The Korean Journal of Applied Statistics (2015) DOI: http://dx.doi.org/10.5351/KJAS.2015.28.1.093
28(1), 93–102

Two-Stage Experimental Design for Multiple Objectives

Dae-Heung Jang[a] · Youngil Kim[b,1]

[a]Department of Statistics, Pukyong National University
[b]School of Business and Economics, ChungAng University

(Received December 3, 2014; Revised January 28, 2015; Accepted February 5, 2015)

Abstract

The D-optimal design for the nonlinear model typically depends on the unknown parameters to be estimated. Therefore, it is strongly recommended in literature to use a sequential experimental design for estimating the parameters. In this paper two stage experimental design is discussed under many different circumstances including estimating parameters. The method is so universal to be applied to any mixture of objectives for any model including linear model. A hybrid approach is suggested to handle more than 2 objectives in two-stage experimental design. The design is discussed in approximate design framework.

Keywords: D-optimal design, sequential experimental design, two-stage experimental design.

1. 서론

반응변수 $y(x)$는 식 (1.1)과 같이 x의 비선형 함수로 연결되어 있다.

$$y = \eta(x, \theta) + \epsilon, \tag{1.1}$$

여기서 $\theta^T = (\theta_1, \ldots, \theta_p)$는 추정하여야 할 모수벡터이다. 실험설계 ξ는 실험영역 Ω 내의 유한 k개의 점 $x_i \in \Omega$, $i = 1, 2, \ldots, k$에 대한 확률질량함수 $\xi(x_i)$, $i = 1, 2, \ldots, k$로 기술된다. 개개의 $\xi(x_i)$, $i = 1, 2, \ldots, k$는 n_i/n의 형태를 가지며 $\sum_i^k \xi(x_i) = 1.0$의 제약조건을 갖는다. $n \times \xi(x_i)$가 정수라는 제약조건이 주어지면 정확실험설계(exact design)라 하고 제약조건이 없는 경우를 근사실험설계(approximate design)라 한다. 본 연구에서는 근사실험계획만 고려한다. 식 (1.1)에서 오차항 ϵ의 구조로서 기댓값 0 그리고, 일반성의 손실 없이 $\sigma^2 = 1$인 비상관의 정규분포를 가정한다면 실험 ξ에 대한 정보행렬(information matrix)은 식 (1.2)와 같이 주어진다.

$$M(\xi, \theta) = \int_\Omega \frac{\partial \eta(x,\theta)}{\partial \theta} \frac{\partial \eta(x,\theta)}{\partial \theta^T}\, d\xi(\mathbf{x}), \tag{1.2}$$

여기서 $\partial \eta(x,\theta)/\partial\theta$는 미분벡터(derivative vector)라 하며 차원은 $p \times 1$이다.

본 연구에서 다루고자 하는 기본적인 실험기준은 식 (1.2)의 정보행렬의 행렬식(determinant)을 최대화하는 실험기준인 $D(\theta)$-최적이다. 그러나 $\eta(x,\theta)$가 비선형모형인 경우 $D(\theta)$-최적은 추정하고자 하는 θ

[1]Corresponding author: School of Business and Economics, ChungAng University, 84 Heukseok-ro, Dongjak-gu, Seoul 156-756, Korea. E-mail: yik01@cau.ac.kr

값에 의존하는 모순이 일어난다. 이런 경우에는 θ 대신 θ의 초기값으로 θ^0를 가정하고 국지(local) 실험을 한 후 추가 실험여부를 결정하는 경우가 많다. 만약 θ를 두 개의 부분집합인 p_1개의 모수 θ_1과 p_2개의 모수 θ_2로 분할한다면 정보행렬 $M(\xi,\theta)$는 다음과 같이 분할된다.

$$M(\xi,\theta) = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}, \tag{1.3}$$

여기서 M_{11}, M_{22}는 각각 모수 θ_1, θ_2에 해당하는 부분행렬이다. 만약 실험자가 p_2개의 부분집합 모수 θ_2의 추정에 관심을 가진다면 $|M(\xi,\theta)|/|M_{11}(\xi,\theta)|$을 최대화하는 $D_s(\theta_2)$-최적을 사용한다. 부연하면 첨자 s는 모수의 부분집합(subset)을 의미하고 그 중 θ_2의 추정에 관심을 두는 경우이다. 자세한 내용은 Atkinson과 Donev (1992)나 Kitos (2013)과 같은 책을 참조하기 바란다.

본 연구에서 정의될 2단계 실험설계를 논하기 전에 임의의 실험설계 ξ가 최적실험에 대하여 가지는 효율을 먼저 정의할 필요가 있다.

