ON APPLYING THE JACK-KNIFE PROCEDURE TO THE RATIO ESTIMATOR

Poduri S.R.S. Rao, University of Rochester

SUMMARY

The merits of a modified method of applying the Jack-knife procedure are evaluated through a model.

1. INTRODUCTION

Quenouille's (1951) method of bias reduction, popularly known as the Jack-knife procedure, has been successfully applied to increase the efficiency of estimators. Let $(\bar{X},\bar{Y})$ denote the population means of two characteristics (x,y) and $(\bar{x},\bar{y})$ denote the means of a random sample of size n. For estimating $R = (\bar{Y}/\bar{X})$, Durbin (1959) compared the classical estimator

$$\hat{R} = \frac{\bar{y}}{\bar{x}} \tag{1}$$

and the Jack-knife estimator

$$\hat{R}^* = g\,\frac{\bar{y}}{\bar{x}} - \frac{(g-1)}{g}\,\Sigma\,\frac{\bar{y}'_j}{\bar{x}'_j} \tag{2}$$

with g = 2 groups. In (2), $(\bar{x}'_j,\bar{y}'_j)$ are the means obtained by deleting the (n/g) observations of the jth group. Subsequently, for estimating the population mean $\bar{Y}$, in Rao (1969) and in Rao and Rao (1971), the corresponding estimators

$$t_1 = \hat{R}\bar{X} \tag{3}$$

$$= \bar{y} + \hat{R}(\bar{X}-\bar{x}) \tag{3a}$$

and

$$t_2 = \hat{R}^*\bar{X} \tag{4}$$

were considered. The estimator with the expression in (3a) is of the 'regression type' and it suggests the possibility of replacing $\hat{R}$ by other suitable estimators. In this paper, for $\bar{Y}$ we consider

$$t_3 = \bar{y} + \hat{R}^*(\bar{X}-\bar{x}). \tag{5}$$

The investigations in the above articles are based on the model

$$y_i = \alpha + \beta x_i + \varepsilon_i, \tag{6}$$

$(i = 1,\ldots,n)$, where ε_i has mean zero and variance δx^{ℓ} $(0\le\ell\le2)$ and is uncorrelated with ε_j. Further it is assumed that the size of the population is large and x has the Gamma distribution with parameter h. In Section 2, we present the biases and the Mean Square Errors (MSE's) for the estimators for the case of g = 2 groups. The results show that t_3 is more efficient than t_2.

Encouraged with the results for two groups, we compared the efficiencies for g = n groups; the biases and MSE's of the estimators for this general case are given in Sections 3 and 4. Summary of the investigation is given in Section 5. Two major conclusions are that in general t_3 is more efficient than t_2 and it is superior to t_1 for a wide range of the values of α and δ.

2. TWO GROUPS

2.1. Biases of the estimators

Writing the mean of x as $E(x) = \mu$ and that of y as $E(y) = \mu_y$, from the model in (6), the parameter that is being estimated is

$$\mu_y = \alpha + \beta\mu .$$

The biases of t_1 and t_2 are derived by Durbin (1959) and the author in Rao (1969) as

$$B_1 = \frac{1}{(u-1)}\,\alpha \tag{7}$$

and

$$B_2 = -\frac{2}{(u-1)(u-2)}\,\alpha, \tag{8}$$

where u = nh. From (5) and (6),

$$t_3 - \mu_y = \alpha\Big[\frac{2}{\bar{x}} - \frac{1}{2}\Big(\frac{1}{\bar{x}_1} + \frac{1}{\bar{x}_2}\Big)\Big]\,(\mu-\bar{x})$$
$$+ \bar{e} + \Big[2\frac{\bar{e}}{\bar{x}} - \frac{1}{2}\Big(\frac{\bar{e}_1}{\bar{x}_1} + \frac{\bar{e}_2}{\bar{x}_2}\Big)\Big](\mu-\bar{x}). \tag{9}$$

In (9), the subscripts 1 and 2 refer to the two groups and $\bar{e}$ is the mean of e_i for the entire sample. From (9), the bias in t_3 is

$$B_3 = E(t_3-\mu_y)$$
$$= \Big[2\Big(\frac{u}{u-1} - 1\Big) - \frac{u}{u-2} + \frac{1}{2}\Big(1 + \frac{u}{u-2}\Big)\Big]\,\alpha$$
$$= \frac{u-3}{(u-1)(u-2)}\,\alpha . \tag{10}$$

From (7), and (8) and (10), we make the following observations: $|B_2|$ and B_3 are smaller than B_1, and $|B_2|$ is smaller than B_3 unless u is five or less.

