
$\xi(\pm 1) = 0.4191, \xi(0) = 0.1618$	0.9155	0.9155

예제 3.3: 단순회귀모형과 2차 회귀모형에 대한 $D(G)$ -최적 최대-최소 실험설계

모형이 두 개인 경우 각 모형에 대한 $D(G)$ -최적 최대-최소 실험설계의 효율은 일치 한다. 각 모형에 대한 91.55%의 효율을 확인하기 바란다. 정의 3.4에서 제약조건식에서 설정된 c 의 값은 이보다 낮은 값으로 설정되어야 한다.

이상과 같은 기본적인 실험설계를 바탕으로 제 2절에서 제안한 정의 2.1, 2.2, 2.3 그리고 2.4에 대한 실험설계를 구하여 본다.

유전알고리즘은 이미 Park 등 (2005)에 의해 그 유용성이 밝혀진바 있는데 특히 Limmun 등 (2013)에 따라 비규칙 실험영역(irregular design space)이라든지 제약조건이 있는 실험설계인 경우는 전통적인 교환알고리즘(exchange algorithm)은 적용하기 어렵다고 판단되어 본 연구에서는 유전알고리즘을 채택하여 최적해를 구하였다. 또한 유전알고리즘의 모집단의 크기는 50으로 설정되었다. 통상적으로 모집단의 크기는 30에서 100으로 하나 연구에서 설정된 문제의 크기가 크지 않아 50이상의 모집단의 크기는 필요하지 않다.

예제 3.4이하의 모든 예제는 유전알고리즘을 이용하여 해를 구하였는데 본 연구에서는 팰리사이드(palisade)사의 evolver란 엑셀 기반의 상용프로그램을 사용하여 최적 해를 구하였다. 교배율과 돌연변이율을 각각 0.5와 0.10으로 설정하였다. 주어진 실험설계의 조건에 맞는 모집단을 구성한 다음 교배와 돌연변이 과정을 통해 목표 값의 향상이 이루어졌는지를 감시하여 수렴전 100단계 전부터 목표 값이 0.001% 이내의 범위로 움직이지 않으면 최적화 계산을 정지시키는 옵션을 설정하였다. 유전알고리즘의 특징상 최적해를 구하는 과정이 교착되는 것을 방지하기 위해 돌연변이율을 점증적으로 높이는 옵션도 돌연변이율 0.2이하에서 추가하였다. 유전알고리즘의 효율성은 모집단의 크기, 교배율, 돌연변이율의 조합에 상당히 의존하지만 본 연구의 예제에서는 문제가 없었다.

예제 3.4: 이미 언급하였듯이 Dette와 Wong (1996)의 논문에서 예제로 제안한 외삽점, $z_1 = 1.1$, $z_2 = 2.0$, $z_3 = 4.0$ 을 그대로 다중외삽점으로 사용한다. 정의 2.1에서 언급한 실험설계이다. Table 3.5는 2차 회귀 모형에 대한 정의 2.1에서 언급된 실험설계이다. 물론 이 실험설계는 2차회귀모형의 타당성이 모든 외삽점에 대해 존재하여야 한다. Table 3.5에 의하면 최소 외삽-효율의 최대값은 양 극단의 외삽점에서 발생한다. 보수적인 실험설계기준이므로 중간에 있는 외삽점에 대한 최소효율은 이미 어느 정도 보장을 받는다고 해석할 수 있다. 즉, 중간 외삽점을 고려하지 않고 양 극단의 외삽점에 대한 최소-최대방법을 시행하여도 같은 결과가 나온다. 결론은 일변 다항회귀 모형인 경우 다중 외삽점에 대한 실험설계라 하더라도 양극단에 위치한 외삽점에 대해 실험자는 관심을 가져야 한다는 의미로 해석된다.

Table 3.5. Maximin extrapolation optimal design for multiple extrapolation points.

ξ	$\phi_{\xi}^{z_1}$	$\phi_{\xi}^{z_2}$	$\phi_{\xi}^{z_3}$
$\xi(-1) = 0.1125, \xi(0.0514) = 0.3124, \xi(1) = 0.5751$	0.8045	0.9227	0.8045

Table 3.6. Maximin model dependent extrapolation design for simple and quadratic regression model.

