

입방형 영역에서의 G-효율이 높은 Model-Robust 실험설계

박유진^{1*} · 이윤주²

¹중앙대학교 상경학부 경영학과 / ²중앙대학교 통계학과 대학원

Model-Robust G-Efficient Cuboidal Experimental Designs

You-Jin Park¹ · Yoon-Ju Yi²

¹Department of Business Administration, College of Social Sciences, Chung-Ang University

²Department of Statistics, College of Natural Science, Chung-Ang University

The determination of a regression model is important in using statistical designs of experiments. Generally, the exact regression model is not known, and experimenters suppose that a certain model form will be fit. Then an experimental design suitable for that predetermined model form is selected and the experiment is conducted. However, the initially chosen regression model may not be correct, and this can result in undesirable statistical properties. We develop model-robust experimental designs that have stable prediction variance for a family of candidate regression models over a cuboidal region by using genetic algorithms and the desirability function method. We then compare the stability of prediction variance of model-robust experimental designs with those of the 3-level face centered cube. These model-robust experimental designs have moderately high G-efficiencies for all candidate models that the experimenter may potentially wish to fit, and outperform the cuboidal design for the second-order model. The G-efficiencies are provided for the model-robust experimental designs and the face centered cube.

Keyword: G-efficiency, genetic algorithm, desirability function method, model-robust experimental design, cuboidal design region

1. 서론

반응표면 분석법(Response Surface Methodology)은 여러 산업 및 연구 분야의 실험에서 널리 사용 되는 체계적인 통계적 분석방법이다. 이처럼 반응 표면 분석법을 여러 산업 및 연구 분야에 적용하기에 앞서, 일반적으로 대부분의 실험자들은 적합 시키고자 하는 특정 회귀모형(반응표면모형)을 가정하고, 이 모형에 적절한 실험을 설계한다(Myers and Montgomery, 2002). 예를 들어, 중심합성설계(Central Composite Design)와 같은 표준설계를 선택하거나 또는 D-최적성과 같은 특정 실험 최적성 기준을 통해 구해진 실험설계를 적용하기도 한다. 그러나 특정 실험

최적성 기준을 통해 구한 실험설계를 선택하는 것은 이미 가정한 회귀모형에 굉장히 의존적이기 때문에, 이 기준과 관련된 실험설계의 효율성은 데이터에 적합 된 최종 모형이 제안된 모형과 다를 경우 달라질 수 있다. 따라서 정확하지 않은 회귀모형을 선택함으로써 부적합한 통계적 특성을 가진 실험이 설계될 오류를 방지하고자, 고려되어야 할 모든 회귀모형에 둔감한 실험 즉, Model-robust 실험들을 설계할 필요가 있다. 이러한 바람직한 통계적 특성을 가진 Model-robust 실험설계는 어떤 형태의 회귀모형이 사용되든지 모든 회귀모형에 대해 전반적으로 적절한 통계적 특성들을 반영한다. 본 논문에서는 다음과 같이 독립변수의 개수가 3, 4, 5개인 경우의 회귀모형들의 집합을 고려할 것이다.

본 연구는 2009년 중앙대학교 교내일반학술연구비에 의해 지원되었음

*연락처 : 박유진 교수, 456-756 경기도 안성시 대덕면 내리 72-1, 중앙대학교 안성캠퍼스 사회과학대학 상경학부 경영학과,

Fax : 031-675-1384, E-mail : eugenepark@cau.ac.kr

투고일(2009년 10월 24일), 심사일(1차 : 2010년 01월 19일), 게재확정일(2010년 03월 23일).

$$y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \epsilon, \quad (1)$$

$$y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{i=1}^{k-1} \sum_{j=i+1}^k \beta_{ij} x_i x_j + \epsilon \quad (2)$$

$$y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{i=1}^{k-1} \sum_{j=i+1}^k \beta_{ij} x_i x_j + \sum_{i=1}^k \beta_{ii} x_i^2 + \epsilon \quad (3)$$

여기서 y 는 관측치, β_0 는 절편, β_i 는 1차 항의 계수들, β_{ii} 와 β_{ij} 는 각각 순수 2차 항의 계수들과 교호작용 항의 계수들이다. 그리고 ϵ 는 오차항을 나타내는 것으로 $NID(0, \sigma^2)$ 라고 가정한다(Myers and Montgomery, 2002). 본 연구에서는 Model-robust 실험설계들을 찾기 위해 탐색적 알고리즘의 하나인 유전자 알고리즘(Genetic algorithm)과 다중 반응 최적화(다목적 함수 최적화)에서 사용되는 만족도 함수 기법(Desirability function method)을 사용하였다(Derringer and Suich, 1980). 그리고 Model-robust 실험설계와 3-수준의 면심입방 실험설계의 척도화된 예측 분산(Scaled prediction variance : SPV)의 비교를 통해 Model-robust 실험설계들이 2차 회귀모형들에 대한 면심입방 실험설계보다 뛰어난 성능을 보이는지의 여부를 살펴볼 것이다.

