품질경영학회지 제23권 제3호 1995년 9월

▩ 연구논문

2-단계 확률화응답모형에 대한 베이즈 선형추정량에 관한 연구

염준근 · 손창균
동국대학교 통계학과

A Study on the Bayes Linear Estimator for the 2-stage Randomized Response Models

Joon-Keun Yum · Chang-Kyoon Son
Dept. of Statistics, Dong Guk University

Abstract

This paper describes the 2-stage randomized response model in the Bayesian view point. The classical Bayesian analysis needs the complete information for a prior density, but the Bayes linear estimator needs only the first and second moments. Therefore, it is convenient to find the estimator and this estimator robusts to a prior density.

We show that MSE's of the Bayes linear estimators for the 2-stage randomized response models are smaller than those of the MLE's for the 2-stage randomized response models.

1. 서론

사회적으로 민감한 조사(탈세, 마약복용, 낙태, 에이즈 등)에서 응답자들에게 직접질문기법을 사용하는 경우에 무응답이나 응답회피 등과 같은 응답편의(response bias)가 발생하게 된다. 이러한 응답편의를 줄이기 위하여 Warner(1965)는 확률장치를 이용하는 간접질문기법인 확률화응답모형을 제안하였다. Warner는 민감한 질문과 관련된 설문을 사용하는 관련질문모형을 제안하였으며, Horvitz, Shah and Simmons(1967)은 민감한 설문과 관계가 없는 설문을 사용하는 무관질문모형을 설계하였다. 이러한 무관질문모형은 Greenberg, Abul-Ela, Simmons and Horvitz(1969)에 의해 이론적인 체계가 완성되었으며, Moors(1971)와 Folsom, Greenberg, Horvitz and Abernathy(1973)은 이를 개

선 · 보완하였고, Drane(1975)은 민감한 설문과 강요된 설문을 사용하는 강요질문모형을 제안하였다.

또한, Mangat와 Singh(1990)은 2개의 확률장치를 이용하는 2단계 확률화응답모형을 제안하였으며, 김종호, 류제복 그리고 이기성(1992)은 2단계 확률화응답모형에 무관질문모형과 강요질문모형을 적용하여 2단계 확률화응답모형을 발전시켰다. Leyseiffer와 Warner(1976)는 이러한 확률화응답모형들이 응답자의 신분보호 측면과 정보 손실의 한계 측면에서 볼 때 어떤 절충이 필요하다는 사실을 지적하였다. 이러한 이유로 많은 학자들이 확률화응답모형에 베이지안 분석방법을 이용하게 되었으며, 실제로 유한 모집단의 경우에 대해 많은 연구가 진행되었다. Winkler와 Franklin(1979)은 베이지안 분석방법을 Warner 모형에 적용하여 사후분포로 매우 복잡한 혼합베타분포를 구하였고, 보다 용이한 방법으로는 사후분포에 대해 2가지의 다른 베타분포의 접근을 제안하였다. 그리고 Pitz(1980)는 무관질문모형을 이용하여 베이지안 접근을 시도하였다. O'Hagan(1987)은 확률화응답모형에 베이지안 분석방법을 적용하는데 있어서 복잡한 사후분포를 이용하여 추정량을 구하기 보다는 이론적으로 간단하고 또한 사전분포에 대해 영향을 받지 않는 베이즈 선형추정량을 이용하는 것이 바람직하다고 하였다. 또한, Erisson(1988)은 초모집단(superpopulation)의 경우 최량선형 불편추정량(BLUE)이 베이즈 선형추정량과 일치함을 보였다.

본 논문에서는 Warner(1976)의 확률화 응답 모형에 베이즈 선형추정량을 적용한 O'Hagan(1987)의 이론을 2단계 확률화응답모형에 적용하여 2단계 확률화응답모형의 베이즈 선형추정량을 구하고, Mangat와 Singh(1990)의 2단계 확률화응답모형의 추정량에 대한 베이즈 선형추정량의 효율성을 비교하고자 한다.

