The Korean Communications in Statistics **Vol. 14** No. 3, 2007, pp. 531–540

Hybrid Approach When Multiple Objectives Exist*

YoungIl Kim[1) and YongBin Lim[2)

Abstract

When multiple objectives exist, there are three approaches exist. These are maximin design, compound design, and constrained design. Still, each of three design criteria has its own strength and weakness. In this paper Hybrid approach is suggested when multiple design objectives exist, which is a combination of maximin and constrained design. Sometimes experimenter has several objectives, but he/she has only one or two primary objectives, others less important. A new approach should be useful under this condition. The genetic algorithm is used for few examples. It has been proven to be a very useful technique for this complex situation. Conclusion follows.

Keywords: Optimal design; criteria; compound design; maximin design; constrained design; hybrid approach; genetic algorithm.

1. 서론

최적실험계획법은 Kiefer (1959), Fedorov (1972)에 의해 이론적인 초석을 다진 통계학의 분야이다. 제시된 각 실험계획기준들은 실험계획기준 선정에 뛰어난 수학적인 배경을 가지고 있다. Fedorov (1972), Silvey (1980), Pukelsheim (1993)등의 책들은 이러한 내용을 담고 있다. 그럼에도 불구하고 최적실험계획법의 이론은 주어진 모형과 가정에 의존하는 특성 때문에 실질적으로 현장에서 적용하기 힘들다는 많은 비평을 받아왔다. Box와 Draper (1975)의 논문은 실험계획이 성공리에 수행되기 위한 14가지 실험계획기준을 제안하였다. 따라서 실험계획의 복합적인 성격을 만족하는 실험계획기준을 제시하는 방법론이 많이 제안되었는데, 초창기에는 Stigler (1971), Läuter (1974)등이 이에 부응하는 실험계획기준들을 제시하기 시작하였으며 80년 이후로는 Studden (1982), Cook과 Fedorov (1995)와 Cook과 Wong (1994)에 의해 이론적인 틀이 마련되었다. 이를 Wong (1999)이 체계적으로 정리하였다. 다중목적의 실험을 위해 제안되는 기준을 설명하기 위한 기본적인 표현방법은 다음과 같이 요약될 수 있다.

* Kim's research was supported by the 2006 ChungAng University Research Fund.
1) Professor, Department of Information System, ChungAng University, KyungGi-Do 456-756, Korea. E-mail : yik01@cau.ac.kr
2) Professor, Department of Statistics, Ewha Women's University, Seoul 120-750, Korea. Correspondence : yblim@ewha.ac.kr

사용자가 정의하는 R^k 공간의 어떤 콤팩트 (compact) 부분집합 X에서 정의된 m개의 서로 독립인 연속회귀함수 (regression function)로 구성된, $f_m^T(x) = (f_1(x), f_2(x), \ldots, f_m(x))$가 있다고 가정하자. X안에 있는 모든 x에 대하여 반응변수 $y(x)$는 다음과 같은 통계적인 모형에 의해 관측이 된다고 하자.

$$y(x) = f^T(x)\beta + \epsilon = \sum_{i=1}^{m} f_i(x)\beta_i + \epsilon. \tag{1.1}$$

여기서 미지의 회귀계수들로 구성된 $\beta^T = (\beta_1, \ldots, \beta_m)$는 추정하여야할 모수이고 오차 ϵ는 서로 독립이고 기댓값이 0, 분산이 σ^2인 분포를 따른다. 앞으로는 일반성의 손실 없이 σ^2는 1로 가정한다. 실험계획의 문제는 실험계획 영역 X안에 있는 s개의 점, $x_i, i = 1, \ldots, s$의 유한집합에 확률을 부여하는 확률 질량함수 $\xi(x)$로 기술될 수 있다. 표본의 크기 n이 작은 실험계획에서는 $n\xi(x_i)$가 정수여야 하는 제약을 지니지만 본 연구에서는 이러한 제약이 없는 실험계획인 근사 실험계획 (approximate design)만 고려하여 본다. 참고로 x_i는 실험계획 ξ의 받힘점 (supporting point) 이라 부른다. 모형 (1.1)에 대한 임의의 실험계획 ξ의 정보행렬 (information matrix)는 식 (1.2)와 같이 정의된다.

