有限多項母集團에 있어서 두 比率의 差의 分散의 不偏推定量

李　　鍾　　厚

Unbiased Estimator of the Variance of the Difference Between any two Proportions in a Finite Multinomial Population

Jong Hoo Lee

目　　次

Abstract

Let us consider a multinomial Population Π which is divided into rth mutually exclusive classes $\Pi_1, \cdots, \Pi_r$, consisting of $N_1, \cdots, N_r (N_1 + \cdots N_r = N)$ units respectively and each element in Π_i numbered a_i.

Then the proportion of Π_i to Π,

$$p_i = \frac{N_i}{N}, \quad 0 \leqq p_i \leqq 1, \quad i = 1, \cdots, r, \quad p_1 + \cdots p_r = 1.$$

Suppose that the simple random sample of size n is drawn from this population without replacement and the outcomes contain n_i elements of Π_i $(i = 1 \cdots, r)$.

The probability that the outcomes contain n_i elements of Π_i is given by

$$p_r(n_1 \cdots n_r) = \frac{\binom{N_1}{n_1}\binom{N_2}{n_2}\cdots\binom{N_r}{n_r}}{\binom{N}{n}},$$

Where

$$0 \leqq n_i \leqq \min(n, \ N_i), \quad i = 1 \cdots r, \quad n_1 + \cdots + n_r = n, \quad N_1 + \cdots + N_r = N.$$

Let $\hat{p}_i$ $(i = 1 \cdots r)$ denote the proportions of Π_i to Π respectively, i. e.

$$\hat{p}_i = \frac{n_i}{n}, \quad 0 \leqq \hat{p}_i \leqq 1, \quad i = 1, \cdots\cdots, r, \quad \sum_{i=1}^{r} \hat{p}_i = 1,$$

Now consider the differences between any two proportions,

$$\left.\begin{array}{l} d=p_i-p_k \\ \hat{d}=\hat{p}_i-\hat{p}_k \end{array}\right\} \ (i\neq k,\ 1\leqq i,\ k\leqq r)$$

In this note we derived the unbiased estimate of the variance of $\hat{d}$, *i.e.*

$$\hat{\sigma}_d^2=\frac{N-n}{N(n-1)n}\sum_{j=1}^{n}[(Y_{ji}-Y_{jk})-(\bar{Z}_i-\bar{Z}_k)]^2,$$

where

$$Y_{ji}=\begin{cases} 1 & X_j=a_i \\ 0 & X_j\neq a_i \end{cases} \quad (j=1,\cdots,\ n;\ i=1,\cdots,r)$$

and

$$\bar{Z}_i=\frac{1}{n}(Y_{1i}+\cdots+Y_{ni}).$$

1. 序 論

크기 N인 有限多項母集團 Π의 構成要素를 相異한 r個의 實數値 $\{a_1,\cdots,a_r\}$, 各 a_i의 個數를 N_i $(i=1,\cdots,r)$ $N_1+\cdots N_r=N$이라 하면 各 a_i의 比率 p_i는

$$p_i=\frac{N_i}{N} \quad (i=1,\cdots,r) \qquad \cdots\cdots(1)$$

이며 母平均 μ는

$$\mu=\frac{1}{N}\ (N_1a_1+\cdots+N_ra_r) \qquad \cdots\cdots(2)$$

이다.

$n(<N)$回 非復元으로 標本을 抽出했을 때 a_i가 나타난 回數를 $n_i(i=1,\cdots,r)$라 하면 $n_1+\cdots+n_r=n$이며 各 標本比率은

$$\hat{p}_i=\frac{n_i}{n} \quad (i=1,\cdots,r) \qquad \cdots\cdots(3)$$

으로 表示된다. 지금

$$\left.\begin{array}{l} d=p_i-p_k \\ \hat{d}=\hat{p}_i-\hat{p}_k \end{array}\right\} \ (i\neq k,\ 1\leqq i,\ k\leqq r) \qquad \cdots\cdots(4)$$

으로 두면

$$E(\hat{d})=p_i-p_k$$

$$V(\hat{d})=\frac{N-n}{N-1}\ \frac{p_i(1-p_i)+p_k(1-p_k)}{n}$$

임을 알 수 있다.(文獻 1)

Yoshiaki는 文獻 2에서 $r=3$ 일 때의 $\hat{p}_i$, $\hat{p}_k$ 및 $\hat{d}$의 信賴領域을 Thebychef의 不等式을 利用하여 評價한 바 있다.

여기서는 $\hat{d}$의 分布를 考察하고 分散 $V(\hat{d})=\frac{N-n}{N-1}\ \frac{p_i+p_k-d^2}{n}$의 不偏推定量

$$\hat{\sigma}_d^2=\frac{N-n}{Nn(n-1)}\sum_{j=1}^{n}\{(Y_{ji}-\bar{Z}_i)-(Y_{jk}-\bar{Z}_k)\}^2$$

을 求하였다.

