Biometrika (1980), **67**, 3, *pp.* 715–8
Printed in Great Britain

Variance estimation for the second-order jackknife

BY CHRISTOS C. FRANGOS

*Department of Statistical and Mathematical Sciences,
London School of Economics and Political Science*

SUMMARY

This paper considers Sharot's family of jackknives (1976) and chooses a member of it using a minimum variance based criterion. The criterion is different from that considered by Sharot, which is claimed to be inappropriate. Results on the variance estimator of the jackknife are derived and compared with those of Hinkley (1978). A Monte Carlo simulation study supports the above conclusion.

Some key words: Correlation; Jackknife; Influence function; Variance reduction.

1. INTRODUCTION

The jackknife is a well-known method for bias reduction and robust interval estimation. An excellent review is given by Miller (1974). Let $x_1, \ldots, x_n$ be independent and identically distributed random variables with cumulative distribution function $F(x, \theta)$. Let $T_n = T_n(x_1, \ldots, x_n)$ be an estimator of θ based on n observations. Define T_{-i} and T_{-i-j} to be estimators of θ based on all but the ith observation and on all but the ith and jth observation respectively. If we write

$$\bar{T}_{.} = n^{-1} \Sigma T_{-i}, \quad \bar{T}_{..} = 2\{n(n-1)\}^{-1} \Sigma_{i<j} T_{-i-j} \quad (i<j),$$

$$g_i = nT_n - (n-1) T_{-i} \quad (i = 1, \ldots, n),$$

then the first- and second-order jackknives are

$$\hat{\theta}^{(1)} = n^{-1} \Sigma g_i, \quad \hat{\theta}^{(2)} = \tfrac{1}{2}\{n^2 T_n - 2(n-1)^2 \bar{T}_{.} + (n-2)^2 \bar{T}_{..}\},$$

respectively. Sharot (1976) suggested the family of jackknives

$$J(p) = 2\{n(n-1)\}^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} g_{ij}(p) \tag{1·1}$$

with variance estimator of $J(p)$

$$S_J^2(p) = 4\{n(n-1)(n-2)\}^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} \{g_{ij}(p) - J(p)\}^2, \tag{1·2}$$

where for $i<j$

$$g_{ij}(p) = pnT_n + \tfrac{1}{2}(1-2p)(n-1)(T_{-i} + T_{-j}) + (p-1)(n-2) T_{-i-j}. \tag{1·3}$$

Sharot suggested the choice of $J(p^*)$, where p^* is the value of p minimizing $S_J^2(p)$. In § 2 of this paper we show, extending work of Hinkley (1978), that the $g_{ij}(p)$ are substantially correlated and therefore $S_J^2(p)$ is inappropriate as a variance estimator of $J(p)$. In § 3 we derive asymptotically unbiased estimators of the variance of first- and second-order influence functions to find the variance of $g_{ij}(p)$, In §§ 4 and 5 we propose a rule for choosing a member from Sharot's family and we investigate the above results by a simulation study.

2. Derivation of the covariance matrix of the $g_{ij}(p)$

Suppose that T_n is a regular differentiable functional of the form $T_n = t(\hat{F}_n)$, where $\hat{F}_n$ is the empirical distribution function. Then $\theta = T(F)$ and, following Hinkley (1978), we have that

$$T_n \sim \theta + n^{-1} \sum_{j=1}^{n} f_1(x_j) + \tfrac{1}{2} n^{-2} \sum_{j=1}^{n} \sum_{l=1}^{n} f_2(x_j, x_l), \qquad (2\cdot1)$$

where $f_1(x)$, $f_2(x,y)$ are the first- and second-order influence functions of T (Hampel, 1974), defined by the identities

$$\left[\frac{d}{d\varepsilon} T\{(1-\varepsilon)F + \varepsilon F_1\}\right]_{\varepsilon=0} = \int f_1(x)\, dF_1(x),$$

$$\left[\frac{d^2}{d\varepsilon^2} T\{(1-\varepsilon)F + \varepsilon F_1\}\right]_{\varepsilon=0} = \iint f_2(x,y)\, dF_1(x)\, dF_1(y), \qquad (2\cdot2)$$

where F_1 is an arbitrary cumulative distribution function. We write

$$(i) = f_1(x_i) \quad (i = 1, \ldots, n), \quad (i,j) = f_2(x_i, x_j) \quad (i,j = 1, \ldots, n).$$

