Jackknifing Linear Estimating Equations: Asymptotic Theory and Applications in Stochastic Processes

Subhash Lele

Journal of the Royal Statistical Society. Series B (Methodological), Volume 53, Issue 1 (1991), 253-267.

J. R. Statist. Soc. B (1991)
53, *No.* 1, *pp.* 253–267

Jackknifing Linear Estimating Equations: Asymptotic Theory and Applications in Stochastic Processes

By SUBHASH LELE†

Johns Hopkins University, Baltimore, USA

[Received November 1987. Final revision October 1989]

SUMMARY

Let $(X_1, X_2, \ldots, X_n)$ be a vector of (possibly dependent) random variables having distribution $F(\mathbf{X}, \theta)$. Let $G(\mathbf{X}, \theta) = \Sigma_{i=1}^{n} g_i(\mathbf{X}, \theta) = 0$ be an estimating equation for θ, e.g. the score function or the maximum pseudolikelihood estimating equation in spatial processes. Let θ_n be the estimator obtained from G such that $\theta_n \rightarrow \theta_0$ in probability and $n^{1/2}(\theta_n - \theta_0) \rightarrow N(0, V)$ in distribution. In many situations, it is difficult to derive an analytical expression for V, e.g. for maximum pseudolikelihood estimators for the spatial processes. In this paper, we give a jackknife estimator of V and show that it is weakly consistent. The method consists of deleting one estimating equation (instead of one observation) at a time and thus obtaining the pseudovalues. The method of proof and conditions are similar to those of Reeds with some modifications. The method applies equally to independent and identically distributed random variables, independent but not identically distributed random variables, time- or space-dependent stochastic processes. Our conditions are less severe than Carlstein's who deals with a similar problem of estimating V for dependent observations. We also give some simulation results.

Keywords: ESTIMATING EQUATIONS; JACKKNIFE; SPATIAL PROCESSES; STOCHASTIC PROCESSES

1. INTRODUCTION

Jackknife, bootstrap and other resampling methods of estimating bias, variance and other distributional properties of the sample statistics are extensively used in many areas of applications. Most of the results in these areas assume that the data come from independent and identically distributed random variables. However, in many interesting applications, the data come from a collection of independent but not identically distributed random variables or from a sequence of dependent random variables such as a time series or a spatial process. Not much work has been done in using jackknife techniques in these areas. We quote Miller (1974), section 3.7: 'An area in which the jackknife has had little or no success is time series analysis. Except for the case $g = 2$, the removal of data segments from a serially correlated sequence of observations causes difficulty for the jackknife.' Frangos (1987) echoes the same opinion. Zeger and Brookmeyer (1986) suggest the use of pseudolikelihood for time series regression with censored data and comment: 'The principal disadvantage . . . (of the maximum pseudolikelihood estimator) . . . is that a consistent estimator of variance is difficult to obtain'. Künsch in the discussion of Besag (1986) comments: 'Note that in order to estimate the variance of $\hat{\beta}$ . . . (the maximum pseudolikelihood

†*Address for correspondence*: Department of Biostatistics, Johns Hopkins University, 615 North Wolfe Street, Baltimore, MD 21205, USA.

 0035–9246/91/53253 $2.00

estimator) . . . similar (Monte Carlo) calculations are needed anyhow'. Guyon (1986) proves that the maximum pseudolikelihood estimators for spatial processes are asymptotically normal. However, the asymptotic variance can be calculated only if the joint probability structure is known! This is, in general, very difficult to find. See Besag (1974) for details. In this paper, we propose an extension of the jackknife technique which provides a nonparametric estimate of variance, thus answering these problems.

As a remark, we note here a paper by Carlstein (1986) which addresses the same issue. He has suggested a consistent estimator of variance of an arbitrary statistic calculated from a strictly stationary α-mixing process. He divides the whole series of size n into several subseries of size m_n such that $m_n \to \infty$ but $m_n/n \to 0$. Loosely speaking, our extension is such that we obtain a consistent estimator of the variance for $m_n = n-1$ and $n \to \infty$. In Section 5 we discuss the effect of this on the bias of the estimator. In our case we do not need stationarity, however, we need conditions on the rates of mixing and statistics given by estimating equations.

The idea involved is very easy. Suppose that we have $(X_1, X_2, \ldots, X_n)$, a vector of possibly dependent random variables having distribution $F(\mathbf{X}, \theta)$. Let

$$G(\mathbf{X}, \theta) = \sum_{i=1}^{n} g_i(\mathbf{X}, \theta) = 0$$

be an estimating equation for θ, e.g. the score function or the maximum pseudolikelihood estimating equation (Besag, 1975). We call an estimating equation of this form a linear estimating equation (Godambe, 1960, 1985). To obtain the pseudovalues, we do not delete observations, but rather delete one component of the estimating equation at a time. We show that the jackknife estimate of variance obtained from these pseudovalues is weakly consistent.

The organization of the paper is as follows. Section 2 discusses the linear estimating equations and introduces the jackknife estimate of variance. Section 3 gives the notation, assumptions and the main theoretical result of the paper. The practical validity of the theorem was checked by a simulation study presented in Section 4. In Section 5 we compare our method with that of Carlstein (1986) and give some theoretical and simulation results. We prove the main theorem in Appendix A.

