한국통계학회논문집
2012, 19권, 1호, 65–75

DOI: http://dx.doi.org/10.5351/CKSS.2012.19.1.065

하이브리드형 제약 외삽실험 계획법

김영일[1,a], 장대흥[b]

[a]중앙대학교 경영학부, [b]부경대학교 통계학과

요 약

실험영역을 벗어난 점에서의 분산을 최소화 하는 외삽-최적의 문제는 모형의 불확실성을 내재하고 있다. 즉, 실험영역을 벗어나게 되면 모형의 불확실성이 높아지므로 모형의 타당성 여부를 진단하여야 한다. 본 연구에서는 외삽(extrapolation) 문제 하에서 기본적인 3가지 실험, 즉 제약 실험과 최대최소 실험, 그리고 복합실험 등을 융합한 새로운 제약조건형의 실험 2가지를 제안하였다. 그리고 실험영역의 바깥에 위치한 점의 위치가 실험영역을 얼마나 벗어나느냐에 따른 실험결과에 대한 영향력도 고려하여 보았다. 문제의 특성상 유전 알고리즘을 이용하여 해를 구하였다.

주요용어: 외삽실험, 모형불확실성, 제약실험, 복합실험, 최소최대실험, 유전 알고리즘.

1. 서론

최적실험계획(optimal experimental design)의 초기 논문들은 주어진 모형 하에서 수학적인 실험기준에 맞는 실험을 사용자에게 제공하는 것들이 주종을 이루었다. 그러나 이런 예외적인 환경은 사실 현실적이지 않다. 왜냐하면 실험자는 모델에 관한 확신이 100% 확실하지 않을 뿐 아니라 실험기준 역시 하나가 아닌 경우가 많기 때문이다. 특히 실험영역을 벗어난 점에서의 외삽을 염두에 둔다면 이러한 문제는 더욱 우려할 만하다. 실험영역에서의 모형은 실험자가 확신을 가지고 설정한다고 해도 실험영역을 벗어나면 모형에 대한 불확실성이 증가하기 때문이다. Box와 Draper (1975)의 논문을 기점으로 이러한 이론적인 최적실험계획법은 다양한 목적을 반영하는 방향으로 수정 보완되었다.

앞으로 전개될 논의와 결과는 식 (1.1)과 같은 m차 다항모형회귀 모형을 가정하였다.

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

여기서 $y(x)$는 반응변수, $f_m^T(x) = (1, x, \ldots, x^m)$이며, $\theta^T = (\theta_0, \theta_1, \ldots, \theta_m)$은 모수벡터이며 그리고 ϵ은 기댓값이 0이고 분산은 일반성의 손실 없이 $\sigma^2 = 1$인 확률변수이다. Ω는 실험영역이다. 본 연구에서는 $\Omega = [-1, 1]$로 설정한다. 실험계획 문제는 실험영역 Ω안에 있는 s개의 점의 유한집합에 질량을 부여하는 확률 질량함수, ξ로 기술할 수 있다. 실험계획에서는 관측 값의 개수 n에 확률 질량함수를 곱한 $n\xi(x_i)$, $i = 1, 2, \ldots, s$가 정수이어야 하는 제약조건을 지닌다면 이를 정확실험(exact design), 그렇지 않은 경우를 근사실험(approximate design)이라 한다. 거의 모든 이론적인 최적실험계획분야의 연구는 근사실험을 기반으로 한다.

이 논문은 2010년 중앙대학교 교내 학술연구비 지원을 받아 수행된 연구임.

[1] 교신저자: (456-756) 경기도 안성시 대덕면 내리 산40-1, 중앙대학교 제2캠퍼스 경영학부, 교수.
Email: yik01@cau.ac.kr

식 (1.1)의 모델 $f_j(x)$하에서 특정한 실험점 z의 분산은 식 (1.2)와 같다.

$$v_j(\xi, z) = f_j^T(z) M_j^{-1}(\xi) f_j(z), \tag{1.2}$$

여기서 $M_j(x) = \int_\Omega f_j(x) f_j^T(x) \xi(dx)$로 정보행렬(information matrix)이다. 이 정보행렬은 실험 ξ내에 포함된 정보의 양을 측정하는데 본 연구에서는 비특이(nonsingular)로 가정한다. 대표적인 최적실험계획기준은 $M_j(x)$의 값을 최대화하는 실험계획, D-최적, $z \in \Omega$인 경우 분산의 최대값, $\max_{z\in\Omega} v_j(\xi, z)$을 최소화하는 실험계획을 G-최적, 그리고 $\mathrm{tr}(M^{-1})$를 최소화 하는 실험기준, A-최적 등을 언급할 수 있다. D-최적과 G-최적은 동격이다. 그리고 G-최적 하에서는 실험영역 내에서는 분산의 최대값은 모수의 개수, $m + 1$과 일치한다. Kiefer와 Wolfowitz (1960)와 Fedorov (1972) 등에 의하면 이는 동격(equivalence)을 확인하는 조건이 된다. 그리고 본 연구에서 고려하고 있는 다른 실험기준은 외삽-최적(extrapolation optimal)이다. 즉, 다항회귀모형이 $f_j(x)$ 하에서 z가 실험영역 바깥에 있는 점, $z \notin \Omega$이라면 식 (1.2)의 값을 최소화하는 실험기준이다.