2. 베이즈 선형추정량

2.1 베이즈 선형추정량

베이즈 선형추정량은 Whittle(1958), Stone(1963), Hartigan(1969), 그리고 Goldstein(1975) 등에 의해 연구되었으며, Smouse(1984)와 O'Hagan(1985)은 이러한 베이지안 분석방법을 표본조사에 응용하였다. Smouse는 초모집단의 개념을 도입하여 접근하였으며, O'Hagan은 보다 간단한 주변분포의 접근을 시도하였다.

Royall(1976)은 non-Bayesian 기법에서 다음과 같은 정리를 만족하는 최량선형불편추정량(BLUE)을 표본 이론에 적용하였다.

[정리]

임의의 확률변수 Y, X가 결합분포가 유한인 2차 적률을 갖는다고 하자.

또한 $E(Y)=\mu_Y$, $Var(Y)=\sigma_Y^2<\infty$ 이며, $E(X)=\mu_X$, $Var(X)=\sigma_X^2$, $Cov(X,Y)=\sigma_{XY}$ 라고 하자.

만일 X와 무관한 상수 a, b에 대해 다음이 성립하면

$$E(Y \mid X) = aX + b$$

식 (2-1)이 최소가 되도록 하는 상수 a^*, b^*를 구하면 다음과 같다.

$$z(a, b) = E\left[\,(E(Y \mid X) - a - b^{T} X)^2\,\right] \tag{2-1}$$

$$a^* = \mu_Y - b^{*T} \mu_X\,,\quad b^* = \sigma_{XY} / \sigma_X^2$$

이 값을 식 (2-1)에 대입하면 다음과 같다.

$$z(a^*, b^*) = \sigma_Y^2 - \sigma^2_{XY} / \sigma_X^2$$

위의 정리로부터 다음의 추정량은 Y의 모든 선형추정량 중에서 최소기대제곱오차손실 (minimal expected squared error loss)을 가진다.

$$\hat{Y}(X) = a^* + b^{*T} X \tag{2.2}$$

이 때, 최량추정량은 사후평균이 되며, 사후평균을 도출하기 위해서는 실제로 완전한 사후분포가 필요하다. 그러나 베이즈 선형추정량은 1차 적률과 2차 적률을 나타내는 μ_Y μ_X, σ_X^2, σ_{XY} 만이 필요하게 된다.

식 (2.2)의 추정치는 다음과 같다.

$$\hat{Y}(x) = \mu_Y + b^{*T}(x - \mu_X) \tag{2.3}$$

$Y(X)$의 정확도를 나타내는 척도로써 다음과 같이 기대제곱오차를 이용할 수 있으며 이는 점근적으로 다음과 같은 사후분산이 된다.

$$V\left[\hat{Y}(x)\right] = \sigma_Y^2 - \sigma^2_{XY} / \sigma_X^2 \tag{2.4}$$

2.2 2단계 확률화응답모형에 대한 베이즈 선형추정량

이 장에서는 베이즈 선형추정량을 Mangat와 Singh(1990)의 2단계 관련질문모형과 김종호외 2인(1992)이 제안한 2단계 무관질문모형과 2단계 강요질문모형에 적용하여 보고자 한다.

2.2.1 2단계 관련질문모형

Mangat와 Singh(1990)은 확률장치를 2개 사용하는 다음과 같은 2단계 확률화응답모형을 제시하였다.

R_1	설문 1 : 당신은 민감한 그룹 A에 속합니까? 설문 2 : 확률장치 R_2로 가시오.

R_2	설문 1 : 당신은 민감한 그룹 A에 속합니까? 설문 2 : 당신은 민감한 그룹 A에 속하지 않습니까?

여기서 응답자들은 1단계에서 설문 1과 설문 2가 선택될 확률이 각각 T와 $1-T$인 확률장치 R_1을 사용하며, 2단계에서 설문 1과 설문 2가 선택될 확률이 각각 p와 $1-p$인 확률장치 R_2를 이용하게 된다.

