The Annals of Statistics
1993, Vol. 21, No. 3, 1166–1181

SMOOTHED EMPIRICAL LIKELIHOOD CONFIDENCE INTERVALS FOR QUANTILES

BY SONG XI CHEN AND PETER HALL

CMA, Australian National University; and CMA, Australian National University, CSIRO Division of Mathematics and Statistics, Sydney

Standard empirical likelihood confidence intervals for quantiles are identical to sign-test intervals. They have relatively large coverage error, of size $n^{-1/2}$, even though they are two-sided intervals. We show that smoothed empirical likelihood confidence intervals for quantiles have coverage error of order n^{-1}, and may be Bartlett-corrected to produce intervals with an error of order only n^{-2}. Necessary and sufficient conditions on the smoothing parameter, in order for these sizes of error to be attained, are derived. The effects of smoothing on the positions of endpoints of the intervals are analysed, and shown to be only of second order.

1. Introduction. Empirical likelihood methods for constructing confidence regions were introduced by Owen (1988, 1990). One of their advantages is that they enable the shape of a region, such as the degree of asymmetry in the case of a confidence interval, to be determined "automatically" by the sample. In certain regular cases, empirical likelihood regions are Bartlett-correctable, meaning that an empirical correction for scale reduces the order of magnitude of coverage error from n^{-1} to n^{-2}, where n denotes sample size. See DiCiccio, Hall and Romano (1991).

Almost all theoretical development of empirical likelihood has focussed on the case where the statistic of interest is a smooth function of means. For example, it is only in this case that coverage error has been shown to be of order n^{-1}, reducible to n^{-2} by Bartlett correction. A survey of developments in this setting has been given by Hall and La Scala (1990). Owen (1988) has noted that, when applied to the problem of constructing confidence intervals for a population quantile (in particular, for the median), empirical likelihood reproduces precisely the so-called sign-test or binomial-method interval. This is reassuring, but it does show that in the context of quantile estimation, straight empirical likelihood has nothing to offer over existing techniques. One of the disadvantages of the sign-test method is that it is usually unable to deliver confidence intervals with coverage accuracy better than $n^{-1/2}$. This is true for both one- and two-sided intervals, and arises because of the discreteness of the binomial distribution, which determines the true coverage probability. By way of comparison, even the most rudimentary of normal-approximation methods for constructing a two-sided confidence interval for the population mean has coverage error of order n^{-1}.

Received August 1991; revised July 1992.
AMS 1991 *subject classifications*. Primary 62G15; secondary 62G30.
Key words and phrases. Bandwidth, Bartlett correction, bootstrap, confidence interval, empirical likelihood, kernel, median, quantile, sign test, smoothing.

Our aim in this paper is to show that by appropriately smoothing the empirical likelihood method, coverage accuracy may be improved from order $n^{-1/2}$ to order n^{-1}. We demonstrate that this improvement is available for a wide range of choices of the smoothing parameter, so that it is not necessary to accurately determine an "optimal" value of the parameter. Furthermore, we show that smoothed empirical likelihood is Bartlett-correctable. That is, an empirical correction for scale can reduce the size of coverage error from order n^{-1} to order n^{-2}.

We also establish a very general version of Wilks' theorem in the context of empirical likelihood for quantiles. This result provides necessary and sufficient conditions on the range within which the smoothing parameter must lie if the asymptotic distribution of the empirical log likelihood ratio statistic is to be central chi-squared. Furthermore, we derive necessary and sufficient conditions on the smoothing parameter for the error in the chi-squared approximation to be $O(n^{-1})$, and also for the error after Bartlett correction to be $O(n^{-2})$. We suggest a particularly simple version of the Bartlett correction that produces confidence intervals with coverage error $o(n^{-1})$, although not quite $O(n^{-2})$.

Section 2 describes empirical likelihood methods for constructing confidence intervals for quantiles. Section 3 presents our results about Wilks' theorem and the order of approximation there and Section 4 discusses analogous results for Bartlett correction. A small simulation study is presented in Section 5. Proofs are deferred to Section 6.

One competing approach to determining confidence intervals for quantiles is based on interpolation of sign-test intervals. This method has only been developed in the case of the median, and no convergence rates have been given for the coverage error; see, for example, Sheather (1987) and Sheather and McKean (1987). Unpublished work of R. Beran and P. Hall shows that the approach may be satisfactorily generalized to general quantiles, and that the convergence rate is $O(n^{-1})$. Most importantly, that rate cannot be reduced to $O(n^{-2})$ by higher-order interpolation. Therefore, in the sense of coverage accuracy, the Bartlett-corrected, smoothed empirical likelihood approach suggested in the present paper is particularly competitive with methods based on interpolated sign-test intervals.

Other techniques may be based on the asymptotic distribution of a sample quantile. However, that involves the unknown value of the probability density at the population quantile. While the latter may be estimated, choice of bandwidth in this problem is a rather tricky matter, and even with an optimal choice of bandwidth the convergence rate of coverage error is of larger order than n^{-1}; see Hall and Sheather (1988). Duffy and Santner (1990) have recently surveyed different approaches to the setting of confidence intervals for quantiles.

An excellent expository paper by Barndorff-Nielsen and Cox (1979) develops the subject of Edgeworth expansion to the point where it produces rather general forms of the Bartlett correction, in parametric problems. These include one of the forms suggested by Hayakawa (1977). Hall and La Scala (1990) have also discussed the implications of Bartlett correction in the context of

Edgeworth expansion, this time for empirical likelihood. However, it should be stressed that there are extensive differences between proofs of the validity of Bartlett correction in parametric and nonparametric (i.e., empirical likelihood) contexts. Indeed, the fact that Bartlett correction is applicable to empirical likelihood is not, to us, entirely obvious.

It is not possible to derive the results in this paper from those of DiCiccio, Hall and Romano (1991), since the latter work relies on the so-called "smooth function model" for a statistical estimator, which does not encompass sample quantiles.

2. Smoothed empirical likelihood for quantiles. Let $X_1, \dots, X_n$ denote a random sample from a distribution with distribution function F. Assume that the qth quantile, $\theta_q = F^{-1}(q)$, is uniquely defined. We wish to construct a confidence interval for θ_q.

To this end, let G_h denote a smoothed version of the degenerate distribution function G_0 defined by $G_0(x) = 1$ for $x \geq 0$, 0 otherwise. Specifically, let K denote an rth-order kernel, of the type commonly used in nonparametric density estimation or regression [e.g., Silverman (1986), page 66ff and Härdle (1990), page 141ff]. That is, for some integer $r \geq 2$ and constant $\kappa \neq 0$, K is a function satisfying

$$\int u^j K(u)\,du = \begin{cases} 1, & \text{if } j = 0, \\ 0, & \text{if } 1 \leq j \leq r-1, \\ \kappa, & \text{if } j = r. \end{cases} \tag{2.1}$$

The case $r = 2$ is the most common, and there we take K to be a symmetric probability density. Larger values of r produce curve estimators with smaller variance. Define $G(x) = \int_{y<x} K(y)\,dy$. In this notation we put $G_h(x) = G(x/h)$. When $r = 2$ and K is a density, G and G_h are proper distribution functions.