2. $V(\hat{d})$의 不偏推定量

非復元으로 $n(<N)$回의 任意抽出한 標本을 $\{x_1, \cdots, x_n\}$이라 하고 이것을 實現値로 하는 確率變數를 各各 $X_1, \cdots, X_n$이라 하면

$$E(X_j)=\frac{1}{N}(N_1a_1+\cdots+N_ra_r)=\mu \qquad (j=1,\cdots,n)$$

$$p(X_j=a_i)=\frac{N_i}{N}=p_i \qquad (i=1,\cdots,r)$$

이다. 여기서

$$Y_{ji}=\begin{cases}1 & X_j=a_i\\ 0 & X_j\neq a_i\end{cases} \qquad (j=1,\cdots,n;\ i=1\ \cdots,r)$$

으로 두면

$$\left.\begin{array}{l}P(Y_{ji}=1)=p_i\\ E(Y_{ji})=p_i\\ V(Y_{ji})=p_i(1-p_i)\end{array}\right\} \qquad (j=1\cdots,n;\ i=1,\cdots,r)$$

이다. 그리고

$$\{Y_{1i},\ Y_{2i},\cdots,Y_{ni}\}\ (i=1,\cdots,r)$$

에 關해서

$$Z_i=Y_{1i}+Y_{2i}+\cdots+Y_{ni}=n_i \qquad (i=1,\cdots,r)$$

으로 두면 n_i도 하나의 確率變數이고

$$\sum_{i=1}^{r}Z_i=\sum_{i=1}^{r}n_i=n$$

이다. 또

$$\bar{Z}_i=\frac{1}{n}(Y_{1i}+\cdots+Y_{ni})=\frac{n_i}{n}=\hat{p}_i, \qquad (i=1,\cdots,r)$$

는 a_i에 關한 標本比率이다.

다음에 $V(\hat{d})$의 不偏推定量을 求하기 위하여 若干의 補助定理를 둔다.

補助定理 1. (1)

$$\begin{aligned}E\{\textstyle\sum_j(Y_{ji}-\bar{Z}_i)^2\}&=E\{\textstyle\sum_jY_{ji}^2\}-nE(\bar{Z}_i^2)\\ &=n\sigma_{p_i}^2+np_i^2-n\frac{N-n}{N-1}\cdot\frac{\sigma_{p_i}^2}{n}-np_i^2\\ &=n\left(1-\frac{N-n}{N-1}\cdot\frac{1}{n}\right)\sigma_{p_i}^2\\ &=\frac{N(n-1)}{N-1}\sigma_{p_i}^2\\ &=\frac{N(n-1)}{N-1}p_i(1-p_i) \qquad (\because\ \sigma_{p_i}^2=V(Y_{ji})=p_i(1-p_i))\end{aligned}$$

(2)

$$\begin{aligned}\mathrm{cov}\ (\bar{Z}_i,\ \bar{Z}_k)&=E(\bar{Z}_i-p_i)(\bar{Z}_k-p_k)\\ &=E[\bar{Z}_i\bar{Z}_k-p_i\bar{Z}_k-p_k\bar{Z}_i+p_ip_k]\\ &=E[\bar{Z}_i\bar{Z}_k]-p_ip_k\\ &=-\frac{N-n}{N-1}\cdot\frac{p_ip_k}{n} \qquad (i\neq k,\ i,\ k=1,\cdots,r) \qquad (\because\ \text{文献 1})\end{aligned}$$

$$\therefore\ E(\bar{Z}_i\bar{Z}_k)=\frac{N(n-1)}{(N-1)n}p_ip_k.$$

補助定理 2. (1) $E(Y_{ji}\cdot Y_{jk})=0 \qquad (i\neq k)$

(2) $E(Y_{\xi i}\cdot Y_{\eta k})=\dfrac{N}{N-1}p_i p_k \qquad (\xi\neq\eta,\ i\neq k)$

證明 (1)

$$\begin{aligned}E(Y_{ji}\cdot Y_{jk})&=1\cdot 0\cdot p(Y_{ji}=1, Y_{jk}=0)\\&+0\cdot 1\cdot p(Y_{ji}=0,\ Y_{jk}=1)+0\cdot 0\cdot p(Y_{ji}=0, Y_{jk}=0)\\&+1\cdot 1\cdot p(Y_{ji}=1,\ Y_{jk}=1)\\&=0\end{aligned}$$