크기가 N인 유한모집단에서 만일 i번째 응답자가 민감한 그룹 A에 속하면 $A_i=1$이라 하고 그렇지 않으면 $A_i=0$라 하자. 이 때, 민감한 그룹 A에 속하는 모비율 Y는 다음과 같다.

$$Y = N^{-1}\sum_{i=1}^{N} A_i \tag{2.5}$$

그리고 단순임의복원추출된 n명의 응답자 중 만일 i번째 응답자가 "예"라고 응답할 경우 $X_i=1$이라 하고, "아니오"라고 응답할 경우에는 $X_i=0$라 하면, 다음을 구할 수 있다. 베이즈 선형추정량을 구하기 위해서는 A_i들에 대한 사전정보를 필요로 하게 되는데, 이때 가정되는 것이 교환가능성(exchangeability)이다. 이러한 개념은 모든 A_i들이 동일한 사전평균과 분산을 가지며, 모든 쌍 (A_i, A_j)이 동일한 사전공분산을 가진다는 것을 의미한다. A_i는 이지변수(binary variable)이므로 그 기대값, 분산 그리고 공분산은 다음과 같다.

$$E(A_i) = m \quad i=1, 2, \cdots, N$$

$$V(A_i) = m(1-m)$$

$$Cov(A_i, A_j) = rm(1-m), i, j=1, 2, \cdots, N; i \neq j$$

여기서, $r \in [-(N-1)^{-1}, 1]$

위의 정의로부터 다음을 구할 수 있다.

$$E(Y) = E(N^{-1}\Sigma A_i) = m$$

$$V(Y) = V(N^{-1}\Sigma A_i)$$

$$= \frac{1}{N^2}\left[V(\Sigma A_i) + N(N-1)\Sigma Cov(A_i, A_j)\right]$$

$$= m(1-m)\{1+(N-1)r\}/N$$

$$E(X_i) = Tm + (1-T)\{(1-p) + (2p-1)m\}$$

$$V(X_i) = \{T + (1-T)(2p-1)\}^2 m(1-m) + (1-p)(1-T)\{T+p(1-T)\}$$

또한, i번째 응답과 j번째 응답을 나타내는 변수를 각각 다음과 같다고 할 때

$$X_i = TA_i + (1-T)[(1-p) + (2p-1)A_i]$$

$$X_j = TA_j + (1-T)[(1-p) + (2p-1)A_j]$$

아래와 같이 공분산을 구할 수 있다.

$$Cov(X_i, X_j) = \{T+(1-T)(2p-1)\}^2 rm(1-m)$$

$$Cov(Y, X_i) = \{T+(1-T)(2p-1)\}m(1-m)\{1+(N-1)r\}/N$$

이 값들은 모두 사전 기대값이 되기 때문에 X_i들이 관찰되었다면, 베이즈 선형추정량은 사후 추정치 안에 있는 2가지 정보인 μ_Y와 μ_X를 선형결합하여 구할 수 있다.

모집단으로부터 크기 n인 단순임의표본을 추출하여 얻은 표본비율은 다음과 같다.

$$\overline{X} = n^{-1}\sum_{j=1}^{n} X_j \tag{2.6}$$

이로부터 $\overline{X}$의 기대값, 분산 그리고 공분산은 다음과 같이 구할 수 있다.

$$E(\overline{X}) = Tm + (1-T)\{(1-p) + (2p-1)m\}$$

$$\begin{aligned} V(\overline{X}) = n^{-1}[&\{T+(1-T)(2p-1)\}^2 m(1-m) \\ &+ (1-T)(1-p)\{T+p(1-T)\} \\ &+ \{T+(1-T)(2p-1)\}^2 m(1-m)r(n-1)\}] = w_1 \end{aligned} \tag{2.7}$$

$$Cov(Y, \overline{X}_i) = \{T+(1-T)(2p-1)\}m(1-m)\{1+r(N-1)\}/N = d_1 \tag{2.8}$$