$D(\theta)$-**효율**: $D(\theta)$-최적에 대한 실험 ξ의 $D(\theta)$-효율 $\phi_\xi(\xi^*_{D(\theta)})$은 다음과 같이 정의된다.

$$\phi_\xi(\xi^*_{D(\theta)}) = \left\{ \frac{\det(M(\xi,\theta))}{\det\left(M\left(\xi^*_{D(\theta)},\theta\right)\right)} \right\}^{\frac{1}{p}},$$

여기서 $\xi^*_{D(\theta)}$는 $\eta(x,\theta)$에 대한 $D(\theta)$-최적실험이며 det는 행렬식을 의미한다.

$D_s(\theta_2)$-**효율**: $D_s(\theta_2)$-최적에 대한 실험 ξ의 $D_s(\theta_2)$-효율 $\phi_\xi(\xi^*_{D_s(\theta_2)})$은 다음과 같이 정의된다.

$$\phi_\xi\left(\xi^*_{D_s(\theta_2)}\right) = \left\{ \frac{\det\left(M(\xi,\theta)\right)/\det\left(M_{11}(\xi,\theta)\right)}{\det\left(M\left(\xi^*_{D_s(\theta_2)},\theta\right)\right)/\det\left(M_{11}\left(\xi^*_{D_s(\theta_2)},\theta\right)\right)} \right\}^{\frac{1}{p_2}}.$$

참고로 위와 같은 효율은 어느 실험기준 ψ이든지 정의가 될 수 있다. 간단히 표현하면 실험 ξ의 ψ-효율이란 실험기준 ψ에 대하여 실험 ξ가 ψ-최적실험 ξ^*_ψ에 대하여 가지는 효율 $\phi_\xi(\xi^*_\psi)$이라 정의가 된다. 여기서는 $D(\theta)$-효율과 $D_s(\theta_2)$-효율만 언급하였으나 향후 발생되는 기타 효율도 이런 기본적인 정의에 의하여 계산될 것이며 필요한 경우 구체적인 효율 정의를 다시 할 것이다.

2. 2단계 실험설계

실험자는 실험의 특성상 다목적의 실험기준이 있는 경우 보유하고 있는 자원을 모두 처음부터 사용하기를 원하지 않는다. 특히, 비선형 모형인 경우 $D(\theta)$-최적처럼 모수의 초기값에 의존하는 경우는 더욱이 그렇다. 이런 경우 문헌에서는 모수 θ의 불확실성으로 인한 대처 방법으로 maximin 방법이나 혹은 θ에 대한 사전확률을 적용한 베이지안(Bayesian) 방법이 사용되나 근본적으로 문헌에서는 Fedorov (1972)와 같이 순차실험(sequential design)을 추천한다. 모수의 추정을 위한 순차실험은 단계별로 모수 값을 갱신하는 과정을 거칠 것이다.

먼저 $\eta(x,\theta)$를 대상으로 한 1단계 실험설계를 ξ_1이라 하고 해당되는 정보행렬을 $M(\theta^0,\xi)$라 한다. 여기서 θ^0는 1단계 실험설계에서 주어지는 모수의 초기값이다. 그리고 1단계에서 나오는 실험을 바탕으로 2단계 실험설계를 한다면 1단계의 실험에서 나온 정보를 포함하여 2단계 실험을 실시하여야 한다. 모수의 초기값 θ^0가 1단계 실험을 거친 다음 θ^1으로 확인이 된다면, 그리고 가지고 있는 자원을 1단계

에서 α_1만큼 사용하고 나머지 α_2만큼을 2단계 실험에 배정하는 방법으로 실험을 k단계 진행한다면 k단계의 실험은 식 (2.1)과 같은 정보행렬에 근거한 실험기준을 가지고 실험설계 ξ을 하여야 한다.

$$\tilde{M}(\theta, \xi) = \alpha_1 M\left(\theta^0, \xi_1\right) + \alpha_2 M\left(\theta^1, \xi_2\right) + \alpha_3 M\left(\theta^2, \xi_3\right) + \cdots + \alpha_k M\left(\theta^{k-1}, \xi_k\right), \tag{2.1}$$

$$0 \le \alpha_i \le 1, \qquad \sum_{i=1}^{k} \alpha_i = 1.0.$$

만약 $\alpha_i = 1/k$, $i = 1, 2, \ldots, k$라 한다면 이는 매 단계 자원을 등분하여 사용하다는 의미이다. 예를 들어 $k = 2, \alpha_1 = 0.5$로 가정하고 1단계 실험기준을 $D(\theta^0)$-최적, 그리고 2단계에서 실험기준을 $D(\theta^1)$-최적으로 한다면 이는 $D(\theta)$-최적을 기준으로 자원을 각 단계에 동등하게 배분한 2단계 실험설계가 된다. 본 연구에서는 순차실험의 단계를 편의상 2단계로 나누어 설명하고자 한다. 물론 순차실험은 2단계로 국한할 필요는 없다.