Write $p_1, \dots, p_n$ for nonnegative numbers adding to unity, and define $p = (p_1, \dots, p_n)$,

$$\hat{F}_{p,h}(\theta) = \sum_{i=1}^{n} p_i G_h(\theta - X_i), \qquad L_h(\theta) = \sup_{p:\, \hat{F}_{p,h}(\theta) = q} \prod_{i=1}^{n} (np_i).$$

Empirical log likelihood is defined by $l_h(\theta) = -2 \log L_h(\theta)$.

A smoothed empirical likelihood confidence interval for θ_q is given by

$$I_{hc} = \{\theta : l_h(\theta) \leq c\}.$$

Here, $c > 0$ is a constant whose value determines the coverage probability, α_{hc}, of I_{hc}:

$$\alpha_{hc} = P(\theta_q \in I_{hc}) = P\{l_h(\theta_q) \leq c\}. \tag{2.2}$$

As we shall see shortly, $l_h(\theta_q)$ typically has an asymptotic χ_1^2 distribution, and this property may be used to select c.

If G_h is a distribution function, then I_{hc} will be an interval, as desired. When higher-order kernels are used to construct G_h, it is possible that for small sample sizes and unusual values of h, I_{hc} might be a union of disjoint intervals.

We may give the definition of l more explicitly, as follows. Put $w_i(\theta) = G_h(\theta - X_i) - q$. The method of Lagrange multipliers may be used to maximise Πp_i subject to $\Sigma p_i w_i(\theta) = 0$ and $\Sigma p_i = 1$. Arguing thus, we may prove that the turning point occurs with $p_i = n^{-1}\{1 + \lambda(\theta) w_i(\theta)\}^{-1}$, whence

$$l(\theta) = 2 \sum_{i=1}^{n} \log\{1 + \lambda(\theta) w_i(\theta)\},$$

where $\lambda(\theta)$ is determined by

$$\sum_{i=1}^{n} w_i(\theta)\{1 + \lambda(\theta) w_i(\theta)\}^{-1} = 0.$$

The standard, unsmoothed version of empirical likelihood has $h = 0$ in the preceding formulae. It produces the interval

$$I_{oc} = \left[X_{(r_1)}, X_{(r_2)}\right],$$

where $X_{(1)} \le \cdots \le X_{(n)}$ denote the order statistics of the sample $X_1, \dots, X_n$, and r_1 and r_2 are respectively the smallest and largest integers such that

$$(q/i)^i\{(1-q)/(n-i)\}^{n-i} \ge c.$$

The exact coverage probability of the interval I_{oc} is given by

$$\alpha_{oc} = P(\theta_q \in I_{oc}) = P(r_1 \le M \le r_2 - 1),$$

where M is a binomial $\mathrm{Bi}(n, q)$ random variable. Compare (2.2). This probability cannot be rendered closer than order $n^{-1/2}$ to any predetermined value, such as 0.95, no matter how the integers r_1 and r_2 are selected. That is because any alterations to r_1 and r_2 produce changes of size $n^{-1/2}$ in $P(r_1 \le M \le r_2 - 1)$, except in the extreme tails. To appreciate why, observe that

$$\sup_{0 \le r \le n} \left| P(M = r) - \{2\pi n q(1-q)\}^{-1/2} \exp\left[-\tfrac{1}{2}(r - nq)^2\{nq(1-q)\}^{-1}\right]\right|$$
$$= o(n^{-1/2}).$$

See Petrov (1975), page 187.

3. Wilks' theorem and coverage accuracy. We adopt throughout the notation introduced in Section 2. Recall that the smoothed empirical likelihood confidence interval I_{hc} is defined in terms of a positive constant c, which determines coverage probability. If the distribution of $l_h(\theta_q)$ were known, then c could be chosen so that I_{hc} had a predetermined level of coverage, such as

0.95. Our first result establishes necessary and sufficient conditions on the choice of bandwidth, h, for $l_h(\theta_q)$ to have an asymptotic χ_1^2 distribution.

Let $f = F'$ denote the first derivative of F, where defined.

THEOREM 3.1. *Assume that*

(3.1) *K satisfies (2.1), and is bounded and compactly supported; that f and $f^{(r-1)}$ exist in a neighbourhood of θ_q and are continuous at θ_q; that $f(\theta_q) > 0$; and that for some $t > 0$, $nh^t \to 0$ as $n \to \infty$.*

Then $l_h(\theta_q)$ has an asymptotic χ_1^2 distribution if $nh^{2r} \to 0$, and this condition is also necessary if $f^{(r-1)}(\theta_q) \neq 0$.

Let us explain the implications of condition (3.1). The first part of (3.1) asks that K be a kernel of order r. The second part asks that the distribution function F be sufficiently smooth in a neighbourhood of θ_q; the condition that r continuous derivatives of the target function (here, F) exist is the usual smoothness assumption imposed when working with an rth-order kernel. Requiring that $f(\theta_q) > 0$ ensures that the asymptotic variance of the sample quantile is of order n^{-1}. Without that assumption the order of magnitude of variance is strictly larger than n^{-1}, and the asymptotic theory is quite different. Finally, asking that $nh^t \to 0$ as $n \to \infty$ ensures that the bandwidth does not converge to zero too slowly.

If K is a second-order kernel (i.e., $r = 2$) and $f'(\theta_q) \neq 0$, then $l_h(\theta_q)$ is asymptotically χ_1^2 if and only if $h = o(n^{-1/4})$. Such a bandwidth is of smaller order of magnitude than that which is usually appropriate for minimising error of a curve estimator; the latter h is of size $n^{-1/5}$, as shown, for example, by Silverman (1986), page 40ff, and Härdle (1990), page 155. When $f^{(r-1)}(\theta_q) = 0$, it is possible for $l_h(\theta_q)$ to have an asymptotic χ_1^2 distribution yet nh^{2r} to be bounded away from zero.

If (3.1) holds and $nh^{2r} \to 0$, then by the theorem,

$$P\{l_h(\theta_q) \le c\} \to P(\chi_1^2 \le c).$$

Therefore, if c is chosen from chi-squared tables to satisfy

$$P(\chi_1^2 \le c) = \alpha, \tag{3.2}$$

the asymptotic coverage of the interval I_{hc} will equal α:

$$\alpha_{hc} = P(\theta_q \in I_{hc}) \to \alpha$$

as $n \to \infty$.