2.2. MSE's of the estimators

The MSE's M_1 and M_2 of t_1 and t_2 are derived by the author in Rao (1969) as

$$M_1 = \alpha^2\,\frac{u+2}{(u-1)(u-2)} + \delta\,\frac{u^2}{n(u'-1)(u'-2)}\,G \tag{11}$$

and

$$M_2 = \alpha^2\,\frac{u^3-5u^2+12u+16}{(u-1)(u-2)^2(u-4)}$$
$$+ \delta\,\frac{u^2(u^2+6u\ell-7u+9\ell^2-27\ell+18)}{n(u'-1)(u'-2)(u'+\ell-2)(u'+\ell-4)}\,G, \tag{12}$$

where $u' = (u+\ell)$ and $G = \Gamma(h+t)/\Gamma h$. Durbin derived (11) and (12) when ℓ is equal to zero.

From (9) the MSE of t_3 is

$$M_3 = \alpha^2 E\Big[\frac{2}{\bar{x}} - \frac{1}{2}\Big(\frac{1}{\bar{x}_1} + \frac{1}{\bar{x}_2}\Big)\Big]^2(\mu-\bar{x})^2$$

$$+ E\left\{\bar{e}^2 + \left[2\frac{\bar{e}}{\bar{x}} - \frac{1}{2}\left(\frac{\bar{e}_1}{\bar{x}_1} + \frac{\bar{e}_2}{\bar{x}_2}\right)\right]^2 (\mu-\bar{x})^2\right.$$

$$\left. + 2\bar{e}\left[2\frac{\bar{e}}{\bar{x}} - \frac{1}{2}\left(\frac{\bar{e}_1}{\bar{x}_1} + \frac{\bar{e}_2}{\bar{x}_2}\right)\right](\mu-\bar{x})\right\}$$

$$= \alpha^2\left[\frac{4(u+2)}{(u-1)(u-2)} + \frac{u^2-u-6}{(u-2)^2(u-4)} - \frac{4(u+2)}{(u-2)^2}\right]$$

$$+ \frac{\delta}{n}\left\{1 + [u+(\ell-1)(\ell-2)]\left[\frac{5}{(u'-1)(u'-2)}\right.\right.$$

$$\left.\left. + \frac{4}{(u'+\ell-2)(u'-2)}\right] + 2(\ell-1)\left(\frac{2}{u'-1} - \frac{1}{u'+\ell-2}\right)\right\} G$$

$$= \alpha^2 \frac{u^3-6u^2+3u+38}{(u-1)(u-2)^2(u-4)}$$

$$+ \delta G[u^4+2(2\ell-3)u^3+(2\ell^2-6\ell+5)u^2$$

$$-2(\ell-1)(\ell-2)(2\ell-5)u+(u-1)^2(u-2)^2] \,/$$

$$n(u'-1)(u'-2)(u'+\ell-2)(u'+\ell-4). \qquad (13)$$

From (11) and (12), as was given by the author in Rao (1969),

$$M_1 - M_2 = \alpha^2 \frac{u(u-16)}{(u-1)(u-2)^2(u-4)}$$

$$+ \delta \frac{u^3(1-2\ell)-5u^2(\ell-1)(\ell-2)}{n(u'-1)(u'-2)(u'+\ell-2)(u'+\ell-4)} \qquad (14)$$

From (11) and (13),

$$M_1 - M_3 = \alpha^2 \frac{(u+2)(2u-11)}{2(u-1)(u-2)^2(u-14)}$$

$$+ \delta G[(2\ell^2-6\ell+3)u^2+2(\ell-1)(\ell-2)(2\ell-5)u$$

$$-(u-1)^2(u-2)^2] \,/$$

$$n(u'-1)(u'-2)(u'+\ell-2)(u'+\ell-4) \qquad (15)$$

and from (12) and (13)

$$M_2 - M_3 = \alpha^2 \frac{(u+11)}{(u-1)(u-2)(u-4)}$$

$$+ \delta G[(2\ell-1)u^3+(7\ell^2-21\ell+13)u^2$$

$$+2(\ell-1)(\ell-2)(2\ell-5)u-(\ell-1)^2(\ -2)^2] \,/$$

$$n(u'-1)(u'-2)(u'+\ell-2)(u'+\ell-4) \qquad (16)$$

From (14) -- (16), we draw the following conclusions:

(i) As Durbin pointed out, for the case of $\ell = 0$, t_2 is more efficient than t_1 when $u > 16$, that is, the coefficient of variation of $\bar{x}$ (CV) is less than 25 percent. For the same case, t_3 is more efficient than t_1 if $u > 7$, that is, the CV is less than 40 percent

(ii) For the case of $0 < \ell < (\frac{1}{2})$, t_1 is more efficient than t_2 if the CV is less than 25 percent, but t_3 is more efficient than t_1 if the CV is less than 30 percent $(u > 13)$.