Dykstra (1971), Evans (1979) 등과 같이 기존 실험에 받힘점을 추가하는 추가실험에 관한 논문들이 존재하는데 이들은 기본적으로 (2.1)의 개념을 선형모형을 대상으로 단계별로 정확실험설계에 적용한 논문들이다.

본 연구에서는 관점을 달리하여 실험자는 복수의 실험 목적을 가지고 있다고 하자. 예를 들어 하나는 모형에 대한 전반적인 모수 추정이고 다른 하나는 부분모수의 추정이라고 하자. 실험자는 두 개의 목적을 되도록 다 충족시키려고 한다. 그러나 모수에 대한 불확실성이 대두가 되는 경우 모수의 초기값인 θ^0에 의존하여 두 실험기준을 동시에 진행하는 것은 별로 바람직하지 않다. 먼저 일차적인 실험을 통하여 θ^0에 대한 확신을 가진 다음 갱신된 모수 값 θ^1에 의거하여 2단계 실험을 진행하는 것이 바람직하다. 따라서 본 연구에서는 1단계 실험설계에서는 1단계 실험 목적에 맞는 실험설계를 구하고 2단계 실험에서는 2차적인 목적의 실험을 구현한다. 이와 같은 시도는 Myers 등 (1996)이나 Fedorov와 Lenov (2014)에서 방법론의 제안만 있었을 뿐 복합적인 실험에 적용된 사례가 많지 않아 본 연구에서는 확장된 응용성을 탐색하고 새로운 복합적인 실험기준을 시도하고자 한다. 다음 절에서는 2단계 실험의 적용을 다양한 예제로 알아보도록 한다.

3. 예제

2절에서 제안된 2단계 실험설계는 주어진 모형이 반드시 비선형일 필요는 없다. 선형인 모형이라도 다목적실험인 경우 적용이 된다. 따라서 아래 예제들에서는 처음 세 예제를 선형에서 그리고 나머지 세 예제는 비선형 모형을 기준으로 2단계 실험설계의 특징을 살펴보고자 한다.

예제 3.1: 두 개의 실험설계기준 - 2차 회귀모형

주어진 모형은 식 (3.1)의 2차 회귀모형이다.

$$\eta(x, \theta) = \theta_0 + \theta_1 x + \theta_2 x^2. \tag{3.1}$$

$D(\theta)$-최적 및 $D_s(\theta_2)$-최적 실험은 각각 잘 알려졌듯이 다음과 같다. Atkinson과 Donev (1992, p.110)을 참조하기 바란다. 참고로 식 (3.1)은 선형이므로 $D(\theta)$-최적은 θ에 의존하지 않는다. $D_s(\theta_2)$는 θ의 부분모수 θ_2에 대한 최적실험이 된다.

$$\xi^*_{D(\theta)} : \xi(\pm 1) = \xi(0) = \frac{1}{3},$$

$$\xi^*_{D_s(\theta_2)} : \xi(\pm 1) = \frac{1}{4}, \quad \xi(0) = \frac{1}{2},$$

여기서는 $\alpha = 1/2$를 가정하고 2단계 실험실계를 시행한다고 가정한다. 1단계에서는 $D(\theta)$-최적을, 그리고 2단계에서는 1, 2단계에서 나오는 복합 정보행렬 $\tilde{M}$의 부분 행렬에 해당하는 모수 θ_2의 추정을 위한 $D_s(\theta_2)$-최적을 추구할 것이다. 1단계 실험설계 ξ_1은 다음과 같이 설정할 것이다.

$$\xi_1 : \xi(\pm 1) = \xi(0) = \frac{1}{3}.$$

그리고 2단계에서 $D_s(\theta_2)$-최적을 추구한다면 2단계 실험설계 ξ_2는 당연히 다음과 같이 설정되어야 한다. 왜냐하면 $D(\theta)$-최적과 $D_s(\theta_2)$-최적은 같은 받힘점을 가지고 있고 2단계 실험기준이 $D_s(\theta_2)$-최적 실험이기 때문이다. 즉, $0.5\xi_1 + 0.5\xi_2$는 $\xi^*_{D_s(\theta_2)}$가 되는 결과를 낳는다.

$$\xi_2 : \xi(\pm 1) = \frac{1}{6}, \quad \xi(0) = \frac{2}{3}.$$

참고로 $D_s(\theta_2)$최적이 $D(\theta)$-최적에 대해 가지는 효율 $\phi_{\xi^*_{D_s(\theta_2)}}(\xi^*_{D(\theta)})$은 0.9449다.

예제 3.1의 경우는 2단계에서의 실험설계가 너무 자명하기 때문에 이를 구하는 알고리즘은 필요 없다. 단지 2단계 실험설계의 특징을 알리고자 이 예제를 설정하였다.

