Biometrika (1976), 63, 2, p. 315–21
Printed in Great Britain

Sharpening the jackknife

BY T. SHAROT
Audits of Great Britain, Eastcote, Middlesex

SUMMARY

The jackknife is now well known as a widely applicable bias-reduction tool, with the added advantages of Tukey's variance estimator, and in certain applications, a gain in precision. In this paper a new family of jackknives is introduced, together with a variance estimator. The family includes the ordinary, first-order, and second-order jackknives as special cases and the variance estimator includes Tukey's as a special case. A sample-based decision rule for choosing a member of the family, intended to achieve a gain in precision over either order of jackknife, is given. The success of the rule is examined by simulation in estimation of the reciprocal of the parent mean and in ratio estimation.

Some key words: Bias reduction; Jackknife; Simulation; Variance estimator; Variance reduction.

1. INTRODUCTION

A brief account of the jackknife technique follows: readers are referred to Miller (1974) for a review of theoretical developments, and to Bissell & Ferguson (1975) for practical aspects.

Let $Y_1, \ldots, Y_n$ be a sample of independent and identically distributed random variables. Let t be an estimator of a parameter θ based on this sample. Let t_i be the corresponding estimator based on the sample of size $n-1$ obtained by omitting Y_i. Define

$$g_i = nt - (n-1)\,t_i$$

to be the so-called pseudovalues. The first-order jackknife is their average:

$$\hat{\theta}^{(1)} = \Sigma g_i/n = nt - (n-1)\,\bar{t}_i.$$

If

$$E(t) = \theta + a_1/n + a_2/n^2 + O(1/n^3) \tag{1·1}$$

then

$$E(\hat{\theta}^{(1)}) = \theta + O(1/n^2). \tag{1·2}$$

Tukey's variance estimator for var $(\hat{\theta}^{(1)})$ is based on the variance of the pseudovalues:

$$s_T^2 = \frac{1}{n(n-1)}\Sigma(g_i - \hat{\theta}^{(1)})^2.$$

If we write $d_i = (n-1)\,(t_i - \bar{t}_i)$, then $s_T^2 = \Sigma d_i^2/\{n(n-1)\}$.

To obtain the second-order jackknife, let t_{ij} be the estimator based on all but Y_i and Y_j $(i \neq j)$. Define

$$g_{ij} = \tfrac{1}{2}\{n^2t - (n-1)^2\,(t_i + t_j) + (n-2)^2\,t_{ij}\}$$

to be the pseudovalues. Then their average

$$\hat{\theta}^{(2)} = \frac{2}{n(n-1)}\sum_{i<j} g_{ij} = \tfrac{1}{2}\{n^2t - 2(n-1)^2\,\bar{t}_{.} + (n-2)^2\,\bar{t}_{..}\},$$

say, is the second-order jackknife, and under (1·1)

$$E(\hat{\theta}^{(2)}) = \theta + O(1/n^3).$$

We shall also need the quantities $d_{ij} = (n-2)(t_{ij} - \bar{t}_{..})$.

2. A FAMILY OF JACKKNIVES

We propose a new family of jackknives consisting of linear combinations of t, $\hat{\theta}^{(1)}$ and $\hat{\theta}^{(2)}$ but removing bias of order $1/n$ only. The unique family with these properties is easily verified to be of the form

$$J(p) = pnt + (1-2p)(n-1)\bar{t}_{.} + (p-1)(n-2)\bar{t}_{..}.$$

In particular, $J(1) = \hat{\theta}^{(1)}$ and $J(\frac{1}{2}n) = \hat{\theta}^{(2)}$. Because $J(p)$ is linear in p, its variance $\sigma_J^2(p)$ is quadratic in p and, except for trivial cases such as $t = \bar{Y}$, possesses a unique minimum at p_0, an unknown value dependent on the parent distribution and the sampling distribution of t. Since only exceptionally will p_0 be 1 or $\frac{1}{2}n$, $J(p_0)$ is in general a sharper jackknife in terms of variance than either $\hat{\theta}^{(1)}$ or $\hat{\theta}^{(2)}$.

