한국통계학회논문집 제4권 1호
1997년 4월 pp. 301-310

잭나이프 방법을 이용한 비추정[1)]

조 길 호[2)], 조 장 식,[3)] 김 상 룡[4)], 이 우 동[5)]

요 약

본 연구에서는 비(ratio)에 대한 2차잭나이프 추정량을 제안하고, 그것의 편의와 분산이 집단의 수에 대한 감소함수임을 보인다. 또한, 이 추정량의 우수성을 편의와 평균제곱오차의 측면에서 기존의 추정량과 비교 분석한다.

1. 서 론

Quenouille(1956)는 주어진 표본을 크기가 같은 여러 개의 집단으로 나누어 반복 사용함으로써 추정량의 편의(bias)를 감소시킬 수 있는 방법을 소개하였다. n개의 표본에서 얻은 모수 θ의 추정량을 $\hat{\theta}$이라 하고, n개의 표본을 각 각의 크기가 p인 g개의 표본집단으로 나눈 후에 j번째 집단을 제거한 p(g-1)개의 표본에서 얻은 $\hat{\theta}$과 같은 모양의 추정량을 $\hat{\theta}_j^*$라 할 때, Quenouille가 제안한 모수 θ의 통상잭나이프 추정량(ordinary jackknife estimator) $J(\hat{\theta})$은

$$J(\hat{\theta})=g\hat{\theta}-\frac{(g-1)}{g}\sum_{j=1}^{g}\hat{\theta}_j^* \tag{1}$$

이다. 만약 $\hat{\theta}$ 이 $cn^{-1}+O(n^{-2})$ 의 편의를 가진다면 $J(\hat{\theta})$의 편의는 $O(n^{-2})$ 이다.

또한 이 추정량은 Tukey(1958)에 의해 모수의 근사적 신뢰구간을 구하기 위한 방법에 이용되었다.

Schucany, Gray, 그리고 Owen(1971)은 $\hat{\theta}$의 편의가 $c_1n^{-1}+c_2n^{-2}+O(n^{-3})$ 이라면 편의를 n^{-2} 항 까지 제거할 수 있는 2차잭나이프 추정량(second-order jackknife estimator)을 제안하였다. 즉, $\hat{\theta}_{ij}^*$를 크기가 p인 i번째와 j번째의 표본집단을 동시에 제거한 나머지 p(g-2)개 표본에서 얻은 $\hat{\theta}$과 같은 모양의 추정량이라 할 때, 2차잭나이프 추정량 $J^{(2)}(\hat{\theta})$은

$$J^{(2)}(\hat{\theta})=\frac{1}{2}[g^2\hat{\theta}-2(g-1)^2(\frac{1}{g}\sum_{i=1}^{g}\hat{\theta}_i^*)+(g-2)^2(\frac{2}{g(g-2)}\sum_{i<j}\hat{\theta}_{ij}^*)] \tag{2}$$

---

1) 이 연구는 1995년도 경북대학교 공모과제 연구비 지원에 의한 결과임.
2) (702-701) 대구광역시 북구 산격동 경북대학교 통계학과 부교수
3) (608-736) 부산광역시 남구 대연동 경성대학교 전산통계학과 전임강사
4) (705-715) 대구광역시 남구 대명2동 대구교육대학교 수학교육과 전임강사
5) (712-240) 경북 경산시 점촌동 경산대학교 통계학과 조교수

로 정의된다. 2차잭나이프 추정 방법은 Gray와 Schucany(1972), Frangos(1980), Shao(1992), Shyamal과 Girish(1992) 등 많은 학자들에 의해 여러분야에 응용되었다.

