

단순선형회귀와 이차형식회귀모형을 중심으로
 D -와 이분산 G -최적에 비교한 I -최적실험기준의
특성연구

Characteristics of I -optimality Criterion compared to the
 D - and Heteroscedastic G -optimality with respect to
Simple Linear and Quadratic Regression

김 영 일*

ABSTRACT

The characteristics of I_λ -optimality, one of the linear criteria suggested by Fedorov (1972) are investigated with respect to the D -and heteroscedastic G -optimality in case of non-constant variance function.

Though having limited results obtained from simple models, we may conclude that I_λ -optimality is sometimes preferred to the heteroscedastic G -optimality suggested newly by Wong and Cook (1992) in the sense that the experimenter's belief in weighting function exists in I_λ -optimality criterion, not to mention its computational simplicity.

* 중앙대학교 산업대학 산업정보학과

1. 소개 및 표현방법

특별한 경우를 명시하지 않는 한 다음의 선형모형을 간주하고자 한다.

$$y_i = f^T(x_i)\theta + \epsilon_i, \quad i = 1, 2, \dots, N$$

..... (1)

x_i 는 예측변수 $q \times 1$ 벡터이고 $f^T(x_i)$ 는 가정된 반응함수의 형태에 의존되는 $p \times 1$ 벡터이며 ϵ_i 는 평균 0 공분산 행렬 $\sigma^2 V$ 을 따르는데, V 는 원소가 v_{ij}, \dots, v_{nn} 로 되어있는 대각행렬이다. 일반성의 손실없이 $\sigma^2 = 1$ 을 채택한다. 또한 개개의 v_{ij} 는 x_i 의 함수로서 가정을 하고, $v_{ij} = w^{-1}(x_i)$ 로 표기한다. 앞으로는 $w(x)$ 와 $w^{-1}(x)$ 는 각각 효율함수 및 오차함수를 불리도록 한다. 실험계획문제는 실험계획영역, X 에서 벡터 $x_i, i = 1, 2, \dots, N$ 를 선택하는 문제이다. 그리고 어떤 주어진 설정된 의미에서 이러한 N 벡터로서 형성된 실험계획이 최적일 수 있게끔 하는 것이 최적실험이다. 이러한 아이디어는 쉽게 확률을 이용한 실험계획으로 연장할 수 있다. 실험계획문제는 (f, X, w) 에 대한 확률적인 실험계획 $\xi \in E_x$ 의 선택이라 할 수 있다. 식(1)에 관련하여 다음과 같은 정보행렬을 정의할 수 있다.

$$M(\xi) = \int_X w(x) f(x) f^T(x) d\xi(x)$$

정의 1.1 실험계획 ξ_D 는 아래 조건이 만족할 때 D -최적이라 한다.

$$\max_{\xi \in E_x} |M(\xi)| = |M(\xi_D)|.$$

정의 1.2 실험계획 ξ_G 는 아래 조건이 만족할

때 G -최적이라 한다.

$$\min_{\xi \in E_x} \max_{x \in X} d(x, \xi) = \max_{x \in X} d(x, \xi_G).$$

여기서 $d(x, \xi) = f^T(x)D(\xi)f(x)$ 는 실험계획 ξ 의 주어진 x 값에서의 예측치의 분산의 함수이다. 그리고 $D(\xi)$ 는 $M(\xi)$ 의 역행렬이 존재할 때 이를 뜻한다.

최적실험계획에서는 이러한 D 및 G -기준이 많이 쓰이고 있는데, 이의 한 이유로서 등분산 오차함수인 ($w(x)=1$)인 경우 이 두 기준의 동치를 들 수 있다. 그러나 이분산오차함수인 경우에는 두 기준의 동치정리가 성립되지 않는다. 최근에 Wong과 Cook(1992)과 Wong(1992)은 위에서 제시한 G -최적성을 위시한 일반적인 Mini-Max 접근방법을 이용한 실험계획을 연구하였고 이를 확인할 수 있는 필요충분 조건을 구하였다. 그러나 이분산오차함수의 경우 G -최적성을 구하는 문제는 일반적으로 D -최적성보다 복잡한 절차를 통하는 것이 보통이다. 자세한 내용은 이들 논문을 참조바란다. 그러나 그들은 아래 제시되는 I_L -최적기준 역시 이분산오차 함수하에서의 실험계획에 도움이 된다는 생각은 간과한 것 같다.

Fedorov(1972)는 $D(\xi)f(x)f^T(x)D(\xi)$ 의 선형함수치(Linear Operational : L)를 최소화하는 선형최적기준 및 이의 동치정리를 제시하였다.

정리 1.1 $\phi(x, \xi) = w(x)L[D(\xi)f(x)f^T(x)D(\xi)]$. 다음의 각 사항은 동치이다.

1. ξ_L 은 $L(D(\xi))$ 을 최소화 한다.

2. ξ_L 은 $\phi(x, \xi)$ 의 최대치를 최소화 한다.
3. $\phi(x, \xi_L)$ 의 최대치는 $L[D(\xi_L)]$ 이다.

선형최적기준의 한 예로서 I_λ -최적을 들수 있다.

정의 1.3 실험계획 ξ_{I_λ} 는 아래 조건이 만족할 때 I_λ -최적이라 한다.

$$\min_{\xi \in E_\lambda} \int_{\Omega} d(x, \xi) d\lambda(x) = \int_{\Omega} d(x, \xi_{I_\lambda}) d\lambda(x).$$

여기서 Ω 는 실험계획자가 λ 라는 가중함수를 설정할 수 있는 영역을 말한다. 그러나 꼭 Ω 가 실험계획영역인 \mathcal{X} 와 일치할 필요는 없다. 이러한 의미에서 이 I_λ -최적은 내삽법(Interpolation) 및 외삽법(Extrapolation) 문제를 연구하는데 적절한 방법일 뿐 아니라 가중치의 함수를 이용할 수 있기 때문에 실험자의 의도를 충분히 감안할 수 있는 장점이 있다. 위의 정의에서 제시된 최적기준은 쉽게 L -최적의 하나의 형태임을 알 수 있다. 왜냐하면 I_λ -최적은 $Tr[D(\xi)\Lambda]$, $\Lambda = \int_{\Omega} f(x)f^T(x)d\lambda(x)$ 을 최소화하는 기준이기 때문이다. 즉 $L[D(\xi)] = Tr[D(\xi)\Lambda]$ 이다.

