THE APPLICATION OF THE JACKKNIFE TO THE ANALYSIS OF SAMPLE SURVEYS

David R. Brillinger
The London School of Economics and Political Science

The jack-knife technique of estimation is due originally to Tukey. It has been applied widely, whenever the effort of making the theoretically desirable calculations has proved complex and laborious. Indeed, the technique takes its name from the fact that, like the Boy Scout's knife, it can be used to do many jobs, like taking stones out of horses hooves. Standard error estimation for complex, multi-stage samples is just such a job.

All that the jackknife method consists of is the dividing of the sample into a number of different sub-samples (chosen in a particular way) and the calculating of the values we are interested in from all the sub-groups, *except one*, to produce what are known as 'pseudo-values.' The procedure is repeated over and over, omitting a different sub-group each time, the variances then being calculated directly from these 'pseudo-values'. Repeated calculations like this can, of course, be handled expeditiously on a computer.

Summary

This paper presents a description of J. W. Tukey's jackknife technique of setting approximate confidence limits. A theorem is presented indicating a class of statistics to which the jackknife may reasonably be applied. The manner of use of the technique on sample survey data is considered with specific reference to stratified sampling, multi-stage sampling and clustered sampling.

1. Introduction

Statisticians have become proficient at developing estimators of parameters of interest, quite possibly because so many statistical problems may be reduced to a form requiring such. Given an estimator, the next step is to find a measure of how good the estimator is. This step frequently reduces to the derivation of an estimate of the standard error of the estimator. J. W. Tukey (1958) has presented a technique for deriving estimates of standard errors simply employing the estimator of the original parameter of interest. He has called this technique the jackknife.

Reprinted from Commentary Vol. 8 No. 2 by kind permission of the Market Research Society, 37 Hertford Street, London W.1
PRINTED BY SWAIN PRESS LIMITED WIDNES AND LONDON

The jackknife proceeds as follows: suppose one has a set of data and an estimator that may be calculated from this data. Separate the data into r groups (a discussion of the properties required by this separation will occupy Section 2 of the paper). Let q be the estimator based on all the data, $q_{(i)}$ that based on all but the ith group and $q_{pi} = r\,q - (r-1)q_{(i)}$. Quenouille (1956) suggested the use of the estimator $q_p = \Sigma\, q_{pi}/r$ as under certain conditions it has smaller bias than q. Tukey suggested that $\Sigma\,(q_{pi} - q_p)^2 / r(r-1)$ may be used as an estimate of the variances of q and q_p. Further he suggests that

$$\frac{q_p - \theta}{\sqrt{\Sigma\,(q_{pi} - q_p)^2 / r(r-1)}}$$

may be taken as having a Student's distribution on r−1 degrees of freedom for the purpose of obtaining confidence intervals for the unknown parameter θ.

We see that the use of the jackknife requires the repetition of the same set of calculations, namely the estimation, r + 1 times. As such it is particularly suited for use with electronic computers. Current trends are for calculations to become cheaper and programming more expensive. If this continues the advantages of the jackknife should become even more irresistible.

Tukey (1958) justifies the use of the jackknife for the class of sequelinear estimates. The following theorem indicates a further class of statistics to which the jackknife may be applied.

In the theorem s refers to a parameter tending to infinity with increasing sample size. We imagine r the number of groups remaining fixed.

*Theorem**. Suppose (i) $\mathrm{var}\; q = \frac{1}{r}\sigma^2_s + o(\sigma^2_s)$, (ii) $\mathrm{ave}\; q = \theta + o(\sigma_s^2)$,

(iii) $\mathrm{var}\; q_{(i)} = \frac{1}{r-1}\sigma^2_s + o(\sigma^2_s)$, (iv) $\mathrm{ave}\; q_{(i)} = \theta + o(\sigma_s^2)$, then

(a) $\mathrm{ave}\,\frac{r-1}{r}(\Sigma\, q_{(i)}^2 - r\,q^2) = \frac{1}{r}\sigma^2_s + o(\sigma^2_s)$.

