Journal of Statistical
Theory & Methods
1994, Vol. 5, pp33~48

올바른 선택의 확률에 대한 추정

손 중 권, 윤 주 영, 김 현 주1)

요약. 무차별 영역에서 올바른 선택을 할 확률을 베이지안 관점으로 추정하였으며 특히 모수의 사전분포를 주는 대신 P(CS) 자체의 사전분포를 정의하여 여러가지 추정량 제안하였다. 또한 제안된 추정량이 사전분포에 어떤 영향을 받는 지를 모의실험을 통해 알아보았다.

1. 여 는 말

$k(\geq 2)$개의 모집단중에서 '최고(Best)'인 모집단을 선택하는 문제에 대해 Bechhofer(1954)는 무차별 영역(Indifference-Zone)선택론을 소개하였다. 무차별 영역 선택절차론에서 올바른 선택(Correct Selection(CS))이란 참 최고 모집단(True best population)을 선택하는 것으로 올바른 선택과 관련된 몇 개의 상수가 실험자에 의해 미리 주어져야 한다. 그러나 선택은 매번 얻어진 표본에 의하여 이루어 지므로 선택이 이루어진 후에도 올바른 선택이 이루어졌을 확률(probability of Correct Selection, P(CS))은 알 수가 없다. 그러므로 사용된 통계적 방법에 대한 검증의 의미로써 P(CS)에 대한 추론이 필요하다.

일반적으로 모수에 대한 정보가 거의 없어 P(CS)의 참값을 추정하는 문제에 관심을 가져왔다. Olkin, Sobel과 Tong(1976)은 표본정보를 이용하여 P(CS)를 추정하는 문제를 연구했다. Faltin(1980)은 동일한 기지의 분산을 가지는 두 정규모집단인 경우에 대해 P(CS)의 추정량을 제안했다. Faltin과 McCulloch(1983)는 분산을 알고 있는 두 정규모집단의 성질을 연구하여 추정량의 단점을 밝혔다. 아울러 McCulloch와 Dechter(1985)는 Olkin, Sobel과 Tong(1976)추정량을 개선하기 위해 경험적 베이스 추정량을 연구하였다. Bofinger(1990)는 계층적 베이스 접근법(Hierarchical Bayes Approach)과 모수적 경험적 베이스 접근법(Parametric Empirical Bayes Approach)을 절충한 추정량을 제안하였다. 그러나, 손과 김(1987)은 모수에 대해서는 모르는 것이

---

1) 경북대학교 통계학과

절충한 추정량을 제안하였다. 그러나, 손과 김(1987)은 모수에 대해서는 모르는 것이 일반적이므로 모수에 대한 사전분포(Prior distribution)보다는 P(CS)에 사전분포를 주는 것이 타당하다고 보고 P(CS)의 사전 분포가 부분균일분포(Locally uniform distribution)를 따르는 경우 P(CS)의 추정량을 베이지안적 입장에서 고찰하였다. 또한 손과 강(1992)은 붓스트랩 방법을 이용한 P(CS)의 점추정과 신뢰구간 추정문제를 다루었고, 두 정규모집단인 경우에 대하여 베이스 추정량을 제안하고 붓스트랩 추정량과 비교하였다.

본 논문에서는 두 정규모집단의 경우에 손과 김(1987)이 제안한 추정량의 성질을 살펴보고, 이 추정량의 단점을 개선하기 위하여 사전분포를 일반화시켜 가면서 베이스 추정량을 비교해 보고자 한다. 2장에서는 무차별 영역 선택론에서 k개의 정규모집단을 순서화하고 선택하는 문제를 소개하고 올바른 선택의 확률과 관련된 알려진 결과들을 서술하고자 한다. 3장에서는 몇가지 사전분포에 대해서 P(CS)의 추정량을 구하여 성질을 살펴보고, 사전분포가 추정량에 어떤 영향을 미치는지 알아보고자 한다. 4장에서는 몬테칼로 모의실험을 통하여 3장에서 살펴본 여러 추정량들을 편의(Bias)와 평균제곱오차(MSE)의 측면에서 비교해 보고자 한다.

2. 기존의 추정량에 대한 소개

$\pi_i(1 \le i \le k)$를 평균이 각각 $\theta_i(1 \le i \le k)$이고, 분산이 σ^2인 k개의 정규모집단이라고 하자. 분산 σ^2은 공통이며 알고 있다고 가정한다. 모평균 $\theta_1, \theta_2, \cdots, \theta_k$을 크기대로 나열하면 다음과 같이 표현할 수 있다.

