The Annals of Mathematical Statistics
1972, Vol. 43, No. 1, 1–30

ON THE JACKKNIFE STATISTIC, ITS EXTENSIONS, AND ITS RELATION TO e_n-TRANSFORMATIONS

By H. L. Gray, T. A. Watkins and J. E. Adams

Texas Tech University, Angelo State University and Stephen F. Austin University

In this paper, a complete overview is given of the theoretical development of various estimators generated by the jackknife statistic. In particular, the jackknife method is extended to stochastic processes by means of two estimators referred to as the J_∞-estimator and the $J_\infty^{(2)}$-estimator. These estimators are studied in some detail and shown to have the same properties as the jackknife when one considers the length of the process record as the sample size. Finally, it is shown that the entire development of the jackknife procedures discussed in this paper can be considered as a direct parallel of earlier developments in numerical analysis surrounding the study of a transformation referred to as the e_n-transformation.

1. Introduction. In the last several years, the jackknife statistic has been and still remains a topic of some interest in the literature. Recently in [35] the "classical" jackknife was extended to a more general type of estimator which was referred to as the generalized jackknife. In that paper it was shown how one could incorporate additional information into the jackknife procedure to enhance its bias reduction properties without destroying its asymptotic properties. Although the results included in [35] are of interest and should be of value to the statistician there is possibly a more interesting, as well as useful, facet of that extension which should be pointed out. That is, it should be observed that the jackknife and generalized jackknife have an exact counterpart in the deterministic realm, i.e., in numerical analysis. In fact it was this counterpart which suggested the generalized jackknife and hence an awareness of the corresponding developments in that area have already proved their worth.

In this paper the jackknife technique is extended in such a manner that the analogy to previous work in numerical analysis is complete and the sense in which the two developments are exact parallels is discussed. In order to accomplish this, the jackknife method is first extended to appropriate stochastic processes and is shown to have some value in that area. The parallel of the current stages of the two developments is then shown to be complete. Before proceeding to these comparisons let us first consider briefly some history regarding the jackknife and the generalized jackknife.

Received March 30, 1971.

2. The jackknife method.

DEFINITION 2.1. Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be any two estimators for θ defined on a random sample $X_1, X_2, \cdots, X_n$. Then for any real number $R \neq 1$ we define the generalized jackknife $G(\hat{\theta}_1, \hat{\theta}_2)$, by the equation

$$G(\hat{\theta}_1, \hat{\theta}_2) = \frac{\hat{\theta}_1 - R\hat{\theta}_2}{1 - R} . \tag{2.1}$$

The function G in (2.1) was first introduced in [35] by Schucany, Gray, and Owen (1970) where it was observed that if

$$E[\hat{\theta}_1] = \theta + b_1(n, \theta) , \tag{2.2}$$

$$E[\hat{\theta}_2] = \theta + b_2(n, \theta) , \tag{2.3}$$

and $b_2(n, \theta) \neq 0$, then

$$E[G(\hat{\theta}_1, \hat{\theta}_2)] = \theta \tag{2.4}$$

when

$$R = \frac{b_1(n, \theta)}{b_2(n, \theta)} . \tag{2.5}$$

In that same paper it was also pointed out that in many cases $b_1(n, \theta)$ will be of the form $f(n)b(\theta)$ and, by proper selection of $\hat{\theta}_2$, $b_2(n, \theta)$ will be of the form $f(n - 1)b(\theta)$, so that if

$$R = \frac{f(n)}{f(n - 1)} , \tag{2.6}$$

$G(\hat{\theta}_1, \hat{\theta}_2)$ is an unbiased estimator for θ.

Although essentially no restriction is placed on $\hat{\theta}_1$ and $\hat{\theta}_2$ in Definition 2.1 the only situation in which any significant theory for $G(\hat{\theta}_1, \hat{\theta}_2)$ has been developed is when $\hat{\theta}_2$ is the estimator obtained by restricting $\hat{\theta}_1$ to subsamples of size $n - 1$ and averaging. That is, by letting

$$\hat{\theta}_1 = \hat{\theta}_1(X_1, X_2, \cdots, X_n) , \tag{2.7}$$

$$\hat{\theta}_1^i = \hat{\theta}_1(X_1, \cdots, X_{i-1}, X_{i+1}, \cdots, X_n) \tag{2.8}$$

and

$$\hat{\theta}_2 = n^{-1} \textstyle\sum_{i=1}^n \hat{\theta}_1^i = \overline{\hat{\theta}_1^i} . \tag{2.9}$$

Actually one can form the above estimators on subsamples of size $n - m$. However, this increased generality is of no interest here and we shall limit ourselves to the case $m = 1$.

For a given estimator $\hat{\theta}$, the particular form of (2.1) studied in some detail in [35] is defined by

$$G(\hat{\theta}) = \frac{\hat{\theta} - R\overline{\hat{\theta}^i}}{1 - R}, \tag{2.10}$$

where we have now removed the subscripts from our notation to stress that we are referring to the case in which the second estimator in (2.1) is completely determined by the first by means of (2.9). As is obvious, when $R = (n-1)/n$, (2.10) reduces to the ordinary jackknife (first introduced by Quenouille (1956)) which we shall denote by $J(\hat{\theta})$. Thus

$$J(\hat{\theta}) = \frac{\hat{\theta} - \frac{n-1}{n}\overline{\hat{\theta}^i}}{1 - \frac{n-1}{n}} \doteq n\hat{\theta} - (n-1)\overline{\hat{\theta}^i}, \tag{2.11}$$

and if $b(n, \theta) = C(\theta)/n$, $J(\hat{\theta})$ is unbiased since in that event R is the ratio of the biases in $\hat{\theta}$ and $\overline{\hat{\theta}^i}$.

The notions set forth above have an immediate extension to the problem of removing additional bias terms in the representation of $b(n, \theta)$. That is, let $\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_{k+1}$ be $k+1$ estimators defined on a random sample of size n. Then we define $G(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, \cdots, \hat{\theta}_{k+1})$ by the equation

$$G(\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_{k+1}) = \frac{\begin{vmatrix} \hat{\theta}_1 & \hat{\theta}_2 & \cdots & \hat{\theta}_{k+1} \\ a_{11} & a_{12} & \cdots & a_{1,k+1} \\ \vdots & & & \\ a_{k1} & a_{k2} & \cdots & a_{k,k+1} \end{vmatrix}}{\begin{vmatrix} 1 & 1 & \cdots & 1 \\ a_{11} & a_{12} & \cdots & a_{1,k+1} \\ \vdots & & & \\ a_{k1} & a_{k2} & \cdots & a_{k,k+1} \end{vmatrix}}, \tag{2.12}$$

where the a_{ij} are any real numbers for which the denominator in (2.12) is not zero. Taking $R = a_{11}/a_{12}$ and trivial algebra shows that when $k = 1$, (2.12) reduces to (2.1). Moreover it is also clear that when

$$E[\hat{\theta}_j] = \theta + \textstyle\sum_{i=1}^{k} f_{ij}(n) b_i(\theta) \tag{2.13}$$

and $G(\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_{k+1})$ is defined, taking $a_{ij} = f_{ij}(n)$ yields

$$E[G(\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_{k+1})] = \theta. \tag{2.14}$$

When the estimators $\hat{\theta}_2, \hat{\theta}_3, \cdots, \hat{\theta}_{k+1}$, in (2.12) are defined in a manner analogous to (2.9), i.e., $\hat{\theta}_2$ is defined by restricting $\hat{\theta}_1$ to subsamples of size $n-1$ and averaging; $\hat{\theta}_3$ is defined by restricting $\hat{\theta}_1$ to subsamples of size $n-2$, and averaging etc., we will adopt the notation $G^{(k)}(\hat{\theta}_1)$ in place of $G(\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_{k+1})$. When this is done it has been shown that a natural extension to what might be referred to as an iterated jackknife results by taking $a_{ij} = (n - j + 1)^{-1}$. When the a_{ij} are so chosen we will use the special notation $J^{(k)}(\hat{\theta}_1)$, and clearly $J^{(1)}(\hat{\theta}_1) = J(\hat{\theta}_1)$. It is trivial to see that if

$$E[\hat{\theta}_1] = \theta + \frac{b_1(\theta)}{n} + \frac{b_2(\theta)}{n^2} + \cdots + \frac{b_k(\theta)}{n^k}, \tag{2.15}$$

then $E[J^{(k)}(\hat{\theta}_1)] = \theta$, which is the justification for considering $J^{(k)}(\hat{\theta}_1)$ as the proper notion for a "higher order" or iterated jackknife.

As is well known, probably the most important properties of $J[\hat{\theta}]$ are not its bias reduction properties but its asymptotic distribution properties. See Arvesen (1969), Brillinger (1964), or Miller (1964). Although, not so well known, the same is true for $G^{(k)}(\hat{\theta})$, when $k = 1$, and no results in this area are available for larger values of k. We will not expand on this here, for the properties to which we are alluding will become apparent in Section 6.

We are now in a position to consider the notion of jackknifing continuous data which in some sense retains the character of a random sample. We will not, however, be so ambitious as to attempt to consider such processes as white noise processes but shall limit our study to those processes which are at least piecewise continuous. This is the subject of the next few sections.

3. The J_n-estimator. As was mentioned above we will limit our discussion to stochastic processes which are piecewise continuous. There is, however, another restriction which we will find it necessary to make. Namely, we will only consider estimators which are functions of an estimator $\hat{\theta}$, $\hat{\theta}$ having the structure indicated below. This is possibly unnecessary but it is the only tractable way in which these authors have been able to construct the desired development.

DEFINITION 3.1. Let $\{G(t) \mid t \in S\}$ be a stochastic process defined over an index set S containing the interval $[a, b]$, and suppose that the probability law of $G(t)$ depends on θ for every $t \in [a, b]$. Then for $t_1, t_2 \in [a, b]$ and $t_1 \neq t_2$, we define $\hat{\theta}(t_1, t_2)$ to be an estimator for θ of the form

$$\hat{\theta}(t_1, t_2) = \frac{I_G(t_2) - I_G(t_1)}{t_2 - t_1}, \tag{3.1}$$

where $\{I_G(t) \mid t \in [a, b]\}$ is a stochastic process, determined by the process $\{G(t) \mid t \in [a, b]\}$, such that almost every realization is piecewise continuous.

To illustrate the above definition we include the following simple examples, the first of which simply points out that for a regular partition, $\hat{\theta}(a, b)$ is the sample mean over the $\hat{\theta}(t_i, t_{i-1})$.

EXAMPLE 1. Let $a = t_0 < t_1 < \cdots < t_n = b$, and let $t_i - t_{i-1} = (b - a)/n$.

Then

$$\hat{\theta}(a, b) = n^{-1} \sum_{i=1}^{n} \hat{\theta}(t_{i-1}, t_i) . \tag{3.2}$$

EXAMPLE 2. Let $\{N(t) \mid t \in [0, T]\}$ be a Poisson process with parameter λ. Then the maximum likelihood estimator, based on the interval $[0, T]$, for λ, is $\lambda = N(T)/T$. Thus in the notation of Definition 3.1,

$$G(t) = N(t) , \qquad I_G(t) = N(t) , \tag{3.3}$$

$$\theta(t_{i-1}, t_i) = \frac{N(t_i) - N(t_{i-1})}{t_i - t_{i-1}} \quad \text{and} \quad \hat{\theta}(0, T) = \frac{N(T)}{T} . \tag{3.4}$$

EXAMPLE 3. Let $E[G(t)] \equiv \theta$. Then a familiar estimator for θ based on the interval $[a, b]$ is

$$\hat{\theta}(a, b) = (b - a)^{-1} \int_a^b G(x)\, dx . \tag{3.5}$$

Thus in the terminology of Definition 3.1

$$I_G(t) = \int_a^t G(x)\, dx . \tag{3.6}$$

We now give a definition which extends the jackknife notions to estimators which are functions of $\hat{\theta}$.