If we have (iii) and (v) $\mathrm{cor}\,(q_{(i)}, q_{(j)}) = \frac{r-2}{r-1} + o(1)$, $i \neq j$, then

(b) $\mathrm{ave}\; \Sigma(q_{pi} - q_p)^2/r(r-1) = \frac{1}{r}\sigma^2_s + o(\sigma^2_s)$.

* $f(s) = o(\sigma_s)$ if $\lim_{s\to\infty} f(s)/\sigma_s = 0$. In practice $\sigma_s^2 = \sigma^2/s$ frequently. The motivation for (v) is that $q_{(i)}$ and $q_{(j)}$ have r-2 groups in common out of r-1. "cor" stands for correlation.

If we have (i), (iii), (v) and (vi) $\mathrm{cor}\,(q, \frac{1}{r}\Sigma\, q_{(i)}) = 1 + o(1)$, then

(c) $\mathrm{var}\; q_p = \frac{1}{r}\sigma^2_s + o(\sigma_s^2)$.

Proof. To prove (a) note that $\mathrm{ave}\; q^2 = \theta^2 + \frac{1}{r}\sigma_s^2 + o(\sigma_s^2)$ and $\mathrm{ave}\; q_{(i)}^2 = \theta^2 + \frac{1}{r-1}\sigma_s^2 + o(\sigma_s^2)$. To prove (b) note that $q_{pi} - q_p = (r-1)\left(\frac{\Sigma q_{(j)}}{r} - q_{(i)}\right)$ and therefore $\mathrm{ave}\;(q_{pi} - q_p)^2 = \frac{r-1}{r}\sigma^2_s$, yielding the stated result. (c) follows since $q_p = r\,q - (r-1)\,\Sigma\, q_{(i)}/r$.

We see that the theorem indicates two plausible estimates of var q, namely $\Sigma\,(q_{pi} - q_p)^2/r(r-1)$ and $\frac{r-1}{r}(\Sigma\, q_{(i)}^2 - r\,q^2)$. The first requires that q_p, $q_{(i)}$ have small bias, while the second requires a specific form of correlation.

Note that no assumption has been made concerning the independence of the observations. Also note that (i) to (vi) indicate properties that one seeks to achieve by the division into groups and the form of $q_{(i)}$.

The next section will be devoted to finding the required relation between the splitting into groups and the sample design employed in the collection of the data.

2. Applications to Sample Survey Means

For many sample designs the variance of an estimated mean has the form,

$$\mathrm{var}\;\bar{x} = \frac{1-f}{a}\sigma^2$$

where f is the sampling fraction, a the number of primary selections and σ^2 the variation among the means per element of primary selections (see Kish (1965), p.254)

Suppose that a = r s and that* f does not vary too much with r, then we see that conditions (i) and (iii) of the theorem are satisfied if we split the primary selections into r groups of s and apply the jackknife to the estimated mean.

In the case of stratified designs the variance frequently has the approximate form,

$$\mathrm{var}\;\bar{x} = \Sigma\, w_k \frac{(1-f_k)}{a_k}\sigma^2_k$$

*—In many situations f = a/A and $\mathrm{var}\;\bar{x} = \frac{\sigma^2}{a} - \frac{\sigma^2}{A}$. In this case the jackknife is estimating σ^2/a, i.e. providing an overestimate.

in an obvious notation. If we write $a_k = r\, s_k$ this becomes,

$$\text{var } \bar{x} = \frac{1}{r} \Sigma\, w_k \frac{(1 - f_k)}{s_k} \sigma^2_k$$

and supposing that the f_k do not vary too much with r we see that we should split the primary selections in each of the strata into r groups of s_k and then collect one group from each stratum to form r global groups.

An essential thing to note here is the requirement of an inverse dependence of the variance upon an *a posteriori* manipulable parameter. We note that the notion of "design effect", Kish (1965), has implications of an inverse dependence on sample size.

Let us turn to the consideration of some specific standard sampling designs.