2. 실험설계 최적성(Design Optimality)과 Model-Robust 실험설계

2.1 최적 설계 이론(Optimum Design Theory)

효율적인 실험을 설계하는 것과 구축된 실험설계들을 비교하기 위한 적절한 평가 기준들을 개발하는 것은 반응표면 분석법에서 예전부터 중요하게 다루어지고 있다. 반응표면 분석법에서 일반적으로 사용되는 실험설계 최적성 기준에 대한 내용은 Myers and Montgomery(2002)에 자세히 설명되어 있다. 실험설계 최적성 기준 중에서 가장 잘 알려져 있고 또한 빈번히 사용되는 기준은 D-최적성(D-optimality) 기준으로, 회귀모형의 추정된 계수들의 분산과 공분산을 최소화하는 실험을 설계하는 것과 관련된 것이다(Kiefer, 1959; Kiefer, 1961; Wald, 1943). 이와는 대조적으로, 반응치의 예측 분산과 관련된 실험설계 최적성 기준은 G-최적성(G-optimality) 기준이라고 하고, 특히 실험설계 공간상에서 최대 예측 분산이 최소인 경우의 실험설계를 G-최적화(G-optimal)가 되었다고 한다. G-최적성 기준은 반응표면 분석법에서 최적 실험설계 이론과 실험설계의 구축에 있어서 상당히 중요한 위치를 차지하고 있는데, 이는 대부분 반응표면 분석법에 관한 연구에서의 궁극적인 목적은 가능한 분산이 적은 반응치를 예측하는 것이기 때문이다. 본 연구에서 고려하는 회귀모형들인 식 (1)~식 (3)은 다음과 같이 행렬의 형태로 나타낼 수 있다.

$$y = X\beta + \epsilon \quad (4)$$

여기서 y 는 관측치들의 벡터, β 는 추정된 회귀모수들의 벡터, X 는 고려하고 있는 회귀모형 형태로 확장된 실험설계 행렬, 그리고 ϵ 는 $NID(0, \sigma^2)$ 를 따르는 랜덤 오차들의 벡터이다. 추정된 회귀모수들의 최소 제곱 예측치는 $\hat{\beta} = (X'X)^{-1}X'y$ 이며, 설계 공간에서 특정한 하나의 실험점 $x' = [x_1, x_2, \dots, x_k]$ 에서 예측된 값은 다음과 같다.

$$\hat{y} = x'^{(m)}\hat{\beta} \quad (5)$$

여기서 $x'^{(m)}$ 은 모형의 형태로 확장된 점인 x' 이다. 점 $x'^{(m)}$ 에서 반응치의 예측 분산과 척도화된 예측 분산은 다음과 같다.

$$Var[\hat{y}(x)] = x'^{(m)}(X'X)^{-1}x^{(m)}\sigma^2 \quad (6)$$

$$v(x) = N[Var(\hat{y}(x))]/\sigma^2 = Nx'^{(m)}(X'X)^{-1}x^{(m)} \quad (7)$$

따라서, 구축된 실험설계의 G-효율은 다음과 같이 정의된다.

$$G_{eff} = \frac{p}{\max_{x \in R}[v(x)]} \quad (8)$$

구축된 실험설계의 G-효율이 가장 좋은 경우에는 $\max[v(x)]$ (척도화된 예측분산의 최대치)의 값이 고려하고 있는 회귀모형의 모수(Parameter)의 값(p)과 같으므로 G-효율은 1 즉, 100%가 된다. 이와 같이 하나의 특정한 실험설계기준을 통해 최적의(혹은 효율적인) 실험을 설계하기 위해서, 다양하고 유용한 알고리즘들이 개발되어 사용되고 있다. D-최적 실험을 설계하기 위해 가장 많이 쓰이는 알고리즘은 교환 형태 알고리즘(Exchange-type algorithms)이다(Fedorov, 1972; Mitchell, 1974; Wynn, 1970). 그러나 정확한 실험들을 구축하는데 바탕이 되는 이론이 없으므로, 현재의 어떠한 알고리즘도 정확한 D-최적 실험설계를 보장하진 못한다. Hamada, Martz, Reese, 그리고 Wilson(2001)은 몇 가지 회귀모형들에 대해 최적에 근사한 베이지안 실험들을 설계하고자 유전자 알고리즘을 사용하였고, 이에 연속형 반응변수와 이진수 형태의 반응변수들을 모두 고려하였다. Borkowski (2003)는 독립변수의 개수 (k)가 1~3인 경우에, 2차 회귀모형들에 대하여, 초입방체의 실험설계영역에서 최적에 근사한 n -점으로 이루어진 D-최적, A-최적, G-최적, 그리고 IV-최적의 실험을 설계하기 위해 유전자 알고리즘을 사용하였다. Heredia-Langner, Carlyle, Montgomery, Borrer, 그리고 Runger(2003)는 혼합물 문제들과 같은 제약된 영역에서의 D-효율 실험들을 설계하고자 유전자 알고리즘을 적용하였다. 이외에 유용한 실험설계들을 구축하기 위한 알고리즘들에 대한 정보는 Cook and Nachtsheim(1980), Welch(1982), Haines(1987), 그리고 Meyer and Nachtsheim(1995) 등에서 찾을 수 있다.