예제 3.2: 2단계 maximin 복합 최적실험기준

예제 3.1을 다시 검토하여 보자. 예제 3.1에서 밝혔듯이 실험자는 $D(\theta)$-최적 및 $D_s(\theta_2)$-최적 두 개의 실험목적을 가지고 있다고 가정하자. 1단계 실험설계의 기준은 $D(\theta)$-최적이다. 다만 예제 3.1과 다르게 2단계에서는 $D_s(\theta_2)$-최적이 주 관심사항이지만 그래도 우리가 가정한 모형에 대한 추정 $D(\theta)$-최적도 여전히 관심을 가지고 있는 경우를 설정하여 보자. 두 개의 실험목적을 동시에 만족시키는 실험설계를 2단계에서 설정하여야 한다. 이런 경우 2단계에서의 실험기준으로 $D(\theta)$-최적과 $D_s(\theta_2)$-최적에 대해 가지는 효율 중 낮은 효율을 끌어 올리는 maximin 방법을 이용한 실험설계 ξ_2를 생각할 수 있다.

$$\max_{\xi_2} \min \left\{ \phi_{\xi_{2satge}}\left(\xi^*_{D(\theta)}\right), \phi_{\xi_{2stage}}\left(\xi^*_{D_s(\theta_2)}\right) \right\},$$

여기서 $\phi_{\xi_{2satge}}(\xi^*_{D(\theta)})$ 및 $\phi_{\xi_{2stage}}(\xi^*_{D_s(\theta_2)})$은 2단계 실험설계 (2stage)가 각각 $D(\theta)$-최적 및 $D_s(\theta_2)$-최적에 대하여 가지는 효율이다. 다음이 $\alpha = 0.5$로 하였을 때 2단계에서의 실험설계 ξ_2이다.

$$\xi_2 : \xi(\pm 1) = 0.2364, \quad \xi(0) = 0.5272.$$

이러한 2단계에서의 실험은 예제 3.1과는 달리 알고리즘에 의하여 구하여 한다. 최적실험에서는 다수의 알고리즘이 존재하는데 본 연구에서는 Fedorov (1972)의 매 단계 두 받힘점의 교환(exchange)을 통해 목적함수가 수렴할 때 까지 시행되는 반복적인 알고리즘(iterative algorithm)을 사용하였다. 자세한 내용은 Fedorov (1972)의 책을 참조하기 바란다. 아래 예제들의 2단계 실험결과는 모두 알고리즘에 의해 구해진 결과들이다.

그리고 이러한 2단계 실험설계가 $D(\theta)$-최적 및 $D_s(\theta_2)$-최적에 대하여 가지는 효율은 똑같이 0.9805이다. 이러한 maximin 방법은 2단계에서도 여전히 첫 번째 목적을 관심을 가지고 있는 경우 유효하다고 본다. 1, 2단계를 거치지 않고 처음부터 maximin과 같은 복합 최적실험을 실시할 수 있으나 이는 모든 자원을 1단계에서 사용하는 문제가 발생하므로 본 연구에서처럼 α값을 정한 후 2단계에서 maximin과 같은 복합실험기준을 사용하는 2단계 실험방법을 제안하는 것이다.

예제 3.3: 모델 확장(model expansion) - 다항회귀모형

1단계의 실험기준을 2차 회귀모형을 대상으로 설정한다. 그리고 실험기준을 $D(\theta)$-최적으로 한다면 다음이 1단계 실험설계 ξ_1이다.

$$\xi_1 : \xi(\pm 1) = \xi(0) = \frac{1}{3}.$$

그러나 1단계 실험이 종료된 후 2차 회귀모형보다는 3차 회귀모형이 더 타당하다고 판단되면 추가로 3차 회귀모형을 대상으로 $D(\theta)$-최적 실험을 시행하는 시나리오를 고려하여 보자. 그러나 2차 회귀모형과 달리 3차 회귀모형인 경우는 $D(\theta)$-최적은 다음과 같이 받힘점 0을 포함하고 있지 않는다.

$$\xi^*_{D(\theta)} : \xi(\pm 1) = \xi\left(\pm\sqrt{\frac{1}{5}}\right) = \frac{1}{4}$$

따라서 2단계 실험설계는 받힘점으로서 0을 포함하고 있지 않을 가능성이 높다. 그리고 이미 1단계 실험설계에서 0을 포함하고 있기 때문에 중간받힘점 $\pm\sqrt{1/5}$와 일치하지 않는 받힘점을 선택할 가능성이 높다고 사전 판단된다. 다음이 $\alpha = 0.5$를 가정하였을 때 2단계에서의 실험설계 ξ_2 결과이다. 역시 중간받힘점으로서 $\pm\sqrt{1/5}$와는 다른 받힘점을 찾았다.

$$\xi_2 : \xi(\pm 1) = 0.1323, \xi(\pm 0.5033) = 0.3677$$

참고로 1, 2단계를 거친 실험이 3차 회귀모형을 가정한 $D(\theta)$-최적에 대해 가지는 효율은 0.9561로 밝혀진다. α의 값을 0.5보다 낮출수록 이 효율은 높아질 것이다. 다수의 실험기준과 마찬가지로 모형이 달라지는 경우라 하더라도 얼마든지 갱신된 모형에 대해서도 2단계 실험을 할 수 있다.

