



## The Jackknife--A Review

Rupert G. Miller

*Biometrika*, Volume 61, Issue 1 (Apr., 1974), 1-15.

Stable URL:

<http://links.jstor.org/sici?&sici=0006-3444%28197404%2961%3A1%3C1%3ATJR%3E2.0.CO%3B2-4>

---

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

*Biometrika* is published by Biometrika Trust. Please contact the publisher for further permissions regarding the use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/bio.html>.

---

*Biometrika*  
©1974 Biometrika Trust

JSTOR and the JSTOR logo are trademarks of JSTOR, and are Registered in the U.S. Patent and Trademark Office. For more information on JSTOR contact [jstor-info@umich.edu](mailto:jstor-info@umich.edu).

©2003 JSTOR

## The jackknife – a review

By RUPERT G. MILLER

Department of Statistics, Stanford University, California

### SUMMARY

Research on the jackknife technique since its introduction by Quenouille and Tukey is reviewed. Both its role in bias reduction and in robust interval estimation are treated. Some speculations and suggestions about future research are made. The bibliography attempts to include all published work on jackknife methodology.

*Some key words:* Bias reduction; Interval estimation; Jackknife; Pseudo-value; Robustness.

### 1. INTRODUCTION

Quenouille (1949) introduced a technique for reducing the bias of a serial correlation estimator based on splitting the sample into two half-samples. In his 1956 paper he generalized this idea into splitting the sample into  $g$  groups of size  $h$  each,  $n = gh$ , and explored its general applicability.

Let  $Y_1, \dots, Y_n$  be a sample of independent and identically distributed random variables. Let  $\hat{\theta}$  be an estimator of the parameter  $\theta$  based on the sample of size  $n$ . Let  $\hat{\theta}_{-i}$  be the corresponding estimator based on the sample of size  $(g-1)h$ , where the  $i$ th group of size  $h$  has been deleted. Define

$$\hat{\theta}_i = g\hat{\theta} - (g-1)\hat{\theta}_{-i} \quad (i = 1, \dots, g). \quad (1.1)$$

The estimator

$$\hat{\theta} = \frac{1}{g} \sum_{i=1}^g \hat{\theta}_i = g\hat{\theta} - (g-1) \frac{1}{g} \sum_{i=1}^g \hat{\theta}_{-i} \quad (1.2)$$

has the property that it eliminates the order  $1/n$  term from a bias of the form

$$E(\hat{\theta}) = \theta + a_1/n + O(1/n^2).$$

In an abstract Tukey (1958) proposed that the  $g$  values (1.1) could be treated as approximately independent and identically distributed random variables in many situations. The statistic

$$\frac{\sqrt{g}(\hat{\theta} - \theta)}{\left\{ \frac{1}{g-1} \sum_{i=1}^g (\hat{\theta}_i - \hat{\theta})^2 \right\}^{1/2}} \quad (1.3)$$

should then have an approximate  $t$  distribution with  $g-1$  degrees of freedom and constitute a pivotal statistic for robust interval estimation. In unpublished work Tukey subsequently called the  $g$  values (1.1) pseudo-values and created the name jackknifed estimator for (1.2) in the hope that it would be a rough-and-ready statistical tool.

The research which has substantiated, amplified, and built upon this original work is surveyed in this article. Since there are the two different aspects of the jackknife technique, namely, bias reduction and interval estimation, the survey is divided along these lines into

separate sections, §§2 and 3, respectively. Section 4 describes some additional developments, and §5 contains some speculations and suggestions about future research on the jackknife. The bibliography attempts to include all published work on jackknife methodology.

Although it will be necessary to mention the case of general  $g$  and  $h$  at certain points, the presentation in this article centres on  $g = n$  and  $h = 1$ . Much of the research on the jackknife has been devoted to this special case. It is the most appealing because it eliminates any arbitrariness in the formation of the groups, and it is probably the best form of the jackknife to use in any problem. For large data bases, however, this may be computationally not feasible. In most if not all instances any result proved for the case  $h = 1$  can be extended to  $h > 1$ .

## 2. BIAS REDUCTION

### 2.1. Second-order jackknife

In 1956 Quenouille also gave a way to eliminate the order  $1/n^2$  term from a bias by jackknifing with weights  $n^2$  the jackknifed estimator. The second-order jackknife estimator is

$$\tilde{\theta}^{(2)} = \frac{n^2\tilde{\theta} - (n-1)^2\sum\tilde{\theta}_{-j}/n}{n^2 - (n-1)^2}, \quad (2.1)$$

where  $\tilde{\theta}_{-j}$  is (1.2) applied to the sample of size  $n-1$  with the  $j$ th observation removed. In terms of the original estimator  $\theta$  the second-order jackknife is expressible as

$$\tilde{\theta}^{(2)} = (n-1)^{-1} \left[ n^3\theta - (2n^2 - 2n + 1)(n-1) \left( \frac{1}{n} \sum \theta_{-i} \right) + (n-1)^2(n-2) \left\{ \frac{2}{n(n-1)} \sum_{i < j} \theta_{-ij} \right\} \right], \quad (2.2)$$

where  $\theta_{-ij}$  denotes the original estimator applied to the sample of size  $n-2$  with the  $i$ th and  $j$ th observations removed.

If  $E(\theta) = \theta + a_1/n + a_2/n^2$ , then  $E(\tilde{\theta}^{(2)}) = \theta + O(1/n^3)$ , but  $\tilde{\theta}^{(2)}$  is not unbiased. Schucany, Gray & Owen (1971) suggested modifying the weights to achieve complete unbiasedness when the bias has only first- and second-order terms in  $1/n$ . Their estimator, which has simpler weights than Quenouille's (2.2), is

$$\tilde{\theta}^{(2)*} = \frac{1}{2} \left[ n^2\theta - 2(n-1)^2 \left( \frac{1}{n} \sum \theta_{-i} \right) + (n-2)^2 \left\{ \frac{2}{n(n-1)} \sum_{i < j} \theta_{-ij} \right\} \right]. \quad (2.3)$$

These ideas can be extended to eliminate even higher-order bias terms if one so desires. The generalization of Quenouille's coefficients in (2.1) should be clear, and the extension of (2.3) is a special case in the next subsection.

### 2.2. Generalized jackknife

In the same 1971 paper Schucany, Gray & Owen generalized the jackknife technique to handle more general forms of bias. Suppose there are two estimators  $\theta_1$  and  $\theta_2$  based on all or parts of the data for which the biases factorize in the following manner:

$$E(\theta_1) = \theta + f_1(n)b(\theta), \quad E(\theta_2) = \theta + f_2(n)b(\theta). \quad (2.4)$$

Then the estimator

$$\tilde{\theta}^* = \frac{\begin{vmatrix} \theta_1 & \theta_2 \\ f_1(n) & f_2(n) \end{vmatrix}}{\begin{vmatrix} 1 & 1 \\ f_1(n) & f_2(n) \end{vmatrix}} \quad (2.5)$$

is exactly unbiased. The usual jackknife (1.2) for  $g = n$  fits into the form (2.5) with  $\hat{\theta}_1 = \hat{\theta}$ ,  $\hat{\theta}_2 = \Sigma\hat{\theta}_{-i}/n$ ,  $f_1(n) = 1/n$  and  $f_2(n) = 1/(n-1)$ .