2.2 Model-Robust 실험설계

Model-robust 실험설계들을 구축하는 문제에 대한 여러 가지 접근법들 중, 특히 회귀모형들을 판별하기 위한 기술들에 대

해 많은 연구가 있어 왔다. Hill과 Reiner(1965)는 부적합한 모형들을 탐색하여 효과적으로 제거하기 위해, 순차적인 실험설계 절차를 제안하였다. Hill과 Hunter(1974)는 회귀모형의 모수들 중 일부만, 가능한 한 정확하게 예측되어야 할 필요가 있는 경우에, 그 모수들의 일부분을 정확하게 예측하기 위하여 베이지안 방법을 사용하였다. Atkinson과 Cox(1974)는 몇몇의 회귀모형들을 고려하여 회귀모형들을 판별하기 위한 순차적 실험설계 절차를 제시하였고, 잠재적으로 정확하다고 여겨지는 여러 회귀모형들에 대해 가장 평균된 D-최적값들을 최적화시켰다. Atkinson과 Fedorov(1975)는 두 개의 회귀모형들 중 하나가 맞을 때, 이들 회귀모형들을 판별해내기 위한 순차적 실험설계 절차를 제시하였다. Hill(1978)은 여러 가지 모형 판별 방법들을 재검토하였고, 잠재적인 회귀모형들로 구성된 모집단의 “엔트로피(Entropy)”를 순차적으로 최적화하는 방법에 대하여 논하였다. Cook과 Nachtsheim(1982)은 회귀모형의 정확한 형태가 알려져 있지 않은 경우, 다중 모형 실험설계를 구축하기 위해 누적 분산 최적성 기준(Integrated variance optimality criterion)에 대한 분석적인 방법을 제시하였다. Li와 Nachtsheim(2000)은 Li와 Wu(1997)의 Columnwise-Pairwise(CP) 알고리즘을 확장한 효율적인 교환 알고리즘을 사용하여 회귀 모형들에 둔 감한 요인실험설계들을 구축하였다(Li and Wu, 1997). Heredia-Langner, Montgomery, Carlyle, 그리고 Borrer(2004)는 여러 가지 반응표면 실험설계 문제에 대한 D-최적(또는 D-효율이 매우 높은) 실험설계들을 구축하고자, 유전자 알고리즘과 만족도 함수 최적화 방법을 적용하였다. Borkowski와 Valeroso (2001)는 2차 회귀모형들과 다양한 형태의 축소된 회귀모형들에 대한 여러 개의 표준 실험 설계들과 Computer-generated 실험설계들의 D-효율, G-효율, A-효율, IV-효율들을 비교하였다. 이러한 실험설계 최적성 기준들은 독립 변수들이 3~5개 인 경우에 대하여, 입방형 실험설계 영역과 구체 실험설계 영역들 모두에서 평가되었다. 본 논문에서는, 유전자 알고리즘(Genetic Algorithm)과 만족도 함수(Desirability Function) 최적화 방법을 사용하여 입방형 영역에서의 현재 고려하고 있는 모든 회귀모형들에 대해 안정적인 예측분산을 제공하는 Model-robust 실험설계들을 구축하고자 한다. 또한 입방형 영역에서의 Model-robust 실험설계와 전형적인 2차 회귀모형에 대한 면심입방체 실험설계 간의 예측 분산을 비교할 것이다. 다음 섹션에서는 유전자 알고리즘과 만족도 함수 최적화 방법의 일반적인 성질들에 대하여 간략히 논의하고자 한다.

3. 유전자 알고리즘과 만족도 함수 최적화 방법

3.1 유전자 알고리즘 (Genetic Algorithm)

전수 탐색 알고리즘(Exhaustive search algorithm)의 적용은 대규모 조합최적화 문제들의 최적해 혹은 최적에 근사한 해를 찾아내는 데 있어 계산적으로 매우 비능률적이다. 따라서 어

려운 조합최적화 문제들을 다룰 때, 발견적 방법들이 종종 사용되는데, 몇 가지 조건 하에서 이러한 발견적 방법들은 때로는 최적해를 도출해 주지 못하기도 하지만 대부분의 경우에 최적이나 최적에 가까운 해를 찾게끔 해준다. 이처럼 다양하고 유용한 발견적 방법들이 어려운 조합최적화 문제들을 다루기 위해 개발되어 왔고, 유전자 알고리즘 역시 그 가운데 하나로, 찾고자 하는 해의 부분적 향상에 근간한 발견적 방법 중의 하나이다. 유전자 알고리즘은 생물학적 진화와 유전법칙을 바탕으로 한 방법으로, 개체집단이 어떠한 변화를 겪는 과정 속에 나타나는, 각각의 개체들이 생존하는 현상을 모방하여 설계된 알고리즘이다. 이러한 유전자 알고리즘은 새로운 개체들을 생성하기 위해서, 유전자형(또는 유전인자)이라고 불리는 문자열 형태의 구조를 수정하는 작업인 유전자 연산(Genetic Operation)을 사용한다. 유전자 알고리즘이 어떠한 최적화 문제에 적용되었을 때, 최적해는 그 문제에 대해 미리 정해진 숫자 척도로 표현이 되는데, 이 척도를 적합도 함수(Fitness measure function) 또는 평가함수(Evaluation function)라고 부른다. 또한 유전자 알고리즘은 반복적으로 개체들의 집단을 생성하고 모든 개체들을 평가한다. 집단내의 모든 개체들이 평가되면, 그들의 적합도는 다음 세대들을 선택하는 데 기준이 되는데, 이 적합도 기준을 바탕으로 확률적 또는 결정적인 방법으로 새로운 개체들이 선택된다. 다양한 구성계획 및 적합도 함수와 함께, 유전자 알고리즘은 다양한 유형의 최적화 문제들에 적용될 수 있다(Forrest, 1993; Mitchell, 1996). 본 논문에서는 Cloning, Crossover, Mutation의 세 가지 형태의 유전자 연산 방법을 사용하고, 해를 표현하는 Encoding 방법으로 Real-value encoding 방법을 사용한다. 유전자 알고리즘의 효율성은 해의 이웃들을 탐색하는 방법뿐 만 아니라, 해와 해의 이웃들의 설계에도 영향을 받는다. 따라서 본 연구에서는 모든 회귀모형에 대해 대부분 높은 G-효율을 지닌 실험설계를 신속히 찾을 확률을 높이고, 부분 최적화된(Local optimal) 해에서 빨리 빠져 나와 더 나은 해로 이동하도록 하기 위해, 여러 개의 다른 개체집단을 생성하여 여러 번의 유전자 연산을 반복하였고, 더 나은 해를 찾기 위한 탐색방법과 유전자 모수(Genetic parameter)들을 선택하였다. 그러나 유전자 알고리즘을 복잡한 조합최적화 문제에 적용하였을 때, 부분 최적화된 해를 제시해 줄 가능성이 있으므로, 항상 최적의 해를 제시한다고 장담할 수는 없다. 하지만, 이러한 문제가 있음에도 불구하고, 유전자 알고리즘은 가능한 해 공간을 확률적인 탐색 방법과 해의 체계적인 향상을 보장하는 직접적인 탐색 방법을 적절하게 사용하므로, 다양한 실제 문제들을 해결하는 데에 널리 사용되고 있다. 본 논문에서 최적 실험설계 기준으로 반응치에 대한 적도화된 분산의 최대값을 최소화 시키는데 관련된 G-최적성 기준을 고려한다. 독립변수들의 개수 (k)가 3, 4, 5일 때, 각 모형에서의 최적 실험설계(Model-best experimental design)와 Model-robust 실험설계의 G-효율성에 대해 제 4장에서 설명한다.

