The Annals of Statistics
1990, Vol. 18, No. 1, 121–140

PSEUDO-LIKELIHOOD THEORY FOR EMPIRICAL LIKELIHOOD

BY PETER HALL[1]

Brown University

It is proved that, except for a location term, empirical likelihood does draw contours which are second-order correct for those of a pseudo-likelihood. However, except in the case of one dimension, this pseudo-likelihood is not that which would commonly be employed when constructing a likelihood-based confidence region. It is shown that empirical likelihood regions may be adjusted for location so as to render them second-order correct. Furthermore, it is proved that location-adjusted empirical likelihood regions are Bartlett-correctable, in the sense that a simple empirical scale correction applied to location-adjusted empirical likelihood reduces coverage error by an order of magnitude. However, the location adjustment alters the form of the Bartlett correction. It is also shown that empirical likelihood regions and bootstrap likelihood regions differ to second order, although both are based on statistics whose centred distributions agree to second order.

1. Introduction and summary. Let θ_0 be an unknown s-vector and $\hat{\theta}$ be an estimator of θ_0. Owen (1988, 1990) has suggested the method of empirical likelihood as a device for drawing "likelihood" contours and thereby constructing "likelihood-based" confidence regions for θ_0. We use quotation marks here because there has not yet been any theory to show that empirical likelihood does, in some sense, draw contours which approximate those of a likelihood function. The purpose of the present paper is to fill that gap.

Empirical likelihood contours may be constructed as follows. Let $\mathscr{X} = \{X_1, \dots, X_n\}$ denote an r-variate random sample which depends in some way on an unknown s-variate quantity $\theta = \theta_0$. For example, θ_0 may be the r-variate mean, $E(X)$, in which case $r = s$. Write $p = (p^1, \dots, p^n)$ for a multinomial distribution on the points $X_1, \dots, X_n$. (We denote vector elements by superscripts.) Let $\theta[p]$ denote the value taken by θ_0 when the true distribution is the multinomial. For example, when $\theta_0 = E(X)$ we have $\theta[p] = \sum p^i X_i$. Should a second sample be drawn, this time from the multinomial distribution conditional on $\mathscr{X}$, then the chance that the second sample (called a resample) is identical to $\mathscr{X}$ equals $\prod p^i$ multiplied by a combinatorial factor. For a wide range of choices of p, the most likely resample is, in fact, $\mathscr{X}$. In particular, this is the case when $p = (n^{-1}, \dots, n^{-1})$. The empirical likelihood of θ_1 is defined to be the "profile likelihood"

$$L_\theta(\theta_1) = \max_{p:\, \theta[p] = \theta_1} \prod p^i, \qquad \theta_1 \in \mathbb{R}^s.$$

Received June 1988; revised January 1989.

[1]On leave from The Australian National University, Canberra.

AMS 1980 *subject classifications*. Primary 62G05; secondary 62G10.

Key words and phrases. Bootstrap likelihood, confidence interval, Cornish–Fisher expansion, coverage, Edgeworth expansion, empirical likelihood, pseudo-likelihood, second-order correct.

Empirical likelihood contours are level sets of the form

$$\mathscr{C}_u = \{\theta_1 \in \mathbb{R}^s \colon L_\theta(\theta_1) = u\}, \qquad u > 0.$$

If $s = 1$, then $\mathscr{C}_u$ consists of two points, being the endpoints of a confidence interval. If $s \geq 2$, then $\mathscr{C}_u$ is a closed, connected $(s-1)$-dimensional surface, being the boundary of the confidence region

$$\mathscr{R}_u = \{\theta_1 \in \mathbb{R}^s \colon L_\theta(\theta_1) \geq u\}.$$

The coverage probability of this region equals $P(\theta_0 \in \mathscr{R}_u)$, which may be adjusted by altering the value of u. It is of interest to note that the bootstrap estimator of θ_0, given by $\hat{\theta} = \theta[n^{-1}, \dots, n^{-1}]$, is also the empirical maximum likelihood estimator. This follows from the fact that $\prod p^i$ is maximized over multinomials p by taking $p = (n^{-1}, \dots, n^{-1})$.