(iii) When $\alpha = 0$ and $\ell = 1$ or 2, t_2 and t_3 have larger MSE's than t_1. However, when $\alpha \neq 0$ and $\ell = 1$, t_2 is more efficient than t_1 if

$$C^2 = \frac{(\delta/n)}{\alpha^2} < \frac{u-16}{(u-2)(u-4)}$$

and t_3 is superior to t_1 if

$$C^2 < \frac{(u+2)(2u-11)}{2(u-2)(u-4)} .$$

Similar limits for C^2 can be found for the other values of ℓ .

(iv) The estimator t_3 is superior to t_2 when $\alpha \neq 0$ and ℓ lies between zero and two.

3. BIASES AND MSE'S WHEN g = n

The biases and MSE's of t_1 and t_2 are derived in Rao and Rao (1971). The procedure of deriving them is given in Rao and Webster (1966) and by the author in Rao (1974). Here we present the biases and MSE's of the three estimators with some detail.

3.1. Biases of the estimators

Let r, r_j, s and s_j denote $(1/\bar{x})$, $(1/\bar{x}_j')$, $(\bar{e}/\bar{x})$ and $(\bar{e}'_j/\bar{x}'_j)$ where $\bar{x}'_j$ and $\bar{e}'_j$ are the means of the $k = (n-1)$ observations. From (3) -- (6),

$$t_1 - \mu_y = \alpha(r\mu-1) + s\mu, \qquad (17)$$

$$t_2 - \mu_y = \alpha\left\{[nr-(n-1)\bar{r}]\mu-1\right\} + [ns-(n-1)\bar{s}]\mu, \qquad (18)$$

and

$$t_3 - \mu_y = \alpha[nr-(n-1)\bar{r}](\mu-\bar{x})$$

$$+ \bar{e} + [ns-(n-1)\bar{s}](\mu-\bar{x}), \qquad (19)$$

where $n\bar{r} = \Sigma r_j$ and $n\bar{s} = \Sigma s_j$.

Denoting nh and (n-1)h by u and v, we find that the expectations of r, r_j and $\bar{x}r_j$ are

$$a_1 = \frac{n}{u-1} , \qquad (20)$$

$$a_2 = \frac{k}{v-1} , \qquad (21)$$

and

$$a_3 = \frac{k(u-1)}{n(v-1)} . \qquad (22)$$

From (17) -- (22), the biases of t_1, t_2 and t_3 can be written as

$$B_1 = \frac{1}{u-1}\alpha , \quad (23)$$

$$B_2 = -\frac{1}{(u-1)(v-1)}\alpha , \quad (24)$$

$$B_3 = \frac{nv-2n+1}{n(u-1)(v-1)}\alpha . \quad (25)$$

We notice that $|B_2|$ and B_3 are smaller than B_1, and $|B_2|$ is smaller than B_3.

3.2 MSE's of the estimators

For finding the MSE's from (17) - (19), here we give the expectations of the different terms; details of the derivations are available with us. Let I(a,b,c) denote the expectations of $(X_1+X_3)^{-1}(X_2+X_3)^{-1}$, where X_1, X_2 and X_3 are independent Gamma variates with parameters a, b and c. The expectations of r^2, r_j^2, $r_j r_k$ and rr_j are

$$a_4 = \frac{n^2}{(u-1)(u-2)} , \quad (26)$$

$$a_5 = \frac{k^2}{(v-1)(v-2)} \quad (27)$$

$$a_6 = n^2 I[h,h,(n-2)h] \quad (28)$$

and

$$a_7 = \frac{nk}{(u-2)(v-1)} . \quad (29)$$

Similarly, the averages of $(\bar{x}r_j)^2$, $\bar{x}^2 r_j r_k$, $\bar{x}r_j^2$ and $\bar{x}r_j r_k$ are

$$a_8 = \left(\frac{k}{n}\right)^2 \frac{(u-1)(u-2)}{(v-1)(v-2)} \quad (30)$$

$$a_9 = \left(\frac{k}{n}\right)^2 \left\{\frac{n}{k} + \frac{h}{v-1} + h(h+1)I[h,h+2,(n-2)h]\right\} \quad (31)$$

$$a_{10} = \frac{k^2}{n} \frac{(u-2)}{(v-1)(v-2)} \quad (32)$$

and

$$a_{11} = \frac{k^2}{n}\left\{\frac{1}{v-1} + hI[h,h+1,(n-2)h]\right\} \quad (33)$$