Of course, we are interested in not just the asymptotic coverage of I_{hc} but also in coverage accuracy. Our next result gives a necessary and sufficient condition for the coverage error of I_{hc} to be of order n^{-1}. It also shows that this size of error is best possible, in the sense that h cannot be chosen to give an error of smaller size than n^{-1}.

THEOREM 3.2. *Assume condition* (3.1) *and in addition that the bandwidth* h *satisfies*

$$nh^{2r} \to 0 \quad \text{and} \quad nh/\log n \to \infty. \tag{3.3}$$

Define $c = c(\alpha)$ *by* (3.2). *Then a sufficient condition for*

$$P(\theta_q \in I_{hc}) = \alpha + O(n^{-1}), \tag{3.4}$$

as $n \to \infty$, *is that* nh^r *is bounded. This condition is also necessary if* $f^{(r-1)}(\theta_q) \neq 0$. *No matter what the value of* $c > 0$, *the right-hand side of* (3.4) *cannot be rendered equal to* $\alpha + o(n^{-1})$ *by appropriately choosing* h.

The first part of assumption (3.3) is just the necessary and sufficient condition for $l_h(\theta_q)$ to have an asymptotic χ_1^2 distribution, noted in Theorem 3.1. The second part asks that h not be too small, and is imposed to ensure a minimal level of smoothness of the statistic $l_h(\theta_q)$. We use it to derive a version of Cramér's smoothness condition from the theory of Edgeworth expansions. It is not difficult to see that if

$$nh \log n \to 0, \tag{3.5}$$

then (3.4) must fail because insufficient smoothness is imposed. [The closest data pairs in the vicinity of θ_q are separated by no more than $(n \log n)^{-1}$, and if h is of smaller order than this, then there is effectively no smoothing of the empirical likelihood method.] Asking that the second part of (3.3) hold is only a little more stringent than insisting that (3.5) fail.

When $f^{(r-1)}(\theta_q) = 0$, it is possible for (3.4) to hold yet nh^r to diverge to $+\infty$.

Our proof of Theorem 3.2 provides an expansion of Edgeworth type for the left-hand side of (3.4); see (6.11) in Section 6. That result may be used to derive the "optimal" value of C when $nh^r \to C$. However, determination of C by this rule is not really a practical proposition, particularly in view of the availability of Bartlett correction. (See Section 4.) Therefore, we do not give further details here.

4. Bartlett correction. It may be proved that if $nh^r \to 0$ then

$$E\{l_h(\theta_q)\} = 1 + n^{-1}\beta + o(n^{-1}),$$

where $\beta = \frac{1}{6}(3\mu_2^{-2}\mu_4 - 2\mu_2^{-3}\mu_3^2)$ and $\mu_j = E[G\{(\theta_q - X_i)/h\} - q]^j$. Thus, the expected values of $l_h(\theta_q)$ and its approximate chi-squared distribution differ by $n^{-1}\beta$. It stands to reason that if this error of scale were corrected, then the chi-squared approximation might be more effective. The striking thing is that the accuracy of the approximation can be improved by an order of magnitude by making such an adjustment. Indeed, the main result in this section shows that if n^3h^{2r} is bounded, then the distribution of $l_h(\theta_q)/(1 + n^{-1}\beta)$ differs from the χ_1^2 distribution only in terms of order n^{-2}, not just terms of order

n^{-1}. In practice, β is unknown and must be estimated. To this end, define

$$\hat{\mu}_j = n^{-1} \sum_{i=1}^{n} \left[G\left\{ \left(\hat{\theta}_q - X_i \right) / h \right\} - q \right]^j$$

and $\hat{\beta} = \frac{1}{6}(3\hat{\mu}_2^{-2}\hat{\mu}_4 - 2\hat{\mu}_2^{-3}\hat{\mu}_3^2)$, where $\hat{\theta}_q$ denotes the usual estimate of θ_q. Let c be the α-level point of the χ_1^2 distribution, defined at (3.2), and put $d(c, \gamma) = c(1 + n^{-1}\gamma)$ where γ is either β or $\hat{\beta}$. Our claim is that, provided h is chosen appropriately, the Bartlett-corrected confidence regions $I_{h, d(c, \beta)}$ and $I_{h, d(c, \hat{\beta})}$ have smaller coverage error than I_{hc}. This is made clear by our next theorem.

THEOREM 4.1. *Assume conditions* (3.1) *and* (3.3), *and define* $c = c(\alpha)$ *by* (3.2). *Then a sufficient condition for*

$$P\left(\theta_q \in I_{h, d(c, \gamma)}\right) = \alpha + O(n^{-2}), \tag{4.1}$$

for either $\gamma = \beta$ *or* $\gamma = \hat{\beta}$, *is that* $n^3 h^{2r}$ *be bounded. If* $f^{(r-1)}(\theta_q) \neq 0$, *then the boundedness of* $n^3 h^{2r}$ *is also necessary for* (4.1). *Since* $\mu_2 = q(1-q) + O(h)$, $\mu_3 = q(1-q)(1-2q) + O(h)$ *and* $\mu_4 = q(1-q)(1 - 3q + 3q^2) + O(h)$, *then* $\beta = \beta_0 + O(h)$, *where* $\beta_0 = \frac{1}{6} q^{-1}(1-q)^{-1}(1 - q + q^2)$. *It is permissible to take* $\gamma = \beta_0$ *when constructing the confidence interval* $I_{h, d(c, \gamma)}$, *but this will not give the same coverage accuracy as* $\gamma = \beta$ *or* $\gamma = \hat{\beta}$, *owing to the relatively large distance between* β *and* β_0. *In particular, result* (4.1) *should be changed to*

$$P\left(\theta_q \in I_{h, d(c, \beta_0)}\right) = \alpha + O(n^{-1}h). \tag{4.2}$$

From a practical viewpoint, this very simple approach to Bartlett correction is particularly attractive. Although it does not enjoy quite the same asymptotic performance as the "full" correction discussed earlier, the simulation study in the next section shows that it performs commendably well in practice. This is presumably because the "full" correction is relatively sensitive to bandwidth choice, and such methods can be rather variable in small samples.