이미 알고 있거나 대표본으로 부터 쉽게 추정할 수 있는 보조정보(X)를 이용한 비추정법(ratio estimation)은 표본 조사에서 유효한 추정을 하는데 중요한 역활을 한다. 예를 들어, 모평균 E(X)와 E(Y)를 갖는 모집단에서 n개의 확률표본 $(X_i,\ Y_i),\ i=1,\ 2,\ \ldots,\ n$이 주어질 때, E(X)는 이미 알거나 쉽게 추정 가능하고 다만 E(Y)에 관심이 있다면, 표본으로 부터 모집단의 비 $\rho=\dfrac{E(Y)}{E(X)}$를 $r=\dfrac{\overline{Y}}{\overline{X}}$로 추정해서 $\widehat{E(Y)}=r\,E(X)$ 의 방법으로 추정하는 것이 표본으로 부터 직접 E(Y)를 추정하는 것보다 더욱 더 바람직한 경우가 있다(Cochran(1977) 참조). 그러나 모집단의 비 ρ에 대한 일반적인 비추정량(customary ratio estimator) r은 편의추정량이므로 잭나이프 방법을 응용할 수 있는 한 분야이다.

비추정량에 잭나이프 방법을 처음으로 응용한 사람은 Durbin(1959)으로 그는 보조정보 X가 정규분포를 따르는 확률변수일 때

$$Y_i=\alpha+\beta X_i+e_i \tag{3}$$

의 모형에서, 집단의 수 g가 2일 때 통상잭나이프 추정량이 일반적인 비추정량 r보다 편의와 분산이 작다는 것을 보였다.

Rao(1965)는 위의 (3)식의 모형에서 g가 증가할 수록 통상잭나이프 추정량의 편의와 분산이 작아지므로 g=n일 때 최적의 통상잭나이프 추정량이 됨을 보였다.

또한 Durbin은 X가 감마분포를 따르는 확률변수인 (3)식의 모형에서 g=2일 때 통상잭나이프 추정량의 분산이 r의 분산보다 약간 큰 편이지만, 편의의 감소로 인해 평균제곱오차가 r의 평균제곱오차보다 작다는 것을 밝혔다. 그리고 Rao와 Webster(1966)는 g가 증가할 수록 통상잭나이프 추정량이 일반적인 비추정량 보다 작은 편의와 분산을 가짐을 보였다.

많은 실제 문제에서 보조확률변수 X는 양의 값을 가지는 경우가 대부분이기 때문에, 여기서는 보조확률변수 X가 감마분포를 따르는 경우에 한해서 살펴보기로 한다.

따라서 본 연구에서는 보조확률변수 X가 감마분포를 따르는 (3)식의 모형에서 비에 대한 2차잭나이프 추정량을 제안하고, 그것의 편의와 분산이 집단의 수에 대한 감소 함수임을 보인다. 또한 이 추정량의 우수성을 편의와 평균제곱오차의 측면에서 기존의 통상잭나이프 추정량 및 일반적인 비추정량과 서로 비교 분석하고자 한다.

2. 잭나이프 방법을 이용한 비추정

확률변수 (X, Y)에 대한 n개의 확률표본을 $(X_i,\ Y_i),\ i=1,\ 2,\ \ldots,\ n$이라 하고, 각 변수의 표본평균을 $\overline{X}$와 $\overline{Y}$라 하자. X_i와 Y_i는 다음과 같은 선형회귀 관계가 있다고 가정한다.

$$Y_i=\alpha+\beta X_i+e_i,\quad i=1,\ 2,\ \ldots,\ n. \tag{4}$$

여기서 $E(e_i|x_i)=0$, $Var(e_i|x_i)=n\delta$ 이며, δ는 $O(n^{-1})$ 인 일정한 양의 상수이다. 또한

$$\frac{X_i}{n} \sim G(h,\ 1)$$

이라 가정하면

$$\overline{X} = \sum_{i=1}^{n} \frac{X_i}{n} \sim G(m,\ 1)$$

이다. 여기서 m=nh 이다.