여기서 시도되는 제안은 모형이 선형이건 비선형이건 적용이 가능하다. 다음 예제 3.4에서는 비선형 모형을 이용하여 이와 같은 개념을 알아보도록 하자.

예제 3.4: 모수의 갱신(updated value of parameter) - Michaelis-Menten 모형(MM)

식 (3.2)는 생명과학분야에서 효소역학(enzyme-kinetics)을 연구하는데 있어 많이 쓰이는 Michaelis-Menten 모형이다. Raaijmakers (1997)에 의해 이러한 모형을 이용한 통계적인 분석이 많이 이루어지고 있음을 알 수 있다. 모형은 통상적으로 반응변수의 속도(velocity)를 기질농축량(substrate concentration)인 x로 설명한다. x의 값은 이론적으로 무한이나 통상적으로 상한값을 가진다.

$$\eta(x, \theta) = \frac{\theta_1 x}{\theta_2 + x}, \quad \theta = (\theta_1, \theta_2),\ x \in [0, B], \tag{3.2}$$

여기서 θ_1은 속력이 도달할 수 있는 최대값이고 θ_2는 Michaelis-Menten 상수이다. θ_2에 따라 이 함수의 성장률이 정해진다. θ_1, θ_2 모두 양수의 값을 가진다. 식 (3.2)에서 보면 θ_1은 선형모수이고 θ_2는 비선형모수이다. 따라서 Hill (1980)에 의하면 식 (3.2)의 모형에 대한 $D(\theta)$-최적은 θ_2의 값에만 의존한다. 즉, θ_1의 값이 달리 주어져도 $D(\theta)$-최적은 변하지 않는다. 그러나 MM모형은 비선형 모형이므로 $D(\theta)$-최적을 구하기 위해서는 θ_1값을 지정하여야 한다. 편의상 본 연구에서는 $\theta_1 = 1$로 가정한다.그리고 Kitos (2001)에 의하여 $D(\theta)$-최적은 받힘점의 위치가 B와 $B\theta_2/(B + 2\theta_2)$로 두 받힘점에 질량 1/2씩 배정하는 등질량(equi-weight)임이 밝혀졌다. 예를 들어 B가 10이고 θ_2가 1이면 다음이 $D(\theta)$-최적에 의한 실험설계이다.

$$\xi^*_{D(\theta)} : \xi(10) = \xi(0.8333) = \frac{1}{2}$$

그러나 θ_2의 초기값을 1로 가정하였으나 참의 값이 10으로 나오는 경우 $D(\theta)$-효율은 0.5207로 결과가 떨어지게 나온다. 문헌에서는 종종 이런 상황을 타개하기 위하여 Dette와 Biedermann (2003)은

θ_2에 대한 maximin 방법을 도입하곤 한다. 참고로 다음이 maximin 실험설계 결과이다. 이러한 실험은 θ_2의 참의 값이 어떻게 나오더라도 효율은 0.8383으로 높게 나와 모수의 불확실성에 대처하는 실험설계가 된다. 자세한 내용은 Kim과 Jang (2014)을 참조하기 바란다.

$$\xi_{maxmin} : \xi(10) = \xi(1.7573) = \frac{1}{2}$$
$$\phi_{\xi_{D(\theta_2=1.0)}}\left(\xi^*_{D(\theta_2=10.0)}\right) = \phi_{\xi_{D(\theta_2=10.0)}}\left(\xi^*_{D(\theta_2=1.0)}\right) = 0.8383.$$

그러나 maximin 역시 두 받힘점 실험설계이다. 따라서 이런 실험설계는 받힘점의 개수가 모형의 모수의 개수와 같을 뿐 크지 않기 때문에 적합결여 검정을 실시할 수 없는 치명적인 단점이 나타난다. 본 연구에서 제안한 2단계 실험설계와 어떤 차이가 있는지 비교하여 보자. 비교를 위해 1단계에서는 θ_2의 초기값 θ_2^0는 1로 가정하고 1단계 실험 후 나온 결과에 의하여 갱신된 θ_2^1의 값은 10으로 한다. 다음이 $\alpha = 0.5$일 때의 1, 2단계 실험설계이다.

$$\xi_1 : \xi(10) = \xi(0.8333) = \frac{1}{2},$$
$$\xi_2 : \xi(10) = 0.3149, \quad \xi(3.3333) = 0.6851.$$