In many cases of practical importance, empirical likelihood satisfies a version of Wilks's theorem,

$$-2\log\{L_\theta(\theta_0)/L_\theta(\hat{\theta})\} \to \chi_s^2$$

in distribution [Owen (1988, 1990)]. Therefore, a practical approach to constructing a $(1-\alpha)$-level empirical likelihood confidence region is to find, from χ^2 tables, a value x such that $P(\chi_s^2 \leq x) = 1 - \alpha$, and put $u = n^{-n}e^{-x/2}$. In this event,

$$\mathscr{R}_u = \{\theta_1 \in \mathscr{R}^s \colon -2\log\{L_\theta(\theta_1)/L_\theta(\hat{\theta})\} \leq x\}. \tag{1.1}$$

More commonly, likelihood-based confidence regions are constructed as follows. Let $\hat{Q}$ be an estimate of the asymptotic variance matrix Q of $n^{1/2}\hat{\theta}$, and put

$$\hat{\eta}_0 = \hat{Q}^{-1/2}(\hat{\theta} - \theta_0).$$

If the density f of $\hat{\eta}_0$ were known, then we would take the confidence region to be the pseudo-likelihood region

$$\{\theta_1 \colon f\{\hat{Q}^{-1/2}(\hat{\theta} - \theta_1)\} \geq v\},$$

where v is chosen so that

$$\int_{\{y \colon f(y) \geq v\}} f(y)\, dy = 1 - \alpha.$$

(The qualifier "pseudo" is used since f is the density of a particular function of the data, and is not the likelihood of the entire data set.) However, the form of f is usually unknown, and so the assumption is often made that f is approximately equal to the standard normal density ϕ. In this case the confidence region becomes

$$\{\theta_1 \colon \|\hat{Q}^{-1/2}(\hat{\theta} - \theta_1)\| \leq r\},$$

where r is chosen so that

$$\int_{\{y \colon \|y\| \leq r\}} \phi(y)\, dy = 1 - \alpha.$$

Clearly, this approach fails to take any account of the skewness, kurtosis etc. of

$\hat{Q}^{-1/2}(\hat{\theta} - \theta_1)$. It produces familiar elliptical (ellipsoidal for $s \geq 3$) confidence regions. One would hope that the empirical likelihood technique somehow captures skewness, at least.

In this paper we argue that empirical likelihood confidence regions are approximately the same as pseudo-likelihood regions, but are not based on $\hat{\eta}_0$. Instead they are based on

$$\hat{\xi}_0 \equiv \left(Q^{1/2}\hat{Q}^{-1}Q^{1/2}\right)^{1/2} Q^{-1/2}\left(\hat{\theta} - \theta_0\right).$$

In a broad class of circumstances which will be made precise later in this section, we shall show that empirical likelihood draws contours which are second-order correct for pseudo-likelihood contours based on the distribution of $\hat{\xi}_0 + n^{-1}\psi$, where ψ is a certain fixed vector. (Here "second-order correct" means that contours at the same probability level drawn by the two different methods are distant $n^{-3/2}$ apart—they agree in terms of order n^{-1}.) Since ψ is usually nonzero, then these contours are not second-order correct for the pseudo-likelihood based on $\hat{\xi}_0$. Nevertheless, the contours are readily location-adjusted, which amounts to recentering by adding $n^{-1}\hat{\psi}$, where $\hat{\psi}$ is an estimate of ψ. This makes the contours second-order correct. They are of the correct size, shape and orientation.

Bootstrap likelihood is usually constructed so as to draw contours of pseudo-likelihood based on the distribution of $\hat{\eta}_0$. See Hall (1987). In the case of $s = 1$ dimension, $\hat{\xi}_0$ and $\hat{\eta}_0$ are identical. But for $s \geq 2$ dimensions, they differ in terms of second order. This means that there are second-order differences in shape between contours of pseudo-likelihood based on $\hat{\xi}_0$ and $\hat{\eta}_0$. Interestingly, the *distributions* of $\hat{\xi}_0 - E(\hat{\xi}_0)$ and $\hat{\eta}_0 - E(\hat{\eta}_0)$ agree to second order, and so in a certain sense the "average" contours of pseudo-likelihood regions based on distributions of $\hat{\xi}_0$ and $\hat{\eta}_0$ are identical in size, shape and orientation, to second order. However, the difference $\hat{\xi}_0 - \hat{\eta}_0 - E(\hat{\xi}_0 - \hat{\eta}_0)$ contains terms of second order, and $E(\hat{\xi}_0 - \hat{\eta}_0)$ is also of second order.

