The Korean Communications
in Statistics Vol. 7, No. 1, 2000
pp. 257-267

Optimal Value Estimation Method with Lower and Upper Bounds

Chong Sun Hong[1], Youn Jong Kim[2], Jong Seok Byun[3]

Abstract

As one of indirect ways to get an optimal answer for sensitive questions, both lower and upper values are sometimes asked and collected. In this paper a statistical method is proposed to analyze this kind of data using graphics. This method could define each sample median and estimate an optimal value between lower and upper bounds. In particular, we find that this method has similar explanations of an equilibrium price with demand and supply functions in Economics.

1. 서론

마케팅조사와 같은 설문조사에서 민감한 사항에 대하여는 직접적인 질문을 회피하고 간접적으로 대답을 유도하는 경향이 있다. 이러한 설문 조사에서 소비자가 희망하는 적정 가격을 질문하는데 사용되는 대표적인 방법으로는 적절한 구간으로 설정된 범주형 변수로 수집된 자료를 분석하는 방법이 있으나, 많은 경우에는 조사된 자료를 범주형 변수보다도 연속형 변수로 다루게 된다. 예를 들면 현재 일반적인 마케팅조사에서는 다음과 같은 형태의 구입 가능한 최소/최대가격, 구입 불가능한 최소/최대가격 등의 분리된 질문을 함으로써 가격 민감도 분석으로 특정 제품에 대한 소비자의 구입 희망 적정 가격을 추정하고 있다.(Kotler (1988), Kinnear and Root (1994), Kinnear and Taylor (1996), 그리고 Star and Urban (1988) 참조)

질문1 : M의 가격은 얼마라고 생각하십니까?
질문2 : M의 가격이 싼 편이라고 생각하시는 가격은 얼마입니까?
질문3 : M의 가격이 너무 싸서 품질이 의심스러워지는 가격은 얼마입니까?
질문4 : M의 가격이 비싼 편이라고 생각하시는 가격은 얼마입니까?
질문5 : M의 가격이 너무 비싸서 구입하실 생각이 없어지는 가격은 얼마입니까?

즉, 위와 같은 형태의 문제를 질문함으로써 특정제품 M에 대한 소비자의 적정 희망 가격을 추정할 수 있지만, 실제로 표본조사에에서는 질문문항 수가 많거나 복잡한 질문에 대한 이해의 부족으로

1) Professor, Department of Statistics, Sungkyunkwan University, cshong@skku.ac.kr.
2) Associate Professor, Department of Computer Science and Statistics, Yongin University.
3) Assistant Professor, Department of Statistics and Information, Hanshin University.

인하여 응답거부, 무응답, 응답의 신뢰성 등의 비표본오차를 증가시키게 된다. 따라서 우리는 위와 같은 복잡한 문제를 축소하여 희망 구입 가격의 상한 가격과 하한 가격, 즉 다음과 같은 구입 가능한 최소 가격과 구입 불가능한 최대 가격을 질문함으로써 소비자가 구입을 희망하는 적정 가격의 추정 방법에 대해 검토해 보고자 하는 것이다. 예를 들어 다음과 같은 설문을 고려하자.

"질문 1 : M의 가격이 최소 어느 정도이면 구입하시겠습니까?"
"질문 2 : M의 가격이 어느 정도 이상이면 구입을 포기하시겠습니까?"

본 논문에서는 위와 같은 질문으로 소비자 성향 조사를 실시했다고 가정한 상황에서 논의를 계속하고자 한다. 물론 소비자의 구매 희망가격은 소비자가 원하는 것 보다 축소하여 대답하는 경향이 있으며, 생산자의 판매 희망가격은 생산자가 원하는 것 보다 확대하여 가격을 결정하는 경향이 있겠다. 이 논문에서는 생산자와 소비자 사이에서 가격을 결정하는 문제를 다루는 것이 아니고, 단지 소비자의 최소와 최대 희망가격 자료가 주어진 상황에서 그래픽을 이용하여 적정한 희망가격을 추정하는 비모수적인 방법을 제안하고자 한다. 이 방법의 특징은 많은 사람들에게 이미 친숙한 경제학의 기초 이론인 수요와 공급함수, 그리고 균형가격의 설명과 유사하다는 점이다. 따라서 경제학의 기초적인 균형가격을 이해하는 많은 사람들에게 본 논문에서 제안한 방법은 설명하기가 쉽고 이해하기가 명확하다고 할 수 있다.

2절에서는 분포함수의 대칭인 부분포함수(負分布函數)를 정의하고 이 분포들을 통하여 중앙값을 정의하고, 상한값들과 하한값들을 나타내는 확률표본들에서 각각의 표본중앙값을 구하는 방법과 상한값들과 하한값들의 사이에 존재하는 최적값을 추정하는 방법에 대하여 논한다. 그리고 3절에서는 1994년도에 실시한 'PCS 사업추진을 위한 시장조사'에 적용시킨 예제를 들어 실증적으로 설명하였다.

