

## A Comparison of Variance Lower Bound between the Optimum Allocation and the Power Allocation

Chang Kyoon Son<sup>1)</sup>

### Abstract

In this paper, we study the efficiency of the stratified estimator in related with the variance lower bound of Horvitz-Thompson estimator subject to the superpopulation model. Especially, we compare the variance lower bound of optimum allocation with that of power allocation subject to Dalenius-Hodges stratification.

Keywords : Superpopulation model, Dalenius-Hodges stratification, Variance lower bound.

### 1. 서론

일반적으로 층화 표본설계는 추정량의 정도를 높이고, 지역 또는 연령과 같은 층화 변수를 사용함으로써 표본 설계가 용이하다는 장점을 가진다. 층화 추출설계에서 사용된 층의 수가 적절할 때, Horvitz-Thompson(:HT) 추정량의 분산에 대한 최대 하한과 그와 연관된 추정량들의 효율성은 Bethel(1989)에 의해 논의된 바 있다.

이 과정에서 초 모집단 모형을 적용하였으며, 추정량으로는 HT추정량을 사용하였고, 이용 가능한 보조정보로는 단변량 보조정보를 가정하였다. 그러나 모집단이 다양한 크기를 가진 부차 모집단들로 구성되어 있는 경우 기존의 표본 배분 방법을 적용하게 되면 부차 모집단의 특성들을 제대로 반영하지 못하는 문제점이 발생한다. 즉, 층화 표본 설계에서 표본 배분을 위해 층별 총합과 전체 모집단 총합을 필요로 하며, 특별히 네이만 배분의 경우 전체 모집단과 큰 층에 대해 필요 이상의 정도를 제공하지만, 작은 층의 경우에는 충분하지 않다.

이러한 문제점을 적절히 해결하기 위한 표본 배분 방법중의 하나가 떡 배분이며, 이 방법은 모집단 전체를 대상으로 하는 표본 배분 방법보다는 전체 모집단을 구성하고 있는 다양한 크기를 가진 부차 모집단에 대한 조사의 경우 표본 배분의 문제에 초점을 맞추고 있다.

따라서 본 논문에서는 떡 배분으로 적용한 경우 층 경계에 대해 Dalenius-Hodges(D-H)(1959)가 제시한 방법을 적용했을 때 분산의 하한을 도출하며, 최적배분방법과의 효율성을 비교하고자 한다.

논문의 구성은 2절에서 초 모집단 모형 하에서 일반적인 HT 추정량의 분산의 하한을 소개하고,

---

1) Full-time Lecturer, Department of Computer Science, Dongshin University, JeonNam 520-714, Korea  
E-mail: ckson85@blue.dongshinu.ac.kr

3절에서는 D-H 층화 방법에 최적배분과 역 배분을 적용했을 때의 분산의 하한을 도출하였으며, 4절에서는 이들 간의 효율성과 분산의 하한을 구하기 위해 모의실험을 실시하였으며, 마지막으로 5절에서는 결론에 대해 언급하였다.

## 2. 모형 가정과 최소 분산

추출설계  $D$ 를 이용하여 유한모집단  $U = \{(y_1, x_1), (y_2, x_2), \dots, (y_N, x_N)\}$ 을 조사한다고 하면,  $(y_k, x_k)$ 는 결합분포를 가진 확률변수  $(Y, X)$ 의 실현치이고,  $(Y, X)$ 는 다음과 같은 선형 모형을 만족한다고 하자.

$$Y = \alpha + \beta X + \gamma(X)\varepsilon \quad (2.1)$$

여기서  $\alpha, \beta$ 는 상수이고,  $X$ 와  $\varepsilon$ 는 서로 독립이며,  $E(\varepsilon) = 0$ ,  $V(\varepsilon) = \sigma^2$ ,  $\gamma(X) > 0$ 이다.