5. Simulation study. Here we summarise the conclusions of a numerical study designed to investigate the performance of simple rules for selecting h. Throughout we smooth using a second-order kernel (i.e., $r = 2$),

$$K(u) = \begin{cases} \dfrac{3}{4\sqrt{5}} \left(1 - \dfrac{1}{5} u^2 \right), & \text{if } |u| \leq \sqrt{5}, \\ 0, & \text{otherwise.} \end{cases}$$

This is the so-called Bartlett or Epanečnikov kernel; see, for example, Silverman (1986), page 42, and Härdle (1990), page 25. We concentrate on confidence intervals for quartiles and the median (i.e., $q = \frac{1}{4}, \frac{1}{2}, \frac{3}{4}$), and take the parent distribution F to be chi-squared with a variety of different degrees of freedom. We choose nominal coverages of $\alpha = 0.90$ and $\alpha = 0.95$, employ a variety of different formulae for h and check on the performance of two different versions of Bartlett's correction. In the latter we take $\gamma = \beta_0$ and $\gamma = \hat{\beta}$ in the formula for the confidence interval $I_{h,\,d(c,\gamma)}$; we do not treat the case $\gamma = \beta$ since that version is not of practical interest. Formulae for I_{hc}, $I_{h,\,d(c,\gamma)}$, β_0 and $\hat{\beta}$ are given in Sections 2 and 4.

Recall from Theorems 3.2 and 4.1 that when $r = 2$, the bounds $h = O(n^{-1/2})$ and $h = O(n^{-3/4})$ define the largest h for which the uncorrected interval has coverage error $O(n^{-1})$ and the Bartlett-corrected interval with $\gamma = \hat{\beta}$ has coverage error $O(n^{-2})$, respectively.

Table 1 summarises results for the χ_3^2 distribution and the sample size $n = 20$. Figure 1 illustrates how coverage accuracy varies over different degrees of freedom and different sample sizes. Each point in the table and figure is based on 10,000 simulations. The chi-squared variables were produced by adding squares of independent normal variables given by the routine in Numerical Recipes [Press, Flannery, Teukolsky and Vetterling (1989)].

Table 1 and Figure 1 summarise a larger simulation study that is given in detail in Chen (1993). The following broad conclusions may be drawn from those results. First, smoothed empirical likelihood intervals have greater coverage accuracy than their unsmoothed counterparts, and further improvement is offered by Bartlett correction. Second, the "theoretical" Bartlett correction (based on the value β_0) performs similarly to the "empirical" Bartlett correction (using $\hat{\beta}$). Since β_0 is simpler than $\hat{\beta}$ to implement, it is to

TABLE 1

Estimated true coverages, from 10,000 simulations, of α-level smoothed empirical likelihood confidence intervals for the qth quantile of the χ_3^2 distribution

		q	0.25		0.50		0.75	
		α	0.90	0.95	0.90	0.95	0.90	0.95
h								
	0		0.874	0.935	0.922	0.973	0.905	0.972
	uncorr.		0.884	0.957	0.887	0.956	0.885	0.960
n^{-1}	β_0		0.887	0.959	0.888	0.957	0.888	0.961
	$\hat{\beta}$		0.891	0.960	0.889	0.971	0.887	0.961
	uncorr.		0.883	0.947	0.890	0.949	0.883	0.941
$n^{-3/4}$	β_0		0.889	0.950	0.894	0.951	0.890	0.942
	$\hat{\beta}$		0.890	0.951	0.894	0.951	0.890	0.943
	uncorr.		0.895	0.944	0.899	0.948	0.889	0.947
$n^{-1/2}$	β_0		0.902	0.948	0.903	0.949	0.896	0.949
	$\hat{\beta}$		0.903	0.950	0.903	0.950	0.896	0.949
	uncorr.		0.890	0.941	0.893	0.944	0.897	0.947
$n^{-1/4}$	β_0		0.896	0.945	0.897	0.948	0.900	0.951
	$\hat{\beta}$		0.897	0.946	0.899	0.949	0.903	0.951

Rows headed "uncorr.," "β_0" and "$\hat{\beta}$" give the uncorrected interval and the Bartlett-corrected intervals computed with $\gamma = \beta_0$ and $\gamma = \hat{\beta}$, respectively.

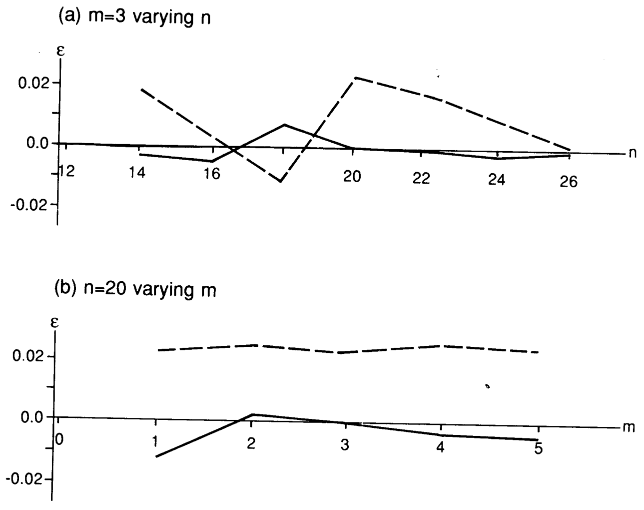

FIG. 1. *The graphs depict coverage error, given by ε = true coverage − 0.95, of smoothed (—) and unsmoothed (- - -) 95% confidence intervals for the median (i.e., $q = \frac{1}{2}$). In the case of the smoothed confidence interval, the bandwidth is $h = n^{-3/4}$ and the Bartlett correction is employed with $\gamma = \beta_0$. Throughout, the underlying distribution is χ^2_m. Panel* (a) *illustrates the case where $m = 3$ is fixed and n varies*; *panel* (b) *illustrates $n = 20$ and varying m.*

be recommended. Third, choices of h in the range $n^{-1/2}, n^{-3/4}$ generally provide quite good coverage accuracy. However, when the underlying distribution is heavily skewed (e.g., χ_1^2), less smoothing than this is desirable.

6. Outlines of proofs.

PROOF OF THEOREM 3.1. Define $w_i = w_i(\theta_q) = G_h(\theta_q - X_i) - q$, and let $\lambda = \lambda(\theta_q)$ denote a solution of the equation

$$\sum_{i=1}^{n} w_i(1 + \lambda w_i)^{-1} = 0. \tag{6.1}$$

Our first step is to prove that

$$\lambda = O_p(n^{-1/2} + h^r). \tag{6.2}$$

The probabilities p_i appearing in the definitions of $L(\theta_q)$ and $l(\theta_q)$ (see Section 2) are given by $p_i = n^{-1}(1 + \lambda w_i)^{-1}$. Therefore, $1 + \lambda w_i \ge 0$, and so

$$|1 + \lambda w_i|^{-1} \ge (1 + |\lambda| \max|w_i|)^{-1}. \tag{6.3}$$

Define $\overline{w}_j = n^{-1}\Sigma w_i^j$ for integers $j \ge 1$. Let C_1 denote an upper bound to $|G|$. Then $C_2 = C_1 + q$ is an upper bound to $|w_i|$, for all i and all h. By definition of λ,

$$0 = n^{-1}\left|\sum_{i=1}^{n}\left\{\lambda w_i^2(1+\lambda w_i)^{-1} - w_i\right\}\right|$$
$$\geq |\lambda|(1 + C_2|\lambda|)^{-1}\overline{w}_2 - |\overline{w}_1|,$$

using (6.3). Therefore, $|\lambda|\overline{w}_2 \leq (1 + C_2|\lambda|)|\overline{w}_1|$, or, equivalently,

$$|\lambda|\left(\overline{w}_2 - C_2|\overline{w}_1|\right) \leq |\overline{w}_1|. \tag{6.4}$$