제 2절에서는 본 연구에서 언급할 외삽최적실험의 문제점과 이를 위한 2가지 다중 제약실험에 대해 정의하고 제 3절에서는 예제를 통해 이를 알아보고 마지막으로 제 4절에서는 결론 및 향후 개선방안을 언급한다.

2. 외삽최적실험과 복합 및 최소최대 제약-외삽실험

1절에서 이야기한 기본적인 최적 실험들은 식 (1.1)의 모형과 수반되는 가정을 기반으로 한다. 즉, 모형이 틀린다든지 가정이 올바르지 않은 경우가 밝혀지게 되면 최적실험계획은 최적이 아닌 실험이 되기 때문이다. 특히 이런 가정은 외삽-최적인 경우에는 치명적일 수 있다. 외삽-최적 실험의 해는 이미 60년대 연구되었다. Kiefer와 Wolfowitz (1964), Hoel과 Levine (1964) 등은 다항회귀모형에 대해 외삽최적실험의 해를 구축하였다. 더 나아가 Studden (1971)은 다변다항회귀모형에 대한 해를 연구한 바 있다. 그러나 실험영역 Ω를 벗어나 있는 점 $z \notin \Omega$에 대한 분산을 최소화 하는 실험은 가정한 모형의 타당성을 의심받을 수 있다. 왜냐하면 실험영역을 벗어나도 모형에 가정이 유지가 되어야 하기 때문이다. 이러한 경우를 염두에 둔 실험이 소위 강건(robust) 실험이다. 추후 밝혀진 모형이 가정한 모형과 다르다고 해도 효율 면에서 그렇게 나쁘지 않는 실험을 사전에 강구하는 방법이다. 이를 위한 첫 번째 시도가 Huang과 Studden (1988)에 의해 시도가 되었는데 주어진 모형에서 이탈되는 모형들을 하나의 집합체의 구성원으로 보고 이탈되는 평균제곱오차(MSE)의 최대값을 최소화하는 이론적인 논문을 발표한 바 있다. 그러나 이 논문은 보편적으로 많이 사용되는 차후 정의될 실험의 효율을 기준으로 논문을 전개한 것이 아니다.

본 연구의 단계적 설명을 위해 식 (1.1)의 모형을 식 (2.1)과 같이 분해하여 보자.

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

여기서 모수벡터 $\theta_j^T = (\theta_0, \theta_1, \ldots, \theta_j)$는 $j + 1$차원을 가진다. 식 (2.1)은 식 (1.1)의 불확실성을 염두에 둔 식으로 참의 모형이 식 (1.1)일 수 도 있으나 경우에 따라서는 식 (2.1)의 부분 모형

$$y = f_j^T(x)\theta_j + \epsilon \tag{2.2}$$

이 참일 수가 있을 것이다. 그러나 실험자는 θ_j에 대한 추정에 관심을 기울인다 하더라도 적합결여 검정을 실시하는 경우를 염두에 둔다면 조금 더 포괄적인 모형인 식 (2.1)의 모형을 사용한다.

일반적으로 모형에 대한 불확실성이 존재 하는 경우 문헌에는 3가지 방법이 존재한다. 제약실험(constrained experimental design), 복합실험(compound experimental design), 그리고 최소-최대실험(mini-max experimental design)이 그것들이다. 제약실험은 Stigler (1971), Cook과 Fedorov (1995) 그리고 Lee (1987) 등에 의해 논의가 된 바 있다. 여러 개의 경쟁적인 실험기준의 가중평균을 최적화하는 복합실험은 Läuter (1974)에 의해 제안되었으며 Cook과 Nachtsheim (1982) 등에 의해 연구되었으며 최소-최대실험은 최악의 경우를 최소화시키는 보수적인 방법으로 Wong (1992) 이후에 시도가 많이 되었다. 이들 방법을 위해 실험의 효율에 대해 정의할 필요가 있다.

외삽-효율: 외삽을 위한 실험 ξ의 외삽-효율은 다음과 같이 정의한다.

$$\phi_\xi^z\left(\xi_j^*\right) = \frac{v_j(\xi_j^*, z)}{v_j(\xi, z)},$$

여기서 ξ_j^*는 $z \notin \Omega$에서 $f_j(x)$에 대한 외삽최적실험(optimal extrapolation design)이다. 이 효율은 0과 1 사이의 값으로 역의 값은 최적실험과 같은 효율을 가지기 위해서는 실험 ξ은 이 만큼의 회수만큼 반복되어야 한다는 의미를 가지고 있다. 비슷하게 D-기준을 위한 실험 ξ의 D-효율와 G-효율은 다음과 같이 정의된다.

