

## ☒ 응용논문

## 혼합물 실험계획에서 실험점의 확장, 결측치, 이상치의 영향을 평가할 수 있는 그래픽 방법\*

장대홍 · 박상현

부경대학교 응용수학과

### A graphical method for evaluating the effect of design augmentation, missing observation, and outlier in mixture experiments

Dae-Heung Jang · Sang-Hyun Park

Dept. of Applied Mathematics, Pukyong National University

#### Abstracts

D-optimality is used often in design augmentation of mixture experiments. Although such alphabetic criteria provide a valuable foundation for generating designs, they often fail to convey the true nature of the design's support of the fitted model in terms of prediction variance over a region of interest. Thus, a graphical method is proposed to evaluate augmented designs in mixture experiments. This method can be used to evaluate the effect of missing observation and outlier in mixture experiments.

#### 1. 서론

실험계획의 확장은 반응표면분석에서 뿐만이 아니라 혼합물 실험에서도 매우 중요한 연구주제가 되고 있다. 실험계획의 확장시 기존의 연구들은 주로 D-최적화 계획을 염두에 두었다. 그러나, 이러한 알파벳 최적화 기법들(예로, A-, D-, E-, G-와 V-최

\* 이 논문은 1995년도 한국학술진흥재단의 지방대 육성공모과제 학술연구 조성비에 의하여 연구되었음.

적화)은 하나의 수치를 이용하는 분산-최소화 기준들로서, 실험계획의 전 영역에 걸친 성능의 정도를 파악하기에는 부족한 점이 많다. Vining, Cornell과 Myers(1993)는 혼합물실험에서 이러한 문제를 지적하고, 제한된 영역에서의 실험계획에 대하여 전 영역에 걸친 성능을 파악할 수 있는 그래픽 방법을 제안하였다. 장대홍과 박상현(1995)은 이 그래픽 방법을 이용하여 혼합물 실험계획의 확장이나 실험계획점들 중 결측값이 발생하였을 때 실험계획의 전 영역에 걸친 변화를 알아 보고, 실험계획들을 상호 비교, 평가할 수 있는 방법을 제시하였다. 본 논문에서는 장대홍과 박상현(1995) 논문의 확장으로서 이 그래픽 방법이 실험계획의 확장이나 실험계획점들 중 결측값이 발생하는 경우와 아울러 실험계획점들 중 이상값이 발생하였을 때 실험계획의 전영역에 걸친 변화를 알아보고, 실험계획점들을 상호 비교, 평가할 수 있는 방법임을 보였다.

## 2. 혼합물 실험계획을 평가하기 위한 그래픽 방법

혼합물실험에서는 혼합물을 구성하는 성분의 상대적 비율에 의해 반응값이 결정된다. 혼합물의 구성성분 갯수를  $q$  개라 하고,  $x_i$ 를  $i$ 번째 구성성분의 비율이라 하면

$$0 \leq x_i \leq 1, \quad i = 1, 2, \dots, q$$

이고,

$$\sum_{i=1}^q x_i = 1$$

이 된다. 그리하여, 실험영역이  $(q-1)$ 차 단체(simplex)가 된다.  $i$ 번째 구성성분의 비율에 제한이 있는

$$0 < L_i \leq x_i \leq U_i < 1, \quad i = 1, 2, \dots, q$$

일 때, 실험영역은 원 단체상의 제한된 영역이 되고, 일반적으로 불규칙한 볼록다면체가 된다. 이런 경우 볼록다면체의 꼭지점, 각 면의 중심점, 각 변의 중심점, 전체 중심점들이 실험계획을 구성할 수 있는 후보점들(candidate points)이 된다.

제한된 영역에서 각 구성성분의 효과를 알아보기 위하여는 Cox 방향이 이용되는데,  $i$ 번째 구성성분에 대한 Cox 방향이란, 전체 중심점에서  $i$ 번째 구성성분에 해당하는 꼭지점에 그은 가상선을 말한다. 전체중심점에서 각 구성성분의 비율을  $\underline{c} = (c_1, c_2, \dots, c_q)$ 라 하면  $\sum_{i=1}^q c_i = 1$ 이 된다.  $i$ 번째 구성성분의 비율을 Cox 방향을 따라  $c_i$ 에서  $\Delta_i$ 만큼 변화시키면,  $i$ 번째 구성성분의 새로운 비율은

$$x_i = c_i + \Delta_i \quad (1)$$

가 되고, 나머지  $(q-1)$ 개 구성성분의 비율들은

$$x_j = c_j \frac{1 - c_i - \Delta_i}{1 - c_i}, \quad j = 1, 2, \dots, q, \quad j \neq i \quad (2)$$

이 된다. 그런데, (2)식처럼 정의되는  $x_j$ 와  $x_k$ 의 비를 보면

$$\frac{x_j}{x_k} = \frac{c_j}{c_k}$$

가 되어 전체 중심점에서의 비와 같아짐을 알 수 있다. 그리하여, 제한된 영역에서 각 구성성분의 효과를 알아보기 위하여는 Cox 방향을 이용하게 되는 것이다.