따라서, 2단계 관련질문모형에서 구하고자 하는 베이즈 선형추정치는 다음과 같다.

$$\hat{Y}_1(\bar{x}) = m + (d_1/w_1)[\bar{x} - Tm - (1-T)\{(1-p) + (2p-1)m\}] \tag{2.9}$$

식 (2.9)를 간단히 표현하면 다음과 같다.

$$\hat{Y}_1(\bar{x}) = \alpha_1 y_1^* + (1-\alpha_1)m \tag{2.10}$$

여기서 $y_1^* = \{\bar{x} - (1-T)(1-p)\}/\{(2p-1) + 2T(1-p)\}$

$\alpha_1 = \{(2p-1) + 2T(1-p)\}(d_1/w_1)$

y_1^*는 non-Bayesian 불편추정량이다.

그리고, 2단계 관련질문모형에서 베이즈 선형추정량의 분산은 다음과 같다.

$$V[\hat{Y}_1(\bar{X})] = (1-\alpha_1)V(Y) \tag{2.11}$$

$$= (1-\alpha_1)m(1-m)\{1 + (N-1)r\}/N$$

2.2.2 2단계 무관질문모형

김종호외 2인(1992)은 2.2.1절에서 설명한 Mangat와 Singh(1990) 모형에서 1단계 확률장치 R_1은 그대로 사용하고 2단계 확률장치 R_2에서 다음과 같은 무관질문모형을 이용하는 2단계 무관질문모형을 제안하였다.

R_2	설문 1 : 당신은 민감한 그룹 A에 속합니까 ? 설문 2 : 당신은 민감한 그룹 B에 속합니까 ?

A_i와 $B_i (i=1, 2, \cdots, N)$를 i번째 응답자에 대한 2가지의 속성을 가지는 값이라고 하자. 실제로 A_i는 민감한 속성을 가지며 B_i는 민감한 속성과 관계없는 무관속성을 갖는다. 또한, A_i와 B_i들은 각각 교환가능(exchangeable)하며, 이들 두 속성간에는 서로 독립이라고 가정하면 $Cov(A_i, B_i)=0$이 된다.

이 절에서 사용되는 기호를 정의하면 다음과 같다.

$$E(A_i) = m_A \quad , \quad E(B_i) = m_B$$

$$V(A_i) = v_A \quad , \quad V(B_i) = v_B$$

$$Cov(A_i, A_j) = c_A \quad , \quad Cov(B_i, B_j) = c_B$$

여기서 그룹 B는 민감한 그룹과 전혀 관계가 없는 그룹을 의미한다. 또한 2.1절의 제 가정을 이용하면, 모집단으로부터 단순임의복원으로 추출된 n명의 응답자들의 응답에 대해 다음을 구할 수 있다.

$$E(X_i) = Tm_A + (1-T)\{pm_A + (1-p)m_B\} = E(\overline{X})$$

$$\begin{aligned} V(X_i) &= \{T+(1-T)p\}v_A + (1-T)(1-p)v_B \\ &\quad + \{T+(1-T)p\}(1-T)(1-p)(m_A - m_B)^2 \end{aligned}$$

$$Cov(X_i, X_j) = \{T+(1-T)p\}^2 c_A + (1-T)^2(1-p)^2 c_B$$

$$\begin{aligned} V(\overline{X}) &= n^{-1}[\{T+(1-T)p\}v_A + (1-T)(1-p)v_B \\ &\quad + \{T+(1-T)p\}(1-T)(1-p)(m_A - m_B)^2 \\ &\quad + (n-1)\{\{T+(1-T)p\}^2 c_A + (1-T)^2(1-p)^2 c_B\}] \\ &= w_2 \end{aligned} \tag{2.12}$$

또한, Y에 대한 1차 적률과 2차 적률은 다음과 같다.