To eliminate  $k$  separate terms in the bias, each of which factorizes into distinct functions of  $n$  and  $\theta$ ,  $k+1$  estimators are required whose expectations are of the form

$$E(\hat{\theta}_i) = \theta + \sum_{j=1}^k f_{ij}(n) b_j(\theta) \quad (i = 1, \dots, k+1). \quad (2.6)$$

The functions  $f_{ij}(n)$  are assumed to be known. The generalization of (2.5) is

$$\hat{\theta}^* = \frac{\begin{vmatrix} \hat{\theta}_1 & \dots & \hat{\theta}_{k+1} \\ f_{11}(n) & \dots & f_{k+1,1}(n) \\ \vdots & & \vdots \\ f_{1k}(n) & \dots & f_{k+1,k}(n) \\ \hline 1 & \dots & 1 \\ f_{11}(n) & \dots & f_{k+1,1}(n) \\ \vdots & & \vdots \\ f_{1k}(n) & \dots & f_{k+1,k}(n) \end{vmatrix}}{\begin{vmatrix} f_{11}(n) & \dots & f_{k+1,1}(n) \\ \vdots & & \vdots \\ f_{1k}(n) & \dots & f_{k+1,k}(n) \end{vmatrix}}. \quad (2.7)$$

The refined second-order estimator (2.3) is a special case of (2.7) with  $\hat{\theta}_1 = \hat{\theta}$ ,  $\hat{\theta}_2 = \Sigma\hat{\theta}_{-i}/n$ ,  $\hat{\theta}_3 = 2 \sum_{i < j} \hat{\theta}_{-ij}/\{n(n-1)\}$  and  $f_{ij}(n) = 1/(n-i+1)^j$  for  $i = 1, 2, 3$  and  $j = 1, 2$ . To remove the next higher-order bias term  $1/n^3$ , the determinants in (2.7) are simply enlarged to include  $\hat{\theta}_4 = 6 \sum_{i < j < l} \hat{\theta}_{-ijl}/\{n(n-1)(n-2)\}$  and  $f_{i3} = 1/(n-i+1)^3$  ( $i = 1, \dots, 4$ ).

Adams, Gray & Watkins (1971) investigate the effect of  $\hat{\theta}$  and  $\hat{\theta}^{(2)*}$  on a general bias term from an asymptotic point of view. In a further paper (Gray *et al.*, 1972) the same authors point out an interesting connexion between the jackknife technique and the  $e_1$ -transformation which is one of a variety of methods in numerical analysis for increasing the speed of convergence of a series. For a slowly converging series of numbers

$$S_n = \sum_{i=1}^n a_i, \quad (2.8)$$

the transformation

$$e_1(S_n) = \frac{S_n - \rho(n) S_{n-1}}{1 - \rho(n)} \quad (2.9)$$

for  $\rho(n) = a_n/a_{n-1} \neq 1$  will increase the rate of convergence to the limit  $S_\infty$  in many instances. The analogy is  $S_n \sim E(\hat{\theta})$ ,  $S_{n-1} \sim E(\Sigma\hat{\theta}_{-i}/n)$ ,  $S_\infty \sim \theta$ ,  $\rho(n) \sim (n-1)/n$ , and similar analogies exist between the generalizations of the  $e_1$ -transformation and the jackknife. The jackknife estimate is the linear extrapolation to  $0 = 1/\infty$  from  $\hat{\theta}$  plotted at  $1/n$  and  $\Sigma\hat{\theta}_{-i}/n$  at  $1/(n-1)$ .

In their book Gray & Schucany (1972) have amalgamated and expanded the results from these aforementioned papers on bias reduction and included some of the material from the next section on interval estimation.

In any specific problem there are usually alternative methods for reducing the bias. For example, in the problems of §§3.3-3.5 estimation of the quadratic term in the Taylor expansion of  $f$  will also eliminate the order  $1/n$  bias. Alternative estimators are also cited in the next section. A systematic comparison of the jackknife's effectiveness in bias reduction in competition with alternative procedures over a wide class of problems has not been made.

### 2.3. Main application: ratio estimation

Ratio estimation occupies an important place in sample surveys, and since the simple estimator  $\bar{Y}/\bar{X}$  is biased, this has become an area of application for the jackknife technique. For a sample  $(X_i, Y_i)$  ( $i = 1, \dots, n$ ) of paired random variables with  $E(X_i) = \mu$  and  $E(Y_i) = \eta$ , the problem is to estimate  $\theta = \eta/\mu$ . In sample surveys the auxiliary population mean  $\mu$  may be considered known, or at least estimated from a much larger sample, in which case  $\hat{\eta} = \hat{\theta}\mu$ , where  $\hat{\theta}$  is a ratio estimate based on  $(X_i, Y_i)$  ( $i = 1, \dots, n$ ), is often a more precise estimator of  $\eta$  than the less sophisticated estimator  $\bar{Y}$ . There are also many instances of ratio estimation in scientific problems which have no connection with sample surveys.

Durbin (1959) pioneered application of the jackknife to ratio estimation by studying the behaviour of (1.2) with  $g = 2$  in the model

$$Y_i = \alpha + \beta X_i + e_i, \quad (2.10)$$

where the  $e_i$  are independently, identically distributed with either a normal or gamma distribution. Durbin established that, neglecting terms of  $O(n^{-4})$ , the jackknife estimator has both smaller bias and smaller variance than the simple estimator  $\bar{Y}/\bar{X}$  for the normal distribution. In the case of the gamma distribution expansions are not necessary, and Durbin proved that for gamma distributions with coefficient of variation less than  $\frac{1}{4}$  the jackknife reduces the bias, increases the variance, but reduces the mean squared error in comparison with  $\bar{Y}/\bar{X}$ .

Rao (1965) proved that the optimum choice of  $g$  in the jackknife is  $g = n$  for the normal auxiliary distribution. Through a combination of theoretical and numerical work J. N. K. Rao & Webster (1966) demonstrated that this also holds true for the gamma distribution. In an abstract Chakrabarty & J. N. K. Rao (1968) announced similar results on the estimate of the jackknife variance.

The reader should not conclude from this discussion that the only alternative estimator to  $\bar{Y}/\bar{X}$  is its jackknifed version. There is in fact a considerable number of competitors including estimators proposed by Mickey, Hartley and Ross, Tin, and Beale. A variety of papers have attempted to unravel which estimator is the best to use. Four papers which include the jackknifed  $\bar{Y}/\bar{X}$  as one of the contestants are Tin (1965), Rao & Beagle (1967), Rao (1969), and Hutchison (1971). The findings favour the jackknife, the Tin, and the Beale estimators. The jackknife does not always win the contest, but it never lags far behind the winner. Although the jackknifed  $\bar{Y}/\bar{X}$  is computationally more difficult than the other estimators when  $g = n$ , it has the advantage of an easily computed estimate of its variability being associated with it; see §3.4.