혼합물 실험모형을 행렬을 이용하여 나타내면,

$$y = X\beta + \varepsilon,$$

이 된다. 여기서,  $y$ 는  $n \times 1$  반응값 벡터이고,  $X$ 는  $n \times p$  행렬로서 선택하는 회귀 모형의 차수에 따라 달라지는 모형행렬이고,  $\beta$ 는  $p \times 1$  모수벡터이고,  $\varepsilon$ 는 오차항을 나타내는  $n \times 1$  확률오차벡터로서, 일반적으로 확률분포를  $N(0, \sigma^2)$ 으로 가정한다. 이 경우에  $\beta$ 의 최소제곱추정량  $b$ 의 분산-공분산 행렬은

$$\text{Var}(b) = (X'X)^{-1}\sigma^2$$

이 되고,  $\sigma^2$ 으로 나눈 추정반응값  $\hat{y}(x) = x_0' b$ 의 분산은

$$V(x) = \frac{\text{Var}(\hat{y}(x))}{\sigma^2} = x_0'(X'X)^{-1}x_0 \quad (3)$$

이 된다. 여기서,  $x_0$ 는 1차모형일 때는  $x_0' = (x_1, x_2, \dots, x_q)$ 이고, 2차모형에서는  $x_0' = (x_1, x_2, \dots, x_q, x_1x_2, x_1x_3, \dots, x_1x_{q-1}, \dots, x_{q-1}x_q)$ 이다.  $i$ 번째 구성성분의 비율을 Cox 방향을 따라  $c_i$ 에서  $\Delta_i$ 만큼 변화시키면, (1)식으로부터

$$\Delta_i = x_i - c_i$$

이므로

$$x_i = c_j \frac{1 - x_i}{1 - c_i} \quad (4)$$

가 된다. (4)식을 (3)식에 대입하여 행렬계산을 행하면, 혼합물 실험모형이 1차모형일 때는  $V(\mathbf{x})$ 가  $x_i$ 에 대하여 2차식이 되고, 2차모형일 때는  $V(\mathbf{x})$ 가  $x_i$ 에 대하여 4차식이 된다.

실험계획의 확장이나 실험계획점들 중 결측값이 발생하는 각각의 경우, 각 구성성분에 대하여 Cox 방향을 따라 양 경계영역 사이를 이동하면서  $V(\mathbf{x})$ 의 값을 구하여 그림으로 그린다. 이 그림을 추정반응값 분산그림(prediction variance trace : PVT)이라 부른다. 이 그림을 이용하면 실험계획의 확장이나 실험계획점들 중 결측값이 발생하는 경우에 제약이 있는 전 실험영역에 걸쳐 추정반응값 분산의 변화를 각 구성성분에 대하여 알아 볼 수 있어 확장 또는 결측의 효과를 알아 볼 수 있고, 추정반응값 분산들의 크기를 비교하므로써 실험계획들을 비교, 평가할 수 있다. 또한, 실험계획점들 중 이상값이 발생하는 경우에는 추정반응값 편의그림(prediction bias trace)을 이용하여 제약이 있는 전 실험영역에 걸쳐 추정반응값 편의의 변화를 각 구성성분에 대하여 알아 볼 수 있어 이상값의 영향을 알아 볼 수 있다.

## 2.1 실험계획의 확장이나 결측이 있는 경우

$\mathbf{x}_a$ 를 후보점들 중 원래 실험계획에 새로이 첨가되는 벡터라 하고,  $\hat{y}_a(\mathbf{x})$ 를 확장된 모형행렬을 이용하여 얻어지는 추정반응값이라고 하면,  $\sigma^2$ 으로 나눈  $\hat{y}_a(\mathbf{x})$ 의 분산은

$$V_a(\mathbf{x}) = \mathbf{x}_0'(X'X)^{-1}\mathbf{x}_0 - \frac{\mathbf{x}_0'(X'X)^{-1}\mathbf{x}_a \mathbf{x}_a'(X'X)^{-1}\mathbf{x}_0}{1 + \mathbf{x}_a'(X'X)^{-1}\mathbf{x}_a} \quad (5)$$