$y_k$ 들의 모평균  $\bar{Y} = 1/N \sum y_k$ 의 추정량으로 HT 추정량인  $\hat{Y} = 1/n \sum y_k / \pi_k$ 를 적용하자. 층화변수  $X$ 가 기지이므로 최적기준은 다음의 예측 분산(anticipated variance)을 최소로 한다.

$$\begin{aligned} v(\hat{Y} - \bar{Y}) &= E_M[v_D(\hat{Y} - \bar{Y})|X] + v_M[E_D(\hat{Y} - \bar{Y})|X] \\ &= E_M[v_D(\hat{Y} - \bar{Y})|X] \\ &= E_M[v_D(\hat{Y})|X] \end{aligned} \quad (2.2)$$

모형 (2.1) 하에서  $\hat{Y} = \alpha + \beta X$ 라 하면 예측 분산은 다음과 같다.

$$\begin{aligned} E_M[v_D(\hat{Y})|X] &= \frac{v_D(\hat{Y})}{N^2} + \frac{\sigma_\varepsilon^2}{N^2} \sum \gamma(X_k)^2 \left( \frac{1}{\pi_k} - 1 \right) \\ &\geq \frac{\sigma_\varepsilon^2}{N^2} \sum \gamma(X_k)^2 \left( \frac{1}{\pi_k} - 1 \right) \\ &\geq \frac{\sigma_\varepsilon^2}{N^2} \left[ \frac{1}{n} \left( \sum \gamma(X_k) \right)^2 - \sum \gamma(X_k)^2 \right] \\ &= \frac{\sigma_\varepsilon^2}{n} \left[ [E(\gamma(X))]^2 - \frac{n}{N} E(\gamma(X)^2) \right] \end{aligned} \quad (2.3)$$

여기서  $E(h(X)) = \sum h(X_k)/N$ 이다.

식(2.3)은 분산의 하한으로 이용될 수 있으며, Godambe와 Joshi(1965)는 설계 비편향 예측 추정량(design unbiased predicted estimator)의 기대 분산의 하한으로 이 값을 도출하였고, Cassel(1977)은 일반화 차의 추정량의 기대분산으로 도출한바 있다.

특별히 표본설계  $D$ 가 층화인 경우  $E_h(\cdot)$ 와  $S_h^2(\cdot)$ 를 각각  $h$ 층의 기대값과 분산이라 하자. 그러면 층화 표본 설계 하에서  $\hat{Y}$ 의 예측 분산은 다음과 같다.

$$\begin{aligned} E_M[v_D(\hat{Y})|X] &= E_M \left[ \sum_h W_h^2 S_h^2(Y) \left( \frac{1}{n_h} - \frac{1}{N_h} \right) | X \right] \\ &= \sum_h \frac{(1-f_h)}{n_h} W_h^2 [\beta^2 S_h^2(X) + \sigma_\varepsilon^2 E_h(\gamma(X)^2)] \end{aligned} \quad (2.4)$$

여기서  $f_h = n_h/N_h$ 이다.

위의 식(2.4)로부터  $n_h$ 가  $N_h$ 에 비해 상대적으로 작다면  $f_h \approx 0$ 으로 무시될 수 있으며, 따라서 식(2.4)는 다음과 같이 다시 표현할 수 있다.

$$E_M[v_D(\widehat{Y})|X] \approx \sum_h \frac{1}{n_h} W_h^2 [\beta^2 S_h^2(X) + \sigma_\varepsilon^2 E_h(\gamma(X)^2)] \quad (2.5)$$

### 3. 분산의 하한

#### 3.1 최적배분

최적배분 하에서 모든 층에 대해 동일한 추출비용을 가정하면, 다음과 같은 최적 표본크기를 구할 수 있다.

$$n_h = n \frac{W_h S_h(Y)}{\sum_h W_h S_h(Y)} \quad (3.1.1)$$

모집단 크기  $N$ 이 크다고 가정하면, 식(2.5)에 식(3.1.1)을 대입하여 예측 분산은 다음과 같다.

$$E_M[v_D(\widehat{Y})|X] \approx \frac{1}{n} \left( \sum_h W_h [\beta^2 S_h^2(X) + \sigma_\varepsilon^2 E_h(\gamma(X)^2)]^{1/2} \right)^2 \quad (3.1.2)$$