For a more general model P. S. R. S. Rao & J. N. K. Rao (1970) also announced results on the comparison of some of these estimators.

Two final papers to be mentioned are Deming (1963) and Brillinger (1966a) which describe the application of the jackknife technique in different types of sample surveys.

## 3. INTERVAL ESTIMATION

### 3.1. General remarks

The next subsections describe general problems in which it has been proved that Tukey's proposal is indeed valid. Namely, the statistic (1.3) has an approximate  $t$  distribution or,

for large  $g$ , an approximate normal distribution. The proofs, which are not included, are all of a similar character. A power series expansion of  $\hat{\theta}$  in terms of the basic random variables  $Y_1, \dots, Y_n$  is derived. It is then shown that the linear term in the expansion gives the correct behaviour and the other terms are asymptotically negligible.

### 3.2. Function of a maximum likelihood estimate

Consider the standard formulation in which the maximum likelihood estimate  $\hat{\theta}$  is a root of the equation

$$0 = \sum_{i=1}^n \frac{\partial \log p(Y_i; \theta)}{\partial \theta}, \quad (3.1)$$

where  $p(y; \theta)$  is the density function or discrete mass function for the random variables  $Y_i$ . Under the usual regularity conditions for the asymptotic normality of  $\hat{\theta}$ , Brillinger (1964) proved that the limiting distribution of (1.3) is exactly a  $t$  distribution with  $g - 1$  degrees of freedom when  $g$  is held fixed and  $h \rightarrow \infty$ . With little extra effort the proof can be extended to the case where  $\hat{\theta} = f(\hat{\phi})$  and  $\hat{\phi}$  is the root of an equation analogous to (3.1) with  $\phi$  replacing  $\theta$ . The function  $f$  needs to have a continuous first derivative.

Note that the above limit was taken by holding the number of groups  $g$  fixed and letting the size of each group become large. It has been shown in ratio estimation that smaller mean squared error is achieved by increasing the number of groups rather than the size of each group, and one might guess this is true generally whenever it is computationally feasible to do so. Also, by letting  $g \rightarrow \infty$  one obtains Gaussian critical values, which are smaller than  $t$  critical values, so that the intervals may be sharper for large  $g$  and small  $h$ . However, no one has succeeded in proving that (1.3) has a limiting normal distribution when  $g = n \rightarrow \infty$  under the usual regularity conditions on  $p(y; \theta)$ . The proof can be pushed through by putting much heavier conditions on  $p(y; \theta)$  such as bounds on derivatives higher than the third, but the details are so unpleasant that no one has seen fit to publish them.

In an abstract Fryer (1970) reports results on the moments of maximum likelihood jackknifing in the multiparameter case.

The maximum likelihood problem is perhaps not so significant an application for the jackknife as the remaining subsections because the distribution of the random variables  $Y_i$  is specified. This means that considerable other distributional machinery can be brought to bear to determine the exact or asymptotic distribution of  $\hat{\theta}$ . It would be quite interesting to know what the jackknife gives when the assumed parametric model does not hold.

### 3.3. Function of a mean

Let  $\theta = f(\mu)$  where  $\mu = E(Y_i)$ . Miller (1964) proved that for  $\hat{\theta} = f(\bar{Y})$  the limiting distribution of (1.3) with  $g = n$  is a unit normal distribution as  $n \rightarrow \infty$ , provided that  $\text{var}(Y_i) = \sigma^2 < \infty$  and  $f$  has a bounded second derivative near  $\mu$ .

The class of statistics which can be put in the form  $f(\bar{Y})$  is rather limited, but extension to similar variables such as  $f(s^2)$  where  $s^2$  is the sample variance (Miller, 1968) was immediate. The class was considerably broadened by the generalization in the next subsection to statistics of the form  $f(U)$  where the argument is a  $U$ -statistic. A  $U$ -statistic is in a sense a fancy mean.

3.4. *Function of a U-statistic*

Any statistic of the form

$$U(Y_1, \dots, Y_n) = \frac{1}{\binom{n}{m}} \sum_c k(Y_{i_1}, \dots, Y_{i_m}), \quad (3.2)$$

where the kernel function  $k(y_1, \dots, y_m)$  is symmetric in its  $m$  arguments and the summation is over all the combinations of  $m$  variables  $Y_{i_1}, \dots, Y_{i_m}$  out of the  $n$  variables  $Y_1, \dots, Y_n$ , is termed a  $U$ -statistic. Let  $\mu = E\{k(Y_1, \dots, Y_m)\}$ . The parameter of interest is  $f(\mu)$ , and the associated estimator to be jackknifed is  $\hat{\theta} = f(U)$ . Then Arvesen (1969) proved that (1.3) with  $g = n$  has a limiting unit normal distribution as  $n \rightarrow \infty$  provided that  $E\{k^2(Y_1, \dots, Y_m)\} < \infty$  and  $f$  has a bounded second derivative near  $\mu$ .

Arvesen also extended this result to the very general case of a real-valued function of several  $U$ -statistics  $f(U_1, \dots, U_r)$  where each  $U$ -statistic  $U_i$  has a different kernel function  $k_i$  for the same set of basic independent and identically distributed variables  $Y_1, \dots, Y_n$  which can now be  $p$ -dimensional vectors. The class of statistics falling into this framework is quite broad and includes, for example, ratios, the  $t$  statistic, the Wilcoxon signed-rank statistic, and the product-moment correlation coefficient.

An attempt was made by Arvesen to generalize these results to the case of nonidentically distributed independent random variables or vectors. Unfortunately, heavy conditions on the proper behaviour of complicated moments are required, so this approach has not led to many fruitful applications with the exception of justifying the use of the jackknife in unbalanced variance component problems; see §3.8.

Earlier Mantel (1967) had noticed that the components used in the jackknife  $\hat{\theta}$ ,  $\sum \hat{\theta}_{-i}/n$ ,  $2 \sum_{i < j} \hat{\theta}_{-ij}/\{n(n-1)\}$ , etc. are  $U$ -statistics. It is not clear how to exploit this observation to use the probability theory developed for  $U$ -statistics unless  $h$  is held fixed and  $\sum \hat{\theta}_{-i}/n$  is generalized to averaging over all possible subsets of size  $(g-1)h$ .

3.5. *Function of regression estimates*

With the one exception noted at the end of the previous subsection the machinery of the jackknife has to date been confined to handling balanced problems with independent and identically distributed random variables. In a forthcoming paper Miller (1974) widens the domain of validated applicability of the jackknife to the full linear model.

Let  $Y = X\beta + e$ , where  $Y = (Y_1, \dots, Y_n)'$ ,  $\beta = (\beta_1, \dots, \beta_p)'$ ,  $e = (e_1, \dots, e_n)'$ , and  $X$  is an  $n \times p$  matrix. For simplicity assume  $\text{rank}(X) = p$ . The random variables  $e_i$  are assumed to be independent and identically distributed with  $E(e_i) = 0$ ,  $\text{var}(e_i) = \sigma^2$ , and  $E(e_i^4) < \infty$ . The  $X$  matrix is assumed known, the  $Y$  vector observed, and the parameters  $\beta$  and  $\sigma^2$  are unknown.