***D*-효율**: D-최적을 위한 실험 ξ의 D-효율은 다음과 같이 정의한다.

$$\phi_\xi^D\left(\xi_j^*\right) = \left(\frac{|M(\xi)|}{\left|M\left(\xi_j^*\right)\right|}\right)^{j+1},$$

여기서 ξ_j^*는 $f_j(x)$에 대한 D-최적실험(D-optimal design)이다.

***G*-효율**: G-최적을 위한 실험 ξ의 G-효율은 다음과 같이 정의한다.

$$\phi_\xi^G\left(\xi_j^*\right) = \frac{j+1}{\max\limits_{x\in\Omega} f_j^T(x) M_j^{-1}(\xi) f_j(x)},$$

여기서 ξ_j^*는 $f_j(x)$에 대한 G-최적실험(G-optimal design)이다.

위와 같은 효율은 어느 실험기준, ψ이든지 정의될 수 있는데 통일하여 본 연구에서는 편의상 $\phi_\xi^\psi(\xi^*)$라 표기한다. 간단히 설명하면 실험기준 ψ을 기준으로 최적실험, ξ^*에 대해 실험 ξ이 가지는 효율이다.

위에서 언급하였듯이 모형을 2개로 가정하고 두 모형에 대한 최적실험, $\xi_i^*, i = 1,2$에 대해 실험 ξ가 가지는 효율을 $\phi_\xi(\xi_i^*), i = 1,2$라 한다면 최적 제약실험계획은 다음과 같은 제약조건 하에서의 목적함수 값을 최대화 하는 실험 ξ을 찾는 문제가 된다.

$$\max \phi_\xi\left(\xi_2^*\right) \quad \text{subject to} \quad \phi_\xi\left(\xi_1^*\right) \ge c,$$

여기서 첨자 인덱스 1은 실험자가 제일 우선적으로 달성하고자 하는 효율을 염두에 둔 모형을 의미하고 인덱스 2는 부차적인 목적에 해당하는 실험기준을 의미한다. 또한 c는 실험자가 우선적으로 달성하고자 하는 효율 값이다. 통상적으로 인덱스 1에 해당되는 모형은 식 (2.1)의 부분모형이고 인덱스 2에 해당하는 모형이 식 (1.1)이나 식 (2.1)의 완전모형이다. 이러한 제약실험계획은 Lee (1987), Cook과 Fedorov (1995) 등에 의해 연구가 진행되었다. 그러나 제약실험의 제일 큰 약점은 조건이 하나가 아닌

두 개 이상으로 확장되는 경우이다. 즉, 참으로 나타날 수 있는 모형의 개수가 3개 이상인 경우는 모든 조건식의 효율, c_i, $i = 1, \ldots, m$, $m \geq 2$을 모두 다 만족시키지는 못하는 비타당성(infeasibility)의 문제가 발생이 되기 때문이다. 이 문제는 Huang과 Wong (1998)과 강명욱과 김영일 (2002) 등이 언급한 축차적인 방법이거나 김영일과 임용빈 (2007)이 제안하였듯이 최소-최대 방법과 제약조건식이 혼합된 형태로 접근이 가능하다.

복합실험기준은 Läuter (1974)에 의해 처음으로 주장되었으며 다음과 같이 설명된다. 위에서 언급하였듯이 모형을 k개로 가정하고 k 모형에 대한 최적실험, ξ_i^*, $i = 1, 2, \ldots, k$에 대해 실험 ξ가 가지는 효율을 $\phi_\xi(\xi_i^*)$, $i = 1, 2, \ldots, k$라 한다면 복합실험 ξ은 다음과 같이 정의되는 실험이다.

$$\max_{\xi} \alpha_1 \phi_\xi(\xi_1^*) + \alpha_2 \phi_\xi(\xi_2^*) + \cdots + \alpha_k \phi_\xi(\xi_k^*),$$

여기서 α_i는 실험자가 명시하여야 하는 0과 1사이의 값이고 $\sum_i^k \alpha_i = 1.0$이다. 참고로 α_i를 $1/k$로 한다고 하여도 실험 ξ가 k개의 모든 모형에 대한 최적실험에 대해 가지는 효율은 같지 않다. 실험자가 적절한 α_i를 명시한다는 것은 어렵다. 그러나 제약실험과 달리 3개 이상의 모형을 염두에 두더라도 가능한 기준이다. 특수하게 Dette와 Huang (2000)의 논문은 다항모형의 차수가 m과 $2m$인 경우 외삽 복합실험을 다루어 보았다.

또한 Cook과 Wong (1994)에 의해 복합실험계획은 적절한 제약조건이 형성되어 있으면 대략적으로 제약실험계획의 하나의 해(solution)로 구현된다. 두 실험기준의 동격성에 관한 자세한 내용은 그들의 논문을 참조하기 바란다.