Observe that $\overline{w}_j - E(\overline{w}_j) = O_p(n^{-1/2})$, $E(\overline{w}_1) = o(h^r)$ and $E(\overline{w}_2) = q(1-q) + o(1)$, the second identity following by Taylor expansion. Therefore, by (6.4),

$$|\lambda|\left\{q(1-q) + o_p(1)\right\} \leq O_p(n^{-1/2} + h^r),$$

which gives the desired result (6.2).

Next we develop Taylor expansions of λ and $l_h(\theta_q)$. These are a little longer than are necessary for the present proof. However, the additional details given here will be needed in the proof of Theorem 3.2.

By (6.1) and (6.2), for each integer $j \geq 1$,

$$0 = n^{-1}\sum_{i=1}^{n} w_i\left\{1 - \lambda w_i + (\lambda w_i)^2 - (\lambda w_i)^3 + \cdots\right\}$$
$$= \sum_{k=1}^{j}(-\lambda)^{k-1}\overline{w}_k + O_p\left\{(n^{-1/2} + h^r)^j\right\}.$$

Solving this polynomial equation for λ gives, for each $j \geq 1$,

$$\begin{aligned}\lambda = {} & \overline{w}_2^{-1}\overline{w}_1 + \overline{w}_2^{-3}\overline{w}_3\overline{w}_1^2 + \left(2\overline{w}_2^{-5}\overline{w}_3^2 - \overline{w}_2^4\overline{w}_4\right)\overline{w}_1^3 \\ & + \sum_{k=4}^{j} R_{1k}\overline{w}_1^k + O_p\left\{(n^{-1/2} + h^r)^{j+1}\right\},\end{aligned} \tag{6.5}$$

where R_{lk} denotes $\overline{w}_2^{-(2k-1)}$ multiplied by a polynomial in $\overline{w}_2, \ldots, \overline{w}_{k+1}$, with constant coefficients. Similarly,

$$\begin{aligned} l_h(\theta_q) &= 2\sum_{i=1}^{n}\log(1 + \lambda w_i) \\ &= 2n\sum_{k=1}^{j+1}(-1)^{k+1}k^{-1}\lambda^k\overline{w}_k + O_p\left\{n(n^{-1/2} + h^r)^{j+2}\right\} \\ &= n\Big\{\overline{w}_2^{-1}\overline{w}_1^2 + \tfrac{2}{3}\overline{w}_2^{-3}\overline{w}_3\overline{w}_1^3 + \left(\overline{w}_2^{-5}\overline{w}_3^2 - \tfrac{1}{2}\overline{w}_2^{-4}\overline{w}_4\right)\overline{w}_1^4 \\ &\qquad + \left(8\overline{w}_2^{-6}\overline{w}_3\overline{w}_4 - 8\overline{w}_2^{-7}\overline{w}_3^3 - \tfrac{8}{5}\overline{w}_2^{-5}\overline{w}_5\right)\overline{w}_1^5\Big\} \\ &\quad + n\sum_{k=5}^{j} R_{2k}\overline{w}_1^{k+1} + O_p\left\{n(n^{-1/2} + h^r)^{j+2}\right\}.\end{aligned} \tag{6.6}$$

The third identity follows on substituting (6.5) into the second identity, and noting that $\bar{w}_1 = \bar{w}_1 - E(\bar{w}_1) + E(\bar{w}_1) = O_p(n^{-1/2} + h^r)$.

Put $\mu_j = E(\bar{w}_j)$ and $Z = n^{1/2}(\bar{w}_1 - \mu_1)\mu_2^{-1/2}$. It is readily proved that under condition (3.2), Z is asymptotically normal $N(0,1)$. Furthermore, $\bar{w}_2 = \mu_2 + o_p(1) = q(1-q) + o_p(1)$, and $\bar{w}_j = O_p(1)$ for $j \geq 3$. Hence by (6.6) we have for any $k > 1$,

$$\begin{aligned} l_h(\theta_q) &= n\mu_2^{-1}\bar{w}_1^2\{1 + o_p(1)\} + O_p\left\{n\left(n^{-1/2} + h^r\right)^k\right\} \\ &= \left(n^{1/2}\mu_2^{-1/2}\mu_1 + Z\right)^2\{1 + o_p(1)\} + O_p\left\{n\left(n^{-1/2} + h^r\right)^k\right\}. \end{aligned}$$

Therefore, since $nh^t \to 0$ for some $t > 0$, $l_h(\theta_q)$ has an asymptotic central chi-squared distribution with one degree of freedom if and only if $n^{1/2}\mu_2^{-1/2}\mu_1 \to 0$, that is, $n^{1/2}\mu_1 \to 0$. Now,

$$\begin{aligned} \mu_1 &= \int_{-\infty}^{\infty}\{F(\theta_q - hu) - F(\theta_q)\}K(u)\,du \\ &= (-h)^r(r!)^{-1}\kappa f^{(r-1)}(\theta_q) + o(h^r). \end{aligned}$$

Therefore, $n^{1/2}\mu_1 \to 0$ if $nh^{2r} \to 0$; and if $f^{(r-1)}(\theta_q) \neq 0$, then $n^{1/2}\mu_1 \to 0$ implies $nh^{2r} \to 0$. □