모집단의 비 $\rho = \dfrac{E(Y)}{E(X)}$에 대한 일반적인 비추정량 $r = \dfrac{\overline{Y}}{\overline{X}}$의 편의는

$$Bias(r) = \frac{a}{m(m-1)} \equiv B_1 a \tag{5}$$

이고, 분산은

$$Var(r) = \frac{a^2}{(m-1)^2(m-2)} + \frac{\delta}{(m-1)(m-2)} \equiv V_1 a^2 + V_2 \delta \tag{6}$$

이다 (Durbin(1959) 참조). 여기서

$$B_1 = \frac{1}{m(m-1)},\quad V_1 = \frac{1}{(m-1)^2(m-2)},\quad V_2 = \frac{1}{(m-1)(m-2)}.$$

크기가 p인 g개의 집단 중에서 j번째 집단을 제거하고 얻은 표본에서의 일반적인 비추정량을

$$r_j^* = \frac{\overline{Y_j^*}}{\overline{X_j^*}},\quad j=1,\ 2,\ \ldots,\ g \tag{7}$$

이라 하자. 여기서

$$\overline{Y_j^*} = \frac{g\overline{Y} - \overline{Y_j}}{g-1},\qquad \overline{Y_j} = \frac{1}{p}\sum_{i=1}^{p} Y_{ji}$$

$$\overline{X_j^*} = \frac{g\overline{X} - \overline{X_j}}{g-1},\qquad \overline{X_j} = \frac{1}{p}\sum_{i=1}^{p} X_{ji}$$

Y_{ji}는 Y의 관측값들 중 j번째 집단 속에 있는 i번째의 자료값

X_{ji}는 X의 관측값들 중 j번째 집단 속에 있는 i번째의 자료값이다.

따라서, 통상잭나이프 추정량 r_Q는

$$r_Q = gr - \frac{g-1}{g}\sum_{j=1}^{g} r_j^* \tag{8}$$

이다.

r_Q의 편의는

$$Bias(r_Q) = -\frac{a}{m(m-1)}\,\frac{g}{m(g-1)-g} \equiv B_2\ a \tag{9}$$

이다. 여기서

$$B_2 = \frac{1}{m(m-1)}\,\frac{g}{m(g-1)-g}.$$

(보조정리 1) (Rao와 Webster(1966))

서로 독립인 두개의 확률변수 X_1과 X_2가 $X_1 \sim G(a,\ 1)$, $X_2 \sim G(b,\ 1)$일 때,

$$E[\frac{1}{X_1(X_1+X_2)}] = \frac{1}{(a-1)(a+b-2)} \tag{10}$$

이다.

(보조정리 2) (Rao와 Webster(1966))

서로 독립인 세개의 확률변수 X_1, X_2, X_3가 $X_1 \sim G(a,\ 1)$, $X_2 \sim G(a,\ 1)$, $X_3 \sim G(b,\ 1)$ 일 때, (단, a와 b는 정수)

$$E[\frac{1}{(X_1+X_3)(X_2+X_3)}]$$

$$= \begin{cases} \frac{\Gamma(2a+b-2)}{\Gamma^2(a)\Gamma(b)}[\sum_{k=0}^{a-2}(-1)^k \frac{\Gamma^2(a-k-1)\Gamma(b+k)}{\Gamma(2a+b-k-2)} \\ +(-1)^{a-1}\{2\sum_{k=1}^{a-2}(-1)^{k+1}\frac{1}{(a+b-k-1)^2}+(-1)^{a+b}\frac{\pi^2}{6}\} \equiv \phi_1(a,b) & \text{if } a \ge 2 \\ 2\sum_{k=1}^{b-1}(-1)^{k+1}\frac{1}{(b-k)^2}+(-1)^{b+1}\frac{\pi^2}{6} \equiv \phi_2(b) & \text{if } a=1 \end{cases} \tag{11}$$

