REVIEW OF THE INTERNATIONAL
STATISTICAL INSTITUTE
Volume 32 : 3, 1964

THE ASYMPTOTIC BEHAVIOUR OF TUKEY'S GENERAL METHOD OF SETTING APPROXIMATE CONFIDENCE LIMITS (THE JACKKNIFE) WHEN APPLIED TO MAXIMUM LIKELIHOOD ESTIMATES[1]

by

D. R. Brillinger

London School of Economics and Political Science[2]

1. INTRODUCTION

There are many situations in which it is possible to derive a simple estimate of some parameter but yet in which it is difficult to obtain an idea of the accuracy of the estimate. J. W. Tukey [4], [5] has suggested a procedure that may be used in a broad class of instances of the above type to obtain an approximate confidence interval for the parameter. The procedure may be used whenever a suitable form of repetition can be identified in the data, and basically involves the linear combination of an estimate based on all the data with a corresponding estimate based on parts thereof.

Tukey has called this procedure the Jackknife[3]. Quoting from [5], "The procedure described here shares two characteristics with a Boy Scout Jackknife:

(1) wide applicability to very many different problems, and
(2) inferiority to special tools for those problems for which special tools have been designed and built."

To be more explicit concerning the technique, suppose that the data is separated into r portions. Let y be the estimate based on all the data, $y_{(i)}$ (read "y not i") that based on all but the i^{th} portion, and y_{pi} (read "y pseudo i") defined by $y_{pi} = ry - (r-1)\, y_{(i)}$.

Previous to Tukey's work M. Quenouille [3] had suggested the use of the estimator $y_p = \Sigma y_{pi}/r$, rather than the estimator y, in view of its reduced bias property in certain situations. Being more explicit if,

$$Ey = \theta + A/n + o\left(\frac{1}{n}\right),$$

then

$$Ey_p = \theta + o\left(\frac{1}{n}\right).$$

Tukey goes further. In addition to suggesting the use of the estimator y_p, he suggests that Student's t may be applied to the y_{pi} as if they were independent observations, (which they generally are not), to obtain an approximate confidence interval. Specifically one may approximate the distribution of,

$$\frac{y_p - \theta}{\sqrt{\Sigma (y_{pi} - y_p)^2 / r(r-1)}}$$

by t on $(r-1)$ degrees of freedom. This is the Jackknife procedure.

---

[1] This research was carried out while the author had the support of a Research Training Fellowship of the Social Science Research Council.
[2] Presently at Bell Telephone Laboratories, Incorporated, and Princeton University.
[3] A large pocketknife with multipurpose blades.

Tukey has justified this procedure for the class of estimates that are functions of linear statistics. In this paper, the procedure is justified asymptotically, and the properties of the y_{pi} are investigated, for the case of a maximum likelihood estimate of a real parameter θ.

2. ASYMPTOTIC DISTRIBUTION

Let $x_1, x_2, \ldots, x_n$ be a sample from a distribution with density function $f(x, \theta)$. Subject to regularity conditions the maximum likelihood estimate $\hat{\theta}$ of θ based on the n observations satisfies,

$$\Sigma \frac{\partial \log f(x_i, \theta)}{\partial \theta} \Bigg|_{\theta = \hat{\theta}} = \Sigma \frac{\partial \log f(x_i, \theta)}{\partial \theta} + (\hat{\theta} - \theta) \, \Sigma \frac{\partial^2 \log f(x_i, \theta^*)}{\partial \theta^2} = 0,$$

where θ^* lies between $\hat{\theta}$ and θ.

Let $I = -E\left[\frac{\partial^2 \log f(x, \theta)}{\partial \theta^2}\right]$ be the Fisher information.

Now,

$$\sqrt{n}\,(\hat{\theta} - \theta) = \Sigma \frac{\partial \log f(x_i, \theta)}{\partial \theta} \Big/ \sqrt{n}\, I + y, \tag{1}$$

where y is a random variable that tends to zero in probability as n tends to infinity because $p \lim \theta^* = \theta$.

Suppose $n = rs$, r and s being integers. Divide the data into r groups, each containing s observations. Let $\hat{\theta}_j$ denote the maximum likelihood estimate derived from the $r\,(s-1)$ observations obtained by omitting the j^{th} group.

As above,

$$\sqrt{r\,(s-1)}\;(\hat{\theta}_j - \theta) = \sum_{(j)} \frac{\partial \log f(x_i, \theta)}{\partial \theta} \Big/ \sqrt{r\,(s-1)}\; I + y_j, \tag{2}$$

$\sum_{(j)}$ denoting the summation over all observations except those in the j^{th} group.