$$\theta_{[1]} \le \theta_{[2]} \le \cdots \le \theta_{[k]}$$

여기서 평균이 제일 큰 모집단 즉, 평균이 $\theta_{[k]}$인 모집단을 '최고'라고 정의한다. 모집단 $\pi_i\,(i = 1, 2, \cdots, k)$로 부터 n개의 표본을 추출하여 얻어진 표본평균을 $\overline{X_i}$라 하자. 최고의 모집단을 선택하기 위해서는 $\theta_{[k]} - \theta_{[k-1]} \ge \delta^*$ 일때 $\text{Inf}\ P(CS) \ge P^*$ 라는 요구조건이 있다. 이를 P^*-조건이라 한다. 여기서 $P^*\left(\frac{1}{k} < P^* < 1\right)$, $\delta^*(>0)$ 그리고 $n \equiv n(k, P^*, \delta^*)$ 은 주어지는 상수이다. $\overline{X_1} \le \overline{X_2}, \cdots, \overline{X_k}$ 를 크기대로 나열하면 $\overline{X_{[1]}} \le \overline{X_{[2]}} \le \cdots \le \overline{X_{[k]}}$ 가 되며 $\overline{X_{(i)}}$ 는 $\theta_{[i]}\ (1 \le i \le k)$ 로부터 얻어진 표본평균이라 하자. $\overline{X_i}$ 와 $\theta_{[l]}\ (1 \le i \le k,\ 1 \le l \le k)$ 사이의 올바른 쌍에 대한 사전정보는 없다고 가정한다. 선택절차론 $\overline{X_{[k]}}$를 가지는 모집단을 선택하고, 이것을 최고의 모집단이라 하며, 이때 올바른 선택의 확률은 다음과

같이 주어진다.

$$P(CS) = \int_{-\infty}^{\infty} \prod_{i=1}^{k-1} \Phi(y + \frac{\sqrt{n}\delta_i}{\sigma}) d\Phi(y) \tag{2.1}$$

여기서 $\Phi(\cdot)$는 표준 정규분포의 누적분포함수이며 $\delta_i = \theta_{[k]} - \theta_{[i]}$이다.

Olkin, Sobel과 Tong(1976)은 P(CS)의 추정량을 다음과 같이 제안하였다.

$$\widehat{P(CS)} = \int_{-\infty}^{\infty} \prod_{i=1}^{k-1} \Phi(y + \frac{\sqrt{n}\,\widehat{\delta_i}}{\sigma}) d\Phi(y) \tag{2.2}$$

여기서 $\widehat{\delta_i} = \overline{X_{[k]}} - \overline{X_{[i]}}$ 이며 δ_i의 추정량이다. 또한 Faltin(1980)은 동일한 분산을 가진 두 정규모집단인 경우에 P(CS)의 분위수 불편 추정량(Quantile unbiased estimator)을 제안했다. Olkin, Sobel 그리고 Tong(1976)추정량의 소표본 성질은 Faltin과 McCulloch(1983)에 의하여 연구되었는데 그것은 평균의 차이가 작을 때는 과대추정을 하고 평균차가 커지면 과소추정을 하는 것으로 밝혀졌다.

Olkin, Sobel과 Tong(1976)의 추정량을 개선하기 위하여 McCulloch와 Detcher(1985)는 P(CS)를 추정하는데 있어 경험적 베이즈 추정량을 제안하였는데 다음과 같다.

$$\widetilde{P(CS)} = \int_{-\infty}^{\infty} \prod_{i=1}^{k-1} \Phi(y + \frac{\sqrt{n}\,\widetilde{\delta_i}}{\sigma}) d\Phi(y) \tag{2.3}$$

여기서

$$\widetilde{\delta_i} = \begin{cases} \left(1 - \dfrac{(k-3)\sigma^2/n}{\sum_{i=1}^{k}(\overline{X_i} - \overline{X})^2}\right)^+ \widehat{\delta_i} & \text{if } k>3 \\ \left(1 - \dfrac{\sigma^2/n}{\sum_{i=1}^{k}(\overline{X_i} - \overline{X})^2}\right)^+ \widehat{\delta_i} & \text{if } k=2,3, \end{cases}$$

그리고 $(t)^+$는 $\max(t,0)$로 정의되며 $\overline{X} = \sum_{i=1}^{k} \overline{X_i} / k$ 이다. 그들은 평균차가 적을 때 식(2.2)보다 식(2.3)이 평균제곱오차와 편의측면에서 더 낫다는 것을 보였지만 식(2.3)은 평균차가 커질수록 좋지 못함이 밝혀졌다.

Bofinger(1990)는 θ_i에 대한 사전분포가 모든 $i(1 \le i \le k)$에 대하여 독립인 $N(\mu, V)$를 따를때 올바른 선택의 사후확률을 다음과 같음을 보였다.

$$P(CS|\eta) = \int_{-\infty}^{\infty} \prod_{i=1}^{k-1} \Phi(y + \frac{D_i\eta}{\sigma^2}) d\Phi(y) \tag{2.4}$$

여기서 η는 V의 함수로 $\eta^{-2} = V^{-1}+\sigma^{-2}$이며 $D^i = \overline{X_{[k]}} - \overline{X_{[k-i]}}$ $(i = 0, \cdots, k-1)$이다. 즉 V의 추정량으로 $V_B = (k-1)S^2/Z_B - \sigma^2$을 제안하였는데 여기서 Z_B는 $(k-1)S^2/\sigma^2$에서 절단되고 자유도가 (k-1)인 카이제곱분포의 중앙값(Median)이다.