$$M(\xi) = \int_{\mathcal{X}} f(x)f^T(x)\xi d(x). \tag{1.2}$$

여기서는 정보행렬이 양정치 행렬 (positive-definite matrix)인 경우만 고려하여 본다. 최적 실험계획을 구하기 위해서는 $(X, f(x))$와 함께 실험자의 실험계획기준을 반영하는 실험계획의 볼록 (convex) 함수인 Φ를 결정하여야 하는데 흔히 많이 쓰이는 기준에는 D-, E- 그리고 A-최적 (optimality) 기준이 있다. D-최적 기준은 모수 β에 대한 추정에 관련된 기준으로 $\Phi(\xi) = -\ln|M(\xi)|$이다. E-최적 기준은 $\Phi(\xi) = \max_{c^T c=1} c^T M(\xi)^{-1} c$이다. 다른 흥미로운 기준으로는 실험계획 영역 X에 있는 임의의 점 x에서 예측치 $\hat{y}(x)$의 분산의 최대값을 최소화하는 G-최적 기준이 있다. 즉, $\bar{d}(\xi) = \max_{x \in \mathcal{X}} d(x, \xi)$을 최소화하는 기준이다. 여기서 $d(x, \xi) = f^T(x)M(\xi)^{-1}f(x)$이다. 이 기준은 오차가 등분산인 경우에는 D-최적 기준과 동격이다. 또한 $\Phi(\xi) = trLM(\xi)^{-1}$으로 정의되는 L-최적 기준이 있다. 여기서 L은 비음정치 (non-negative definite) 행렬인데 L이 단위행렬인 경우에는 A-최적 기준과 동일하게 된다. 주어진 점이 z이고 $L = f(z)f^T(z)$이면 주어진 점 z에서 예측치 $\hat{y}(z)$의 분산을 최소화하는 기준이 된다. 이러한 모든 실험계획기준은 정보행렬 공간 (space of information matrices)에서 볼록함수이다. 이에 관련된 최근 문헌으로는 Pukelsheim (1993)을 들 수 있다.

위에서 언급한 실험계획기준들은 모두 회귀함수 $f(x)$가 알려져 있다고 가정하는데 실제 문제의 적용에 있어서는 그렇지 않은 경우가 많다. 또한 회귀모형에 대한 불확실성 뿐 아니라 실험계획기준도 문제가 된다. D-최적 기준이 많이 쓰이나 같은 모형에 대해서도 여러 개의 실험계획기준의 동시 최적화를 시도하는 경우가 있다. Wong (1995)은 차수가 낮은 다항회귀모형에 대해 D-최적 기준과 A-최적 기준을 결합하여 모수추정을 시도하였다. 그리고 실험자의 관점에 따르면, 모형에 포함된 모수들이 모두 동일한 중요성을 가지는 것이 아니다. 이러한 상황 하에서 일괄적으로 D-최적 기준을 부여한다

는 것은 바람직하지 않다. 따라서 여러 목적을 가지고 있는 실험계획기준의 필요성은 실험자의 관점에서 자연스럽게 대두가 되고, 주어진 실험계획에 대해서 여러 실험 기준에 대한 성능을 평가하기 위한 지표가 필요하게 된다. 이 지표를 활용하여 실험계획기준이 주어진 경우에 임의의 실험계획이 최적실험계획에 대해 가지는 효율성 (efficiency)을 정의하고자 한다.

Fedorov (1972), Silvey (1980), Pukelsheim (1993)등에 소개된 바와 같이 주어진 모형 f하에서 임의의 실험계획 ξ가 Φ-최적실험계획, ξ^*_Φ에 대해 가지는 Φ-효율성 (efficiency)은 $e_f(\xi, \xi^*_\Phi) = \Phi_f(\xi^*_\Phi)/\Phi_f(\xi)$로 정의된다. 만약 정의된 모형이 차수가 m인 다항회귀모형인 경우, ξ^*_Φ이 D-최적 실험계획이라면 임의의 실험계획 ξ의 D와 G-효율성은 각각 $\{|M_m(\xi)/|M(_m(\xi^*)|\}^{1/(m+1)}$과 $(m+1)/\bar{d}_m(\xi)$이다. 여기서 효율성은 실험의 크기와 관련해서 다음과 같이 해석할 수 있다. 어떤 실험계획의 효율성이 0.5라면 최적실험계획과 같은 효율성을 가지기 위해서는 실험계획의 효율성의 역수인 2회의 실험이 반복되어야 한다는 의미이다.