Denote $(u+\ell)$ by u' and $(v+\ell)$ by v'. Averages of the expressions involving s and s_j are as follows. All the terms should be multiplied by δG, where $G = \Gamma(h+t)/\Gamma h$, as defined earlier. The expectations of s^2, s_j^2, $s_j s_k$ and ss_k are

$$d_1 = \frac{n}{(u'-1)(u'-2)} , \quad (34)$$

$$d_2 = \frac{k}{(v'-1)(v'-2)} , \quad (35)$$

$$d_3 = (n-2)I[h,h,(n-2)h+\ell] \quad (36)$$

and

$$d_4 = \frac{k}{(v'-1)(u'-2)} \quad (37)$$

Similarly, the averages of $\bar{e}^2$, $\bar{x}^2 s_j^2$, $\bar{x}^2 s_j s_k$ and $\bar{x}\,\bar{e}\,s_j$ are

$$d_5 = \frac{1}{n} , \quad (38)$$

$$d_6 = \frac{k(u'-1)(u'-2)}{n^2(v'-1)(v'-2)} \quad (39)$$

$$d_7 = \frac{(n-2)}{n^2}\left\{1 + \frac{h}{v'-1} + \frac{h}{v'} + h(h+1)I[h,h+2,(n-2)h+\ell]\right\} \quad (40)$$

and

$$d_8 = \frac{k(u'-1)}{n^2(v'-1)} . \quad (41)$$

The averages of $\bar{x}s^2$, $\bar{x}s_j^2$, $\bar{x}s_j s_k$ and $\bar{e}s_j$ are

$$d_9 = \frac{1}{(u'-1)} , \quad (42)$$

$$d_{10} = \frac{k}{n}\frac{(u'-2)}{(v'-1)(v'-2)} , \quad (43)$$

$$d_{11} = \frac{(n-2)}{n}\left\{\frac{1}{v'-1} + hI[h,h+1,(n-2)h+\ell]\right\} \quad (44)$$

and

$$d_{12} = \frac{k}{n}\frac{1}{(v'-1)} . \quad (45)$$

From (17)--(22) and (26)--(45), the MSE's of t_1, t_2 and t_3 can be expressed as follows:

$$M_1 = MSE(t_1) = \alpha^2 A_1 + \delta D_1 , \quad (46)$$

$$M_2 = MSE(t_2) = \alpha^2 A_2 + \delta D_2 , \quad (47)$$

and

$$M_3 = MSE(t_3) = \alpha^2 A_3 + \delta D_3 , \quad (48)$$

where

$$A_1 = h^2 a_4 + 1 - 2ha_1 ,$$

$$D_1 = h^2 d_1 ;$$

$$A_2 = (n^2 a_4 + \frac{k^2}{n} a_5 + \frac{k^3}{n} a_6 - 2nka_7)h^2 + 1 - 2(na_1 + ka_2)h ,$$

$$D_2 = (n^2 d_1 + \frac{k^2}{n} d_2 + \frac{k^3}{n} d_3 - 2nkd_4)h^2 ;$$

$$A_3 = (n^2 a_4 + \frac{k^2}{n} a_5 + \frac{k^3}{n} a_6 - 2nka_7)h^2 ,$$

and

$$D_3 = d_5 + [n^2 d_1 + \frac{k^2}{n} d_2 + \frac{k^3}{n} d_3 - 2nkd_4]h^2$$

$$+\left[n^2 d_5 + \frac{k^2}{n} d_6 + \frac{k^3}{n} d_7 - 2nkd_8\right]$$

$$-2\left[n^2 d_9 + \frac{k^2}{n} d_{10} + \frac{k^3}{n} d_{11} - 2nkd_{12}\right]h$$

$$+ 2\left[nd_9 - kd_{12}\right]h - 2\left[nd_5 - kd_8\right].$$

4. RELATIVE EFFICIENCIES

For values of n ranging from 5 to 50 and h from 1 to 4, we have computed the MSE's derived in the previous section on CDC 6600 with double precision. We present them in Table 1 for some values of n and h.

The three MSE's can be expressed as

$$M_i = (A_i/nc^2 + D_i)\delta \ , \qquad (49)$$

where $c^2 = (\delta/n\alpha^2)$ as defined earlier. We note that c is the coefficient of variation of y in the model

$$y_i = \alpha + e_i \qquad (50)$$

with $E(\varepsilon_i) = 0$ and $V(\varepsilon_i) = \delta$. In practical situations it may be possible to have some knowledge of c. We computed the MSE's in (49) for c ranging from (¼) to 2. The following conclusions can be drawn from our investigation.