The context of our work is that where θ_0 is a vector-valued function of a vector of means. That is to say, $\theta_0 = K(EX)$, where $K: \mathbb{R}^r \to \mathbb{R}^s$ is a smooth function and EX denotes the r-variate population mean. Examples include vectors θ_0 whose elements are means, variances, standard deviations, correlation coefficients or functions (such as ratios) of any of these quantities. For the sake of definiteness we take our estimate $\hat{Q}$ of $\mathrm{var}(n^{1/2}\hat{\theta})$ to be the asymptotic variance matrix with unknowns replaced by their sample estimates. This prescription agrees to second order with both bootstrap and jackknife variance estimates—the differences between these estimates are only of third order. Therefore, all our results about second-order correctness continue to hold if $\hat{Q}$ is taken to be the bootstrap variance estimate.

Sufficient regularity conditions for all of these results would be that the underlying distributions have sufficiently many finite moments and satisfy Cramér's condition. That is to say, $E(|X|^c) < \infty$ for c sufficiently large, and

$$\limsup_{\|t\| \to \infty} |E \exp(it^T X)| < \infty,$$

where $i = \sqrt{-1}$ and $t = (t^1, \ldots, t^p)^T$ is a real vector. We do not go into details here because they are not at all relevant to our conclusions.

Section 2 will introduce notation connected with empirical likelihood, and Section 3 will describe and prove our main results about contours of pseudo-likelihood regions. In Section 4 we shall show how to adjust empirical likelihood regions for location and prove that location-adjusted regions may be Bartlett-corrected. Bartlett correction amounts to an empirical $O(n^{-1})$ adjustment to the value x in (1.1) and has the effect of reducing coverage error from $O(n^{-1})$ to $O(n^{-2})$. Section 5 will show that the statistics $\hat{\xi}_0 - E(\hat{\xi}_0)$ and $\hat{\eta}_0 - E(\hat{\eta}_0)$ differ to second order but have distributions which agree to second order.

Statisticians are usually particularly interested in shape and orientation of a confidence region, and the slight error in location which empirical likelihood commits might be overlooked for many purposes. However, we should stress that empirical likelihood does not do a good job contouring likelihood when the latter is defined strictly parametrically; see DiCiccio, Hall and Romano (1988a).

There is an element of arbitrariness in our definition of the statistic on which pseudo-likelihood is based, and that we freely acknowledge. However, by describing pseudo-likelihood relative to $\hat{\xi}_0$ we are able to establish a connection between empirical likelihood and likelihood in a more classical meaning of the term.

In the remainder of this paper we shall usually omit the qualifier "pseudo" affixed to "likelihood." It is to be understood that in so doing we intend "likelihood" to indicate the probability density of a function of the data, such as the function $\hat{\xi}_0$, and not the density of the data themselves.

2. Definitions. In this section we define the empirical likelihood with which we shall work and introduce other notation. Let $X_1, \dots, X_n$ be independent and identically distributed r-vectors with common mean $\mu_0 \equiv E(X)$ and nonsingular variance matrix $\Sigma_0 \equiv \mathrm{var}(X)$. Let $p \equiv (p^1, \dots, p^n)$ denote a multinomial distribution on the points $X_1, \dots, X_n$, and put

$$L(\mu) \equiv \max_{p:\, \Sigma p^i X_i = \mu} \prod p^i,$$

called the empirical likelihood of the r-vector μ. The function L is maximized at $\mu = \overline{X} \equiv n^{-1}\Sigma_i X_i$, where it attains the value n^{-n}. The empirical log-likelihood ratio is $L(\mu)/L(\overline{X})$. Twice its negative logarithm is the empirical log-likelihood,

$$\ell(\mu) \equiv -2\log\{L(\mu)/L(\overline{X})\}.$$

A little calculus of variations shows that

$$\ell(\mu) = 2\sum_i \log\{1 + t^T(X_i - \mu)\}, \tag{2.1}$$

where $t = t(\mu)$ is the solution of

$$\sum_i \{1 + t^T(X_i - \mu)\}^{-1}(X_i - \mu) = 0. \tag{2.2}$$

Let $\theta(\mu)$ denote an s-variate function of μ, where $s \le r$. Denote elements of vectors by superscripts; thus, μ^j is the jth element of μ. The log-likelihood of θ is

$$\ell_\theta\{\theta(\mu_1)\} \equiv \min_{\mu:\, \theta(\mu) = \theta(\mu_1)} \ell(\mu). \tag{2.3}$$