2. LINEAR ESTIMATING EQUATIONS AND JACKKNIFING

Let $(X_1, X_2, \ldots, X_n)$ be a vector of (possibly dependent) random variables. Let the distribution of this vector be F with parameter θ. θ may possibly be vector valued. For simplicity of notation we give all the proofs for univariate θ. The proofs for vector-valued θ are straightforward extensions. We assume that the estimating equation for θ can be written as $\Sigma_{i=1}^{n} g_i(\mathbf{X}, \theta) = 0$. See Godambe (1960, 1985) for general conditions on the g_i.

2.1. *Some Examples of Linear Estimating Equations*

Example 1. $X_1, X_2, \ldots, X_n$ are independent but not identically distributed random variables with densities $f_1(\theta), f_2(\theta), \ldots, f_n(\theta)$ respectively. Then the score function may be written as

$$\sum_{i=1}^{n} \frac{\mathrm{d}}{\mathrm{d}\theta} \{\log f_i(X_i, \theta)\} = \sum_{i=1}^{n} g_i(\mathbf{X}, \theta) = 0.$$

Example 2. A conditional least squares estimating equation for a Markov process (Klimko and Nelson, 1978; Godambe, 1985) can be written as

$$\sum_{i=1}^{n} \{X_i - E(X_i|X_{i-1}, \theta)\} \left\{\frac{\mathrm{d}}{\mathrm{d}\theta} E(X_i|X_{i-1}, \theta)\right\} = \sum_{i=1}^{n} g_i(\mathbf{X}, \theta) = 0.$$

Example 3. A maximum pseudolikelihood estimating equation for spatial processes (Besag, 1975) can be written as

$$\sum_{i=1}^{n} \frac{\mathrm{d}}{\mathrm{d}\theta} \{\log f_i(X_i|X(N(i)), \theta)\} = \sum_{i=1}^{n} g_i(\mathbf{X}, \theta) = 0$$

where $N(i)$ denotes the neighbourhood of the site i.

2.2. *Jackknifing*

In the usual jackknife, we delete one observation at a time and estimate θ from the remaining observations. See Miller (1974) and Efron (1982) for more details. As noted by Miller (1974), removal of data segments from a serially correlated sequence of observations causes difficulty for the jackknife. However, we may conceivably delete a component of a linear estimating equation. In doing this, we are using the information in X_j conditionally but not unconditionally. As we show in the next section, this idea works.

Let

$$G(\mathbf{X}, \theta) = \sum_{i=1}^{n} g_i(\mathbf{X}, \theta) = 0$$

be the original estimating equation and θ_n be the estimate thereof.

Let

$$G^j(\mathbf{X}, \theta) = \sum_{i \neq j} g_i(\mathbf{X}, \theta) = 0$$

and $\theta_{n,-j}$ be the estimate thereof.

The jackknife estimate of θ is

$$\mathrm{JK}\ \theta_n(X_1, X_2, \ldots, X_n) = \theta_n - \frac{n-1}{n} \sum_j (\theta_{n,-j} - \theta_n)$$

$$= \theta_n - \frac{n-1}{n} \sum_j R_{nj} \tag{2.1}$$

where $R_{nj} = \theta_{n,-j} - \theta_n$.

If the estimating equations are uncorrelated in the sense that $E(g_i g_j) = 0$ for all i and j, then the jackknife estimate of variance is

$$\text{JKV } \theta_n = (n-1) \sum_j (R_{nj} - \bar{R}_n)^2 \tag{2.2}$$

where $\bar{R}_n = (1/n)\Sigma_j R_{nj}$. This estimates V, the asymptotic variance of θ_n.

However, modification is needed when the estimating equations are not uncorrelated. Consider example 3 of Section 2.1. It is clear that

$$E\left(\frac{\mathrm{d}}{\mathrm{d}\theta}\{\log f(X_i|X(N(i)),\theta)\}\frac{\mathrm{d}}{\mathrm{d}\theta}\left\{\log f(X_j|X(N(j)),\theta)\right\}\right) \neq 0$$

for $j \in N(i)$. In this case, $E(g_i g_j) \neq 0$ if i and j are neighbours; otherwise it is zero. In such a situation, the jackknife estimate of variance has to estimate these covariances also. The modified estimate is given by

$$\text{JKV}^* \, \theta_n = (n-1) \sum_{i=1}^{n} \sum_{j \in N(i)} (R_{ni} - \bar{R}_n)(R_{nj} - \bar{R}_n) \tag{2.3}$$

where $N(i)$ is the set of sites for which $E(g_i g_j) \neq 0$.

In the first part of the paper we assume that $E(g_i g_j) \neq 0$ for only a finite number of js for a fixed i, i.e. each site has only a finite number of neighbours. In the last section we relax this condition.

The next section gives conditions under which this estimator is consistent.

3. NOTATION, ASSUMPTIONS AND MAIN RESULT

Let $X_1, X_2, \ldots$ be random elements of some measurable space (x, A). Let $\theta \subseteq R^q$. We take $q=1$ for notational simplicity. The results hold for $q > 1$ also. Let E_θ denote expectation for a particular θ.

We make the following assumptions. Let the distribution P of $X_1, X_2 \ldots$ be such that $E_\theta|g_i(\theta)| < \infty$ for all θ, all i, and there is an interior point $\theta_0 \in \Theta$ with $E_\theta(g_i(\theta_0)) = 0$, $E_\theta(g_i(\theta)) \neq 0$ $(\theta \neq \theta_0)$ for all i.