통상잭나이프 추정량 r_Q의 분산은

$$Var(r_Q) = E(r_Q-\beta)^2 - E^2(r_Q-\beta)$$

$$= \alpha^2 E[\frac{g}{X} - \frac{g-1}{g}\sum_{j=1}^{g}\frac{1}{X_j^*}]^2 + E[\frac{g\bar{e}}{X} - \frac{g-1}{g}\sum_{j=1}^{g}\frac{\overrightarrow{e_j^*}}{X_j^*}]^2 - E^2(r_Q-\beta) \tag{12}$$

$$\equiv V_3\alpha^2 + V_4\delta$$

으로 표현된다. 여기서

$$\bar{e} = \frac{\sum_{i=1}^{n} e_i}{n},\quad \overline{e_j^*} = \frac{1}{g-1}(g\bar{e} - \bar{e_j}),\quad \bar{e_j} = \frac{1}{p}\sum_{i=1}^{p} e_{ji},$$

e_{ji}는 j번째 집단에서 i번째 오차항,

V_3는 (12)식의 계산에서 α^2항의 계수,

V_4 는 (12)식의 계산에서 δ항의 계수이다.

또한,

$$E[\frac{g}{X} - \frac{g-1}{g}\sum_{j=1}^{g}\frac{1}{X_j^*}]^2 = g^2 E(\frac{1}{X^2}) - 2g(g-1)E(\frac{1}{X\,X_j^*})$$

$$+ \frac{(g-1)^2}{g}E(\frac{1}{X_j^*})^2 + \frac{(g-1)^3}{g}E(\frac{1}{X_i^*\ X_j^*}).$$

그리고,

$$E(\ \bar{e}^2 |\ x)=\delta,\quad E(\ \overrightarrow{e_j}^{*2} |\ x)=\frac{g}{g-1}\delta,\quad E(\ \overrightarrow{e_i}^{*}\ \overrightarrow{e_j}^{*} |\ x)=\frac{g(g-2)}{(g-1)^2}\delta,\quad E(\bar{e}\ \overrightarrow{e_j}^{*} |\ x)=\delta$$

이므로

$$E[\ \frac{g\bar{e}}{\bar{X}}-\frac{g-1}{g}\sum_{j=1}^{g}\frac{\overrightarrow{e_j}^{*}}{\overline{X_j}^{*}}\ ]^2$$

$$=\delta[g^2E(\frac{1}{\bar{X}^2})-2g(g-1)E(\frac{1}{\bar{X}\ \overline{X_j}^{*}})$$

$$+(g-1)E(\frac{1}{\overline{X_j}^{*}})^2+(g-1)(g-2)E(\frac{1}{\overline{X_i}^{*}\ \overline{X_j}^{*}})]$$

이다. (보조정리 1)과 (보조정리 2)를 이용하여 $Var(r_Q)$를 구할 수 있다.

Rao와 Webster는 $|Bias(r_Q)|$와 $Var(r_Q)$는 g의 감소함수임을 보였다. 따라서 $g=n$ 일 때 $|Bias(r_Q)|$와 $Var(r_Q)$는 최소가 됨을 알 수 있다.

크기가 p인 g개의 집단에서 i번째 집단과 j번째 집단을 동시에 제거하고 얻은 P(g-2)개의 표본에서 구한 일반적인 비추정량을

$$r_{ij}^{*}=\frac{\overline{Y_{ij}}^{*}}{\overline{X_{ij}}^{*}},\qquad i,\ j=1,\ 2,\ \ldots,\ g\ (i\neq j) \tag{13}$$

라 하자. 여기서

$$\overline{Y_{ij}}^{*}=\frac{g\bar{Y}-(\overline{Y_i}+\overline{Y_j})}{g-2},\quad \overline{X_{ij}}^{*}=\frac{g\bar{X}-(\overline{X_i}+\overline{X_j})}{g-2}$$

이다. 그리고

$$\widehat{\theta_1}=r, \tag{14}$$

$$\widehat{\theta_2}=\frac{1}{g}\sum_{i=1}^{g}r_i^{*}, \tag{15}$$

$$\widehat{\theta_3}=\frac{1}{\binom{g}{2}}\sum_{i<j}r_{ij}^{*} \tag{16}$$

라 두면, 2차잭나이프 추정량 r_S는 다음과 같이 주어진다.

$$r_S=\frac{1}{2}[g^2\widehat{\theta_1}+2(g-1)^2\widehat{\theta_2}+(g-2)^2\widehat{\theta_3}] \tag{17}$$

2차잭나이프 추정량 r_S의 편의와 분산은 다음과 같다.