We take the minimum here since $\ell(\mu)$ was defined with a minus sign. Put $\hat{\theta} \equiv \theta(\overline{X})$ and $\theta_0 \equiv \theta(\mu_0)$. Define Θ to be the $s \times r$ matrix whose (u, j)th element is

$$\partial \theta^u / \partial \mu^j|_{\mu=\mu_0}.$$

Put

$$\theta^u_{jk} \equiv \partial \theta^u / \partial \mu^j \, \partial \mu^k|_{\mu=\mu_0}, \qquad \theta_{jk} \equiv \left(\theta^u_{jk}\right).$$

We assume that Θ is of full rank, s. Now,

$$\hat{\theta} - \theta_0 = \Theta\left(\overline{X} - \mu_0\right) + O_p(n^{-1}),$$

and so

$$n \operatorname{var}(\hat{\theta}) = \Theta \Sigma_0 \Theta^T + O(n^{-1}).$$

Therefore the asymptotic variance matrix of $n^{1/2}\hat{\theta}$ is

$$Q \equiv \Theta \Sigma_0 \Theta^T.$$

As its estimate we take

$$\hat{Q} \equiv \hat{\Theta} \hat{\Sigma} \hat{\Theta}^T,$$

where

$$\hat{\Theta}^{uj} \equiv \partial \theta^u / \partial \mu^j|_{\mu=\overline{X}}, \qquad \hat{\Sigma}^{jk} \equiv n^{-1} \sum_i \left(X_i^j X_i^k - \overline{X}^j \overline{X}^k\right).$$

Put $\hat{H} \equiv (Q^{1/2}\hat{Q}^{-1}Q^{1/2})^{1/2}Q^{-1/2}$, $A \equiv \overline{X}$, $B \equiv \Theta^T Q^{-1} \Theta \overline{X}$,

$$\hat{\xi}_0 \equiv \hat{H}\left(\hat{\theta} - \theta_0\right), \qquad \hat{\eta}_0 \equiv \hat{Q}^{-1}\left(\hat{\theta} - \theta_0\right) \quad \text{and} \quad \hat{\xi} = \hat{\xi}(\mu_1) = \hat{H}\left\{\hat{\theta} - \theta(\mu_1)\right\}.$$

We shall usually take $\mu_0 = 0$ and $\Sigma_0 = I$, both of which assumptions may be made without loss of generality, since they amount to only a change of variable in the function θ. Of course, the condition $\Sigma_0 = I$ does not mean that we may drop the matrix $\hat{\Sigma}$ from our definition of $\hat{Q}$, but it does permit a simpler definition of Q:

$$Q \equiv \Theta \Theta^T.$$

Throughout this paper we use the terms "first-order," "second-order" etc. in the usual way for statistical theory: first-order properties are those connected with simple normal approximation; second-order properties are those connected with basic skewness and asymmetry corrections to the first-order approximations; third-order properties are those concerned with kurtosis and squared skewness corrections, and so on. Therefore, in an Edgeworth expansion of the density of $n^{1/2}\hat{\xi}_0$ or of $n^{1/2}\hat{\eta}_0$, second-order terms are those of order $n^{-1/2}$, third-order terms are those of order n^{-1}, etc. In a confidence region, first-order terms are of order $n^{-1/2}$, second-order terms are of order n^{-1}, third-order terms are of order $n^{-3/2}$, and so on.

We write ϕ for the s-variate standard normal density and use summation notation in Sections 3, 4 and 5. That is to say, any index which is repeated anywhere in a product, as either a subscript or a superscript, is intended to be summed over.

3. Relationship between empirical log-likelihood and true log-likelihood of $\hat{\xi}_0$.

3.1. *Introduction and summary.* Let ψ be a certain constant vector, to be defined shortly. Our aim in this section is to prove that, neglecting a remainder of third order, the log-likelihood of $\hat{\xi} + n^{-1}\psi$ is identical to empirical log-likelihood when the latter is expressed as a function of $\hat{\xi}$. Therefore, except for the location term $n^{-1}\psi$, empirical likelihood captures the contours of the true likelihood of $\hat{\xi}$, to second order.

Our proof of this result is comprised of five parts. First, in Section 3.2 we develop an approximation to the probability density of $\hat{\xi}_0 - E(\hat{\xi}_0)$. Then, in Section 3.3 we approximate the density of $\hat{\xi}_0 + n^{-1}\psi$. Section 3.4 derives an expansion of empirical log-likelihood, Section 3.5 employs that result to express empirical log-likelihood as a function of $\hat{\xi}$ and Section 3.6 draws these conclusions together to obtain the result described in the previous paragraph.