식(3.1.2)로부터 분산의 하한을 결정하기 위해 층의 수  $H$ 가 증가함에 따라 식(3.1.2)의 인수들을 테일러 정리를 이용하여 다음과 같이 유도할 수 있다.

$$\begin{aligned} & \sum_h W_h [\beta^2 S_h^2(X) + \sigma_\varepsilon^2 E_h(\gamma(X)^2)]^{1/2} \\ &= O(1) \sum_{h=1}^H W_h S_h^2(X) + \sum_{h=1}^H W_h [\sigma_\varepsilon^2 E_h(\gamma(X)^2)]^{1/2} \end{aligned} \quad (3.1.3)$$

그런데 추정량의 분산이 층의 경계에 좌우되므로 적절하게 층 경계가 결정된다면, 분산을 최소로 하는 층의 경계로부터 분산의 하한을 결정할 수 있을 것이다. 이를 위해 다음과 같이 D-H의 층화방법을 적용하여  $a_h$ 와  $a_{h-1}$ 를  $h$ 층의 상한과 하한이라 하면, 층의 경계점은 다음의 식을 만족한다.

$$\int_{a_{h-1}}^{a_h} f_x(t)^{1/2} dt = c/H$$

여기서,  $c = \int f_x(t)^{1/2} dt$ 이다.

$M_h = (a_h + a_{h-1})/2$ 라 하고, 층수  $H$ 이 크면, 다음과 같은 근사식이 성립한다.

$$S_h^2(X) \approx \frac{(a_h - a_{h-1})^2}{12},$$

$$c/H \approx f_x(M_h)^{1/2} (a_h - a_{h-1}),$$

$$W_h = \int_{a_{h-1}}^{a_h} f_x(t) dt \approx f_x(M_h) (a_h - a_{h-1}) \approx \left( \frac{c}{H} \right)^2 (a_h - a_{h-1})^{-1}$$

이와 같은 성질로부터 식(3.1.3)의 오른쪽 첫 항은 근사적으로 다음과 같다.

$$\begin{aligned}\sum_{h=1}^H W_h S_h^2(X) &\approx \sum_{h=1}^H \frac{(c/H)^2}{a_h - a_{h-1}} \frac{(a_h - a_{h-1})^2}{12} \\ &= O(H^{-2})\end{aligned}$$

이와 유사한 방법으로 식(3.1.3)의 오른쪽 두 번째 항은 다음과 같이 표현할 수 있다.

$$\sum_{h=1}^H W_h [E_h(\gamma(X))^2]^{1/2} \approx \sum_{h=1}^H W_h E_h(\gamma(X)) = E(\gamma(X))$$

결과적으로 식(3.1.2)는 근사적으로 다음과 같은 식으로 표현이 가능하다.

$$\begin{aligned}E_M[v_D(\widehat{Y})|X] &\approx \frac{1}{n} [O(H^{-2}) + \sigma_e^2 E(\gamma(X))]^2 \\ &= O(H^{-2}) + \frac{1}{n} \sigma_e^2 [E(\gamma(X))]^2 \\ &\rightarrow \frac{1}{n} \sigma_e^2 [E(\gamma(X))]^2\end{aligned}\quad (3.1.4)$$

### 3.2 떡 배분

층화 단순임의 표본을 추출한다고 할 때,  $N_h$ 는  $h=1, 2, \dots, H$  층의 모집단 크기이다. 층화 모집단으로부터  $\sum_h n_h = n$ 의 조건하에서 다음의 손실함수를 최소로 하는  $h$ 층의 표본 수  $n_h$ 를 결정하고자 한다.

$$L = \sum_h (X_h^q cv(\widehat{Y}_h))^2 \quad (3.2.1)$$

식(3.2.1)에서  $cv(\widehat{Y}_h) = \sqrt{v(\widehat{Y}_h)}/\widehat{Y}_h$ 로서 변동계수를 나타내며,  $X_h$ 는  $h$ 층에 대한 층화 변수의 총합이고,  $q$ 는  $0 \leq q \leq 1$ 인 범위를 갖는 상수이다. 또한,  $\widehat{Y}_h = N_h \sum_i y_{hi}/n_h$ 로서  $h$ 층의 총합  $Y_h$ 의 추정량이다.