(정리 1) r_S의 편의는

$$Bias(r_S)=\frac{\alpha}{m(m-1)}[\frac{g^2}{(m(g-1)-g)(m(g-2)-g)}]\equiv B_3\,\alpha \tag{18}$$

이다. 여기서

$$B_3 = \frac{1}{m(m-1)}\left[\frac{g^2}{(m(g-1)-g)(m(g-2)-g)}\right].$$

$$(\text{증명})\quad E\left(\frac{1}{\overline{X}}\right) = \frac{1}{m-1},\quad E\left(\frac{1}{\overline{X}_j^*}\right) = \frac{g-1}{m(g-1)-g}$$

이며,

$$\frac{g-2}{g}\ \overline{X}_{ij}^* \sim G\left(m\left(\frac{g-2}{g}\right),\ 1\right)$$

이므로

$$E\left(\frac{1}{\overline{X}_{ij}^*}\right) = \frac{g-2}{m(g-2)-g}$$

이다. 따라서 r_S의 기대값은

$$E(r_S) = \beta + \frac{\alpha}{m-1}\left[\frac{(g-1)(m-1)(m(g-2)-3g)+g^2m}{(m(g-1)-g)(m(g-2)-g)}\right]$$

이므로, r_S의 편의는

$$\begin{aligned} Bias(r_S) &= E(r_S) - \rho \\ &= \alpha\left[\frac{(g-1)(m-1)(m(g-2)-3g)+g^2m}{(m-1)(m(g-1)-g)(m(g-2)-g)} - \frac{1}{m}\right] \\ &= \frac{\alpha}{m(m-1)}\left[\frac{g^2}{(m(g-1)-g)(m(g-2)-g)}\right] \end{aligned}$$

이다.

<표 1>을 살펴보면 $Bias(r_S)$는 g의 감소함수이므로, $g=n$ 일 때 r_S는 가장 작은 편의를 가짐을 알 수 있다.

(보조정리 3) 세개의 확률변수 X_1, X_2, X_3가 서로 독립이고 $X_1 \sim G(a,\ 1)$, $X_2 \sim G(b,\ 1)$, $X_3 \sim G(c,\ 1)$일 때, (단, a, b, c는 정수)

$$E\left[\frac{1}{(X_1+X_3)(X_2+X_3)}\right]$$

$$= \begin{cases} \Gamma\dfrac{(a+b+c-2)}{\Gamma(a)\Gamma(b)\Gamma(C)}\Big[\displaystyle\sum_{k=0}^{a-2}(-1)^k\frac{\Gamma(a-k-1)\Gamma(b-k-1)\Gamma(c+k)}{\Gamma(a+b-c-k-2)} \\ \quad +(-1)^{b-1}\Big\{\displaystyle\sum_{k=1}^{a-b-1}(-1)^{k+1}\frac{\Gamma(a-b-k)\Gamma(b+c+k-1)}{(b+c+k-1)\Gamma(a+c-1)} \\ \quad +(-1)^{a-b}\Big(2\displaystyle\sum_{k=1}^{a+c-2}\frac{(-1)^{k+1}}{(a+c-k-1)^2}+(-1)^{a+c}\frac{\pi^2}{6}\Big)\Big\}\Big] \equiv \xi_1(a,b,c) & \text{if } a \geq b \geq 2 \\ 2\displaystyle\sum_{k=1}^{c-1}(-1)^{k+1}\frac{1}{(c-k)^2}+(-1)^{c+1}\frac{\pi^2}{6} \equiv \xi_2(c) & \text{if } a=b=1 \end{cases} \tag{19}$$

(정리 2) r_S의 분산은 다음과 같다.