The main contributions of this section may be quantified as follows. Define ψ as at (3.15), put $\tilde{\xi} \equiv \hat{\xi}_0 + n^{-1}\psi$ and let g denote the density of $n^{1/2}\tilde{\xi}$. To first order, $n^{1/2}\tilde{\xi}$ is asymptotically s-variate normal $N(0,1)$. To second order, g is well-approximated by an Edgeworth expansion,

$$g(y) = \phi(y)\{1 + n^{-1/2}q(y) + O(n^{-1})\}$$
$$= (2\pi)^{-s/2}\exp\{-\tfrac{1}{2}y^T y + n^{-1/2}q(y) + O(n^{-1})\},$$

where ϕ is the s-variate standard normal density and q is a polynomial containing only terms of precise degree 3. Therefore, twice the negative of the logarithm of the density of $\tilde{\xi}$ equals

$$T(x) \equiv -2\log\{n^{s/2}g(n^{1/2}x)\} = nx^T x - n2q(x) + s\log(2\pi/n) + O_p(n^{-1}),$$

where x is taken to be of order $n^{-1/2}$. We shall prove that

$$\ell_\theta\{\theta(\mu_1)\} = T\{\hat{\xi}(\mu_1)\} + s\log(n/2\pi) + O_p(n^{-1}), \tag{3.1}$$

where $\hat{\xi}(\mu_1) = \hat{H}\{\hat{\theta} - \theta(\mu_1)\}$ is intended to be regarded as a function of μ_1. The fact that T on the right-hand side is evaluated at $\hat{\xi}(\mu_1)$ and not $\hat{\xi}(\mu_1) + n^{-1}\psi$, even though T is proportional to the log of the density of $\hat{\xi}(\mu_0) + n^{-1}\psi$, indicates the necessity of the location adjustment $n^{-1}\psi$.

3.2. *Approximation to density of* $\hat{\xi}_0 - E\hat{\xi}_0$. Without any loss of generality we fix location and scale at the values $E(X) = 0$ and $\mathrm{var}(X) = I$. Should these conditions not hold, they may be made to hold by introducing the obvious change of variable to the function θ. Under the conditions, the standardized third moment

$$\alpha^{jkl} \equiv E\left[\{\Sigma^{-1/2}(X - EX)\}^j\{\Sigma^{-1/2}(X - EX)\}^k\{\Sigma^{-1/2}(X - EX)\}^l\right]$$

simplifies to $\alpha^{jkl} = E(X^j X^k X^l)$.

Recall from Section 2 that

$$\hat{\xi}_0 = \left(Q^{1/2}\hat{Q}^{-1}Q^{1/2}\right)^{1/2}Q^{-1/2}\left(\hat{\theta} - \theta_0\right).$$

Our principal aim in this subsection is to derive an Edgeworth approximation to the likelihood, or probability density, of $n^{1/2}(\hat{\xi}_0 - E\hat{\xi}_0)$. We shall prove that if f denotes that density then

$$f(y) = \phi(y)\left\{1 + n^{-1/2}\tfrac{1}{6}(\lambda^{uvw}y^u y^v y^w - 3\lambda^{uuv}y^v) + O(n^{-1})\right\}, \tag{3.2}$$

where

$$\lambda^{uvw} \equiv -\Big\{2\left(Q^{-1/2}\Theta\right)^{uj}\left(Q^{-1/2}\Theta\right)^{vk}\left(Q^{-1/2}\Theta\right)^{wl}\alpha^{jkl} + \left(Q^{-1/2}\Theta\right)^{uj}\left(Q^{-1/2}\Theta\right)^{vk}\left(Q^{-1/2}\theta_{jk}\right)^{w}[3]\Big\} \tag{3.3}$$

and [3] indicates that the previous term should be included for each of the three arrangements of its superscripts. We shall also derive an approximation to $E(\hat{\xi}_0)$:

$$E\left(\hat{\xi}_0\right) = n^{-1}\psi_1 + O(n^{-2}), \tag{3.4}$$

where

$$\psi_1^u \equiv \left\{Q^{-1/2}\left(\tfrac{1}{2}\theta_{jj} - \theta_{jk}R^{jk}\right)\right\}^u - \tfrac{1}{2}\left(Q^{-1/2}\Theta\right)^{uj}R^{kl}\alpha^{jkl} \tag{3.5}$$

and $R \equiv \Theta^T Q^{-1}\Theta$. This result is required in Section 3.3.