그러면 식(3.2.1)의 손실함수  $L$ 은  $n_h$ 가 다음과 같을 때 최소가 된다.

$$n_h = n \frac{X_h^q cv(\widehat{Y}_h)}{\sum_h X_h^q cv(\widehat{Y}_h)} = n \frac{S_h(Y) X_h^q / \overline{Y}_h}{\sum_h S_h(Y) X_h^q / \overline{Y}_h} \quad (3.2.2)$$

이 때  $v(\widehat{Y}_h) = (1/n_h - 1/N_h) S_h^2(Y)$ 이며,  $S_h^2(Y) = 1/(N_h - 1) \sum (y_{hi} - \overline{Y}_h)^2$ ,  $\overline{Y}_h = Y_h/N_h$ 이다.

즉, 손실함수 (3.2.1)을 최소로 하는  $h$ 층의 표본크기는 식(3.2.2)로서 이러한 배분 방법을 떡 배분(power allocation)이라 한다.

떡 배분의 가장 큰 특징은 앞서서도 언급한 바와 같이 전체 정도는 약간 떨어지더라도, 상대적으로 작은 층의 정도를 높이는 방법으로 표본을 배분하는 방법이다. 떡 배분의 경우  $q(0 \leq q \leq 1)$ 의 선택 값에 따라 특정한 표본 배분이 될 수 있으며, 특별히  $q=1$ 와  $X_h = Y_h$ 일 경우 Neyman 배분과 일치한다. 만일 층별로  $cv(\widehat{Y}_h)$ 가 일정하다면, 식(3.2.2)의 떡 배분은 다음과 같이 축소된다.

$$n_h \propto (X_h)^q \quad (3.2.3)$$

이 때,  $q$ 의 값은 일반적으로 1/2 또는 1/3이 적절한 것으로 알려져 있다.

식(2.5)에 식(3.2.2)를 대입하여 정리하면 다음과 같다.

$$\begin{aligned}
 E_M[v_{DP}(\widehat{Y})|X] &\approx \sum_h \frac{1}{n_h} W_h^2 [\beta^2 S_h^2(X) + \sigma_\epsilon^2 E_h(\gamma(X)^2)] \\
 &\approx \frac{1}{n} \left( \sum_h \frac{\bar{Y}_h}{X_h^q} W_h^2 [\beta^2 S_h^2(X) + \sigma_\epsilon^2 E_h(\gamma(X)^2)]^{1/2} \right) \\
 &\quad \times \left( \sum_h \frac{X_h^q}{\bar{Y}_h} [\beta^2 S_h^2(X) + \sigma_\epsilon^2 E_h(\gamma(X)^2)]^{1/2} \right)
 \end{aligned} \tag{3.2.4}$$

**정리1]** 멱 배분의 예측분산  $E_M[v_{DP}(\widehat{Y})|X]$ 과 최적 배분의 예측분산  $E_M[v_D(\widehat{Y})|X]$ 은 다음과 같은 관계가 성립한다.

$$E_M[v_{DP}(\widehat{Y})|X] \geq E_M[v_D(\widehat{Y})|X]$$

증명) 식(3.2.4)로부터 Cauchy-Schwartz 부등식을 이용하기 위해 다음과 같이  $a_h$ 와  $b_h$ 을 정의 하자. 또한 전개를 간단히 하기 위해  $S_h(Y) = [\beta^2 S_h^2(X) + \sigma_\epsilon^2 E_h(\gamma(X)^2)]^{1/2}$ 라 하고,

$a_h = W_h \sqrt{\frac{\bar{Y}_h}{X_h^q}} S_h(Y)$ ,  $b_h = \sqrt{\frac{X_h^q}{\bar{Y}_h}} S_h(Y)$  라 하면, 식(3.2.4)로부터 다음의 부등식을 만족한다.