DEFINITION 3.2. Let $\hat{\theta}(a, b)$ be defined by Definition 3.1 and let $a = t_0 < t_1 < t_2 < \cdots < t_n = b$ be a regular partition of the interval $[a, b]$, i.e. $(t_i - t_{i-1}) = (b - a)/n$. Then for any real valued function f, we define the estimator $J_n[f(\hat{\theta})]$ by the following equation:

$$J_n(f(\hat{\theta})) = nf(\hat{\theta}) - \frac{(n-1)}{n} \sum_{i=1}^{n} f(\hat{\theta}_n{}^i) , \tag{3.7}$$

where

$$\hat{\theta} = \hat{\theta}(a, b) , \qquad \hat{\theta}_i = \hat{\theta}(t_{i-1}, t_i) \tag{3.8}$$

and

$$\hat{\theta}_n{}^i = \frac{n}{n-1} \hat{\theta} - \frac{1}{n-1} \hat{\theta}_i . \tag{3.9}$$

One should note that when the process $I_G(t)$ has stationary independent increments, (3.7) is the classical jackknife obtained by *considering* $\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_n$ a random sample. However the notions are not identical as is quickly seen by observing that $J_n[f(\hat{\theta})]$ has the undesirable property that two different users could get different results from the same data simply by choosing different partition sizes for the interval $[a, b]$. This problem, of course, does not arise in the estimator $J[\hat{\theta}]$, pointing out a distinction between the roles of the parameter n in the two estimators.

Although $J_n[f(\hat{\theta})]$ has only been considered in the literature for the particular process of Example 2 with f the reliability function, (Gaver and Hoel (1970)) this appears to be sufficient to indicate the appropriate means of eliminating its dependence on n. That is, in [15] it was suggested that the bias in $J_n[e^{-\hat{\lambda}x}]$ is a decreasing function n, and this immediately suggests considering the estimator which one would obtain from $\lim_{n\to\infty} J_n[\hat{\theta}]$ with a and b held fixed. This is the purpose of the next section where we shall first obtain this limit and then justify our efforts.

4. The J_∞-estimator. Let $\{I_G(t) \mid t \in [a, b]\}$ be defined as in Definition 3.1 and let the set $\mathscr{G}$ and $\mathscr{G}'$ be sets of realizations defined by

$$\mathscr{G} = \{g \mid g \text{ is a realization of } \{G(t) \mid t \in S \supset [a, b]\}\}, \tag{4.1}$$

and

$$\mathscr{G}' = \{g \mid g \in \mathscr{G}, I_g \text{ is piecewise continuous on } [a, b]\}, \tag{4.2}$$

where I_g is the realization of $\{I_G(t) \mid t \in [a, b]\}$ determined by g. Then for each $g \in \mathscr{G}'$ we define

$$H_g(a) = I_g(a^+) - I_g(a), \qquad H_g(b) = I_g(b) - I_g(b^-), \tag{4.3}$$

and

$$H_g(t) = I_g(t^+) - I_g(t^-) \tag{4.4}$$

for each $t \in (a, b)$, where $+$ and $-$ indicate limits from the right and left respectively. Also, for $g \notin \mathscr{G}'$ we define

$$H_g(t) = 0 \tag{4.5}$$

for each $t \in [a, b]$. It is clear that for each $t \in [a, b]$, the functional $H_G(t)$, defined on $\mathscr{G}$ by (4.3), (4.4), and (4.5), is a random variable since

$$\begin{aligned} &\{g \mid g \in \mathscr{G}', H_g(t) \in [x_1, x_2]\} \\ &\quad = \textstyle\bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} \{g \mid g \in \mathscr{G}', I_g(t + n^{-1}) - I_g(t - n^{-1}) \\ &\quad \in [x_1 - m^{-1}, x_2 + m^{-1}]\}, \end{aligned} \tag{4.6}$$

where $[x_1, x_2]$ is an arbitrary, closed interval.

Now let R_t denote the range of $H_G(t)$ and let Γ be the set of possible values of jumps in I_G, i.e.,

$$\Gamma = \textstyle\bigcup_{t\in[a,b]} R_t - \{0\}. \tag{4.7}$$

Also, for each $\gamma \in \Gamma$, let N_γ be defined on $\mathscr{G}$ by

$$N_\gamma(g) = \mathscr{C}(\{t \mid t \in [a, b], H_g(t) = \gamma\}), \tag{4.8}$$

where $\mathscr{C}(\cdot)$ denotes cardinality, and let N be defined by

$$N(g) = \textstyle\sum_{\gamma\in\Gamma} N_\gamma(g), \tag{4.9}$$

where by $\gamma \in \Gamma$ we mean the summation is over all γ such that $N_\gamma(g) \neq 0$. Thus we note that $N_\gamma(g)$ is simply the number of points in $[a, b]$ where I_g has a jump of size γ and that $N(g)$ is the total number of discontinuities in I_g on $[a, b]$. We should also observe that if $g \in \mathscr{G}'$, $N(g)$ is finite so that only a finite number of the $N_\gamma(g)$'s are nonzero, and if $g \in \mathscr{G} - \mathscr{G}'$, $N(g) = 0$, so that the series in (4.9) is always finite.

We are now prepared to define the estimator which, we will shortly show, under appropriate conditions, is the limit of $J_n(f(\hat{\theta}))$.

Definition 4.1. Let $\hat{\theta}$ be defined over the interval $[a, b]$ as in Definition 3.1, and let N and N_γ be random variables defined by (4.9) and (4.8) respectively. Then if f is a real valued function defined and differentiable on the range of $\hat{\theta}$, we define the estimator $J_\infty[f(\hat{\theta})]$ by

$$J_\infty[f(\hat{\theta})] = f(\hat{\theta}) - \sum_{\gamma \in \Gamma} N_\gamma \left[f\left(\hat{\theta} - \frac{\gamma}{T}\right) - f(\hat{\theta}) + \frac{\gamma}{T} f'(\hat{\theta}) \right], \tag{4.10}$$

where $T = b - a$ and

$$f'(\hat{\theta}) = \left. \frac{df(\theta)}{d\theta} \right|_{\theta = \hat{\theta}} . \tag{4.11}$$

Before establishing the conditions under which $\lim_{n \to \infty} J_n[f(\hat{\theta})] = J_\infty[f(\hat{\theta})]$ let us consider some simplifications of (4.10) and a simple example.

Case 1. Almost every realization of $\{I_G(t) \mid t \in [a, b]\}$ is a step function. In this event

$$\sum_{\gamma \in \Gamma} N_\gamma \frac{\gamma}{T} = \hat{\theta} \tag{4.12}$$

and

$$J_\infty[f(\hat{\theta})] = f(\hat{\theta}) - \hat{\theta} f'(\hat{\theta}) - \sum_{\gamma \in \Gamma} N_\gamma \left[f\left(\hat{\theta} - \frac{\gamma}{T}\right) - f(\hat{\theta}) \right]. \tag{4.13}$$

Case 2. If $\Gamma = \{\gamma_0\}$, then

$$J_\infty[f(\hat{\theta})] = f(\hat{\theta}) - N\left[f\left(\hat{\theta} - \frac{\gamma_0}{T}\right) - f(\hat{\theta}) + \frac{\gamma_0}{T} f'(\hat{\theta}) \right], \tag{4.14}$$

where N is defined by the total number of discontinuities in I_g on $[a, b]$.

Case 3. Suppose Case 1 and Case 2 hold, then

$$J_\infty[f(\hat{\theta})] = f(\hat{\theta}) - \hat{\theta} f'(\hat{\theta}) - N\left[f\left(\hat{\theta} - \frac{\gamma_0}{T}\right) - f(\hat{\theta}) \right]. \tag{4.15}$$

Case 4. If almost every realization of $\{I_G(t) \mid t \in [a, b]\}$ is continuous, then

$$J_\infty[f(\hat{\theta})] = f(\hat{\theta}) . \tag{4.16}$$

An illustrative example is now in order.

EXAMPLE 1. Consider once again the Poisson process of Example 2, Section 3 with $f(\lambda) = e^{-\lambda x}$, $x > 0$. Taking $\hat{\lambda} = N(T)/T$ as before, and noting that $\Gamma = \{1\}$, and $f'(\lambda) = -xe^{-\lambda x}$ we have, by Case 3,

$$J_\infty[f(\hat{\lambda})] = e^{-\hat{\lambda}x}\left\{1 - N(T)\left[e^{x/T} - 1 - \frac{x}{T}\right]\right\}. \tag{4.17}$$

The estimator obtained in (4.17) was first introduced in [15] where it was investigated with regard to its robustness to violations of the Poisson assumption. We will not discuss it further at the present since we have included it at this point to simply demonstrate the terminology. We now establish the sense in which $J_\infty[f(\hat{\theta})]$ is the limit of $J_n[f(\hat{\theta})]$.

THEOREM 4.1. *Let the stochastic process* $\{I_G(t) \mid t \in [a, b]\}$ *be as in Definition* 3.1 *and also assume that almost every realization is of bounded variation on* $[a, b]$. *In addition suppose that for each* $t \in [a, b]$

$$P[H_G(t) = 0] = 1,$$

where $H_G(t)$ *is defined by* (4.3), (4.4), *and* (4.5). *Then as* $n \to \infty$, $J_n[f(\hat{\theta})]$ *converges to* $J_\infty[f(\hat{\theta})]$ *with probability one.*

PROOF. Let

$$A = \bigcup_{n=1}^{\infty}\left\{a + \frac{m}{n}(b-a) \,\middle|\, m \text{ is an integer with } 0 \le m \le n\right\}. \tag{4.18}$$

Then, since A is a countable set, it is clear that

$$P[H_G(t) \equiv 0 \quad \text{for} \quad t \in A] = 1. \tag{4.19}$$

Now let $\mathscr{T}$ denote the set of all realizations of $\{G(t) \mid t \in S\}$ such that for each $g \in \mathscr{T}$

(i) I_g is piecewise continuous on $[a, b]$,
(ii) I_g is of bounded variation on $[a, b]$, and
(iii) $H_g(t) = 0$ for every $t \in A$,

where I_g is the realization of $\{I_G(t) \mid t \in [a, b]\}$ determined by g. Then since $\{I_G(t) \mid t \in [a, b]\}$ is piecewise continuous and of bounded variation on $[a, b]$ and (4.19) holds, it is easy to see that $\mathscr{T}$ is a set of realizations which has probability one. Thus, it is sufficient to show that $J_n(f(\hat{\theta}))|_g$ converges to $J_\infty(f(\hat{\theta}))|_g$ for each $g \in \mathscr{T}$.

Now let $g \in \mathscr{T}$, $t \in [a, b]$, $\alpha(t) = \{x : H_g(x) \neq 0, a \le x \le t\}$, and S_g and C_g be defined on $[a, b]$ by

$$S_g(t) = \textstyle\sum_{\alpha(t)} H_g(x), \tag{4.20}$$

and

$$C_g(t) = I_g(t^+) - S_g(t). \tag{4.21}$$

Obviously $S_g(t)$ is finite, since H_g is nonzero at only a finite number of points of $[a, b]$, and it is clear that S_g is a step function. Also observe that C_g is continuous, and therefore uniformly continuous, on $[a, b]$. Also, for $x, y \in A$,

$$I_g(x) - I_g(y) = C_g(x) + S_g(x) - S_g(y) - C_g(y) . \tag{4.22}$$

Let $a = t_0 < t_1 < \cdots < t_n = b$, with $t_i - t_{i-1} = T/n$. Then, given $\varepsilon > 0$, there exists a positive integer n_ε, independent of t_i, such that for $n > n_\varepsilon$

$$|C_g(t_i) - C_g(t_{i-1})| < \varepsilon , \tag{4.23}$$

and

$$\mathscr{C}(\{t \mid t \in [t_{i-1}, t_i], H_g(t) \neq 0\}) \leq 1 , \tag{4.24}$$

where $\mathscr{C}(\cdot)$ denotes cardinality.