Let δ^{jk} denote the Kronecker delta. Observe that

$$\hat{Q} = Q + \Delta + O_p(n^{-1}) \tag{3.6}$$

where, with $A \equiv \bar{X}$ and $A^{jk} \equiv n^{-1}\Sigma(X_i^j X_i^k - \delta^{jk})$, we have

$$\Delta^{uv} \equiv \left(\theta_{jk}^u\Theta^{vj} + \theta_{jk}^v\Theta^{uj}\right)A^k + \Theta^{uj}\Theta^{vk}A^{jk}. \tag{3.7}$$

In this notation,

$$\hat{Q}^{-1} = Q^{-1} - Q^{-1}\Delta Q^{-1} + O_p(n^{-1}),$$

so that

$$\left(Q^{1/2}\hat{Q}^{-1}Q^{1/2}\right)^{1/2}Q^{-1/2} = Q^{-1/2} - \tfrac{1}{2}Q^{-1/2}\Delta Q^{-1} + O_p(n^{-1}).$$

Furthermore,

$$\hat{\theta} - \theta_0 = \Theta A + \tfrac{1}{2}\theta_{jk}A^j A^k + O_p(n^{-3/2}), \tag{3.8}$$

$$\begin{aligned}\hat{\xi}_0 &= Q^{-1/2}\left(\Theta A + \tfrac{1}{2}\theta_{jk}A^jA^k\right) - \tfrac{1}{2}Q^{-1/2}\Delta Q^{-1}\Theta A + O_p(n^{-3/2})\\ &= Q^{-1/2}\Theta A + \left(\tfrac{1}{2}Q^{-1/2}\theta_{jk} - Q^{-1/2}\theta_{kl}R^{lj}\right)A^jA^k \\ &\quad - \tfrac{1}{2}Q^{-1/2}\Theta A^* R A + O_p(n^{-3/2}),\end{aligned} \tag{3.9}$$

where $R \equiv \Theta^T Q^{-1}\Theta$ and $A^* \equiv (A^{jk})$.

Next we compute an Edgeworth expansion for a statistic admitting the Taylor expansion (3.9). Put $U \equiv V + W$ where $V \equiv Q^{-1/2}\Theta A$ and

$$W \equiv \left(\tfrac{1}{2}Q^{-1/2}\theta_{jk} - Q^{-1/2}\theta_{kl}R^{lj}\right)A^jA^k - \tfrac{1}{2}Q^{-1/2}\Theta A^*RA. \tag{3.10}$$

Then V and W are respectively linear and quadratic functions of sample means, those means all having zero expected values. Let the means be components of the vector $\bar{Z} \equiv n^{-1}\Sigma_i Z_i$, where each Z_i is distributed as the generic Z and $E(Z) = 0$, $E(Z^aZ^b) = \zeta^{ab}$, $E(Z^aZ^bZ^c) = \zeta^{abc}$. In this notation we may write $V^u = \beta_a^u\bar{Z}^a$ and $W_{ab}^u = \gamma_{ab}^u\bar{Z}^a\bar{Z}^b$, for suitable constants β_a^u and γ_{ab}^u. Write δ^{uv} for the Kronecker delta. Then

$$\begin{aligned}
E(U^u) &= E(W^u) = n^{-1}\gamma_{ab}^u\zeta^{ab},\\
E(V^uV^v) &= n^{-1}\,\delta^{uv},\\
E(U^uV^v) &= E(V^uV^v) + O(n^{-2})\\
&= n^{-1}\,\delta^{uv} + O(n^{-2}),\\
E(V^uV^vV^w) &= n^{-2}\beta_a^u\beta_b^v\beta_c^w\zeta^{abc}, \qquad (3.11)\\
E(V^uV^vV^w) &= \beta_a^u\beta_b^v\gamma_{cd}^wE(\bar{Z}^a\bar{Z}^b\bar{Z}^c\bar{Z}^d)\\
&= \beta_a^u\beta_b^v\gamma_{cd}^w\{E(\bar{Z}^a\bar{Z}^b)E(\bar{Z}^c\bar{Z}^d)\\
&\qquad + E(\bar{Z}^a\bar{Z}^c)E(\bar{Z}^b\bar{Z}^d)\\
&\qquad + E(\bar{Z}^a\bar{Z}^d)E(\bar{Z}^b\bar{Z}^c)\} + O(n^{-3}).
\end{aligned}$$