In the absence of knowledge of p_0, the natural way to proceed is to substitute for it the value p^* which minimizes some variance estimator for $J(p)$. Such an estimator is now derived.

3. A VARIANCE ESTIMATOR FOR $J(p)$

We may write

$$J(p) = \frac{2}{n(n-1)} \Sigma g_{ij}(p),$$

where throughout this section sums are for $i < j$ and where

$$g_{ij}(p) = pnt + \tfrac{1}{2}(1-2p)(n-1)(t_i + t_j) + (p-1)(n-2)t_{ij}$$

are the pseudovalues. Following Tukey, let

$$\begin{aligned} s_J^2(p) &= c\Sigma\{g_{ij}(p) - J(p)\}^2 \\ &= c\Sigma[p(d_{ij} - d_i - d_j) - \{d_{ij} - \tfrac{1}{2}(d_i + d_j)\}]^2. \end{aligned}$$

It may be verified that $c = 4\{n(n-1)(n-2)\}^{-1}$ yields $s_J^2(1) = s_T^2$, that is, Tukey's estimator for var $(\hat{\theta}^{(1)})$. Writing

$$u_{ij} = d_{ij} - d_i - d_j, \quad v_{ij} = d_{ij} - \tfrac{1}{2}(d_i + d_j),$$

we have

$$s_J^2(p) = \frac{4}{n(n-1)(n-2)} \Sigma(pu_{ij} - v_{ij})^2, \tag{3·1}$$

so that the minimizing value p^* is given by

$$p^* = (\Sigma u_{ij} v_{ij})/(\Sigma u_{ij}^2). \tag{3·2}$$

The proposed jackknife is $J(p^*)$ with variance estimator $s_J^2(p^*)$. The expression

$$s_J^2(p^*) = \frac{4}{n(n-1)(n-2)} \frac{(\Sigma u_{ij}^2)(\Sigma v_{ij}^2) - (\Sigma u_{ij} v_{ij})^2}{\Sigma u_{ij}^2}$$

is obtained by substituting (3·2) into (3·1).

Note that, although

$$E\{J(p)\} = \theta + O(1/n^2)$$

for any fixed p, this is not necessarily true for variable p. For

$$J(p) = \bar{a}(y) + p(y)\,\bar{b}(y),$$

where y represents the data and $a(y)+p(y)\,b(y)$ is the form of each pseudovalue $g_{ij}(p)$. Now, under (1·1)

$$E(\bar{a}) = \theta + O(1/n^3), \quad E(\bar{b}) = O(1/n^2),$$

so that

$$E(J) = \theta + O(1/n^2) + \text{cov}\,\{p(y), \bar{b}(y)\}.$$

The last term is naturally zero for any fixed p. For variable p, its value will depend on the decision rule for choosing p. It may be verified that $p(y) = p^*$ implies that

$$p(y) = -\{\Sigma(a-\bar{a})\,(b-\bar{b})\}/\{\Sigma(b-\bar{b})^2\}.$$

It seems intuitively acceptable that this expression should be independent of $\bar{b}$ under weak assumptions. This has not been proved, though it is upheld by the following Monte Carlo studies.

4. Monte Carlo studies

The performance of the new jackknife was investigated by Monte Carlo simulation in two specific applications.

A standard estimator for $\theta = 1/\mu_Y$ $(\mu_Y \neq 0)$ is $t = 1/\bar{Y}$. Attention will be restricted to nondegenerate distributions over $Y \geqslant 0$; then t is biased upwards. Let $\bar{Y} = \mu_Y + \delta$. Then, by expansion in powers of δ, $E(t) = 1/\mu_Y + \sigma^2/(n\mu_Y^3) + \ldots$, so that $\hat{\theta}_c = 1/\bar{Y} - s_c^2/(n\bar{Y}^3)$, where

$$s_c^2 = \Sigma(Y_i - \bar{Y})^2/(n-1)$$

is approximately unbiased. Very similar is Jaeckel's infinitesimal jackknife J_∞ (Miller, 1974), which in this application is given by

$$J_\infty = 1/\bar{Y} - s_u^2/(n\bar{Y}^3),$$

where

$$s_u^2 = \Sigma(Y_i - \bar{Y})^2/n.$$

The model used for comparing the estimators was $f(Y) = \theta e^{-\theta Y}$. If this is the true density, then the estimator $\tilde{\theta} = (n-1)/(n\bar{Y})$ is minimum variance unbiased and was included in the study as a yardstick for the other estimators.