$\xi(-1) = 0.1658, \xi(0.0242) = 0.2241, \xi(1) = 0.6101$		
$z_1 = 1.1:$	$\phi_{\xi}^{z_1}(f_m) = 0.8366$	$\phi_{\xi}^{z_1}(f_j) = 0.6966$
$z_2 = 2.0:$	$\phi_{\xi}^{z_2}(f_m) = 0.8042$	$\phi_{\xi}^{z_2}(f_j) = 0.7709$
$z_3 = 4.0:$	$\phi_{\xi}^{z_3}(f_m) = 0.6966$	$\phi_{\xi}^{z_3}(f_j) = 0.6966$

여기서 $\phi_{\xi}^{z_i}(f_m), \phi_{\xi}^{z_i}(f_j)$ 는 각각 2차 회귀모형과 단순회귀모형에 대해 외삽점 z_i 에서 실험설계 ξ 가 가지는 외삽-효율이다.

Table 3.7. Constrained extrapolation design for simple and quadratic regression model.

$c = 0.8: \xi(-1) = 0.0822, \xi(0.2879) = 0.2122, \xi(1) = 0.6746,$		
$c = 0.7: \xi(-1) = 0.1115, \xi(0.1996) = 0.2913, \xi(1) = 0.5972.$		
$z_1 = 1.1:$	$\phi_{\xi}^{z_1}(f_j) = 0.8 \geq 0.8,$	$\phi_{\xi}^{z_1}(f_j) = 0.7 \geq 0.7$
$z_2 = 2.0:$	$\phi_{\xi}^{z_2}(f_m) = 0.6601,$	$\phi_{\xi}^{z_2}(f_m) = 0.8503$
$z_3 = 4.0:$	$\phi_{\xi}^{z_3}(f_m) = 0.5559,$	$\phi_{\xi}^{z_3}(f_m) = 0.7530$

여기서 $\phi_{\xi}^{z_i}(f_m), \phi_{\xi}^{z_i}(f_j)$ 는 각각 2차 회귀모형과 단순회귀모형에 대해 외삽점 z_i 에서 실험설계 ξ 가 가지는 외삽-효율이다.

예제 3.5: Table 3.6은 정의 2.2에 의한 실험설계결과이다. 모형 불확실성과 다중외삽점을 모두 고려한 maxi-min-min 방법으로 매우 보수적인 결과를 초래 할 것이다. Table 3.6의 결과에 의하면 모든 모형에 대한 최소효율의 최대값은 약 69.66%로 나타났다. 그러나 이 실험설계는 외삽 효율만을 고려한 결과이다. 당연히 예제 3.4의 효율보다는 낮으나 이는 모형을 지정하지 않은 결과이므로 의외적인 결과는 아니다. 오히려 Table 3.3에서 언급된 최저효율보다 훨씬 높은 효율을 유지하고 있다.

예제 3.6: 모형이 불확실한 경우이다. $z = 1.1$ 에 대해서는 단순회귀모형이 적절하나 실험영역을 많이 벗어나는 $z_2 = 2.0, z_3 = 4.0$ 인 경우 2차 회귀모형을 가정하고 외삽에 관한 실험설계를 하여야 한다. 정의 2.3에 의한 실험설계를 시행하여 보자. 단순회귀모형에 대한 외삽점에서의 외삽-효율은 적어도 80%와 70%를 초과하는 두 개의 제약조건으로 다른 2 외삽점에서의 효율을 최소최대도 한 실험설계를 시행하면 Table 3.7과 같다. 이 경우 정의 2.3에 의한 실험설계는 두 제약조건의 하한값을 정확히 만족시킨다. 그리고 단순회귀모형에 대한 외삽-효율을 높이면 다른 두 외삽점에서의 외삽-효율은 떨어짐을 알 수 있다. c 값을 실험자가 변하게 함으로써 다른 외삽점에서의 2차회귀모형에 대한 외삽-효율을 조절할 수 있는 장점이 생긴다.