마지막으로 최소-최대(maxi-min) 실험은 다음과 같이 설명된다. k개의 모형에 대한 최적실험, ξ_i^*, $i = 1, 2, \ldots, k$에 대해 실험 ξ가 가지는 효율을 $\phi_\xi(\xi_i^*)$, $i = 1, 2, \ldots, k$라 한다면 최소-최대실험 ξ은 다음과 같이 정의된다.

$$\max_{\xi} \min_{i} \phi_\xi(\xi_i^*), \quad i = 1, \ldots, k.$$

이 실험은 매우 보수적인 방법으로 모형이 $k = 2$개인 경우는 Imhof와 Wong (2000)이 주장하였듯이 $\phi_\xi(\xi_1^*) = \phi_\xi(\xi_2^*)$가 되어 버린다. 이 방법은 위의 두 방법에 비해 실험자가 확신하는 모형에 대한 가중치를 부여하지 못하는 단점이 있다.

외삽-최적실험은 이미 Hoel과 Levine (1964)에 의해 사전에 다항회귀모형 $f_m(x)$을 알고 있다는 가정 하에서 실험영역을 벗어난 점, $z \notin \Omega$에 대한 외삽-최적 실험 ξ_m^*의 해를 제공한바 있다. 외삽-최적의 실험은 $m + 1$개의 받힘점에서 질량을 가지며 이 실험점은 다항식 $(1 - x^2)U_{m-1}(x)$의 근(roots)으로 주어진다. 여기서

$$U_{m-1}(x) = \frac{\sin\{m \arccos x\}}{\sin\{\arccos x\}}$$

는 2종(second kind)의 $(m - 1)$번째 Chebyshev 다항식이다. 그리고 $L_j(x)$를 j번째 받힘점에서 라그랑지 내삽 다항식(Lagrange interpolating polynomial)이라 한다면 j번째 받힘 점에서의 질량은 다음과 같이 주어진다.

$$\frac{|L_j(z)|}{\sum_{i=0}^m |L_i(z)|}, \quad (j = 0, 1, \ldots, m).$$

그러나 위에서 언급하였듯이 이러한 해는 $f_m(x)$이 실험영역을 벋어나더라도 타당한 경우여야만 한다. 그러나 $z \notin \Omega$가 Ω를 많이 벗어난 위치에 있다면 이러한 가정은 현실적으로 불안정해진다. 예를 들어

$\Omega = [-1, 1]$이고 $z = 1.1$에서 $z = 2.0$으로 멀어진다면 주어진 모형이 계속 유지가 된다는 것은 어렵다. Dette와 Wong (1996)은 주어진 다항회귀모형의 차수가 조금만 이탈하여도 $z \notin \Omega$가 실험영역을 멀리 벗어나면 날수록 외삽-효율이 급속히 나빠짐을 보인바 있다. 따라서 Kussmaul (1969)의 높은 차수를 염두에 둔 실험 방법은 외삽문제인 경우 소용이 없다. Wong (1995)는 외삽-최적만의 문제에서 벗어나 외삽-최적과 D-최적(혹은 A-최적)으로 구성된 복합실험을 고려하여 보았다. 그러나 고려된 D-최적은 주어진 모형에 대한 최적실험기준이라 모형이 불확실한 경우는 다른 실험기준이 필요로 하다. 따라서 Dette와 Wong (1996)이 여러 모형에 대한 외삽-최적들의 복합실험기준을 발표하고 이론적인 해를 구축하였다. 즉, 다음과 같은 목적함수를 최대화 하는 실험기준에 대한 이론적인 해를 구한 바 있다.

$$\alpha_1\phi_\xi^z\left(\xi_j^*\right) + \alpha_2\phi_\xi^z\left(\xi_{j+1}^*\right) + \cdots + \alpha_{k-j+1}\phi_\xi^z\left(\xi_k^*\right),$$

여기서 k는 주어진 모형 $f_j(x)$ 보다는 차수가 높은 $f_k(x)$을 위한 첨자이다. 일반적으로 k는 j보다 1이나 2가 큰 숫자이다. 그러나 이러한 실험기준은 외삽만을 위한 실험이므로 모형이 불확실한 경우에는 원래 Wong (1995)이 제안하였듯이 외삽-최적만의 문제에서 벗어나 외삽-최적과 $D(G)$-최적과의 복합실험을 강건성(robustness)을 고려하여 다음과 같이 제안한다.