Let  $\theta = f(\beta)$  be the parameter of interest, where  $f$  is a real-valued function with bounded second derivatives near the true  $\beta$ . The customary *ad hoc* estimator of  $\theta$  would be  $\hat{\theta} = f(\hat{\beta})$ , where  $\hat{\beta}$  is the least squares estimator  $(X'X)^{-1}X'Y$ . The jackknife is applied in the usual fashion by successively deleting each row of  $X$  and  $Y$  to obtain  $\hat{\theta}_{-i} = f(\hat{\beta}_{-i})$  ( $i = 1, \dots, n$ ) and hence the corresponding pseudo-values. Then, under the condition  $X'X/n \rightarrow \Sigma$ , a positive-definite matrix, as  $n \rightarrow \infty$ , it can be proved that (1.3) with  $g = n$  has a limiting unit normal distribution.

This result extends the valid use of the jackknife to the estimation of linear or nonlinear  $f(\beta)$  in unbalanced analyses of variance or regression problems. Although the details have not been checked, the proofs seem to generalize to the case of nonlinear least squares  $\Sigma\{Y_i - g(x_i; \beta)\}^2$ .

### 3.6. Stochastic processes with stationary, independent increments

Gaver & Hoel (1970) were interested in estimating the reliability parameter  $\theta = e^{-\lambda\tau}$  for fixed  $\tau > 0$  where  $\lambda$  is the intensity parameter of a Poisson process  $\{Y_t\}$ . If the process is observed over the interval  $[0, T]$ , then  $\hat{\lambda} = Y_T/T$  is everyone's estimator for  $\lambda$ . A jackknife method of estimating  $\theta = e^{-\lambda\tau}$  is to divide the time interval  $[0, T]$  into  $n$  equal-length subintervals and to jackknife the *ad hoc* estimator  $\hat{\theta} = e^{-\hat{\lambda}\tau}$ . If  $\Delta Y_i = Y_{id} - Y_{(i-1)d}$ , where  $d = T/n$ , then the estimator with the  $i$ th subinterval removed is  $\hat{\theta}_{-i} = e^{-\hat{\lambda}_{-i}\tau}$ , where  $\hat{\lambda}_{-i} = (Y_T - \Delta Y_i)/(T - d)$ .

Unlike the previous examples it is possible to pass to the limit  $n \rightarrow \infty$ . This gives the limiting estimator

$$\lim_{n \rightarrow \infty} \hat{\theta} = e^{-\hat{\lambda}\tau} \left\{ 1 - Y_T \left( e^{\tau/T} - 1 - \frac{\tau}{T} \right) \right\}. \quad (3.3)$$

This is not, as one might naively suppose, an unbiased estimator of  $\theta$ . Gaver and Hoel compared this estimator with the unjackknifed  $\hat{\theta}$ , the minimum variance unbiased estimator, and several Bayesian estimators. Although (3.3) is not dominated by any other estimator, it can be considerably improved upon by some of the other estimators for certain values of  $\theta$ . However, which estimator and which values depend on the ratio of  $\tau$  to  $T$ .

Gray, Watkins & Adams (1972) generalized this idea to stochastic processes  $\{Y_t\}$  with stationary, independent increments. Since they need to restrict  $\{Y_t\}$  to processes whose path functions are piecewise continuous and of bounded variation, the Wiener process component is eliminated and  $\{Y_t\}$  reduces essentially to a sum of independent Poisson processes with different jump sizes  $\gamma$  and intensity parameters  $\lambda_\gamma$ . Let  $\theta = f(\lambda)$ , where  $E(Y_t) = \lambda t$  and  $\hat{\lambda} = Y_T/T$ . Then, by dividing the interval into  $n$  equal-length subintervals, jackknifing, and passing to the limit  $n \rightarrow \infty$ , one obtains the estimator

$$\lim_{n \rightarrow \infty} \hat{\theta} = f(\hat{\lambda}) - \sum_{\gamma} N_{\gamma} \left\{ f\left(\hat{\lambda} - \frac{\gamma}{T}\right) - f(\hat{\lambda}) + \frac{\gamma}{T} f'(\hat{\lambda}) \right\}, \quad (3.4)$$

where  $N_{\gamma}$  is the number of jumps of size  $\gamma$  in  $[0, T]$  and  $f'$  is the derivative of  $f$ .

Under the conditions that  $\Gamma = \{\gamma\}$  is a bounded set,  $f$  has a bounded second derivative near  $\lambda$ , and

$$\frac{1}{T} \sum_{\gamma} \gamma^2 N_{\gamma} \rightarrow \sigma^2 = \text{var}(Y_1) < \infty \quad (3.5)$$

in probability as  $T \rightarrow \infty$ , the estimator (3.4) is asymptotically normally distributed with mean  $\theta$  and variance  $\sigma^2\{f'(\lambda)\}^2/T$ , as  $T \rightarrow \infty$ . The limit as  $n \rightarrow \infty$  of the jackknife variance estimate  $\tilde{s}^2/n = \Sigma(\hat{\theta}_i - \hat{\theta})^2/(n(n-1))$  is

$$\lim_{n \rightarrow \infty} \frac{\tilde{s}^2}{n} = \sum_{\gamma} N_{\gamma} \left\{ f\left(\hat{\lambda} - \frac{\gamma}{T}\right) - f(\hat{\lambda}) \right\}^2. \quad (3.6)$$

Under the conditions (3.5),  $\Gamma$  bounded, and  $f'$  continuous near  $\lambda$ , (3.6) multiplied by  $T$  converges in probability to  $\sigma^2\{f'(\lambda)\}^2$  as  $T \rightarrow \infty$ . Thus,  $T^{\frac{1}{2}}(\lim \hat{\theta} - \theta)/(\lim \tilde{s}^2/n)^{\frac{1}{2}}$  has a limiting unit normal distribution as  $T \rightarrow \infty$  under the stated conditions.

### 3.7. Counterexamples

The reader should not acquire the impression that the jackknife always works. This is far from true. For the jackknife to operate correctly the estimator  $\hat{\theta}$  has to have a locally linear quality. Miller (1964) demonstrated that for  $\hat{\theta} = Y_{(n)}$ , the largest order statistic, the limiting behaviour of (1.3) with  $g = n$  can be degenerate or nonnormal. Similarly, in unpublished notes L. E. Moses showed that when  $\hat{\theta}$  is the sample median the asymptotic distribution of (1.3) with  $g = n$  is normal but with variance 4 when  $n$  is even.

For a truncation point problem with  $\hat{\theta} = Y_{(n)}$  Robson & Whitlock (1964) modify the definition of the jackknife because of the particular bias expansion and derive  $2Y_{(n)} - Y_{(n-1)}$  as an estimator. Gray & Schucany (1972, pp. 13–8, 44–51) show how this and the higher order Robson–Whitlock truncation point estimators can be obtained from the generalized jackknife (2.5) and (2.7). In obtaining approximate confidence intervals Robson and Whitlock abandon the ratio (1.3) and give arguments for the approximate confidence statement

$$\text{pr} \left\{ Y_{(n)} + \frac{1-\alpha}{\alpha} (Y_{(n)} - Y_{(n-1)}) > \theta \right\} \simeq 1 - \alpha. \quad (3.7)$$

An area in which the jackknife has had little or no success is time series analysis. This is an ironic twist because it was for a time series problem that Quenouille originally proposed the idea. Except for the case  $g = 2$ , the removal of data segments from a serially correlated sequence of observations causes difficulty for the jackknife. For example, no one has successfully found a way to make it provide valid estimates of the variability of smoothed estimates of the spectral density.