두 정규모집단인 경우에 대해 P(CS)의 사전분포가 부분균일분포를 따르는 경우 손과 김(1987)이 제안한 추정량은

$$\widehat{P(CS)}^B = \int_{\frac{1}{2}}^{1} Z \cdot f(z|\overline{x_1} - \overline{x_2})\, dz \tag{2.5}$$

와 같다. 여기서 Z=P(CS)이고

$$f(z|\overline{x_1} - \overline{x_2}) = \frac{1}{\sqrt{\pi}\phi(\Phi^{-1}(z))} \times [\exp\{-\frac{n}{2}(\sqrt{\frac{2}{n}}\Phi^{-1}(Z) - \frac{\overline{x_1} - \overline{x_2}}{2})^2\} + \exp\{-\frac{n}{2}(\sqrt{\frac{2}{n}}\Phi^{-1}(Z) + \frac{\overline{x_1} - \overline{x_2}}{2})^2\}], \quad \frac{1}{2} \le Z \le 1$$

이며, $\phi(\cdot)$는 표준 정규분포의 확률밀도함수이다. 그리고 손과 강(1992)은 모의실험을 통하여 제안한 붓스트랩 추정량와 신뢰구간의 성질을 살펴보았는데, 붓스트랩 추정량은 평균제곱오차측면에서 좋고, 붓스트랩 신뢰구간은 붓스트랩 추정치에 대하여 대칭이지만 치우치는 성질을 가지고 있었다. 그러나 어떤 경우에는 유도된 구간이 P(CS)의 참값을 포함하지 않기도 하며, 이 경우에는 대부분 P(CS)를 과대추정 내지는 과소추정을 하는 경우였음을 보였다.

3. 여러 사전분포의 경우 올바른 선택의 확률에 대한 추정

두 개의 모집단의 각각 평균이 θ_1과 θ_2이고, 분산이 σ^2으로 동일하며 독립인 정규모집단이라고 가정한다. 그러면 식(2.1)은

$$\begin{aligned} P(CS) &= P[\overline{X_{(1)}} \le \overline{X_{(2)}}] \\ &= \Phi(\sqrt{\frac{n}{2\sigma^2}}\,|\theta_1-\theta_2|) \end{aligned}$$

와 같이 나타낼 수 있다.

분산을 알고 있다고 가정한다면 $\sigma^2 = 1$로 둘 수 있다. 그러면

$$P(CS) = \Phi(\sqrt{\frac{n}{2\sigma^2}}|\theta_1-\theta_2|) \equiv Z$$

가 된다.

이 장에서는 몇몇 사전 분포의 경우에 대하여, 표본평균의 차이 $\overline{x_1}-\overline{x_2}$가 주어진 상태에서의 사후분포식을 유도하며, 제곱오차손실(Square-error loss)하에서의 베이스 추정량인

$$E(z|\overline{x_1}-\overline{x_2}) = \int_{\frac{1}{2}}^{1} z \cdot f(z|\overline{x_1}-\overline{x_2})\, dz \tag{3.1}$$

을 P(CS)의 추정량으로 구한다.

3.1 P(CS)의 사전분포가 부분 균일분포를 따르는 경우

P(CS)에 대한 사전분포가 구간(0.5, 1)에서 균일한 값을 가지는 경우 즉,

$$P(CS) \sim \mathrm{U}(\frac{1}{2},1)$$

로 주어질때 표본평균의 차이가 주어진 상태에서의 사후분포는 $f(Z|\overline{x_1}-\overline{x_2})$는 다음과 같이 얻어진다.

$$f(z|\overline{x_1}-\overline{x_2}) = \frac{1}{\sqrt{\pi}\phi(\Phi^{-1}(z))} \times [\exp\{-\frac{n}{2}(\sqrt{\frac{2}{n}}\Phi^{-1}(Z)-\frac{\overline{x_1}-\overline{x_2}}{2})^2\} + \exp\{-\frac{n}{2}(\sqrt{\frac{2}{n}}\Phi^{-1}(z)+\frac{\overline{x_1}-\overline{x_2}}{2})^2\}],\quad \frac{1}{2} \le z \le 1$$

이 경우 표본의 크기나 표본평균의 차이가 커지면 사후분포가 1로 퇴화(Degenerate)하는 형태가 되며, $P(Z=1|\overline{x_1}-\overline{x_2}) \approx 1$이 됨을 알 수 있다.

3.2 P(CS)의 사전분포가 가중 절단 베타분포를 따르는 경우

P(CS)에 대한 사전 정보가 없는 경우에는 P(CS)의 사전분포로 부분균일 분포를 취할 수 있다. 그러나 P(CS)가 0.5와 1 사이의 값을 가지므로 절단된 베타분포를 사전분포로 취해야 한다. 이 경우 사전분포는 증가형태 혹은 감소 형태로 나타나는데 증가 형태의 경우에는 표본평균의 차이가 작을 때 과대추정을 하고 감소형태의 경우에는 표본평균의 차이가 클때 과소추정을 하게 된다. 그러므로 이 절에서는 감소형태의 사전분포와 증가형태의 사전분포에 적절한 가중치를 주어 이러한 문제점들을 최소화시키는 사전분포를 구하여 보고자 한다.

사전분포식은

$$\omega \cdot T_z(\alpha_1,\beta_1) + (1-\omega)\cdot T_Z(\alpha_2,\beta_2)$$

와 같이 주어진다. 여기서 ω는 가중치이며

$$T_{z(\alpha,\beta)} = \frac{\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} z^{\alpha-1}(1-z)^{\beta-1}}{\int_0^{\frac{1}{2}} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} z^{\alpha-1}(1-z)^{\beta-1} dz}$$

$$= \frac{\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} z^{\alpha-1}(1-z)^{\beta-1}}{1-I_{\frac{1}{2}}(\alpha,\beta)}$$

이다.