PROOF OF THEOREM 3.2. Our starting point is formula (6.6). Taking the signed square root of the right-hand side, we see that we may write

$$l_h(\theta_q) = \left(n^{1/2}S_j'\right)^2,$$

where

$$\begin{aligned} S_j' &= \bar{w}_2^{-1/2}\Bigg\{\bar{w}_1 + \tfrac{1}{3}\bar{w}_2^{-2}\bar{w}_3\bar{w}_1^2 + \left(\tfrac{4}{9}\bar{w}_2^{-4}\bar{w}_3^2 - \bar{w}_3^{-3}\bar{w}_4\right)\bar{w}_1^3 \\ &\qquad + \left(\tfrac{112}{27}\bar{w}_2^{-6}\bar{w}_3^3 + \tfrac{97}{12}\bar{w}_2^{-5}\bar{w}_3\bar{w}_4 - \tfrac{4}{5}\bar{w}_2^{-4}\bar{w}_5\right)\bar{w}_1^4 \\ &\qquad + \sum_{k=5}^{j} T_k\bar{w}_1^k\Bigg\} + U_{1j} \\ &= S_j + U_{1j} \end{aligned}$$

say, where T_k denotes $\bar{w}_2^{-2(k-1)}$ multiplied by a polynomial in $\bar{w}_2, \ldots, \bar{w}_k$ with constant coefficients, and $U_{1j} = O_p\{(n^{-1/2} + h^r)^{j+1}\}$. Noting that $nh^t \to 0$ for some $t > 0$, a little additional analysis shows that by choosing j sufficiently large we may ensure that for $l = 1$,

$$P\left(|U_{lj}| > n^{-5/2}\right) = O(n^{-2}). \tag{6.7}$$

Hence, for $x > 0$,

$$\begin{aligned} P\{l_h(\theta_q) \leq x^2\} &= P\left(-x \leq n^{1/2}S_j' \leq x\right) \\ &\left\{\begin{matrix}\leq \\ \geq\end{matrix}\right\} P\left(-x \mp n^{-2} \leq n^{1/2}S_j \leq x \pm n^{-2}\right) + O(n^{-2}), \end{aligned} \tag{6.8}$$

where the inequalities and plus/minus signs are to be taken respectively in the indicated orders.

The next step is to develop an Edgeworth expansion for the distribution of $n^{1/2}S_j$. Observe that S_j is a function of $\bar{w}_1, \dots, \bar{w}_j$. Denote that function by s_j. Put $\mu_k = E(\bar{w}_k)$, $\mu = (\mu_1, \dots, \mu_j)$, $u = (u_1, \dots, u_j)$, $V_k = \bar{w}_k - \mu_k$, $V = (V_1, \dots, V_j)$,

$$d_{k_1, \dots, k_m} = \left(\prod_{l=1}^{m} \partial/\partial u_{k_l} \right) s_j(u_1, \dots, u_j) \bigg|_{u=\mu},$$

$$p(u) = s_j(\mu) + \sum_{m=1}^{6} (m!)^{-1} \sum_{k_1, \dots, k_m \in \{1, \dots, j\}} d_{k_1, \dots, k_m} u_{k_1} \dots u_{k_m}.$$

Then p is a polynomial, and $p(V)$ represents a Taylor approximation to S_j with an error of order n^{-3}:

$$S_j = p(V) + U_{2j},$$

where $U_{2j} = O_p(n^{-3})$. A little additional analysis shows that (6.7) holds for $l = 2$, and so

$$P\left(n^{1/2}S_j \le x\right)\begin{Bmatrix} \le \\ \ge \end{Bmatrix}P\{n^{1/2}p(V) \le x \pm n^{-2}\} + O(n^{-2}). \tag{6.9}$$

By painstakingly developing Taylor expansion formulae for the quantities $d_{k_1, \dots, k_m}$, and for the cumulants of V, we may prove that the cumulants $k_1, k_2, \dots$ of $n^{1/2}p(V)$ satisfy the following formulae:

$$k_1 = n^{1/2}s_j(\mu) - \tfrac{1}{6}\mu_2^{-3/2}\mu_3 n^{-1/2} + O(n^{-1/2}h^r + n^{-3/2}),$$

$$k_2 = \sigma^2 + \left(\tfrac{1}{2}\mu_2^{-2}\mu_4 - \tfrac{13}{36}\mu_2^{-3}\mu_3^2\right)n^{-1} + O(n^{-1}h^r + n^{-2}),$$

$$k_3 = O(n^{-1/2}h^r), \qquad k_4 = O(n^{-1}h^r), \qquad k_l = O(n^{-(l-2)/2}) \quad \text{for } l \ge 5,$$

where

$$\sigma^2 = \sum_{k_1} \sum_{k_2} d_{k_1} d_{k_2} E\left\{\left(w_1^{k_1} - \mu_{k_1}\right)\left(w_1^{k_2} - \mu_{k_2}\right)\right\}$$

$$= 1 + \tfrac{1}{3}\mu_2^{-2}\mu_1\mu_3 + \left(\tfrac{7}{9}\mu_2^{-4}\mu_3^2 - \tfrac{1}{4}\mu_2^{-1} - \tfrac{7}{12}\mu_2^{-3}\mu_4\right)\mu_1^2 + O(h^{3r}).$$

(Details are outlined in the Appendix.) This allows us to develop a formal Edgeworth expansion for the distribution of $p(V)$: assuming $nh^{2r} \to 0$,

$$\begin{aligned} &P\{n^{1/2}p(V) \le x\} \\ &\quad = \Phi(x) - \tfrac{1}{12}n^{-1}\left\{6\mu_2^{-1}(n\mu_1)^2 + 3\mu_2^{-2}\mu_4 - 2\mu_2^{-3}\mu_3^2\right\}x\phi(x) \\ &\qquad + (\text{even polynomial in } x)\phi(x) + o(nh^{2r}) + O(n^{-2}), \end{aligned} \tag{6.10}$$

where Φ and ϕ denote the standard normal distribution and density functions, respectively. Now,

$$\mu_1 = (-h)^r (r!)^{-1} \kappa f^{(r-1)}(\theta_q) + o(h^r),$$
$$\mu_2 = q(1-q) + o(1),$$
$$\mu_3 = q - 3q^2 + 2q^3 + o(1),$$
$$\mu_4 = q - 4q^2 + 6q^3 - 3q^4 + o(1).$$

Hence, for $x > 0$,

$$(6.11)\quad P\{-x \le n^{1/2}p(V) \le x\} = 2\Phi(x) - 1 - n^{-1}\Big\{(r!)^{-2}\kappa^2 f^{(r-1)}(\theta_q)^2 (nh^r)^2 q^{-1}(1-q)^{-1} + \tfrac{1}{6}q^{-1}(1-q)^{-1}(1-q+q^2)\Big\}x\phi(x) + o(n^{-1} + nh^{2r}).$$

Accepting that the formal expansion (6.10), and hence (6.11), may be justified, it follows that (a) $\sup_n nh^r < \infty$ implies

$$(6.12)\quad P\{-x \le n^{1/2}p(V) \le x\} - \{2\Phi(x) - 1\} = O(n^{-1}) \quad \text{for all } x;$$

(b) if $f^{(r-1)}(\theta_q) \ne 0$, then (6.12) implies $\sup_n nh^r < \infty$; (c) if $nh^r \to C$, $0 \le C < \infty$, then

$$(6.13)\quad P\{-x \le n^{1/2}p(V) \le x\} - \{2\Phi(x) - 1\} = o(n^{-1}) \quad \text{for } x > 0$$

if and only if

$$(r!)^{-2}\kappa^2 f^{(r-1)}(\theta_q)^2 C^2 + \tfrac{1}{6}(1-q+q^2) = 0.$$

The left-hand side is strictly positive for all $0 < q < 1$. In view of (6.8) and (6.9), conclusions (a)–(c) continue to hold if, on the left-hand sides of (6.8) and (6.9), $P\{-x \le n^{1/2}p(V) \le x\}$ is replaced by $P\{l_h(\theta_q) \le x^2\}$. This proves Theorem 3.2.