From (1) and (2)

$$y_{pj} - \theta = r \sum_j \frac{\partial \log f(x_i, \theta)}{\partial \theta} \Big/ nI + ry / \sqrt{n} - y_j \sqrt{r\,(r-1)/n}, \tag{3}$$

$\sum_j$ denoting summation over the j^{th} group. Thus

$$y_p - \theta = \Sigma \frac{\partial \log f(x_i, \theta)}{\partial \theta} \Big/ nI + ry / \sqrt{n} - \bar{y} \sqrt{r\,(r-1)/n}, \tag{4}$$

where $\bar{y} = (y_1 + \ldots + y_r)/r$.

If r is held constant while $n \to \infty$, it follows from the Central Limit Theorem applied to (4) that,

$$\sqrt{n}\,(y_p - \theta) \to N(0, 1/I)$$

since $ry - \sqrt{r\,(r-1)}\;\bar{y}$ tends to zero in probability.

It is possible to say even more than this. From (3), because the $ry - \sqrt{r\,(r-1)}\; y_j$ tend to zero in probability, the $\sqrt{n\,(y_{pj} - \theta)}$ tend to be independent $N(0, r/I)$ variates. This in turn implies, invoking a theorem in [8], that,

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

tends to t_{r-1}. An asymptotic justification of Tukey's procedure has consequently been derived for this case.

It must be noted that another, basically simpler technique, is available for the construction of approximate confidence intervals. Namely evaluate the r maximum likelihood estimates based on the s observations in each of the r subdivisions. These estimates are certainly independent and asymptotically normal. Consequently they may be used to construct a t statistic and an associated approximate confidence interval.

It is seen therefore that in order to carry out comparisons, terms of a higher order will have to be included in the expansions for $\hat{\theta}$, y_{pi} and y_p. In the next section these higher order terms will be investigated and the estimates $\hat{\theta}$ and y_p will be compared as regards asymptotic means, variances, mean-squared errors and skewnesses.

3. ASYMPTOTIC MEANS AND VARIANCES

Under regularity conditions, in a neighborhood of θ,

$$0 = \Sigma \frac{\partial \log f(x_j, \theta)}{\partial \theta} \Big/ n + (\hat{\theta} - \theta) \Sigma \frac{\partial^2 \log f(x_j, \theta)}{\partial \theta^2} \Big/ n + \frac{(\hat{\theta} - \theta)^2}{2n} \Sigma \frac{\partial^3 \log f(x_j, \theta^*)}{\partial \theta^3}, \tag{5}$$

where θ^* lies between $\hat{\theta}$ and θ.

Let

$$a_i = \Sigma \frac{\partial \log f(x_j, \theta)}{\partial \theta} \quad \text{over the } i^{\text{th}} \text{ group;}$$

$$b_i = \Sigma \left(\frac{\partial^2 \log f(x_j, \theta)}{\partial \theta^2} + I \right) \quad \text{" " " " ;}$$

$$c_i = \Sigma \left(\frac{\partial^3 \log f(x_j, \theta^*)}{\partial \theta^3} - J \right) \quad \text{" " " " ;}$$

where

$$J = E \left[\frac{\partial^3 \log f(x, \theta)}{\partial \theta^3} \right].$$

Inverting (5) about the origin,

$$\hat{\theta} - \theta = \Sigma a_i / nI + \Sigma a_i \Sigma b_i / n^2 I^2 + J (\Sigma a_i)^2 / 2n^2 I^3 + o \left[(\Sigma a_i)^2 / n^2 + (\Sigma b_i)^2 / n^2 + (\Sigma c_i)^2 / n^2 \right], \tag{6}$$

and similarly,

$$\hat{\theta}_j - \theta = \sum_{(j)} a_i / (n - n / r) I + \sum_{(j)} a_i \sum_{(j)} b_i / (n - n / r)^2 I^2 + J (\sum_{(j)} a_i)^2 / (n - n/r)^2 2 I^3 + o [\ldots].$$

Therefore,

$$y_{pj} - \theta = r a_j / nI + \{ r \Sigma a_i \Sigma b_i - \Sigma a_i \Sigma b_i r^2 / (r-1) + a_j \Sigma b_i r^2 / (r-1) + b_j \Sigma a_i r^2 / (r-1) - a_j b_j r^2 / (r-1) \} \frac{1}{n^2 I^2} + \ldots \tag{7}$$

And finally,

$$y_p - \theta = \Sigma a_i / nI + (\sum_{i \neq j} a_i b_j) r / n^2 I^2 (r-1) + J (\sum_{i \neq j} a_i a_j) r / 2n^2 I^3 (r-1) + o [\ldots]. \tag{8}$$

Because the expansions (6) and (8) are valid only in a neighborhood of the origin, they may not be used to find the true large sample means, variances etc. of $\hat{\theta} - \theta$ and $y_p - \theta$. At many points in the statistical literature writers have proceeded to find "large sample means" etc. from such local expansions without questioning the validity of the operation. One must however proceed cautiously.