등분산의 오차함수인 경우는 Studden (1977) 및 Cook과 Nachtsheim (1982)의 연구가 있다. 유의할 사항은 등분산오차함수가 아닐 때에도 Fedorov가 제시한 L -최적성은 그 의미를 유지하기 때문에 이분산성의 오차함수 일때에는 G -최적성보다 실험자의 의도를 반영할 수 있고 최적성을 확인할 수 있는 절차가 G -최적성보다 간편한 I_λ -최적이 더 선호될 수 있다. 이러한 맥락에서 다음 두절에서는 I_λ -최적에 대한 유용성을 단순선형 및 이차형식의 회귀모형을 중심으로 따져보고자 한다. 물론

오차함수가 등분산의 형태를 제공하지 못할 경우를 포함한다.

2. 등분산오차함수하에서의 I_λ -최적기준

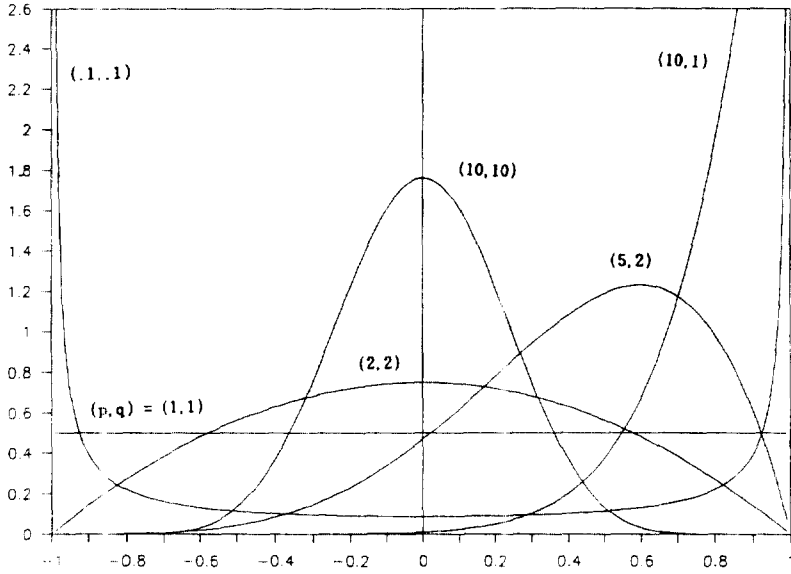
실험계획자가 예측치의 분산함수, $d(x, \xi)$ 의 가중치를 반영하기 위한 함수로 다양한 함수형태를 제시하는 Beta(p, q) 함수를 생각할 수 있다.

$$\begin{aligned} \lambda'_{p,q}(x) &= 1/(b-a)^{p+q-1} \Gamma(p, q) / \Gamma(p) \Gamma(q) \\ &\quad (x-a)^{p-1} (b-x)^{q-1}; \\ &\quad a < x < b, p, q > 0 \end{aligned}$$

이렇게 임의의 구간으로 설정된 Beta함수는 내삽법 및 외삽법을 가능케 한다. 또한 Studden (1977)는 다항회귀모형에서 I_λ -최적성을 구하는 근사적인 방법을 $(1-x^2)\lambda'_{p,q}(x)$ 에 직교하는 다항식으로 구하고자 하였는데 λ'_x 가 Beta함수인 경우는 이를 Jacobi다항식, $P_n^{p,q}(x)$ ($p, q > 1$)으로 표시할 수가 있기 때문에 여기서는 Beta함수만 가중함수로 고려키로 한다. 다양한 p, q 를 갖는 Beta함수를 구간 -1 과 1 에서 그림 2.1에 그려보았다.

2.1 단순선형회귀모형

정리 2.1 : 위에서 정의한 Beta함수와 $\mathcal{X} = [-1, 1]$, $\Omega = [a, b]$ 및 $f^T = (1, x)$ 그리고 등분산의 오차함수인 경우는 $\xi(-1) = d$ 는 이차방정식 $4\mu_1 d^2 + (2-4\mu_1 + 2\mu_2)d + (2\mu_1 - \mu_2 - 1) = 0$ 의 실근중 $0 < d < 1$ 에 있는 d 로 결정이 되고 $\xi(-1) = d$, $\xi(1) = 1-d$ 는 이 경우 I_λ -최적이다. 여기서 μ_i 는 가중치함수로 결정되



다양한 Bata(p, q)함수

<그림 2.1>

는 중심(Central) i 차 적률임을 뜻한다.

증명 : 정리 2. 1을 포함한 아래 모든 정리는 Fedorov의 동치정리를 이용하면 된다.

동치정리 2번째 조건식은 이 경우 $\chi = [-1, 1]$ 에서 아래로 볼록한 2차 다항함수인 관계로 -1 과 1 에서의 극치만 관심을 둔다고 생각하면 $\phi(1, \xi) = \phi(-1, \xi)$ 을 만족시키는 d 가 주어진 d 임을 알 수 있다.

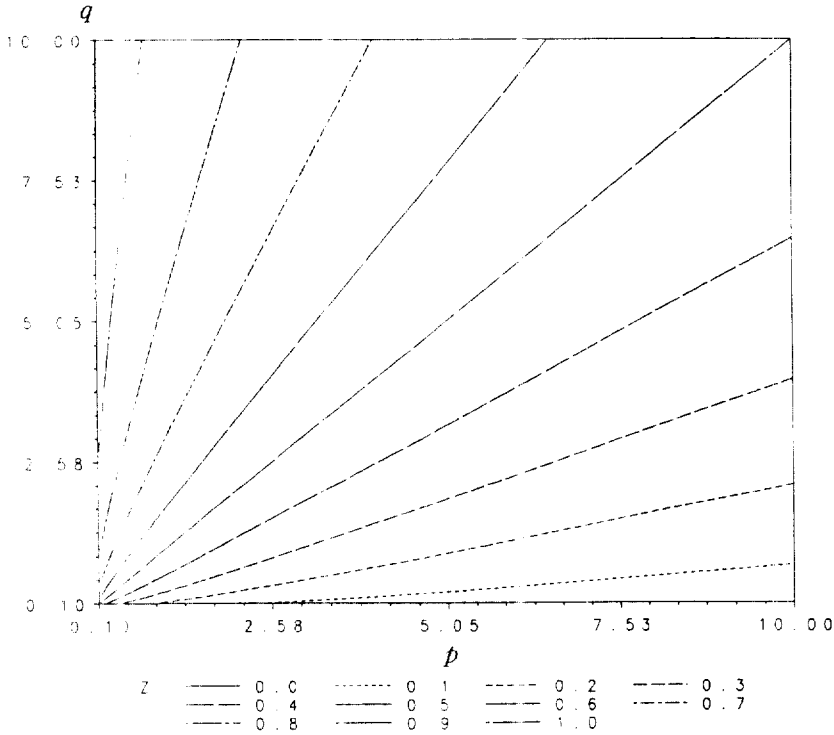
보조정리 2. 1 : 위에서 정의한 Bata함수 ($p = q$)와 $\chi = [-1, 1]$, $\Omega = [-g, g]$ 및 $f^T = (1, x)$ 그리고 등분산오차함수인 경우는 D -최적실험방법, $\xi(\pm 1) = 0.5$ 는 I_1 -최적이다.