10 000 samples were generated for each of $n = 6$, 12 and 24; without loss of generality, $\theta = 1$. The results are summarized in Table 1, and the conclusions are as follows.

(*a*) The bias in t is substantially removed by all the other estimators.

(*b*) The variance of t is also greater than that of the other estimators, excepting $\hat{\theta}^{(2)}$ when $n = 6$.

(*c*) Apart from $\tilde{\theta}$, $\hat{\theta}^{(2)}$ is least biased, as expected.

(*d*) The bias and variance of $J(p^*)$ are less than those of $\hat{\theta}^{(1)}$.

(*e*) Those estimators which are based on knowledge of the parent density or of the properties of t, that is $\tilde{\theta}$ and $\hat{\theta}_c$, respectively, outperform the jackknives.

The variance estimators for $\hat{\theta}^{(1)}$, $\hat{\theta}^{(2)}$ and $J(p^*)$ given in Table 2 show the following.

(a) The variance estimator $s_J^2(p)$ is biased upwards for all p and n. Given p, the bias decreases with increasing n; given n, the bias is considerable for $p = \frac{1}{2}n$. That is, the quadratic $s_J^2(p)$ lies above the quadratic $\sigma_J^2(p)$ in expectation and the two diverge for large p. For $p = 1$ and $p = p^*$ the bias is not too bad and being positive it is therefore conservative.

(b) The standard deviations of the variance estimators are even larger than their biases. Alternative variance estimators may therefore be more attractive, where available.

Table 1. *Estimation of* $1/\mu_Y$ *: performance of estimators*

	$n = 6$			$n = 12$			$n = 24$		
	Mean	Rel. var.	Rel. MSE	Mean	Rel. var.	Rel. MSE	Mean	Rel. var.	Rel. MSE
$\hat{\theta}$	0·999	1·000	1·000	1·004	1·000	1·000	0·999	1·000	1·000
t	1·199	1·440	1·603	1·096	1·190	1·280	1·043	1·089	1·129
$\hat{\theta}_c$	1·029	1·121	1·125	1·011	1·035	1·036	1·001	1·008	1·008
J_∞	1·057	1·162	1·175	1·018	1·046	1·049	0·997	1·011	1·011
$\hat{\theta}^{(1)}$	0·951	1·312	1·322	0·955	1·035	1·035	0·997	1·003	1·004
$\hat{\theta}^{(2)}$	1·021	1·961	1·963	1·005	1·062	1·062	0·999	1·008	1·008
$J(p^*)$	0·972	1·261	1·264	0·996	1·035	1·035	0·997	1·004	1·004

Table 2. *Estimation of* $1/\mu_Y$ *: variance estimator* $s_J^2(p)$

n	p	$\sigma_J^2(p)$	$E\{s_J^2(p)\}$	Stand. dev. of $s_J^2(p)$	RMSE $\{s_J^2(p)\}$
6	1	0·317	0·610	1·939	1·961
	3	0·474	3·420	23·656	23·839
	p^*	0·305	0·552	1·749	1·766
12	1	0·105	0·149	0·201	0·206
	6	0·108	0·532	1·126	1·191
	p^*	0·105	0·147	0·196	0·200
24	1	0·0454	0·0544	0·0402	0·0412
	12	0·0456	0·1744	0·1794	0·2208
	p^*	0·0454	0·0543	0·0399	0·0409

In fact, $\sigma_J^2(p)$ was evaluated for $p = -2(0{\cdot}05)3$. For $n = 6$, the minimum falls at $p_0 = 0{\cdot}65$, though no standard error is available for this figure. For comparison, the mean and standard deviation of p^* are 1·13 and 0·58 respectively. Thus p^* is not an unbiased estimator of p_0, though its use still gives a gain in precision.