2. 추정 방법

분포함수 $F(\cdot)$를 따르는 확률변수 X를 고려하자. 이때 확률변수 X의 부분포함수(negative distribution function) $G(\cdot)$를 다음과 같이 정의한다.

<정의>

확률변수 X의 부분포함수 $G(\cdot)$는 임의의 실수 x에 대하여 다음과 같이 정의한다.

$$G(x) = 1 - F(x). \tag{1}$$

부분포함수 $G(\cdot)$의 성질은 분포함수 $F(\cdot)$의 성질과 유사하게 언급할 수 있는데 이를 포함하여 부분포함수에 관한 많은 이론 중에서 우리의 관심 대상인 부분인 중앙값에 대하여 우선 언급하여 보자. 확률변수 X의 중앙값(median)은 분포함수와 부분포함수를 이용하여 다음과 같이 표현할 수 있다.

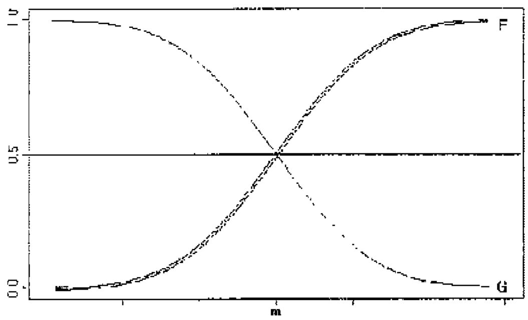
<정리 1>

확률변수 X의 중앙값, m은 다음을 만족한다.

$$G(m) \le 1/2 \le F(m) \text{ 와 } F(m^-) \le 1/2 \le G(m^-), \tag{2}$$

여기서 m^-는 X의 값이 m의 왼쪽으로부터의 극한값을 의미한다. 즉, $F(m^-) = \lim_{h \to 0^+} F(m-h) = P(X < m)$이며, $G(m^-) = P(X \ge m)$이다.

(증명) Bickel and Doksum (1977, p54), Hogg and Craig (1995, p44), Lindgren (1993, p51) 그리고 Degroot (1986, p207)를 참조하면, 중앙값에 대한 정의는 모두 동일한 것임이 확인되며 식 (2)를 유도할 수 있다. ■



<그림 1>

Mood, Graybill, and Boes(1974, p73)는 연속형 확률변수의 경우에 중앙값을 정의하였는데 특히, 연속형인 경우에 분포함수와 부분포함수를 사용하여 중앙값을 묘사하는 그림을 <그림 1>로 나타내었다. <그림 1>로부터 $F(\cdot)$와 $G(\cdot)$가 교차하는 점의 수직축 값은 0.5를 나타내며, 수평축의 좌표가 중앙값을 의미한다는 사실을 얻을 수 있다. 즉, 식 (2)는 다음과 같이 변환된 조건으로 나타난다.

$$F(m) = G(m) = 1/2.$$

그리고 이산형 확률변수의 경우에는 확률표본의 분포함수와 부분포함수인 경우와 일치하므로 <그림 2>와 <그림 3>에서 설명하기로 한다.

분포함수 $F(\cdot)$를 따르는 모집단에서 두 개의 확률표본(random sample) $\{X_1, X_2, \ldots, X_n\}$과 $\{Y_1, Y_2, \ldots, Y_m\}$를 추출했다고 가정하자. 이때 각각의 표본 크기 n과 m은 서로 다른 상수이나 본 논문에서는 논문의 간결성을 유지하고자 동일한 표본크기 n을 취하고자 한다. 이와 같은 확률표본의 예로, 이미 1절에서 소개한 마케팅조사에서 소비자가 원하는 희망가격과 같이 민감한 질문을 조사하는 간접적인 질문의 방법으로 제시된 '질문 1'과 '질문 2'에 대하여 임의로 추출된 n명에게 조사되어 수집된 자료 중 '질문 1'의 하한값 자료와 '질문 2'의 상한값 자료를 각각 $\{x_1, x_2, \ldots, x_n\}$과, $\{y_1, y_2, \ldots, y_n\}$으로 표현하자.

하한값과 상한값에 대한 확률표본의 표본분포함수(sampling distribution function)를 다음과 같

이 각각 $F^L{}_n(\cdot)$과 $F^U{}_n(\cdot)$으로 정의한다. 임의의 실수 x에 대하여,

$$F^L{}_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \le x), \quad F^U{}_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(Y_i \le x), \tag{3}$$

여기서 $I(\cdot)$는 지시함수(indicator function)이다.

따라서 (1)식을 이용하여 (3)식에서 정의된 표본분포함수들에 대한 부표본분포함수(negative sampling distribution function)는 다음과 같이 표현할 수 있다. 임의의 실수 x에 대하여,

$$G^L{}_n(x) = 1 - F^L{}_n(x), \quad G^U{}_n(x) = 1 - F^U{}_n(x). \tag{4}$$

우리는 (3)과 (4)에서 언급한 두 개의 표본분포함수와 두 개의 부표본분포함수를 하나의 좌표에 모두 표현할 수 있는데 이에 대한 전체적인 설명은 뒤로 미루며 우선 표본중앙값(sample median)을 구하는 방법을 살펴보자.