Third-order cumulants of U have the form

$$\begin{aligned}
\kappa^{uvw} &= E(U^uU^vU^w) - E(U^uU^v)E(U^w)[3] + 2E(U^u)E(U^v)E(U^w)\\
&= E(V^uV^vV^w) + \{E(V^uV^vW^w) - E(V^uV^v)E(W^w)\}[3] + O(n^{-3}),
\end{aligned}$$

where [3] indicates that the previous term should be included for each of the three distinct arrangements of its superscripts u, v, w. Hence, using the moment formulae derived just above,

$$\begin{aligned}
\kappa^{uvw} &= n^{-2}\{\beta_a^u\beta_b^v\beta_c^w\zeta^{abc} + \beta_a^u\beta_b^v\gamma_{cd}^w(\zeta^{ac}\zeta^{bd} + \zeta^{ad}\zeta^{bc})[3]\} + O(n^{-3})\\
&= n^{-2}(\beta_a^u\beta_b^v\beta_c^w\zeta^{abc} + 2\beta_a^u\beta_b^v\gamma_{cd}^w\zeta^{ac}\zeta^{bd}[3]) + O(n^{-3}).
\end{aligned} \tag{3.12}$$

At this stage we substitute explicit formulae for the constants β, γ and ζ, using formulae for U and V given at the beginning of the previous paragraph. We have $\beta_a^u = (Q^{-1/2}\Theta)^{ua}$ if $\bar{Z}^a$ is of the form A^a, and $\beta_a^u = 0$ otherwise;

$$\gamma_{ab}^u = \tfrac{1}{2}\left(Q^{-1/2}\theta_{ab}\right)^u - \left(Q^{-1/2}\theta_{al}\right)^uR^{lb}$$

if $(\bar{Z}^a, \bar{Z}^b)$ is of the form (A^a, A^b), $\gamma_{ab}^u = -\tfrac{1}{2}(Q^{-1/2}\Theta)^{ui}R^{jb}$ if $(\bar{Z}^a, \bar{Z}^b)$ is of the form (A^{ij}, A^b) and $\gamma_{ab}^u = 0$ otherwise. See (3.10). Furthermore, $\zeta^{abc} = \alpha^{abc} = E(X^aX^bX^c)$ if $(\bar{Z}^a, \bar{Z}^b, \bar{Z}^c) = (A^a, A^b, A^c)$; $\zeta^{ab} = \delta^{ab}$ if $(\bar{Z}^a, \bar{Z}^b) = (A^a, A^b)$;

and $\zeta^{ab} = \alpha^{ijb}$ if $(\bar{Z}^a, \bar{Z}^b) = (A^{ij}, A^b)$. Substituting into (3.12) we deduce that

$$\begin{aligned}
\kappa^{uvw} &= n^{-2}\Big[\big(Q^{-1/2}\Theta\big)^{ua}\big(Q^{-1/2}\Theta\big)^{vb}\big(Q^{-1/2}\Theta\big)^{wc}\alpha^{abc} \\
&\quad + \big(Q^{-1/2}\Theta\big)^{ua}\big(Q^{-1/2}\Theta\big)^{vb}\Big\{\big(Q^{-1/2}\theta_{ab}\big)^{w} - 2\big(Q^{-1/2}\theta_{al}\big)^{w}R^{lb}\Big\}[3] \\
&\quad - \big(Q^{-1/2}\Theta\big)^{ua}\big(Q^{-1/2}\Theta\big)^{vb}\big(Q^{-1/2}\Theta\big)^{wi}R^{jb}\alpha^{aij}[3]\Big] + O(n^{-3}) \\
&= -n^{-2}\Big\{2\big(Q^{-1/2}\Theta\big)^{ua}\big(Q^{-1/2}\Theta\big)^{vb}\big(Q^{-1/2}\Theta\big)^{wc}\alpha^{abc} \\
&\quad + \big(Q^{-1/2}\Theta\big)^{ua}\big(Q^{-1/2}\Theta\big)^{vb}\big(Q^{-1/2}\Theta_{ab}\big)^{w}[3]\Big\} + O(n^{-3}) \\
&= n^{-2}\lambda^{uvw} + O(n^{-3}),
\end{aligned} \tag{3.13}$$

where λ is given by (3.3).