정의 1. 다음과 같은 실험 ξ을 모형 $f_j(x)$에 대한 복합-제약 외삽실험(compound-restricted extrapolation design)이라 한다.

$$\max_\xi \alpha_1\phi_\xi^z\left(\xi_j^*\right) + \alpha_2\phi_\xi^z\left(\xi_{j+1}^*\right) + \cdots + \alpha_{k-j+1}\phi_\xi^z\left(\xi_k^*\right)$$
$$\text{subject to } \alpha_1\phi_\xi^D\left(\xi_j^*\right) + \alpha_2\phi_\xi^D\left(\xi_{j+1}^*\right) + \cdots + \alpha_{k-j+1}\phi_\xi^D\left(\xi_k^*\right) \geq c.$$

정의 2. 다음과 같은 실험 ξ을 모형 $f_j(x)$에 대한 최소최대-제약 외삽실험(maximin-restricted extrapolation design)이라 한다.

$$\min_\xi \max_l \phi_\xi^z\left(\xi_l^*\right), \quad l = j, \ldots, k$$
$$\text{subject to } \max_\xi \min_l \phi_\xi^D\left(\xi_l^*\right) \geq c, \quad l = j, \ldots, k.$$

정의 1과 2의 제약조건식에서 나오는 우변의 값 c는 실험자가 설정하여야 하는데 실험기준이 2개뿐인 일반적인 제약조건 실험의 상수 c, $0 \leq c \leq 1$와 같이 설정하면 안 된다. 이를 설정하기 위해서는 제약조건식에 설정된 실험기준의 최대값을 미리 확보하여야 한다. 그리고 이 값보다 작게 c를 설정하여야 한다. 이는 3절에서 예를 통해 설명하도록 한다.

두 정의의 제약조건에 나오는 실험기준은 D-최적으로 설정하였지만 위에서 언급하였듯이 D-최적과 G-최적은 동격이라고 하였기 때문 G-효율로도 설정할 수 있다. 또한 Atwood (1969)는 주어진 모형에 대해 G-효율은 D-효율의 하한 값을 제공한다고 하였으므로 D-효율을 직접 언급하는 것보다는 G-효율을 명시하여 D-효율을 올리는 방법도 생각하여 볼 수 있다. 그러나 본 연구에서는 D-효율만을 기준으로 한다.

정의 1과 정의 2의 실험은 문헌상에서 나오는 강건 실험을 강구하는 3가지 방법을 융합한 실험기준이라 볼 수 있는데 외삽문제를 다룰 때는 외삽문제 자체뿐 아니라 모형에 관한 불확실성을 고려하여야 하기 때문이다. 따라서 전통적인 교환알고리즘(exchange algorithm)은 적용키 어렵다고 판단된다. 따라서 논 연구에서는 유전 알고리즘을 채택하여 그 해를 구하였다. 유전알고리즘은 이미 그 유용성을 검증받은바 있다. Yum과 Nam (2000), Park 등 (2005)을 참고하기 바란다. 참고적으로 본 연구에서는 임의로 교배율과 돌연변이율을 각각 0.10, 0.05로 설정하였다. 제 3절에서는 정의 1과 정의 2에서 언급된 실험에 대한 예제를 $z > 1$, $z = 1.1, 2, 4$, $\Omega = [-1, 1]$을 중심으로 알아보도록 한다. 이 세 개의 z점의 선택은 Dette와 Wong (1996)의 논문을 따른 것이다.

표 1: 단순회귀모형 $D(G)$-최적 및 외삽-최적

$D(G)$-최적	$\xi(-1) = \xi(1) = 0.5$
외삽최적, $z = 1.1$	$\xi(-1) = 0.0455$, $\xi(1) = 0.9545$
외삽최적, $z = 2$	$\xi(-1) = 0.25$, $\xi(1) = 0.75$
외삽최적, $z = 4$	$\xi(-1) = 0.375$, $\xi(1) = 0.625$

표 2: 2차 회귀모형 $D(G)$-최적 및 외삽-최적

$D(G)$-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, $z = 2$	$\xi(-1) = 0.1429$, $\xi(0) = \xi(1) = 0.4286$
외삽최적, $z = 4$	$\xi(-1) = 0.1935$, $\xi(0) = 0.4839$, $\xi(1) = 0.3226$

표 3: 단순회귀와 2차 회귀모형 D-최적 복합실험

α	ξ
0.2	$\xi(0) = 0.3018$, $\xi(\pm 1) = 0.3491$
0.5	$\xi(0) = 0.2296$, $\xi(\pm 1) = 0.3852$
0.8	$\xi(0) = 0.1005$, $\xi(\pm 1) = 0.4497$

표 4: 설정된 모형에 대해 가지는 효율 및 최대 복합효율

α	$\phi_\xi^D(\xi_1^*)$	$\phi_\xi^D(\xi_2^*)$	$\alpha_1\phi_\xi^G(\xi_1^*) + (1-\alpha)\phi_\xi^G(\xi_2^*)$
0.2	0.8356	0.9976	0.9653
0.5	0.8776	0.9726	0.9251
0.8	0.8188	0.9484	0.9225

3. 예제

본 연구에서 고려하는 모형의 개수는 2개로 단순화한다. 실험자는 단순회귀모형과 2차 다항회귀모형 중 어느 모형이 참의 모형이 될지 모르는 상태이다. 이러한 설정은 Dette와 Wong (1996)의 논문에서 벤치마크 기준으로 사용되었다. 그리고 $z \notin \Omega$는 $z = 1.1, 2, 4$의 3가지 값을 고려한다. 2절에서 정의한 두 개의 새로운 실험에 대한 해를 구하기전에 이해에 도움을 주는 몇 가지 실험을 예제 1부터 예제 4에서 명기하여 본다.