$$\begin{aligned}
 \left( \sum_h \frac{\bar{Y}_h}{X_h^q} W_h^2 S_h(Y) \right) \left( \sum_h \frac{X_h^q}{\bar{Y}_h} S_h(Y) \right) &\geq \left( \sum_h W_h \sqrt{\frac{\bar{Y}_h}{X_h^q}} S_h(Y) \sqrt{\frac{X_h^q}{\bar{Y}_h}} S_h(Y) \right)^2 \\
 &= \left[ \sum_h W_h S_h(Y) \right]^2
 \end{aligned}$$

따라서 예측분산은 다음과 같다.

$$\begin{aligned}
 E_M[v_{DP}(\widehat{Y})|X] &\geq \frac{1}{n} \left[ \sum_h W_h [\beta^2 S_h^2(X) + \sigma_\epsilon^2 E_h(\gamma(X)^2)]^{1/2} \right]^2 \\
 &= E_M[v_D(\widehat{Y})|X]
 \end{aligned}$$

■

**따름정리]** 만일 층별 가중치  $W_h \propto (X_h^q / \bar{Y}_h)$  이면, 멱 배분의 분산은 최적배분의 분산과 같다.

$$E_M[v_{DP}(\widehat{Y})|X] \approx \frac{1}{n} \left[ \sum_h W_h [\beta^2 S_h^2(X) + \sigma_\epsilon^2 E_h(\gamma(X)^2)]^{1/2} \right]^2 \tag{3.2.5}$$

증명) 정리1)의 증명 참조.

■

**정리2]** D-H 층화의 경우 멱 배분의 예측분산  $E_M[v_{DP}(\widehat{Y})|X]$ 의 하한은 다음과 같다.

$$E_M[v_{DP}(\widehat{Y})|X] \rightarrow \frac{1}{n} \sigma_\epsilon^2 (E(\gamma(X)))^2 \tag{3.2.6}$$

증명) 예측분산의 식(3.2.4)로부터 첫 번째 항은 테일러 정리를 이용하여 정리하면 다음과 같다.

$$O(1) \sum_h W_h^2 S_h^2(X) \left( \frac{\bar{Y}_h}{X_h^q} \right) + \sum_h W_h^2 [\sigma_\varepsilon^2 E_h(\gamma(X)^2)]^{1/2} \left( \frac{\bar{Y}_h}{X_h^q} \right) \quad (3.2.7)$$

또한 두 번째 항은 같은 방법으로 정리하면, 다음과 같다.

$$O(1) \sum_h S_h^2(X) \left( \frac{X_h^q}{\bar{Y}_h} \right) + \sum_h \sigma_\varepsilon^2 E_h(\gamma(X)^2)^{1/2} \left( \frac{X_h^q}{\bar{Y}_h} \right) \quad (3.2.8)$$

식(3.2.7)에 D-H 총화를 적용하면,

$$\sum_h W_h^2 S_h^2(X) \left( \frac{\bar{Y}_h}{X_h^q} \right) \approx \sum_h \frac{(c/H)^4}{(a_h - a_{h-1})^2} \frac{(a_h - a_{h-1})^2}{12} \left( \frac{\bar{Y}_h}{X_h^q} \right) \approx O(H^{-4}) \quad (3.2.9)$$

그런데 총의 수  $H$ 가 적절히 크면,  $[E_h(\gamma(X)^2)]^{1/2} \approx E_h(\gamma(X))$  가 성립하므로 다음과 같다.

$$\sum_h W_h^2 [E_h(\gamma(X)^2)]^{1/2} \left( \frac{\bar{Y}_h}{X_h^q} \right) \approx \sum_h W_h^2 E_h(\gamma(X)) \left( \frac{\bar{Y}_h}{X_h^q} \right) \quad (3.2.10)$$

따라서 식(3.2.9)와 (3.2.10) 그리고 (3.2.8)을 결합하여 정리하면 다음과 같다.