A well-defined way of proceeding is the following. Let R be a bounded open interval containing the origin. The expected value of a random variable x_n truncated to the interval R may often be calculated to a certain order in $1/n$, and it may happen that this expected value is independent of the particular interval R selected, to the given order. If this occurs call $E(x_n \| R)$ the asymptotic mean of x_n about the origin, ($x_n \| R$ denotes the random variable x_n truncated to the interval R), and denote it by $\overrightarrow{\text{ave}}\, x_n$. (Note that $\overrightarrow{\text{ave}}\, x_n$ is only defined up to a certain order.) A case may be made and has been made elsewhere, [1], [6], [7] for the use of truncated characteristics of this form as more representative when considering the large sample properties of sequences of random variables. In the particular case under consideration it turns out that with regularity conditions similar to those in [2] the random variables $\hat{\theta}$ and y_p may be truncated effectively in the manner described above, and now the expansions (6) and (8) may be used to obtain the following asymptotic average values,

$$\overrightarrow{\text{ave}}\, \hat{\theta} = \theta + \mu_{11} / n I^2 + J / 2 n I^2 + o(1/n)$$

where,

$$\mu_{ij} = E\left[\left(\frac{\partial \log f(x, \theta)}{\partial \theta} \right)^i \left(\frac{\partial^2 \log f(x, \theta)}{\partial \theta^2} + I \right)^j \right]$$

$$\overrightarrow{\text{ave}}\, y_p = \theta + o(1/n).$$

Note the reduced bias of y_p.

$$\overrightarrow{\text{ave}}\, (\hat{\theta} - \theta)^2 = 1/nI + 2\mu_{21} / n^2 I^3 + J\mu_{30} / n^2 I^4$$
$$+ \mu_{02} / n^2 I^3 + 2(\mu_{11})^2 / n^2 I^4 + 3\mu_{11} J / n^2 I^4 + 3J^2 / 4 n^2 I^4 + o(1/n^2)$$
$$\overrightarrow{\text{ave}}\, (y_p - \theta)^2 = 1/nI + [I\mu_{02} + (\mu_{11})^2 + J^2/2 + 2J\mu_{11}]\, r / (r-1) n^2 I^4 \qquad (9)$$
$$+ o(1/n^2).$$

Note that the mean squared error of y_p will not always be smaller than that of $\hat{\theta}$, but in many situations it will.

$$\overrightarrow{\text{ave}}\, (y_p - \theta)^3 = \mu_{30}\, n^2 I^3 + 6\mu_{11}\, n^2 I^3 + 3J\, n^2 I^3 + o(1/n^2)$$

$$\overrightarrow{\text{ave}}\, (\hat{\theta} - \theta)^3 = \mu_{30} / n^2 I^3 + 9\mu_{11} / n^2 I^3 + 9J / n^2 I^3 2 + o(1/n^2)$$

Note that in many situations y_p will be closer to asymptotic normality than $\hat{\theta}$.

It is also instructive to see to what degree one can approximate $\overrightarrow{\text{var}}\, y_p$ by $\Sigma (y_p - y_{pi})^2 / r(r-1)$. Employing the expansion (7) one finds,

$$\overrightarrow{\text{ave}}\, \Sigma (y_{pi} - y_p)^2 / r(r-1) = 1/nI + 2r\mu_{21} / (r-1) n^2 I^3 + rJ\mu_{30} / (r-1) n^2 I^4$$
$$+ 2r(r-2)[I\mu_{02} + (\mu_{11})^2] / (r-1)^2 n^2 I^4 + 2r(r-2) J\mu_{11} / (r-1)^2 n^2 I^4$$
$$+ r(r-2) J^2 / (r-1)^2 n^2 I^4 + o(1/n^2).$$

This should be compared with expression (9), which gives $\overrightarrow{\text{var}}\, y_p$ to $o(1/n^2)$.