또한 모형의 불확실성에 대비해 $f(x)$를 $f(x) = (f_1^T(x), f_2^T(x))$로 분해하여 보자. 예를 들어 실험자가 차수 r의 다항회귀모형에 대해 확신을 하고 있다 하더라도 모형이 r보다 차수가 큰 $m(>r)$인 다항회귀모형일 가능성에도 대비를 하여야 한다면 실험자는 차수가 r인 회귀모형만을 염두에 둔 실험계획기준은 고려하지 않을 것이다. 실험자는 차수가 r인 회귀모형은 물론, 차수가 m인 회귀모형에 대해서도 좋은 모수 추정 값을 제공하는 실험계획을 구하고자 할 것이다. 이 경우는 $f_1^T(x) = (1, x, \ldots, x^r), f_2^T(x) = (x^{r+1}, \ldots, x^m)$ 그리고 $\beta^T = (\beta_1^T, \beta_2^T)$이다. 다시 언급하면 바람직한 실험계획은 β_1뿐 아니라 모형이 확장이 될 것에 대비해 β_2에 대한 좋은 추정도 제공하여야 할 것이다. 실험계획기준이 Φ-최적이라면 실험계획 ξ이 f_1에 대해 가지는 효율성 $e_\Phi(\xi, \xi^*_{f_1})$과 f에 대해 가지는 효율성 $e_\Phi(\xi, \xi^*_f)$사이에서 실험자는 고민할 것이다. 이러한 문제를 다루는 방법들은 문헌에서 제약조건 실험계획법, 여러 최적기준에 대해 가지는 최소의 효율을 최대화 하는 실험계획을 구하는 방법과 Läuter의 결합실험계획법으로 파악된다. 2절에서는 이와 같은 방법론들에 대한 개요를 알아보고 각각 장단점을 파악한다. 여러 개의 실험계획기준을 동시에 최적화하는 문제 중에서 우리가 관심이 있는 경우는 실험자가 다수의 실험계획기준 $\Phi_1, \Phi_2, \cdots, \Phi_k$들 중에서 한 개의 실험계획기준은 확실한 선호도를 가지고 있어서 바람직한 실험계획으로 이 실험계획기준에 대해서는 최소의 효율을 보장하는 실험계획들 중에서 나머지 실험계획기준들을 동시에 최적화하는 실험계획을 찾는 것이다. 이를 위한 방법론으로 제약조건 실험계획의 장점과 최소최대 방법의 장점을 혼합하여 쓰는 방법인 Hybrid 접근방법을 제안코자 한다. 3절에서는 예제를 통해서 근사 실험계획들 중에서 최소최대 방법에 의한 최적 실험계획과 Hybrid 접근방법에 의한 최적 실험계획을 유전자 알고리즘을 적용하여 구하고, 각각의 효율들을 비교해 본다.

2. 다중실험계획기준 방법

문제의 단순성을 위해 실험자가 두 개의 실험계획기준인 Φ_1, Φ_2를 동시에 최적화하

는 실험계획을 찾는 데에 관심이 있고, 이 두 개의 실험계획기준에 대해 우선순위를 부여할 수 있다고 가정을 하자. Φ_1이 Φ_2에 비해 더 중요하다고 판단된다면 첫 번째 다중 실험계획기준은 $\Phi_1(\xi) \leq c$하에서 $\Phi_2(\xi)$을 최소화 하는 제약실험계획법이 될 것이다. 그러나 실험자가 지정하여야 하는 상수 c는 명기하기가 불편하므로 편의상 제약조건은 위에서 언급한 효율성을 이용하여 $e(\xi, \xi^*_{\Phi_1}) \geq e$로 바꿀 필요가 있다. e는 0과 1사이의 값으로 실험자가 얻고자 하는 최소한의 Φ_1-효율성이며 실험자가 지정하여야 하는 값이다. 요약하면 e의 값을 지정만 하면 부차적인 실험계획기준의 목적함수를 최소화 하는 방법이다. 이러한 제약조건 실험계획법은 Φ_1, Φ_2의 선형결합으로 이루어지는 목적함수, $\Phi(\xi|\lambda) = \lambda\Phi_1(\xi) + (1-\lambda)\Phi_2(\xi), \lambda \in [0,1]$를 최소화하는 Läuter (1974)의 복합최적실험 (compound optimal design), ξ_λ와 동격임이 Cook과 Wong (1994)에 의해 증명되었다. 대략적으로 말하면 적절한 제약조건이 형성되어 있는 실험계획법에서 복합최적실험계획법으로 목적이 구현된다는 것이다. Läuter가 제안한 실험계획기준에 대한 알고리즘은 쉽게 구할 수 있다. 왜냐하면 볼록함수의 결합 역시 볼록하기 때문에 알고리즘 적으로는 새로운 실험계획기준을 구하는 것은 단일 최적 실험계획기준의 해를 구하는 것과 별반 다를 바 없다. 설사 여러 개의 실험계획기준이 있다 하더라도 λ가 주어진 상태라면 어렵지 않게 $\Phi(\xi|\lambda)$를 최소화하는 실험을 구할 수 있다. 따라서 Cook과 Wong (1994)는 효율성 그림 (efficiency plot)이라는 도구를 이용하여 제약실험계획의 구성을 시도하였다. 이는 x축은 λ로 y축은 ξ_λ의 Φ_1, Φ_2의 효율성을 그리는 것이다. 제약조건 실험계획법에서 주어진 상수 e를 y축에서 수평으로 움직여 $e(\xi_\lambda, \xi^*_\Phi)$와 만나면 바로 밑으로 수선을 그어 λ를 확인하고 실험계획법을 구하는 방법이다. 그러나 Läuter가 제안한 방법은 제약조건실험계획법과 동격인 사실은 밝혀졌으나 단독으로 적용하기는 매우 힘든 기준이 된다. 가중치 λ의 선택에 대한 일반적인 기준이 마련되어 있지 않다. 즉, $\lambda = 0.5$로 설정한다 하더라도 두 기준에 대해서 같은 효율성을 얻지 못하기 되기 때문이다. 즉, 복합실험계획기준에서 λ를 실험자가 명기하도록 하는 것은 실험자가 원하지 않는 효율성을 얻을 가능성이 매우 높다. 그렇다고 제약조건 실험계획법이 대안이 될 수는 없다. 또한 실험계획기준이 k개로 확장이 될 경우 제약조건 실험계획법은 식 (2.1)과 같이 표현된다.