$$Var(r_S) = V_5\ \alpha^2 + V_6\ \delta \tag{20}$$

여기서,

$$\begin{aligned}
V_5 &= \frac{g^4}{(m-1)(m-2)} + \frac{4(g-1)^6}{g\{m(g-1)-g)\}\{m(g-1)-2g\}} \\
&= \left(\begin{array}{ll} \frac{4(g-1)^7}{g^3}\phi_1(\frac{m}{g}, \frac{m(g-2)}{g}); & \frac{m}{g} \geq 2 \\ \frac{4(g-1)^7}{g^3}\phi_2(m-2); & m=g \end{array}\right) \\
&\quad + \frac{2(g-2)^6}{g(g-1)\{m(g-2)-g\}\{m(g-2)-2g\}} \\
&\quad + \left(\begin{array}{ll} \frac{4(g-2)^7}{g^3(g-1)}\phi_1(\frac{m}{g}, \frac{m(g-3)}{g}); & \frac{m}{g} \geq 2 \\ \frac{4(g-2)^7}{g^3(g-1)}\phi_2(m-3); & m=g \end{array}\right) \\
&\quad + \frac{(g+1)(g-2)^7}{g^3(g-1)}\phi_1(\frac{m}{g}, \frac{m(g-4)}{g}) \\
&\quad - \frac{4g^2(g-1)^3}{(m-2)\{g(m-1)-m\}} + \frac{2g^2(g-2)^3}{(m-2)\{m(g-2)-g\}} \\
&\quad - \frac{8(g-1)^3(g-2)^3}{g\{m(g-2)-g\}\{m(g-1)-2g\}} \\
&\quad - \left(\begin{array}{ll} \frac{4(g-1)^3(g-2)^4}{g^3}\xi_1(\frac{2m}{g}, \frac{m}{g}, \frac{m(g-3)}{g}); & \frac{m}{g} \geq 2 \\ \frac{4(g-1)^3(g-2)^4}{g^3}\xi_2(m-3); & m=g \end{array}\right) \\
&\quad -4\left\{\frac{(g-1)(m-1)(m(g-2)-3g)+g^2 m}{(m-1)(m(g-1)-g)(m(g-2)-g)}\right\}^2
\end{aligned} \tag{21}$$

$$\begin{aligned}
V_6 &= \frac{g^4}{(m-1)(m-2)} + \frac{4(g-1)^5}{\{m(g-1)-g)\}\{m(g-1)-2g\}} \\
&= \left(\begin{array}{ll} \frac{4(g-1)^5(g-2)}{g^2}\phi_1(\frac{m}{g}, \frac{m(g-2)}{g}); & \frac{m}{g} \geq 2 \\ \frac{4(g-1)^5(g-2)}{g^2}\phi_2(m-2); & m=g \end{array}\right) \\
&\quad + \frac{2(g-2)^5}{(g-1)\{m(g-2)-g\}\{m(g-2)-2g\}} \\
&\quad + \left(\begin{array}{ll} \frac{4(g-2)^5(g-3)}{g^2(g-1)}\phi_1(\frac{m}{g}, \frac{m(g-3)}{g}); & \frac{m}{g} \geq 2 \\ \frac{4(g-2)^5(g-3)}{g^2(g-1)}\phi_2(m-3); & m=g \end{array}\right) \\
&\quad + \frac{2(g-2)^5(g-3)(g-4)}{g^2(g-1)}\phi_1(\frac{2m}{g}, \frac{m(g-4)}{g}) \\
&\quad - \frac{4g^2(g-1)^3}{(m-2)\{g(m-1)-m\}} + \frac{2g^2(g-2)^3}{(m-2)\{m(g-2)-g\}}
\end{aligned} \tag{22}$$

$$
-\frac{8(g-1)^2(g-2)^3}{\{m(g-2)-g\}\{m(g-1)-2g\}}
$$
$$
-\begin{cases} \dfrac{4(g-1)^2(g-2)^3(g-3)}{g^2}\xi_1\left(\dfrac{2m}{g}, \dfrac{m}{g}, \dfrac{m(g-3)}{g}\right); & \dfrac{m}{g}\geq 2 \\ \dfrac{4(g-1)^2(g-2)^3(g-3)}{g^2}\xi_2(m-3); & m=g \end{cases}
$$