이 된다. 실험계획의 확장에 의한 결과는 (5)식의 오른쪽 두번째 항으로 나타나, 분산이 감소하게 된다. 각 구성성분에 대하여 Cox 방향을 따라 양 경계영역 사이를 이동하면서  $V_a(\mathbf{x})$ 의 값을 구하여 PVT를 그리면, 제한이 있는 전 실험영역에 걸쳐 실험계획의 확장으로 인한 추정반응값 분산의 변화를 쉽게 알 수 있고, 추정반응값 분산들의 크기를 비교하므로써 실험계획들을 비교, 평가할 수 있게 된다.

한편,  $k$ 번째 실험계획점이 결측값이 되는 경우에는,  $\sigma^2$ 으로 나눈 추정반응값의 분산은

$$V_{-k}(x) = x_0'(X'X)^{-1}x_0 + \frac{x_0'(X'X)^{-1}x_k x_k'(X'X)^{-1}x_0}{1-h_{kk}} \quad (6)$$

가 된다. 여기서,  $x_k$ 는 결측값 벡터이고,  $h_{kk}$ 는 hat matrix  $H = X(X'X)^{-1}X'$ 의  $k$ 번째 대각선 원소이다.  $k$ 번째 실험계획점이 결측되는 경우에 대한 결과는 (6)식의 오른쪽 두번째 항으로 나타나, 분산이 증가하게 된다. 각 구성성분에 대하여 Cox 방향을 따라 양 경계영역 사이를 이동하면서  $V_{-k}(x)$ 의 값을 구하여 PVT를 그리면, 제한이 있는 전 실험영역에 걸쳐 실험계획의 결측으로 인한 추정반응값 분산의 변화를 쉽게 알 수 있고, 추정반응값 분산들의 크기를 비교하므로써 실험계획들을 비교, 평가할 수 있게 된다. 여러 실험점들이 확장되거나 결측되는 경우는 첨가 벡터  $x_a$ 나 결측값 벡터  $x_k$ 를 각각 첨가행렬이나 결측값 행렬로 바꾼 후 추정반응값 분산을 구하면 된다. 여기서, 주의할 사항은 실험계획점들의 확장은 통제가 가능한 반면 결측은 통제가 불가능하므로 실험을 행한 후 결측값이 발생하는 경우, 바로 그 점에서 실험점을 확장하는 것이 순리일 것이다.

## 2.2 실험계획점 중 이상값이 발생하는 경우

우리는 회귀진단을 통하여, 즉 표준화 잔차나 스튜던트화 잔차를 통하여 특정 실험계획점이 이상값인지의 여부를 진단할 수 있다. 이상값을 나타내는 모형으로서는 Cook과 Weisberg(1982)의 mean shift outlier model과 Cook, Holschuh, 그리고 Weisberg(1982)의 variance inflation model이 있는데, 이 중 mean shift outlier model을 사용하면

$$y = X\beta + d_k\phi + \varepsilon$$

이 된다. 여기서,  $d_k$ 는  $k$ 번째 위치의 값이 1이고, 나머지는 0인 index vector이고  $\phi$ 는 mean shift의 양이다. 그러므로,  $\phi$ 가 0이 아니면  $k$ 번째 실험계획이 이상값이 된다. 이 모형을 이용하면,  $\hat{y}(x)$ 의 평균제곱오차는

$$\frac{MSE(\hat{y}(x))}{\sigma^2} = x_0'(X'X)^{-1}x_0 + \frac{\phi^2}{\sigma^2} x_k'(X'X)^{-1}x_0 x_0'(X'X)^{-1}x_k \quad (7)$$

이 된다. O'Gorman(1984)은  $\frac{\phi^2}{\sigma^2} > \frac{1}{1-h_{kk}}$ 이면 이상값인  $k$ 번째 실험계획을 포함하여 행하는 추정이  $k$ 번째 실험계획을 제거하고 행하는 추정보다 좋지 않음을 보이고, 다음과 같은 검정을 실시할 수 있는 검정통계량을 제시하였다.

$$H_0 : \frac{\phi^2}{\sigma^2} \leq \frac{1}{1-h_{kk}} \quad H_1 : \frac{\phi^2}{\sigma^2} > \frac{1}{1-h_{kk}}$$