$$\begin{aligned} E_M[v_{DP}(\hat{Y})|X] &\approx \frac{1}{n} [O(H^{-4}) + \sigma_\varepsilon \sum_h W_h^2 E_h(\gamma(X))] [O(1) + \sigma_\varepsilon \sum_h E_h(\gamma(X))] \\ &\approx \frac{1}{n} \sigma_\varepsilon^2 [\sum_h W_h^2 E_h(\gamma(X))] [\sum_h E_h(\gamma(X))] \\ &\rightarrow \frac{1}{n} \sigma_\varepsilon^2 (E(\gamma(X)))^2 \end{aligned}$$

■

#### 4. 수치적 예제

3절에서 언급한 떡 배분의 분산의 하한에 대한 성질을 알아보기 위해 다음과 같은 모집단 가정 하에서 모의실험을 실시하였다.

확률변수  $X$ 의 분포를 평균이  $\mu$ 이고, 분산이  $\theta^2$ 인 로그 정규분포를 가정하고, 이때  $\gamma(X) = X^g$ 라 하였다. 그러면, 확률변수  $X$ 의 확률밀도함수는 다음과 같다.

$$f(x) = \frac{1}{x\sqrt{2\pi\theta}} \exp\left[-\frac{(\ln x - \mu)^2}{2\theta^2}\right] \quad (4.1)$$

그리고  $\gamma(X)$ 의 기대값은 다음과 같다(Johnson 과 Kotz,1970).

$$E_M(\gamma(X)) = E_M(X^g) = \exp(g\mu + g^2\theta^2/2) \quad (4.2)$$

이다.

D-H 총화의 경우  $X$ 의 분포가 연속형 분포인 로그정규분포이므로 임의의  $h$ 의 총 경계점은 다음의 식을 만족한다.

$$\Phi\left[\frac{\ln a_h - (\mu + \theta^2)}{\theta\sqrt{2}}\right] = \frac{h}{H} \quad (4.3)$$

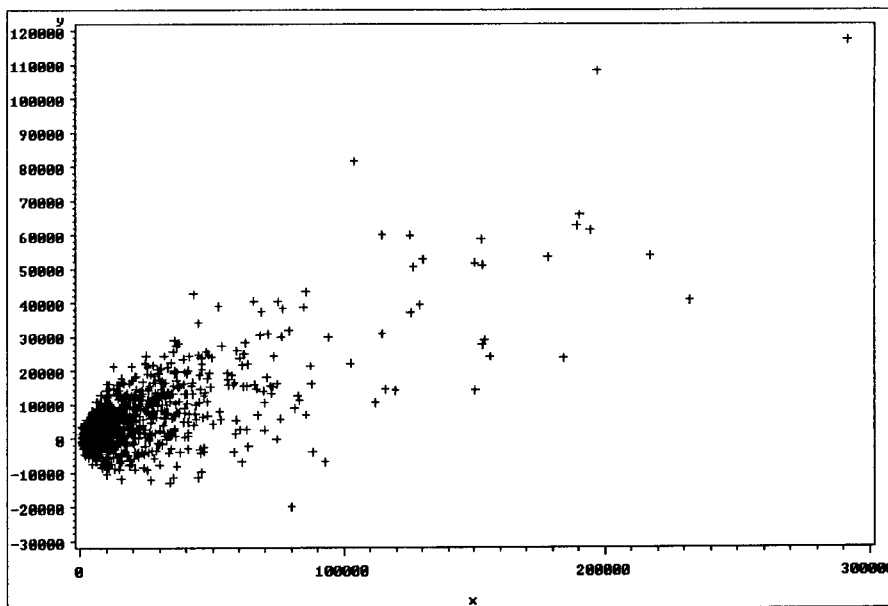
여기서  $1 \leq h \leq H-1$  이다.

수치적 예를 위해 평균  $\mu=9.5$ , 표준편차  $\theta=1.0$ 를 가정하자. 또한 모형(2.1)을 다음과 같다고 가정하자.

$$Y=0.25X + X^{\theta}\epsilon$$

이때  $\sigma_{\epsilon}^2=2,500$ 이라 하자.