Also, let i_0 and i_γ be defined by

$$i_0 = \{i \mid H_g(t) \equiv 0 \quad \text{for} \quad t \in [t_{i-1}, t_i]\} , \tag{4.25}$$

and

$$i_\gamma = \{i \mid H_g(t) = \gamma \neq 0 \quad \text{for some} \quad t \in [t_{i-1}, t_i]\} . \tag{4.26}$$

Then for $n > n_\varepsilon$, the sets defined by (4.25) and (4.26) are mutually disjoint sets whose union is $\{1, 2, \cdots, n\}$ and we have (using the same notation for the estimate as the estimator)

$$\begin{aligned} J_n[f(\hat\theta)] &= f(\hat\theta) + \frac{n-1}{n} \textstyle\sum_{i=1}^n [f(\hat\theta) - f(\hat\theta_n{}^i)] \\ &= f(\hat\theta) + \frac{n-1}{n} \textstyle\sum_{i \in i_0} [f(\hat\theta) - f(\hat\theta_n{}^i)] \\ &\quad + \frac{n-1}{n} \textstyle\sum_{\gamma \in \Gamma} \sum_{i \in i_\gamma} [f(\hat\theta) - f(\hat\theta_n{}^i)] \\ &= f(\hat\theta) + \frac{n-1}{n} \textstyle\sum_{i \in i_0} \Big[f(\hat\theta) - f\Big(\hat\theta - \frac{n}{n-1} \frac{\Delta C_g(t_i)}{T} + \frac{1}{n-1} \hat\theta\Big)\Big] \\ &\quad + \frac{n-1}{n} \textstyle\sum_{\gamma \in \Gamma} \sum_{i \in i_\gamma} \Big[f(\hat\theta) \\ &\quad - f\Big(\hat\theta - \frac{n}{n-1} \frac{\gamma}{T} - \frac{n}{n-1} \frac{\Delta C_g(t_i)}{T} + \frac{1}{n-1} \hat\theta\Big)\Big] , \end{aligned} \tag{4.27}$$

where $\Delta C_g(t_i) = C_g(t_i) - C_g(t_{i-1})$. Now since $t_i \in A$ we have $N_\gamma(g) = \mathscr{C}(i_\gamma)$ for all $n > n_\varepsilon$, and

$$\begin{aligned} \lim_{n\to\infty} \textstyle\sum_{i \in i_\gamma} \Big[f(\hat\theta) - f\Big(\hat\theta - \frac{n}{n-1} \frac{\gamma}{T} - \frac{n}{n-1} \frac{\Delta C_g(t_i)}{T} + \frac{1}{n-1} \hat\theta\Big)\Big] \\ = N_\gamma(g) \Big[f(\hat\theta) - f\Big(\hat\theta - \frac{\gamma}{T}\Big)\Big] . \end{aligned} \tag{4.28}$$

Thus, since only a finite number of the $N_\gamma(g)$'s are nonzero,

$$(4.29)\quad \lim_{n\to\infty} \frac{n-1}{n} \sum_{\gamma\in\Gamma} \sum_{i\in i_\gamma} [f(\hat\theta) - f(\hat\theta_n{}^i)] = \sum_{\gamma\in\Gamma} N_\gamma(g)\left[f(\hat\theta) - f\left(\hat\theta - \frac{\gamma}{T}\right)\right],$$

the latter series of course being finite, since $g \in \mathscr{T}$. Now

$$(4.30)\quad \frac{n-1}{n} \sum_{i\in i_0}\left[f(\hat\theta) - f\left(\hat\theta - \frac{n}{n-1}\frac{\Delta C_g(t_i)}{T} + \frac{1}{n-1}\hat\theta\right)\right]$$

$$= -\frac{1}{n}\hat\theta \sum_{i\in i_0} \frac{f(\hat\theta) - f\left(\hat\theta - \frac{n}{n-1}\frac{\Delta C_g(t_i)}{T} + \frac{1}{n-1}\hat\theta\right)}{\frac{n}{n-1}\frac{\Delta C_g(t_i)}{T} - \frac{1}{n-1}\hat\theta}$$

$$+ \sum_{i\in i_0}\left[\frac{f(\hat\theta) - f\left(\hat\theta - \frac{n}{n-1}\frac{\Delta C_g(t_i)}{T} + \frac{1}{n-1}\hat\theta\right)}{\frac{n}{n-1}\frac{\Delta C_g(t_i)}{T} - \frac{1}{n-1}\hat\theta}\right]\frac{\Delta C_g(t_i)}{T}.$$

Since, for $n > n_\varepsilon$,

$$(4.31)\quad \left|\frac{n}{n-1}\frac{\Delta C_g(t_i)}{T} - \frac{1}{n-1}\hat\theta\right| < \frac{n}{n-1}\frac{\varepsilon}{T} + \frac{|\hat\theta|}{n-1},$$

and $|\theta|/(n-1)$ can be made arbitrarily small by choosing n sufficiently large, it is clear that, given $\delta > 0$,

$$(4.32)\quad \lim_{n\to\infty}\left|\frac{1}{n}\hat\theta \sum_{i\in i_0}\left[\frac{f(\hat\theta) - f\left(\hat\theta - \frac{n}{n-1}\frac{\Delta C_g(t_i)}{T} + \frac{1}{n-1}\hat\theta\right)}{\frac{n}{n-1}\frac{\Delta C_g(t_i)}{T} - \frac{1}{n-1}\hat\theta} - f'(\hat\theta)\right]\right|$$

$$\le \lim_{n\to\infty} |\hat\theta|\delta\frac{n - N(g)}{n} = |\hat\theta|\delta,$$

and therefore,

$$(4.33)\quad \lim_{n\to\infty} -\frac{\hat\theta}{n} \sum_{i\in i_0} \frac{f(\hat\theta) - f\left(\hat\theta - \frac{n}{n-1}\frac{\Delta C_g(t_i)}{T} + \frac{1}{n-1}\hat\theta\right)}{\frac{n}{n-1}\frac{\Delta C_g(t_i)}{T} - \frac{1}{n-1}\hat\theta} = -\hat\theta f'(\hat\theta).$$

Similarly,

$$(4.34)\quad \lim_{n\to\infty}\left|\sum_{i\in i_0}\left[\frac{f(\hat\theta) - f\left(\hat\theta - \frac{n}{n-1}\frac{\Delta C_g(t_i)}{T} + \frac{1}{n-1}\hat\theta\right)}{\frac{n}{n-1}\frac{\Delta C_g(t_i)}{T} - \frac{1}{n-1}\hat\theta} - f'(\hat\theta)\right]\frac{\Delta C_g(t_i)}{T}\right|$$

$$\le \lim_{n\to\infty} \frac{\delta}{T}\sum_{i\in i_0} |\Delta C_g(t_i)| \le \frac{\delta}{T} V(I_g),$$

where $V(I_g)$ denotes the total variation of I_g on $[a, b]$. Hence,

$$\lim_{n\to\infty} \sum_{i\in i_0} \left[\frac{f(\hat{\theta}) - f\left(\hat{\theta} - \frac{n}{n-1}\frac{\Delta C_g(t_i)}{T} + \frac{1}{n-1}\hat{\theta}\right)}{\frac{n}{n-1}\frac{\Delta C_g(t_i)}{T} - \frac{1}{n-1}\hat{\theta}} \right]$$

$$\begin{aligned}
&= \frac{f'(\hat{\theta})}{T} \lim_{n\to\infty} \sum_{i\in i_0} \Delta C_g(t_i) \\
(4.35)\qquad &= \frac{f'(\hat{\theta})}{T} \lim_{n\to\infty} [C_g(b) - C_g(a) - \sum_{\gamma\in\Gamma} \sum_{i\in i_\gamma} \Delta C_g(t_i)] \\
&= \frac{f'(\hat{\theta})}{T} [C_g(b) - C_g(a)] \\
&= f'(\hat{\theta}) \left[\hat{\theta} - \sum_{\gamma\in\Gamma} N_\gamma(g) \frac{\gamma}{T}\right],
\end{aligned}$$

since $N_\gamma(g)$ is finite for each $\gamma \in \Gamma$ and only a finite number of the $N_\gamma(g)$'s are nonzero, and $a,b \in A$.

Thus, combining the results of (4.29), (4.33) and (4.35), we obtain

$$(4.36)\qquad \begin{aligned} &\lim_{n\to\infty} J_n(f(\hat{\theta})) \\ &\quad = f(\hat{\theta}) - \sum_{\gamma\in\Gamma} N_\gamma(g)[f(\hat{\theta} - \gamma/T) - f(\hat{\theta}) + (\gamma/T)f'(\hat{\theta})], \end{aligned}$$

and the theorem follows.

5. Bias reducing properties of the J_∞-estimator. As was mentioned at the start, our intention is to produce the counterpart of the jackknife for certain types of stochastic processes. It is the purpose of this and the next section to show that the J_∞-estimator is precisely that. In what follows one will quickly see that $J_\infty[f(\hat{\theta})]$ enjoys essentially the same properties as the jackknife with the sample size n replaced by the record length T. The first few results are exact counterparts of similar results given by Adams, Gray and Watkins in [2].

THEOREM 5.1. *For $T > T_0$ let*

$$(5.1)\qquad E[f(\hat{\theta})] = f(\theta) + b(\theta, T),$$

and

$$(5.2)\qquad E[f(\hat{\theta}_n^{\,i})] = f(\theta) + b_i(\theta, T).$$

If

$$(5.3)\qquad E[f(\hat{\theta}_n^{\,i})] = E[f(\hat{\theta}_n^{\,j})] \qquad \textit{for all } i \textit{ and } j,$$

$$(5.4)\qquad \lim_{n\to\infty} E[J_n(f(\hat{\theta}))] = E[J_\infty(f(\hat{\theta}))],$$

and $(\partial/\partial T)b(\theta, T)$ *exists for* $T > T_0$, *then*

$$E[J_\infty(f(\hat{\theta}))] = f(\theta) + b(\theta, T) + T\frac{\partial}{\partial T} b(\theta, T)\,, \tag{5.5}$$

when $T > T_0$.

PROOF. Since

$$J_n[f(\hat{\theta})] = f(\hat{\theta}) + \frac{n-1}{n} \sum_{i=1}^n [f(\hat{\theta}) - f(\hat{\theta}_n{}^i)]\,, \tag{5.6}$$

we have

$$E[J_n(f(\hat{\theta}))] = f(\theta) + b(\theta, T) + \frac{n-1}{n} \sum_{i=1}^n [b(\theta, T) - b_i(\theta, T)]\,. \tag{5.7}$$

But

$$E[f(\hat{\theta}_n{}^n)] = f(\theta) + b\left(\theta, \frac{n-1}{n} T\right), \tag{5.8}$$

and hence, from (5.3)

$$E[J_n(f(\hat{\theta}))] = f(\theta) + b(\theta, T) + \frac{n-1}{n} \sum_{i=1}^n \left[b(\theta, T) - b\left(\theta, \frac{n-1}{n} T\right)\right\}. \tag{5.9}$$

Taking the limit of (5.9) as $n \to \infty$ and imposing (5.4) yields the result of (5.5).

It should be noted that condition (5.3) could be replaced by the condition that $\{I_G(t) \mid t \in [a, b]\}$ has stationary independent increments and the theorem would remain valid. An immediate consequence of Theorem 5.1 is the following corollary which we might have expected.