There exists a compact neighbourhood K of θ_0, such that, for all θ in K,

(a) $E_\theta|g_i(\theta_0)|^2 < \infty \qquad \forall\, i,$

(b) $E_\theta \left| \frac{\partial}{\partial\theta} g_i(\theta)|_{\theta_0} \right| < \infty \qquad \forall\, i,$

(c) $\bar{I}_n = \frac{1}{n}\sum_{i=1}^{n} E_\theta\left(\frac{\partial}{\partial\theta} g_i(\theta)\right), \qquad \bar{J}_n = \frac{1}{n} E_\theta\left(\sum_{i=1}^{n} g_i(\theta)\right)^2,$
$\bar{I} = \lim_{n\to\infty}\bar{I}_n$ is non-singular, $\bar{J} = \lim_{n\to\infty}\bar{J}_n$ and $V = \bar{I}^{-1}\bar{J}\bar{I}^{-1}$,

(d) $\left|\frac{\partial}{\partial\theta} g_i(s) - \frac{\partial}{\partial\theta} g_i(t)\right| \leqslant m(x)\,|s-t|^\lambda$
uniformly in i for some random variable X (this is a Lipshitz-type condition),

(e) $E_\theta(g_i g_j) = 0$ for $d(i,j) > P$ for some $P < \infty$ (this says that the estimating equations should become uncorrelated after certain finite distance; here $d(i,j)$ is distance between site i and site j),

(f) $n^{1/2}(\theta_n - \theta_0) \to N(0, V)$ in distribution and

(g) $\theta_n \to \theta_0$ in probability.

We do not state the additional conditions needed to derive assumptions (f) and (g). They are assumed to be satisfied in each individual case. For the regularity conditions in spatial processes, see Guyon (1986). For the regularity conditions in stochastic processes, see Hall and Heyde (1980) or Klimko and Nelson (1978).

Theorem 1. Under these assumptions, with JK θ_n as given in equation (2.1) and JKV θ_n as given in equation (2.2) or equation (2.3),

(a) $n^{1/2}(\text{JK } \theta_n - \theta_n) \to 0$ in probability,
(b) JKV $\theta_n \to V$ in probability and
(c) $n^{1/2}(\text{JKV } \theta_n)^{-1/2}(\text{JK } \theta_n - \theta_0) \to N(0, I)$ in distribution.

For the proof see Appendix A.

4. SIMULATION RESULTS

We applied this method of jackknifing to three different situations: the first-order Gaussian autoregressive process, bilinear time series and a two-neighbour line process. In all these cases we know the asymptotic variance and hence we can check the performance of the procedure. To generate the random numbers we used International Mathematical and Statistical Libraries (1984) subroutine GGNML. The simulation results indicate that the jackknife estimate of variance is consistent.

4.1. *Gaussian Autoregressive Process*

The Gaussian first-order autoregressive (AR(1)) process that we consider here is

$$X_{i+1} = \rho X_i + \epsilon_{i+1}$$

where $|\rho| < 1$ and ϵ_i are independent identically distributed normal random variables with zero mean and unit variance. We consider the maximum likelihood estimating equation for estimating ρ

$$\sum_{i=1}^{n} (X_i - \rho X_{i-1}) X_{i-1} = 0.$$

It is well known that the asymptotic variance of this estimator is $1 - \rho^2$ (Basawa and Prakasa Rao, 1980). Table 1 contains the simulation results. We can easily see that the jackknife estimate of variance is consistent.

4.2. *Bilinear Time Series*

For a bilinear time series (Granger and Anderson, 1978) we consider the model

$$X_{i+1} = \rho X_i + \epsilon_{i+1} + \beta X_i \epsilon_{i+1}$$

where the ϵ_i are independent identically distributed Gaussian random variables with zero mean and unit variance.

We consider the conditional least squares estimating equation for ρ,

$$\sum_{i=1}^{n} (X_i - \rho X_{i-1}) X_{i-1} = 0,$$

which is identical with the estimating equation in the previous case. However, the asymptotic variance of $\hat{\rho}$ is different and can be calculated using standard methods of

TABLE 1
Jackknife estimate of variance for a first-order autoregressive process†

ρ	*True asymptotic variance* $(1-\rho^2)$	*Jackknife variance JKV*‡	*Standard deviation of JKV*§
0.0	1.00	0.9622	0.1995
0.1	0.99	1.0028	0.2109
0.2	0.96	0.9654	0.1965
0.3	0.91	0.9436	0.1977
0.4	0.84	0.8584	0.1849
0.5	0.75	0.7642	0.1871
0.6	0.64	0.6360	0.1543
0.7	0.51	0.5581	0.1833
0.8	0.36	0.4057	0.1319
0.9	0.19	0.2286	0.0964

†100 simulations, $n = 100$.
‡JKV is the average of the estimates of V from 100 simulations.
§The standard deviation of JKV is the standard deviation of the estimates of V from 100 simulations.

TABLE 2
Jackknife estimate of variance for a bilinear time series†

ρ		*Estimates for the following values of* β: 0.1	0.15	0.20
0.5	AV‡	0.8150	0.9170	1.0585
	JKV§	0.8171	0.8780	1.0360
	SD§§	0.2876	0.3488	0.4524
0.75	AV‡	0.5279	0.6672	0.9063
	JKV§	0.5324	0.6127	0.8027
	SD§§	0.1559	0.2274	0.4549

†100 simulations, $n = 100$ for the model $X_n = \rho X_{n-1} + \beta \epsilon_n X_{n-1} + \epsilon_n$.
‡Theoretical asymptotic variance.
§JKV, the jackknife estimate of variance, is the average of the estimates of V from 100 simulations.
§§SD, the standard deviation of JKV, is the standard deviation of the estimates of V from 100 simulations.