3.2 만족도 함수 최적화 방법(Desirability Function(DF) Optimization Method)

만족도 함수(Desirability function)는 다양한 형태의 다중 반응 최적화 문제(Multiple response optimization problem)들에 널리 사용된다. 이 접근법의 기본적인 원리는 고려하고 있는 여러 목적함수들의 각각의 속성들을 동시에 최적화하기 위하여, 목적함수들을 하나의 함수 형태로 변환하는 것이다. i 번째의 특정 반응변수 y_i 를 최대로 할지, 최소로 할지, 아니면 어떠한 목표 값에 근접한 값을 찾을 지의 여부에 따라서 다른 만족도 함수 $d_i(y_i)$ 가 사용된다. 다중 반응 최적화 문제를 풀기 위해 사용되는 유용한 만족도 함수들은 Derringer와 Suich(1980)에 의해 자세히 제시되고 있다. 만족도 함수에서 사용되는 L_i, U_i, T_i, s 는 각각 반응 값 y_i 의 하위 값, 상위 값, $L_i \leq T_i \leq U_i$ 의 조건에서 만족시킬 목표 값, 그리고 중요도를 나타내는 가중치를 나타낸다. 만약 반응변수를 최소화 하고자 한다면 다음과 같은 만족도 함수를 사용할 수 있다.

$$d_i(\hat{y}_i) = \begin{cases} 1 & \text{if } \hat{y}_i < T_i \\ \left(\frac{\hat{y}_i - U_i}{T_i - U_i} \right)^s & \text{if } T_i \leq \hat{y}_i \leq U_i \\ 0 & \text{if } \hat{y}_i > U_i \end{cases} \quad (9)$$

각 목적함수의 반응치 y_i 에 대하여, 만족도 함수 $d_i(y_i)$ 는 y_i 의 가능한 값으로 0과 1사이의 값을 할당한다. 예를 들어, 여기서 $d_i(y_i) = 0$ 인 경우는 y_i 의 값이 전혀 바람직하지 않음을 나타내며, $d_i(y_i) = 1$ 인 경우는 반응치 y_i 가 이상적인 값을 가짐을 나타낸다. 또한, 일반적으로 전체 시스템이나 전체 프로세스의 만족도는 개별적인 만족도 함수를 이용하여 다음과 같은 기하평균의 형태로 표현된다. 여기서 k 는 반응변수의 수를 나타낸다.

$$D = (d_1(y_1) \cdot d_2(y_2) \cdot d_3(y_3) \cdots d_k(y_k))^{1/k} \quad (10)$$

만족도 함수의 적용방법과 전반적인 만족도 함수들에 대한 자세한 내용은 Harrington(1965)에서 찾을 수 있다.

4. 입방형 영역에서의 Model-Robust G-최적 실험설계의 구축

다양한 설계영역에 대해 2차 회귀모형을 기준으로 구축된 많은 G-효율 실험설계들이 분석되었다. 입방형 영역들과 2차 회귀모형 기준을 바탕으로 한 실험설계들은 Mitchell과 Bayne의 DETMAX 알고리즘(Mitchell and Bayne, 1978; Mitchell, 1974)과 Hoke(1974)의 알고리즘을 통해 구축되었다. 독립변수의 개수 $k=4$ 일 때, DETMAX 알고리즘에 의해 구축된 실험설계는 가

장 효율적이라고 알려져 있다. 독립변수의 개수가 $k=3$ 과 $k=5$ 인 경우에는 Hoke에 의해 제안된 실험설계들이 G-효율성 기준 측면에서 가장 뛰어나다고 알려져 있다. 본 연구에서 제안된 입방형 Model-robust G-효율 실험설계들은 유전자 알고리즘과 만족도 함수 최적화 방법을 이용해 생성된 것으로서, 독립변수의 개수가 $k=3, 4$ 그리고 5인 경우의 Model-robust G-효율 실험설계가 <표 1>에 나타나 있다. 2차 회귀모형에 대해, 현재 최고의 G-효율 입방형 실험설계라고 알려진 Hoke 실험설계와 MB-D 실험설계의 효율성과 Model-robust 실험설계의 G-효율을 비교하기 위하여, 같은 수의 설계 점들을 고려하였으며, 관측 오차들은 독립적으로 평균이 0이고 상수의 분산을 가지는 정규 분포를 따른다고 가정한다.