COROLLARY 5.1. *Under the conditions of Theorem* 5.1 *a necessary and sufficient condition that* $J_\infty[f(\hat{\theta})]$ *be unbiased is that* $b(\theta, T) = C(\theta)/T$, *where* $C(\theta)$ *is an arbitrary function of* θ.

PROOF. The result follows at once by setting

$$b(\theta, T) + T\frac{\partial}{\partial T} b(\theta, T) = 0\,, \qquad T > T_0\,,$$

and solving for $b(\theta, T)$.

Another result which is of interest and is obvious from Theorem 5.1 is the following.

COROLLARY 5.2. *Under the hypothesis of Theorem* 5.1, *if* $f(\hat{\theta})$ *is unbiased for* $f(\theta)$, *then* $J_\infty[f(\hat{\theta})]$ *is unbiased for* $f(\theta)$.

It is possible under the conditions of Theorem 5.1 to give a complete characterization of the bias reduction properties of the J_∞-estimator in an asymptotic sense. This is the purpose of the next sequence of definitions and theorems.

DEFINITION 5.1. Let $f_1(\hat{\theta})$ and $f_2(\hat{\theta})$ be biased estimators of $f(\theta)$ such that

$$\left|\lim_{T\to\infty} \frac{E[f_2(\hat{\theta})] - f(\theta)]}{E[f_1(\hat{\theta})] - f(\theta)]}\right| = L \leq 1 . \tag{5.10}$$

If $L = 1$ we shall say that $f_2(\hat{\theta})$ and $f_1(\hat{\theta})$ are "same order bias estimators" of $f(\theta)$, denoted by $f_2(\hat{\theta})$ S.O.B.E. $f_1(\hat{\theta})$. If $0 < L < 1$ we shall say that $f_2(\hat{\theta})$ is a "better same order bias estimator" than $f_1(\hat{\theta})$, denoted $f_2(\hat{\theta})$ B.S.O.B.E. $f_1(\hat{\theta})$. If $L = 0$, we shall say that $f_2(\hat{\theta})$ is a "lower order bias estimator than $f_1(\hat{\theta})$", denoted $f_2(\hat{\theta})$ L.O.B.E. $f_1(\hat{\theta})$.

In case $f_2(\hat{\theta})$ is unbiased and $f_1(\hat{\theta})$ is biased we shall say $f_2(\hat{\theta})$ L.O.B.E. $f_1(\hat{\theta})$, and if $f_1(\hat{\theta})$ and $f_2(\hat{\theta})$ are both unbiased we shall say that $f_2(\hat{\theta})$ S.O.B.E. $f_1(\hat{\theta})$.

THEOREM 5.2. *Let the conditions of Theorem* 5.1 *be satisfied and suppose that there exists* $k > 0$ *such that*

$$\lim_{T\to\infty} T^k b(\theta, T) = C(\theta) \neq 0 \quad \text{or} \quad \pm\infty , \tag{5.11}$$

and $\lim_{T\to\infty} T^{k+1}(\partial b(T, \theta)/\partial T)$ *exists. Then*

(i) *if* $k = 1$, $J_\infty(f(\hat{\theta}))$ L.O.B.E. $f(\hat{\theta})$,
(ii) *if* $k < 2$ *and* $k \neq 1$, $J_\infty(f(\hat{\theta}))$ B.S.O.B.E. $f(\hat{\theta})$,
(iii) *if* $k = 2$, $J_\infty(f(\hat{\theta}))$ S.O.B.E. $f(\hat{\theta})$,
(iv) *if* $k > 2$, $f(\hat{\theta})$ B.S.O.B.E. $J_\infty(f(\hat{\theta}))$.

PROOF. First we note that

$$C(\theta) = \lim_{T\to\infty} T^k b(\theta, T) = \lim_{T\to\infty} \left(-\frac{1}{k}\, T^{k+1} \frac{\partial}{\partial T}\, b(\theta, T)\right) , \tag{5.12}$$

and by applying (5.12) it is easily shown, using (5.5), that

$$\left|\lim_{T\to\infty} \frac{E[J_\infty(f(\hat{\theta}))] - f(\theta)}{E[f(\hat{\theta})] - f(\theta)}\right| = |1 - k| . \tag{5.13}$$

From (5.13) it follows that (i), (ii), (iii), and (iv) hold.

Some trivial examples are now included for clarity.

EXAMPLE 1. Let us again consider the process $\{N(t) \mid t \in [0, \infty)\}$, where $N(t)$ has the Poisson distribution with parameter λt. Then $\hat{\lambda}(0, T) = N(T)/T$ is an unbiased estimator of λ, and a simple-minded estimator of λ^2 is $f(\hat{\lambda}) = (\hat{\lambda})^2$. However,

$$J_\infty(f(\hat{\lambda})) = \left(\frac{N(T)}{T}\right)^2 - \frac{N(T)}{T^2} , \tag{5.14}$$

and $E[f(\hat{\lambda})] = \lambda^2 + \lambda/T$. Thus, by Corollary 5.1, $J_\infty(f(\hat{\lambda}))$ is an unbiased estimator of λ^2.

EXAMPLE 2. Consider Example 1 with k an integer greater than 2. As an estimator for λ^k, let $f(\hat{\lambda}) = (N(T)/T)^k$. Then

$$E[f(\hat{\lambda})] = \lambda^k + \sum_{i=1}^{k-1} \frac{a_i(k)}{T^i}, \tag{5.15}$$

where the $a_i(k)$ are functions of λ but not of T. Thus, by Theorem 5,2, $J_\infty(f(\hat{\lambda}))$ L.O.B.E. $f(\hat{\lambda})$, and, by Theorem 5.1.

$$\begin{aligned} E[J_\infty(f(\hat{\lambda}))] &= \lambda^k + \sum_{i=1}^{k=1} \frac{a_i}{T^i} - \sum_{i=1}^{k-1} \frac{ia_i}{T^i} \\ &= \lambda^k + \sum_{i=2}^{k-1} \frac{(1-i)a_i}{T^i}. \end{aligned} \tag{5.16}$$

EXAMPLE 3. We now return to the estimator $f(\hat{\lambda}) = e^{-(N(T)/T)x}$ defined in Example 1 of Section 4. We have already pointed out that in this case

$$J_\infty(f(\hat{\lambda})) = e^{-(N(T)/T)x}\{1 - N(T)[e^{x/T} - 1 - x/T]\}. \tag{5.17}$$

It is easy to show that $E[f(\hat{\lambda})] = e^{-\lambda T(1-e^{-x/T})}$, so that by Theorem 5.1

$$E[J_\infty(f(\hat{\lambda}))] = e^{-\lambda T(1-e^{-x/T})}[1 - \lambda T + \lambda T e^{-x/T}(1 + x/T)]. \tag{5.18}$$

But then

$$\begin{aligned} \lim_{T\to\infty} Tb(\theta, T) &= \lim_{T\to\infty} T(e^{-\lambda T(1-e^{-x/T})} - e^{-\lambda x}) \\ &= \tfrac{1}{2}\lambda x^2 e^{-\lambda x}. \end{aligned} \tag{5.19}$$

Thus, by Theorem 5.2, $J_\infty[f(\hat{\lambda})]$ L.O.B.E. $f(\hat{\lambda})$.

Before leaving our discussion of the bias reduction properties of $J_\infty[f(\hat{\theta})]$ we include a final result which shows under reasonable assumptions J_∞ is indeed more effective as a bias reduction tool than J_n for any finite n.

THEOREM 5.3. *If the conditions of Theorem* 5.1 *are satisfied and* $J_\infty(f(\hat{\theta}))$ *is an asymptotically unbiased estimator of* $f(\theta)$ *such that* $b(\theta, T) + T\partial b(\theta, T))\partial T$ *is a monotone function of* T, *then*

$$|E[J_\infty(f(\hat{\theta})) - f(\theta)]| \leq |E[J_n(f(\hat{\theta})) - f(\theta)]| \tag{5.20}$$

for all n.

PROOF. First note that $b(\theta, T) + T\partial b(\theta, T)/\partial T = \partial(Tb(\theta, T))/\partial T$. Since $\partial(Tb(\theta, T))/\partial T$ is monotone and $\lim_{T\to\infty} \partial(Tb(\theta, T))/\partial T = 0$, it is easily seen that either

(i) $\partial(Tb(\theta, T))/\partial T = 0$, or
(ii) $\partial(Tb(\theta, T))/\partial T$ is negative and increasing, or
(iii) $\partial(Tb(\theta, T))/\partial T$ is positive and decreasing.

Now observe that

$$(5.21)\qquad E[J_n(f(\theta)) - f(\hat{\theta})] = nb(\theta, T) - (n-1)b\left(\theta, \frac{n-1}{n}T\right)$$

$$= \frac{Tb(\theta, T) - \frac{n-1}{n}Tb\left(\theta, \frac{n-1}{n}T\right)}{\frac{1}{n}T}.$$

Thus, by the mean value theorem for derivatives, there exists x_n, with $n^{-1}(n-1)T < x_n < T$, such that

$$(5.22)\qquad E[J_n(f(\hat{\theta})) - f(\theta)] = \frac{\partial}{\partial t}(tb(\theta, t))\Big|_{t=x_n}.$$

Therefore, in view of (i), (ii), and (iii), we obtain

$$(5.23)\qquad |E[J_\infty(f(\hat{\theta})) - f(\theta)]| \le |E[J_n(f(\hat{\theta})) - f(\theta)]|\,.$$

6. Asymptotic properties of the J_∞-estimator. This section will be devoted to a study of the limiting distribution of $J_\infty(f(\hat{\theta}))$ as the length, T, of the interval $[a, b]$, becomes large. In view of this fact, it is possibly necessary to strengthen our notation for the J_∞-estimator to indicate the dependency of N and N_γ, of Definition 4.1, on the interval $[a, b]$. With this in mind, let

$$(6.1)\qquad N = N(a, b) \quad\text{and}\quad N_\gamma = N_\gamma(a, b)$$

for each $\gamma \in \Gamma$. Then a more appropriate notation for the J_∞-estimator becomes

$$(6.2)\qquad J_\infty(f(\hat{\theta}(a, b))) = f(\hat{\theta}(a, b)) - \sum_{\gamma\in\Gamma} N_\gamma(a, b)\left[f\left(\hat{\theta}(a, b) - \frac{\gamma}{T}\right) - f(\hat{\theta}(a, b)) + \frac{\gamma}{T}f'(\hat{\theta}(a, b))\right].$$

The latter is, however, quite awkward, and hence we will continue to use the more convenient notation of (4.10), and assume this short discussion will eliminate any confusion.

In order to establish the desired asymptotic theory for $J_\infty[f(\hat{\theta})]$ we will need a suitable estimator for its variance. This can be accomplished in a natural way. That is, return to the interval $[a, b]$, subdivide it by letting $t_i = a + (i-1)/n$, $i = 1, \cdots, [nT]$, and $t_n = b$, and assume $\{I_G(t) \mid t \in [a, \infty)\}$ has stationary independent increments. Then, since $\operatorname{Var}[\hat{\theta}(a + (i-1)/n, a + i/n)] = n\operatorname{Var}[\hat{\theta}(a, a+1)]$, for fixed n, a consistent estimator as $T \to \infty$ for $\operatorname{Var}[\hat{\theta}(a, a+1)]$ is

$$(6.3)\qquad \hat{\sigma}^2_{n,T} = \frac{1}{n([nT]-1)}\sum_{i=1}^{[nT]}\left[\hat{\theta}\left(a + \frac{i-1}{n}, a + \frac{i}{n}\right) - \hat{\theta}\left(a, a + \frac{[nT]}{n}\right)\right]^2,$$

where $[nT]$ is the greatest integer not exceeding nT. Our comments leading to (6.3) therefore suggest that a portion of the quantity we seek may be obtained by taking the limit of (6.3) as $n \to \infty$. When this is done one obtains

$$\lim_{n\to\infty} \hat{\sigma}^2_{n,T} = \frac{1}{T} \sum_{\gamma\in\Gamma} \gamma^2 N_\gamma, \text{ a.e. .} \tag{6.4}$$

We will denote the right side of (6.4) by $\hat{\sigma}_T{}^2$ and will not prove the result. The proof is not difficult but serves no purpose here since $\hat{\sigma}_T{}^2$ is the quantity we shall need. That is, we have introduced $\hat{\sigma}_T{}^2$ through (6.4) simply to indicate how it arises. We should note before leaving this discussion that if $\sigma^2 = \operatorname{Var}[\hat{\theta}(a, a+1)]$, then for any $\varepsilon > 0$

$$\lim_{n\to\infty} \lim_{T\to\infty} P[|\hat{\sigma}^2_{n,T} - \sigma^2| < \varepsilon] = 1 , \tag{6.5}$$

and hence when these limits can be reversed, $\hat{\sigma}_T{}^2$ converges to $\operatorname{Var}[\hat{\theta}(a, a+1)]$ in probability. A number of conditions could therefore be given which would guarantee this convergence. We will not do this however, but simply assume $\hat{\sigma}_T{}^2 \to_P \sigma^2$.