### 3.8. Main application: inference on variances

Standard textbook methodology uses the chi-squared distribution for setting a confidence interval on  $\sigma^2$  from  $s^2$ , the sample variance. Similarly, the  $F$  distribution is employed to make inferences on  $\sigma_1^2/\sigma_2^2$  from  $s_1^2/s_2^2$ , the ratio of variances from independent samples. Pearson (1931) and Box (1953) sounded the alarm on the use of these procedures as well as their  $k$ -sample cousins, Bartlett's test, Hartley's test, and Cochran's test. The distribution theory for these techniques is precisely correct when the observations are normally distributed, but the error probabilities can be grossly inaccurate for nonnormal distributions.

Mosteller & Tukey (1968) and Miller (1968) studied the application of the jackknife to  $\log s^2$  or  $\log s_1^2 - \log s_2^2$ . The jackknife variance does correctly estimate the variability of  $\log s^2$  or  $\log s_1^2 - \log s_2^2$  for all underlying distributions, whereas the previously mentioned techniques rely upon a theoretical variability which is valid only when  $\mu_4/\sigma^4 = 3$ . For two-sample hypothesis testing the jackknife seems to perform about the same as the Box–Andersen technique, in which a beta approximation uses the fourth-sample moments to adjust the degrees of freedom. Both outperform Levene's test and the variants of the Box technique where the sample is divided into subsamples and comparison is made between the  $\log s^2$  computed for each subsample. In the Box-type tests power is lost in the arbitrary division into subsamples. Shorack (1969) makes additional comparisons among competing variance tests which are relevant to the preceding brief discussion.

Layard (1973) examined the jackknife's performance in the  $k$ -sample hypothesis testing problem. The jackknife and an asymptotic  $\chi^2$  test, which, like Box and Andersen's test,

estimates the fourth moment directly, perform about the same, and Box's test lags farther behind than in the two-sample problem.

Inference on variance components in problems such as

$$Y_{ij} = \mu + a_i + e_{ij} \quad (i = 1, \dots, I; j = 1, \dots, n),$$

with  $a_i \sim N(0, \sigma_a^2)$  and  $e_{ij} \sim N(0, \sigma_e^2)$ , all random variables being independent, is also extremely sensitive to normality assumptions. Arvesen (1969) considers application of the jackknife to inference on  $\sigma_a^2$ , and Arvesen & Schmitz (1970) to inference on  $\sigma_a^2/\sigma_e^2$ . In the latter paper their Monte Carlo results reaffirm the unsuitability of normal theory techniques for general distributions, and demonstrate the robustness of the jackknife when applied to the log of the variance ratio. The results are not as good without the log transformation. More will be said in §5.3 about the need for transformations in conjunction with the jackknife. In a technical report Arvesen and Layard give the theoretical machinery needed to justify the use of the jackknife in unbalanced variance component problems.

#### 4. OTHER DEVELOPMENTS

##### 4.1. Multivariate analysis

Normal theory tests on covariance matrices are also sensitive to nonnormality. Layard (1972) asymptotically described the nonrobustness of the usual tests for the equality of two covariance matrices and the effectiveness of the jackknife in dealing with this problem. In a recent paper Duncan & Layard (1973) investigate by Monte Carlo sampling the special case of the correlation coefficient in a bivariate population. It is dangerous to use Fisher's transformation  $Z = \tanh^{-1}r$  for testing whether a correlation coefficient equals a specified nonzero value or for testing the equality of two correlation coefficients because of the extreme sensitivity of the variance of  $Z$  to a lack of normality. Jackknifing  $Z$  does, however, produce a correct asymptotic variance.

In an earlier paper Dempster (1966) proposed a modified jackknife when dealing with canonical correlations. His proposal is to delete single degrees of freedom rather than single-vector observations. The procedure seems harder to follow than the ordinary jackknife, but this may be illusory.

Another area of multivariate analysis in which the jackknife has found application is discriminant analysis. One can jackknife the discriminant coefficients to assess their variability, but the more interesting application is in the estimation of the error or misclassification probabilities. The method of testing each vector observation in the samples, with the discriminant function computed from all the observations, in order to estimate the error rates has been known for some time to be subject to serious bias. There are various normal and nonparametric procedures alternative to this, and one is to test each observation with the discriminant function computed from the data with that particular observation removed. This has been termed the  $U$  method by Lachenbruch & Mickey (1968), and in the same journal issue Cochran (1968) referred to this method as an application of the jackknife principle although this is not precisely correct. Mosteller & Tukey (1968) examined the performance of the  $U$  method on the discrimination problem created by the Federalist Papers. A synopsis of the Mosteller-Tukey work is given by Gray & Schucany (1972, pp. 115–36).

#### 4.2. Infinitesimal jackknife

In a recent Bell Telephone Laboratories technical memorandum L. B. Jaeckel has introduced the concept of an infinitesimal jackknife. At the moment it does not appear to be as practically useful as the ordinary jackknife, but it establishes what could be an important bridge between the jackknife and the recently developed theory for robust estimation of the location of a symmetric distribution (Andrews *et al.* 1972; Huber, 1972).

To understand the connexion it is necessary to briefly summarize some relevant aspects of the theory for robust estimation. Many estimators  $\hat{\theta}$  are equal to, or are asymptotically equivalent to, functions  $T(\hat{F})$  of the sample cumulative distribution function  $\hat{F}$ , where  $T$  is defined over a wide class of distribution functions. In particular, the unknown parameter  $\theta$  is  $T(F)$ , the value of the function at the true  $F$ . Under regularity conditions estimators of this type can be expressed in the form

$$T(\hat{F}) = T(F) + \int T'(F, y) d(\hat{F} - F)(y) + o_p(n^{-\frac{1}{2}}), \quad (4.1)$$

where  $T'(F, y)$  is a von Mises (1947) derivative, defined by

$$\lim_{\epsilon \rightarrow 0} \frac{T(F + \epsilon G) - T(F)}{\epsilon} = \int T'(F, y) dG(y). \quad (4.2)$$

The term influence curve has been attached to  $T'(F, y)$  by F. R. Hampel because it measures to what degree a change in the mass at  $y$  will change the estimate.