그러면 $\overline{x_1}-\overline{x_2}$ 가 주어진 상태에서 P(CS)의 사후밀도 함수는

$$\begin{aligned} f(z|\overline{x_1}-\overline{x_2}) = & \frac{1}{\sqrt{\pi}\phi(\Phi^{-1}(z))} \\ & \times \frac{C_1 z^{\alpha_1-1}(1-z)^{\beta_1-1}+C_2 z^{\alpha_2-1}(1-z)^{\beta_2-1}}{C_1E_1+C_2E_2} \\ & \times (\exp\{-\frac{n}{2}(\sqrt{\frac{2}{n}}\Phi^{-1}(Z)-\frac{\overline{X_1}-\overline{X_2}}{2})^2\} \\ & + \exp\{-\frac{n}{2}(\sqrt{\frac{2}{n}}\Phi^{-1}(Z)+\frac{\overline{X_1}-\overline{X_2}}{2})^2\}) \end{aligned}$$

$$\begin{aligned} = & \frac{1}{\sqrt{\Pi}\phi(\Phi^{-1}(z))} \\ & \times \frac{C_1 z^{\alpha_1-1}(1-z)^{\beta_1-1}+C_2 z^{\alpha_2-1}(1-z)^{\beta_2-1}}{C_1E_1+C_1E_2} \\ & \times (\exp\{-\Phi^{-1}(z)^2+Q\Phi^{-1}(z)-\frac{Q^2}{4}\} \\ & + \exp\{-\Phi^{-1}(z)^2-Q\Phi^{-1}(z)-\frac{Q^2}{4}\}), \quad \frac{1}{2} \le Z \le 1 \end{aligned}$$

와 같이 얻어지며, 여기서

$$Q = \sqrt{\frac{n}{2}}(\overline{x_1}-\overline{x_2}),$$

$$C_1 = \frac{\omega}{1-I_{\frac{1}{2}}(\alpha_1,\beta_1)} \frac{\Gamma(\alpha_1+\beta_1)}{\Gamma(\alpha_1)\Gamma(\beta_1)},$$

$$C_2 = \frac{\omega}{1-I_{\frac{1}{2}}(\alpha_2,\beta_2)} \frac{\Gamma(\alpha_2+\beta_2)}{\Gamma(\alpha_2)\Gamma(\beta_2)},$$

$$E_1 = E([\Phi(\sqrt{\frac{n}{2}}|\theta_1-\theta_2|)]^{\alpha_1-1}[1-\Phi(\sqrt{\frac{n}{2}}|\theta_1-\theta_2|)]^{\beta_1-1}),$$

$$E_2 = E([\Phi(\sqrt{\frac{n}{2}}|\theta_1-\theta_2|)]^{\alpha_2-1}[1-\Phi(\sqrt{\frac{n}{2}}|\theta_1-\theta_2|)]^{\beta_2-1}),$$

이다.

가중치 ω가 0.5인 경우 α_1, β_1, α_2, β_2의 값을 바꾸어 가면서 P(CS)의 추정치를 계산해 본 결과 대체적으로 표본평균의 차이가 크거나 작은 경우에는 α_1, β_2는 클수록 β_1, α_2는 작을수록 P(CS)의 참값과 근사한 값을 얻을 수 있었고 사전분포의 모수에 많은 영향을 받는다는것을 알 수 있었다.

3.3 P(CS)의 사전분포가 두점 이산형 분포인 경우

모집단이 두 개인 경우 올바르게 선택할 확률은 0 혹은 1 인 이산형의 결과를 얻게 된다. 그러나 P(CS)는 0.5와 1 사이의 값을 취하므로 이 절에서는 앞 절에서 언급한 바와같이 P(CS)의 사전분포로 0.5와 1에서만 값을 가지는 두점 이산형 분포인 경우에 대하여 알아보고자 한다.

두점 분포의 함수식은 다음과 같이 주어진다.

$$f(z) = p \cdot I_{(z=0.5)} + (1-p) \cdot I_{(z=1)}$$

p의 값을 0.1에서 0.9까지 변화시켜가면서 살펴 보았다. 표본평균의 차이가 주어진 상태에서 올바른 선택확률의 사후분포는

$$f(z|\overline{x_1}-\overline{x_2}) = \frac{(p \cdot I_{(z=0.5)}+(1-p) \cdot I_{(z=1)}) \times (K1+k2)}{2 \cdot p \cdot K3+(1-p)}, \quad z=0.5,\ 1$$

이다. 여기서

$$K1 = \exp\{-\frac{n}{4}(\sqrt{\frac{2}{n}}\Phi^{-1}(Z)-(\overline{x_1}-\overline{x_2}))^2\},$$

$$K2 = \exp\{-\frac{n}{4}(\sqrt{\frac{2}{n}}\Phi^{-1}(Z)+(\overline{x_1}-\overline{x_2}))^2\},$$

$$K3 = \exp\{-\frac{n}{4}(\overline{x_1}-\overline{x_2})^2\}$$

와 같다.

3.4 모평균의 차이가 정규분포를 따르는 경우

이 절에서는 모수에 대한 사전분포를 준 기존의 경우와 P(CS)에 사전분포를 준 경우가 어떤 연관성이 있는지를 알아보기 위하여 사전분포가

$$\theta_1-\theta_2 \sim N(\mu,\eta^2)$$

인 경우에 대하여 생각해 보기로 하자.