It remains to check that the formal expansion (6.10) is valid. This may be done by developing an Edgeworth expansion of the multivariate distribution of $n^{1/2}V$, with the form

$$(6.14)\quad P(n^{1/2}V \in A) = \Phi_{0,\Sigma}(A) + \sum_{k=1}^{m} n^{-k/2}\int_A p_k(x)\phi_{0,\Sigma}(x)\,dx + O(n^{-(m+1)/2})$$

uniformly in j-variate sets A from any class $\mathscr{A}$ satisfying

$$\sup_{A\in\mathscr{A}} \Phi_{0,\Sigma}\{(\partial A)^{\varepsilon}\} = O(\varepsilon)$$

as $\varepsilon \downarrow 0$. In these formulae, $\Phi_{0,\Sigma}$ and $\phi_{0,\Sigma}$ denote the distribution and density functions of the $N(0,\Sigma)$ distributions; p_k is a polynomial of degree $k+2$ with uniformly bounded coefficients; $m \ge 1$ is any integer; and $(\partial A)^{\varepsilon}$ is the set of all points distant at most ε from the boundary of A. Noting that V is a mean of a sum of independent and identically distributed random variables, this result may be proved using techniques from Bhattacharya and Rao (1976), page 192ff, provided we establish an analogue of Cramér's condition. This result states that for each $\varepsilon > 0$ there exists a constant $C(\varepsilon) > 0$ such that for all sufficiently small h,

$$\sup_{t_1, \dots, t_j:\, \Sigma|t_k| > \varepsilon} \left| h \int_{-\infty}^{\infty} \exp\left\{ i \sum_{k=1}^{j} t_k G(u)^k \right\} f(\theta_q - hu)\, du \right| \le 1 - C(\varepsilon) h,$$

where $i = \sqrt{-1}$ and $G(x) = \int_{y<x} K(y)\, dy$. The methods used to establish this result, indeed to prove (6.14), are those given by Hall (1991).

The method of deriving (6.10) and (6.14) is identical to that described by Bhattacharya and Ghosh (1978), page 443ff. □

PROOF OF THEOREM 4.1. Put $\beta = \frac{1}{6}(3\mu_2^{-2}\mu_4 - 2\mu_2^{-3}\mu_3^2)$. Then by (6.10),

$$\begin{aligned} P\{np(V)^2 &\le x^2(1 + \beta n^{-1})\} \\ &= P\{-x(1 + \tfrac{1}{2}\beta n^{-1}) \le n^{1/2} p(V) \\ &\qquad \le x(1 + \tfrac{1}{2}\beta n^{-1})\} + O(n^{-2}) \\ &= P(\chi_1^2 \le x) - n\mu_2^{-1}\mu_1^2 x\phi(x) + o(nh^{2r}) + O(n^{-2}) \\ &= P(\chi_1^2 \le x) - nh^{2r}(r!)^{-2}\kappa^2 f^{(r-1)}(\theta_q)^2 q^{-1}(1-q)^{-1} \\ &\qquad + o(nh^{2r}) + O(n^{-2}). \end{aligned}$$

Therefore, (a) $\sup_n n^3 h^{2r} < \infty$ implies

$$P\{np(V)^2 \le x^2(1 + \beta n^{-1})\} - P(\chi_1^2 \le x^2) = O(n^{-2}) \quad \text{for all } x, \tag{6.15}$$

and (b) if $f^{(r-1)}(\theta_q) \neq 0$, then (6.15) implies $\sup_n n^3 h^{2r} < \infty$. In view of (6.8) and (6.9), these conclusions continue to hold if on the left-hand side of (6.15), $P\{np(V)^2 \le x^2(1 + \beta n^{-1})\}$ is replaced by $P\{l_h(\theta_q) \le x^2(1 + \beta n^{-1})\}$.

The case where β is replaced by $\hat{\beta}$ may be handled similarly, although the analysis is far more tedious. □

APPENDIX

Calculation of cumulants k_l. In this Appendix we calculate the cumulants $k_1, k_2, \dots$ of $n^{1/2}P(V)$ which were used in the proof of Theorem 3.2. Let $k^{i_1, i_2, \dots, i_p}$ be the pth-order multivariate cumulants of $V = (V_1, \dots, V_j)$. According to results given by James and Mayne (1962), the k_j's may be expressed as follows:

$$\text{(A.1)}\quad k_1 = n^{1/2}\left\{S_j(u) + \tfrac{1}{2}d_{ij}k^{ij} + \tfrac{1}{6}d_{ijk}k^{ijk} + \tfrac{1}{8}d_{ijkl}k^{ij}k^{kl}\right\} + O(n^{-5/2}),$$

$$\text{(A.2)}\quad k_2 = n\left\{d_i d_j k^{ij} + d_{ij}d_k k^{ijk} + \left(d_{ijk}d_l + \tfrac{1}{2}d_{ik}d_{jl}\right)k^{ij}k^{kl}\right\} + O(n^{-2}),$$

$$\text{(A.3)}\quad \begin{aligned} k_3 = n^{3/2}\Big\{& d_i d_j d_k k^{ijk} + 3d_{ik}d_j d_l k^{ij}k^{kl} + \tfrac{3}{2}d_{ij}d_k d_l k^{ijkl} \\ &+ \left(3d_{ijl}d_k d_m + \tfrac{3}{2}d_{ilm}d_j d_k \right. \\ &\left. + 3d_{ij}d_{kl}d_m + 3d_{il}d_{jm}d_k\right)k^{ijk}k^{lm} + \tfrac{3}{2}d_{ijkm}d_l d_n k^{ij}k^{kl}k^{mn} \\ &+ (3d_{ijk}d_{lm}d_n + 3d_{ikm}d_{jl}d_n + d_{ik}d_{jm}d_{in})k^{ij}k^{kl}k^{mn}\Big\} \\ & + O(n^{-5/2}), \end{aligned}$$

$$\text{(A.4)}\quad \begin{aligned} k_4 = n^2\Big\{& d_i d_j d_k d_l k^{ijkl} + 12 d_{il}d_j d_k d_m k^{ijk}k^{lm} \\ &+ (4d_{ikm}d_j d_l d_n + 12 d_{ik}d_{jm}d_l d_n)k^{ij}k^{kl}k^{mn}\Big\} + O(n^{-2}), \\ k_l = O(n&^{-(l-2)/2}), \qquad l \ge 5, \end{aligned}$$

where

$$d_{j_1,\dots,j_m} = \left(\prod_{l=1}^{m} \partial/\partial u_{j_l}\right) S_j(u_1,\dots,u_j)\Big|_{u=\mu}.$$