If the function  $T$  is assumed to be normalized so that  $T(cF) = T(F)$  for all  $F$  and  $c > 0$ , then  $\int T'(F, y) dF(y) = 0$ , so that

$$T(\hat{F}) = T(F) + \frac{1}{n} \sum_{i=1}^n T'(F, Y_i) + o_p(n^{-\frac{1}{2}}). \quad (4.3)$$

The average of independent and identically distributed random variables in (4.3) is asymptotically normally distributed with mean zero and variance

$$\frac{1}{n} \int \{T'(F, y)\}^2 dF(y). \quad (4.4)$$

If  $T'$  is known, an empirical estimate of the asymptotic variance is

$$\frac{1}{n^2} \sum_{i=1}^n \{T'(\hat{F}, Y_i)\}^2. \quad (4.5)$$

To define the infinitesimal jackknife think of the estimator  $\hat{\theta}$  as a function  $T(Y; w)$  of the observations  $Y = (Y_1, \dots, Y_n)'$  and arbitrary weights  $w = (w_1, \dots, w_n)'$ . If  $w_i \equiv 1/n$ , then  $\hat{\theta} = T(\hat{F})$ . Also, suppose that the function of the observations and weights is self-normalizing in the weights so that  $T(Y; cw) = T(Y; w)$  for all  $c > 0$ .

For the ordinary jackknife

$$\hat{\theta}_{-i} = T\left(Y_1, \dots, Y_n; \frac{1}{n}, \dots, 0, \dots, \frac{1}{n}\right), \quad (4.6)$$

but for the infinitesimal jackknife the weight of the  $i$ th observation is reduced only to  $1/n - \epsilon$  instead of to 0, namely,

$$\hat{\theta}_{-i}(\epsilon) = T\left(Y_1, \dots, Y_n; \frac{1}{n}, \dots, \frac{1}{n} - \epsilon, \dots, \frac{1}{n}\right). \quad (4.7)$$

By analogy with the ordinary jackknife the infinitesimal jackknife estimate of the asymptotic variance of  $\hat{\theta}$ , or its infinitesimally jackknifed version, is defined to be

$$\frac{\tilde{s}^2(\epsilon)}{n} = \frac{(1-\epsilon)}{n^2\epsilon^2} \sum_{i=1}^n \left\{ \hat{\theta}_{-i}(\epsilon) - \frac{1}{n} \sum \hat{\theta}_{-j}(\epsilon) \right\}^2. \quad (4.8)$$

When  $\epsilon = 1/n$ , this coincides with the ordinary jackknife estimate.

Suppose that  $T(y; w)$  is differentiable with respect to the weights. Then, let

$$\hat{D}_i = \frac{\partial T(y; w)}{\partial w_i}, \quad \hat{D}_{ii} = \frac{\partial^2 T(y; w)}{\partial w_i^2}, \quad (4.9)$$

where the derivatives are evaluated at  $y = Y$ ,  $w = (1/n, \dots, 1/n)'$ . The self-normalizing condition on the weights in  $T$  implies that  $\sum \hat{D}_i = 0$ . From the expansion

$$\hat{\theta}_{-i}(\epsilon) = \hat{\theta} - \epsilon \hat{D}_i + \frac{1}{2} \epsilon^2 \hat{D}_{ii} - \dots, \quad (4.10)$$

it follows that

$$\frac{\tilde{s}^2(0)}{n} = \lim_{\epsilon \rightarrow 0} \frac{\tilde{s}^2(\epsilon)}{n} = \frac{1}{n^2} \sum_{i=1}^n \hat{D}_{ii}^2. \quad (4.11)$$

But  $\hat{D}_i$  is precisely  $T'(\hat{F}, Y_i)$ , so (4.11) equals (4.5). Thus, the infinitesimal jackknife variance estimate (4.11) provides an estimate of the asymptotic variance (4.4).

For the ordinary jackknife

$$\hat{\theta} - \tilde{\theta} = (n-1) \left( \frac{1}{n} \sum \hat{\theta}_{-i} - \hat{\theta} \right) \quad (4.12)$$

estimates the bias of  $\hat{\theta}$ . Correspondingly, let

$$\tilde{b}(\epsilon) = \frac{1-\epsilon}{n\epsilon^2} \left\{ \frac{1}{n} \sum \hat{\theta}_{-i}(\epsilon) - \hat{\theta} \right\}. \quad (4.13)$$

Expression (4.13) equals (4.12) when  $\epsilon = 1/n$ . From the power series expansion (4.10) it follows that

$$\tilde{b}(0) = \lim_{\epsilon \rightarrow 0} \tilde{b}(\epsilon) = \frac{1}{2n} \sum_{i=1}^n \hat{D}_{ii}. \quad (4.14)$$

The infinitesimal jackknife estimate is defined to be  $\tilde{\theta}(0) = \hat{\theta} - \tilde{b}(0)$ .

Jaeckel proves under general conditions, which will not be cited here, that for estimators of the form (4.3)  $\tilde{s}^2(0)$  and  $n\tilde{b}(0)$  converge to the correct asymptotic constants as  $n \rightarrow \infty$ . It is not at present clear how easy it is to check these conditions in particular examples. Jaeckel also shows that under the same conditions the ordinary jackknife estimates behave correctly asymptotically.

#### 4.3. *Miscellanea*

Salsburg (1971) gives an application of the jackknife to testing in quantal response bioassay where the probabilities of success at the observed dose levels are all near 1 (or 0). This application is also described in a review paper by Arvesen & Salsburg (1973).

In a recent paper Gray, Watkins & Schucany (1973) explore the connexion between the jackknife and uniform minimum variance unbiased estimation of  $f(\mu)$ ,  $f(\sigma^2)$  and  $f(\mu, \sigma^2)$  for special  $f$ , where  $\mu$  and  $\sigma^2$  are the parameters of a normal distribution.

In the discussion of a paper by P. Sprent on linear functional relationships Brillinger

(1966b) proposes the use of the jackknife to assess the variability of the estimators. In another forthcoming paper P. O. Anderson has studied the jackknife's effect on regression estimators which minimize the orthogonal distance.

## 5. NEW DIRECTIONS

### 5.1. General remarks

Where should research on the jackknife go from here? What worthwhile questions on the jackknife still remain unanswered? Listed below are some thoughts on this which are by no means all-inclusive.

#### 5.2. Linear combinations of order statistics

Estimators based on single order statistics such as the median or maximum (§ 3.7) do not behave properly under jackknifing. What about smooth functions of the order statistics such as  $\theta = \sum J(i/n)Y_{(i)}/n$ , where  $J$  is a continuous function? Estimators like this arise in robust estimation of the location of a symmetric distribution and overlap with the class of estimators (4.3) where the jackknife is known to work under regularity conditions; see § 4.2. How broad is the class of estimators for which (4.3) and the needed regularity conditions are satisfied? Will the jackknife accurately assess the variability of the more complicated adaptive estimators? Obtaining good estimators of the variability of robust estimators seems to be an open and important question.

#### 5.3. Transformations

Most advocates of the jackknife would suggest using a variance stabilizing transformation on the estimator in conjunction with the jackknife. Examples are jackknifing  $\log s^2$  and  $\tanh^{-1} r$  instead of  $s^2$  and  $r$ . However, the connexion between transformations and the jackknife is more than just a nicety. Transformations are needed to keep the jackknife on scale and thus prevent distortion of the results.