이 경우, 표본평균의 차이 $\overline{x_1}-\overline{x_2}$가 주어진 상태에서의 P(CS)의 사후분포는 다음과 같다.

$$\begin{aligned} f(z|\overline{x_1}-\overline{x_2}) &= \sqrt{\frac{2+n\eta^2}{2\pi\eta^2 n}} \cdot \frac{1}{\phi}(\Phi^{-1}(z)) \\ &\times[\exp\{-\frac{2+n\eta^2}{4\eta^2}(\sqrt{\frac{2}{n}}\Phi^{-1}(Z)-\frac{2\mu+n\mu^2(\overline{x_1}-\overline{x_2})}{2+n\eta^2})^2\} \\ &+\exp\{-\frac{2+n\eta^2}{4\eta^2}(\sqrt{\frac{2}{n}}\Phi^{-1}(Z)+\frac{2\mu+n\mu^2(\overline{x_1}-\overline{x_2})}{2+n\eta^2})^2\}], \quad \frac{1}{2}\le z\le 1. \end{aligned}$$

$\mu = 0$ 이고 $\eta^2 = 2/n$일때 사후분포 함수는 P(CS)가 부분 균일분포를 따르는 경우와 일치함을 알 수 있었다.

4. 몬테칼로 모의실험

이 장에서는 3장에서 살펴본 여러 추정량들을 모의실험을 통하여 편의와 평균제곱오차 측면에서 비교해 보고자 한다.

모평균의 차 $\theta_1-\theta_2$가 δ인 두 정규모집단에서 n개의 난수(Random number)를 발생시키기 위하여 IMSL의 DRNNOA, DSCAL, DADD, RNSET, RNGET 부프로그램을 사용하였다. 모의실험은 두 경우로 나누어 실행해 보았다. 우선 P(CS)의 사전분포가 가중 절단 베타분포를 따르는 경우 추정량이 사전분포의 모수에 얼마나 영향을 받는지를 알아보기 위하여 다음과 같이 세가지 경우를 비교해 보았다.

ω	α_1	β_1	α_2	β_2
0.5	1	1	1	1*
0.5	10	1	1	10
0.5	50	1	1	50

여기서 *는 P(CS)의 사전분포가 부분 균일분포를 따르는 경우와 일치한다. 이 모의실험에서는 모평균차이 δ가 0.1, 0.3인 경우는 표본의 크기 n을 4, 9, 16, 25인 각 경우에 대하여 1000번 반복을 하였으며, δ가 0.5, 0.1일때는 표본의 크기 n이 4, 9, 16, 25인 각 경우에 대하여 500번 반복을 하였다. 모의실험결과는 그림 1과 그림 2에 나타나 있다. 편의와 평균제곱오차의 두 측면에서 $\sqrt{n}\cdot\delta$가 아주 작거나 큰 경우에는 $\widehat{P(CS)}^{50}$이 가장 좋으며 $\widehat{P(CS)}^{10}$, $\widehat{P(CS)}^{1}$의 순서이다. 중간부분에선 반대로 작용함을 알 수 있었다.

또한, P(CS)의 사전분포가 두점 이산형 분포를 따르는 경우에 추정량이 사전분포의 모수 p에 얼마나 영향을 받는지 알아보기 위하여 확률 p를 0.1, 0.3, 0.5, 0.7, 0.9로 두고, 모평균의 차이 δ는 0.1, 0.3, 0.5, 1.0이고 표본의 크기 n은 4, 9, 16, 25인 경우에 대하여 500번의 반복을 하였다. 이 모의실험의 결과는 그림3과 그림 4에서와 같으며, 편의와 평균제곱오차의 측면에서 비교해 보면 $\sqrt{n}\cdot\delta$가 아주 작은 경우를 제외하고는 p의 값이 작을수록 더 좋은 추정량임을 알 수 있었다.

5. 닫 는 말

이제까지 두 정규모집단의 경우에 몇가지 사전분포에 대하여 P(CS)의 추정량을 구하여 보았다. P(CS)의 사전분포가 가중 절단 베타분포를 따르는 경우 가중치를 변화시키면 더 다양한 결과를 얻을 수 있으며, 가중치와 다른 모수들을 적절히 변화시키면 두점 이산형 사전 분포와 유사한 모양을 얻을 수도 있다. 그러나 이 경우 추정치를 구하려면 여러 모수에 대하여 계산하여야 하므로 복잡하고, 사후분포가 1로 퇴화하는 형태가 됨에 따라 계산 시간이 많이 소요되는 단점이 있으므로 두 모집단에 대하여 P(CS)를 추정하는 문제에 대해서는 P(CS) 의 사전분포를 두점 이산형 분포를 선택하는 것도 좋은 추정치를 얻을 수 있음을 알 수 있다.

또한 모의실험을 통하여, P(CS)의 사전분포의 모수는 P(CS)의 추정량에 많은 영향을 미친다는 것을 알 수 있었으며, 표본평균의 차이에 관계없이 편의와 평균제곱오차를 작게해 주는 추정량은 얻을 수 없었다. 그러나 표본평균의 차이가 커지면 편의와 평균제곱오차가 모두 0으로 수렴하게 되므로 P(CS)의 추정문제는 표본평균의 차

이가 작을 때 의미가 있다고 할 수 있다.