Taylor series expansions. We want to check whether the jackknife method suggested earlier is robust against model specification provided that the g_i remain uncorrelated. The results are shown in Table 2. The jackknife technique is estimating the variance of $\hat{\rho}$ correctly.

4.3. *Spatial Process on Real Line*

In this case we consider a spatial process on the real line. The model that we consider is such that

$$E(X_i \mid X_{i-1}, X_{i+1}) = \alpha(X_{i-1} + X_{i+1}),$$
$$\mathrm{var}(X_i \mid X_{i-1}, X_{i+1}) = \sigma^2.$$

We estimate α using Besag's method of maximum pseudolikelihood. The estimating equation is

$$\sum_{i=1}^{n} \{X_i-\alpha(X_{i-1}+X_{i+1})\}(X_{i-1}+X_{i+1}) = 0.$$

Guyon (1986) and Besag (1977) give the analytic expression for the asymptotic variance.

We give the simulation results in Table 3. It shows that the jackknife estimate of variance is consistent.

5. JACKKNIFING ESTIMATING EQUATIONS WITH CORRELATED COMPONENTS

In this section, we deal with the question: what if $E(g_i g_j) \neq 0$ for all i and j? To show that this is not vacuous, we consider the estimation of the mean in the Gaussian AR(1) process. The model under consideration is

$$X_{i+1} = \mu + \rho X_i + \epsilon_{i+1}.$$

Let $\overline{X}$ estimate μ. Then the corresponding estimating equation may be written as

$$\sum_{i=1}^{n} (X_i-\mu) = \sum_{i=1}^{n} g_i(\mathbf{X}, \mu) = 0.$$

It is obvious that $E(g_i g_j) \neq 0$ for all i and j. In such a situation we modify our estimator as follows:

$$\mathrm{JKV}_{M_n}\, \theta_n = (n-1) \sum_{i=1}^{n} \sum_{j \in M_n(i)} (R_{ni}-\overline{R}_n)(R_{nj}-\overline{R}_n) \tag{5.1}$$

TABLE 3
Jackknife estimate of variance for a spatial process on the real line with two neighbours†

α	*Asymptotic variance AV*	*Jackknife variance JKV*‡	*Standard deviation of JKV*§
0.00000	1.00000	0.97086	0.1375
0.09900	0.94910	0.92747	0.1769
0.19231	0.78779	0.78946	0.1744
0.27523	0.58665	0.59872	0.1375
0.34483	0.38970	0.38660	0.1179
0.40000	0.23040	0.22367	0.0800
0.44118	0.11974	0.13081	0.0548
0.46980	0.05277	0.06130	0.0316
0.48781	0.01792	0.01995	0.0130
0.49724	0.00336	0.00374	0.0053

†50 simulations, $n=300$.
‡JKV is the average of the estimates of V from 50 simulations.
§The standard deviation of JKV is the standard deviation of the estimates of V from 50 simulations.

where $M_n(i)$ denotes a neighbourhood of i with $2M_n+1$ sites. Now if we let $n \to \infty$, $M_n \to \infty$ such that $M_n/n \to 0$, it can be seen very easily that $\mathrm{JKV}_{M_n}\,\theta_n \to V$. This scheme is very similar to that of Carlstein (1986). We now compare our scheme with that of Carlstein. We consider the case of estimating the variance of $\bar{X}$ where X_i are observations from a Gaussian AR(1) process, with $\mu = 0$. We know that

$$(\bar{X}-\mu)\sqrt{n} \xrightarrow{\mathrm{D}} N(0, 1/(1-\rho)^2).$$

5.1. *Carlstein's Scheme*

Let $X_1, X_2, \ldots, X_n$ be the observations from a Gaussian AR(1) process.

(a) Divide the series into k_n subseries each of length M_n.

(b) Calculate $\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_{K_n}$ and $\bar{\bar{X}} = (1/K_n)\sum_{i=1}^{K_n}\bar{X}_i$.

(c) $\hat{\sigma}^2 = \dfrac{M_n}{K_n}\displaystyle\sum_{i=1}^{K_n}(\bar{X}_i - \bar{\bar{X}})^2$.

In his paper, Carlstein shows that

$$\sigma^2 - E(\hat{\sigma}^2) = O\left(\frac{1}{M_n}\right)$$

for fixed M_n. Now we consider the scheme presented in this paper. The jackknife estimator of variance is

$$\mathrm{JKV}\ \theta_n = (n-1)\sum_{i=1}^{n}\sum_{j\in M_n(i)}(\bar{X}_{(i)}-\bar{X})(\bar{X}_{(j)}-\bar{X})$$

where

$$\bar{X}_{(i)} = \frac{1}{n-1}\sum_{k\neq i}X_k, \qquad \bar{X}_{(j)} = \frac{1}{n-1}\sum_{k\neq j}X_k$$

$$M_n(i) = \{X_{i-M_n}, X_{i-M_n+1}, X_{i-1}, X_i, X_{i+1}, \ldots, X_{i+M_n}\}.$$

After some algebra, we can write

$$\mathrm{JKV}\ \theta_n = \frac{1}{n-1}\sum_{i=1}^{n}\sum_{j\in M_n(i)}(\bar{X}_i-\bar{X})(X_j-\bar{X}).$$