A detailed analysis of $g_{ij}(p)$ using all terms given in (2·1) gives

$$g_{ij}(p) \sim \theta + \tfrac{1}{2}(i) + \tfrac{1}{2}(j) + T_1(i,j), \qquad (2\cdot3)$$

where

$$T_1(i,j) = a_1\left[2n^{-1} \sum_{l=1}^{n} \left\{\sum_{m=1}^{n} (m,l) - n(l,i) - n(l,j)\right\} + (i,i) + (j,j)\right] + a_2(i,j)$$

and $a_1 = (2p-n)/\{4(n-1)(n-2)\}$, $a_2 = (p-1)/(n-2)$. We define $\sigma_{11} = \text{var}\{(i)\}$, $\sigma_{22} = \text{var}\{(i,l)\}$ $(i \neq l)$ and we choose $p = kn$, where k is a constant, independent of n. This means that asymptotically we are working with a family of jackknives more like a second-order jackknife, because $\hat{\theta}^{(1)} = J(1)$, $\hat{\theta}^{(2)} = J(\tfrac{1}{2}n)$. After some algebra we find from (2·3), omitting terms of $o(1/n)$, that

$$\text{var}\{g_{ij}(p)\} = \tfrac{1}{2}\sigma_{11} + k^2\sigma_{22} \quad (i \neq j), \quad \text{cov}\{g_{ij}(p), g_{il}(p)\} = \tfrac{1}{4}\sigma_{11} \quad (i<j,\ i<l,\ l \neq j),$$
$$\text{cov}\{g_{ij}(p), g_{li}(p)\} = \tfrac{1}{4}\sigma_{11} \quad (i<j,\ l<i,\ l \neq j), \qquad (2\cdot4)$$
$$\text{cov}\{g_{ij}(p), g_{lm}(p)\} = 0 \quad (i<j;\ l<m;\ l \neq i,j;\ m \neq i,j).$$

Expressions (2·4) reveal that asymptotically there is nonzero correlation between some of the $g_{ij}(p)$. Using (2·4) and the definitions of $J(p)$ and $S_J^2(p)$ we find that

$$\text{var}\{J(p)\} = n^{-1}(\sigma_{11} + 2k^2 n^{-1}\sigma_{22}), \quad E\{S_J^2(p)\} \sim n^{-1}(\sigma_{11} + 2k^2\sigma_{22}), \qquad (2\cdot5)$$

so that Sharot's estimator is positively biased asymptotically, when p is chosen to give something like a second-order jackknife. In his paper Sharot suggests choosing p to minimize $S_J^2(p)$. One can see from the results presented here that the effect of this procedure is to move p away from values close to the second-order jackknife towards those similar to the first-order jackknife. This is due to the positive bias of $S_J^2(p)$ for large p, and its lack of bias for $p = 1$. The choice of p by this criterion is not based on any valuable property of $J(p)$, but simply gets rid of p values for which $S_J^2(p)$ is badly biased. Only if σ_{22} is very small compared with σ_{11} would $S_J^2(p)$ be a good estimator of $\text{var}\{J(p)\}$ for p of order n.

3. New estimators of σ_{11} and σ_{22} unbiased to the first order

Hinkley (1978) estimates σ_{11}, σ_{22} by the corresponding sample moments of the estimators of (i) and (i,j) for $i \neq j$. Using (2·1) we find that his estimator of σ_{22} is biased to $o(1/n)$

because

$$\operatorname{var}\{(i,j)^*\} \sim \sigma_{22} - 4n^{-1}\sigma_{22}, \quad \operatorname{cov}\{(i,i-1)^*,(i-1,i-2)^*\} \sim -2n^{-1}\sigma_{22},$$