표1. P(CS)의 사전분포가 부분균일분포를 따르는 경우 P(CS)에 대한 추정치

n	$\overline{x_1} - \overline{x_2}$	P(CS)	$\widehat{P(CS)}$
4	0.1	0.5562	0.6966
	0.3	0.6643	0.7025
	0.5	0.7602	0.7140
	1.0	0.9213	0.7615
	1.5	0.9830	0.8229
	2.0	0.9976	0.8804
	3.0	0.9999	0.9585
6	0.1	0.5687	0.6907
	0.3	0.6983	0.7058
	0.5	0.8067	0.7226
	1.0	0.9583	0.7882
	1.5	0.9583	0.8625
	2.0	0.9997	0.9226
	3.0	0.9999	0.9830
10	0.1	0.5884	0.6977
	0.3	0.7488	0.7122
	0.5	0.8682	0.7389
	1.0	0.9873	0.8322
	1.5	0.9996	0.9161
	2.0	0.9999	0.9669
	3.0	1.0000	0.9969
20	0.1	0.6240	0.6996
	0.3	0.8286	0.7276
	0.5	0.9430	0.7753
	1.0	0.9992	0.9040
	1.5	0.9999	0.9736
	2.0	0.9999	0.9950
	3.0	1.0000	0.9999

표2. P(CS)의 사전분포가 가중 절단 베타분포를 따르는 경우 P(CS)에 대한 추정치

$\widehat{P(CS)}^1$는 $\alpha_1, \beta_1, \alpha_2, \beta_2 = 1$인 경우

$\widehat{P(CS)}^{10}$는 $\alpha_1, \beta_2 = 10$이고 $\alpha_2, \beta_1 = 1$인 경우

$\widehat{P(CS)}^{50}$는 $\alpha_1, \beta_2 = 50$이고 $\alpha_2, \beta_1 = 1$인 경우

Q	$P(CS)$	$\widehat{P(CS)}^1$	$\widehat{P(CS)}^{10}$	$\widehat{P(CS)}^{50}$
0.000	0.50000	0.69591	0.62815	0.55524
0.100	0.53983	0.69629	0.99491	0.59029
0.300	0.61791	0.69927	0.71368	0.65592
0.500	0.69146	0.70513	0.77148	0.75584
0.800	0.78814	0.71888	0.83727	0.87321
1.000	0.84134	0.73094	0.88092	0.93326
1.645	0.95000	0.78072	0.93002	0.97299
1.960	0.97500	0.80813	0.94570	0.97967
2.000	0.97725	0.81163	0.95023	0.98102
3.000	0.99865	0.89289	0.96969	0.98719

표3. P(CS)의 사전분포가 두점 이산형 분포를 따르는 경우 P(CS)에 대한 추정치 $\widehat{P(CS)}^{0.2}$ 는 p가 0.2인 경우, $\widehat{P(CS)}^{0.5}$는 p가 0.5인 경우, $\widehat{P(CS)}^{0.8}$ 는 p가 0.8인 경우

n	$\overline{x_1} - \overline{x_2}$	P(CS)	$\widehat{P(CS)}^{0.2}$	$\widehat{P(CS)}^{0.5}$	$\widehat{P(CS)}^{0.8}$
4	0.1	0.5562	0.8344	0.6677	0.5560
	0.3	0.6643	0.8431	0.6788	0.5601
	0.5	0.7602	0.8598	0.6954	0.5691
	1.0	0.9213	0.9223	0.7880	0.6268
	1.5	0.9830	0.9749	0.9129	0.7712
	2.0	0.9976	0.9954	0.9823	0.9361
	3.0	0.9999	0.9999	0.9998	0.9995
6	0.1	0.5687	0.8349	0.6683	0.5563
	0.3	0.6983	0.8498	0.6819	0.5625
	0.5	0.8067	0.8721	0.7105	0.5769
	1.0	0.9583	0.9498	0.8457	0.6795
	1.5	0.9583	0.9915	0.9679	0.8925
	2.0	0.9997	0.9993	0.9975	0.9902
	3.0	0.9999	0.9999	0.9999	0.9999
10	0.1	0.5884	0.8360	0.6694	0.5568
	0.3	0.7488	0.8573	0.6952	0.5676
	0.5	0.8682	0.8944	0.7414	0.5946
	1.0	0.9873	0.9802	0.9294	0.8018
	1.5	0.9996	0.9990	0.9964	0.9859
	2.0	0.9999	0.9999	0.9999	0.9998
	3.0	1.0000	1.0000	0.9999	0.9999
20	0.1	0.6240	0.8388	0.6722	0.5580
	0.3	0.8286	0.8791	0.7197	0.5819
	0.5	0.9430	0.9373	0.8178	0.6518
	1.0	0.9992	0.9983	0.9933	0.9744
	1.5	0.9999	0.9999	0.9999	0.9999
	2.0	0.9999	0.9999	0.9999	0.9999
	3.0	1.0000	1.0000	1.0000	1.0000