예제 1. 단순회귀모형에 대해 $D(G)$-최적과 외삽-최적 실험.
실험영역 $\Omega = [-1, 1]$에서 단순회귀모형에 대해 $D(G)$-최적과 외삽-최적 실험은 표 1과 같다. $z \notin \Omega$가 1에 가까우면 가까울수록 $\xi(1)$의 값이 커짐을 알 수 있다. $z \notin \Omega$가 1에서 멀어 질수록 실험은 $D(G)$-최적, $\xi(\pm 1) = 0.5$에 가까워진다.

예제 2. 2차 회귀모형에 대해 $D(G)$-최적과 외삽-최적 실험.
실험영역 $\Omega = [-1, 1]$에서 2차 회귀모형에 대해 $D(G)$-최적과 외삽-최적 실험은 표 2와 같다.

예제 3. 단순회귀모형과 2차 회귀모형에 대해 D-최적 복합실험.
복합실험기준 $\alpha_1\phi_\xi^D(\xi_1^*) + (1-\alpha)\phi_\xi^D(\xi_2^*)$을 최대화하는 실험기준의 실험은 표 3과 같다. 첫 번째 인덱스는 단순 회귀모형 두 번째 인덱스는 2차 회귀모형이다. 그리고 설정된 모형에 가지는 효율 및 구해진 최대 복합효율이다. 후에 나오는 c의 값은 표 4의 최대 복합효율보다 낮게 설정된다.

예제 4. 단순회귀모형과 2차 회귀모형에 대해 D-최적 최소최대실험.

표 5: 단순회귀모형과 2차 회귀모형에 대해 D-최적 최소최대실험

ξ	$\phi^D_\xi(\xi^*_1)$	$\phi^D_\xi(\xi^*_2)$
$\xi(\pm1) = 0.4191$, $\xi(0) = 0.1618$	0.9155	0.9155

표 6: 복합-제약 외삽실험 $\alpha = 0.2$

z	ξ
1.1	$\xi(-1) = 0.1653$, $\xi(-0.0747) = 0.1828$, $\xi(1) = 0.6519$
2	$\xi(-1) = 0.1504$, $\xi(0) = 0.3987$, $\xi(1) = 0.4509$
4	$\xi(-1) = 0.2052$, $\xi(0) = 0.4528$, $\xi(1) = 0.3419$

표 7: 복합-제약 외삽실험 $\alpha = 0.5$

z	ξ
1.1	$\xi(-1) = 0.1761$, $\xi(-0.06281) = 0.1797$, $\xi(1) = 0.6442$
2	$\xi(-1) = 0.1702$, $\xi(0.0003) = 0.317$, $\xi(1) = 0.5119$
4	$\xi(-1) = 0.2392$, $\xi(0.0003) = 0.3619$, $\xi(1) = 0.3989$

표 8: 복합-제약 외삽실험 $\alpha = 0.8$

z	ξ
1.1	$\xi(-1) = 0.2158$, $\xi(-0.0694) = 0.0787$, $\xi(1) = 0.7055$
2	$\xi(-1) = 0.2296$, $\xi(-0.0003) = 0.0872$, $\xi(1) = 0.6887$
4	$\xi(-1) = 0.3646$, $\xi(0) = 0.0280$, $\xi(1) = 0.6075$

표 9: 각 모형에 대해 복합제약실험이 가지는 D-효율과 외삽 효율, $\alpha = 0.2$

z	$\phi^D_\xi(\xi^*_1)$	$\phi^D_\xi(\xi^*_2)$	$\phi^z_\xi(\xi^*_1)$	$\phi^z_\xi(\xi^*_2)$
1.1	0.7711	0.8072	0.7283	0.9005
2	0.7148	0.9004	0.6013	0.9963
4	0.7270	0.9502	0.5472	0.9961

표 10: 각 모형에 대해 복합제약실험이 가지는 D-효율과 외삽 효율, $\alpha = 0.5$

z	$\phi^D_\xi(\xi^*_1)$	$\phi^D_\xi(\xi^*_2)$	$\phi^z_\xi(\xi^*_1)$	$\phi^z_\xi(\xi^*_2)$
1.1	0.7825	0.8173	0.7221	0.8893
2	0.7514	0.9068	0.6821	0.9450
4	0.7827	0.9769	0.6381	0.9395

또한 $\min_\xi \max_l \phi^D_\xi(\xi^*_l)$, $l = 1, 2$를 기준으로 한다면 실험 및 각 모형에 대한 효율은 표 5와 같다. 모형이 두 개인 경우는 각 모형에 대한 효율은 일치한다. 후에 나오는 c의 값은 표 5의 0.9155보다 낮게 설정된다. 2절에서 제안한 실험을 살펴본다.

예제 5. 복합-제약 외삽실험.