확률변수  $X$ 와  $\epsilon$ 는 각각  $Normal(seed=1234567)$ 를 이용하여 로그 정규 분포 하에서 난수를 발생시켜 모의실험을 실시하였다. 다음의 [그림4.1]은 모집단의 특성을 나타낸 그래프로서  $y$ 의 작은 값들이 상대적으로 많이 분포한 형태를 갖는다.



[그림 4.1] 모집단의 분포형태

이와 같은 모집단에 대해 D-H 층화를 적용하여 각 층별 경계점을 식(4.3)을 이용하여 구할 수 있으며, 그에 따르는 층별 평균과 표준편차, 그리고 층별 가중치는 다음의 [표 4.1]과 같다. 이때 층의 수를  $H=10$ 개로 하고, 모집단 크기  $N=1,000$ , 표본 수  $n=1$ 로 하였다. 참고로 표본 수를  $n=1$ 로 한 이유는 초모집단 모형하에서 최적배분과 역배분간의 분산의 하한을 비교하기 위한 것으로 표본변동을 제외시키기 위해서이다.

[표4.1]로부터 모집단의 형태에 따라 작은 값을 갖는 층(층1과 층2)의 층별 가중치가 다른 층들에 비해 크게 나타났으며, 반대로 큰 값을 갖는 층(층9와 층10)의 경우 층별 가중치가 거의 0에 가깝도록 층을 나누었다.

이와 같이 D-H 층화 방법으로 층별경계점을 결정하고, 다음으로 각 층에 최적배분 방법과 역 배 분 방법을 각각 적용했을 때, 모형 분산의 하한과 이들 표본 배분 방법들 간의 상대 효율을 구한 결과 [표 4.2]와 같다. 이때 상대 효율은 다음과 같이 정의하였다.

$$Reff = \frac{\hat{v}_{\text{하한}}}{\hat{v}_{\text{방법}}} \times 100(\%) \quad (4.4)$$

상대효율 식(4.4)의 첨자인 “방법”은 최적배분과 먹 배분의 분산의 하한을 의미하며, “하한”은 모형가정에 따라 계산된 분산을 의미한다.

[표4.1] 로그 정규분포 하에서 D-H 층화의 층별 평균과 표준편차

층	층 경계	$W_h$	$E_h(X)$	$E_h(X^{1/2})$	$E_h(X^2)$	$S_h(X)$	$S_h(Y)$
1	0 - 5,942	0.227	3,588.83	58.605	1,477,5169.4	1,379.80	2,869.42
2	5,942 - 11,071	0.210	8,366.69	91.088	72,329,682.24	1,529.46	4,940.71
3	11,071 - 17,162	0.174	13,762.54	117.072	192,536,593	1,774.00	5,894.02
4	17,162 - 25,142	0.117	20,661.41	143.523	432,086,480.4	2,288.51	7,177.28
5	25,142 - 36,316	0.118	30,457.24	174.278	937,910,982.8	3,217.96	8,363.83
6	36,316 - 52,454	0.066	42,645.35	206.280	1,834,637,744	4,032.10	10,712.55
7	52,454 - 76,844	0.047	64,426.93	253.453	4,199,326,631	7,039.29	12,150.80
8	76,844 - 192,677	0.036	120,681.91	344.047	15,720,278,548	34,484.62	23,005.94
9	192,677 - 221,947	0.003	202,915.00	450.322	41,277,646,405	12,438.84	29,481.48
10	221,947 -	0.002	261,828.27	510.867	69,434,620,266	41,966.12	54,291.43