where $(i,j)^*$ is the estimator of (i,j). Further estimators of σ_{11}, σ_{22} can be found by choosing that quadratic form in $g_{ij}(p)$ which treats the $g_{ij}(p)$ symmetrically and is unbiased to the first order. Using (2·4) we find the estimators

$$\hat{\sigma}_{11} = 4\{n(n-1)(n-2)\}^{-1}S_2 - 16\{n(n-1)(n-2)(n-3)\}^{-1}S_3,$$

$$\hat{\sigma}_{22} = 2\{k^2n(n-1)\}^{-1}S_1 - 2\{k^2n(n-1)(n-2)\}^{-1}S_2 + 4\{k^2n(n-1)(n-2)(n-3)\}^{-1}S_3, \quad (3·1)$$

where

$$S_2 = \Sigma_{i<j<l}\{g_{ij}(p)\,g_{il}(p) + g_{il}(p)\,g_{jl}(p) + g_{ij}(p)\,g_{jl}(p)\}, \quad S_1 = \Sigma_{i<j}\,g_{ij}^2(p),$$

$$S_3 = \Sigma_{i<j<k<l}\{g_{ij}(p)\,g_{kl}(p) + g_{ik}(p)\,g_{jl}(p) + g_{jk}(p)\,g_{il}(p)\}.$$

To the first order $E(\hat{\sigma}_{11}) = \sigma_{11}$, $E(\hat{\sigma}_{22}) = \sigma_{22}$. The estimators $\hat{\sigma}_{11}$ and $\hat{\sigma}_{22}$ are, to the first order, minimum norm quadratic unbiased estimators, MINQUE, of σ_{11} and σ_{22} (Rao, 1972).

4. Reduction of variance estimator in a new jackknife

One might think, as did Sharot, of choosing k to minimize $\hat{\sigma}_{11}$ and so consequently $\operatorname{var}\{J(p)\}$ in (2·5), to the first order. Another possible choice of p is to minimize the variance of $\hat{\sigma}_{11}$. If we assume that the $g_{ij}(p)$ are asymptotically multivariate normal and use (2·4), the best choice of k is zero to the first order. In the simulations of § 5 only the case $k = 0$ is considered since the computation of p to minimize $\hat{\sigma}_{11}$ is complicated and the results in the case $k = 0$ are clearly better.

5. Monte Carlo simulation studies

We consider estimation from the negative exponential distribution with probability density function $f(x,\theta) = \theta e^{-\theta x}$ and cumulative distribution function $F(x,\theta)$. We estimate θ by $\bar{x}^{-1} = \{\int x\,d\hat{F}_n(x)\}^{-1}$. Then, $\theta = T(F) = \{\int x\,dF(x)\}^{-1}$. For the functional $\theta = T(F)$ we find from (2·2)

$$f_1(x) = \theta(1-\theta x), \quad f_2(x,y) = 2\theta(1-\theta x)(1-\theta y), \quad \sigma_{11} = E_F(f_1^2) = \theta^2, \quad \sigma_{22} = E_F(f_2^2) = 4\theta^2. \quad (5·1)$$

For each of n = 12, 24, 30, 40, 50, 60, 10,000 samples are generated from the negative exponential distribution with mean $\mu_x = 1$, on the CDC7600 of London University. Our aim is to compare the performances of the following two members of Sharot's family $J(p)$ of estimators: $\hat{J}^{(1)} = J(p)$ with $p = 0$ and $J(p^*)$ with p^* chosen to minimize $S_J^2(p)$. We choose as a 'yardstick' the minimum variance unbiased estimator: $\tilde{\theta} = (n-1)n^{-1}\bar{x}^{-1}$. The results are summarized in Tables 1 and 2.