제약조건에 부여되는 상수 c는 표 4에서 언급한 D-최적 복합실험의 복합 효율보다 낮은 0.8에서 임의로 설정하였다. $\alpha = 0.2, 0.5, 0.8$에 따른 복합-제약 외삽실험은 각각 표 6, 표 7, 그리고 표 8과 같다. 표 9, 표 10, 그리고 표 11은 설정된 모형에 대해 가지는 효율이다.

가운데 받힘점은 더 이상 0이 아니다. 이는 Wong (1995)이 지적하였듯이 외삽-최적과 D-최적이 복합적으로 빚어내는 현상이다. α의 값이 지나치게 한쪽 모형으로 치우쳐 있는 경우 $z > 1$점이 1에서

표 11: 각 모형에 대해 복합제약실험이 가지는 D-효율과 외삽 효율, $\alpha = 0.8$

z	$\phi_{\xi}^{D}(\xi_1^*)$	$\phi_{\xi}^{D}(\xi_2^*)$	$\phi_{\xi}^{z}(\xi_1^*)$	$\phi_{\xi}^{z}(\xi_2^*)$
1.1	0.8297	0.6843	0.7798	0.8404
2	0.8409	0.7175	0.9128	0.4058
4	0.9554	0.5512	0.9720	0.1158

표 12: 최소최대-제약 외삽실험, $c = 0.8$

z	ξ
1.1	$\xi(-1) = 0.1984,\ \xi(-0.0331) = 0.2149,\ \xi(1) = 0.5868$
2	$\xi(-1) = 0.2019,\ \xi(-0.0009) = 0.2142,\ \xi(1) = 0.5839$
4	$\xi(-1) = 0.2818,\ \xi(-0.0194) = 0.2422,\ \xi(1) = 0.4760$

표 13: 최소최대-제약 외삽실험, $c = 0.9$

z	ξ
1.1	$\xi(-1) = 0.3770,\ \xi(-0.0873) = 0.1900,\ \xi(1) = 0.4323$
2	$\xi(-1) = 0.3781,\ \xi(-0.0769) = 0.1859,\ \xi(1) = 0.4324$
4	$\xi(-1) = 0.3865,\ \xi(-0.0336) = 0.1892,\ \xi(1) = 0.4243$

표 14: 각 모형에 대해 최소최대-제약 외삽실험이 가지는 D-효율과 외삽-효율, $c = 0.8$

z	$\phi_{\xi}^{D}(\xi_1^*)$	$\phi_{\xi}^{D}(\xi_2^*)$	$\phi_{\xi}^{z}(\xi_1^*)$	$\phi_{\xi}^{z}(\xi_2^*)$
1.1	0.8	0.8767	0.7852	0.7852
2	0.8	0.8802	0.7852	0.7852
4	0.8497	0.9570	0.7579	0.7579

표 15: 각 모형에 대해 최소최대-제약 외삽실험이 가지는 D-효율과 외삽-효율, $c = 0.9$

z	$\phi_{\xi}^{D}(\xi_1^*)$	$\phi_{\xi}^{D}(\xi_2^*)$	$\phi_{\xi}^{z}(\xi_1^*)$	$\phi_{\xi}^{z}(\xi_2^*)$
1.1	0.9	0.9338	0.6954	0.6927
2	0.9	0.9309	0.6964	0.6926
4	0.9	0.9420	0.7827	0.6333

이탈되면 될수록 각 모형에 대한 효율이 낮아지는 경향이 있다. 따라서 실험자가 모형에 대한 확신이 없는 경우 α의 값을 0.5로 하던지 혹은 본 연구에서 제안한 두 번째 실험기준이 더 나을 수 있을 것이다.

예제 6. 최소최대-제약 외삽실험.

제약조건에 부여되는 상수 c는 표 5에서 언급한 최대최소실험에서 나온 효율, 0.9155보다 작은 값으로 0.8과 0.9를 임의로 두 값을 설정하였다. 표 12와 표 13이 $c = 0.8$, $c = 0.9$에 해당되는 최소최대-제약 외삽실험이다. 표 14, 그리고 표 15는 설정된 모형에 대해 가지는 효율이다.

그리고 위 표 14와 표 15를 비교하여 보면 높은 c값은 D-효율을 높이고자 함이므로 외삽-효율이 낮아진다는 의미이다. 이와 같은 논리는 지면상 생략되었지만 D-제약 외삽 복합 실험에서도 적용이 된다. 또한 표 14에서 보다시피 낮은 c값에서 최소최대-제약 외삽실험 하에서는 두 개의 외삽-효율은 기대했던 대로 동일하다. 복합-제약 외삽실험과 비교하면 각 모형에 대해 4가지 효율이 어느 정도 안정되어 있음을 알 수 있을 것이다. 최소최대-제약 외삽실험 $(z = 4, c = 0.8)$에서 나온 값 0.8497, 0.9570, 0.7579, 0.7579 (표 14의 마지막 줄)를 비교대상이 되는 $(\alpha = 0.5, z = 4, c = 0.8)$일 때 복합-제약 외삽실험 하에서 나온 값, 0.7827, 0.9769, 0.6381, 0.9395 (표 10 마지막 줄)과 비교하면 최소최대-제약 외삽실험의 효율이 더 안정적임을 알 수 있다.