이상값의 영향은 (7)식의 오른쪽 두번째 항으로 나타나므로, 이상점의 크기가 일정하다고 가정한다면,  $\frac{\phi^2}{\sigma^2}$  을 제거시킨 이상값의 영향은

$$B_k(x) = x_k'(X'X)^{-1}x_0 x_0'(X'X)^{-1}x_k \quad (8)$$

으로 평가할 수 있다.  $i$ 번째 구성성분의 비율을 Cox 방향을 따라  $c_i$ 에서  $\Delta_i$ 만큼 변화시키면,  $x_j$ 는 (4)식 처럼 나타낼 수 있으므로 (4)식을 (8)식에 대입하여 행렬계산을 행하면, 혼합물 실험모형이 1차모형일 때는  $B_k(x)$ 가  $x_i$ 에 대한 2차식이 되고, 혼합물 실험모형이 2차모형일 때는  $B_k(x)$ 가  $x_i$ 에 대한 4차식이 된다. 각 구성성분에 대하여 Cox 방향을 따라 양 경계영역 사이를 이동하면서  $B_k(x)$ 의 값을 구한 그림을 추정반응값 편 의그림(prediction bias trace: PBT)라 부르며, 이 그림을 이용하면 실험계획점들 중 이상값이 있는 경우에 제약이 있는 전 실험영역에 걸쳐 추정반응값 편 의의 변화를 각 구성성분에 대하여 알아 볼 수 있어 이상값의 영향을 알 수 있게 된다.

### 3. 수치 예

Snee(1975)는 윤활제 혼합실험을 통하여 4개의 구성성분들에 아래와 같은 제약조건을 주어 다음 <표 1>과 같은 33개의 후보점을 제시하고, 이 중 다음 <표 2>와 같은 18개의 실험점들로 구성된 Snee 실험계획을 제시하고, 이 실험계획을 이용하여 2차모형을 추정하였다. 이 Snee 실험계획은 McLean과 Anderson(1966)이 제안한 꼭지점 실험계획과는 실험계획점이 하나만(변 중심점인 19번 실험점) 다른 실험계획이다.

$$\begin{aligned} 0.07 &\leq x_1 \leq 0.18 \\ 0.00 &\leq x_2 \leq 0.30 \\ 0.37 &\leq x_3 \leq 0.70 \\ 0.00 &\leq x_4 \leq 0.15 \end{aligned}$$

여기서,  $x_1$ 은 첨가제,  $x_2, x_3, x_4$ 는 성분 A, B, C이다.

< 표 1 > Snee의 윤활제 혼합실험에서의 후보점들

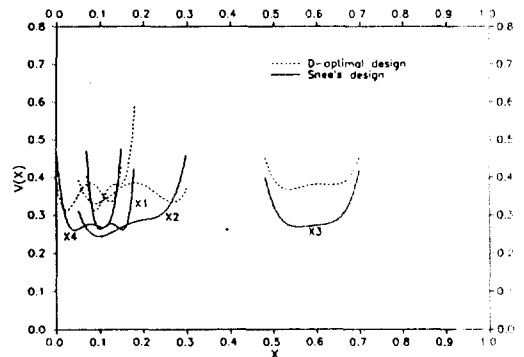
	꼭지점				면 중심점				변 중심점					
	$x_1$	$x_2$	$x_3$	$x_4$	$x_1$	$x_2$	$x_3$	$x_4$	$x_1$	$x_2$	$x_3$	$x_4$		
1	0.150	0.0000	0.7000	0.150	11	0.070	0.2275	0.6275	0.075	19	0.180	0.1500	0.5200	0.150
2	0.180	0.3000	0.3700	0.150	12	0.180	0.1440	0.5920	0.084	20	0.070	0.1550	0.7000	0.075
3	0.070	0.2300	0.7000	0.000	13	0.170	0.0000	0.6900	0.140	21	0.180	0.3000	0.4450	0.075
4	0.070	0.0800	0.7000	0.150	14	0.125	0.3000	0.5000	0.075	22	0.070	0.3000	0.5550	0.075
5	0.180	0.1200	0.7000	0.000	15	0.130	0.0860	0.7000	0.084	23	0.125	0.1750	0.7000	0.000
6	0.070	0.3000	0.6300	0.000	16	0.125	0.2375	0.6375	0.000	24	0.125	0.3000	0.4250	0.150
7	0.070	0.3000	0.4800	0.150	17	0.130	0.1360	0.5840	0.150	25	0.125	0.3000	0.5750	0.000
8	0.180	0.0000	0.6700	0.150	18*	0.133	0.1630	0.6170	0.087	26	0.070	0.1900	0.5900	0.150
9	0.180	0.3000	0.5200	0.000						27	0.180	0.0600	0.7000	0.060
10	0.180	0.0000	0.7000	0.120						28	0.180	0.2100	0.6100	0.000
										29	0.110	0.0400	0.7000	0.150
										30	0.165	0.0000	0.7000	0.135
										31	0.070	0.2650	0.6650	0.000
										32	0.165	0.0000	0.6850	0.150
										33	0.180	0.0000	0.6850	0.135