Assuming that for each i all the neighbouring values are available, we can see that

$$\begin{aligned}
E(\mathrm{JKV}\ \theta_n) &= \frac{1}{n-1}\left\{\frac{n}{1-\rho^2} + \frac{2n(\rho+\rho^2+\ldots+\rho^{M_n})}{1-\rho^2}\right\}\\
&= \frac{n}{n-1}\left\{\frac{1+2(\rho+\ldots\rho^{M_n})}{1-\rho^2}\right\}\\
&= \frac{n}{n-1}\left\{\frac{2(1-\rho^{M_{n+1}})/(1-\rho)-1}{1-\rho^2}\right\}\\
&= \frac{n}{n-1}\left\{\frac{1+\rho-2\rho^{M_{n+1}}}{(1-\rho^2)(1-\rho)}\right\}
\end{aligned}$$

$$\rightarrow \frac{1+\rho-2\rho^{M_{n+1}}}{(1-\rho^2)(1-\rho)}$$

Thus, for the jackknife estimator, the asymptotic bias for fixed M_n is $o(1/M_n)$. We are able to reduce the bias because we are using more information that is available in the data than Carlstein's scheme does.

To confirm our claim we did some simulation studies for this model. The results are displayed in Tables 4 and 5. The jackknife estimate of variance clearly has less bias. The calculations of bias and mean-squared error were done following the formulae on pp. 1176 and 1178 of Carlstein (1986).

In applying the jackknife technique we have chosen the value of M_n arbitrarily. This choice is subjective in general. The same comment applies to Carlstein's method. In practice we may draw a correlogram to choose a reasonable value.

6. MISCELLANEOUS REMARKS

Remark 1. It can be easily seen that, when the estimating equations are correlated, the jackknife estimate of variance can be negative. This problem will not be acute when there is a large sample; however, this is still a somewhat troubling fact.

TABLE 4
Jackknife estimate of variance of X for a first-order autoregressive process†

ρ	*Asymptotic variance*		*Jackknife variance* $n=100$	$n=200$	$n=300$
0.3	2.04	JKV‡	1.90	1.89	1.9918
		SD§	0.8062	0.5196	0.4850
		Bias	0.14	0.15	0.0482
		Mean-squared error	0.6696	0.2925	0.2375
0.5	4.00	JKV‡	3.51	3.54	3.75
		SD§	1.4663	0.9434	0.8775
		Bias	0.49	0.46	0.25
		Mean-squared error	2.3901	1.1016	0.8325

†100 simulations, $M_n=50$.
‡JKV is the average of the estimates of V from 100 simulations.
§SD, the standard deviation of JKV, is the standard deviation of the estimates of V from 100 simulations.

TABLE 5
Carlstein's method with optimal M_n

ρ	*Asymptotic variance*		$n=100$	$n=200$	$n=300$
0.3	2.04	Bias	0.3337	0.2685	0.2685
		Mean-squared error	0.4323	0.2717	0.2071
0.5	4.00	Bias	0.8750	0.7559	0.6641
		Mean-squared error	2.7858	1.7631	1.3491

A referee has suggested the following estimator with smooth weights instead of the hard cut-off in our paper:

$$\mathrm{JKV} = (n-1) \sum_{i} \sum_{j} (R_{n_i} - \bar{R}_n)(R_{n_j} - \bar{R}_n)\, \lambda(d(i,j)).$$

From experience with spectral analysis an estimate of this form may be preferable.

Remark 2. An obvious comment on the jackknife method of estimation of variance is: 'why not use the analogue of observed information?', i.e. $V = \hat{I}_n^{-1} \hat{J}_n \hat{I}_n^{-1}$ where

$$\hat{I}_n = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial \theta} g_i(X; \theta) \big|_{\theta=\theta_n}$$

and

$$\hat{J}_n = \frac{1}{n} \sum_{i=1}^{n} \sum_{j \in N(i)} g_i(X, \theta_n)\, g_j(X, \theta_n).$$

See Royall (1986) for a similar estimator in the independent identically distributed random variables set-up. We give below conclusions from a small simulation study. The simulation study was done for the autoregressive process of order 1 as described in Section 4. The sample size n was 100 and the number of simulations was also 100. The basic conclusions are as follows.

(a) The jackknife estimator overestimates the variance whereas the observed information underestimates it. See Efron and Stein (1981) for a similar result in the independent and identically distributed random variables set-up.
(b) The confidence intervals obtained by using the jackknife estimate of variance have more accurate nominal coverage probability than those obtained by using the observed information.

These findings raise two possible issues. Does Efron–Stein inequality hold true for jackknifing linear estimating equations? Is JKV a better 'Studentizing' variance estimator than observed information? To date we do not have answers to these questions.

ACKNOWLEDGEMENTS

The author is grateful to Professor B. G. Lindsay and Dr M. B. Rajarshi for many useful comments. The author would also like to thank the referees and the Editor for a very careful reading of the manuscript and useful comments.

APPENDIX A

We prove the main result of the paper. The proof is very similar to that given in Reeds (1978) for jackknifing maximum likelihood estimates. To simplify the presentation of our proof, we follow his notation closely and use some of his results without proof.

The basic approach of the proof is as follows. By using the expansions of Reeds (1978), we show that

$$\text{JK } \theta_n = \theta_n + R_n \ldots \quad (A.1)$$

and

$$\text{JKV } \theta_n = V_n + R_n^* \ldots . \quad (A.2)$$

These expressions are obtained in equations (A.4)–(A.6).