(증명) $Var(r_S)=E(r_S-\beta)^2-E^2(r_S-\beta)$

$$
=\frac{\alpha^2}{4}E\left[\frac{g^2}{\overline{X}}-2\frac{(g-2)^2}{g}\sum_{i=1}^{g}\frac{1}{X_i^*}+\frac{(g-2)^2}{\binom{g}{2}}\sum_{i<j}\frac{1}{X_{ij}^*}\right]^2
$$

$$
+\frac{1}{4}E\left[\frac{g^2}{\overline{X}}\bar{e}-\frac{2(g-1)^2}{g}\sum_{i=1}^{g}\frac{\overline{e_i^*}}{X_i^*}+\frac{(g-2)^2}{\binom{g}{2}}\sum_{i<j}\frac{\overline{e_{ij}^*}}{X_{ij}^*}\right]^2-E^2(r_S-\beta)
$$

(보조정리 1), (보조정리 2), (보조정리 3)을 이용하여 기대값을 구하면 증명될 수 있다.

$Var(r_s)$는 m과 g의 함수이며 <표 1>에서 알 수 있듯이 m에 관계없이 집단의 수가 많을수록 감소함을 알 수 있다. 따라서 2차잭나이프 추정량은 $g=n$ 일 때 (즉, 각 각의 표본을 하나의 집단으로 생각할 때), $Bias(r_S)$와 $Var(r_S)$는 최소가 됨을 알 수 있다.

3. 비교 분석

2차잭나이프 추정량(r_S)의 우수성을 통상잭나이프 추정량(r_Q)과 일반적인 비추정량(r)의 편의와 평균제곱오차의 측면에서 서로 비교 분석해 보고자 한다.

추정량들의 편의와 평균제곱오차에 대한 표현식이 너무 복잡해서 직접 비교하기가 어렵기 때문에 편의에 대한 계수 ($B_1\sim B_3$)와 평균제곱오차의 α^2항과 δ항의 계수 ($M_1\sim M_6$)를 <표 2>에 나타내었다. <표 2>로 부터 다음과 같은 결과를 얻을 수 있다.

(1) r_S는 r_Q 와 r보다 편의가 비교적 작으며, $g\geq 20$인 경우에는 거의 불편추정량이 된다.

(2) r_S의 평균제곱오차는 r_Q의 평균제곱오차 보다 비교적 작으며, r의 평균제곱오차 보다는 매우 작다. 특히, g가 적을 수록 그 차이가 커짐을 알 수 있다. 따라서 표본의 크기가 적은 경우 ($g\leq 20$)에 2차잭나이프 추정량 r_S를 비추정량으로 사용하는 것이 더욱 더 바람직하다.

(3) $g\geq 20$일 경우에는 r_S와 r_Q의 평균제곱오차는 거의 일치한다.

결론적으로 2차잭나이프 추정량 r_S는 통상잭나이프 추정량 r_Q와 일반적인 비추정량 r 보다 편의와 평균제곱오차의 측면에서 더욱 더 바람직한 비추정량임을 알 수 있으며, 표본의 크기가 적은 경우일 수록 더욱 더 좋은 추정량임을 알 수 있다.

<표 1> r_S의 편의와 분산

m	g	V_5	V_6	B_3
10	5	.00158	.01233	.00032
10	10	.00120	.01108	.00020
12	6	.00079	.00799	.00012
12	12	.00070	.00749	.00008
20	5	.00019	.00270	.00002
20	10	.00015	.00266	.00001
20	20	.00014	.00262	.00001
24	6	.00011	.00184	.00001
24	8	.00009	.00183	.00001
24	12	.00008	.00183	.00001
24	24	.00008	.00181	.00000
30	6	.00007	.00144	.00000
30	10	.00006	.00122	.00000
30	15	.00004	.00116	.00000
30	30	.00004	.00115	.00000