<표 2>는 독립변수의 개수가 3, 4, 5인 경우에, Hoke 실험설계와 MB-D 실험설계들, 그리고 입방형 Model-robust G-효율 실험설계들의 척도화된 최대 예측 분산(Maximum Scaled Prediction Variances: Maximum SPV)과 G-효율에 대한 요약이다. 예를 들어, $k=3$ 인 경우를 보면, 식 (3)과 같은 2차 회귀모형을 가정한 후 구한 Hoke 실험설계는 1차 선형회귀모형, 2-요인 선형교호작용모형, 그리고 2차 회귀모형 각각에 대해 0.7588, 0.6396, 그리고 0.8304와 같은 G-효율을 갖는다. 때때로, 1차 선형회귀모형에 대해 가장 효율적인 실험설계들이 다른 형태의 회귀모형들에 대해 매우 낮은 G-효율을 갖거나 효율을 계산할 수 없는 경우가 발생하게 되는데, 그 이유는 구축된 실험설계들이 Singularity인 경우, 즉 $(X'X)^{-1}$ 가 존재하지 않기 때문이다. <표 2>에서 우측에 있는 표는 유전자 알고리즘과 만족도 함수 최적화 방법에 의해 생성된 Model-robust 실험설계의 척도화된 최대 예측 분산과 G-효율들을 나타낸다. $k=5$ 인 경우, Model-robust 실험설계는 각각 0.7267, 0.6471, 0.7561의 G-효율들을 갖는다. 여기서 보듯이 척도화된 최대 예측 분산(Maximum SPV) 값과 G-효율의 개선 정도가 매우 적다고 판단할 수 있다. 그러나 통계학적인 측면에서 보면 이러한 작은 차이가 때로는 심각한 문제를 발생시키는 경우가 있기 때문에 이러한 작은 개선도 의미 있는 결과라고 볼 수 있다. 추가적으로 $k=3, 4, 5$ 인 경우에 관하여, Model-robust 실험설계들과 Hoke 실험설계들, 그리고 MB-D 실험설계의 척도화된 최대 예측 분산은 <그림 1> ~ <그림 3>에 나타나 있다.

독립변수의 개수가 3인 경우에, 표2와 그림1에서 보여 지는 것처럼, 2차 회귀모형에 대해 효율적인 Hoke 실험설계가 1차 선형회귀모형과 2-요인 선형교호작용모형에 적용되면, 이 실험설계는 1차 선형회귀모형과 2-요인 선형교호작용모형에 대해서는 다소 낮은 G-효율값을 가진다. 하지만 $k=3$ 인 경우의 Model-robust 실험설계는 Hoke 실험설계에 비해 모든 회귀모형들에 대해 대부분 높은 G-효율들을 갖는다. 독립변수의 개수가 4와 5인 두 경우 모두에 대해서는, MB-D 실험설계와 Hoke 실험설계들이 2차 회귀모형에 대하여 높은 G-효율을 갖는 반면, 1차 선형회귀모형과 2-요인 선형교호작용모형들에 대해서는 낮은 G-효율을 가진다. 하지만 $k=4$ 와 $k=5$ 인 경우

표 1. $k = 3, 4, 5$ 인 경우에 대한 입방형 Model-robust G-효율 실험설

$k = 3, n = 14$			$k = 4, n = 24$				$k = 5, n = 26$				
X_1	X_2	X_3	X_1	X_2	X_3	X_4	X_1	X_2	X_3	X_4	X_5
-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
-1	-1	1	-1	-1	-1	1	1	1	-1	-1	-1
-1	0	0	-1	-1	1	-1	1	-1	1	-1	-1
-1	1	-1	-1	-1	1	1	-1	1	1	-1	-1
-1	1	1	-1	0	-1	0	0	0	0	0	-1
0	-1	0	-1	0	0	-1	1	-1	-1	1	-1
0	0	-1	-1	1	-1	-1	-1	1	-1	1	-1
0	0	1	-1	1	-1	1	-1	-1	1	1	-1
0	1	0	-1	1	1	-1	1	1	1	1	-1
1	-1	-1	-1	1	1	1	0	1	0	-1	0
1	-1	1	0	-1	0	0	0	0	-1	0	0
1	0	0	0	0	-1	-1	0	-1	0	0	0
1	1	-1	0	0	0	1	-1	0	0	0	0
1	1	1	0	0	1	0	1	0	0	0	0
			0	1	0	0	0	1	0	0	0
			1	-1	-1	-1	0	0	1	1	0
			1	-1	-1	1	1	-1	-1	-1	1
			1	-1	1	-1	-1	1	-1	-1	1
			1	-1	1	1	-1	-1	1	-1	1
			1	0	0	0	1	1	1	-1	1
			1	1	-1	-1	0	0	1	0	1
			1	1	-1	1	-1	-1	-1	1	1
			1	1	1	-1	1	1	-1	1	1
			1	1	1	1	0	0	0	1	1
							1	-1	1	1	1
							-1	1	1	1	1

표 2. 2차 회귀모형에 대한 Hoke 실험설계, MB-D 실험설계, Model-robust 실험설계들의 척도화된 최대 예측 분산(Maximum Scaled Prediction Variances)과 G-효율