표4. 모평균의 차이가 정규분포를 따르는 경우 P(CS)에 대한 추정치

$n = 20, \overline{x_1} - \overline{x_2} = 0.3, P(CS) = 0.8286$인 경우

평균(μ)	분산(η^2)	$\widehat{P(CS)}^N$
-1.0	0.1	0.8302
	0.3	0.7278
	0.5	0.7419
	1.0	0.7696
-0.5	0.1	0.7105
	0.3	0.7374
	0.5	0.7605
	1.0	0.7835
-0.3	0.1	0.6959
	0.3	0.7497
	0.5	0.7707
	1.0	0.7897
0.0	0.1	0.7276
	0.3	0.7751
	0.5	0.7883
	1.0	0.7994
0.3	0.1	0.8026
	0.3	0.8061
	0.5	0.8079
	1.0	0.8096
0.5	0.1	0.8569
	0.3	0.8283
	0.5	0.8216
	1.0	0.8166
1.0	0.1	0.9535
	0.3	0.8825
	0.5	0.8566
	1.0	0.8347

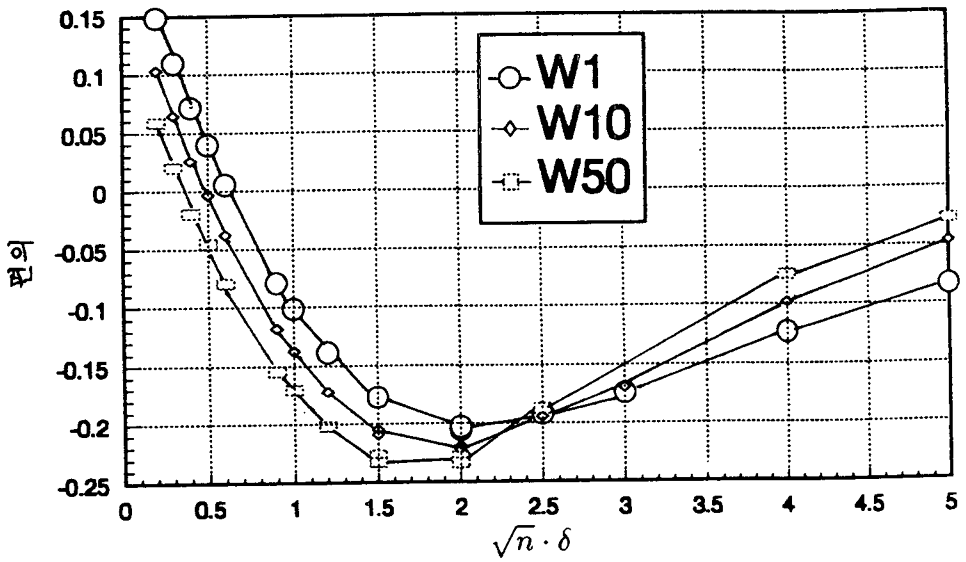

그림 1. P(CS)의 사전분포가 가중 절단 베타분포를 따르는 경우 편의

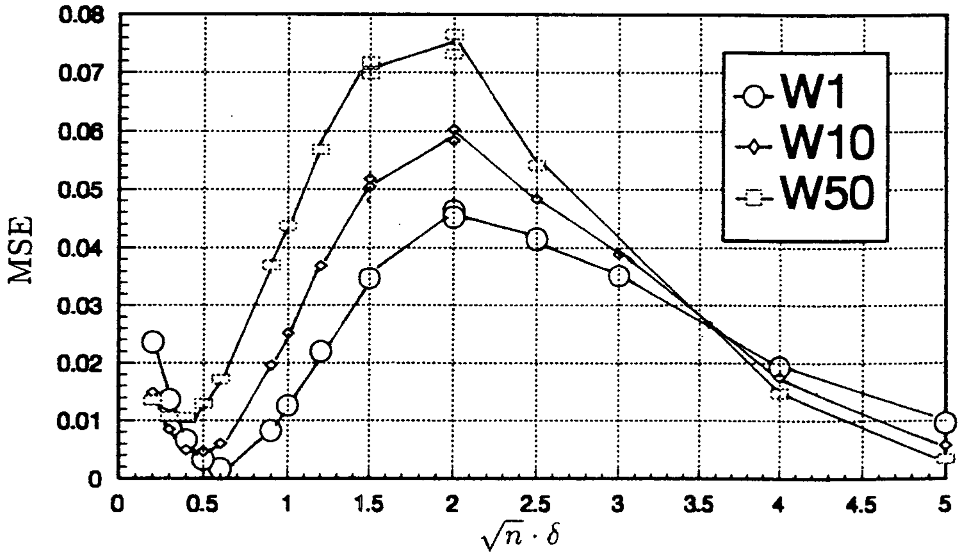

그림 2. P(CS)의 사전분포가 가중 절단 베타분포를 따르는 경우 MSE

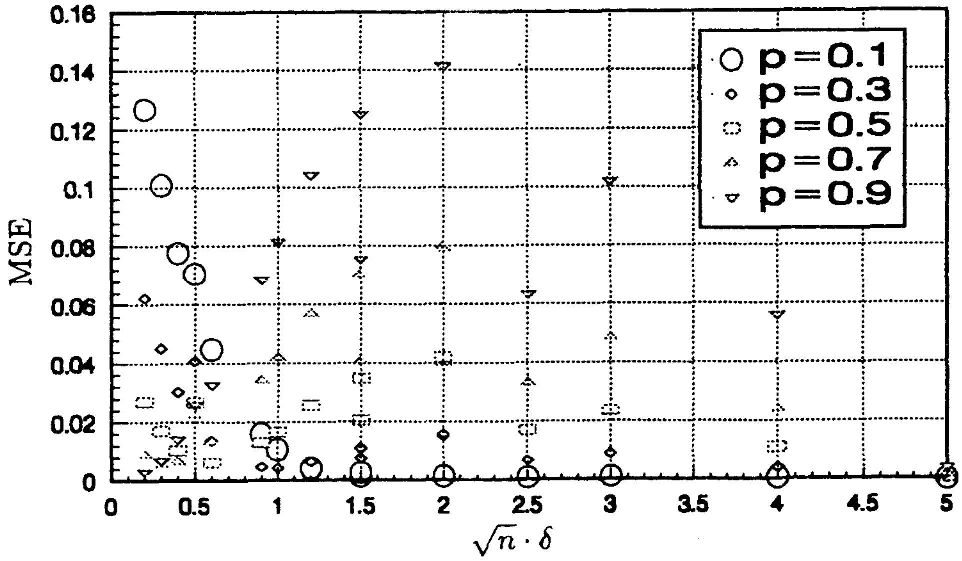

그림 3. P(CS)의 사전분포가 두점 이산형 분포를 따르는 경우 편의

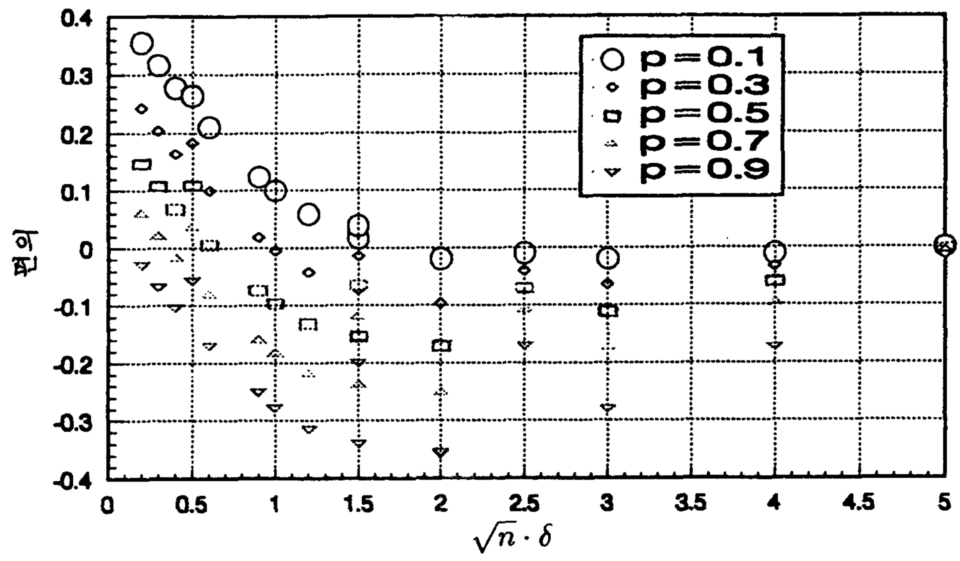

그림 4. P(CS)의 사전분포가 두점 이산형 분포를 따르는 경우 MSE

참 고 문 헌

1. Bechhofer, R. O.(1954), A Single Sample Multiple Decision Procedure for Ranking Means of Normal Populations with Known Variances, *Annals of Mathematical Statistics*, 25, 16-39.
2. Bofinger, E. (1990), Posterior Probability of Correct Selection, *Communications in Statistics - Theory and Methods*, **19(2)**, 599-616.
3. Faltin, F. W. (1980), A Quantile Unbiased Estimator of the Probability of Correct Selection Achieved by Bechhofer's Single-Stage procedure for the Two population Normal Means Problems, *Abstract 80t-60, IMS Bulletin*, **9**, 180-181.