그리고 이러한 2단계 실험이 $\theta_2 = 10$을 기준으로 할 때 $D(\theta)$-효율은 0.8202이다. 2단계 실험설계에서는 받힘점 0.8333에 대한 추가질량은 확보하지 않는 것으로 나타난다. 이미 2단계 실험기준의 목적에 맞지 않는 받힘점이기 때문이다. 경우에 따라서는 θ_2에 대한 확신 없이 1단계에서 실험을 종료한다든지 혹은 maximin 실험을 하기 보다는 단계별로 실험을 추진하는 것이 바람직할 것이다. 비록 maximin 방법에 비하여 $\theta_2 = 10$을 기준으로 효율이 약간 떨어지기는 하지만 무엇보다도 이 실험설계는 3개의 받힘점으로 구성되어 향후 적합결여검정을 실시할 수 있다는 큰 장점이 있다.

예제 3.5에서는 갱신된 θ_2에 대해서 추가실험을 하는 것 이외에도 2단계 실험설계에서 추구하는 실험기준이 바뀌는 경우를 고려하도록 하여 보자.

예제 3.5: 모수의 갱신 및 두 최적실험기준 - Extended Michaelis-Menten 모형(EMM)

Dunn (1988)은 기본적인 Michaelis Model을 확장하여 식 (3.3)과 같은 모형을 소개하였다.

$$\eta(x, \theta) = \frac{\theta_1 x}{\theta_2 + x} + Fx, \quad x \in [0, B]. \tag{3.3}$$

편의상 θ는 기호를 θ_1, θ_2와 F로 나누어, 즉 $\theta^T = (\theta_1, \theta_2, F)$로 표기한다. 식 (3.3)에서 보면 θ_1, F는 선형모수이고 θ_2는 비선형모수이다. 따라서 EMM모형 역시 Hill (1980)에 의하면 θ_2에 의해서만 $D(\theta)$-최적이 결정되는 부분비선형 (partially nonlinear model)모형이다. $F = 0$이면 위의 모형은 Michaelis-Menten 모형이 된다. 복잡하기는 하지만 Trandafir와 Lopez-Fidalgo (2004)에 의하여 $D(\theta)$-최적에 대한 해가 알려져 있다. 최적실험은 3개의 받힘점으로 구성되어 있다. 자세한 받힘점의 구조식은 Trandafir와 Lopez-Fidalgo (2004)의 논문을 참조하기 바란다. 참고로 예제 3.4와 마찬가지로 $D(\theta)$-최적은 모수 θ_1의 값에 의존하지 않으나 비선형 모형의 모수이므로 θ_1은 지정하여야 하나 θ_1과 달리 F는 선형모형의 선형모수이므로 F값을 명시할 필요가 없다. 여기서는 편의상 θ_1은 예제 3.4와 같이 1로 가정한다. 그리고 실험영역의 상한값 B는 MM 모형과 마찬가지로 10으로 한다. 특히 식 (3.3)은 비선형 모형에 선형모형이 추가되는 가법형태의 비선형(additive partially nonlinear model)인 모형이므로 Kitos (2013)에 의하면 (θ_1, θ_2)에 대한 추정에 관한 $D_s(\theta)$-최적도 θ_2에만 의존한다.

본 연구에서는 1단계 실험설계기준으로 $D(\theta)$-최적을, 2단계 실험기준으로는 θ_1, θ_2의 추정에 관심을 두는 $D_s(\theta)$-최적을 실험기준으로 하는 2단계 실험을 통한 복합 실험설계를 설정하여 본다. 먼저 θ_2에 대

한 초기값인 θ_2^0가 1인 경우 1단계 실험계획 ξ_1을 알아보면 다음과 같다.

$$\xi_1 : \xi(10) = \xi(3.7800) = \xi(0.5993) = \frac{1}{3}.$$

이와 같은 실험을 바탕으로 2단계에서는 식 (2.1)에서 정의된 복합 정보행렬 $\tilde{M}$을 대상으로 $D_s(\theta)$-최적을 구할 것이다. 추가로 실험 전 θ_2에 대한 초기값으로 $\theta_2^0 = 1$을 가정하였으나 1단계 실험을 거치는 과정에서 $\theta_2^1 = 2$로 그 값이 갱신되었다고 가정하자. 참고로 $\theta_2^0 = 2$일 경우 $D_s(\theta)$-최적은 다음과 같다.

$$\xi^*_{D_s(\theta)} : \xi(10) = 0.1426, \quad \xi(3.9708) = 0.3595, \quad \xi(0.8794) = 0.4979.$$

그리고 다음이 2단계 실험설계 ξ_2이다.

$$\xi_2 : \xi(10) = 0, \quad \xi(3.9708) = 0.3667, \quad \xi(0.8794) = 0.6333,$$

여기서 반힘점 10에서 질량이 부여되지 않는 이유는 이미 1단계 실험설계에서 많은 질량이 $D_s(\theta)$-최적에 필요한 질량보다 많이 부여되었기 때문이다. α를 1/2보다 낮추게 되면 어느 순간에 반힘점 10에 대한 질량이 2단계에서 발생이 될 것이다. 1, 2단계를 거친 실험 $\xi_{2stage} = 1/2\xi_1 + 1/2\xi_2$는 $\theta_2^1 = 2$라고 가정할 때 $D_s(\theta)$-최적에 대하여 가지는 효율 $\phi_{\xi_{2stage}}(\xi^*_{D_s(\theta_2)})$은 0.9079이다.