<보조정리 1>

표본분포함수 $F^L{}_n(\cdot)$과 부표본분포함수 $G^L{}_n(\cdot)$를 따르는 확률표본 $\{X_1, X_2, \ldots, X_n\}$에 대하여 임의의 상수 m_L이 다음의 조건을 만족한다면,

$$G^L{}_n(m_L) \le F^L{}_n(m_L) \text{ 와 } G^L{}_n(m_L^-) \ge F^L{}_n(m_L^-), \tag{5}$$

m_L는 확률표본자료 $\{x_1, x_2, \ldots, x_n\}$의 표본중앙값이다. 그리고 확률표본 $\{Y_1, Y_2, \ldots, Y_n\}$의 표본분포함수 $F^U{}_n(\cdot)$와 부표본분포함수 $G^U{}_n(\cdot)$에 대하여

$$G^U{}_n(m_U) \le F^U{}_n(m_U) \text{ 와 } G^U{}_n(m_U^-) \ge F^U{}_n(m_U^-) \tag{6}$$

의 조건을 만족하는 상수 m_U는 확률표본자료 $\{y_1, y_2, \ldots, y_n\}$의 표본중앙값이다.

(증명) <보조정리 1>의 (5)와 (6)은 <정리 1>의 관계 조건 (2)와 동일하다. ■

<보조정리 1>에 대하여 표본중앙값 존재의 유일성에 따라 자세히 나누어 서술하면 다음과 같다. (j)번째 순서통계량과 (j+1)번째 순서통계량 사이의 임의의 상수 m에 대하여

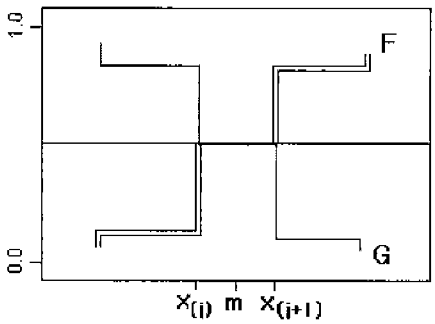
$$F_n(m) = G_n(m) = 1/2$$

을 만족한다면, (j)번째 순서통계량과 (j+1)번째 순서통계량 사이의 구간은 표본중앙값이다. 계단함수(step function)인 $F_n(\cdot)$과 $G_n(\cdot)$을 평면 위에 그림으로 나타낼 때, 계단의 높이 부분에 해당되는 부분을 선으로 연결하여 명백한 계단 모양으로 묘사한 <그림 2>를 살펴보자. <그림 2>에서 구간으로 존재하는 표본중앙값은 $F_n(\cdot)$을 기반으로 하는 올라가는 계단 모양과 $G_n(\cdot)$을 기반으로 하는 내려가는 계단 모양이 서로 만나는 부분인 구간 $[x_{(j)}, x_{(j+1)})$과 일치함을 파악할 수 있다. 그리고 임의의 상수 m에 대하여

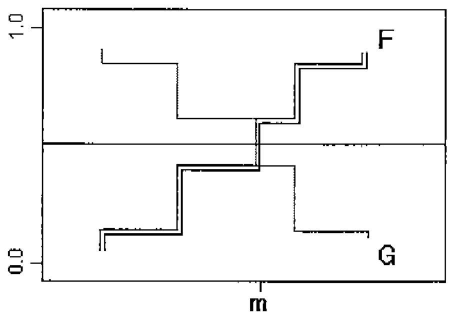
$$G_n(m) < 1/2 < F_n(m) \text{ 와 } G_n(m^-) > 1/2 > F_n(m^-)$$

이 만족하면, 이 때의 상수 m은 확률표본자료의 유일한 표본중앙값이다. <그림 3>도 계단함수인 $F_n(\cdot)$과 $G_n(\cdot)$을 계단 모양으로 표현하여 두 계단이 올라가고 내려가면서 서로 교차되는 점

이 표본중앙값임을 나타내어 준다. 특히 이때의 표본중앙값은 유일하게 존재하는 경우이다.



<그림 2>



<그림 3>

이 그림들을 통하여 확률표본의 표본중앙값은 표본분포함수 $F_n(\cdot)$와 부표본분포함수 $G_n(\cdot)$을 기반으로 하는 계단 모양이 겹쳐지거나 서로 교차할 때의 값임을 쉽게 알 수 있다. 그러므로 표본중앙값을 구하는 방법을 정리하면 다음과 같다.

