The Annals of Mathematical Statistics
1969, Vol. 40, No. 6, 2076–2100

JACKKNIFING *U*-STATISTICS[1]

By James N. Arvesen

Purdue University

1. Introduction. Several recent papers have shown the versatility of a relatively new estimation procedure. The *jackknife* was originally proposed by Quenouille in [17], and later expanded on by the same author in [18]. Subsequently, Tukey [15] proposed the use of it to obtain approximate t-statistics to be used in testing or construction of confidence intervals. Two papers of Miller [13], and [14] are rigorous justifications of situations where Tukey's method proves valid (as well as situations where it is grossly invalid).

Let θ be an unknown parameter, and let $X_1, \cdots, X_N$ be N independent, identically distributed observations from the cdf F_θ. The essence of the jackknife is to divide the N observations into n groups of k observations each ($N = nk$). Let $\hat{\theta}_n{}^0$ be the estimate of θ based on all N observations, and $\hat{\theta}^i_{n-1}$, $i = 1, \cdots, n$, denote the estimate obtained after deletion of the ith group of observations.

Let

$$(1) \qquad \hat{\theta}_i = n\hat{\theta}_n{}^0 - (n-1)\hat{\theta}^i_{n-1}, \qquad i = 1, \cdots, n.$$

These are called pseudo-values by Tukey. Then the jackknife estimate of θ is the average of the $\hat{\theta}_i$, $i = 1, \cdots, n$,

$$(2) \qquad \hat{\theta} = n^{-1} \textstyle\sum_{i=1}^n \hat{\theta}_i.$$

When originally proposed, Quenouille considered $n = 2$, and found that the technique eliminated the $1/N$ term from any bias. This result holds for all values of n, for if

$$(3) \qquad E(\hat{\theta}_n{}^0) = \theta + a/kn + b/(kn)^2 + \cdots$$

one can show that

$$(4) \qquad E(\hat{\theta}) = \theta - b/[k^2 n(n-1)] + \cdots.$$

Perhaps more importantly, Tukey has proposed that in most instances $\hat{\theta}_1, \cdots, \hat{\theta}_n$ can be treated as n approximately *independent*, identically distributed observations from which an approximate t_{n-1} statistic can be constructed (note that if $X_1, \cdots, X_N$ are independent, identically distributed random variables, then $\hat{\theta}_1, \cdots, \hat{\theta}_n$ are interchangeable random variables for each n). Equivalently, Tukey conjectured that

$$(5) \qquad n^{\frac{1}{2}}(\hat{\theta} - \theta)/((n-1)^{-1} \textstyle\sum_{i=1}^n (\hat{\theta}_i - \hat{\theta})^2)^{\frac{1}{2}}$$

Received 3 February 1969.

[1] Adapted from the author's doctoral dissertation at Stanford University. The research was supported in part by Public Health Service Grant USPHS-2T1 GM 25-11.