4. Faltin, F. W. and McCulloch, C. E. (1983), On the Small Sample Properties of the Olkin-Soble-Tong Estimator of the Probability of Correct Selection, *Journal of the American Statistical Association*, **78**, 464-467.
5. Gupta, S. S. and Panchapakesan, S. (1979), *Multiple Decision Procedures : Theory and Methodology of Selecting and Ranking Popultions*, John Wiley & Sons, New York.
6. Gibbons, J. D., Olkin, I. and Sobel, M. (1977), *Selecting and Ordering Populations : A new Statistical Methodology*, John Wiley & Sons, New York.
7. McCulloch, C. E. and Dechter, A. (1985), An Empirical Bayes Approach to Estimationg the Probability of Correct Selection, *Communications in Statistics-Simulation and Computation*, **14**, 173-186.
8. Olkin, I., Sobel, M. and Tong, Y. L. (1976), Estimating the True Probability of a Correct Selection for Location and Scale Parameter Families, *Technical Report 110*, Stanford University, Department of Statistics.
9. Sohn, J. K. and Kang, S. G. (1992), On the Estimation of the Tre Probability of a Correct Selection, *Communications in Statistics - Simulation and Computation*, **21(2)**, 445-462.
10. Sohn, J. K. and Kim. Y. H. (1987), A Bayes Approach for Estimating the Probability of a Correct Selection, *Reseach Review of Kyungpook National Univ.*, **44**, 187-190.