$$\begin{aligned} &\min \Phi_k(\xi) \\ \text{subject to} \quad & e(\xi, \xi^*_{\Phi_i}) \geq e_i, i = 1, \ldots, k-1. \end{aligned} \tag{2.1}$$

그러나 실험계획기준이 3개 이상인 다중목적을 염두에 둔 경우에는 예상치 못한 난제가 존재한다. $k-1$개의 실험계획기준이 달성하여야 할 효율성을 실험자가 명시하는 것은 그렇게 쉬운 일이 아니다. 간혹 문헌에서 이러한 경우가 취급되어 왔지만 일반적으로 이론적인 해를 구하는 것은 절대 불가능하기 때문에 수치 해석적으로 해를 구해야 하는데 이 또한 소수의 제약조건하에서만 해를 찾을 수 가 있다. 왜냐하면 제약조건이 2개 이상 늘어나면 모든 제약조건을 만족하는 실험계획이 존재 하지 않는 비타당성 (infeasibility)의 문제가 발생이 된다. Pukelsheim (1993)에 의하면 이론적으로 제약조건식이 2개 이상인 경우 제약최적실험에 대한 구체적인 해는 일반적으로 존재하지 않는

다고 알려져 있다. Huang과 Wong (1998)과는 별도로 강명욱과 김영일 (2002, 2006)은 타당하지 않은 해의 위험을 피해갈수 있는 순차적인 알고리즘을 제안하였다. 개괄적으로 이야기하면 m개의 실험계획기준에 대한 우선순위를 부여한다. 제일 중요한 실험계획기준을 제약조건에 배치하고, 그 다음으로 중요한 실험계획기준을 목적함수에 배치한 다음에 제약 실험계획법을 실행한다. 다음으로 첫 번째와 두 번째로 중요했던 목적함수를 제약조건식으로 배치를 하되 e_2의 값은 첫 번째 단계에서 나온 목적함수의 값보다 작은 값으로 설정하여 비타당성의 문제를 해소한다. 세 번째로 중요한 실험계획기준을 목적함수에 배치한 다음 제약실험계획에 대한 해를 구한다. 자세한 내용은 위에서 언급한 논문을 참조하기 바란다. 그러나 이러한 방법의 단점은 비타당성의 문제를 해소하였으나 우선순위에 대한 설정을 실험자가 하여야 하는 부담이 생긴다. 실험계획기준이 많아지면 다소 사용하기 편치 않은 면이 있다. 보통의 경우 실험자가 두 가지 혹은 여러 가지 실험계획기준에 대해 선호도를 가지고 있다 하더라도 그 정도를 정확하게 실험에 반영하기는 어렵기 때문이다. 문제점을 해결하기 위한 방법의 하나로 실험자가 염두에 두고 있는 여러 최적기준에 대해 가지는 최소의 효율을 최대화 하는 실험계획기준을 생각해 볼 수 있다. 이 기준은 실험자가 가지고 있는 최적기준들의 집합을 Γ, 그중 하나의 원소를 γ라 한다면 모든 $\gamma \in \Gamma$에 대해 발생할 수 있는 최악의 효율을 최대로 끌어 올려 보겠다는 것이 기본적인 의미이다. 이러한 방법이 최소최대 (maximin) 방법이다. 이러한 실험계획 ξ^*는 식 (2.2)와 같은 조건을 만족한다.