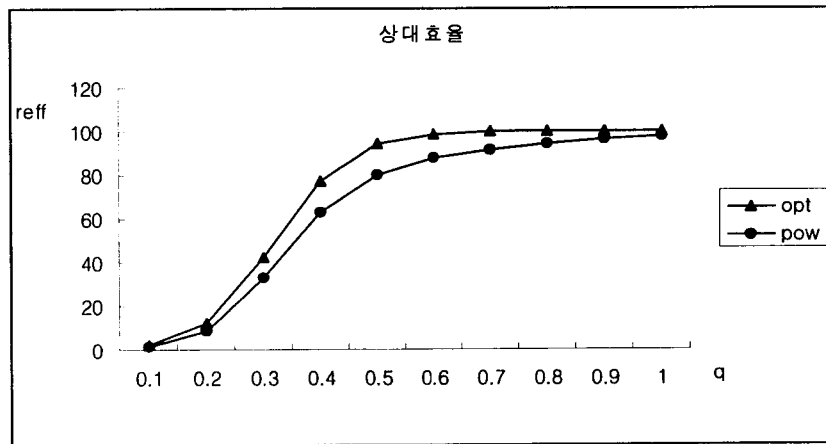
[표 4.2]로부터  $q$ 의 변화에 따라 두 방법간의 효율성을 알아본 결과 먹 배분의 분산의 하한이  $q$ 가 1에 근접함에 따라 최적배분의 분산의 하한과 거의 같아짐을 알 수 있다.

[표4.2] 최적배분과 먹배분의 분산의 하한에 대한 상대효율(%)

방법 \ $q$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
최적배분	2.06	11.82	42.01	76.88	94.08	98.91	99.76	99.89	99.89	99.89
먹배분	1.49	8.89	32.60	62.54	80.21	87.67	91.38	94.03	96.13	97.73

이를 그래프로 표현하면 [그림4.2]와 같고, 이와 같은 결과는 이미 3절에서 언급한 먹배분의 성질과 일치하는 결과이다.





[그림4.2] 최적배분과 역 배분의 효율성 비교

## 5. 결론

본 논문에서는 층화 표본 설계의 경우 Dalenous-Hodges의 층 경계 결정 방법에 역배분을 적용했을 때 분산의 성질을 살펴보고, 모의실험을 통해  $q$ 의 값의 변화에 따라 상대효율을 알아본 결과 역 배분의 경우  $q$ 의 값이 1에 근접함에 따라 최적 배분의 분산의 하한에 근사함을 알 수 있었다. 이와 같은 결과는 3절에서 언급한 최적배분과 역배분간의 관계와도 일치하는 사실이다.

## 참고문헌

- [1] 손창균, 홍기학, 이기성(2000), "Minimum Variance Estimation in Stratified Sampling by an Extended Ekman Rule", 한국통계학회 호남제주지회 발표 논문.
- [2] 손창균, 홍기학, 이기성 (2002), "Minimum Variance Estimation for the Power Allocation in Stratified Sampling", 한국통계학회 추계 학술 발표 논문.
- [3] Bankier, M. D.(1988), "Power Allocations : Determining Sample Sizes for Subnational Area", *The American Statistician*, Vol, 42, No. 3, pp. 174-177.
- [4] Bethel, J.(1989), "Minimum Variance Estimation in Stratified Sampling", *Journal of American Statistics Association*, Vol. 84, pp. 260-265.
- [5] Cassel, C. M., Särndal C. E., and Wretman, J. H.(1977), *Foundations of Inference in Survey Sampling*, New York : John Wiley.
- [6] Dalenius, T.(1950), "The Problem of Optimum Stratification," *Skandinavisk Aktuarietidskrift*.
- [7] Dalenius, T., and Hodges, J. L.(1959), "Minimum Variance of Stratification," *Journal of American Statistics Association*, Vol. 54, pp. 88-101.

- [8] Godambe, V. P., and Joshi, V. M.(1965), "Adimissibility and Bayes Estimation in Sampling Finite Populations I", *Annals of Mathematical Statistics*, Vol. 36, pp. 1701-1722.
- [9] Särndal, C. E.(1980), "On  $\pi$ -Inverse Weighting Versus Best Linear Unbiased Weighting in Probability Sampling", *Biometrika*, Vol. 67, pp. 639-650.
- [10] Wright, R. L.(1983a), "Finite Population Sampling With Multivariate Auxiliary Information," *Journal of the American Statistical Association*, Vol. 78, pp. 879-884.

[ 2002년 11월 접수, 2003년 3월 채택 ]