$$E(Y) = m_A$$

$$V(Y) = \{v_A + (N-1)c_A\}/N$$

$$Cov(Y, \overline{X}) = \{T+(1-T)p\}\{v_A + (N-1)c_A\}/N = d_2 \tag{2.13}$$

이로부터 2단계 무관질문모형에서 구하고자 하는 베이즈 선형추정량은 다음과 같다.

$$\begin{aligned} \hat{Y}_2(\bar{x}) &= m_A + (d_2/w_2)\{\bar{x} - Tm_A - (1-T)pm_A - (1-T)(1-p)m_B\} \\ &= \alpha_2 y_2^* + (1-\alpha_2)m_A \end{aligned} \tag{2.14}$$

여기서 $y_2^* = \{\bar{x} - (1-T)(1-p)m_B\}/\{p + T(1-p)\}$

$\alpha_2 = \{p + T(1-p)\}(d_2/w_2)$

y_2^* 는 non-Bayesian 불편추정량이다.

그리고, 2단계 무관질문모형에서 베이즈 선형추정량의 분산은 다음과 같다.

$$V[\acute{Y}_2(\overline{X})] = (1-\alpha_2)V(Y)$$
$$= (1-\alpha_2)\{v_A + (N-1)c_A\}/N \tag{2.15}$$

2.2.3 2단계 강요질문모형

김종호외 2인(1992)은 2.2.1절에서 설명한 Mangat와 Singh(1990) 모형에서 2단계 확률장치 R_1은 그대로 사용하고 2단계 확률장치 R_2에서 다음과 같은 강요질문모형을 이용하는 2단계 강요질문모형을 제안하였다.

R_2	설문 1 : 당신은 민감한 그룹 A에 속합니까 ? 설문 2 "예" 라고 응답한다.

또한 2.2.1절의 제가정을 이용하면, 모집단으로부터 단순임의복원으로 추출된 n명의 응답자들의 응답에 대해 다음을 구할 수 있다.

$$E(X_i) = Tm + (1-T)\{pm + (1-p)\} = E(\overline{X})$$

$$V(X_i) = \{T+(1-T)p\}^2 m(1-m) + (1-T)(1-p)\{T+p(1-T)\}$$

$$Cov(X_i, X_j) = \{T+(1-T)p\}^2 rm(1-m)$$

$$Cov(Y, X_i) = [T+(1-T)p]m(1-m)\{1+r(N-1)\}/N$$

$$\begin{aligned} V(\overline{X}) = n^{-1}[&\{T+(1-T)p\}^2 m(1-m) \\ &+(1-T)(1-p)\{T+p(1-T)\} \\ &+\{T+(1-T)p\}^2 m(1-m)(n-1)r] = w_3 \end{aligned} \tag{2.16}$$

또한, 민감한 속성을 가진 모비율 Y에 대한 1차 적률과 2차 적률은 다음과 같다.

$$E(Y) = m$$

$$V(Y) = m(1-m)\{1+r(N-1)\}/N$$

$$Cov(Y, \overline{X}) = \{T+(1-T)p\}m(1-m)\{1+r(N-1)\}/N = d_3 \tag{2.17}$$

이로부터 2단계 강요질문모형에서 구하고자 하는 베이즈 선형추정치는 다음과 같다.

$$\hat{Y}_3(\bar{x}) = m+(d_3/w_3)\{\bar{x}-Tm-(1-T)pm-(1-T)(1-p)\}$$
$$= \alpha_3 y_3^* + (1-\alpha_3)m \qquad (2.18)$$

여기서 $y_3^* = \{\bar{x}-(1-T)(1-p)\}/\{p+T(1-p)\}$

$\alpha_3 = \{p+T(1-p)\}(d_3/w_3)$

y_3^* 는 non-Bayesian 불편추정량이다.