예제 1 : Bata함수의 p 및 q 가 변할때 $\chi =$

$[-1, 1]$, $\Omega = [-1, 1]$ 및 $f^T = (1, x)$ 그리고 등분산의 오차함수인 경우 어떻게 $\xi(-1)$ 이 변하는지 p, q 를 0.1 에서 10 까지 변화시키면서 등고선을 그림 2. 2에 그려보았다.

기대하였던대로 $p = q$ 인 경우는 $\xi(-1)$ 의 값은 보조정리 2. 1에 의해 0.5 이고 $p > q$ 인 경우는 실험계획 우측에 가중치를 두는 관계로 $\xi(-1)$ 의 값이 적어지고 $p < q$ 인 경우는 $\xi(1)$ 의 값이 적어진다. Bata확률함수의 균형성에 의해 $\xi(-1)$ 값과 $\xi(1)$ 값은 $p = q$ 을 기준으로 대칭임을 알 수 있다.

다음 예제들에서는 위에서 언급한 내삼법 및 외삼법문제를 $\Omega = [a, b]$ 의 구간을 $[0, g]$, $[1, g]$, $[1-g, 1]$ 및 $[-g, g]$ 으로 나누어 고려키로 하되 $[p, q]$ 의 값은 $[1, 1]$, $[.1, .1]$,



p, q 변화에 따른 $x = -1$ 에서의 질량등고선(등분산오차함수의 단순선형회귀모형)

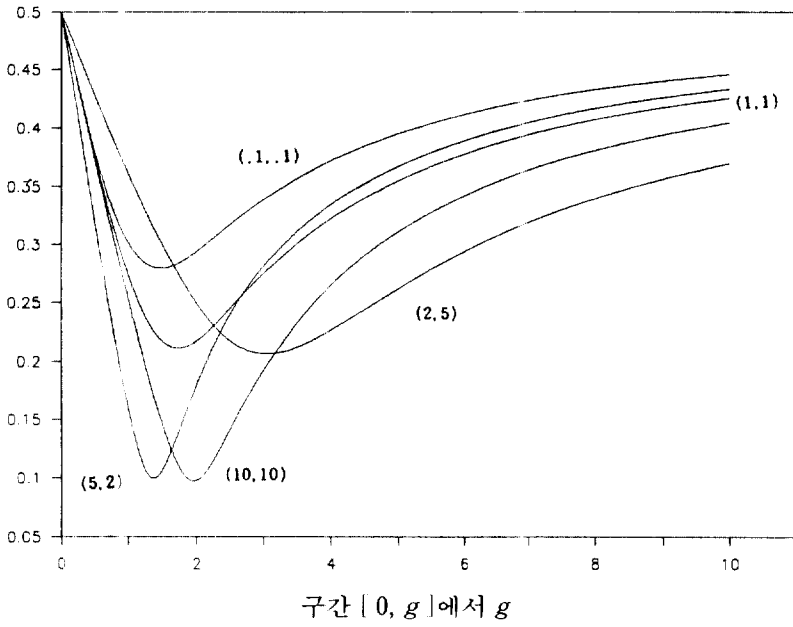
<그림 2.2>

[5, 2], [2, 5], [10, 10]으로 나누어 생각한다.

예제 2 : $\Omega = [0, g]$ 구간 : g 가 0에서 멀어
 심과 동시에 발생하는 $\zeta(-1)$ 의 값을 각각의 p, q 에 대해 그림 2.3에 제시하였다. p, q 에 따른
 약간의 가중치의 차이는 있지만 모든 p, q 에 상
 관 없이 내삽법과 외삽법이 혼선이 있는 영역
 (실험계획구간 상한인 값1을 우측으로 약 1정
 도)을 벗어 나면 다시 $\zeta(-1)$ 의 값이 회복되고
 0.5에 접근됨을 알 수 있다.

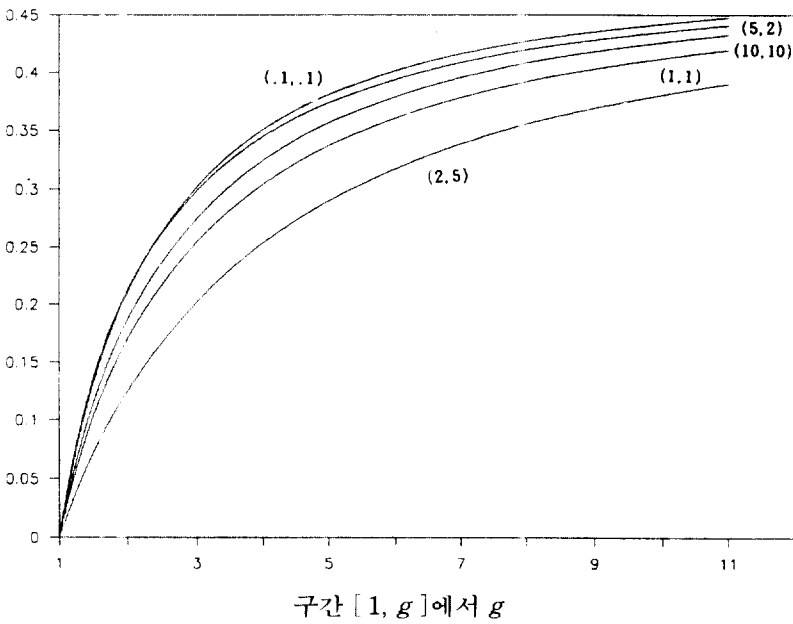
예제 3 : $\Omega = [1, g]$ 구간 : g 가 1에서 멀어