예제 3.6: 수의 갱신 및 다수의 최적실험기준 - Michaelis-Menten 모형(MM)

여기서는 식 (3.2)의 MM 모형에서 $x_e > B$인 외삽점에서의 예측의 분산을 최소화 하는 기준을 2단계에서 설정하고자 한다. 외삽의 문제는 모형에 대한 확신이 없는 경우 잘 사용치 않는다. 모형이 주어진 영역에서는 확인이 되더라도 그 영역을 넘어서는 같은 모형이 이어지리라고는 생각지 않기 때문이다. 그러나 MM 모형인 경우는 최대 기질농축량의 크기를 임으로 B로 한 것이므로 모형 자체가 변하지 않는다고 보아도 무방하다. 그러나 θ_2에 대한 불확실성이 남아 있다면 1단계 실험을 한 후 갱신된 θ_2^1에 기초하여 2단계에서 외삽최적을 시행하는 것이 바람직 할 것이다.

예제 3.4와 마찬가지로 θ_2의 초기값 θ_2^0은 1로 가정하였으나 2단계에서는 갱신된 θ_2값인 θ_2^1는 2로 가정한다. 관심을 가지는 외삽점은 두개로 $x_{e1} = 12$과 $x_{e2} = 20$이다. 외삽점이 하나인 경우는 예제 3.4와 같은 방법으로 처리하면 되겠지만 2개인 경우는 복합적인 실험을 2단계에서 하여야 한다. 복합적인 실험은 예제 3.2와 마찬가지로 maximin 방법을 제안한다. 즉, 2단계의 복합정보행렬 $\tilde{M}(\xi, \theta)$에 근거한 두 외삽점에서의 예측의 분산이 각각의 외삽최적 ξ^*_{ext}의 분산값에 대해 가지는 두 개의 효율 $\phi_\xi(\xi^*_{ext})$ 중 낮은 효율을 최대화 시키는 다음과 같은 2단계에서의 실험설계 ξ_2가 될 것이다.

$$\max_{\xi_2} \min_{x_e} \left\{ \phi_\xi(\xi^{x_{e1}}_{ext}), \phi_\xi(\xi^{x_{e2}}_{ext}) \right\}.$$

참고로 다음 두 실험은 각각 $\theta_2 = 2$를 가정하였을 때 외삽최적이다.

$$\xi^*_{ext} : x_e = 12, \quad \xi(1.0543) = 0.1009, \quad \xi(10) = 0.8991,$$
$$\xi^*_{ext} : x_e = 20, \quad \xi(1.0543) = 0.2449, \quad \xi(10) = 0.7551.$$

좌측의 반힘점은 외삽점의 위치와 상관없이 θ_2의 값에 의해서만 결정되는 특징이 있다. 그리고 외삽점이 10을 벗어날수록 극단점 10에 배치되는 질량은 감소된다.

다음은 본 연구에서 제안한 1, 2단계의 실험설계 결과이다.

$$\xi_1 : \xi(10) = \xi(0.8333) = \frac{1}{2},$$
$$\xi_2 : \xi(10) = 0.4649, \quad \xi(0.1212) = 0.5351.$$

이러한 실험설계가 두 외삽최적에 대해 가지는 효율은 0.6816이 나온다. 70%보다 작은 값이라 하더라도 이는 결코 적은 숫자가 아니다. 왜냐하면 제안된 maximin 방법을 이용치 않고 개별적으로 외삽최적을 2단계에서 수행하는 경우에는 2단계 실험이 각각의 외삽점에서 다음과 같이 2단계 실험계획 ξ_2가 나오는 데

$$\xi_2 : x_e = 12, \quad \xi(0.0387) = 0.5432, \quad \xi(10) = 0.4567,$$
$$\xi_2 : x_e = 20, \quad \xi(0.0387) = 0.7396, \quad \xi(10) = 0.26044.$$

외삽점 $x_e = 12$에 대한 최적실험이 $x_e = 20$에 대하여 가지는 효율은 0.5476이며 외삽점 $x_e = 20$에 대한 최적이 $x_e = 12$에 대해 가지는 효율은 0.4772로 매우 낮게 나오기 때문이다. 따라서 2단계에서 외삽점이 2개 이상으로 관심이 있는 경우에는 제안한 maximin 방법도 매우 유효하다고 본다.