그리고, 2단계 강요질문모형에서 베이즈 선형추정량의 분산은 다음과 같다.

$$V[\hat{Y}_3(\bar{X})] = (1-\alpha_3)V(Y)$$
$$= (1-\alpha_3)m(1-m)\{1+r(N-1)\}/N \qquad (2.19)$$

3. 2단계 확률화응답모형에 대한 MLE와 BLE의 효율성 비교

3.1 2단계 관련질문모형에 대한 MSE의 비교

이 절에서는 O'Hagan(1987)이 제안한 베이즈 선형추정량을 2단계 확률화응답모형에 적용하여 2단계 확률화응답모형의 베이즈 선형추정량(BLE)에 대한 MSE를 구해 보고자 한다. 민감한 속성을 가진 모집단의 비율을 $\pi_A = 0.1 (= m)$로 안다고 가정하여 N이 무한히 크다면 식(2.8)로부터 다음과 같은 d_1를 얻을 수 있다.

$$d_1 = \{T+(1-T)(2p-1)\}rm(1-m) \qquad (3.1$$

또한 베이즈 선형추정량을 구하기 위해서는 사전평균 m과 속성들간의 상관을 나타내는 값인 r의 규정이 필요하다. 이를 위해 O'Hagan(1987)은 사전분포를 베타분포로 가정하여, 1차 적률과 2차 적률로 부터 얻어진 다음과 같은 m과 r을 이용하였다.

$$m = \alpha/(\alpha+\beta), \quad r = 1/(\alpha+\beta+1)$$

모의실험은 균등분포에서 $n = 100$개의 난수를 발생시켜 T와 p, 사전평균의 변화에 따라 각각의 경우에 대한 MSE를 구하여 이를 비교하여 보았다.

3.2 2단계 무관질문모형에 대한 MSE의 비교

민감한 그룹 A에 속한 비율을 $m_A = 0.1$로 알고 있으며, 또한 이와 무관한 그룹 B에 속할 비율을 $m_B = 0.15$로 가정하였는데, 이러한 이유는 무관질문모형에서 표본평균 $\bar{X}$의 분산인 $V(\bar{X})$의 항에서 m_A과 m_B의 차를 작게 함으로써 표본으로 부터의 오차를 줄이기 위

〈 표 3.1 〉 2단계 관련질문모형에 대한 BLE와 MLE의 MSE

Estimator	prior mean	$p=0.6$ $T=0.1$	$p=0.6$ $T=0.3$
BLE	0.05	0.00187	0.00104
	0.10	0.00019	0.00103
	0.15	0.00420	0.00739
	0.20	0.01332	0.01866
	0.25	0.02715	0.03399
	0.30	0.04541	0.05288
	0.35	0.06793	0.07506
	0.40	0.09459	0.10043
	0.45	0.12542	0.12897
	0.50	0.16039	0.16081
MLE		0.20484	0.15578

함이다. 물론 m_A과 m_B의 차를 작게 하면 응답자의 신분보호 측면에서 응답을 꺼리게 되어 무응답오차가 발생될 수 있다.

3.1절의 경우와 마찬가지로 모의실험은 균등분포에서 $n=100$개의 난수를 발생시켜 T와 p, 민감한 그룹에 속할 비율인 사전평균 m_A의 변화에 따라 각각의 경우에 대한 MSE를 구하여 이를 비교하여 보았다.

〈 표 3.2 〉 2단계 무관질문모형에 대한 BLE와 MLE의 MSE

Estimator	prior mean	$p=0.6$ $T=0.1$	$p=0.6$ $T=0.3$
BLE	0.05	0.00139	0.00101
	0.10	0.00012	0.00100
	0.15	0.00276	0.00434
	0.20	0.01145	0.01686
	0.25	0.02682	0.03123
	0.30	0.04311	0.05112
	0.35	0.06382	0.06616
	0.40	0.09152	0.10033
	0.45	0.12654	0.12038
	0.50	0.15641	0.15644
MLE		0.12553	0.10574

이지 모집단에 대한 2단계 확률화응답 모형의 베이즈 선형추정량(BLE)과 최우추정량(MLE)의 MSE를 비교하였다. BLE와 MLE를 비교한 결과 다음과 같은 결론을 도출할 수 있었다.