For $n = 6$, $\sigma_J^2(p^*) = 0{\cdot}305$, $\sigma_J^2(p_0) = 0{\cdot}314$ and $\sigma_J^2(1) = 0{\cdot}317$. As in Table 1, the relative sizes of these values are relatively error-free, since the same 10000 samples were used for all of them. Thus $J(p^*)$ is sharper than $J(p_0)$, itself sharper than $\hat{\theta}^{(1)}$.

For $n = 12$ and 24 the quadratic is very nearly flat and its minimum lies some undetermined distance below $p = -2$. The flattening necessarily occurs with increasing n in situations such as this where all the members of $J(p)$ are asymptotically efficient. Nevertheless, p^* does not follow p_0 below -2; for $n = 12$ its mean and standard deviation are 1·23 and 0·36, and for $n = 24$, 1·26 and 0·24. This is curious, though $J(p^*)$ still 'works'. It is not known whether as $n \to \infty$ the asymptotic limit of $E(p^*)$ has any significance, or even if it necessarily exists.

As a second application, consider ratio estimation, where the mean μ_Y is estimated by

$r\mu_X = \mu_X(\bar{Y}/\bar{X})$, in situations in which a concomitant variable X with known mean μ_X is available. Several alternatives to r as an estimator of $\theta = \mu_Y/\mu_X$ are available. Hutchison (1971) did a Monte Carlo study of five of these, due respectively to Hartley & Ross, Mickey, Quenouille $\{\hat{\theta}^{(1)}\}$, Beale and Tin. On the basis of Hutchison's results, the last three estimators are generally superior to the first two.

We have therefore repeated Hutchison's study with the estimators $t = r$, 'Beale', 'Tin', $\hat{\theta}^{(1)}$, $\hat{\theta}^{(2)}$, J_∞ and $J(p^*)$. The appropriate formulae for infinite populations are for Beale's and Tin's estimators respectively:

$$r\left(1+\frac{s_{xy}}{n\bar{X}\bar{Y}}\right)\Big/\left(1+\frac{s_{xx}}{n\bar{X}^2}\right),\quad r\left\{1+\frac{1}{n}\left(\frac{s_{xy}}{\bar{X}\bar{Y}}-\frac{s_{xx}}{\bar{X}^2}\right)\right\},$$

where

$$s_{xx} = \frac{1}{n-1}\Sigma(X-\bar{X})^2,\quad s_{xy} = \frac{1}{n-1}\Sigma(X-\bar{X})(Y-\bar{Y}),$$

and

$$J_\infty = r\left\{1+\frac{n-1}{n^2}\left(\frac{s_{xy}}{\bar{X}\bar{Y}}-\frac{s_{xx}}{\bar{X}^2}\right)\right\}.$$

The following models were used for the study.

Model 1. Here $Y = \beta X + \epsilon$, where β is a constant, $\log X$ is distributed as $N(\mu, \sigma^2)$ and ϵ is distributed as $N(0, kX^\gamma)$, where k and γ are constants. All the estimators, including r, are unbiased for this model. The variances of the estimators relative to var (r) are independent of μ, β and k; we used the values 0, 1, 1 respectively. For the other parameters the same values as Hutchison's were chosen and 1000 samples drawn for each combination; a subset of the results is presented in Table 3. The conclusions are as follows.