$$\max_{\xi} \min_{\gamma \in \Gamma} e(\xi, \xi_{\gamma}^*) = \min_{\gamma \in \Gamma} E(\xi^*, \xi_{\gamma}^*). \tag{2.2}$$

여기서 $e(\xi, \xi_{\gamma}^*)$는 ξ가 ξ_{γ}^*에 대해 가지는 효율을 의미한다. 이러한 실험계획기준은 위에서 언급한 두 가지 실험계획기준과는 달리 실험계획기준에 대한 선호도를 사전에 명시할 필요가 없는 특징을 가진다. 실험계획기준에 대한 명확한 선호도가 존재하지 않는 경우에는 이러한 방법은 효과적일 수 있으나 실험자가 여러 개의 실험계획기준에 대해 선호도가 존재하는 경우는 이러한 방법은 매우 보수적인 방법일수 밖에 없다. 참고로 Läuter의 결합최적기준에서 λ를 결정하는 방법으로 최적실험계획인 ξ_{λ}^*에서 각각의 실험계획기준에 대한 효율을 계산한 후에, 각 효율값이 같은 결과가 나오는 가중치 λ를 찾는 문제는 이러한 최소최대 방법과 같은 맥락에서 이해가 된다. 실제로 Imhof와 Wong (2000)은 두 개의 실험계획기준이 있을 때 효율성 그림 (efficiency plot)에서 두 효율성이 만나는 교점이 최소최대를 생성하는 가중치가 부여되는 점임을 수학적으로 밝힌바 있다. 그러나 이와 같은 성질은 실험계획기준이 3개 이상으로 확장되는 경우에는 성립되지 않는다.

단일 실험계획기준에 대한 최적실험계획에 대한 성질을 밝히는 작업이 초창기의 최적실험계획법의 문제였다면 최근에는 다중실험계획기준에 대한 연구가 활발하다. 문헌에서 밝힌 위의 세 가지 방법들은 각자의 장단점을 가지고 있다. 최소최대 방법은 매우 보수적이며 제약조건 실험계획법은 실험계획기준이 다수 존재하는 경우에는 우선순위를 정하고 순차적으로 실험계획을 설정하여야 하기 때문에 실험자가 기피할 수 있다. Läuter의 결합실험계획은 기존의 알고리즘을 적용하여 구하는 장점이 있으나 의미가 있

는 가중치를 부여하는 것은 매우 어렵다. 따라서 본 논문에서는 하나의 방법을 고수하기 보다는 최소최대 방법과 제약조건 실험계획법을 혼합시킨 방법을 제안하고자 한다. 물론 Läuter의 방법은 여러 개의 실험계획기준 하에서도 적용이 되나 가중치의 부여의 객관성 때문에 제외하였다.

실험자가 다수의 실험계획기준 $\Phi_1, \Phi_2, \ldots, \Phi_k$에 관심이 있다고 하자. 그러나 실험자는 이중 한 개의 실험계획기준은 확실한 선호도를 가지고 있으나 나머지 기준에 대한 선호도는 없다고 간주한다. 편의상 Φ_1이 제일 선호도가 높은 즉, 가장 중요한 실험계획기준이라고 가정한다. 그렇다면 바람직한 실험계획을 찾는 방법으로 실험계획 ξ가 Φ_1-최적에 대해 가지는 효율성은 e만큼 보존한다는 가정 하에서 나머지 기준에 대해서는 식 (2.3)와 같은 변형된 최소최대 방법을 생각하여 볼 수 있을 것이다. 즉,

$$\begin{aligned} \max_{\xi} \min_{\gamma \in \Gamma} e(\xi, \xi^*_\gamma) &= \min_{\gamma \in \Gamma} e(\xi^*, \xi^*_\gamma), \Gamma = [\Phi_2, \ldots, \Phi_k] \\ \text{subject to} \quad & e(\xi, \xi^*_{\Phi_1}) \geq e \end{aligned} \tag{2.3}$$