<추정 방법 1>

표본분포함수 $F_n(\cdot)$과 부표본분포함수 $G_n(\cdot)$에서 계단의 높이가 되는 부분을 선으로 연결하고 계단 모양으로 표현하여 이차원 평면에 나타내어 보자. 이러한 두 계단 모양의 분포함수가 겹치거나 교차되는 점의 X축 값은 표본중앙값이다. 그러므로 확률표본 $\{X_1, X_2, \ldots, X_n\}$과 $\{Y_1, Y_2, \ldots, Y_n\}$에 대하여, 표본분포함수 $F^L{}_n(\cdot)$을 기반으로 하는 올라가는 계단 모양과 부표본분포함수 $G^L{}_n(\cdot)$을 기반으로 하는 내려가는 계단 모양이 서로 만나거나 교차될 때의 값은 표본자료 $\{x_1, x_2, \ldots, x_n\}$의 표본중앙값, m_L,이며, $F^U{}_n(\cdot)$를 기반으로 하는 계단 모양과 $G^U{}_n(\cdot)$를 기반으로 하는 계단 모양이 서로 만나거나 교차될 때의 값은 표본자료 $\{y_1, y_2, \ldots, y_n\}$의 표본중앙값, m_U이다. ■

뒤에서 언급할 <그림 5>를 살펴보면서 m_L과 m_U를 이해하여 보자. 이와 같은 방법으로 하한값의 중앙값 m_L과 상한값의 중앙값 m_U사이에 존재하는 최적값에 대하여 토론하여 보자.

<정리 2>

하한값을 나타내는 확률변수 X의 분포함수 $F^L(\cdot)$과 상한값을 나타내는 확률변수 Y의 부분포함수 $G^U(\cdot)$에 대하여

$$G^U(m_0) \le F^L(m_0) \text{ 와 } G^U(m_0^-) \ge F^L(m_0^-) \tag{7}$$

을 만족하는 m_0가 존재한다면, 이 값은 확률변수 X의 부분포함수 $G^L(\cdot)$과 확률변수 Y의 분포함수 $F^U(\cdot)$에 대하여 다음을 만족한다.

$$G^L(m_0) \le F^U(m_0) \text{ 와 } G^L(m_0^-) \ge F^U(m_0^-). \tag{8}$$

그리고 (7)과 (8)의 조건을 만족하는 m_0는 반드시 존재한다.

(증명) (7)과 (8)의 조건을 만족하는 값 m_0는 동일하다는 것은 쉽게 보일 수 있으며, 비록 확률변수 X와 Y의 꼬리 확률(tail probability),

$$G^L(m_0^-) = P(X \ge m_0) \text{와 } F^U(m_0) = P(Y \le m_0) \ ,$$

의 값이 0에 가깝다 하더라도, $F(\cdot)$는 비감소함수(non-decreasing function)이며 $G(\cdot)$는 비증가함수(non-increasing function)이므로 $F^L(\cdot)$과 $G^U(\cdot)$또는 $G^L(\cdot)$과 $F^U(\cdot)$는 교차하거나 수직축의 값이 최대 1 또는 최소 0에서 겹친다. 이 경우에는 m_0는

$$G^U(m_0) = F^L(m_0) = 1 \text{ 또는 } G^L(m_0) = F^U(m_0) = 0$$

을 만족하게 된다. 그러므로 (7)과 (8)의 조건을 만족하는 m_0는 반드시 존재한다.

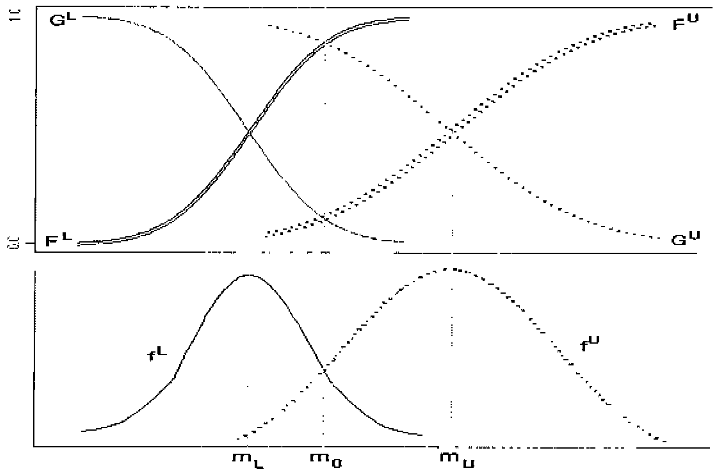
만약 확률변수 X와 Y가 연속형이라면, (7)과 (8)을 만족하는 m_0는 다음과 같은 성질로 변환된다.

$$G^L(m_0) = F^U(m_0) \text{ 또는 } G^U(m_0) = F^L(m_0) \ .$$

그리고 이를 확률로 정리하면 다음과 같이 동일한 크기를 갖는 꼬리 확률로 표현된다.

$$P(X \ge m_0) = P(Y \le m_0) \text{ 또는 } P(Y \ge m_0) = P(X \le m_0) \ .$$

연속형 확률변수의 경우에 X와 Y의 분포함수 $F^L(\cdot)$와 $F^U(\cdot)$와 부분포함수 $G^L(\cdot)$와 $G^U(\cdot)$를 동시에 표현하고 이에 대응하는 확률밀도함수 $f^L(\cdot)$와 $f^U(\cdot)$들을 동일한 수평축에 대하여 표현한 <그림 4>를 살펴보자.