Then to prove part (a) of theorem 1, we show that $R_n\sqrt{n} \to 0$ in probability. To show part (b), we show that $V_n \to V$ in probability and $R_n^* \to 0$ in probability.

The convergences of R_n and R_n^* are proved by using lemma 2 repeatedly. The proof is straightforward but needs careful bookkeeping.

Some Preliminaries

To avoid unnecessary repetition we follow the notation of Reeds (1978) closely and point out the relevant sections and lemmas from his paper. We shall state and prove only those results which are different from Reeds's.

We assume all the preliminaries on spaces as given in Reeds (1978), section 3. Let $G = \Sigma_{i=1}^n g_i(\theta) = 0$ be the estimating function. Then in Reeds's notation, we obtain the result that the random functions $g_i(\theta)$ belong to B and that $E_{\theta_0}\|g_i(\theta)\| < \infty$. Hence

$$E_0(g_i(\theta)) = \gamma_i(\theta) < \infty \qquad (\gamma_i(\theta_0) = 0),$$

$$Y_i = \alpha_2(g_i - \gamma_i) - g_i(\mathbf{x}, \theta_0) - \gamma_i(\theta_0) = g_i(\mathbf{x}, \theta_0) \in \mathbb{V},$$

$$Z_i = \alpha_2(g_i - \gamma_i) = g_i'(\mathbf{x}, \theta_0) - \gamma_i'(\theta_0) \in \mathbb{V} \times \mathbb{V}^*,$$

$$\phi_i = \alpha_3(g_i - \gamma_i) = g_i(\mathbf{x}, \theta) - \gamma_i - g_i(\mathbf{x}, \theta_0) - \{g_i'(\mathbf{x}, \theta_0) - \gamma_i'(\theta_0)\}(\theta - \theta_0) \in B_0.$$

Here

$$\gamma_i'(\theta_0) = \frac{\partial}{\partial\theta}\gamma_i(\theta)|_{\theta_0}.$$

Let $\overline{U}_n = (\overline{Y}_n, \overline{Z}_n, \overline{\phi}_n)$. The equation $\overline{g}_n(\theta) = 0$ can be written as

$$\gamma(\theta) + \overline{Y}_n + (\theta - \theta_0)\overline{Z}_n + \overline{\phi}_n(\theta) = 0. \quad (A.3)$$

For sufficiently small vectors in $U = (y, z, \phi) \in \mathbb{U}$, the equation

$$\gamma(\theta) + y + (\theta - \theta_0)z + \phi(\theta) = 0$$

has a solution in θ which we denote by $\theta = f(y, z, \phi)$.

Now we assume the knowledge of section 4 of Reeds, particularly his lemmas 1 and 2 which give the necessary expansions. We also assume all his notation. None of this section depends on the independent and identically distributed nature of the random variables. Now we state two lemmas which facilitate us to assert that O statements about A_{nj}, B_{nj}, C_{nj} and D_{nj} on p. 736 of Reeds's paper can be replaced by O_p statements. Also in our case $f_y(0) = \overline{I}_0^{-1}$.

Let $\overline{U}_n$, $\overline{Y}_n$ etc. and $\overline{U}_{nj}$, $\overline{Y}_{nj}$ etc. be as defined in section 3, p. 732, and section 5, p. 735, of Reeds (1978). Then

$$\text{JK } \theta_n = f(\overline{U}_n) - \frac{n-1}{n}\sum_{j=1}^n R_{nj}.$$

Lemma 1.

(a) $\lim_{n\to\infty} P(\overline{U}_n \in \mathbb{U}) = 1$ and

(b) $\lim_{n\to\infty} P(\overline{U}_{nj} \in \mathbb{U}) = 1$

where $\mathbb{U}$ is as in lemma 1, section 4, of Reeds (1978).

Proof. Assertion (a) is trivial, considering the fact that $\bar{U}_n$ has mean zero and the weak law of large numbers can be applied. The fact that the U_j are bounded in probability and $\bar{U}_{nj} = n\bar{U}_n/(n-1) - U_j/(n-1)$ together give (b).

Lemma 2. Suppose that

(a) $(X_n, Y_n) \to 0$ in probability and
(b) $(X_n, Y_n) \in \mathbb{U}$ implies that $f(X_n) - g(Y_n) = O(h(X_n, Y_n))$ for specified functions f, g and h:

then

$$f(X_n) - g(Y_n) = O_p(h(X_n, Y_n)).$$

Proof. Part (a) implies that

$$P((X_n, Y_n) \in \mathbb{U}) \to 1.$$

This and part (b) together imply that

$$P\left(\left| \frac{f(X_n) - g(Y_n)}{h(X_n, Y_n)} \right| \leqslant M \right) \to 1.$$

Hence the proof follows.

Consider

$$\text{JK } \theta_n = f(\bar{U}_n) - \frac{n-1}{n} \sum_j R_{nj}. \tag{A.4}$$

Again following Reeds (1978), we find that

$$R_{nj} = f_y(\bar{U}_n)(\bar{Y}_{nj} - \bar{Y}_n) + f_z(\bar{U}_n)(\bar{Z}_{nj} - \bar{Z}_n) + A_{nj} + B_{nj} + C_{nj} + D_{nj} \tag{A.5}$$

where $A_{nj} = O_p(\|\phi_{nj} - \phi_n\| \, |\bar{Y}_{nj}|^{1+\lambda})$ etc.