Concerning the comparison of the estimate y_p, and the estimated alluded to at the end of Section 2, i.e. the one based on the simple sample split, the latter estimate's mean etc. are easily obtained from the above formulas pertaining to the maximum likelihood estimate since the estimate is the mean of r independent maximum likelihood estimates each based on s observations. When this comparison is carried out, y_p is seen to be superior with respect to mean squared error and asymptotic normality in many situations.

REFERENCES

[1] Brillinger, D. R. (1961). *Asymptotic means and variances in the k-dimensional case.* (Ph. D. Thesis submitted to Princeton University.

[2] Brillinger, D. R. (1962). A note on the rate of convergence of a mean. *Biometrika*, 49, 574-576.

[3] Quenouille, M. H. (1956). Notes on bias in estimation. *Biometrika*, 43, 353-360.

[4] Tukey, J. W. (1958). Bias and confidence in not-quite large samples. Abstract. *Annals of Mathematical Statistics*, 29, 614.

[5] Tukey, J. W. (1959). Approximate confidence limits for most estimates. Unpublished manuscript.

[6] Tukey, J. W. (1960). A survey of sampling from contaminated distributions. *Contributions to Probability and Statistics*, Olkin and Others, Stanford University Press, 448-485.

[7] Tukey, J. W. (1961). Asymptotic means and variances. Unpublished manuscript.

[8] Mann, H. B. and Wald, A. (1943). On stochastic limit and order relationships. *Annals of Mathematical Statistics*, 14, 217-226.

[9] Brillinger, D. R. and Tukey, J. W. (1964). Asymptotic variances, moments, cumulants and other average values. Unpublished manuscript.

RESUME

L'objet de cet article est de donner une justification asymptotique d'une technique générale due à Tukey pour construire des intervalles de confiance. L'auteur traite surtout des estimateurs obtenus par la méthode du maximum de vraisemblance.

La technique est la suivante. Etant donné un échantillon, on le divise en r parties. Soit y un estimateur calculé à partir de l'échantillon total, $y_{(i)}$ des estimateurs calculés, à partir de cet échantillon où la partie i est supprimée, et $y_{pi} = ry - (r - 1) y_{(i)}$. La technique de Tukey consiste à construire une intervalle de confiance en considérant les y_{pi} comme des estimateurs indépendants du paramètre intéressé. L'auteur traite des caractéristiques de cette technique.

mean etc. are easily obtained from the above formulas pertaining to the maximum likelihood estimate since the estimate is the mean of r independent maximum likelihood estimates each based on s observations. When this comparison is carried out, y_p is seen to be superior with respect to mean squared error and asymptotic normality in many situations.

REFERENCES

[1] Brillinger, D. R. (1961). *Asymptotic means and variances in the k-dimensional case*. (Ph. D. Thesis submitted to Princeton University.

[2] Brillinger, D. R. (1962). A note on the rate of convergence of a mean. *Biometrika*, 49, 574-576.

[3] Quenouille, M. H. (1956). Notes on bias in estimation. *Biometrika*, 43, 353-360.

[4] Tukey, J. W. (1958). Bias and confidence in not-quite large samples. Abstract. *Annals of Mathematical Statistics*, 29, 614.

[5] Tukey, J. W. (1959). Approximate confidence limits for most estimates. Unpublished manuscript.

[6] Tukey, J. W. (1960). A survey of sampling from contaminated distributions. *Contributions to Probability and Statistics*, Olkin and Others, Stanford University Press, 448-485.

[7] Tukey, J. W. (1961). Asymptotic means and variances. Unpublished manuscript.

[8] Mann, H. B. and Wald, A. (1943). On stochastic limit and order relationships. *Annals of Mathematical Statistics*, 14, 217-226.

[9] Brillinger, D. R. and Tukey, J. W. (1964). Asymptotic variances, moments, cumulants and other average values. Unpublished manuscript.

RESUME

L'objet de cet article est de donner une justification asymptotique d'une technique générale due à Tukey pour construire des intervalles de confiance. L'auteur traite surtout des estimateurs obtenus par la méthode du maximum de vraisemblance.

La technique est la suivante. Etant donné un échantillon, on le divise en r parties. Soit y un estimateur calculé à partir de l'échantillon total, $y_{(i)}$ des estimateurs calculés, à partir de cet échantillon où la partie i est supprimée, et $y_{pi} = ry - (r-1)y_{(i)}$. La technique de Tukey consiste à construire une intervalle de confiance en considérant les y_{pi} comme des estimateurs indépendants du paramètre intéressé. L'auteur traite des caractéristiques de cette technique.