\* 전체 중심점

Cornell(1990)은 Snee의 윤활제 혼합실험을 이용하여 다음 <표 2>와 같은 D-최적화 계획을 제안하였다. 다음 <표 2>에 Snee계획과 D-최적화 계획을 제시하였고, 각 계획의  $\{\det(X'X)^{-1}\}^{1/10}$ 의 값을 나타내었다. 다음 <그림 1>은 <표 2>에 제시한 2가지 계획에 대한 PVT이다.

< 표 2 > 표 1을 이용한 혼합물 실험계획

	Snee 계획	D-최적화 계획
꼭지점 1-10	전부	1-7, 9, 10
면 중심점 11-17	11,12,14-17	14 17
변 중심점 19-33	19	19-24
전체 중심점 18	18	18
$\{\det(X'X)^{-1}\}^{1/10}$	286.5263	273.5506



< 그림 1 > Snee 계획과 D-최적화 계획에 대한 추정반응값 분산그림

### 3.1 실험계획의 확장이 있는 경우

18개로 구성된 Snee 계획에다 나머지 15개의 후보점들 중 하나를 첨가하여 19개로 구성된 확장 Snee 계획을 만들었을 때  $\{\det(X'X)^{-1}\}^{\frac{1}{10}}$ 의 값은 다음 <표 3>과 같다. 그러므로, D-최적화 입장에서는 21번 후보점을 첨가한 확장 Snee 계획이 선택되나, PVT를 이용하면 추정반응값 분산의 입장에서는 20번 또는 23번 후보점을 첨가한 확장 Snee 계획이 행렬식의 값이 더 큼에도 불구하고 21번 후보점을 첨가한 확장 Snee 계획과 대등한 성능을 가짐을 알 수 있다. <그림 2>는 21번 후보점을 첨가한 확장 Snee 계획과 20번 후보점을 첨가한 확장 Snee 계획을 비교한 PVT이다. 이 그림을 통하여 위의 사실을 확인할 수 있고, 특히  $x_1$ 구성성분에 대하여서는 20번 후보점을 첨가한 확장 Snee 계획이 21번 후보점을 첨가한 확장 Snee 계획보다 나음을 알 수 있다. 그러므로, Snee 계획의 확장시 20, 21, 23번 후보점들을 최우선 고려하는 것이 좋다.

18개로 구성된 D-최적화 계획에다 나머지 15개의 후보점들 중 하나를 첨가하여 19개로 구성된 확장 D-최적화 계획을 만들었을 때  $\{\det(X'X)^{-1}\}^{\frac{1}{10}}$ 의 값은 다음 <표 4>와 같다. 그러므로, D-최적화 입장에서는 8번 후보점을 첨가한 확장 D-최적화 계획이 선택되나, PVT를 이용하면 추정반응값 분산의 입장에서는 15번 후보점을 첨가한 확장 D-최적화 계획이 행렬식의 값이 더 큼에도 불구하고 더 좋은 계획임을 알 수 있다. <그림 3>은 8번 후보점을 첨가한 확장 D-최적화 계획과 15번 후보점을 첨가한 확장 D-최적화 계획을 비교한 PVT이다. 이 그림을 통하여 모든 구성성분에 대하여 15번 후보점을 첨가한 확장 D-최적화 계획이 8번 후보점을 첨가한 확장 D-최적화 계획보다 더 나음을 알 수 있다. 그러므로, D-최적화 계획의 확장시 15번 후보점을 최우선 고려하는 것이 좋다.