Although $\hat{\sigma}_T{}^2$ is not the desired estimator for the variance of $J_\infty[f(\hat{\theta})]$, we will be able to make use of it shortly to obtain our goal. First, however, let us give the following two theorems. We state the first without proof, since it is a trivial extension of a known result.

THEOREM 6.1. *Let $\{I_G(t) \mid t \in [a, \infty)\}$ be a stochastic process with stationary independent increments such that $E[\hat{\theta}(a, t)] = \theta$ for each $t \in (a, \infty)$. Also let* $\operatorname{Var}[\hat{\theta}(a, a+1)] = \sigma^2 < \infty$ *and assume f is differentiable in a neighborhood of θ. Then*

$$T^{\frac{1}{2}}[f(\hat{\theta}(a, a+T)) - f(\theta)] \to_L \mathcal{N}(0, \sigma^2(f'(\theta))^2) \tag{6.6}$$

as $T \to \infty$, where L denotes convergence in law.

THEOREM 6.2. *Suppose the conditions of Theorem 6.1 are satisfied and that f has a bounded second derivative in a neighborhood of θ. Moreover suppose that $\hat{\sigma}_T{}^2 \to_P \sigma^2$ and that Γ is a bounded set. Then*

$$T^{\frac{1}{2}}[J_\infty(f(\hat{\theta}(a, a+T)) - f(\theta)] \to_L \mathcal{N}(0, \sigma^2(f'(\theta))^2) , \tag{6.7}$$

as $T \to \infty$.

PROOF. Recall that

$$J_\infty(f(\hat{\theta})) = f(\hat{\theta}) - \sum_{\gamma\in\Gamma} N_\gamma \left[f\left(\hat{\theta} - \frac{\gamma}{T}\right) - f(\hat{\theta}) + \frac{\gamma}{T} f'(\hat{\theta}) \right]. \tag{6.8}$$

In view of Theorem 6.1, it is sufficient to show that

$$T^{\frac{1}{2}} \sum_{\gamma\in\Gamma} N_\gamma \left[f\left(\hat{\theta} - \frac{\gamma}{T}\right) - f(\hat{\theta}) + \frac{\gamma}{T} f'(\hat{\theta}) \right] \to_P 0 \qquad \text{as} \quad T \to \infty . \tag{6.9}$$

Moreover, it is easily seen that $\hat{\theta}(a, a + T) \to_P \theta$ as $T \to \infty$.

Now suppose that $|f''(t)| < M$ for all $t \in (\theta - \delta, \theta + \delta)$, where $\delta > 0$. Then, given $\varepsilon > 0$, there exists $T_\varepsilon > 0$ such that $P[\hat{\theta}(a, a + T) \in (\theta - \delta, \theta + \delta)] > 1 - \varepsilon$ when $T > T_\varepsilon$. Thus, for sufficiently large T,

$$f\left(\hat{\theta} - \frac{\gamma}{T}\right) = f(\hat{\theta}) - \frac{\gamma}{T} f'(\hat{\theta}) + \frac{\gamma^2}{2T^2} f''\left(\hat{\theta} - \alpha_\gamma \frac{\gamma}{T}\right) \tag{6.10}$$

and $|f''(\hat{\theta} - \alpha_\gamma \gamma/T)| < M$ with probability greater than $1 - \varepsilon$, where α_γ is a random variable such that the range of α_γ is a subset of $[-1, 1]$ with probability one. Thus, when T is sufficiently large,

$$\begin{aligned} &\left| T^{\frac{1}{2}} \sum_{\gamma \in \Gamma} N_\gamma \left[f\left(\hat{\theta} - \frac{\gamma}{T}\right) - f(\hat{\theta}) + \frac{\gamma}{T} f'(\hat{\theta}) \right] \right| \\ &\qquad \leqq \left| T^{-\frac{3}{2}} \sum_{\gamma \in \Gamma} \gamma^2 N_\gamma f''\left(\hat{\theta} - \alpha_\gamma \frac{\gamma}{T}\right) \right| \\ &\qquad \leqq \frac{M}{T^{\frac{1}{2}}} \hat{\sigma}_T^{\,2} \end{aligned} \tag{6.11}$$

with probability greater than $1 - \varepsilon$, and since $T^{-\frac{1}{2}} \hat{\sigma}_T^{\,2} \to_P 0$ as $T \to \infty$, we obtain

$$T^{\frac{1}{2}} \sum_{\gamma \in \Gamma} N_\gamma \left[f\left(\hat{\theta} - \frac{\gamma}{T}\right) - f(\hat{\theta}) + \frac{\gamma}{T} f'(\hat{\theta}) \right] \to_P 0 \tag{6.12}$$

as $T \to \infty$.

In view of Theorem 6.2, an appropriate estimator for the variance of $J_\infty[f(\hat{\theta})]$ will complete the asymptotic results we have been working toward. From our previous comments, one possibility would be to use $\hat{\sigma}_T^{\,2}[f'(\hat{\theta})]^2$. We will not, however, use this but take the following approach which is closer to the procedure used in jackknifing random samples.

Recall that

$$J_n(f(\hat{\theta})) = nf(\hat{\theta}) - \frac{n-1}{n} \sum_{i=1}^n f(\hat{\theta}_n^{\,i}), \tag{6.13}$$

and define the estimator $J_n^{\,i}(f(\hat{\theta}))$ by

$$J_n^{\,i}(f(\hat{\theta})) = nf(\hat{\theta}) - (n-1) f(\hat{\theta}_n^{\,i}). \tag{6.14}$$

Then, following the notions of jackknifing on a random sample, let

$$S_n^{\,2}(\hat{\theta}) = \frac{1}{n(n-1)} \sum_{i=1}^n [J_n^{\,i}(f(\hat{\theta})) - J_n(f(\hat{\theta}))]^2 \tag{6.15}$$

be an estimator for the variance of $J_n(f(\hat{\theta}))$. Under reasonable conditions it would appear that, $\lim_{n\to\infty} S_n^{\,2}(\hat{\theta})$ (when the limit is taken in some appropriate sense) should be a reasonable estimator for the variance of $J_\infty(f(\hat\theta))$. This turns out to be the case, and under the appropriate conditions one can show that

$$\text{(6.16)} \qquad \lim_{n\to\infty} S_n^2(\hat\theta) = \sum_{\gamma\in\Gamma} N_\gamma\Big[f\Big(\hat\theta - \frac{\gamma}{T}\Big) - f(\hat\theta)\Big]^2$$

with probability one. Again we will not prove a result such as (6.16) since it is only the right side of the equation that we will need. However, the discussion leading to (6.16) is germane for the reader to understand how it arises and how it should be interpreted. A proof of the result can be found in [39]. The next theorem shows the value of (6.16).

THEOREM 6.3. *Let* $\{I_G(t)\,|\,t\in[a,\infty)\}$ *be a stochastic process with stationary independent increments such that* $E[\hat\theta(a,t)] = \theta$ *for each* $t\in(a,\infty)$ *and* $\operatorname{Var}[\hat\theta(a,a+1)] = \sigma^2 < \infty$. *If* $\hat\sigma_T^2 \to_P \sigma^2$ *as* $T\to\infty$, f *has a continuous first derivative in a neighborhood of* θ, *and if* Γ *is a bounded set, then*

$$\text{(6.17)} \qquad T\sum_{\gamma\in\Gamma} N_\gamma[f(\hat\theta - \gamma/T) - f(\hat\theta)]^2 \to_P \sigma^2[f'(\theta)]^2$$

as $T\to\infty$.

PROOF. Suppose that f' is continuous in the interval $(\theta-\delta, \theta+\delta)$, where $\delta>0$. Since $\hat\theta(a, a+T)\to_P\theta$ as $T\to\infty$, given $\varepsilon>0$, there exists $T_\varepsilon>0$ such that $P[\hat\theta(a,a+T)\in(\theta-\delta,\theta+\delta)] > 1-\varepsilon$ when $T>T_\varepsilon$. Thus, since Γ is bounded, for sufficiently large T,

$$\text{(6.18)} \qquad P\Big[f\Big(\hat\theta - \frac{\gamma}{T}\Big) = f(\hat\theta) - \frac{\gamma}{T}f'\Big(\hat\theta - \alpha_\gamma\frac{\gamma}{T}\Big)\Big] > 1-\varepsilon$$

for every $\gamma\in\Gamma$, where α_γ is a random variable such that the range of α_γ is a subset of $[-1, 1]$ with probability one. Therefore, for sufficiently large T,

$$\text{(6.19)} \qquad T\sum_{\gamma\in\Gamma} N_\gamma\Big[f\Big(\hat\theta - \frac{\gamma}{T}\Big) - f(\hat\theta)\Big]^2 = \frac{1}{T}\sum_{\gamma\in\Gamma}\gamma^2 N_\gamma\Big[f'\Big(\hat\theta - \alpha_\gamma\frac{\gamma}{T}\Big)\Big]^2$$

with probability greater than $1-\varepsilon$. Now, since Γ is bounded,

$$\text{(6.20)} \qquad \hat\theta(a, a+T) - \alpha_\gamma\frac{\gamma}{T} \to_P \theta\,, \qquad \frac{1}{T}\sum_{\gamma\in\Gamma}\gamma^2 N_\gamma(a, a+T) \to_P \sigma^2$$

as $T\to\infty$, and f' is continuous in the interval $(\theta-\delta,\theta+\delta)$, it follows that

$$\text{(6.21)} \qquad T\sum_{\gamma\in\Gamma} N_\gamma[f(\hat\theta - \gamma/T) - f(\hat\theta)]^2 \to_P \sigma^2[f'(\theta)]^2$$

as $T\to\infty$, as was to be shown.

Combining Theorems 6.2 and 6.3 we obtain the following theorem.

THEOREM 6.4. *Under the conditions of Theorem* 6.2,

$$\frac{J_\infty(f(\hat\theta)) - f(\theta)}{[\sum_{\gamma\in\Gamma} N_\gamma[f(\hat\theta - \gamma/T) - f(\hat\theta)]^2]^{\frac12}} \to_L \mathcal{N}(0, 1) \tag{6.22}$$

as $T \to \infty$.