<i>SPA and G-efficiency of Second-Order Designs</i>			<i>SPA and G-efficiency of Model-Robust Designs</i>		
<i>Hoke Design : k = 3, n = 14</i>			<i>Model-Robust Design : k = 3, n = 14</i>		
Candidate Model	SPV	G-eff	Candidate Model	SPV	G-eff
Linear	5.2715	0.7588	Linear	5.2000	0.7692
2-Factor Interaction	10.9440	0.6396	2-Factor Interaction	10.4500	0.6699
Second Order(SO)	12.0420	0.8304	Second Order(SO)	11.2000	0.8929

<i>MB-D Design : k = 4, n = 24</i>			<i>Model-Robust Design : k = 4, n = 24</i>		
Candidate Model	SPV	G-eff	Candidate Model	SPV	G-eff
Linear	6.6890	0.7475	Linear	6.3739	0.7844
2-Factor Interaction	18.8120	0.5847	2-Factor Interaction	15.5840	0.7059
Second Order(SO)	20.8130	0.7207	Second Order(SO)	19.5980	0.7654

<i>Hoke Design : k = 5, n = 26</i>			<i>Model-Robust Design : k = 5, n = 26</i>		
Candidate Model	SPV	G-eff	Candidate Model	SPV	G-eff
Linear	8.3469	0.7188	Linear	8.2570	0.7267
2-Factor Interaction	24.8440	0.6440	2-Factor Interaction	24.7250	0.6471
Second Order(SO)	29.3020	0.7167	Second Order(SO)	27.7760	0.7561

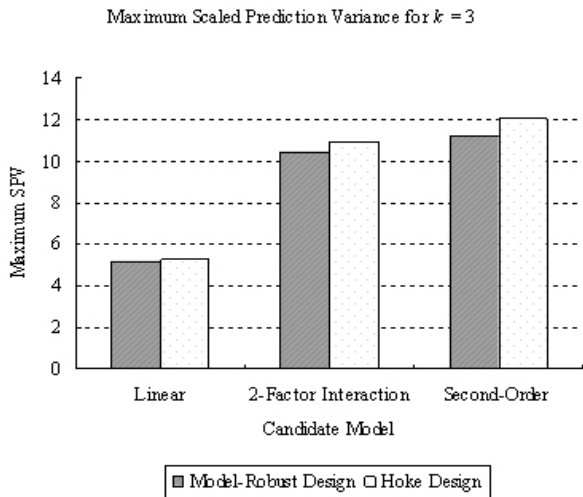


그림 1. $k = 3$ 에 대한 Model-robust 실험설계와 Hoke 실험설계의 척도화된 최대 예측 분산(Maximum Scaled Prediction Variances)

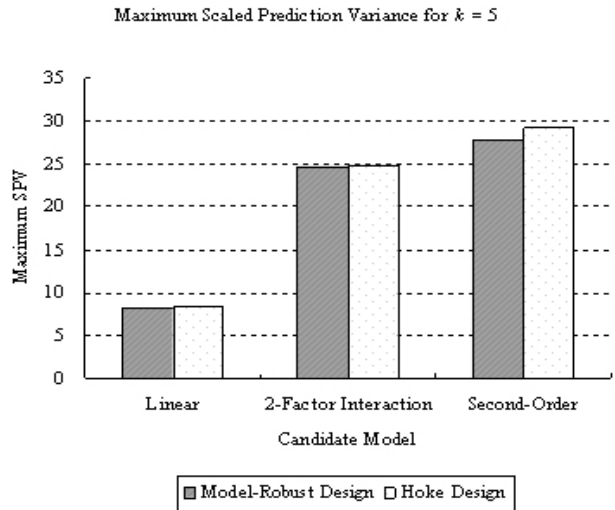


그림 3. $k = 5$ 에 대한 Model-robust 실험설계와 Hoke 실험설계의 척도화된 최대 예측 분산(Maximum Scaled Prediction Variances)

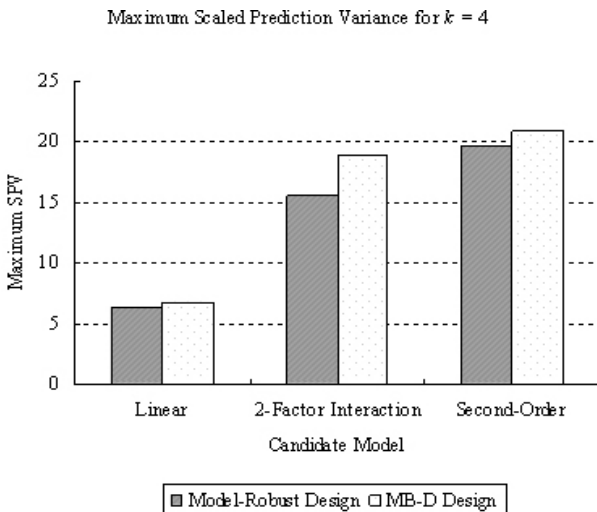


그림 2. $k = 4$ 에 대한 Model-robust 실험설계와 MB-D 실험설계의 척도화된 최대 예측 분산(Maximum Scaled Prediction Variances)

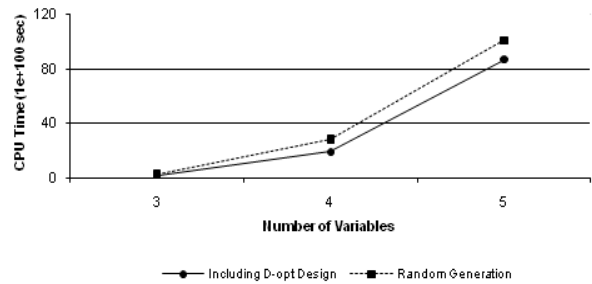


그림 4. 초기 개체집단의 D-최적 실험설계를 포함시킨 실험과 초기 개체집단의 멤버로 랜덤하게 생성된 실험설계들을 포함한 실험 간의 CPU 시간 비교

의 Model-robust 실험설계들은 $k = 3$ 일 때의 Model-robust 실험설계처럼, 모든 회귀모형들에 대하여 모두 높은 G-효율을 갖는다. 특히, 본 연구에서 제시된 독립변수의 개수가 3과 4인 경우에 대한 Model-robust 실험설계들은 Park *et al.*(2005)의 연구에서 제시된 G-효율 실험설계들과 동일하지만, 이 연구에서 발견된 $k = 5$ 일 때의 Model-robust 실험설계는 Hoke 실험설계와 Park *et al.*(2005)의 연구에서 제시된 GA-실험설계보다 훨씬 더 효율적임을 알 수 있다.