예제 3.7: 정의 2.4에 의한 실험설계이다. 정의 2.4에 의한 실험설계는 외삽-효율의 최적화를 위한 후보 실험설계들의 제약조건식에 모형에 관한 D -효율을 고려하였다. 제약 조건의 하한값 c 값은 0.8로 설정하였으며 최적실험설계는 Table 3.8과 같다. 역시 정의 2.4에 의한 실험설계는 실험설계의 제약조건의 하한값을 정확히 만족시킨다. D -효율을 80% 보장을 받는 상태에서 최소효율은 모형에 상관없이 최소 외삽-효율의 최대값을 약 67% 달성하는 결과를 낳았다.

Table 3.8. Model-robust maxi-min constrained extrapolation design for simple and quadratic regression model.

$\xi(-1) = 0.2144, \xi(0.135197) = 0.2164, \xi(1) = 0.5692$		
$z_1 = 1.1:$	$\phi_{\xi}^{z_1}(f_m) = 0.7665$	$\phi_{\xi}^{z_1}(f_j) = 0.6719$
$z_2 = 2.0:$	$\phi_{\xi}^{z_2}(f_m) = 0.7577$	$\phi_{\xi}^{z_2}(f_j) = 0.7874$
$z_3 = 4.0:$	$\phi_{\xi}^{z_3}(f_m) = 0.6769$	$\phi_{\xi}^{z_3}(f_j) = 0.7466$

여기서 $\phi_{\xi}^{z_i}(f_m), \phi_{\xi}^{z_i}(f_j)$ 는 각각 2차 회귀모형과 단순회귀모형에 대해 외삽점 z_i 에서 실험설계 ξ 가 가지는 외삽-효율이다.

4. 결론과 토의

본 연구에서는 외삽-최적이 가지고 있는 근본적인 문제점 즉, 외삽점 $z \notin \Omega$ 가 실험영역인 Ω 를 벗어날수록 모형에 관한 불확실성이 증대하는 경우 실험설계의 문제점을 알아보았다. 더욱이 외삽점이 여러 개 일 때 이러한 문제를 어떻게 처리할 것인가에서 본 연구는 출발하였다. 문헌상에 존재하는 방법론들을 각색하여 네 가지 실험설계기준을 제안하였다. 이러한 시도는 다중 외삽점에서의 문제뿐만 아니라 모형 불확실성에 관한 문제까지 감안한 실험설계기준이므로 다소 복잡하게도 보일 수 있지만 그 문제점을 이해하는 데는 도움을 줄 수 있다. 실험자의 시험의도를 정확히 파악하여야 하는 문제점을 안고 있지만 실험자의 의도를 최대한 반영하는 실험설계기준이다. 실험적인 벤치마킹 모형인 단순회귀모형과 2차 회귀모형의 결과만 가지고 제안된 실험설계기준을 평가하기는 이르다고 볼 수 있지만 결과에서 보이는 의미는 실험자의 의도를 적극적으로 반영하지 않으면 효율이 낮은 실험설계를 초래한다는 점이다. 향후 연구결과는 일반적인 다항회귀모형으로 확장시키는 것이며 근사실험설계가 아닌 정확실험설계로 시스템을 만드는 것이 될 것이다. 다변 다항회귀모형의 문제도 향후 연구과제이다.