<그림 4>

<그림 4>의 하단부분의 확률밀도함수의 그림을 통하여 $P(X \ge m_0) = P(Y \le m_0)$을 만족하는

m_0 값을 구하는 방법은 판별 및 분류분석(discrimination and classification analysis)의 기법에서 사용되는 기본 개념과 매우 유사함을 파악할 수 있다(자세한 것은 Johnson and Wichern (1992) 참조).

<보조정리 2>

하한값을 나타내는 확률표본 $\{X_1, X_2, \ldots, X_n\}$의 표본분포함수 $F^L{}_n(\cdot)$을 기반으로 하는 계단 모양과 상한값을 나타내는 확률표본 $\{Y_1, Y_2, \ldots, Y_n\}$의 부표본분포함수 $G^U{}_n(\cdot)$을 기반으로 하는 계단 모양이 서로 교차하거나 만나는 점을 m_0이라면, 이 값은 다음과 같은 성질을 갖고 있다.

$$G^U{}_n(m_0) \le F^L{}_n(m_0) \text{ 와 } G^U{}_n(m_0^-) \ge F^L{}_n(m_0^-). \tag{9}$$

그리고 이때의 m_0에서는 하한값을 나타내는 확률표본의 부표본분포함수 $G^L{}_n(\cdot)$을 기반으로 하는 계단 모양과 상한값을 나타내는 확률표본의 표본분포함수 $F^U{}_n(\cdot)$을 기반으로 하는 계단 모양이 교차하거나 만나면서 다음을 만족한다.

$$G^L{}_n(m_0) \le F^U{}_n(m_0) \text{ 와 } G^L{}_n(m_0^-) \ge F^U{}_n(m_0^-). \tag{10}$$

그리고 (9) 또는 (10)의 조건을 만족하는 m_0는 하한값의 표본중앙값 m_L과 상한값의 표본중앙값 m_U사이에 반드시 존재한다.

(증명) <보조정리 2>의 (9), (10)의 관계 조건은 <정리 2>의 관계 부등식 (7), (8)과 각각 동일하다. m_L과 m_U사이에 존재하는 m_0의 존재성을 증명하기에 앞서 확률변수 X의 분포함수와 부분포함수는 다음과 같은 성질을 갖고 있음을 상기하자. 임의의 실수 x에 대하여

$$F^U(x) \le F^L(x) \text{ 와 } G^L(x) \le G^U(x).$$

그리고 (9)와 (10)의 조건을 만족하는 m_0가 하한값의 표본중앙값 m_L과 상한값의 표본중앙값 m_U사이에 존재하지 않는다고 가정하자. 우선 m_0가 m_L^-보다 작다면 (즉 $m_0 < m_L^- < m_L$), (10)의 첫 번째 부등식의 왼쪽 항은 비증가함수인 $G_n(\cdot)$의 성격을 이용하여

$$G^L{}_n(m_L^-) \le G^L{}_n(m_0)$$

을 얻었고, 오른쪽 항은 다음과 같은 부등식을 얻을 수 있다.

$$F^U{}_n(m_0) \le F^L{}_n(m_0) \le F^L{}_n(m_L^-)$$

그러므로

$$G^L{}_n(m_L^-) \le F^L{}_n(m_L^-)$$

이 성립하는데 이는 (5)의 두 번째 부등식에 모순이 된다. 다음으로 m_0가 m_U보다 크다고 가정하자(즉 $m_U < m_0^- < m_0$). (10)의 두 번째 부등식은

$$G^U{}_n(m_U) \ge G^U{}_n(m_0^-) \ge G^L{}_n(m_0^-) \ge F^U{}_n(m_0^-) \ge F^U{}_n(m_U)$$

을 만족하는데 이는 (6)의 첫 번째 부등식에 모순이 된다. 그러므로 (9)와 (10)을 만족하는 m_0

는 반드시 하한값의 표본중앙값 m_L과 상한값의 표본중앙값 m_U사이에 존재한다.

위의 <보조정리 2>에서 $F^U(\cdot)$와 $G^L(\cdot)$을 기반으로 하는 계단 모양들이 교차하여 발생할 수 있는 모든 경우를 고려하면, m_0값이 유일하게 존재하는 경우와 구간으로 존재하는 경우로 나눌 수 있다.