또한, 독립변수 개수가 5인 경우의 Model-robust 실험설계에 관한 척도화된 평균 예측 분산(Average Scaled Prediction Variances)은 경계 영역을 제외하고, Hoke 실험설계나 Park *et al.*(2005)의 연구에서 제시된 GA-실험설계보다 더 낮게 나타났다(Park

et al., 2005). 따라서, 독립변수 개수가 5인 경우에 대한 Model-robust 실험설계는 또한 I-효율 실험설계라고 할 수 있다. 이것은 유전자 알고리즘에 의해 구해진 Model-robust 실험설계들이 기존에 잘 알려진 작은 실험점들로 구성된 실험설계들보다 모든 회귀모형들에 대하여 대부분 더 높은 G-효율을 가짐을 말해준다. 결과적으로, 유전자 알고리즘을 통해 구축된 모든 Model-robust 실험설계들이 일반적으로, Hoke 실험설계와 MB-D 실험설계와 같은 기존의 G-효율이 높은 실험설계들과 비교했을 때, 고려하고 있는 회귀모형들에 대하여 높은 G-효율을 갖는다고 결론지을 수 있다.

추가적으로 본 연구에서는 G-효율이 높은 Model-robust 실험설계들을 찾거나 적용한 유전자 알고리즘의 계산 효율성을 검사해보고자 랜덤하게 생성된 개체집단을 초기 집단으로 설정한 경우와 개체집단에 D-최적 실험설계를 포함시킨 개체집단을 초기 집단으로 설정한 경우를 고려하였다. <그림 4>는 D-최적 실험설계들을 초기 개체집단에 포함시켰을 경우의 계산 효율성이 초기 개체 집단에 랜덤하게 생성된 실험설계들을 포함시켰을 때보다

현재까지 높음을 알 수 있다. 그러므로, 초기 개체집단의 구성원으로서 D-최적 혹은 D-효율적인 실험설계를 포함시키는 것은 유전자 알고리즘의 계산상 효율성을 높일 수 있다고 할 수 있다.

5. 결론

본 연구에서 독립변수의 개수가 3, 4 그리고 5인 경우와 입방형 영역에 대해 3가지 회귀모형들을 고려하여 반응치의 안정적인 예측 분산을 가지는 Model-robust G-효율 실험설계들을 구축하고자 유전자 알고리즘과 만족도 함수 최적화 방법을 사용하였다. Hoke 실험설계와 MB-D 실험설계들과 같이 현재 최적이라고 알려진 2차 회귀모형을 기준으로 한 실험설계들은 모든 회귀모형들에 대해 항상 효율적이지는 않음으로 나타났지만, 유전자 알고리즘과 만족도 함수 최적화 방법에 의해 구해진 Model-robust G-효율 실험설계들은 매우 효율적이라고 나타났다. 이 연구에서 제안된 독립변수 $k = 3, 4$ 인 경우에 대한 Model-robust G-효율 실험설계들은 Park *et al.*(2005)의 연구에서 제시된 G-효율 실험설계들과 동일하다. 그러나 $k = 5$ 인 경우의 Model-robust G-효율 실험설계는 Hoke 실험설계나 Park *et al.*(2005)의 연구에서 제시된 GA-실험설계보다 훨씬 더 효율적으로 나타났다(Park *et al.*, 2005). 또한, 독립변수의 개수가 5인 경우, Model-robust 실험설계에 대한 척도화된 평균 예측 분산(Average Scaled Prediction Variances)이 Hoke 실험설계나 Park *et al.*(2005)의 연구에서 제시된 GA-실험설계에 비해 낮은 것으로 나타났으며, 따라서 이 Model-robust 실험설계는 I-효율 실험설계가 될 수 있음으로 나타났다. 유전자 알고리즘이 순차적인 방식으로 실험설계를 구축할 수 없음에도 불구하고, 여러 회귀모형들이 고려되었을 경우에, 효율적으로 최적의 혹은 최적에 가까운 해 즉, Model-robust G-효율 실험설계를 찾는데 있어 매우 유용하다는 것을 본 연구를 통해 알 수 있었다. 결론적으로, 만약 실험자가 사전에 회귀모형의 형태에 대한 어떠한 정보도 없다면, Model-robust G-효율 실험설계를 사용할 것을 제안한다. 추가적인 연구로써, 특이한 형태의 회귀모형들이 주어질 경우, 예를 들어, 주요 항과 순수 2차 항만이 포함된 축소된 2차 회귀모형이 고려될 가능성이 있는 경우에, Model-robust G-효율 실험설계들을 구축하는 것과 관심 있는 회귀모형들에 다른 가중치를 주거나, 실험의 수를 다르게 배정하였을 경우의 G-효율 실험설계를 구축하는 것도 필요하다.