(i) When $\alpha = 0$ and $\ell = 1$ or 2, the classical estimator is more efficient than t_2 and t_3; for these cases, the difference between the MSE's of t_1 and t_3 is negligible.

(ii) When $\alpha = 0$ and $\ell = 0$, t_2 is more efficient than t_3 which in turn is more efficient than t_1.

(iii) The result in (ii) for $\ell = 0$ holds even when $\alpha \neq 0$ for c smaller than 2.

(iv) When $\alpha \neq 0$ and $\ell = 1$ or 2, t_3 is more efficient than t_1 and t_2 when c is smaller than 2. For these cases t_2 may not be more efficient than t_1.

ACKNOWLEDGEMENTS

The author would like to thank Dr. Barbara A. Bailar for her interest in this investigation. This research is partially supported by Grant No. 77-27 from the Bureau of the Census, U.S. Department of Commerce. Thanks to Mrs. Nancy Hufsmith for her skillful and meticulous typing of the manuscript.

REFERENCES

DURBIN, J. (1959). A note on the application of Quenouille's method of bias reduction to the estimation of ratios. Biometrika 46, 477-80.

QUENOUILLE, M.H. (1956). Notes on bias in estimation. Biometrika 43, 353-60.

RAO, J.N.K. & WEBSTER, J.T. (1966). On two methods of bias reduction in the estimation of ratios. Biometrika 53, 571-7.

RAO, J.N.K. & RAO, P.S.R.S. (1971). Small sample results for ratio estimators. Biometrika 58, 625-30.

RAO, P.S.R.S. (1969). Comparison of four ratio-type estimates under a model. J. Am. Statist. Ass. 64, 574-80.

RAO, P.S.R.S. (1974). Jack-knifing the ratio estimator. Sankhya, Series C, 36, 84-97.

TABLE 1. MSE's of t_1, t_2 and t_3 when $g = n$; original values multiplied by 1000. Coefficients of α^2 are given below the values for the coefficient of δ when $\ell = 0$.

		$\ell = 0$			$\ell = 1$			$\ell = 2$		
		h = 1	h = 2	h = 3	h = 1	h = 2	h = 3	h = 1	h = 2	h = 3
n = 10	D_1	138.89	116.96	110.84	111.11	210.53	310.35	181.82	571.43	1161.29
	A_1	166.67	64.33	39.41						
	D_2	112.71	105.81	103.80	114.08	212.46	311.98	220.89	632.58	1243.87
	A_2	132.39	56.50	36.06						
	D_3	133.25	115.79	110.35	111.41	210.62	310.40	182.45	571.78	1161.55
	A_3	132.55	57.77	36.74						
n = 15	D_1	82.42	73.89	71.35	71.43	137.93	204.55	125.00	387.10	782.61
	A_1	93.41	39.41	24.84						
	D_2	71.65	69.09	68.27	72.20	138.46	205.00	142.60	414.15	818.94
	A_2	78.35	35.94	23.34						
	D_3	60.92	73.56	71.21	71.48	137.95	204.56	125.14	387.17	782.66
	A_3	80.65	36.74	23.73						
n = 20	D_1	58.48	53.98	52.60	52.63	102.56	152.54	95.24	292.68	590.16
	A_1	64.33	28.34	18.12						
	D_2	52.69	51.33	50.88	52.94	102.78	152.73	105.18	307.85	610.50
	A_2	56.04	26.40	17.28						
	D_3	57.88	53.85	52.54	52.65	102.57	152.55	95.28	292.71	590.18
	A_3	57.76	26.91	17.51						
n = 25	D_1	45.29	42.52	41.65	41.67	81.63	121.62	76.92	235.29	473.68
	A_1	48.91	22.11	14.25						
	D_2	41.69	40.84	40.56	41.82	81.74	121.71	83.30	244.98	486.65
	A_2	43.70	20.87	13.72						
	D_3	45.00	42.45	41.62	41.67	81.64	121.62	76.94	235.30	473.69
	A_3	44.93	21.22	13.87						
n = 30	D_1	36.95	35.07	34.47	34.48	67.80	101.12	64.52	196.72	395.60
	A_1	39.41	18.12	11.75						
	D_2	34.49	33.91	33.72	34.57	67.86	101.18	68.95	203.44	404.59
	A_2	35.83	17.26	11.37						
	D_3	36.78	35.03	34.46	34.49	67.80	101.12	64.53	196.73	395.61
	A_3	36.74	17.51	11.48						