직접적인 대답을 유도하기 곤란한 민감한 질문에 대하여 우리가 구하기 원하는 최적값은 상한과 하한값을 대신 질문하여 간접적인 대답으로부터 얻고자 한다. 이런 경우에 상한값의 표본자료 $\{y_1, y_2, \dots, y_n\}$의 표본중앙값 m_U는 구입 가능한 상한가의 금액 및 상한가의 구입의사 비율에 대한 중앙값이며, 하한값의 표본자료 $\{x_1, x_2, \dots, x_n\}$의 표본중앙값 m_L은 구입 가능한 하한가의 금액 및 하한가의 구입의사 비율에 대한 중앙값이 된다. 그리고 m_U와 m_L사이에 존재하는 m_0를 구입 가능한 하한가의 금액 및 구입의사 비율과 구입 가능한 상한가의 금액과 구입의사 비율과의 교차점으로 소비자가 구입을 희망하는 적정 가격으로써 우리가 원하는 민감한 질문에 대답이 되는 최적값으로 간주할 수 있다. 그러므로 상한값의 표본중앙값과 하한값의 표본중앙값 사이에 존재하는 최적값을 추정하는 방법을 다음과 같이 설정한다.

<추정 방법 2>

하한값과 상한값을 나타내는 확률표본 $\{X_1, X_2, \dots, X_n\}$과 $\{Y_1, Y_2, \dots, Y_n\}$의 표본분포함수 ${F^L}_n(\cdot)$과 ${F^U}_n(\cdot)$, 그리고 두 확률표본의 부표본분포함수 ${G^L}_n(\cdot)$과 ${G^U}_n(\cdot)$에서 계단의 높이가 되는 부분을 선으로 연결하여 두 개의 표본분포함수는 올라가는 계단 모양으로, 그리고 두 개의 부표본분포함수는 내려가는 계단 모양으로 묘사하자. 이때 두 분포 ${F^L}_n(\cdot)$과 ${G^U}_n(\cdot)$ (또는 ${G^L}_n(\cdot)$과 ${F^U}_n(\cdot)$)를 기반으로 하는 계단 모양이 서로 만나거나 교차되는 값을 하한값의 표본중앙값과 상한값의 표본중앙값 사이의 최적값으로 추정한다. ■

하한값과 상한값에 대한 확률표본의 표본분포함수 ${F^L}_n(\cdot)$와 ${F^U}_n(\cdot)$, 그리고 본 논문에서 제안한 부표본분포함수 ${G^L}_n(\cdot)$, ${G^U}_n(\cdot)$을 기반으로 하는 4개의 계단 모양들을 모두 한 좌표에 표현하면 3절의 <그림 5>와 같이 나타낼 수 있다. 여기에서 ${F^L}_n(\cdot)$과 ${G^L}_n(\cdot)$을 기반으로 하는 계단 모양이 교차하는 점은 하한값의 표본자료 $\{x_1, x_2, \dots, x_n\}$의 표본중앙값 m_L이며, ${F^U}_n(\cdot)$과 ${G^U}_n(\cdot)$을 기반으로 하는 계단 모양이 교차하는 점은 상한값의 표본자료 $\{y_1, y_2, \dots, y_n\}$의 표본중앙값 m_U를 나타낸다. 그리고 ${F^L}_n(\cdot)$과 ${G^U}_n(\cdot)$ 또는 ${G^L}_n(\cdot)$과 ${F^U}_n(\cdot)$을 기반으로 하는 계단 모양이 서로 교차하는 값은 동일하고, 하한값의 표본중앙값 m_L과 상한값의 표본중앙값 m_U사이의 최적값 m_0로 추정됨을 살펴볼 수 있다. 만약 표본자료 $\{x_1, x_2, \dots, x_n\}$의 최대값 $x_{(n)}$이 표본자료 $\{y_1, y_2, \dots, y_n\}$의 최소값 $y_{(1)}$보다 작으면, ${F^L}_n(\cdot)$과 ${G^U}_n(\cdot)$ 또는 ${G^L}_n(\cdot)$과 ${F^U}_n(\cdot)$을 기반으로 하는 계단 모양이 1 또는 0의 수직축 값에

서 만난다. 이 때의 최적값 m_0는 $x_{(n)}$과 $y_{(1)}$ 사이의 구간 $[x_{(n)}, y_{(1)})$으로 설명되고, 일반적으로 $(x_{(n)}+y_{(1)})/2$을 최적값으로 취한다.