4. 결론

본 연구에서는 외삽-최적에서 나타나는 문제점을 파악하고 새로운 실험기준 2가지를 제안하였다. 외삽-최적은 실험영역 Ω의 점의 분산을 염두에 둔 실험기준이 아니기 때문에 항상 모형에 대한 불확실성이 대두되어 왔다. $z \notin \Omega$가 Ω에 아주 근접되어 있다면 모르겠으나 실험영역에서 멀어지면 멀어질수록 이러한 문제점은 더 대두가 된다. 본 연구는 Dette와 Wong (1996)이 시도한 복합실험기준의 새로운 시도를 확장하여 보았다. 이러한 시도는 외삽-최적뿐 아니라 모형에 대한 불확실성도 감안하기 위한 조치로 이해를 할 수 있다. 실험적인 벤치마킹 모형인 단순회귀모형과 이차회귀모형에서 나타나 결과만 가지고 제안된 실험기준을 평가하기는 이르다고 볼 수 있으나 본 연구에서 나타난 결과는 실험자의 의도를 반영하는 것으로 판단된다. 또한 사용된 알고리즘은 유전 알고리즘으로 정확실험(exact design)을 요하는 실험에서는 개선의 여지가 보인다. 추후 연구에서는 실험자가 여러 개의 z점들에서의 분산에 관심을 가지고 있는 다중목적 실험인 경우를 강명욱과 김영일 (2002)이 제안한 순차적인 방법을 이용하여 실험자에게 가이드라인을 줄 수 있는 연구결과도 생각하여 볼 수 있을 것이다.

참고 문헌

강명욱, 김영일 (2002). Multiple constrained optimal experimental design, *The Korean Communications in Statistics*, **9**, 619–627.

김영일, 임용빈 (2007). Hybrid approach multiple objective experimental design, *The Korean Communications in Statistics*, **9**, 619–627.

Atwood, C. L. (1969). Optimal and efficient designs of experiments, *The Annals of Mathematical Statistics*, **40**, 1570–1602.

Box, G. E. P. and Draper, N. R. (1975). Robust design, *Biometrika*, **62**, 347–352.

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

Cook, R. D. and Nachtsheim, C. J. (1982). Model-robust, linear optimal designs, *Technometrics*, **24**, 49–54.

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.

Dette, H. and Huang, M. (2000). Convex optimal designs for compound polynomial extrapolation, *Annals of the Institute of Statistical Mathematics*, **52**, 557–573.

Dette, H. and Wong, W. K. (1996). Robust optimal extrapolation designs, *Biometrika*, **83**, 667–680.

Fedorov, V. V. (1972). *Theory of Optimal Experiments*, Academic Press, Inc, New York.

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.

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

Kiefer, J. and Wolfowitz, J. (1960). The equivalence of two extremum problems, *Canadian Journal of Mathematics*, **12**, 363–366.

Kiefer, J. and Wolfowitz, J. (1964). Optimum extrapolation and interpolation designs, *Annals of the Institute of Statistical Mathematics*, **16**, 79–108.

Kussmaul, K. (1969). Protection against assuming the wrong degree in polynomial regression, *Technometrics*, **11**, 677–682.

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

Lee, C. M. S. (1987). Constrained optimal designs for regression models, *Communications in Statistics, Part A-theory and Methods*, **16**, 765–783.

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.

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. (1992). A unified approach to the construction of minimax designs, *Biometrika*, **79**, 611–619.

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.

Yum, J. K. and Nam, K. S. (2000). A study on D-optimal design using the genetic algorithm, *The Korean Communications in Statistics*, **7**, 357–366.

2011년 8월 접수; 2011년 10월 채택

Hybrid Constrained Extrapolation Experimental Design

Young Il Kim[1,a], DaeHeung Jang[b]

[a]Division of Business Administration, ChungAng University
[b]Department of Statistics, PuKyung University

Abstract

In setting an experimental design for the prediction outside the experimental region (extrapolation design), it is natural for the experimenter to be very careful about the validity of the model for the design because the experimenter is not certain whether the model can be extended beyond the design region or not. In this paper, a hybrid constrained type approach was adopted in dealing model uncertainty as well as the prediction error using the three basic principles available in literature, maxi-min, constrained, and compound design. Furthermore, the effect of the distance of the extrapolation design point from the design region is investigated. A search algorithm was used because the classical exchange algorithm was found to be complex due to the characteristic of the problem.

Keywords: Extrapolation design, model uncertainty, constrained design, compound design, maxi-min design, genetic algorithm.

This research was supported by ChungAng University's 2010 research fund.

[1] Corresponding author: Professor, Division of Business Administration, ChungAng University, KyungGi-Do 456-756, Korea. E-mail : yik01@cau.ac.kr