Before closing this section we should make a final comment with regard to our extension of the jackknife to stochastic processes. The question which naturally arises is in regard to extending $G(\hat\theta)$. In connection with this, note that the definition of say $G_n[f(\hat\theta)]$ corresponding to $J_n[f(\hat\theta)]$ is

$$\begin{aligned} G_n[f(\hat\theta)] &= \frac{f(\hat\theta) - R(n)\overline{f(\hat\theta_n{}^i)}}{1 - R(n)} \\ &= f(\hat\theta) + \frac{R(n)}{[1 - R(n)](n-1)}\{(n-1)[f(\hat\theta) - \overline{f(\hat\theta_n{}^i)}]\}. \end{aligned} \tag{6.23}$$

Thus, assuming the indicated limits exist,

$$\lim_{n\to\infty} G_n[f(\hat\theta)] = f(\hat\theta) + \alpha \lim_{n\to\infty}\{(n-1)[f(\hat\theta) - \overline{f(\hat\theta_n{}^i)}]\}, \tag{6.24}$$

where

$$\alpha = \lim_{n\to\infty} R(n)/[1 - R(n)](n-1).$$

But

$$\lim_{n\to\infty}\{(n-1)[f(\hat\theta) - \overline{f(\hat\theta_n{}^i)}]\} = J_\infty[f(\hat\theta)] - f(\hat\theta), \tag{6.25}$$

and hence

$$\lim_{n\to\infty} G_n[f(\hat\theta)] = (1-\alpha)f(\hat\theta) + \alpha J_\infty[f(\hat\theta)]. \tag{6.26}$$

But, except for the fact that we do not have a random sample, the right side of (6.26) is just a special case of (2.1) with $R = -\alpha/(1-\alpha)^{-1}$. Thus rather than going through similar extensions for $G(\hat\theta)$ one simply takes $\hat\theta_1 = f(\hat\theta)$, $\hat\theta_2 = J_\infty[f(\hat\theta)]$ and selects R, as best possible, to approximate the ratio of the biases in these two estimators.

7. The $J_\infty{}^{(2)}$-estimator. Although $J(\hat\theta)$ is by far the most extensively studied of the estimators we have considered, it is not necessarily the best, even from a bias reduction viewpoint. In fact, it was pointed out in [2] that from a reduction of bias point of view, $J^{(2)}(\hat\theta)$ is more robust to departures from the assumed form of the bias than $J(\hat\theta)$. Moreover, one can give examples where the mean square error in $J^{(2)}(\hat\theta)$ is smaller than the mean square error in $J(\hat\theta)$. For these reasons it seems logical to attempt to extend the concept of $J^{(2)}(\hat\theta)$ to stochastic processes. This is the purpose of this section. To accomplish our goal we take the obvious approach.

DEFINITION 7.1. Let $\hat{\theta}$ be defined as before. Then we define $J_n^{(2)}[f(\hat{\theta})]$ over the interval $[a, b]$ by

$$J_n^{(2)}[f(\hat{\theta})] = \tfrac{1}{2}\Big[n^2 f(\hat{\theta}) - 2(n-1)^2 \frac{1}{n} \textstyle\sum_{i=1}^{n} f(\hat{\theta}_n^{\,i}) + (n-2)^2 \frac{1}{n(n-1)} \sum_{i\neq j} f(\hat{\theta}_n^{\,ij})\Big], \tag{7.1}$$

where

$$\hat{\theta}_n^{\,i} = \frac{n\hat{\theta}}{n-1} - \frac{\hat{\theta}_i}{n-1}; \qquad \hat{\theta}_n^{\,ij} = \frac{n\hat{\theta}}{n-2} - \frac{\hat{\theta}_i + \hat{\theta}_j}{n-2}, \qquad i \neq j. \tag{7.2}$$

It should be noted that $J_n^{(2)}[f(\hat{\theta})]$ is simply $J^{(2)}[f(\hat{\theta})]$ when one considers, as before, the $\hat{\theta}^i$ as independent identically distributed random variables. That is, when one makes this assumption, (7.1) is just the estimator defined in (2.12) with $k = 2$, $a_{ij} = (n - j + 1)^{-1}$ and $\hat{\theta}_2$ and $\hat{\theta}_3$ determined by the drop out rule previously described.

Due to our previous development, our interest centers on establishing the $\lim_{n\to\infty} J_n^{(2)}[f(\hat{\theta})]$. This leads to the following definition.

DEFINITION 7.2. Let the conditions in the definition of $J_\infty[f(\hat{\theta})]$ be satisfied and further assume f is defined and twice differentiable on the range of $\hat{\theta}$. Then the second order J_∞-estimator is defined as

$$\begin{aligned} J_\infty^{(2)}[f(\hat{\theta})] &= f(\hat{\theta}) + \tfrac{1}{2} f''(\hat{\theta})\Big[\textstyle\sum_{\alpha=\Gamma} \frac{\alpha N_\alpha}{T}\Big]^2 \\ &\quad - \textstyle\sum_{\alpha\in\Gamma} N_\alpha\Big[\Big(\frac{\alpha}{T}\Big)(N+1) f'(\hat{\theta}) - 2f(\hat{\theta}) + 2f\Big(\hat{\theta} - \frac{\alpha}{T}\Big)\Big] \\ &\quad + \tfrac{1}{2}\textstyle\sum_{\alpha\in\Gamma} N_\alpha(N_\alpha - 1)\Big[f(\hat{\theta}) - 2f\Big(\hat{\theta} - \frac{\alpha}{T}\Big) \\ &\quad + f\Big(\hat{\theta} - \frac{2\alpha}{T}\Big) + \Big(\frac{2\alpha}{T}\Big) f'\Big(\hat{\theta} - \frac{\alpha}{T}\Big)\Big] \\ &\quad + \tfrac{1}{2}\textstyle\sum_{\alpha\neq\beta} N_\alpha N_\beta\Big[f(\hat{\theta}) - 2f\Big(\hat{\theta} - \frac{\alpha}{T}\Big) \\ &\quad + f\Big(\hat{\theta} - \frac{\alpha+\beta}{T}\Big) + \Big(\frac{2\alpha}{T}\Big) f'\Big(\hat{\theta} - \frac{\beta}{T}\Big)\Big], \end{aligned} \tag{7.3}$$

where $f''(\hat{\theta}) = d^2f(\theta)/d\theta^2|_{\theta=\hat{\theta}}$ and $T = b - a$.

Under the conditions of Theorem 4.1 and the assumption that f has a continuous second derivative over the range of $\hat{\theta}$ one can show that

$$\lim_{n\to\infty} J_n^{(2)}[f(\hat{\theta})] = J_\infty^{(2)}[f(\hat{\theta})] \tag{7.4}$$

with probability one. We shall not prove the result here since it is along exactly the same line as Theorem 4.1 and is excruciatingly long. For a detailed proof of the result see [1].

Although $J^{(2)}[f(\hat{\theta})]$ appears quite unmanageable this is not necessarily the case as we shall demonstrate shortly. First, however, let us list some additional theorems concerning the properties of $J_\infty^{(2)}[f(\hat{\theta})]$ which justify it as the appropriate extension of $J_\infty[f(\hat{\theta})]$. Again we will not prove the results since they are so similar to previous results of this paper.

THEOREM 7.1. *If the following conditions are satisfied*

(i) $\{I_G(t) \mid t \in [a, b]\}$ *has stationary independent increments*
(ii) $\lim_{n\to\infty} E[J_n^{(2)}[f(\hat{\theta})]] = E[J_\infty^{(2)}[f(\hat{\theta})]]$
(iii) $E[f(\hat{\theta})] = f(\theta) + b(\theta, T)$

and

(iv) $\partial b(\theta, T)/\partial T$ *and* $\partial^2 b(\theta, T)/\partial T^2$ *exist,*

then

$$E[J_\infty^{(2)}[f(\hat{\theta})]] = f(\theta) + b(\theta, T) + 2T\frac{\partial b(\theta, T)}{\partial T} + \frac{T^2}{2}\frac{\partial^2 b(\theta, T)}{\partial T^2}. \tag{7.5}$$

COROLLARY 7.1. *If the theorem above is satisfied and*

$$b(\theta, T) = \frac{C_1(\theta)}{T} + \frac{C_2(\theta)}{T^2} + \frac{C_3(\theta)}{T^3} + \cdots, \tag{7.6}$$

then

$$E[J_\infty^{(2)}[f(\hat{\theta})]] = f(\theta) + \frac{D_1(\theta)}{T^3} + \cdots,$$

where $D_1(\theta) = D_2(\theta) = \cdots = 0$ *if and only if* $C_3(\theta) = C_4(\theta) = \cdots = 0$.

To exemplify the above theory let us continue to call on the Poisson process of several of our previous examples by considering $f(\hat{\lambda}) = (N(T)/T)^3$ as an estimator for λ^3. In this event one obtains

$$J_\infty[f(\hat{\lambda})] = \left(\frac{N(T)}{T}\right)^3 - \frac{3}{T}\left(\frac{N(T)}{T}\right)^2 + \frac{1}{T^2}\left(\frac{N(T)}{T}\right), \tag{7.7}$$

and

$$J_\infty^{(2)}[f(\hat{\lambda})] = \left(\frac{N(T)}{T}\right)^3 - \frac{3}{T}\left(\frac{N(T)}{T}\right)^2 + \frac{2}{T^2}\left(\frac{N(T)}{T}\right). \tag{7.8}$$

Equation (7.8) demonstrates quite vividly our earlier comment regarding the manageability of $J_\infty^{(2)}[f(\hat{\theta})]$, and thus it is clear that (7.3) may be somewhat misleading in this respect. Simple calculation shows

$$E[f(\hat{\lambda})] = \lambda^3 + \frac{3\lambda^2}{T} + \frac{\lambda}{T^2}, \tag{7.9}$$

$$E[J_\infty(f(\hat{\lambda}))] = \lambda^3 - \frac{\gamma}{T^2} \tag{7.10}$$

and

(7.11) $$E[J_\infty^{(2)}(f(\hat{\lambda}))] = \lambda^3 .$$

Moreover

$$E[\{J_\infty[f(\hat{\lambda})] - \lambda^3\}^2] - E[\{J_\infty^{(2)}[f(\hat{\lambda})] - \lambda^3\}^2] = \frac{26\lambda^4}{T^2} - \frac{6\lambda^3}{T^3} + \frac{\lambda^2}{T^4} + \frac{\lambda}{T^5},$$

which is positive for λ, $T \geqq 1$ and hence the mean square error in $J_\infty^{(2)}[f(\hat{\lambda})]$ is less in this case than the mean square error in $J_\infty[f(\lambda)]$ which can be shown to be smaller than the mean square error in $f(\hat{\lambda})$. The example is of course rather artificial but it does dispel the misconception, which one has a natural tendency toward, that the $J_\infty^{(2)}$-estimator will always induce too much variability.

In order to complete our development of $J_\infty^{(2)}[f(\hat{\theta})]$ to the same level as $J_\infty[f(\hat{\theta})]$ we list two final results. The first establishes (as does Theorem 5.2 for $J_\infty[f(\hat{\theta})]$) the range of the order of the bias for which one might expect $J_\infty^{(2)}[f(\hat{\theta})]$ to reduce bias and the second establishes the asymptotic normality theory for $J_\infty^{(2)}[f(\hat{\theta})]$.

THEOREM 7.2. *If the conditions of Theorem* 7.1 *are satisfied and there exists a* $k > 0$ *such that*

$$\lim_{T\to\infty} T^k b(\theta, T) = C(\theta) \neq 0, \pm\infty ,$$

and

$$\lim_{T\to\infty}\left[T^{k+2}\,\frac{\partial^2 B(T, \theta)}{\partial T^2}\right]$$

exists, then

(i) *if* $k = 2$ *or* $k = 1$, *then* $J_\infty^{(2)}[f(\hat{\theta})]$ L.O.B.E. $f(\hat{\theta})$

(ii) *if* $k < 3$, $k \neq 1, 2$, *then* $J_\infty^{(2)}[f(\hat{\theta})]$ B.S.O.B.E. $f(\hat{\theta})$

(iii) *if* $k = 3$, *then* $J_\infty^{(2)}[f(\hat{\theta})]$ S.O.B.E. $f(\hat{\theta})$

and

(iv) *if* $k > 3$, *then* $f(\hat{\theta})$ B.S.O.B.E. $J_\infty^{(2)}[f(\hat{\theta})]$.