Consider inference on  $\sigma^2$ . Without the log transformation the pseudo-values  $ns^2 - (n-1)s_{-i}^2$  can be negative. The jackknife does not know variances must be positive. Sizeable or frequent negative pseudo-values can produce distorted point and interval estimates. With the log transformation a negative pseudo-value corresponds to a small variance, and this will not pull the aggregate so far to the left.

Table 1. *Control blood flow data*

DYE	1.15	1.70	1.42	1.38	2.80	4.70	4.80	1.41	3.90
EFP	1.38	1.72	1.59	1.47	1.66	3.45	3.87	1.31	3.75

The following data illustrate the analogous point with correlation coefficients. A medical investigator at the Stanford Medical School wanted a measure of association between two techniques for measuring blood flow. A standard method (DYE) is to inject dye into the pulmonary artery and sample it from the aorta. A computer integrates the experimentally determined curve of dye concentration to obtain a blood-flow measurement. The electromagnetic flow probe method (EFP) is newer. In it a cuff placed around the aorta creates an electrical field to measure the blood flow. The DYE method is extremely variable, and the EFP has serious calibration problems. To assess the amount of agreement between the two methods, essentially simultaneous measurements were made on nine dogs. The data in appropriate units are displayed in Table 1.

The number of observations may be rather small to be trying the jackknife, but it is interesting to see what it gives. The correlation coefficients and pseudo-values with and without the transformation  $\tanh^{-1}$  appear in Table 2.

The correlation coefficient estimate  $\tilde{r}$ , is quite remarkable! The untransformed jackknifed estimate is  $\tilde{r} = 0.9452$  and its estimated standard error is 0.0408. The jackknife has increased the estimated value of  $\rho$  slightly, which is probably in the wrong direction. Because of the size of the standard error, a confidence interval for  $\rho$  will extend well beyond the upper limit 1.0. The transformed values give  $\tilde{\tanh^{-1}r} = 1.593$  which corresponds to a  $\rho$  value of 0.9206. The estimated standard error of  $\tilde{\tanh^{-1}r}$  is 0.471 so that a 95% confidence interval for  $\rho$  is (0.4672, 0.9906). The peculiar pseudo-value -1.724 is what makes the interval so broad.

Table 2. Correlation coefficients and pseudo-values for control blood flow data with and without transformation

$i$	$r_{-i}$	$\tilde{r}_i$	$\tanh^{-1}r_{-i}$	$\tilde{\tanh^{-1}r_i}$	$i$	$r_{-i}$	$\tilde{r}_i$	$\tanh^{-1}r_{-i}$	$\tilde{\tanh^{-1}r_i}$
0	0.9448	—	1.780	—	5	0.9766	0.6904	2.218	-1.724
1	0.9406	0.9784	1.743	2.076	6	0.9396	0.9864	1.735	2.140
2	0.9434	0.9560	1.768	1.876	7	0.9205	1.1392	1.593	3.276
3	0.9434	0.9560	1.768	1.876	8	0.9396	0.9864	1.735	2.140
4	0.9408	0.9768	1.745	2.060	9	0.9583	0.8368	1.925	0.620

Table 3. Propranolol blood flow data

DYE	0.77	0.87	1.05	1.21	3.80	0.55	2.33
EFP	1.03	0.96	1.25	1.66	3.39	0.60	3.12

The connexion between transformations and the jackknife is worth more exploration than a single numerical example and a few published Monte Carlo results (Arvesen & Schmitz, 1970). Is there an optimal way to select a transformation for use with the jackknife?

#### 5.4. Outliers

The jackknife is not a device for correcting outliers. The following numerical example illustrates this. On seven of the dogs in the experiment described in §5.3, first a stimulant, isuprel, and then a depressant, propranolol, were administered after the data listed in Table 1 had been obtained under control conditions. Dual measurements on DYE and EFP were taken under these induced states. The propranolol data are displayed in Table 3.

The fifth point could be considered an outlier because in all other pairs  $\text{DYE} < \text{EFP}$ . The transformed correlation coefficients and pseudo-values are presented in Table 4.

The jackknifed estimate is  $\tilde{\tanh^{-1}r} = 0.375$  which corresponds to a value of  $\rho$  of 0.359. The jackknife has pulled the estimate down from 0.948 as one might hope, but it seems to have gone beyond the bounds of reason. The estimated standard error for  $\tilde{\tanh^{-1}r}$  is 1.077 so that a 95% confidence interval for  $\rho$  is (-0.978, 0.995).

It would be interesting to study theoretically how outliers perturb the jackknife estimator. Is the examination and correction of pseudo-values a good way of handling outliers?

#### 5.5. Higher-order variance estimate

At the moment the variance estimate for the jackknife involves only the pseudo-values from the first-order jackknife. Can the higher-order jackknife expressions be used in any way to improve the variance estimate?

Table 4. *Transformed correlation coefficients and pseudo-values for propranolol blood flow data*

$i$	$r_{-i}$	$\tanh^{-1} r_{-i}$	$\tilde{\tanh^{-1}} r_i$
0	0.9480	1.812	—
1	0.9436	1.770	2.064
2	0.9457	1.789	1.950
3	0.9463	1.795	1.914
4	0.9519	1.852	1.572
5	0.9950	2.995	-5.286
6	0.9423	1.758	2.136
7	0.9837	2.401	-1.722

### 5.6. Multisample jackknives

To illustrate this question consider a two-sample problem. One way to jackknife is to compute an estimate of the unknown parameter from the two samples and then jackknife by successively deleting each observation in the first sample with the second sample intact and then deleting observations in the second sample with the first intact. An alternative method of jackknifing is to jackknife each sample separately and combine the results. Both methods are valid asymptotically, but is either one better than the other?

Similar questions arise when there are three or more samples and when the observations come from an experimental design of some sort.

This work was partially supported by a fellowship from the John S. Guggenheim Memorial Foundation and by a grant from the National Science Foundation. The paper was written while the author was visiting Imperial College and London School of Hygiene and Tropical Medicine. D. R. Cox and D. V. Hinkley contributed a number of the suggestions for future research in § 5, and the author thanks them for their assistance in the preparation of the paper.