여기서 Γ는 Φ_1를 제외한 집합이다. 이와 같은 방법은 다수의 실험계획기준에 대한 선호도를 정하지 않더라도 선호도가 뚜렷한 실험계획기준에 대한 선택만 하면 된다는 점에서 제약조건 실험계획의 장점과 최소최대 방법의 장점을 혼합하여 쓰는 방법이고, 우리는 이 방법을 Hybrid 접근방법이라 부르기로 한다. Hybrid 접근방법은 물론 선호도를 명시하여야할 실험계획기준이 여러 개인 경우라 하더라도 강명욱과 김영일 (2002, 2006)이 제안한 순차적인 방법을 적용하면 되기 때문에 매우 유연성이 있다. 다만 이와 같은 실험계획기준을 구현하기 위한 알고리즘의 존재여부인데 본 연구에서는 전통적인 방법보다는 유전자 알고리즘에 의거 문제를 해결하였다. 유전자 알고리즘은 특히 최소최대 방법에서 그 효율성을 인정받은바 있다. 자세한 알고리즘의 작동 법에 대해서는 지면관계상 염준근과 남기성 (2000)과 Park 등 (2005)등을 참조하기 바란다. 참고적으로 본 논문에서는 임의로 교배율과 돌연변이율은 각각 0.10, 0.05로 두었다. 물론 이러한 비율의 차이로 인한 알고리즘의 성과가 나타날 수 있으나 본 연구에서는 이러한 목적은 배제되어 있다. 다음 절에서는 2절에서 제안된 하이브리드 (Hybrid) 접근방법을 이용하여 간단한 예제를 들어 그 응용성을 보고자 한다.

3. 예제

예제 3.1 실험자는 모형에 대한 확신이 없다. 다만 참의 모형은 단순회귀모형, f_1과 이차 회귀모형, f_2중 하나로 생각한다. 그리고 실험계획기준은 G-최적 기준이다. $E = \{f_1, f_2\}$ 의 각각의 회귀모형에서 G-최적 실험계획은 0을 중심으로 대칭인 실험계획이고, 예측치의 분산을 최대로 하는 점에서 G-최적 실험계획이 받힘점을 갖기 때문에 바람직한 실험계획의 후보를 받힘점 $\{-1, 0, 1\}$을 갖는 대칭인 근사 실험계획들로 제한하는 것을 생각할 수 있다. 최소최대 방법의 실험계획, ξ^*은 Imhof와 Wong (2000)에 의하여 각각의 회귀모형에 대한 G-효율성을 일치시켜 구해볼 수 있다. 즉, 실험점 -1과 1에서 확률 α를, 실험점 0에서 확률 $1-2\alpha$를 갖는 실험계획이 단순회귀모형 f_1의 최적실험

계획에 대해 가지는 G-효율성과 이차회귀모형 f_2의 최적실험계획에 대해 가지는 G-효율성을 일치시킴으로서 해석적으로 구하여 보았다. 이 경우는 최소최대 실험계획은

$$\xi_E^*(\pm 1) = (\sqrt{10} - 1)/6, \quad \xi_E^*(0) = (4 - \sqrt{10})/3$$

이고, 각각의 모형에 대해 가지는 G-효율성은 83.8% 이다. 이와 같은 성질은 2절에서 언급하였듯이 실험계획기준이 2개인 경우는 유효하다. 삼차 회귀모형이 추가된 경우인 $\Gamma = \{f_1, f_2, f_3\}$를 고려하여 보자. 바람직한 실험계획의 후보로 삼차 회귀모형의 G-최적 실험계획의 받힘점인 $\{-1, -1/\sqrt{5}, 1/\sqrt{5}, 1\}$을 갖는 대칭인 근사 실험계획들로 제한하는 것을 생각할 수 있다. 이러한 성질을 만족하는 실험계획들 중에서 최소최대 실험계획을 유전자 알고리즘에 의해 구해보면,

$$\xi_\Gamma^*(\pm 1) = 0.287, \quad \xi_\Gamma^*(\pm 1/\sqrt{5}) = 0.213$$

이다. 이 실험계획이 각각의 모형에 대해 가지는 G-효율성은 79.5%, 79.5%, 그리고 85.2% 이다. 그런데 여기서 실험자는 삼차회귀모형인 f_3의 가능성을 가정 높이 평가해서, 가장 중요한 모형이라 생각하고, 바람직한 실험계획은 삼차의 회귀모형에 대해 가지는 효율성이 적어도 90% 이상이어야 한다고 가정하자. 이러한 제약조건하에서 논문에서 제안된 Hybrid 접근방법에 의한 실험계획을 유전자 알고리즘에 의해 구해보면,

$$\xi_H^*(\pm 1) = 0.275, \quad \xi_H^*(\pm 1/\sqrt{5}) = 0.225$$

이다. 그리고 이 실험계획이 각 모형에 대해 가지는 G-효율성은 각각 78.0%, 83.7%, 90.0%이다. 보다시피 3차 모형에 대한 효율성을 더 보장받으려면 다른 두 모형에 대한 효율성의 균형이 무너진다. 그러나 첫 번째 모형에 대한 효율성은 1.5%만 떨어졌지만 다른 두 모형에 대한 효율성은 상당히 올라갔음을 주지하기 바란다.