3. PCS의 가입비에 대한 시장조사 실증 예제

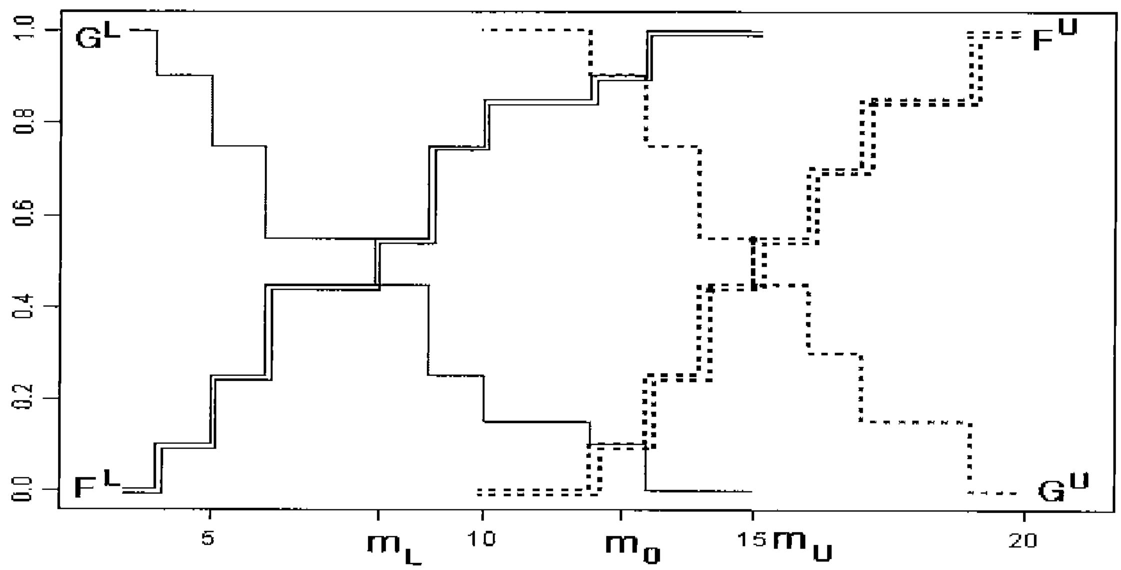
1994년도에 실시된 'PCS 사업추진을 위한 시장조사'에서 소비자가 희망하는 적정 가격의 추정 문제에 적용하여 살펴보기로 하자. 이 조사에서 PCS의 가입비에 대한 소비자의 희망가격 수준을 1절에서 예로 제시한 '질문 1'과 '질문 2'와 같이 희망최소가격(하한값)과 희망최대가격(상한값)에 대하여 따로따로 질문을 하여 수집하였으며 오름차순으로 재배열한 자료는 <표 1>과 같다. 그리고 <표 1>에 희망최소가격과 최대가격에 대한 표본분포함수와 부표본분포함수를 같이 구하였으며, <표 1>에서 $F^L{}_n(x)$는 하한 가격에 대한 표본분포함수로써 금액 x까지의 구입의사 비율을, $G^U{}_n(y)$는 상한 가격에 대한 부표본분포함수로써 금액 y까지의 구입의사 비율을 나타낸 것으로 생각할 수 있다. 이 네 가지 함수들을 기반으로 하는 계단 모양들을 <그림 5>와 같이 하나의 이차원 평면상에 표현하였다.

<표 1>

<table>
<tr><th>질문1
(하한값, x)</th><th>$F^L{}_n(x)$</th><th>$G^L{}_n(x)$</th><th>질문2
(상한값, y)</th><th>$F^U{}_n(y)$</th><th>$G^U{}_n(y)$</th></tr>
<tr><td>$x<4$</td><td>0</td><td>1</td><td>$y<8$</td><td>0</td><td>1</td></tr>
<tr><td>4</td><td rowspan="2">2/20</td><td rowspan="2">18/20</td><td>12</td><td>1/20</td><td>19/20</td></tr>
<tr><td>4</td><td>12</td><td>2/20</td><td>18/20</td></tr>
<tr><td>5</td><td rowspan="3">5/20</td><td rowspan="3">15/20</td><td>13</td><td rowspan="3">5/20</td><td rowspan="3">15/20</td></tr>
<tr><td>5</td><td>13</td></tr>
<tr><td>5</td><td>13</td></tr>
<tr><td>6</td><td rowspan="4">9/20</td><td rowspan="4">11/20</td><td>14</td><td rowspan="4">9/20</td><td rowspan="4">11/20</td></tr>
<tr><td>6</td><td>14</td></tr>
<tr><td>6</td><td>14</td></tr>
<tr><td>6</td><td>14</td></tr>
<tr><td>8</td><td rowspan="2">11/20</td><td rowspan="2">9/20</td><td>15</td><td rowspan="2">11/20</td><td rowspan="2">9/20</td></tr>
<tr><td>8</td><td>15</td></tr>
<tr><td>9</td><td rowspan="4">15/20</td><td rowspan="4">5/20</td><td>16</td><td rowspan="3">14/20</td><td rowspan="3">6/20</td></tr>
<tr><td>9</td><td>16</td></tr>
<tr><td>9</td><td>16</td></tr>
<tr><td>9</td><td>17</td><td rowspan="3">17/20</td><td rowspan="3">3/20</td></tr>
<tr><td>10</td><td rowspan="2">17/20</td><td rowspan="2">3/20</td><td>17</td></tr>
<tr><td>10</td><td>17</td></tr>
<tr><td>12</td><td>18/20</td><td>2/20</td><td>19</td><td rowspan="4">1</td><td rowspan="4">0</td></tr>
<tr><td>13</td><td rowspan="3">1</td><td rowspan="3">0</td><td>19</td></tr>
<tr><td>13</td><td>19</td></tr>
<tr><td>$x>13$</td><td>$y>19$</td></tr>
</table>

(단위: 만원)