짐과 동시에 발생하는 $\zeta(-1)$ 의 값을 각각의 p, q 에 대해 그림 2.4에 제시하였다. p, q 에 따른
 약간의 가중치의 차이는 있지만 모든 p, q 에 상
 관 없이 g 가 커짐과 동시에 $\zeta(-1)$ 의 값은 단순
 증가 하면서 0.5에 접근됨을 알 수 있다. 이는
 예제 2와 틀리게 내삽법과 외삽법의 혼선이 있
 는 영역이 없이 단순 외삽법의 문제이기 때문
 이다. 실험영역인 상한값 주위에 가중치를 많
 이 주면 줄수록 즉 g 값이 적으면 적을수록 ζ
 (-1) 의 값은 적어짐을 알 수 있다. 그러나 주
 어진 p, q 값의 크기에는 그렇게 영향을 받지 않
 음을 알 수 있다.



구간 $[0, g]$ 에서 g
단순선형회귀모형에서 $x = -1$ 에서의 질량

<그림 2.3>



구간 $[1, g]$ 에서 g
단순선형회귀모형에서 $x = -1$ 에서의 질량

<그림 2.4>

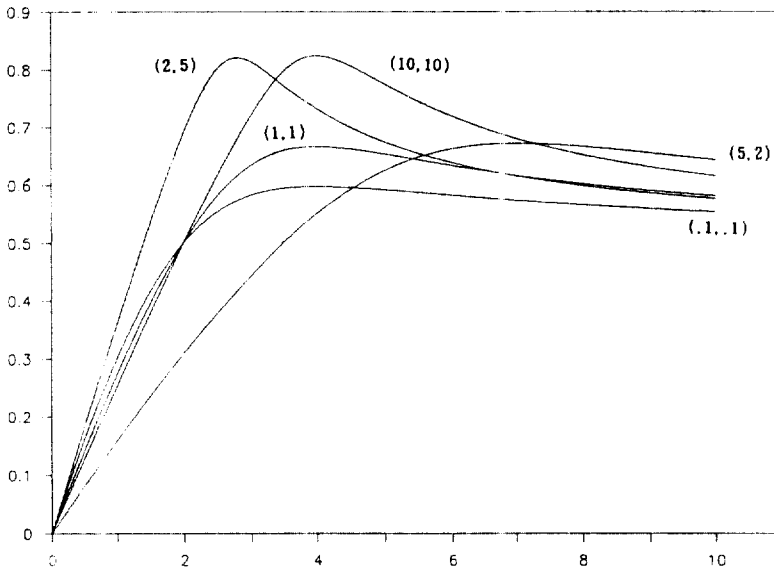
예제 4 : $\Omega = [1-g, 1]$ 구간 : g 가 0에서 멀어짐과 동시에 발생하는 $\xi(-1)$ 의 값을 각각의 p, q 에 대해 그림 2.5에 제시하였다. p, q 에 따른 약간의 가중치의 차이는 있지만 모든 p, q 에 상관 없이 내삽법과 외삽법이 혼선이 있는 영역(실험계획구간 하한인 값 -1 의 좌측의 영역, g 값으로 약 2와 4사이)을 벗어나면 다시 $\xi(-1)$ 의 값이 떨어지고 0.5로 접근함을 알 수 있는데 혼선이 있는 영역이 위의 예제 2의 경우에 비해 더 지속됨을 알 수 있다. 이는 출발점이 실험 영역 상한인 값 1에서 시작하였기 때문이다.

같은 경우에는 $\xi(\pm 1) = 0.5$ 이므로 p, q 가 각각 5, 2 그리고 2, 5인 경우만 고려한다. 지금까지는 p, q 로 주어지는 가중치의 차이가 그림의 전반적인 틀에는 큰 영향이 없다고 여겨졌지만 이 경우는 대칭인 $[-g, g]$ 인 구간이므로 당연히 p, q 의 값에 따른 대칭의 그림을 기대할 수 있다. 역시 내삽법과 외삽법의 혼선이 있는 영역을 벗어나면 다시 $(p, q) = (5, 2)$ 인 경우는 $\xi(-1)$ 의 값이 떨어졌다가 회복되고 0.5로 접근함을 알 수 있고 $(p, q) = (2, 5)$ 인 경우는 그 반대이다.

예제 5 : $\Omega = [-g, g]$ 구간 : g 가 커짐과 동시에 발생하는 $\xi(-1)$ 의 값을 각각의 p, q 에 대해 그림 2.6에 제시하였다. 여기서는 p 가 q 와