1. 2단계 관련질문모형 하에서 MLE는 사전정보와는 관계없이 p가 0.6으로 고정된 조건에서 T의 변화에 따라 감소하는 반면, BLE는 사전정보의 변화에 민감한 것으로 나타났다. 만일 사전정보가 효율적으로 모집단의 정보를 대표할 수 있다면 BLE는 MLE보다 더욱 효과적인 것이다.
2. 2단계 관련질문모형 하에서 BLE의 MSE는 사전평균 m의 범위가 0.1을 기준으로 변화할 때 MSE는 현저하게 작아짐을 알 수 있다. 이는 정확한 사전정보 하에서는 당연히 BLE가 MLE 보다 좋음을 나타낸다.
3. 2단계 무관질문의 경우 BLE의 MSE는 사전평균 m_A 의 범위가 0.1을 기준으로 변화할 때 MSE는 현저하게 작아짐을 볼 수 있었다.
4. T가 증가함에 따라 2단계 관련질문모형과 2단계 무관질문모형을 비교해 보면 BLE의 MSE는 관련질문모형의 경우가 더욱 큰 것으로 나타났으며, 관련질문모형에 대한 MLE의 MSE는 무관질문모형의 경우 보다 큰 것으로 나타나 non-Bayesian의 경우와 일치함을 볼 수 있다.

결과적으로 2단계 확률화응답모형에 대한 베이즈 선형추정량은 사전평균 m의 변화에 민감함을 알 수 있으며, 반면 최우추정량의 경우에는 설문 선택확률에 더욱 민감한 것으로 나타났다.

5. 결론

일반적으로 베이지안 분석방법은 사전확률분포에 대한 완전한 정보를 필요로 하지만 베이즈 선형추정량은 단지 1차 적률과 2차 적률만을 필요로 하기 때문에 추정량을 구하는 데 있어서 매우 간편하며 사전분포를 알 필요가 없다는 장점을 가지고 있다.

Winkler와 Franklin(1979)은 Warner(1965)의 관련질문보형과 무관질문모형에 대한 베이즈 선형추정량을 구하였다.

본 논문에서는 Mangat와 Singh(1990)의 2단계 관련질문모형과 김종호외 2인(1992)이 제안한 2단계 무관질문모형모형과 2단계 강요질문모형에 대한 베이즈 선형추정량을 구했으며, 사전평균 m의 변화에 따른 추정량의 MSE와 설문 선택 확률 p와 T의 변화에 따른 추정량의 MSE를 비교 해 보았다.

그 결과 민감한 속성에 관한 조사에 있어서 2단계 확률화응답모형에 대한 베이즈 선형추정량의 MSE가 기존의 Mangat와 Singh(1990)의 MLE에 대한 MSE가 작아짐을 알 수 있었다. 또한 베이즈 선형추정량의 사전정보의 변화에 매우 민감한 것으로 나타나 사전정보가 모집단의 정보와 차이가 없을 경우 베이즈 선형추정량을 이용하는 것이 효과적임을 알 수 있었다. 또한 무관질문모형의 경우에 대한 MSE의 비교에 있어서도 위의 결과와 마찬가지로 사전정보가 정확할 때 2단계 확률화응답모형에 대한 베이즈 선형추정량의 MSE가 작음을 알 수 있었다.

참고문헌

[1] 류제복, 홍기학, 이기성 (1993), 「확률화응답모형」, 자유아카데미.

[2] 김종호, 류제복, 이기성 (1992), "A New Two-Stage Randomized Response Model," 응용통계연구, 제5권 제2호, pp. 157-167.

[3] 박진우 (1989), "A Study on Bayes Linear Estimation for Multi-Proportions Randomized Response Model," 박사학위논문, 서울대학교.

[4] 이영진 (1992), 「확률응답모형」, 박사학위논문, 인하대학교.