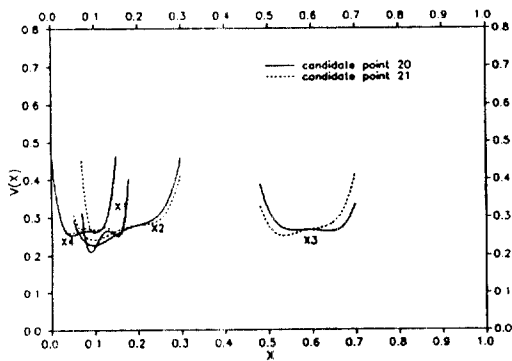
<표 3> 확장 Snee 계획의 행렬식값

<표 4> 확장 D-최적화 계획의 행렬식값

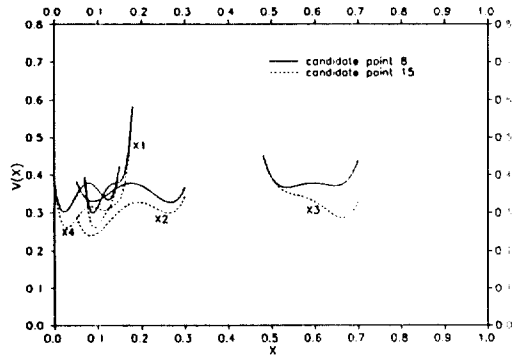
첨가되는 후보점	$\{\det(X'X)^{-1}\}^{\frac{1}{10}}$
13	278.6216
20	272.9759
21	270.6734
22	273.5752
23	273.3536
24	270.9272
25	274.0566
26	273.8824
27	274.1671
28	275.1056
29	274.8459
30	278.4454
31	276.3288
32	277.8139
33	277.3819

첨가되는 후보점	$\{\det(X'X)^{-1}\}^{\frac{1}{10}}$
8	259.3440
11	265.4750
12	261.1651
13	262.4463
15	263.4343
16	264.6034
25	260.6779
26	261.3499
27	259.8243
28	261.9146
29	261.7092
30	262.7353
31	263.8623
32	261.8289
33	260.7839





< 그림 2 > 후보점 20번과 21번이 각각  
첨가된 확장 Snee 계획의  
추정반응값 분산그림



< 그림 3 > 후보점 8번과 15번이 각각  
첨가된 확장 D-최적화 계획의  
추정반응값 분산그림

### 3.2 실험계획점 중 결측값이 발생하는 경우

18개로 구성된 Snee 계획에서 실험계획점을 하나 제거하여 17개로 구성된 축소 Snee 계획을 만들었을 때,  $\{ \det(X'X)^{-1} \}^{1/10}$ 의 값은 다음 <표 5>와 같다. 그러므로, D-최적화의 입장에서는 2번 실험계획점이 결측되는 경우가 제일 나쁘지만, PVT를 이용하면 15번이나 18번 실험계획점이 결측되는 경우가 추정반응값 분산의 변화가 가장 심함을 알 수 있다. 특이한 것은 18번 실험계획점이 결측되는 경우가 가장 행렬식의 값의 변화가 작음에도 불구하고, 추정반응값 분산의 변화가 15번과 아울러 가장 심하다는 것이다. <그림 4>는 2번 실험점이 결측된 축소 Snee 계획과 15번 실험계획점이 결측된 축소 Snee 계획을 비교한 PVT이다. 이 그림을 통하여  $x_3$  구성성분 제한영역의 앞부분을 제외하고는 15번 실험계획점이 결측된 축소 Snee 계획이 2번 실험계획점이 결측된 축소 Snee 계획보다 추정반응값 분산이 더 큼을 알 수 있다. 그러므로, Snee 계획에서 결측값이 발생하는 경우 되도록 15번이나 18번 실험계획이 결측되지 않도록 하여야 할 것이다.

18개로 구성된 D-최적화 계획에서 실험계획점을 하나 제거하여 17개로 구성된 축소 D-최적화 계획을 만들었을 때,  $\{ \det(X'X)^{-1} \}^{1/10}$ 의 값은 다음 <표 6>과 같다. 그러므로, D-최적화의 입장에서는 2번 실험계획점이 결측되는 경우가 제일 나쁘지만, PVT를 이용하면 20번이나 24번 실험계획점이 결측되는 경우가 추정반응값 분산의 변화가 가장 심함을 알 수 있다. 특이한 것은 24번 실험계획점이 결측되는 경우가 가장 행렬식의 값의 변화가 작음에도 불구하고, 추정반응값 분산의 변화가 20번과 아울러 가장 심하다는 것이다. 다음 <그림 5>는 2번 실험계획점이 결측된 축소 D-최적화 계획과 24번 실험계획점이 결측된 축소 D-최적화 계획을 비교한 PVT이다. 이 그림을

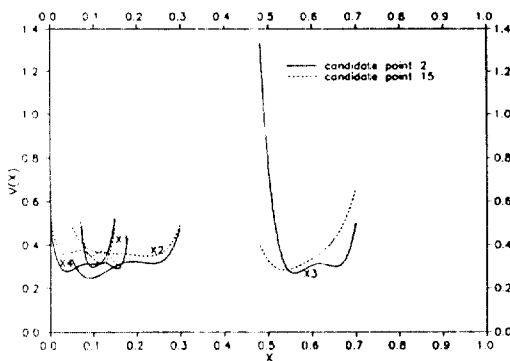
통하여 전 구성성분에 대하여 24번 실험계획점이 결측된 축소 D-최적화 계획이 2번 실험계획점이 결측된 축소 D-최적화 계획보다 추정반응값 분산이 더 크음을 알 수 있다. 그러므로, D-최적화 계획에서 결측값이 발생하는 경우 되도록 20번이나 24번 실험계획점이 결측되지 않도록 하여야 할 것이다.