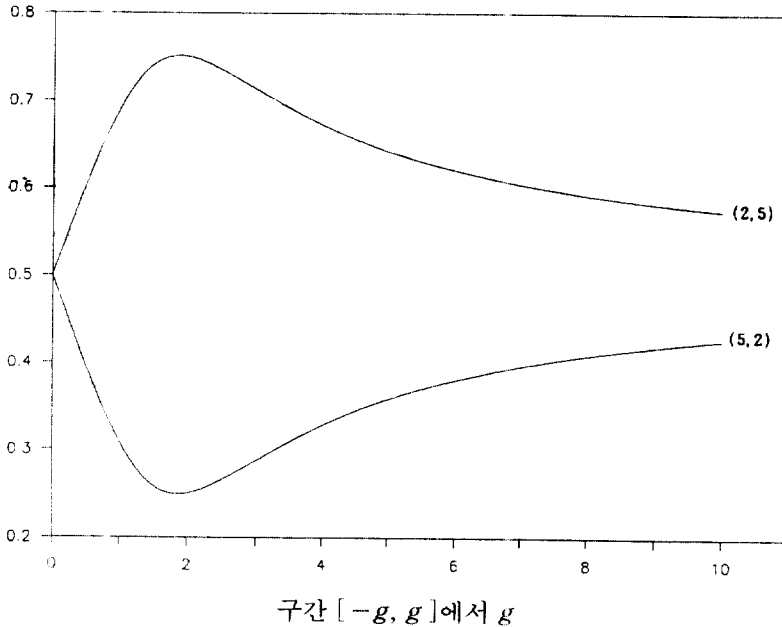
2.2 이차형식의 회귀모형

정리 2.2 위에서 정의한 Bata함수($p=q$ 인



구간 $[1-g, 1]$ 에서 g
단순선형회귀모형에서 $x = -1$ 에서의 질량

<그림 2.5>



단순선형회귀모형에서 $x = -1$ 에서의 질량

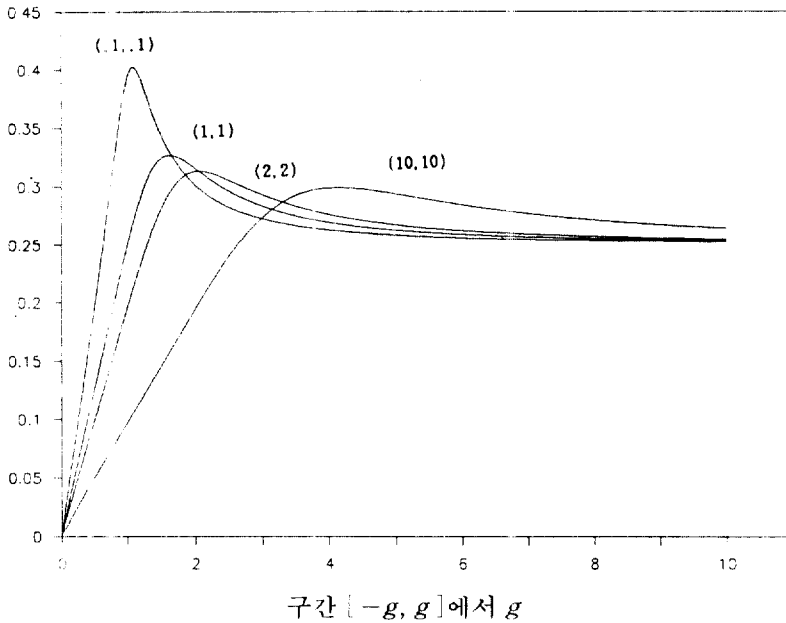
<그림 2.6>

경우에 국한), $X = [-1, 1]$, $\Omega = [-g, g]$ 및 $f^T = (1, x, x^2)$ 그리고 등분산의 오차함수인 경우는 $\xi(\pm 1) = d$ 는 이차방정식 $(3\mu_2 - 1)d^2 - (\mu_2 + \mu_4)d + (\mu_2 + \mu_4)$ 의 실근중 $0 < d < 1/2$ 에 있는 d 로 결정이 되고 $\xi(\pm 1) = d$, $\xi(0) = 1 - 2d$ 는 I_1 -최적이다.

이 과정에서 p, q 가 다른 경우의 이차 형식의 회귀분석모형에서는 중간 반힘점을 찾는 문제가 복잡하여지므로 동치정리중 2번째의 식이 $x = 0$ 을 중심으로 대칭이 되게하는 $p = q$ 인 경우만 고려하였다. 정리 1.1의 두번째 식인 $\phi(x, \xi)$ 이 $x = 0$ 을 중심으로 대칭인 4차의 다항식인 관계로 $x = -1, 0$ 및 1 에서 극대점이 발생한다. $\phi(\pm 1, \xi) = \phi(0, \xi)$ 인 $\xi(-1) = d$ 이다.

예제 6 : 위에서 정의한 Bata함수 ($p = q$ 인 경우만 포함)의 p 가 변하고, $\Omega = [-g, g]$ 에서의 g 가 변할때 $f^T = (1, x, x^2)$ 그리고 등분산의 오차함수인 경우 어떻게 $\xi(-1)$ 이 어떻게 변하는지 과정을 보도록 한다. 먼저 그림 2.7에서 $p(=q)$ 가 1로 정해져 있을때 즉 모든 구간, $\Omega = [-g, g]$ 에 일양분포 가중치함수를 적용시킬때 어떻게 $\xi(-1)$ 가 변하는지 보고자 한다.

약 $g = 1.6$ 일때 $\xi(-1) = 0.328$ 가 최고로 높이 솟아 있음을 알 수 있고 $g = 1$ 일때에는 $\xi(-1) = 0.25$ 임을 확인 할 수 있고 g 가 커짐에 따라 $\xi(-1)$ 는 0.25로 점근함을 알 수 있다. 그리고 p 가 1보다 커짐(적어짐)과 동시에 최고치의 $\xi(-1)$ 을 제공하는 g 는 우측(좌측)으로 쏠림을



이차형식회귀모형에서 $x = \pm 1$ 에서의 질량

<그림 2.7>

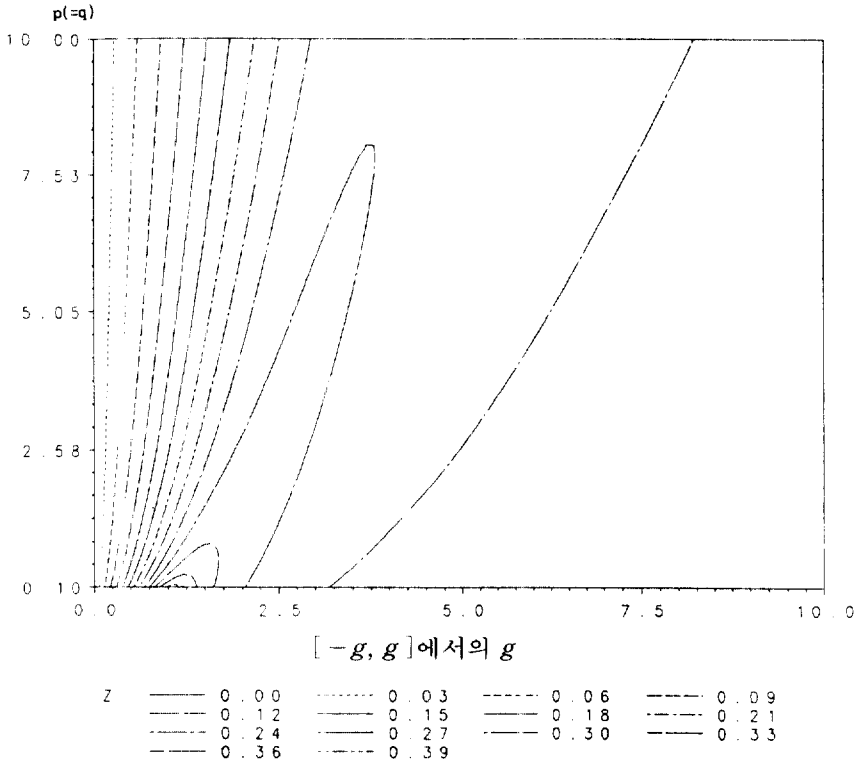
알 수 있다. 이러한 사실을 $p(=q)$ 까지 변화했을 때의 등고선 그림 <그림 2.8>에서 살펴보면 $x = -1$ 에서의 질량이 $1/3$ 보다 큰 영역은 하단 조그마한 지역에서만 발생됨을 알 수 있다. 즉 I_λ -최적은 보통의 경우에는 실험계획의 양극점보다는 중앙에 더 질량을 치우침을 알 수 있다.