The only difference here is that Os are replaced by O_ps, using lemmas 1 and 2. This is the expression needed in equation (A.1). Hence to prove part (a) of theorem 1, we need to prove that

$$\sqrt{n} \sum_j R_{nj} = \sqrt{n} \sum_j (A_{nj} + B_{nj} + C_{nj} + D_{nj}) \to 0 \tag{A.6}$$

in probability.

Now consider the jackknife estimate of the variance. We note the following facts:

$$\text{JKV } \theta_n \approx \frac{1}{n} \sum_{j=1}^{n} (nR_{nj})^2;$$

$$\text{JKV* } \theta_n \approx \frac{1}{n} \sum_{i=1}^{n} \sum_{j \in N(i)} n^2 R_{ni} R_{nj}.$$

In the uncorrelated case,

$$V_n = \frac{1}{n} \sum_{j=1}^{n} f_y^2(\bar{U}_n)(Y_j - \bar{Y}_n)^2 \to V$$

in probability and, in the correlated case,

$$V_n^* = \frac{1}{n} \sum_{i=1}^{n} \sum_{j \in N(i)} f_y^2(\overline{U}_n)(Y_i - \overline{Y}_n)(Y_j - \overline{Y}_n) \to V$$

in probability. Thus to prove part (b) of theorem 1 we need

(a) JKV $\theta_n - V_n \approx (1/n) \Sigma_j (nR_{nj})^2 - (1/n) \Sigma_j f_y^2(\overline{U}_n)(Y_j - \overline{Y}_n) \to 0$ in probability and
(b) JKV* $\theta_n - V_n^* \approx (1/n) \Sigma_i \Sigma_{j \in N(i)} n^2 R_{ni} R_{nj} - (1/n) \Sigma_i \Sigma_{j \in N(i)} f_y^2(\overline{U}_n)(Y_i - \overline{Y}_n) \to 0$ in probability.

These expressions are similar to equation (A.2).

We now state and prove lemma 3. This lemma is the most important part of the proof. It facilitates the proof of $n^{1/2} \Sigma_i |A_{ni}|$ etc. $\to 0$ in probability. Reeds (1978) needs a lemma similar to this (his lemma 3 on p. 736). He proves it under the assumption of independent identically distributed random variables; we extend it to the non-independent identically distributed random variables case. The method of proof used here can be applied to show some of the convergences claimed earlier.

Lemma 3. Let $(V_1, W_1), (V_2, W_2), \ldots,$ be a sequence of random vectors. Let

(a) $T_n = n^{\alpha} \sum_{j=1}^{n} \left| \frac{V_j - \overline{V}_n}{n} \right|^{\beta} \left| \frac{W_j - \overline{W}_n}{n} \right|^{\nu}$ and

(b) $T_n^* = n^{\alpha} \sum_{i=1}^{n} \sum_{j \in N(i)} \left| \frac{V_i - \overline{V}_n}{n} \right|^{\beta} \left| \frac{W_j - \overline{W}_n}{n} \right|^{\nu}$

where β and ν are non-negative real numbers.

Let Z_{ij} be $|V_i|^{\beta} |W_j|^{\nu}$ or $|V_i|^{\beta}$ or $|W_j|^{\nu}$ and β and ν in (0, 2]. Then $T_n \to 0$ ($T_n^* \to 0$) in probability provided that

(a) $|\overline{V}_n|$ and $|\overline{W}_n|$ converge to a finite quantity,

(b) $\frac{1}{n} \sum_{i=1}^{n} \sum_{j \in N(i)} E(Z_{ij}) \to c < \infty$,

(c) $\operatorname{var}\left(\sum_{i=1}^{n} \sum_{j \in N(i)} Z_{ij} \right) = O(n)$ and

(d) $\beta \geqslant 2\alpha$ and $\beta + \nu > \alpha + 1$.

Proof. We prove the convergence of T_n^*. Convergence of T_n follows on the same lines. Note that

$$|V_i - \overline{V}_n|^{\beta} |W_j - \overline{W}_n|^{\nu} \leqslant 2^{\beta+\nu}(|V_i|^{\beta} + |V_n|^{\beta})(|W_j|^{\nu} + |W_n|^{\nu}).$$

Hence

$$\sum_{i=1}^{n} \sum_{j \in N(i)} |V_i - \overline{V}_n|^{\beta} |W_j - \overline{W}_n|^{\nu} \leqslant \sum_{i=1}^{n} \sum_{j \in N(i)} 2^{\beta+\nu}(|V_i|^{\beta} - |\overline{V}_n|^{\beta})(|W_j|^{\nu} - |\overline{W}_n|^{\nu}).$$

Consider

$$T_n^{**} = n^{\alpha} \sum_{i=1}^{n} \sum_{j \in N(i)} \left| \frac{V_i'}{n} \right|^{\beta} \left| \frac{W_j'}{n} \right|^{\nu}.$$

Then taking (V_i', W_j') equal to $(1, 1)$, $(1, |W_j|)$, $(|V_i|, 1)$ and $(|V_i|, |W_j|)$, we can dominate T_n^* by a weighted linear combination of T_n^{**} with weights being monomial functions of $|\overline{V}_n|$ and $|\overline{W}_n|$. Thus it suffices to show that $T_n^{**} \to 0$ in probability.