Table 3. *Ratio estimation for model* 1: *variance of estimators relative to* var (r)

		$n = 4$			$n = 8$		
σ		$\gamma = 0{\cdot}0$	$\gamma = 1{\cdot}0$	$\gamma = 2{\cdot}0$	$\gamma = 0{\cdot}0$	$\gamma = 1{\cdot}0$	$\gamma = 2{\cdot}0$
0·5	'Beale'	0·922	1·030	1·097	0·944	1·004	1·070
	'Tin'	0·918	1·035	1·107	0·943	1·005	1·074
	$\hat{\theta}^{(1)}$	0·912	1·052	1·140	0·937	1·007	1·088
	$\hat{\theta}^{(2)}$	0·926	1·073	1·172	0·937	1·010	1·098
	J_∞	0·934	1·022	1·077	0·949	1·004	1·064
	$J(p^*)$	0·926	1·025	1·085	0·939	1·006	1·077
1·0	'Beale'	0·846	1·028	1·190	0·845	1·008	1·178
	'Tin'	0·829	1·044	1·253	0·828	1·010	1·219
	$\hat{\theta}^{(1)}$	0·871	1·182	1·610	0·790	1·031	1·431
	$\hat{\theta}^{(2)}$	1·188	1·406	1·996	0·816	1·063	1·635
	J_∞	0·859	1·025	1·182	0·846	1·007	1·189
	$J(p^*)$	0·869	1·084	1·227	0·794	1·022	1·288

(*a*) When r has least variance, the least loss of precision is exhibited mainly by J_∞, occasionally by Beale's estimator. In this situation, $J(p^*)$ always reduces the loss of precision of $\hat{\theta}^{(1)}$.

(*b*) When r does not have least variance, the greatest gain in precision is usually made by $\hat{\theta}^{(1)}$, as in Hutchison's study.

(*d*) The highest variance is for $\hat{\theta}^{(2)}$. In bad conditions, i.e. for small n, large γ and large σ, $J(p^*)$ reduces the loss of precision of $\hat{\theta}^{(1)}$ and $\hat{\theta}^{(2)}$ by factors of up to 3 and 5 respectively.

Model 2. In this model $X = ZX_p$, where Z is Poisson with mean μ (= 10, 5, 2·5) and X_p is a perturbing variable centred on unity, distributed as $\chi^2_{(m)}/m$ ($m = \infty, 20, 10$). Given X, Y is distributed as $\chi^2_{(Z)}$; $Z = 0$ is not observed.

When m is infinite ($X_p = 1$) all the estimators are unbiased and r has minimum variance. The variances relative to var (r), based on 1000 samples, are given in Table 4; J_∞ is generally the best alternative to r here and $J(p^*)$ performs badly for $n = 4$.

Table 4. *Ratio estimation for model 2, $m = \infty$: variance of estimators relative to* var (r)

	$n = 4$			$n = 6$		
	$\mu = 10{\cdot}0$	$\mu = 5{\cdot}0$	$\mu = 2{\cdot}5$	$\mu = 10{\cdot}0$	$\mu = 5{\cdot}0$	$\mu = 2{\cdot}5$
'Beale'	0·997	1·008	1·020	1·000	1·009	1·008
'Tin'	0·997	1·011	1·025	1·000	1·010	1·008
$\theta^{(1)}$	0·997	1·018	1·038	1·001	1·011	1·010
$\theta^{(2)}$	0·994	1·031	1·050	1·000	1·009	1·011
J_∞	0·997	1·005	1·014	1·000	1·008	1·006
$J(p^*)$	1·014	1·037	1·391	1·000	1·010	1·008

Table 5. *Ratio estimation for model 2, with finite m: mean squared error of estimators relative to mean squared error of r*