$Bias(r_s) = B_3\ \alpha \qquad Var(r_S) = V_5\ \alpha^2 + V_6\ \delta$

<표 2> $r,\ r_Q,\ r_S$ 의 편의와 평균제곱오차 ($g=n$ 일 때)

g	B_1	M_1	M_2	B_2	M_3	M_4	B_3	M_5	M_6
8	.01786	.00372	.02381	−0.00298	.00292	.01850	.00060	.00235	.01797
10	.01111	.00167	.03189	−0.00139	.00132	.01127	.00020	.00120	.01108
12	.00758	.00088	.00909	−0.00076	.00072	.00763	.00008	.00070	.00749
14	.00549	.00052	.00641	−0.00046	.00043	.00552	.00004	.00042	.00549
16	.00417	.00033	.00476	−0.00030	.00028	.00418	.00002	.00027	.00412
18	.00327	.00023	.00368	−0.00020	.00020	.00327	.00001	.00020	.00320
20	.00263	.00016	.00292	−0.00015	.00014	.00263	.00001	.00014	.00262
22	.00216	.00012	.00238	−0.00011	.00010	.00217	.00001	.00010	.00217
24	.00181	.00009	.00198	−0.00008	.00008	.00181	.00000	.00008	.00181
26	.00154	.00007	.00167	−0.00006	.00006	.00154	.00000	.00006	.00154
28	.00132	.00005	.00142	−0.00005	.00005	.00132	.00000	.00005	.00132
30	.00115	.00004	.00123	−0.00004	.00004	.00115	.00000	.00004	.00115
34	.00089	.00003	.00095	−0.00003	.00003	.00089	.00000	.00003	.00089
38	.00071	.00002	.00075	−0.00002	.00002	.00071	.00000	.00002	.00071
42	.00058	.00002	.00061	−0.00001	.00001	.00058	.00000	.00001	.00058
46	.00048	.00001	.00051	−0.00001	.00001	.00048	.00000	.00001	.00048
50	.00041	.00001	.00043	−0.00001	.00001	.00041	.00000	.00001	.00041

$Bias(r) = B_1\ \alpha \qquad Bias(r_Q) = B_2\ \alpha \qquad Bias(r_S) = B_3\ \alpha$

$MSE(r) = M_1\ \alpha^2 + M_2\ \delta \qquad MSE(r_Q) = M_3\ \alpha^2 + M_4\ \delta \qquad MSE(r_S) = M_5\ \alpha^2 + M_6\ \delta$

참 고 문 헌

[1] Cochran, W.G. (1977). *Sampling Technique*, John Wiley & Sons, New York.

[2] Durbin, J. (1959). A note on the application of Quenouille's method of bias reduction to the estimation of ratios, *Biometrika*, Vol. 46, 477-480.

[3] Frangos, C.C. (1980). Variance estimation for the second-order jackknife, *Biometrika*, Vol. 67, 715-718.

[4] Gray, H.L. and Schucany, W.R. (1972). *The generalized jackknife statistics,* Marcel Dekker, New York.

[5] Quenouille, M. (1956). Notes on bias in estimation, *Biometrika*, Vol. 43, 353-360.

[6] Rao, J.N.K. (1956). A note on estimation of ratios by Quenouille's method, *Biometrika*, Vol. 52, 647-649.

[7] Rao, J.N.K. and Webster, J.T. (1966). On two methods of bias reduction in the estimation of ratios, *Biometrika*, Vol. 53, 571-577.

[8] Schucany, W.R., Gray, H.L. and Owen, D.B. (1971). On bias reduction in estimation, *Journal of the American Statistical Association*, Vol. 66, 524-533.

[9] Shao, J. (1992). Jackknifing in generalized linear models, *Annals of the institute of Statistical Mathematics*, Vol. 44, 673-686.

[10] Shyamal, D.P. and Girish, P. (1992). Jackknife variance estimators in linear models, *Biometrika*, Vol. 79, 654-657.

[11] Turkey, J.W. (1958). Bias and confidence in not quite large samples, (abstract), *The Annals of Mathematical Statistics*, Vol. 29, 614.