참고문헌

- Atkinson, A. C. and Cox, D. R. (1974), Planning Experiments for Discriminating between Models, *Journal of the Royal Statistical Society B*, 36, 321-348.
- Atkinson, A. C. and Fedorov, V. V. (1975), The Design of Experiments for Discriminating Between Two Rival Models, *Biometrika*, 62, 27-70.
- Borkowski, J. J. (2003), Using a Genetic Algorithm to Generate Small Exact Response Surface Designs, *Journal of Probability and Statistical Science*, 1, 65-88.
- Borkowski, J. J. and Valeroso, E. S. (2001), Comparison of Design Optimality Criteria of Reduced Models for Response Surface Designs in the Hypercube, *Technometrics*, 43, 468-477.
- Cook, R. D. and Nachtsheim, C. J. (1980), A Comparison of Algorithms for Constructing Exact D-Optimal Designs, *Technometrics*, 3, 315-324.
- Cook, R. D. and Nachtsheim, C. J. (1982), Model Robust, Linear-Optimal Designs, *Technometrics*, 24, 49-54.
- Derringer, G. and Suich, R. (1980), Simultaneous Optimization of Several Response Variables, *Journal of Quality Technology*, 12, 214-219.
- Fedorov, V. V. (1972), *Theory of Optimal Experiments*. Academic Press.
- Forrest, S. (1993), Genetic Algorithms: Principles of Natural Selection Applied to Computation, *Science*, 261, 872-878.
- Haines, L. M. (1987), The Application of the Annealing Algorithm to the Construction of Exact Optimal Designs for Linear Regression Models, *Technometrics*, 37, 439-447.
- Hamada, M., Martz, H. F., Reese, C. S., and Wilson, A. G. (2001), Finding Near-Optimal Bayesian Experimental Designs via Genetic Algorithms, *The American Statistician*, 55, 175-181.
- Harrington, E. C. Jr. (1965), The Desirability Function. *Industrial Quality Control*, 21, 494-498.
- Heredia-Langner, A., Carlyle, W. M., Montgomery, D. C., Borror, C. M., and Runger, G. C. (2003), Genetic Algorithms for the Construction of D-Optimal Designs, *Journal of Quality Technology*, 35, 28-46.
- Heredia-Langner, A., Montgomery, D. C., Carlyle, W. M., and Borror, C. M. (2004), Model-Robust Optimal Designs: A Genetic Algorithm Approach, *Journal of Quality Technology*, 36, 263-279.
- Hill, P. D. H. (1978), A Review of Experimental Design Procedures for Regression Model Discrimination, *Technometrics*, 1, 15-21.
- Hill, W. G. and Reiner, A. M. (1965), Designs for Discriminating Between Two Rival Models, *Technometrics*, 7, 307-323.
- Hill, W. J. and Hunter, W. G. (1974), Design of Experiments for Subsets of Parameters, *Technometrics*, 16, 425-434.
- Hoke, A. T. (1974), Economical Second-Order Designs Based on Irregular Fractions of the $3n$ Factorial, *Technometrics*, 16, 375-384.
- Kiefer, J. (1959), Optimum experimental designs, *Journal of the Royal Statistical Society B*, 21, 272-319.
- Kiefer, J. (1961), Optimum Designs in Regression Problems, *Annals of Mathematical Statistics*, 32, 298-325.
- Li, C. and Nachtsheim, C. J. (2000), Model-Robust Factorial Designs, *Technometrics*, 4, 345-352.
- Li, W. and Wu, C. F. J. (1997), Columnwise-Pairwise Algorithms with Application to the Construction of Supersaturated Designs, *Technometrics*, 39, 171-179.
- Meyer, R. K. and Nachtsheim, C. J. (1995), The Coordinate-Exchange Algorithm for Constructing Exact Optimal Experimental Designs, *Technometrics*, 37, 60-67.
- Mitchell, M. (1996), *An Introduction to Genetic Algorithms*, MIT Press.
- Mitchell, T. J. (1974), An Algorithm for the Construction of D-optimal Experimental Designs, *Technometrics*, 16, 203-210.
- Mitchell, T. J. and Bayne, C. K. (1978), D-Optimal Fractions of Three-Level Factorial Designs, *Technometrics*, 20, 369-380.
- Myers, R. H. and Montgomery, D. C. (2002), *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, Wiley.
- Park, Y., Richardson, D. E., Montgomery, D. C., Ozol-Godfrey, A., Borror, C. M., and Anderson-Cook, C. M. (2005), Prediction Variance Properties of Second-Order Designs for Cuboidal Regions, *Journal of Quality Technology*, 37, 253-266.
- Wald, A. (1943), On the Efficient Design of Statistical Investigations, *Annals of Mathematical Statistics*, 14, 134-140.

Welch, W. J. (1982), Branch and Bound Search for Experimental Designs Based on D-Optimality and Other Criteria, *Technometrics*, 1, 41-48.

Wynn, H. P. (1970), The Sequential Generation of D-optimum Experimental Designs, *Annals of Mathematical Statistics*, 41, 1655-1664.



박유진

한양대학교 산업공학과에서 학사 및 석사학위를 취득하고 Arizona State University에서 산업공학 박사학위를 취득

현재 : 중앙대학교 사회과학대학 경영학과 교수
관심분야 : 실험계획법, 반응표면분석법, 최적화, 품질경영 관련



이윤주

중앙대학교 통계학과에서 학사학위를 취득하였고, 동대학교 통계학과 석사과정에 재학 중이다.

관심분야 : 실험계획법, 반응표면분석법, 시계열분석법이다.