예제 3.2 Huang (1996)이 시도한 3가지 최적실험계획기준을 다시 분석하여보았는데 이는 다음과 같다. 2차 회귀모형에 대해 3가지 실험계획기준을 설정하였다. 하나는 D-최적 기준이며 두 번째는 실험영역 X에 대한 균일분포를 부여한 I-최적 (integrated optimal) 기준, 그리고 마지막 실험계획기준으로 $z = 2$인 점에서의 외삽 (extrapolation) 기준이다. 이 세 가지 최적기준에 대한 Φ_i는

$$\Phi_1(\xi) = -\ln\left[\frac{|M(\xi)|}{|M(\xi_1^*)|}\right]^{1/3}, \quad \Phi_2(\xi) = \frac{trM^{-1}(\xi)W}{trM^{-1}(\xi_2^*)W}, \quad \Phi_3 = \frac{v(z,\xi)}{v(z,\xi_3^*)}$$

이다. 여기서 $W = \int_{\chi} f(x)f^T(x)U(dx)$이며 $v(x,\xi) = f^T(x)M(\xi)^{-1}f(x)$이다. 제약조건이 없는 최소최대 방법의 실험계획을 구해보면,

$$\xi_\Gamma^*(-1) = 0.220, \quad \xi_\Gamma^*(-0.043) = 0.435, \quad \xi_\Gamma^*(1) = 0.345$$

이다. 이 실험 계획이 각각의 기준에 대해 가지는 효율성은 96.13%로 나타났다. 만약 제일 선호도는 D-최적 기준에 있으며 나머지 두 기준에 대해서는 선호도가 동일하고,

D-최적 기준에 대한 효율성은 적어도 95%이상을 보장받는다는 가정 하에서 위에서 언급한 Hybrid기준을 적용하여 실험계획을 구해보면,

$$\xi^*_{H1}(-1) = 0.207, \quad \xi^*_{H1}(-0.039) = 0.453, \quad \xi^*_{H1}(1) = 0.340$$

이다. 이 실험계획이 각각의 기준에 대해 가지는 효율성은 각각 95%, 96.4%, 96.4%로 나타났다. 첫 번째 효율성이 다른 두 효율성에 비해 낮게 설정되어 효율성의 하한값 95%를 97%로 다시 설정하여 Hybrid기준을 적용하여 실험계획을 구해보면,

$$\xi^*_{H2}(-1) = 0.232, \quad \xi^*_{H2}(-0.037) = 0.419, \; \xi^*_{H2}(1) = 0.349$$

이다. 이 실험계획이 각각의 기준에 대해 가지는 효율성은 각각 97%, 95.7%, 95.7%로 나타났다. 예제 3.1과 달리 다른 두 실험계획기준에 대한 효율성은 균형을 이루고 있다.

예제 3.3 실험계획기준은 3가지이다. Φ_1는 단순 회귀모형에 대해 D-최적 기준이고, Φ_2는 이차모형에 대한 D-최적 기준이며, Φ_3는 이차모형에 대한 A-최적 기준이다. Φ_1에 대한 효율성이 적어도 90%를 보장하는 Hybrid 접근방법에 의한 실험계획을 구해보면,

$$\xi^*_{H1}(-1) = 0.405, \quad \xi^*_{H1}(0) = 0.19, \quad \xi^*_{H1}(1) = 0.405$$

이다. 그리고 이 실험계획이 각 최적실험계획기준에 대한 효율성은 각각 90.0%, 94.4%, 61.5%로 나타나는 반면 Φ_1효율성을 70%를 보장하는 Hybrid 접근방법에 의한 실험계획은

$$\xi^*_{H2}(-1) = 0.285, \quad \xi^*_{H2}(0) = 0.430, \quad \xi^*_{H2}(1) = 0.285$$

이고, 이 실험계획이 각 최적기준에 대한 효율성은 각각 75.4%, 98.0%, 98.0%로 나타난다. 선형회귀모형에 대한 70%의 효율성은 쉽게 취득되는 결과이기에, 나머지 두 기준에 대한 효율성은 높게 나타난다.