### REFERENCES

- The numbers in brackets at the end of each reference give the sections of the paper in which the reference is mentioned.
- ADAMS, J. E., GRAY, H. L. & WATKINS, T. A. (1971). An asymptotic characterization of bias reduction by jackknifing. *Ann. Math. Statist.* **42**, 1606–12, [§ 2.2].
- ANDREWS, D. F., BICKEL, P. J., HAMPEL, F. R., HUBER, P. J., ROGERS, W. H. & TUKEY, J. W. (1972). *Robust Estimates of Location: Survey and Advances*. Princeton University Press, [§ 4.2].
- ARVESEN, J. N. (1969). Jackknifing  $U$ -statistics. *Ann. Math. Statist.* **40**, 2076–100, [§§ 3.4, 3.8].
- ARVESEN, J. N. & SALSBURG, D. S. (1973). Approximate tests and confidence intervals using the jackknife. In *Perspectives in Biometrics*, Ed. R. M. Elashoff. New York: Academic Press, [§ 4.3].
- ARVESEN, J. N. & SCHMITZ, T. H. (1970). Robust procedures for variance component problems using the jackknife. *Biometrics* **26**, 677–86, [§§ 3.8, 5.3].
- BOX, G. E. P. (1953). Non-normality and tests on variances. *Biometrika* **40**, 318–35, [§ 3.8].
- BRILLINGER, D. R. (1964). The asymptotic behaviour of Tukey's general method of setting approximate confidence limits (the jackknife) when applied to maximum likelihood estimates. *Rev. Inst. Int. Statist.* **32**, 202–6, [§ 3.2].
- BRILLINGER, D. R. (1966a). The application of the jackknife to the analysis of sample surveys. *Commentary* **8**, 74–80, [§ 2.3].
- BRILLINGER, D. R. (1966b). Discussion of paper by P. Sprent. *J. R. Statist. Soc. B* **28**, 294, [§ 4.3].
- CHAKRABARTY, R. P. & RAO, J. N. K. (1968). The bias and stability of the jack-knife variance estimator in ratio estimation. *J. Am. Statist. Assoc.* **63**, 748–9, [§ 2.3].
- COCHRAN, W. G. (1968). Commentary on 'Estimation of error rates in discriminant analysis.' *Technometrics* **10**, 204–5, [§ 4.1].

- DEMING, W. E. (1963). On the correction of mathematical bias by use of replicated designs. *Metrika* **6**, 37-42, [§ 2·3].
- DEMPSTER, A. P. (1966). Estimation in multivariate analysis. In *Multivariate Analysis*, Ed. P. R. Krishnaiah, pp. 315-34. New York: Academic Press, [§ 4·1].
- DUNCAN, G. T. & LAYARD, M. W. J. (1973). A Monte-Carlo study of asymptotically robust tests for correlation coefficients. *Biometrika* **60**, 551-8, [§ 4·1].
- DURBIN, J. (1959). A note on the application of Quenouille's method of bias reduction to the estimation of ratios. *Biometrika* **46**, 477-80, [§ 2·3].
- FRYER, J. (1970). Jackknifing maximum likelihood estimates in the multi-parameter case (Abstract). *Ann. Math. Statist.* **41**, 1392, [§ 3·2].
- GAVER, D. P., JR. & HOEL, D. G. (1970). Comparison of certain small-sample Poisson probability estimates. *Technometrics* **12**, 835-50, [§ 3·6].
- GRAY, H. L. & SCHUCANY, W. R. (1972). *The Generalized Jackknife Statistic*. New York: Marcel Dekker, [§§ 2·2, 3·7, 4·1].
- GRAY, H. L., WATKINS, T. A. & ADAMS, J. E. (1972). On the jackknife statistic, its extensions, and its relation to  $e_n$ -transformations. *Ann. Math. Statist.* **43**, 1-30, [§§ 2·2, 3·6].
- GRAY, H. L., WATKINS, T. A. & SCHUCANY, W. R. (1973). On the jackknife statistic and its relation to umvu estimators in the normal case. *Commun. Statist.* **2**, 285-308, [§ 4·3].
- HUBER, P. J. (1972). Robust statistics: A review. *Ann. Math. Statist.* **43**, 1041-67, [§ 4·2].
- HUTCHISON, M. C. (1971). A Monte Carlo comparison of some ratio estimators. *Biometrika* **58**, 313-21, [§ 2·3].
- LACHENBRUCH, P. A. & MICKEY, M. R. (1968). Estimation of error rates in discriminant analysis. *Technometrics* **10**, 1-11, [§ 4·1].
- LAYARD, M. W. J. (1972). Large sample tests for the equality of two covariance matrices. *Ann. Math. Statist.* **43**, 123-41, [§ 4·1].
- LAYARD, M. W. J. (1973). Robust large-sample tests for homogeneity of variances. *J. Am. Statist. Assoc.* **68**, 195-8, [§ 3·8].
- MANTEL, N. (1967). Assumption-free estimators using  $U$  statistics and a relationship to the jackknife method. *Biometrics* **23**, 567-71, [§ 3·4].
- MILLER, R. G. (1964). A trustworthy jackknife. *Ann. Math. Statist.* **35**, 1594-605, [§§ 3·3, 3·7].
- MILLER, R. G., (1968). Jackknifing variances. *Ann. Math. Statist.* **39**, 567-82, [§§ 3·3, 3·8].
- MILLER, R. G. (1974). An unbalanced jackknife. *Ann. Statist.* **2**. To appear. [§ 3·5].
- VON MISES, R. (1947). On the asymptotic distribution of differentiable statistical functions. *Ann. Math. Statist.* **18**, 309-48, [§ 4·2].
- MOSTELLER, F. & TUKEY, J. W. (1968). Data analysis, including statistics. *Handbook of Social Psychology*, Eds. G. Lindzey and E. Aronson. Reading, Mass: Addison-Wesley, [§§ 3·8, 4·1].
- PEARSON, E. S. (1931). The analysis of variance in cases of non-normal variation. *Biometrika* **21**, 114-33, [§ 3·8].
- QUENOUILLE, M. H. (1949). Approximate tests of correlation in time-series. *J.R. Statist. Soc. B* **11**, 68-84, [§ 1].
- QUENOUILLE, M. H. (1956). Notes on bias in estimation. *Biometrika* **43**, 353-60, [§§ 1, 2·1].
- RAO, J. N. K. (1965). A note on the estimation of ratios by Quenouille's method. *Biometrika* **52**, 647-9, [§ 2·3].
- RAO, J. N. K. & BEEGLE, L. D. (1967). A Monte Carlo study of some ratio estimators. *Sankhyā B* **29**, 47-56, [§ 2·3].
- RAO, J. N. K. & WEBSTER, J. (1966). On two methods of bias reduction in the estimation of ratios. *Biometrika* **53**, 571-7, [§ 2·3].
- RAO, P. S. R. S. (1969). Comparison of four ratio-type estimates under a model. *J. Am. Statist. Assoc.* **64**, 574-80, [§ 2·3].
- RAO, P. S. R. S. & RAO, J. N. K. (1970). Some small sample results for ratio estimators (Abstract). *Ann. Math. Statist.* **41**, 1141-2, [§ 2·3].
- ROBSON, D. S. & WHITLOCK, J. H. (1964). Estimation of a truncation point. *Biometrika* **51**, 33-9, [§ 3·7].
- SALSBURG, D. (1971). Testing dose responses on proportions near zero or one with the jackknife. *Biometrics* **27**, 1035-41, [§ 4·3].
- SCHUCANY, W. R., GRAY, H. L. & OWEN, D. B. (1971). On bias reduction in estimation. *J. Am. Statist. Assoc.* **66**, 524-33, [§§ 2·1, 2·2].
- SHORACK, G. R. (1969). Testing and estimating ratios of scale parameters. *J. Am. Statist. Assoc.* **64**, 999-1013, [§ 3·8].
- TIN, M. (1965). Comparison of some ratio estimators. *J. Am. Statist. Assoc.* **60**, 294-307, [§ 2·3].
- TUKEY, J. W. (1958). Bias and confidence in not-quite large samples (Abstract). *Ann. Math. Statist.* **29**, 614, [§ 1].

[Received February 1973. Revised May 1973]