			$h = 1$			$h = 30$		
m	n		$\mu = 10{\cdot}0$	$\mu = 5{\cdot}0$	$\mu = 2{\cdot}5$	$\mu = 10{\cdot}0$	$\mu = 5{\cdot}0$	$\mu = 2{\cdot}5$
20	4	'Beale'	0·955	0·982	0·995	0·875	0·971	0·910
		'Tin'	0·955	0·982	0·998	0·880	0·982	0·905
		$\theta^{(1)}$	0·957	0·990	1·018	0·902	1·023	0·905
		$\theta^{(2)}$	0·961	1·012	1·060	0·912	1·042	0·925
		J_∞	0·960	0·982	0·992	0·874	0·960	0·916
		$J(p^*)$	0·953	0·992	0·969	0·883	0·976	0·907
20	8	'Beale'	0·984	0·988	0·996	0·930	0·930	0·971
		'Tin'	0·984	0·988	0·996	0·932	0·931	0·972
		$\theta^{(1)}$	0·985	0·988	0·997	0·937	0·927	0·979
		$\theta^{(2)}$	0·986	0·990	0·998	0·938	0·927	0·984
		J_∞	0·985	0·988	0·995	0·929	0·928	0·970
		$J(p^*)$	0·985	0·988	0·996	0·935	0·929	0·977
10	4	'Beale'	0·925	0·926	0·955	0·547	0·636	0·680
		'Tin'	0·924	0·923	0·957	0·530	0·619	0·654
		$\theta^{(1)}$	0·939	0·931	0·997	0·518	0·625	0·634
		$\theta^{(2)}$	0·964	0·965	1·088	0·538	0·657	0·693
		J_∞	0·934	0·934	0·959	0·590	0·667	0·707
		$J(p^*)$	0·934	0·936	0·983	0·551	0·636	0·650
10	8	'Beale'	0·958	0·968	0·984	0·740	0·705	0·744
		'Tin'	0·957	0·968	0·985	0·742	0·698	0·734
		$\theta^{(1)}$	0·959	0·971	0·991	0·752	0·688	0·711
		$\theta^{(2)}$	0·962	0·976	0·995	0·752	0·693	0·714
		J_∞	0·961	0·970	0·986	0·743	0·716	0·755
		$J(p^*)$	0·959	0·970	0·989	0·749	0·690	0·716

When m is finite, r is biased, increasingly so as m decreases. Following Hutchison, it is assumed that stratification with h strata is performed. If the within-stratum bias and variance are constant across strata, then an overall measure of mean squared error is $\{h^2(\text{bias})^2 + h\,\text{variance}\}$. This quantity, relative to that for r and based on 1000 samples, is presented in Table 5; $J(p^*)$ shows gains over $\hat{\theta}^{(1)}$ when stratification is absent, and occasionally when it is present. Overall, however, there is little difference between any of the estimators, except that $\hat{\theta}^{(2)}$ is once again worst.

5. Conclusions

On the basis of the Monte Carlo studies, it would appear that the desired gain in precision of $J(p^*)$ over $\hat{\theta}^{(1)}$ is often achieved. Alternative estimators designed for a particular application may, not surprisingly, do better still. The infinitesimal jackknife is seen to yield just such an estimator in many cases.

Two points seem relevant to such comparisons. First, the infinitesimal jackknife is not a true jackknife in the sense that it is not based on reapplications of the original estimator t to subsamples of the data and thus its applicability may not be as wide as for the other jackknives. Secondly, only one *ad hoc* rule for choosing p has been suggested, based on a not altogether satisfactory variance estimator. An alternative rule, based on a different variance estimator, or on theoretical grounds, would be very attractive. Finally, $\hat{\theta}^{(2)}$ may be theoretically less biased than other estimators; see Sharot (1976) for a comparison with $\hat{\theta}^{(1)}$. It is, however, a comparatively 'blunt' jackknife and $J(p^*)$ offers a very real improvement for little more computational effort.

The comments of a referee are gratefully acknowledged. The paper was written at University of Aberdeen.

References

Bissell, A. F. & Ferguson, R. A. (1975). The jackknife–toy, tool or two-edged weapon? *The Statistician* **24**, 79–100.

Hutchison, M. C. (1971). A Monte Carlo comparison of some ratio estimators. *Biometrika* **58**, 313–21.

Miller, R. G. (1974). The jackknife – a review. *Biometrika* **61**, 1–15.

Sharot, T. (1976). The generalized jackknife – finite samples and subsample sizes. *J. Am. Statist. Assoc.* **71**. To appear.

[*Received October* 1975. *Revised January* 1976]