Simple Random Sampling. Here we sample without replacement n items from N with equal probabilities. Suppose the population elements are denoted $X_1, \ldots, X_N$ the observations $x_1, \ldots, x_n$ and we take $q = \bar{x} = \Sigma\, x_i/n$. Then

$$\text{var } q = \frac{1 - f}{n} \sigma^2$$

where $\sigma^2 = \Sigma\,(X_\alpha - \bar{X})^2/(N - 1)$ and $f = n/N$. Supposing $n = r\, s$ we see, as in the preceding discussion, that the conditions of the theorem are satisfied if one randomly splits the sample into r groups of s.

Selection with Probabilities Proportional to Size. Here we sample X with probability π_α and estimate the mean by

$$q = \frac{1}{n\,N} \Sigma\, \frac{x_i}{\pi_i}.$$

If sampling is with replacement the variance of q is $\frac{1}{n}\sigma^2$ where $\sigma^2 = \frac{1}{N^2} \Sigma\pi_\alpha \left(\frac{X_\alpha}{\pi_\alpha} - X\right)^2$. We see that to apply the jackknife one should split the sample randomly into equal sized groups. In addition for this situation we see that the jackknife technique is admirably suited for use with the common procedure of preparing duplicate cards to achieve the proper weighting.

Two-stage sampling. Suppose there are N primaries of which n are selected without replacement and equal probability. Suppose each primary contains M secondaries of which m are selected without replacement. Let x_{ij} denote the jth element in the ith primary sample and X_{ij} the corresponding population value.

As an estimator of the mean one takes

$$q = \frac{\Sigma\,\Sigma\, x_{ij}}{m\,n} = \bar{x}$$

and has

$$\text{var } q = \frac{1}{n}\left[\left(1 - \frac{n}{N}\right)\sigma_1^2 + \left(1 - \frac{m}{M}\right)\frac{\sigma_2^2}{m}\right]$$

where $\sigma_1^2 = \Sigma\,(\bar{X}_\alpha - \bar{X})^2/(N-1)$ and $\sigma_2^2 = \Sigma\,\Sigma\,(X_{\alpha\beta} - \bar{X}_\alpha)^2/N(M-1)$

If $n = r\,s$ and we neglect the dependence of n/N upon r, then the variance has the form required for the theorem and we see that we should split the sample by forming r groups of s primaries each.

The multistage extension of this result is immediate and essentially of the identical form.

Clustered Sampling with Unequal Clusters. Suppose one has N clusters the ith having M_i elements. Let $M_o = \Sigma\, M_i$ and suppose n clusters are sampled with replacement, the ith with probability M_i/M_o. Suppose that m_i elements are selected at random from the ith cluster if the cluster has been sampled.

Consider the estimator $q = \frac{1}{n} \Sigma\, \bar{x}_i$.

$$\text{var } q = \frac{1}{n} \Sigma\, \frac{M_i}{M_o} (\bar{X}_i - \bar{X})^2 + \frac{1}{n} \Sigma\, \frac{M_i}{M_o} \cdot \frac{1 - f_i}{m_i} \cdot \sigma^2_{2i}$$

(see Cochran (1953), p.308).

This variance has the form required by the theorem and we see that the n selected clusters should be divided randomly into r equi-sized groups in order to apply the jackknife.

Stratified Sampling. This case was touched upon at the beginning of the section. Suppose the population is divided into I strata with relative sizes W_i, $\Sigma\, W_i = 1$. The population mean is estimated by $q = \Sigma\, W_i\, \bar{x}_i$, $\bar{x}_i$ denoting the sample mean in the ith stratum. var q is given by

$$\Sigma\, W_i^2\, \frac{\sigma_i^2}{n_i}\,(1 - f_i)$$

$\sigma_i^2 = \Sigma\,(X_{ij} - \bar{X}_i)^2/(N_i - 1)$ where N_i denotes the stratum size, n_i the sample size in the ith stratum and $f_i = n_i/N_i$. Letting $n_i = r\, s_i$ this becomes

$$\frac{1}{r}\, \Sigma\, W_i^2\, \frac{\sigma_i^2}{s_i}\,(1 - f_i).$$

Assuming f_i doesn't change too much with r, this is of the form of the theorem and we see that we should randomly split the sample elements in the ith stratum into r groups of size s_i and then randomly piece together a group from each stratum to form r global groups.