[5] Drane, W. (1975). "Randomized Response to More One Question," Proceedings of Social Statistics Section, *American Statistical Association*, pp. 395-397.

[6] Eriksson, S. A. (1973), "A New Model for Randomized Response," *International Statistical Review*, Vol. 41. pp. 101-113.

[7] Erisson, W. A. (1988), "Bayesian Inference in Finite Populations," *Handbook of Statistics*, Vol. 6, pp. 213-246.

[8] Folsom, Greenberg, Horvitz and Abernathy (1973), "The two Alternative Questions Randomized Response Model for Human Surveys," *Journal of American Statistical Association*, Vol. 68, pp. 525-530.

[9] Godambe, V. P. and Thompson, M. E. (1973), "Estimation in Sampling Theory with Exchangeable Prior Distributions," *The Annals of Statistics*, pp 1212-1221.

[10] Goldstein, M. (1975), "A Note on Some Bayesian Non-parametric Estimates," *The Annals of Statistics*, Vol. 3, pp. 736-740.

[11] Greenberg, Abul-Ela, Simmons and Horvitz (1969), "The Unrelated Question Randomized Response Model : Theoretical Framework," *Journal of American Statistical Association*, Vol. 64, pp. 520-539.

[12] Hartigan, J. A. (1969), "Linear Bayesian Methods," *Journal of Royal Statistical Society*, Ser. B, Vol. 31, pp. 446-454.

[13] Horvitz, D. G., Shah, B. V. and Simmons, W. R. (1967), "The Unrelated Randomized Response Model," *Proceedings of the Social Statistics Section American Statistical Association*, pp. 65-72.

[14] LaMotte, L. R (1978), "Bayes Linear Estimators," *Technometrics*, Vol. 20, pp. 281-290.

[15] Leyseiffer, F. W. and Warner, S. L. (1976), "Respondent Jeopardy and Optimal Designs in Randomized Response Models," *Journal of American Statistical Association*, Vol. 71, pp. 649-656.

[16] Lindley, D. V. and Novick, M. R. (1981), "The Role of Exchangeavility in Inference," *The Annals of Statistics*, Vol. 9, pp. 45-58.

[17] Liu, P. T. and Chow, L. P. (1976), "The Efficiency of the Multiple Trial

Randomized Response Technique," *Biometrics*, Vol. 32, pp. 607.

[18] Mangat, N. S. and Singh, R. (1990), "An Alternative Randomized Response Procedure," *Journal of the American Statistical Association*, Vol. 29, pp. 127−131.

[19] O'Hogan, A. (1987), "Bayes Linear Estimatiors for Randomized Response Models," *Journal of American Statistical Association*, Vol. 82, pp. 580−585.

[20] Pitz, G. F. (1980), "Bayesian Analysis of Randomized Response Models," *Psychological Bulletin*, Vol. 87, pp. 209−212.

[21] Royall, R. M. (1976), "The Linear Least-Squares Prediction Approach to Two-Stage Sampling," *Journal of American Statistical Association*, Vol. 71, pp. 657−664.

[22] Smouse, E. P. (1984), "A Note on Bayesian Least Squares Inference for Finite Population Models," *Journal of American Statistical Association*, Vol. 79, pp. 390−392.

[23] Stone, M. (1963), "Robustness of Non-ideal Decision Procedures," *Journal of American Statistical Association*, Vol. 38, pp. 480−486.

[24] Warner, S. L. (1965), "Randomized Response : A Survey Technique for Eliminating Evasive Answer Bias," *Journal of American Statistical Association*, Vol. 60, pp. 63−69.

[25] Whittle, P. (1958), "On the Smoothing of Probability Density Functions," *Journal of Royal Statistical Society*, Ser. B, Vol. 20, pp. 334−343.

[26] Winkler, R. L. and Franklin, L. A. (1979), "Warner's Randomized Response Model : A Bayesian Approach," *Journal of American Statistical Association*, Vol. 74, pp. 207−214.