Let $Z_{ij} = |V_i'|^\beta |W_j'|^\nu$. Then by assumptions (b)–(d) it follows that

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j\in N(i)}\{Z_{ij}-E(Z_{ij})\}\to 0$$

in probability. Now

$$T_n^{**} = n^\alpha \sum_{i=1}^{n}\sum_{j\in N(i)}\left|\frac{V_i'}{n}\right|^\beta \left|\frac{W_j'}{n}\right|^\nu.$$

$$= n^{\alpha-\beta-\lambda}\sum_{i=1}^{n}\sum_{j\in N(i)}|V_i'|^\beta |W_j'|^\nu.$$

Let $\lambda = \beta + \nu - \alpha - 1 > 0$. Then

$$T_n^{**} - n^{-(1+\lambda)}\sum_{i=1}^{n}\sum_{j\in N(i)} Z_{ij}\to 0$$

in probability. Since $(1/n)\,\Sigma_{i=1}^n \Sigma_{j\in N(i)} Z_{ij}\to c<\infty$ in probability and $n^{-\lambda}\to 0$ almost surely, hence the proof follows.

In the proof of the main theorem, the (V_j, W_j) vectors are (ϕ_j, Z_j), (ϕ_j, Y_j) and (Z_j, Y_j). By lemma 1 we know that $|\overline{V}_n'|$, $|\overline{W}_n'|$ are convergent since $\overline{\phi}_n\to 0$, $\overline{Z}_n\to 0$ and $\overline{Y}_n\to 0$ in probability. Conditions (b) and (c) are essentially those that are needed to prove the consistency of the estimator θ_n. In the stochastic processes and spatial processes setting these correspond to strong or uniform mixing of the process with exponential rate. Some pertinent references are Hall and Heyde (1980) and Guyon (1986).

Now the convergences of $n^{1/2}\Sigma_{i=1}^n|A_n|\to 0$ etc. follow as in Reeds (1978). The proof of part (a) of theorem 1 is now immediate. The proof of part (b) in the uncorrelated case is complete if we show that

$$S_n = n\sum_{j=1}^{n}|f_z(\overline{U}_n)(\overline{Z}_{nj}-\overline{Z}_n)^2|\to 0$$

in probability. By lemma 2(ii) of Reeds's section 4, we know that

$$S_n = O_p\left(|n^{1/2}\,\overline{Y}_n|^2\sum_{j=1}^{n}|\overline{Z}_{nj}-\overline{Z}_n|^2\right)$$

and $|n^{1/2}\,\overline{Y}_n|^2\to\chi^2$, random variable, $\Sigma_{j=1}^n|\overline{Z}_{nj}-\overline{Z}_n|^2\to 0$ in probability. Hence $S_n\to 0$ in probability. The proof for the correlated case is essentially the same.

REFERENCES

Basawa, I. V. and Prakasa Rao, B. L. S. (1980) *Statistical Inference for Stochastic Processes*. New York: Academic Press.

Besag, J. (1974) Spatial interaction and the statistical analysis of lattice systems (with discussion). *J. R. Statist. Soc.* B, **36**, 192–236.

—— (1975) Statistical analysis of non-lattice data. *Statistician*, **14**, 179–195.

—— (1977) Efficiency of pseudo-likelihood estimators for simple Gaussian fields. *Biometrika*, **64**, 616–618.

—— (1986) On the statistical analysis of dirty pictures (with discussion). *J. R. Statist. Soc.* B, **48**, 259–302.

Carlstein, E. (1986) The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *Ann. Statist.*, **14**, 1171–1179.

Efron, B. (1982) *The Jackknife, the Bootstrap and Other Resampling Plans*. Philadelphia: Society for Industrial and Applied Mathematics.

Efron, B. and Stein, C. (1981) The jackknife estimate of variance. *Ann. Statist.*, **9**, 586–596.

Frangos, C. C. (1987) An updated bibliography on the jackknife method. *Communs Statist.*, **16**, 1543–1584.

Godambe, V. P. (1960) An optimum property of regular maximum likelihood estimation. *Ann. Math. Statist.*, **31**, 1208–1211.

—— (1985) The foundations of finite sample estimation in stochastic processes. *Biometrika*, **72**, 419–428.

Granger, C. W. J. and Anderson, A. P. (1978) *An Introduction to Bilinear Time Series Models*. Göttingen: Vandenhoeck and Ruprecht.

Guyon, X. (1986) Estimation d'un champ par pseudo-vraisemblance conditionneue: étude asymptotique et application au cas Markovian. *Act. Geme Rec. Fr.-Belg. Statist.*, to be published.

Hall, P. and Hyde, C. C. (1980) *Martingale Limit Theory and its Application*. New York: Academic Press.

International Mathematical and Statistical Libraries (1984) *Fortran Subroutines for Mathematics and Statistics*, vol. 2. Houston: IMSL.

Klimko, L. A. and Nelson, P. I. (1978) On conditional least squares estimation for stochastic processes. *Ann. Statist.*, **6**, 629–642.

Miller, Jr, R. G. (1974) The jackknife—a review. *Biometrika*, **61**, 1–17.

Reeds, J. (1978) Jackknifing maximum likelihood estimates. *Ann. Statist.*, **6**, 727–739.

Royall, R. M. (1986) Model robust intervals using maximum likelihood estimators. *Int. Statist. Rev.*, **54**, 221–226.

Zeger, S. L. and Brookmeyer, R. (1986) Regression analysis with censored autocorrelated data. *J. Am. Statist. Ass.*, **81**, 722–729.