In this case, if the division into r groups of size S_i proves difficult because of unequal stratum size, it may prove appropriate to build up a variance estimate by estimating the variances of the individual strata and adding these with appropriate multiples.

3. Applications to Non-Linear Estimators

The reader will have noted that the estimators considered in the previous section were of the linear unbiased type. There are however many estimators in sample survey that are neither unbiased not linear, for example ratio estimators, medians and regression coefficients. The use of non-linear estimates will probably increase in the future, trimmed estimates, (Tukey and McLaughlin, 1963), coming in for example.

It is for the non-linear estimators in a sample design structure that the jackknife will undoubtedly prove most useful. In this case the error terms of the theorem, o (σ^2_s), will be required. Turning to specific estimators.

Ratio Estimators. In Cochran (1953), p.158 it is shown that for a ratio estimator.

$$\text{var}\,\hat{R} = \frac{1-f}{n}(\sigma^2_y + R^2\sigma^2_x - 2R\rho\sigma_y\sigma_x) + o\left(\frac{1}{n}\right)$$

while

$$\text{ave}\,\hat{R} = R + \frac{1-f}{n\bar{X}^2}(R\sigma^2_x - \rho\sigma_y\sigma_x) + o\left(\frac{1}{n}\right)$$

indicating that one should simply split the sample at random into r equal sized groups.

Regression Estimates In Cochran (1963), p 194 it is found that the regression estimate is unbiased if the relation is actually linear, while its variance is

$$\frac{1-f}{n}\sigma^2_y(1-\rho^2) + o\left(\frac{1}{n}\right)$$

and we proceed as in the case of the ratio estimate.

If the data is collected in a stratified, two stage etc. design the use of the jackknife mimics its use in the corresponding design of Section 2, for under appropriate regularity conditions the asymptotic variance formulae mimic the corresponding linear formulae.

It is interesting to note that situations where the variance formulae are quite different may well lead to the same procedure with the jackknife. For example consider the separate and combined ratio estimates. Cochran (1963), p.167-170, in stratified sampling.

In the first case the variance is

$$\Sigma\frac{N_h^2(1-f_h)}{n_h}(\sigma_{yh}^2 + R^2_h\sigma^2_{xh} - 2R_h\rho_h\sigma_{yh}\sigma_{xh}) + o\left(\frac{1}{n}\right)$$

and in the second

$$\Sigma_h\frac{N_h^2(1-f_h)}{n_h}(\sigma_{yh}^2 + R^2\sigma^2_{xh} - 2R\rho_h\sigma_{yh}\sigma_{xh}) + o\left(\frac{1}{n}\right)$$

We see that we are led to the identical jackknife calculations, splitting the sample in each stratum into r groups and then piecing together r global groups.

4. Concluding Remarks

So far nothing has been said about the choice of a value for r. One notes that one is estimating a variance and that the pseudo values q_{pi} are approximately independent; therefore values of r between 10 and 20 seem sensible.

Also we note that mathematical experimentation should be carried out with any estimator to which it is proposed to apply the jackknife in order to verify the inverse sample size variance relation required in the theorem. The error plots of Yates (1960) would appear to be useful in this context.

Finally we mention that the technique extends in an obvious manner to the calculation of variances in a multivariate situation.

Two additional references to the jackknife are Miller (1964) and Brillinger (1964).

REFERENCES

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. Int. Stat. Inst.* 3, 202-206.

COCHRAN, W. G. (1963). *Sampling Techniques*, New York, Wiley. 2nd Ed.

KISH, L. (1965). *Survey Sampling*. New York, Wiley.

MILLER, R. G. Jr. (1964). A trustworthy jackknife. *Ann. Math. Statist.* 35, 1594-1605.

TUKEY, J. W. (1958). Bias and confidence in not-quite large samples. Abstract. *Ann. Math. Statist.* 29, 614.

TUKEY, J. W. and MCLAUGHLIN, D. (1963). Less vulnerable confidence and significance procedures for location based on a single sample: trimming/winsorisation. I *Sankhya* A 25, 331-352.

YATES, F. (1960. *Sampling Methods for Censuses and Surveys*, 3rd Ed. London, Chas. Griffin and Company.