References

- Box, G. E. P. and Draper, N. R. (1975). Robust design, *Biometrika*, **62**, 347–352.
- Cook, R. D. and Wong, W. K. (1994). On the equivalence between constrained and compound optimal designs, *Journal of the American Statistical Association*, **89**, 687–692.
- Detle, H. and Wong, W. K. (1996). Robust optimal extrapolation designs, *Biometrika*, **83**, 667–680.
- Hoel, P. G. and Levine, A. (1964). Optimal spacing and weighting in polynomial prediction, *Annals of Statistics*, **35**, 1553–1560.
- Huang, M. L. and Studden, W. J. (1988). Model robust extrapolation designs, *Journal of Statistical Planning and Inference*, **18**, 1–24.
- Imhof, L. and Wong, W. K. (2000). A graphical method for finding maximin designs, *Biometrics*, **56**, 113–117.
- Jones, B. and Nachtsheim, C. (2011). A class of three-level designs for definitive screening in the presence of second-order Effects, *Journal of Quality Technology*, **43**, 1–15.
- Kahng, M. Y. and Kim, Y. I. (2002). Multiple constrained optimal experimental design, *The Korean Communications in Statistics*, **9**, 619–627.
- Kiefer, J. and Wolfowitz, J. (1964). Optimum extrapolation and interpolation designs, *Annals of the Institute of Statistical Mathematics*, **16**, 79–108.
- Kim, Y. I. and Lim, Y. B. (2007). Hybrid approach when multiple objectives exist, *The Korean Communications in Statistics*, **14**, 531–540.
- Kim, Y. I. and Jang, D. H. (2012). Hybrid Constrained Extrapolation Experimental design, *The Korean Communications in Statistics*, **19**, 65–75.
- Läuter, E. (1974). Experimental planning in a class of models, *Mathematische Operationsforschung und Statistik*, **5**, 673–708.
- Limmun, W., Borkowski, J. J. and Chomtee, B. (2013). Using a Genetic Algorithm to Generate D -optimal Designs for Mixture Experiments, *Quality and Reliability Engineering International*, **29**, 1055–1068.

- Park, Y. J., Montgomery, D. C., Folwer, J. W. and Borrer, C. M. (2005). Cost- constrained G -efficient response surface designs for cuboidal regions, *Quality and Reliability Engineering International*, **22**, 121–139.
- Stigler, S. M. (1971). Optimal experimental design for polynomial regression, *Journal of the American Statistical Association*, **66**, 311–318.
- Studden, W. J. (1971). Optimal designs for multivariate polynomial extrapolation, *The Annals of Mathematical Statistics*, **42**, 828–832.
- Wong, W. K. (1995). A graphical approach for the construction of constrained D and L -optimal designs using efficiency plots, *Journal of Statistical Computation and Simulation*, **53**, 143–152.

다중 외삽점에서의 최적 실험설계법을 위한 실험설계기준

김영일^{a,1} · 장대흥^b

^a중앙대학교 경영경제대학, ^b부경대학교 통계학과

(2014년 6월 23일 접수, 2014년 8월 2일 수정, 2014년 8월 15일 채택)

요약

실험영역을 벗어나는 다중 외삽점들에 관한 실험설계를 기획하는 경우 실험자는 종종 어느 외삽점에 더 많은 노력을 집중하여야 하는지 주어진 모형이 있다하더라도, 고민하는 경우가 있다. 본 연구에서는 이러한 상황에 관한 실험설계 문제를 다루었다. 첫 번째는 주어진 모형이 실험영역을 벗어나더라도 모형이 타당한 경우 다중 외삽점에 관한 실험설계고 다른 하나는 그렇지 않은 경우이다. 첫 번째인 경우는 비교적 기존 문헌에서 알려진 방법들이 적용될 수 있으나 그렇지 않은 경우 즉, 모형의 타당성이 의심되는 경우는 다른 실험설계기준을 제시하여야 한다, 본 연구는 이와 관련 다양한 하이브리드 방법을 제시하여 다중 외삽점에서의 문제가 어떻게 모형 불확실성하에서 전개되어야 하는지 다루어 보았다, 이를 위해 서치알고리즘의 하나인 유전알고리즘을 적용하였다. 왜냐하면 전통적인 교환알고리즘의 복잡성보다는 유전알고리즘의 효율성이 더 뛰어났다고 보기 때문이다.

주요용어: 외삽실험설계, 모형불확실성, 제약실험설계, 복합실험설계, 최대최소 실험설계, 하이브리드 방법, 유전알고리즘.

이 논문은 2014년도 중앙대학교 연구년 결과물로 제출됨.

¹교신저자: 서울특별시, 동작구 흑석동 84, 156-756, 중앙대학교 경영경제대학. E-mail: yik01@cau.ac.kr