<표 5> 축소 Snee 계획의 행렬식값

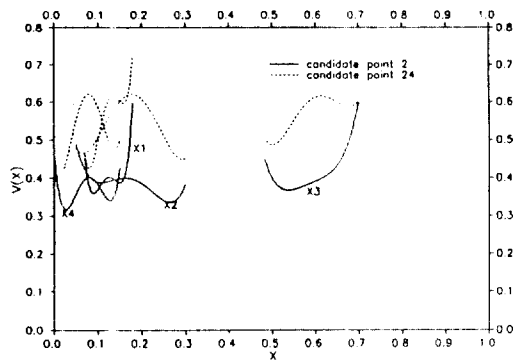
<표 6> 축소 D-최적화 계획의 행렬식값

결측되는 실험계획점	$\{ \det(X'X)^{-1} \}^{\frac{1}{10}}$
1	303.7350
2	372.1460
3	308.8564
4	338.9274
5	334.0396
6	308.6413
7	351.2517
8	303.0896
9	341.3769
10	304.3362
11	302.1102
12	303.0288
14	305.6625
15	301.0223
16	301.4604
17	305.0653
18	295.8741
19	310.8858

결측되는 실험계획점	$\{ \det(X'X)^{-1} \}^{\frac{1}{10}}$
1	296.7972
2	313.0530
3	290.5876
4	303.6475
5	305.9718
6	311.0630
7	311.5841
9	310.8967
10	299.3081
14	287.9610
17	292.8708
18	291.0476
19	289.8432
20	311.7623
21	290.2737
22	287.1621
23	288.8781
24	286.9773



<그림 4> 실험계획점 2번과 15번이 각각 결측된 축소 Snee 계획의 추정반응값 분산그림

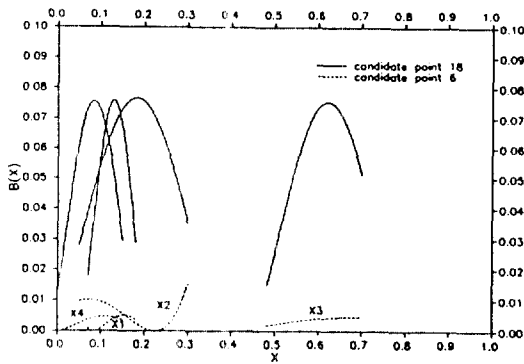


<그림 5> 실험계획점 2번과 24번이 각각 결측된 축소 D-최적화 계획의 추정반응값 분산그림

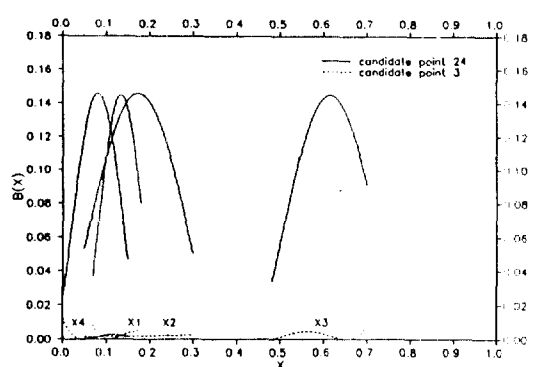
### 3.3 실험계획점 중 이상값이 발생하는 경우

Snee 계획의 실험계획점 중 이상값이 존재하는 경우, 이상값의 크기가 일정하다고 가정하여 각 실험계획점들에 대하여 PBT를 그려보면 추정반응값 편이가 크게 나타나는 실험계획점들은 15번이나 18번 실험계획점들이고, 추정반응값 편이가 작게 나타나