THEOREM 7.3. *Let f have a bounded third derivative in a neighborhood of* θ *and suppose the conditions of Theorem* 6.4 *are satisfied. Then*

(7.12) $$\frac{J_\infty^{(2)}[f(\hat{\theta})] - f(\theta)}{\{\sum_{\alpha\in\Gamma} N_\alpha[f(\hat{\theta} - \alpha/T) - f(\hat{\theta})]^2\}^{\frac{1}{2}}} \to_L \mathcal{N}(0, 1),$$

as $T \to \infty$.

8. The e_n-transformation. In the previous sections we have considered the jackknife from its original conception by Quenouille through several modifications until finally establishing the J_∞-estimator as an extension of this concept to stochastic processes. In this section we will introduce a transformation which has been studied rather extensively in numerical analysis. We will refer to it as the e_n-transformation and point out the manner in which our development of the jackknife is related to this transformation and its extensions.

DEFINITION 8.1. Let S_n be a sequence of real numbers defined by

$$S_n = \sum_{k=1}^{n} a_k . \tag{8.1}$$

Then we define the e_1-transformation of S_n by

$$e_1(S_n) = \frac{S_n - \rho(n)S_{n-1}}{1 - \rho(n)} , \tag{8.2}$$

where

$$\rho(n) = \frac{a_n}{a_{n-1}} \neq 1 . \tag{8.3}$$

The e_1-transformation was first introduced by Aitken [3] in 1926 under the name of the δ^2-process, although its roots go back to the middle 1800's with Kummer. Considerable interest in the subject was stimulated in the 1950's from papers by Shanks [36] and Lubkin [21] which represent the first extensive investigations of the transformation. The primary purpose of the e_1-transformation is to increase the rate of convergence of slowly convergent sequences. In fact, it was shown in [21] that if $\rho(n)$ is an analytic function of n^{-1} and $\lim_{n\to\infty} \rho(n) \neq 1$, then $e_1{}^m(S_n)$, where

$$e_1{}^m(S_n) = \underbrace{e_1 e_1 \cdots e_1(S_n)}_{m \text{ applications}} , \tag{8.4}$$

converges more rapidly than $e_1{}^{m-1}(S_n)$ to the same limit for every m for which the quantity is defined. We demonstrate this effect by the following example which was first given in [36].

EXAMPLE 1. Consider the expansion for π generated by the Taylor series expansion of arc tan θ, i.e.

$$\pi = 4 - \tfrac{4}{3} + \tfrac{4}{5} - \tfrac{4}{7} + \cdots . \tag{8.5}$$

Direct summation of (8.5) requires in excess of 40 million terms to produce the accuracy obtained below by repeated application of e_1 to the sequence of partial sums, S_n, generated from only the first 10 terms of (8.5).

n	S_n	$e_1(S_n)$	$e_1{}^2(S_n)$	$e_1{}^3(S_n)$	$e_1{}^4(S_n)$
0	4.0000000				
1	2.6666667	3.1666667			
2	3.4666667	3.1333333	3.1421053		
3	2.8952381	3.1452381	3.1414502	3.1415993	
4	3.3396825	3.1396825	3.1416433	3.1415909	3.1415928
5	2.9760462	3.1427129	3.1415713	3.1415933	3.1415927
6	3.2837385	3.1408814	3.1416029	3.1415925	
7	3.0170718	3.1420718	3.1415873		
8	3.2523659	3.1412548			
9	3.0418396				

As mentioned already, due to results such as Example 1, much interest has been stimulated in this area. Before we discuss the theory which resulted from this stimulation let us point out one other aspect of the transformation, namely, the fact that in many instances the original sequence need not converge for the transformation to be effective. Only a sparse amount of theory exists with regard to this aspect of the transformation but numerous examples can be found to demonstrate it. We shall give such an example here since we shall mention the statistical counterpart of this notion later.

EXAMPLE 2. Consider the series

$$\log(3) = 0 + 2 - (\tfrac{1}{2})2^2 + (\tfrac{1}{3})2^3 - (\tfrac{1}{4})2^4 + \cdots$$

which arises by erroneously setting $x = 2$ in the Taylor's expansion of $\log(1 + x)$. The series diverges rather rapidly but note the effect of applying e_1 repeatedly. This is given in the following table where the last figure is correct for log 3 to every digit shown.

n	S_n	$e_1(S_n)$	$e_1^2(S_n)$	$e_1^3(S_n)$	$e_1^4(S_n)$	$e_1^5(S_n)$
0	0.0000000					
1	2.0000000	1.0000000				
2	0.0000000	1.1428571	1.0931677			
3	2.6666667	1.0666667	1.1007092	1.0984266		
4	−1.3333333	1.1288421	1.0974359	1.0986841	1.0986080	
5	5.0666667	1.0666667	1.0994536	1.0985761	1.0986141	1.0986122
6	−5.6000000	1.1368421	1.0989008	1.0986346	1.0986114	1.0986123
7	12.6857143	1.0493507	1.0992921	1.0985862	1.0986128	
8	−19.314286	1.1657143	1.0978997	1.0986254		
9	37.574603	1.0031746	1.0994152			
10	−64.825397	1.2391193				
11	121.35642					

One of the earliest results shown concerning e_1 was the fact that if S_n is a geometric series then $e_1(S_n) \equiv S$ for every $n \geqq 3$. Thus e_1 was said to be exact on geometric series. Note that

$$S_n = S - \textstyle\sum_{k=n+1}^{\infty} a_k, \tag{8.6}$$

and at least for geometric series $a_n/a_{n-1} = \sum_{k=n+1}^{\infty} a_k / \sum_{k=n}^{\infty} a_k$. Thus we see, by considering $-\sum_{k=n+1}^{\infty} a_k$ as "bias" in the approximation S_n, that the jackknife is the precise counterpart of $e_1(S_n)$. This is clear when we recall that

$$J(\hat{\theta}) = \frac{\hat{\theta} - R\overline{\hat{\theta}^i}}{1 - R}, \tag{8.7}$$

where $R = (n - 1)/n$ is the ratio of the biases in $\hat{\theta}_n$ and $\hat{\theta}_{n-1}$ when the bias in $\hat{\theta}_n$ is $C(\theta)/n$ (the subscript n is included here for clarity).

Now the e_1-transform was generalized in [17] by Gray and Clark (1969) to the general form defined by

$$T[S(n); g(n)] = \frac{S(n) - \rho(n; g(n))S(g(n))}{1 - \rho(n; g(n))}, \tag{8.8}$$

where g is an integer valued function selected so that the transformation is exact on the type of series of interest. Thus the e_1-transformation is obtained by selecting g so that the transformation is exact on geometric series just as the jackknife is obtained from the generalized jackknife by selecting R to remove bias of the form $C(\theta)/n$. By noting that

$$e_1(S_n) = \frac{\begin{vmatrix} S_{n-1} & S_n \\ \Delta S_{n-1} & \Delta S_n \end{vmatrix}}{\begin{vmatrix} 1 & 1 \\ \Delta S_{n-1} & \Delta S_n \end{vmatrix}}, \tag{8.9}$$

$\Delta S_n = S_n - S_{n-1}$, a natural extension of e_1 is the e_k-transformation defined by

$$e_k(S_n) = \frac{\begin{vmatrix} S_{n-k} & S_{n-k+1} & \cdots & S_n \\ \Delta S_{n-k} & \Delta S_{n-k+1} & \cdots & \Delta S_n \\ \vdots & & & \\ \Delta S_{n-1} & \Delta S_n & \cdots & \Delta S_{n-1+k} \end{vmatrix}}{\begin{vmatrix} 1 & 1 & \cdots & 1 \\ \Delta S_{n-k} & \Delta S_{n-k+1} & \cdots & \Delta S_n \\ \vdots & & & \\ \Delta S_{n-1} & \Delta S_n & \cdots & \Delta S_{n-1+k} \end{vmatrix}}. \tag{8.10}$$

This transformation greatly extended the range of useful application of the basic notions set forth by the e_1-transformation. Indeed, it was shown in [36] that if S_n is the nth partial sum of a power series expansion of a rational function, then e_k is exact for some k. Further results in this area were then obtained by P. Wynn (1956), who established an efficient algorithm for computing $e_k(S_n)$. This algorithm is referred to as the epsilon algorithm and most current studies concerning e_k are related to it.

A reasonable application of the e_1-transformation would appear to be the problem of evaluating improper integrals. As was pointed out by Gray, Atchison, and McWilliams (1971) in [16] one approach to this is as follows.

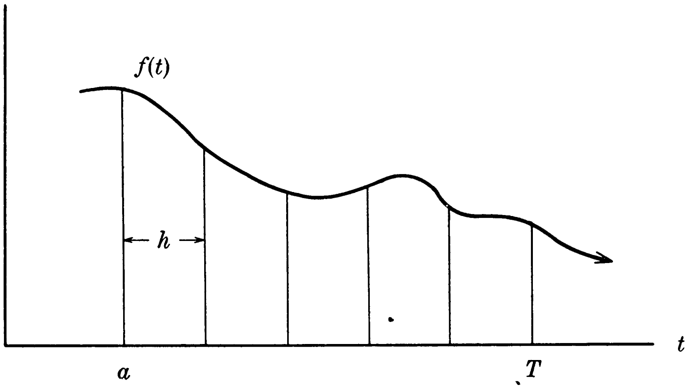

FIG. 1.

Let

$$F(T; a) = \int_a^T f(x)\,dx \tag{8.11}$$

and suppose $F(\infty; a)$ is finite. Now partition the interval $[a, T]$ as shown in figure 1. Then

$$F(T; a) = \sum_{i=1}^{n} a_i(h) = S_n(h)\,, \tag{8.12}$$

where

$$a_i(h) = F(a + ih; a) - F(a + (i-1)h; a)\,, \tag{8.13}$$

and

$$F(\infty; a) = \sum_{i=1}^{\infty} a_i(h)\,. \tag{8.14}$$

Thus one could apply e_1 to the sequence $S_n(h)$ in hopes of obtaining a better approximation to $F(\infty; a)$. With a slight bit of reflection one can see that forming $J_n[f(\hat{\theta})]$ in Section 3 was a direct analogy of applying e_1, in the fashion above, to continuous data.

In general, it was found that $e_1(S_n(h))$ was often an increasingly better approximation as $h \to 0$. This suggested considering the $\lim_{h\to 0} e_1[S_n(h)]$ which was done in [19] and extended to $e_k[S_n(h)]$ in [16]. These extensions resulted in a transformation defined as follows. Let T be fixed and $h = T/n$. Then define the B_k-transformation by

$$B_k[F(T; a)] = \lim_{h\to 0} e_k[S_n(h)] = \frac{\left|\begin{matrix} F & f & f^{(1)} & \cdots & f^{(k-1)} \\ f & f^{(1)} & f^{(2)} & \cdots & f^{(k)} \\ \vdots & & & & \\ f^{(k-1)} & f^{(k)} & f^{(k+1)} & \cdots & f^{(2k-1)} \end{matrix}\right|_T}{\left|\begin{matrix} f^{(1)} & f^{(2)} & \cdots & f^{(k)} \\ f^{(2)} & f^{(3)} & \cdots & f^{(k+1)} \\ \vdots & & & \\ f^{(k)} & f^{(k+1)} & \cdots & f^{(2k-1)} \end{matrix}\right|_T}\,, \tag{8.15}$$

where $f^{(k)}$ denotes the kth derivative of f.

The result (8.15) has been shown to be of value and has yielded a number of new approximation functions for tail probabilities. See [18], [19] for example.