4. 결론

본 연구에서는 2단계 실험설계를 통한 복합적인 실험기준의 설정을 예제들을 통해 알아보았다. 실험자가 우선순위를 배정할 수 있는 2개 이상의 실험기준을 가정하고 순차적으로 실험을 할 수 있는 상황이라면 본 연구에서 제안한 방법은 매우 효과적이다. 순차적인 실험이 반드시 하나의 목적을 위한 실험이 될 필요는 없기 때문이다. 또한 본 연구에서는 2단계에 걸쳐 실험하는 예제를 들었지만 제안된 실험은 다단계로 확장을 하여도 무방하다. 그리고 본 연구에서는 실험자가 가지고 있는 자원의 배분 문제를 α라는 값을 통해 실험자가 임의로 정할 수 있게 하였다. 마지막으로 종합적인 실험의 효율성 측면에서 강건성 (robustness)도 언급하였다. 그리고 실험이 3개 이상의 다목적 실험인 경우 그리고 실험이 2단계에서 종료가 되어야 한다면 2단계에서 추구하는 실험은 maximin을 이용한 다목적 실험기준을 이용할 수 있을 것이다. 물론 본 연구에서 제안한 maximin 방법만을 고집할 이유는 없다. 2단계에서 선형결합형태의 실험이라든지 제약조건 실험 등을 실시할 수 있는 것이다. 이는 2단계 실험과 다목적실험을 위한 기존문헌의 방법들이 융합된 형태가 된다. 이는 향후 연구과제로 남기는 바이다.

References

Atkinson, A. C. and Donev, A. N. (1992). *Optimal Experimental Designs*, Oxford Press, Oxford.

Dette, H. and Biedermann, S. (2003). Robust and efficient designs for the Michaelis-Menten Model, *Journal of the American Statistical Association*, **98**, 679–686.

Dunn, G. (1988). Optimal designs for drug, neurotransmitter and hormone receptor assays, *Statistics in Medicine*, **7**, 805–815.

Dykstra, O. (1971). The augmentation of experimental data to maximize $|X^T X|$, *Technometrics*, **13**, 682–688.

Evans, J. W. (1979). Computer augmentation of experimental designs to maximize $|X^T X|$, *Technometrics*, **21**, 321–330.

Fedorov, V. V. (1972). *Optimal Experimental Design*, Academic Press, New York.

Fedorov, V. V. and Leonov, A. N. (2014). *Optimal Design for Nonlinear Response Models*, Chapman & Hall/CRC, New York.

Hill, P. D. (1980). D-optimal designs for partially nonlinear models, *Technometrics*, **22**, 275–276.

Kim, Y. I. and Jang, D. H. (2014). The maximin robust design for the uncertainty of parameters of Michaelis-Menten Model, *The Korean Journal of Applied Statistics*, **27**, 1269–1278.

Kitos, C. P. (2001). Design aspects for the Michaelis-Menten model, *Biometrical Letters*, **38**, 53–66.

Kitos, C. P. (2013). *Optimal Experimental Design for Non-Linear Models*, Springer, New York.

Myers, W. R., Myers, R. H., Carter, W. H. JKr., and White, K. L. (1996). Two-stage designs for the logistic regression model in a single agent bioassay, *Journal of Biopharmaceutial Statistics*, **6**, 283–301.

Raaijmakers, J. G. W. (1997). Statistical analysis of the Michaelis-Menten equation, *Biometrics*, **43**, 780–793.

Trandafir, C. and Lopez-Fidalgo, J. (2004). Locally optimal designs for an extension of the Michaelis-Menten model, *Advances in Model-Oriented Design Analysis, Communication to Statistics*, 173–181.

다수목적을 위한 2단계 실험

장대흥[a] · 김영일[b,1]

[a]부경대학교 통계학과, [b]중앙대학교 경영학부

(2014년 12월 3일 접수, 2015년 1월 28일 수정, 2015년 2월 5일 채택)

요 약

D-최적 등을 위시한 최적실험은 비선형모형인 경우 추정을 하여야할 모수에 의존하는 문제점이 존재한다. 따라서 기본적으로 문헌에서는 모수추정을 위해서는 순차실험을 제안한다. 본 연구에서는 2단계 실험설계를 모수추정의 사례를 포함한 다양한 환경 하에서의 사용방법을 알아보았다. 본 연구에서 제안한 내용은 단계의 수나 구체적인 실험기준의 숫자에 상관없이 적용되는 범용적인 기준이다. 본 연구는 2단계 실험에서 3개 이상의 실험목적을 가지고 있는 경우 하이브리드(hybrid)방법을 제안하였다. 모든 실험은 근사실험설계의 형태로 논의되었다.

주요용어: D-최적, 순차실험설계, 2단계 실험설계.

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