It may be shown after some calculations that

$$d_1 = \mu_2^{-1/2} + \tfrac{2}{3}\mu_3\mu_1 + O(n^{-1/2}h^r + n^{-3/2}),$$

$$d_2 = -\tfrac{1}{2}\mu_3^{-3/2}\mu_1 + O(n^{-1/2}h^r + n^{-3/2}), \qquad d_l = O(h^{2r}), \qquad l \ge 3,$$

$$d_{11} = \tfrac{2}{3}\mu_2^{-5/2}\mu_3 + O(h^r), \qquad d_{12} = -\tfrac{1}{2}\mu_2^{-3/2} + O(h^r),$$

$d_{lm} = O(h^r)$, for all other second derivatives,

$$d_{111} = -\tfrac{3}{2}\mu_2^{-7/2}\mu_4 + \tfrac{8}{3}\mu_2^{-9/2}\mu_3^2 + O(h^r), \qquad d_{112} = -\tfrac{5}{2}\mu_2^{-7/2}\mu_3 + O(h^r),$$

$$d_{113} = \tfrac{2}{3}\mu_2^{-5/2} + O(h^r), \qquad d_{122} = \tfrac{3}{4}\mu_2^{-5/2} + O(h^r),$$

$d_{ijk} = O(h^r)$, for all other third derivatives.

Moreover, we have

$$k^{11} = n^{-1}(\mu_2 - \mu_1^2), \qquad k^{12} = n^{-1}(\mu_3 - \mu_1\mu_2),$$

$$k^{13} = n^{-1}(\mu_4 - \mu_1\mu_3), \qquad k^{22} = n^{-1}(\mu_4 - \mu_2^2),$$

$$k^{111} = n^{-2}(\mu_3 - 3\mu_1\mu_2 + 2\mu_1^3), \qquad k^{112} = n^{-2}(\mu_4 - 2\mu_1\mu_3 + 2\mu_1\mu_2 - \mu_2^2),$$

$$k^{1111} = n^{-3}(\mu_4 - 2\mu_2^2) + O(n^{-3}\mu_1), \qquad k^{1112} = n^{-3}(\mu_5 - 2\mu_2\mu_3) + O(n^{-3}\mu_1).$$

Substituting the above derivatives and the multivariate cumulants of V into

(A.1)–(A.4), we are able to obtain

$$k_1 = n^{1/2}s_j(\mu) - \tfrac{1}{6}\mu_2^{-3/2}\mu_3 n^{-1/2} + O(n^{-1/2}h^r + n^{-3/2}),$$
$$k_2 = \sigma^2 + \left(\tfrac{1}{2}\mu_2^{-2}\mu_4 - \tfrac{13}{36}\mu_2^{-3}\mu_3^2\right)n^{-1} + O(n^{-1}h^r + n^{-2}),$$
$$k_3 = O(n^{-1/2}h^r), \qquad k_4 = O(n^{-1}h^r), \qquad k_l = O(n^{-(l-2)/2}) \quad \text{for } l \geq 5.$$

Acknowledgment. We are grateful to two reviewers, whose helpful comments have enabled us to significantly improve the presentation.

REFERENCES

BARNDORFF-NIELSEN, O. and COX, D. R. (1979). Edgeworth and saddle-point approximations with statistical applications (with discussion). *J. Roy. Statist. Soc. Ser. B* **41** 279–312.

BHATTACHARYA, R. N. and GHOSH, J. K. (1978). On the validity of the formal Edgeworth expansion. *Ann. Statist.* **6** 434–451.

BHATTACHARYA, R. N. and RAO, R. R. (1976). *Normal Approximation and Asymptotic Expansions.* Wiley, New York.

CHEN, S.-X. (1993). Empirical likelihood method. Ph.D. dissertation, Australian National Univ.

DICICCIO, T. J., HALL, P. and ROMANO, J. P. (1991). Empirical likelihood is Bartlett-correctable. *Ann. Statist.* **19** 1053–1061.

DUFFY, D. E. and SANTNER, T. J. (1990). *The Statistical Analysis of Discrete Data.* Springer, New York.

HALL, P. (1991). Edgeworth expansions for non-parametric density estimators, with applications. *Statistics* **22** 215–232.

HALL, P. and LA SCALA, B. (1990). Methodology and algorithms of empirical likelihood. *Internat. Statist. Rev.* **58** 109–127.

HALL, P. and SHEATHER, S. J. (1988). On the distribution of a Studentized quantile. *J. Roy. Statist. Soc. Ser. B* **50** 381–391.

HÄRDLE, W. (1990). *Applied Nonparametric Regression.* Cambridge Univ. Press.

HAYAKAWA, T. (1977). The likelihood ratio criterion and the asymptotic expansion of its distribution. *Ann. Inst. Statist. Math.* **29** 359–378.

JAMES, G. S. and MAYNE, A. J. (1962). Cumulants of functions of random variables. *Sankhyā Ser. A* **24** 47–54.

OWEN, A. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75** 237–249.

OWEN, A. (1990). Empirical likelihood ratio confidence regions. *Ann. Statist.* **18** 90–120.

PETROV, V. V. (1975). *Sums of Independent Random Variables.* Springer, New York.

PRESS, W. H., FLANNERY, B. F., TEUKOLSKY, S. A. and VETTERLING, W. T. (1989). *Numerical Recipes: The Art of Scientific Computing.* Cambridge Univ. Press.

SHEATHER, S. J. (1987). Assessing the accuracy of the sample median: estimated standard errors and interpolated confidence intervals. In *Statistical Data Analysis Based on the L_1 Norm and Related Methods* (Y. Dodge, ed.) 203–215. North-Holland, Amsterdam.

SHEATHER, S. J. and MCKEAN, J. W. (1987). A comparison of testing and confidence interval methods for the median. *Statist. Probab. Lett.* **6** 31–36.

SILVERMAN, B. W. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, London.

CENTRE FOR MATHEMATICS
AND ITS APPLICATIONS
AUSTRALIAN NATIONAL UNIVERSITY
GPO BOX 4
CANBERRA ACT 2601
AUSTRALIA