표 1에서는 몇가지 p, q 의 값에 따른 $\Omega = [-1, 1]$ 에서의 $f^T = (1, x, x^2)$ 에 대한 I_λ -최적을 제시하였는데 $p=q$ 가 아닌 경우를 제외하면 중앙 받힘점이 더 이상 0이 아님을 볼 수 있다. 여기서 $p=q$ 가 1보다는 크면서 $\xi(-1)$ 이 $1/3$ 에 근접하는 $p=q$ 를 수치적으로 찾아보면 약 .33임(정확한 p 값은 이론적으로 불가능)을 알 수 있다. 여기서 지적이 되는 것은 D -최적실험은 양끝을 중요시하는 실험계획인 것을

알 수 있다. 다음절에서는 이분산오차함수인 경우에는 어떻게 I_λ -최적 기준이 변하는지 간단한 이분산오차함수의 형태하에서 알아보도록 하겠다.

3. 이분산오차함수하에서의 I_λ -최적 기준

어떠한 현상의 메카니즘을 설명하고자 할때 실험자들은 실험조건외 동질성을 보통 의심받기 쉬운 비교적 넓은 실험영역에 관심을 갖는 경우가 있는데 이러한 경우는 등분산을 가정한 실험계획의 적용이 타당한 경우가 아니라 할 수 있다.



$p(=q)$ 와 g 의 변화에 따른 $x = \pm 1$ 에서의 질량등고선(등분산 오차함수인 이차형식회귀모형)

<그림 2.8>

오차항에 대한 분산의 지식은 다음과 같다고 가정을 하자. 제일 작은 분산의 값과 제일 큰 값의 비는 고정된 값 $\gamma > 1$ 이면 실험영역 X 에서는 분산은 선형으로 증가하고 $0 < \gamma < 1$ 이면 실험영역 $X = [-1, 1]$ 에서는 분산은 감소한다. 이를 수학적으로 규정하면 $w^{-1}(x) \propto (\gamma - 1)x + (\gamma + 1)$ 이다. 등분산($\gamma = 1$)일 경우 문제의 단순성을 위하여 $w^{-1}(x) = [(\gamma - 1)x + (\gamma + 1)] / 2$ 로 한다.

즉각적인 I_1 -최적의 실험계획은 분산이 큰 영역에 더 많은 질량을 배치케 하고자 할 것이다. 모든 오차함수의 형태와 모든 모형에 대

해 I_1 -최적의 해를 계산법에 의해 구한다는 것은 불가능한 관계로 다음의 단순선형회귀모형만 고려코자 한다.

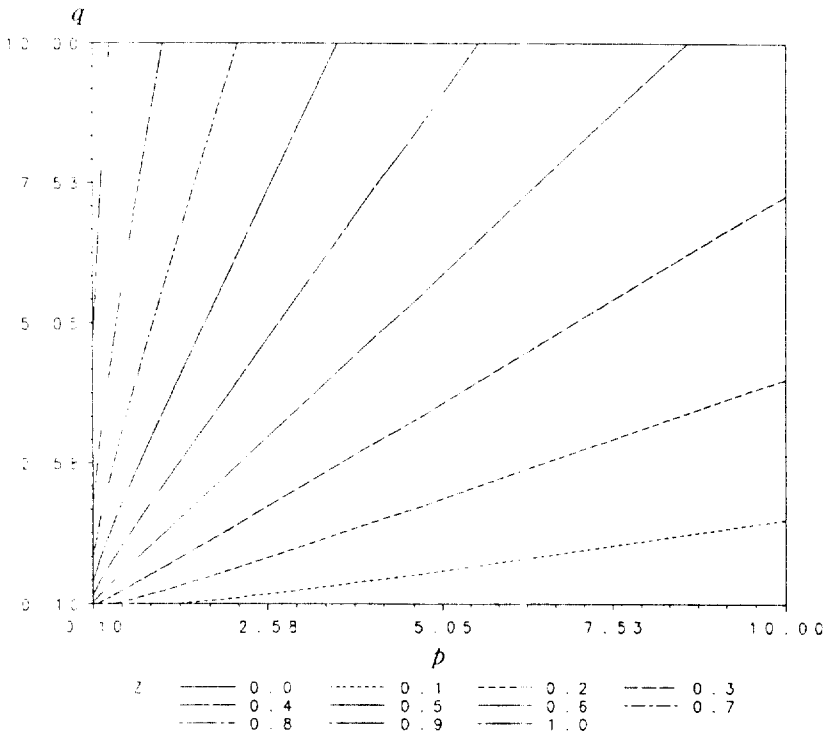
정리 3.1 위에서 정의한 Bata함수($p \neq q$ 인 경우 포함) $X = [-1, 1]$, $\Omega = [a, b]$ 및 $f^{-1} = (1, x)$ 그리고 이분산의 오차함수, $w^{-1}(x) = [(\gamma - 1)x + (\gamma + 1)] / 2$ 인 경우는 $\xi(-1) = d$ 는 이차방정식 $(\gamma - 1 + 2(\gamma + 1)\mu_1 + (\gamma - 1)\mu_2)d + (2 - 4\mu_1 + 2\mu_2)d + (2\mu_1 - \mu_2 - 1) = 0$ 의 실근중 $0 < d < 1$ 에 있는 d 로 결정이 되고 $\xi(-1) = d, \xi(1) = 1 - d$ 는 이 경우 I_1 -최적이다.