3절에서 설명된 <그림 5>는 <표 1>의 자료를 그래픽으로 구현한 그림이다. <그림 5>를 통하여 이미 2절에서 언급하였듯이 하한값(질문 1)에 대한 표본분포함수와 부표본분포함수가 만나는 점을 하한값의 표본중앙값(m_L)으로 추정할 수 있으며, 상한값(질문 2)에 대한 표본분포함수와 부표본분포함수가 교차하는 점이 상한값의 표본중앙값(m_U)이다. 또한 하한값의 표본분포함수와 상한값의 부표본분포함수(또는 하한값의 부표본분포함수와 상한값의 표본분포함수)가 교차하는 부분을 소비자가 원하는 최적의 희망가격(m_0)으로 추정할 수 있다. 이 실증 예제에서 하한값에 대한 적정 최소값으로 m_L =8만원이며, 상한값에 대한 적정 최대값으로는 m_U =15만원으로 추정된다. 그리고 소비자가 희망하는 적정 가격은 12만원에서 13만원 사이의 구간으로 나타나기 때문에 이 적정가격대의 중앙값인 m_0 = 12만 5천원을 적정 최적가격으로 결정할 수 있으며, 하한 가격과 상한 가격에 대한 구입의사 비율을 동시에 최대로 하는 금액이 되기 때문에 적정 최적 가격으로 결정할 수 있다.



<그림 5>

4. 결론

본 논문에서 제안한 상한값과 하한값들의 표본중앙값과 상한값과 하한값 사이의 최적값을 추정하는 방법은 경제학에서 사용하는 수요(supply)와 공급(demand)함수의 설명 그리고 균형가격(equilibrium price)의 생성 과정과 설명이 유사하다는 점이 특징이다. 이 이론은 김대식, 노영기, 안국신(1994, p69-70), 이학용(1996, p44), 정현식(1988, p466-469), 조순, 정운채(1993, p69-70), 주명건(1996, p42-52) 등에 자세히 설명되어 있다. 따라서 경제학의 균형가격을 이해할 정도로 경제학에 친숙한 많은 사람들에게 본 논문에서 제안한 방법은 설명하기가 쉽고 이해하기가 명확하여

설득력이 있다고 할 수 있다.

주식시장에서 매수호가(買收呼價)의 확률표본을 $\{X_1, X_2, \ldots, X_n\}$으로 그리고 매도호가(賣渡呼價)의 확률표본을 $\{Y_1, Y_2, \ldots, Y_m\}$으로 간주한다면, 본 논문에서 논의한 최적값 m_0는 동시호가(同時呼價)에서의 체결가(締結價)이다. 그러므로 본 논문의 내용은 주식의 체결가를 결정하는 문제로 확장 및 응용될 수 있다. 이와 같이 본 논문에서 다루는 최적값의 추정 방법은 선물시장(先物市場)과 경매 등의 현실 생활에서 많이 응용될 수 있겠다.

마지막으로 본 논문에서 제안한 최적값을 추정하는 방법은 다변량분석 기법들 중의 하나인 판별 및 분류분석의 기법에서 두 집단을 분류하는데 사용되는 기본 개념과 유사함을 발견하였다.

참고문헌

[1] 김대식, 노영기, 안국신 (1994). 「현대경제학원론(3판)」, 박문사, 서울.
[2] 이학용 (1996). 「미시경제이론(6판)」, 다산출판사, 서울.
[3] 정현식 (1988). 「미시경제학」, 박영사, 서울.
[4] 조순, 정운채 (1993). 「경제학원론(5판)」, 박문사, 서울.
[5] 주명건 (1996). 「경제학원론(6판)」, 박영사, 서울.
[6] Bickel, P. J., and K. A. Doksum (1977), *Mathematical statistics*, Holden-Day, Inc., San Francisco.
[7] Degroot, M. H. (1986), *Probability and statistics*, 2nd edi., Addison-Wesley, Reading.
[8] Hogg, R. V., and A. T. Craig (1995), *Introduction to mathematical statistics*, 5th edi., Prentice Hall, Inc., New Jersey.
[9] Kinnear, T. C., and A. R. Root (1994), *1994 Survey of Marketing Research*, Chicago: American Marketing Association, 42.
[10] Kinnear, T. C., and J. R. Taylor (1996), *Marketing Research*, 5th edi., Mcgraw-hill, Inc., London.
[11] Kotler, P. (1988), *Marketing Management*, 6th edi., Prentice-Hall, Inc., New Jersey.
[12] Johnson, R, and D. Wichern (1992), *Applied Multivariate Statistical Analysis*, 3rd edi., Prentice Hall, Inc., New Jersey.
[13] Lindgren, B. W. (1993), *Statistical theory*, 4th edi., Chapman-Hall, New York.
[14] Mood, A. M., F. A. Graybill, and D. C. Boes (1974), *Introduction to the theory of statistics*, 3rd edi., Mcgraw Hill, London.
[15] Star, S. H., and Urban, G. L. (1988), *The Case of the Test Market Toss-up*, Harvard Business Review, 4-7, September-October 1988.