(2) $$\begin{aligned}E\{(Y_{\xi i}-p_i)(Y_{\eta k}-p_k)\}&=E\{Y_{\xi i}\ Y_{\eta k}\}-p_i p_k\\&=\frac{1}{N-1}p_i p_k \qquad (\because\ \text{文献 1})\end{aligned}$$

$$\therefore\ E\{Y_{\xi i}\cdot Y_{\eta k}\}=\frac{N}{N-1}p_i p_k\ (\xi\neq\eta,\ i\neq k)$$

補助定理 3. $E\{\sum_{j=1}^{n}(Y_{ji}-\bar{Z}_i)(Y_{jk}-\bar{Z}_k)\}=-\dfrac{N(n-1)}{N-1}p_i p_k.$

證明 $$\begin{aligned}E\{\bar{Z}_i\cdot Y_{jk}\}&=\frac{1}{n}E\{Y_{1i}Y_{jk}+Y_{2i}Y_{jk}+\cdots+Y_{ni}Y_{jk}\}\\&=\frac{N(n-1)}{n(N-1)}p_i p_k \quad (\because\ \text{補助定理 2})\end{aligned}$$

같이하여

$$E\{Y_{ji}\cdot\bar{Z}_k\}=\frac{N(n-1)}{n(N-1)}p_i p_k$$

補助定理 2와 위의 結果에서

$$\begin{aligned}E\{\sum_{j=1}^{n}(Y_{ji}-\bar{Z}_i)(Y_{jk}-\bar{Z}_k)\}&=E\{\sum_{j}^{n}(Y_{ji}Y_{jk}-\bar{Z}_iY_{jk}-Y_{ji}\bar{Z}_k+\bar{Z}_i\bar{Z}_k\}\\&=-\frac{N(n-1)}{N-1}p_i p_k.\end{aligned}$$

以上의 結果에서 다음의 定理를 얻는다.

定理. 個体 a_i, a_k가 나타날 比率의 差를

$\hat{d}=\hat{p}_i-\hat{p}_k=\bar{Z}_i-\bar{Z}_k$라 하면

$$V(\hat{d})=V(\bar{Z}_i-\bar{Z}_k)=\frac{N-n}{N-1}\frac{1}{n}[p_i(1-p_i)+p_k(1-p_k)+2p_i p_k]$$

이다. 이 $\hat{d}$의 分散 $V(\hat{d})$의 不偏推定量 $\hat{\sigma}_d{}^2$은

$$\hat{\sigma}_d{}^2=\frac{N-n}{N(n-1)n}\sum_j[(Y_{ji}-Y_{jk})-(\bar{Z}_i-\bar{Z}_k)]^2$$

이다.

〔證明〕 補助定理 1,2,3에 依하여

$$\begin{aligned}&E\{\sum_{j=1}^{n}[(Y_{ji}-Y_{jk})-(\bar{Z}_i-\bar{Z}_k)]^2\}\\&=E\{\sum_{j=1}^{n}[(Y_{ji}-\bar{Z}_i)-(Y_{jk}-\bar{Z}_k)]^2\}\\&=E\{\sum_{j=1}^{n}[(Y_{ji}-\bar{Z}_i)^2+(Y_{jk}-\bar{Z}_k)^2-2(Y_{ji}-\bar{Z}_i)(Y_{jk}-\bar{Z}_k)]\}\end{aligned}$$

$$=E\{\sum_j(Y_{ji}-\bar{Z}_i)^2\}+E\{\sum_j^n(Y_{jk}-\bar{Z}_k)^2\}-2E\{\sum_j(Y_{ji}-\bar{Z}_i)(Y_{jk}-\bar{Z}_k)\}$$

$$=\frac{N(n-1)}{N-1}p_i(1-p_i)+\frac{N(n-1)}{N-1}p_k(1-p_k)+2\frac{N(n-1)}{N-1}p_ip_k$$

$$=\frac{N(n-1)}{N-1}\{p_i(1-p_i)+p_k(1-p_k)+2p_ip_k\}$$

그리고 $V(\hat{d})=\frac{N-n}{N-1}\cdot\frac{1}{n}[p_i(1-p_i)+p_k(1-p_k)+2p_ip_k]$ 이므로 $V(\hat{d})$의 不偏推定量 $\hat{\sigma}_d{}^2$은

$$\hat{\sigma}_d{}^2=\frac{N-n}{N(n-1)n}\sum_{j=1}^n[(Y_{ji}-Y_{jk})-(\bar{Z}_i-\bar{Z}_k)]^2$$

이다.

參考文獻

1) 李鍾厚 : 同一標本 內의 두 比率의 差의 分布, 東亞論叢 第三輯 1966. pp. 681~7.

2) Yoshiaki Funatsu : Estimation of the Difference Between Two Proportions in A Finite Multinomial Population. Journ. Japan. Statist. Soc. 9. 1. 1979. 29~35.