The B_k-transformation does possess an exactness theorem just as the previous transformations we have mentioned. That is, it has been shown that if $f^{(n)}(t) + c_n f^{(n-1)}(t) + \cdots + c_0 f(t) = 0$, then $B_k[F(T; a)] = F(\infty)$ for all $T \geqq a$ and all $k \geqq n$. This and several other closely related results are established in [16] where the properties of $B_k[F(T; a)]$ are studied in some detail.

At this point, the correspondence between the J_∞-estimator and $B_1[F(T; a)]$ as well as the correspondence between the $J_\infty^{(2)}$-estimator and $B_2[F(t; a)]$ should be clear, and our comparison is complete. We summarize our comments in the following tabular form.

SEQUENCE TO SEQUENCE TRANSFORMATION	CORRESPONDING ESTIMATOR
1. The e_1-transformation.	1. The jackknife statistic, $J(\hat{\theta})$.
2. The generalized e_1-transformation.	2. The generalized jackknife statistic, $G(\hat{\theta})$.
3. The e_k-transformation.	3. The higher-order or iterated jackknife, $J^{(k)}(\hat{\theta})$.
4. $B_k[F(T; a)]$.	4. $J_\infty[f(\hat{\theta})]$ and $J_\infty^{(2)}[f(\hat{\theta})]$.

One should note that the statistical analogy for $B_k[F(t; a)]$ has only been established for $k = 1, 2$. Undoubtly this analogy could be obtained for all k but in view of the complexity of the form of $J_\infty^{(2)}[f(\hat{\theta})]$ this does not appear feasible.

9. Concluding remarks. In the previous section we have tried to point out the manner in which the development of the jackknife, generalized jackknife, etc. is directly analogous to the development surrounding the e_1-transformation. At times, our analogies may have seemed vague but this was necessary in order to eliminate an extremely lengthy discussion. In any event, the analogies are there and sufficient references have been included for those who feel some further insight is necessary.

One point which was mentioned in Section 8, but not pursued, is the corresponding notion in jackknifing to that of summing divergent series by means of the e_1-transformation. We shall make some comments concerning that subject now. First, it should be noted that no statements regarding the consistency of $G(\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_{k+1})$ have been made. In order to consider this property let us restrict our discussion to $G(\hat{\theta})$ and the reader can easily extrapolate our remarks to the general case.

It is quite clear if $\hat{\theta}$ is consistent for θ and $\lim_{n\to\infty} R \neq 1$ that $G(\hat{\theta})$ is also consistent for θ. However, as we have observed $e_1(S_n)$ may converge even though S_n diverges. Thus it seems reasonable to expect that $G(\hat{\theta})$ may be consistent for θ even though $\hat{\theta}$ is not consistent for θ. No theory has been established with regard to this conjecture and that is not our intent here. However, we include the following trivial example for the purpose of stimulating some thought in this area.

Consider $\hat{\theta} = \sum_{i=1}^{n} (x_i - \bar{x})^2$ as an estimator for σ^2 (admittedly a foolish thing to do). Obviously $\hat{\theta}$ is not consistent, but taking $R = (n - 2)/(n - 3)$ in $G(\hat{\theta})$ gives $G(\hat{\theta}) = \sum_{i=1}^{n} (x_i - \bar{x})^2/(n - 1)$. Hence by knowing the form of the bias in $\hat{\theta}$ we were able to generate the consistent-unbiased estimator $G(\hat{\theta})$. It would seem that the notion of generating consistent estimators from inconsistent estimators is the corresponding problem to summing divergent series, and that some study making use of $G(\hat{\theta})$ for that purpose would be interesting. A word of caution, however, should be injected here. That is, if $\lim_{n\to\infty} R = 1$ it is possible for $G(\hat{\theta})$ to be inconsistent even though $\hat{\theta}$ is consistent. This need not alarm us too much. For example $J(\hat{\theta})$ falls in this category (i.e. $\lim_{n\to\infty} (n - 1)/n = 1$) but in practice it has caused no real difficulty. Nevertheless it is a possibility, as the following example demonstrates. Suppose $\hat{\theta}_n = \bar{x} + (-1)^n/n$. Then clearly $\hat{\theta}_n$ is consistent for $E[x]$ and $b(n, \theta) = (-1)^n/n$. However $J(\hat{\theta}_n) = \bar{x} + 2(-1)^n$ and hence is not consistent. Note that taking $R = -(n - 1)/n$ in $G(\hat{\theta})$ gives $G(\hat{\theta}) = \bar{x}$ and hence in this case we see the source of the difficulty with $J(\hat{\theta})$ lies completely in the fact that $R = (n - 1)/n$ is a very unsatisfactory choice for R. For the e_1-transformation such behavior has essentially been characterized (see [17]) however no nontrivial results are known for $G(\hat{\theta})$ in the corresponding situation, i.e. $\lim_{n\to\infty} R = 1$.

As a final remark let us make some comments concerning the assumed form of $\hat{\theta}(a, b)$ in Sections 4 through 7. It would seem that this form may be unnecessarily restrictive, but there are certainly a large class of problems to which the resulting developments apply. Moreover, the results of those sections should serve as a guide for further extensions. One of these extensions, which seems most obvious at this stage, is to consider the $\hat{\theta}_i$'s as a random sample and assume that $\hat{\theta}(a, b)$ is a U-statistic defined on that sample. There obviously would be some difficulty in obtaining this extension, but a close examination of this paper suggests it is feasible. Moreover, such an extension would certainly greatly enlarge the class of problems to which the theory of jackknifing stochastic processes is applicable and thus be of some value.

REFERENCES

[1] ADAMS, J. E. (1971). Jackknifing stochastic processes using the second order J-estimator. Ph. D. dissertation, Texas Tech Univ.

[2] ADAMS, J. E., GRAY, H. L., and WATKINS, T. A. (1971). An asymptotic characterization of bias reduction by jackknifing. *Ann. Math. Statist.* **42** 1606–1612.

[3] AITKEN, A. C. (1926). On Bernoulli's numerical solution of algebraic equations. *Proc. Roy. Soc. Edinburgh Sect. A.* **46** 289-305.

[4] ARVESEN, J. N. (1969). Jackknifing U-statistics. *Ann. Math. Statist.* **40** 2076-2100.

[5] ARVESEN, J. and SCHMITZ, T. (1970). Robust procedures for variance component problems using the jackknife. *Biometrics* **26** 677-686.

[6] BRILLINGER, D. (1964). The asymptotic behavior of Tukey's general method of setting approximate confidence limits (the jackknife) when applied to maximum likelihood estimates. *Rev. Inst. Internat. Statist.* **32** 202-206.

[7] BRILLINGER, D. (1966). The application of the jackknife to the analysis of consumer surveys. *Commentary.* **8** 74-80.

[8] BURDICK, D. S. (1961). Stage by stage modification of polynomial estimators by the jackknife method. Ph. D. dissertation, Princeton Univ.

[9] CHAKRABARTY, R. P. and RAO, J. N. K. (1968). The bias and stability of the jackknife variance estimator in ratio estimation (Abstract). *J. Amer. Statist. Assoc.* **63** 748.

[10] CHATTERJEE, S. and BARCUN, S. (1970). A nonparametric approach to credit screening. *J. Amer. Statist. Assoc.* **65** 150-154.

[11] DEMING, W. E. (1963). On the correction of mathematical bias by use of replicated designs. *Metrika* **6** 37-42.

[12] DEMPSTER, A. (1966). Estimation in multivariate analysis, in *Multivariate Analysis* **1**. Academic Press, New, York.

[13] DURBIN, J. (1959). A note on the application of Quenouille's method of bias reduction to the estimation of ratios. *Biometrika* **46** 477-480.

[14] FRYER, J. (1970). Jackknifing maximum likelihood estimates in the multiparameter case (preliminary report) (Abstract). *Ann. Math. Statist.* **41** 1392.

[15] GAVER, D. P. and HOEL, D. G. (1970). Comparison of certain small-sample Poisson probability estimates. *Technometrics* **12** 835-850.

[16] GRAY, H. L., ATCHISON, T. A. and MCWILLIAMS, G. V. (1971). Higher Order G-Transformations, *SIAM J. Numer. Anal.* **8** 000-000.

[17] GRAY, H. L. and CLARK, W. D. (0000). On a Class of nonlinear Transformations and their applications to the evaluation of infinite series. *J. Res. Nat. Bur. Standards Sect. B.* **73** 251-274.

[18] GRAY, H. L. and LEWIS T. O. (1970). Approximation of tail probabilities by means of the H_n-transformation. *Texas Tech Univ. Math. Ser.* **6** 237-255.

[19] GRAY, H. L. and SCHUCANY, W. R. (1969). Some limiting cases of the G-transformation. *Math. Comp.* **23** 849-860.

[20] JONES, H. L. (1963). The jackknife method. *Proc. IBM Scientific Computing Symposium on Statistics*, 185-197.

[21] LUBKIN, S. A. (1952). A method of summing infinite series. *J. Res. Nat. Bur. Standards Sect. B.* **48** 228-254, 1952.

[22] MANTEL, N. (1967). Assumption-free estimators using U-statistics and a relationship to the jackknife method. *Biometrics* **23** 567-571.

[23] MILLER, R. (1964). A trustworthy jackknife. *Ann. Math. Statist.* **35** 1594-1605.

[24] MILLER, R. (1968). Jackknifing variances. *Ann. Math. Statist.* **39** 567-582.

[25] MOSTELLER, F. and TUKEY, J. W. (1968). Data analysis, including statistics, *Handbook of Social Psychology*. (G. Lindzey and E. Aronson, eds.) Addison-Wesley, Reading.

[26] QUENOUILLE, M. (1949). Approximate tests of correlation in time-series. *J. Roy. Statist. Soc. Ser. B.* **11** 68-84.

[27] QUENOUILLE, M. (1956). Notes on bias in estimation. *Biometrika* **43** 353-360.

[28] RAO, J. N. K. (1965). A note on the estimation of ratios by Quenouille's method. *Biometrika* **52** 647-649.

[29] RAO, J. N. K. (1969). Ratio and regression estimators. *New Developments in Survey*

Sampling. (J. L. Johnson and Harry Smith, eds.) Wiley-Interscience, New York, 213-234.

[30] RAO, J. N. K. and WEBSTER, J. (1966). On two methods of bias reduction in the estimation of ratios. *Biometrika* **53** 571-577.

[31] RAO, P. S. R. S. (1966). Comparison of four ratio-type estimates under a model. *J. Amer. Statist. Assoc.* **64** 574-580.

[32] RAO, P. S. R. S. and RAO, J. N. K. (1970). Some small sample results for ratio estimators (Abstract). *Ann. Math. Statist.* **41** 1141-1142.

[33] ROBSON, D. S. and WHITLOCK, J. H. (1964). Estimation of a truncation point. *Biometrika* **51** 33-39.

[34] SALSBURG, D. (1971). Testing dose responses on proportions near zero or one with the jackknife. To appear in *Biometrics*.

[35] SCHUCANY, W. R., GRAY, H. L. and OWEN, D. B. (1971). On bias reduction in estimation. *J. Amer. Statist. Assoc.* 00 000-000.

[36] SHANKS, DANIEL. (1955). Nonlinear transformations of divergent and slowly convergent sequences. *J. Math. and Phys.* **34** 1-42.

[37] SHORAK, G. R. (1969). Testing and estimating ratios of scale parameters. *J. Amer. Statist. Assoc.* **64** 999-1013.

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

[39] WATKINS, T. A. (1971). Jackknifing stochastic processes. Ph. D. dissertation, Texas Tech Univ.

[40] WYNN, P. (1956). On a device for computing the $e_m(S_n)$ transformation. *MTAC* **10** 91-96.

Note added in proof. A definition similar to (2.12) has been posed independently by K. P. Burnham in an unpublished paper.