예제 3.1 : Bata함수($p \neq q$ 인 경우 포함)의 p 및 q 가 변할때 $X=[-1, 1]$, $\Omega=[-1, 1]$ 및 $f^T=(1, x)$ 그리고 이분산오차함수인 경우 (여기서는 $\gamma=3$ 을 기준으로 삼는다.) 어떻게 $\xi(-1)$ 이 변하는지 p, q 를 0.1에서 10까지 변화하는 과정을 등고선으로 그림 3.1에 제시하였다.

그림 2.2와는 달리 $p=q$ 를 중심으로 대칭이 아님을 쉽게 알 수 있다. 그러나 p, q 가 같으면 $\xi(-1)=.366025$ 은 일정하다. 분산이 $\gamma=3$ 인 경우처럼 증가하면 $x=1$ 에 질량이 더 치우침을 알 수 있다. G -최적인 경우 $x=1$ 에서의 질량은 $\gamma/(\gamma+1)$ 임을 참고바란다. G -최적실

험계획은 가중치 없이 예측치의 분산의 최대치를 최소화하는데 있고 이 경우에는 가중치를 실험계획좌우로 대칭을 두었기 때문이다. 참고로 그림 3.1에서는 $\gamma=3$ 인 경우 G -최적실험계획 ($\xi(-1)=1/4, \xi(1)=3/4$)이 어떠한 p, q 의 조합하에서 I_1 -최적실험계획($\gamma=3$)하에서의 질량과 같은지를 쉽게 파악이 될 것이다. 이와 유사하게 그림 2.2에서는 어떠한 p, q 의 가중치를 두어야 I_1 -최적실험계획하($\gamma=1$)에서 $\gamma=3$ 인 경우의 G -최적의 실험계획과 같아지는지 볼 수 있다.

예제 3.2 : Bata함수(다양한 p, q 인 경우,



p, q 변화에 따른 $x=-1$ 에서의 질량등고선(이분산오차함수, $\gamma=3$ 인 단순선형모형)

<그림 3.1>

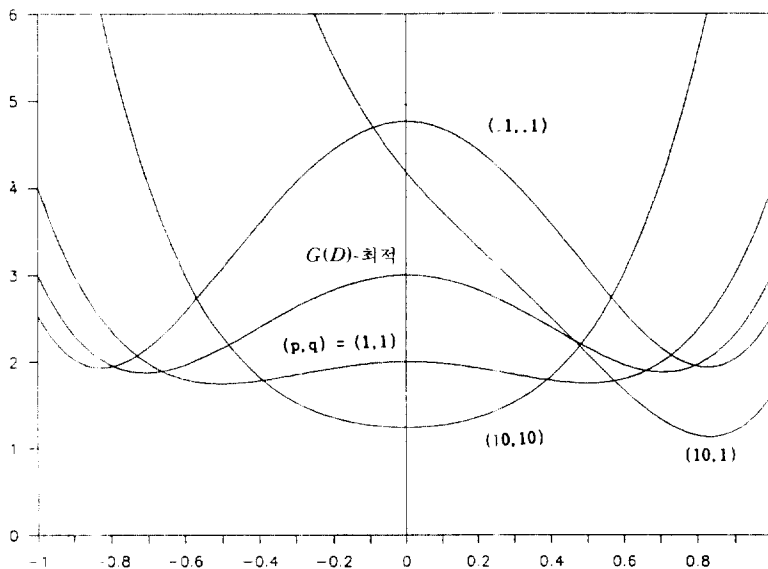
$\Omega = [-1, 1]$, $f^T = (1, x, x^2)$ 그리고 이분산오차함수인 경우 ($\gamma = 3, 5, 0.5$)에서 어떻게 $\xi(\pm 1)$ 및 ξ (중앙반합점)이 어떻게 변하는지 표 1에서 제시하였다. 이 경우는 중앙점의 p, q 에 따른 변화를 수학적으로 표현키 어렵기 때문에 수치적 알고리즘에 의해 작성된 표만 제시하였다. 이 표 1과 G -최적실험기준하에서의 실험계획, $\xi(-1) = (1/(1+\gamma))2/3$, $\xi(0) = 1/3$, $\xi(1) = (\gamma/(1+\gamma))2/3$ 을 비교 바란다. G -최적실험기준은 분산이 큰 지역에 질량을 옮김으로서 최대치의 예측치의 분산을 내리는 경향이 있는데 반해 I_λ -최적은 가중치 함수와 합성적으로 실험기준이 설정되기 때문에 γ 의 변화에 따른 질량의 변화가 상대적으로 심하지 않은 것을 알 수 있다. 즉 I_λ -최적을 이용하여 실험을 하면 실사 가정된 γ 가 잘못된 경우라 할지라도 효율성면에서는 G -최적보다 높다 할 수

있다.

4. 예측치의 분산의 함수

I_λ -최적실험계획법은 주어진 영역하에서의 예측치의 분산의 값을 최소화하는 G -최적과는 달리 실험계획자의 가중치함수에 따른 일종의 평균된 예측치의 분산의 값에 중점을 두었기 때문에 과연 가중치로 중점을 둔 영역에서 예측치의 분산의 값이 어떻게 변하는지 보자.

예제 4. 1 : Beta함수(p, q 가 각각 (1, 1), (10, 10), (.1, .1), (10, 1)인 경우), $\Omega = [-1, 1]$, $f^T = (1, x, x^2)$ 그리고 등분산오차함수인 경우 어떻게 예측치의 분산함수의 값이 변하는지 그림 4. 1에 도시하였다. 이를 G -최적기준에 비



이차형식의 회귀모형에서의 등분산 오차함수일때 I_λ -최적의 예측치의 분산함수

<그림 4. 1>

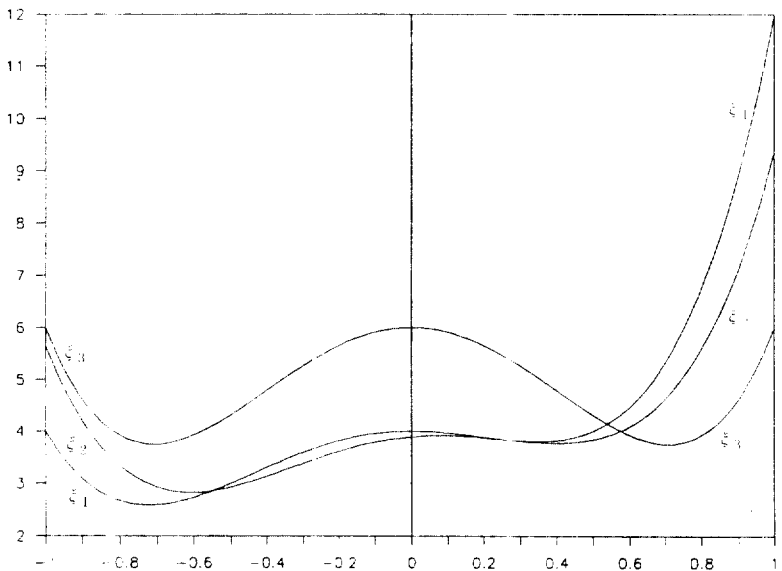
추어 비교하였다. 가중치가 부여된 영역에서는 예측치의 분산함수가 G -최적의 경우보다 낮음을 볼 수 있다.