Table 1. *Estimation of $\theta = 1$ in negative exponential distribution*

	Performance of $\hat{J}^{(1)}$, $J(p^*)$				Variance estimator $S_J^2(p)$				
n	Estimator	Mean	Rel. var.	Rel. MSE	σ^2	σ_J^2	$E\{S_J^2(p)\}$	St. dev. of $S_J^2(p)$	$[\text{MSE}\{S_J^2(p)\}]^{\frac{1}{2}}$
12	$\hat{J}^{(1)}$	0·989	1·017	1·019	0·084	0·103	0·117	0·181	0·182
	$J(p^*)$	0·987	1·022	1·024	0·174	0·104	0·147	0·205	0·209
24	$\hat{J}^{(1)}$	1·002	1·003	1·004	0·042	0·047	0·048	0·004	0·004
	$J(p^*)$	1·002	1·004	1·005	0·065	0·048	0·055	0·040	0·041
40	$\hat{J}^{(1)}$	1·000	1·001	1·001	0·025	0·028	0·029	0·001	0·001
	$J(p^*)$	1·000	1·001	1·001	0·033	0·028	0·030	0·016	0·016
60	$\hat{J}^{(1)}$	1·000	1·000	1·000	0·017	0·017	0·018	0·007	0·007
	$J(p^*)$	1·000	1·000	1·000	0·020	0·018	0·019	0·008	0·008

Table 2. 95% *large sample confidence intervals for the estimators* $\hat{J}^{(1)}$, $J(p^*)$ *and their variances*

n	Estimator	Means of estimators	$E\{S_J^2(p)\}$
12	$\hat{J}^{(1)}$	(0·979, 0·992)	(0·144, 0·120)
	$J(p^*)$	(0·982, 0·995)	(0·142, 0·150)
24	$\hat{J}^{(1)}$	(0·996, 1·004)	(0·047, 0·049)
	$J(p^*)$	(0·996, 1·005)	(0·053, 0·055)
40	$\hat{J}^{(1)}$	(0·997, 1·003)	(0·028, 0·030)
	$J(p^*)$	(0·997, 1·004)	(0·029, 0·030)
60	$\hat{J}^{(1)}$	(0·997, 1·002)	(0·017, 0·019)
	$J(p^*)$	(0·997, 1·002)	(0·018, 0·019)

From Table 1 we find that the estimator $\hat{J}^{(1)}$ is superior to $J(p^*)$ with respect to bias, variance and mean squared error. The column σ^2 represents the true population variance to $o(n^{-1})$ for each of the estimators and is $n^{-1}(\sigma_{11}+2k^2n^{-1}\sigma_{22}) = n^{-1}(\sigma_{11}+2p^2n^{-3}\sigma_{22})$, where we find from (5·1) that $\sigma_{11} = 1$, $\sigma_{22} = 4$; the sampling variance for each of the estimators over 10,000 samples is σ_J^2. We may compare σ^2 and σ_J^2 with the corresponding estimators of the expected value of the variance estimators, labelled $E\{S_J^2(p)\}$ in the table. The variance estimators are given from (1·2) for the estimator $J(p^*)$ and from (2·5) for $\hat{J}^{(1)}$, with (3·1) used for calculation of σ_{11} and σ_{22}. The conclusions from Table 1 are that, for every value of n considered, the variance estimator of $J^{(1)}$ is less biased and has less standard deviation and mean squared error than the variance estimator of $J(p^*)$. Hence, the estimator $J^{(1)} = J(p)$ with $p = 0$ is more efficient than Sharot's estimator $J(p^*)$.

Table 2 contains 95% large sample confidence intervals for the mean and the variance of the estimators $\hat{J}^{(1)}$ and $J(p^*)$.

Additional research is needed on the relation between Hinkley's and MINQUE estimators of σ_{11} and σ_{22} as n increases.

The author would like to thank Dr M. Knott for his very valuable help throughout the preparation of this paper. Helpful comments from a referee were also gratefully received.

References

HAMPEL, F. R. (1974). The influence curve and its role in robust interval estimation. *J. Am. Statist. Assoc.* **69**, 383–93.

HINKLEY, D. V. (1978). Improving the jackknife with special reference to correlation estimation. *Biometrika* **65**, 13–21.

MILLER, R. G. (1974). The jackknife: A review. *Biometrika* **61**, 1–15.

RAO, C. R. (1972). Estimation of variance and covariance components in linear models. *J. Am. Statist. Assoc.* **67**, 112–5.

SHAROT, T. (1976). Sharpening the jackknife. *Biometrika* **63**, 315–21.

[*Received October* 1979. *Revised April* 1980]