는 실험계획점들은 2번이나 6번 실험계획점들이다. <그림 6>은 각각 실험계획점 6번과 18번이 이상값인 경우의 PBT이다. 이 그림에서 알 수 있는 것은 실험계획점 18번이 이상값인 경우는 각 성분의 제한영역 가운데에서 추정반응값 편이가 제일 크게 나타남을 알 수 있다. 그러므로, 15번이나 18번 실험계획점이 이상점인지의 여부와 이상값인 경우의 처리에 대하여 특히 신경을 써야 한다. D-최적화 계획의 실험계획점 중 이상값이 존재하는 경우, 이상값의 크기가 일정하다고 가정하여 각 실험계획점들에 대하여 PBT를 그려보면 추정반응값 편이가 크게 나타나는 실험계획점들은 20번이나 24번 실험계획점들이고, 추정반응값 편이가 작게 나타나는 실험계획점들은 1번이나 3번 실험계획점들이다. <그림 7>은 각각 실험계획점 3번과 24번이 이상값인 경우의 PBT이다. 이 그림에서 알 수 있는 것은 실험계획점 24번이 이상값인 경우는 각 성분의 제한영역 가운데에서 추정반응값 편이가 제일 크게 나타남을 알 수 있다. 그러므로, 20번이나 24번 실험계획점이 이상점인지의 여부와 이상값인 경우의 처리에 대하여 특히 신경을 써야 한다.



<그림 6> Snee 계획에서 실험계획점 6번과 18번이 각각 이상값인 경우 추정반응값 편이그림



<그림 7> D-최적화 계획에서 실험계획점 3번과 24번이 각각 이상값인 경우 추정반응값 분산그림

#### 4. 결론

혼합물 실험계획의 확장시 기존의 연구들은 D-최적화 계획을 중심으로 전개되어 왔다. 이 논문에서는 이러한 알파벳 최적화 기법들의 문제점을 해결할 수 있는 그래픽 방법을 제안하였다. 이 그래픽 방법인 추정반응값 분산그림 및 추정반응값 편이그림을 이용하여 실험계획의 확장시나 실험계획점들 중 결측값이나 이상값이 발생하는 경우, 제한이 있는 전 실험계획 영역에 걸쳐 추정반응값 분산 및 추정반응값 편이의 변화를 알아 볼 수 있고, 실험계획들을 서로 비교, 평가할 수 있다. 특히, 서로 경쟁적인 혼합물 실험계획에 대하여 제한이 있는 전 실험계획 영역에 걸쳐 추정반응값 분산

및 추정반응값 편이의 변화를 한 눈에 확인할 수 있으므로 실험계획들의 장, 단점을 쉽게 파악할 수 있다. 또한, 한 실험계획에 대하여 실험계획의 확장 또는 결측의 결과를 추정반응값 분산의 변화를 보면서 확인할 수 있다. 또한, 이상값이 존재하는 경우 그 영향을 추정반응값 편이의 변화를 보면서 알 수 있다. 이 과정을 통하여 각 구성 성분들에 대하여 추정반응값 분산 및 추정반응값 편이의 크기 변화를 알아내어 구성 성분들의 중요도에 시열을 매길 수 있다. 이 그래픽 방법은 혼합물 실험계획 영역에 제한이 없는 경우에도 활용할 수 있다.

## 5. 참고문헌

- [ 1 ] 장대홍, 박상현(1995), “순차적 혼합물 실험계획을 평가하기 위한 그래픽 방법”, 「한국통계학회 논문집」 제2권 1호.
- [ 2 ] Cook, R.D. and Weisberg, S.(1982), *Residuals and Influence in Regression*, Chapman and Hall, London.
- [ 3 ] Cook, R.D., Holschuh, N., and Weisberg, S.(1982), “A Note on an Alternative Outlier Model,” *Journal of the Royal Statistical Society*, B44, pp. 370-376.
- [ 4 ] Cornell, J.A.(1990), *Experiments with Mixtures*, John Wiley and Sons, Vol.1, New York.
- [ 5 ] McLean, R.A. and Anderson, V.L.(1966), “Extreme Vertices Design of Mixture Experiments,” *Technometrics*, Vol. 8, pp. 447-454.
- [ 6 ] O’Gorman, M.A.(1984), “Outliers and Robust Response Surface Designs,” Unpublished Ph.D. Thesis, Virginia Polytechnic Institute and State University, Virginia.
- [ 7 ] Snee, R.D.(1975), “Experimental Designs for Quadratic Models in Constrained Mixture Spaces,” *Technometrics*, Vol. 8, pp. 149-159.
- [ 8 ] Vining, G.G., Cornell, I.A. and Myers, R.H.(1993), “A Graphical Approach for Evaluating Mixture Designs,” *Applied Statistics*, Vol. 42, pp. 127-138.