예제 4.2 : Beta함수(p, q 가 (1, 1)경우), $\Omega = [-1, 1]$, $f^T = (1, x, x^2)$ 그리고 이분산오차함수인 경우 ($\gamma=3$) 어떻게 예측치의 분산함수의 값이 변하는지 $\gamma=1$ 인 I_λ -최적기준 실험법 ($\xi(\pm 1)=0.25, \xi(1)=0.5$), $\gamma=3$ 인 I_λ -최적기준 실험계획법 ($\xi(-1)=0.177, \xi(-0.0803)=0.503, \xi(1)=0.320$) 및 이분산 G -최적실험계획법 ($\xi(-1)=1/6, \xi(0)=1/3, \xi(1)=1/2$)을 그림 4.2에 비교 도시하였다. G -최적기준 실험법에 대해서 두 실험법은 실험계획영역의 우측부분을 제외한 모든 영역에서 우위를 점하고 있다. 이는 G -최적실험의 특성상 이 우측영역

에서는 G -최적보다 우위에 설 수 없다. 그리고 G -최적실험을 제외한 두 실험중에서는 약 80% 가량의 영역에서 $\gamma=3$ 인 경우의 올라탄 실험법이 우위에 있음을 알 수 있으나 그 차이는 크지 않다. 이는 표 1에서 본 바와 같이 I_λ -최적은 가중치함수와 γ 에 의해 합성으로 실험계획이 설정되기 때문이다.

5. 결 어

이상 단순선형 및 이차형식의 선형회귀모형을 갖고 지금까지 그 실용성이 잘 알려져 있지 않은 I_λ -최적실험기준을 알아보았다. 그리고 D -나 G -최적성과의 비교를 통하여 그 특성을 알아 보았는데 I_λ -기준은 이분산 G -최적성에



ξ_1 : $\gamma=1$ 을 가정한 I_λ -최적, ξ_2 : $\gamma=3$ 을 가정한 I_λ -최적, ξ_3 : $\gamma=3$ 을 가정한 G -최적

이차형식의 회귀모형에서의 이분산 오차함수($\gamma=3$)일때 예측치의 분산함수

<그림 4.2>

비해서 최적성의 확인절차가 간편할 뿐 아니라
가중치함수의 성질을 감안하기 때문에 효율적
이다 할 수 있다. 비록 간단한 예제를 통한 제
한적인 결과이지만 등분산 G-최적과 동치인

D-최적이거나 이분산 G-최적보다는 이분산 오차
함수인 경우에는 I_{λ} -최적실험기준이 경우에 따
라서는 더 유용하다고 본다.

<표 1> 이차형식에 있어서의 I_{λ} -최적실험계획

가중치함수		$\gamma=1$ (등분산 오차함수)			$\gamma=3$		
λ_1	반함점	-1	0	1	-1	-.0803	1
	질량	.25	.5	.25	.177	.503	.320
λ_2	반함점	-1	0	1	-1	-.0667	1
	질량	.196	.608	.196	.138	.610	.252
λ_3	반함점	-1	0	1	-1	-.0390	1
	질량	.098	.804	.071	.069	.805	.126
λ_4	반함점	-1	0	1	-1	-.1121	1
	질량	.395	.210	.395	.287	.214	.5
λ_5	반함점	-1	.1169	1	-1	.1002	1
	질량	.071	.607	.322	.049	.570	.381
λ_6	반함점	-1	.3275	1	-1	.2240	1
	질량	.027	.356	.617	.020	.310	.670

〈표 1 계속〉 이차형식에 있어서의 I_k -최적실험계획

가중치함수		$\gamma=5$			$\gamma=0.5$		
λ_1	받힘점	-1	-.1105	1	-1	.0508	1
	질량	.147	.507	.346	.295	.501	.204
λ_2	받힘점	-1	-.0911	1	-1	.0424	1
	질량	.114	.613	.273	.233	.609	.159
λ_3	받힘점	-1	.0541	1	-1	.0248	1
	질량	.057	.806	.137	.117	.804	.079
λ_4	받힘점	-1	-.1607	1	-1	.0724	1
	질량	.241	.217	.542	.463	.212	.326
λ_5	받힘점	-1	.0730	1	-1	.2015	1
	질량	.040	.558	.401	.086	.637	.277
λ_6	받힘점	-1	.1802	1	-1	.3729	1
	질량	.018	.294	.688	.028	.403	.568

$$\lambda_1 \quad \lambda_2 \quad \lambda_3 \quad \lambda_4 \quad \lambda_5 \quad \lambda_6$$

$$(p, q) = (1, 1) (2, 2) (10, 10) (.1, .1) (5, 2) (10, 1)$$

$$\chi = \Omega = [-1, 1], f^T = (1, x, x^2) \text{에 국한}$$

參 考 文 獻

1. Cook, R. D. and Nachtsheim, C. J. (1982), Model Robust, Linear-Optimal Designs, *Technometrics*, Vol. 24, No. 1., 49-54.
2. Fedorov, V. V. (1972), *Theory of Optimal Experiments*, New York, Academic Press, Inc.
3. Studden, W. J. (1977), Optimal Designs for Integrated Variance for Polynomial Regression, In *Statistical Decision Theory and Related Topics II.*, New York, Academic Press.
4. Wong, W. K., and Cook, R. D. (1992), Heteroscedastic *G*-optimal Designs, Technical Manuscript, Dept. of Bio-Statistics, University of California Los Angeles, submitted to the *Annals of Statistics* for publication.
5. Wong, W. K. (1992), Unified Approach to Mini-Max Designs Constructions, Technical Manuscript, Dept. of Bio-Statistics, University of California Los Angeles.