예제 3.2와 예제 3.3의 결과로부터 Hybrid 접근방법은 가장 중요한 실험계획기준에 대해 실험자가 부여하는 효율성의 크기에 따라 실험의 형태가 달라짐을 확인할 수 있다.

4. 결론

본 연구에서는 다중목적을 가지고 있는 실험환경 하에서 기존의 방법들이 가지고 있는 장점들을 혼합하여 새로운 Hybrid방법을 실험적으로 제안하였다. 이러한 방법은 한두 개의 실험계획기준을 제외한 실험계획기준에 대해서는 선호도가 명확하지 않을 때 사용할 수 있다고 본다. 즉, 제약조건에서 가장 중요시 하는 최적의 실험계획기준의 효율성을 보장하는 한편 목적함수에는 최소최대 방법을 써 보수적인 방법을 융합하였다. 다만 Läuter의 방법은 가중치를 설정하는데 있어 주관성이 내재되어 있기에, 융합이 어려운 점이 있어 배제하였다. 기존의 알고리즘을 적용하여 새로운 방법론을 이용한 실험계획의 구성은 어렵다. 대신 유전자알고리즘을 이용하여 이를 성공적으로 구성하였다.

이러한 유전자 알고리즘의 활성화는 많은 방법론의 구현을 위해 기여를 많이 할 것으로 짐작된다. 또한 본 논문에서는 근사 실험계획 (approximate design) 개념으로 몇 가지 예제들에서 시험적으로 근사 실험계획을 구성하였으나 실제로 현장에서 적용이 되기 위해서는 각각의 실험점에서 몇 회의 실험 실시가 정확하게 결정되는 정확 실험계획 (exact design)의 형태로 다양한 반응표면 모형에 대한 실험의 구조를 밝힐 필요가 있다. 추후 연구과제이다.

참고문헌

강명욱, 김영일 (2002). Multiple constrained optimal experimental design. <한국통계학회논문집>, **9**, 619–627.

강명욱, 김영일(2006). Strategical issues in multiple-objective optimal experimental design. <한국통계학회논문집>, **13**, 1–10.

염준근, 남기성(2000). A study on D-optimal design using the genetic algorithm. <한국통계학회논문집>, **7**, 357–366.

Box, G. E. P. and Draper, N. R. (1975). A basis for the selection of a response surface design. *Journal of the American Statistical Association*, **54**, 622–654.

Cook, R. D. and Fedorov, V. V. (1995). Constrained optimization of experimental design (with discussion). *Statistics*, **26**, 129–178.

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.

Fedorov, V. V. (1972). Theory of Optimal Experiments. Translated and edited by W. J. Studden and E. M. Klimko, Academic Press, New York.

Kiefer, J. (1959). Optimum experimental design. *Journal of the Royal Statistical Society*, Ser. B, **21**, 272–319.

Huang, Y. C. (1996). Multiple-objective optimal designs. Doctor of Public Health Dissertation, Department of Biostatistics, School of Public Health, UCLA.

Huang, Y. C. and Wong, W. K. (1998). Multiple-objective designs. *Journal of Biopharmaceutical Statistics*, **8**, 635–643.

Imhof, L. and Wong, W. K. (2000). A graphical method for finding maximin designs. *Biometrics*, **56**, 113–117.

Läuter, E. (1974). Experimental planning in a class of models. *Mathematishe Operationsforshung und Statistik*, **5**, 673–708.

Park, Y. J., Montgomery, D. C., Folwer, J. W. and Borror, C. M. (2005). Cost-constrained G-efficient response surface designs for cuboidal regions. *Quality and Reliability Engineering International*, **22**, 121–139.

Pukelsheim, F. (1993). *Optimal Design of Experiments*. John Wiley & Sons, New York.

Silvey, S. D. (1980). *Optimal Design*. Chapman & Hall/CRC.

Stigler, S. M. (1971). Optimal experimental design for polynomial regression. *Journal of the American Statistical Association*, **66**, 311–318.

Studden, W. J. (1982). Some robust type D-optimal designs in polynomial regression. *Journal of the American Statistical Association*, **77**, 916–921.

Wong, W. K. (1995). A graphical approach for constructing constrained D- and L-optimal designs using efficiency plot. *Journal of Statistical Simulation and Computations*, **53**, 143–152.

Wong, W. K. (1999). Recent advances in multiple-objective design strategies. *Statistica Neerlandica*, **53**, 257–276.

[Received July 2007, Accepted August 2007]