First-order cumulants of $Y \equiv n^{1/2}(U - EU)$ are identically zero; the second-order cumulant of type (u, v) of Y equals $\delta^{uv} + O(n^{-1})$, using (3.11); the third-order cumulant of type (u, v, w) of Y equals $n^{3/2}$ multiplied by the corresponding cumulant of U, and so is $n^{-1/2}\lambda^{uvw} + O(n^{-3/2})$, by (3.13); and fourth- and higher-order cumulants are all $O(n^{-1})$ [James (1955, 1958)]. Hence, the characteristic function of Y, expressed as a function of the s-vector t, equals

$$\begin{aligned}
&\exp\big\{-\tfrac{1}{2}t^u t^u + n^{-1/2}\tfrac{1}{6}\lambda^{uvw}(it)^u(it)^v(it)^w + O(n^{-1})\big\} \\
&\qquad = \exp\big(-\tfrac{1}{2}t^u t^u\big)\big\{1 + n^{-1/2}\lambda^{uvw}(it)^u(it)^v(it)^w + O(n^{-1})\big\}.
\end{aligned}$$

Since Y and $n^{1/2}(\hat{\xi}_0 - E\hat{\xi}_0)$ differ only in terms of $O_p(n^{-1})$, as shown by (3.9), then this is also the characteristic function of $n^{1/2}(\hat{\xi}_0 - E\hat{\xi}_0)$ or, equivalently, the Fourier transform of the density of $n^{1/2}(\hat{\xi}_0 - E\hat{\xi}_0)$. Inverting that Fourier transform we conclude that $n^{1/2}(\hat{\xi}_0 - E\hat{\xi}_0)$ has density given by (3.2).

To obtain (3.4) for $E(\hat{\xi}_0)$, observe that by (3.9) and (3.10),

$$E\big(\hat{\xi}_0^u\big) = E(W^u) + O(n^{-2}) = n^{-1}\gamma_{ab}^u\zeta^{ab} + O(n^{-2}).$$

Now use formulae given above for γ_{ab}^u and ζ^{ab}.

3.3. *A statistic whose density admits a certain Edgeworth expansion.* Define q to be the cubic

$$q(y) \equiv \tfrac{1}{6}\lambda^{uvw}y^u y^v y^w, \tag{3.14}$$

where λ^{uvw} is given by (3.3). Let ψ denote the s-vector whose uth element is given by

$$\begin{aligned}
\psi^u &\equiv -\tfrac{1}{2}\big(Q^{-1/2}\Theta\big)^{uj}R^{kl}\alpha^{jkl} + \tfrac{1}{2}\Big\{Q^{-1/2}\Big(\theta_{jk}R^{jk} - \theta_{jj}\Big)\Big\}^u \\
&\quad - \big(Q^{-1/2}\Theta\big)^{uj}\Big(\Theta^T Q^{-1}\theta_{jk}\Big)^k,
\end{aligned} \tag{3.15}$$

where $R \equiv \Theta^T Q^{-1}\Theta$. We shall prove in this subsection that the density g of the statistic $n^{1/2}\hat{\xi}_0 + n^{-1/2}\psi$ admits the Edgeworth expansion

$$g(y) = \phi(y)\big\{1 + n^{-1/2}q(y) + O(n^{-1})\big\}. \tag{3.16}$$

We know from (3.2) that the density f of $n^{1/2}(\hat{\xi}_0 - E\hat{\xi}_0)$ admits the expansion

$$f(y) = \phi(y)\left[1 + n^{-1/2}\{q(y) - \tfrac{1}{2}\lambda^{uuv}y^v\} + O(n^{-1})\right].$$

Therefore the density g of $n^{1/2}\hat{\xi}_0 + n^{-1/2}\psi$ is given by

$$\begin{aligned} g(y) &= f\left(y - n^{-1/2}\psi - n^{1/2}E\hat{\xi}_0\right) \\ &= \phi(y)\left[1 + n^{-1/2}\left\{q(y) + \left(-\tfrac{1}{2}\lambda^{uuv} + \psi^v + nE\hat{\xi}_0^v\right)y^v\right\} + O(n^{-1})\right]. \end{aligned}$$

This gives us (3.16) provided

$$\psi^v = \tfrac{1}{2}\lambda^{uuv} - nE\hat{\xi}_0^v + O(n^{-1/2}).$$

If ψ is given by (3.15) then the above relation follows from (3.3)–(3.5).

3.4. *Approximation to empirical log-likelihood.* We begin with notation. Since we have stipulated that $\Sigma = \operatorname{var}(X) = I$, then the (j,k)th element of Σ is δ^{jk}, the Kronecker delta. Define

$$A^{jk} \equiv n^{-1}\sum_i \left(X_i^j X_i^k - \delta^{jk}\right), \qquad \theta_{jk}^u \equiv \partial^2\theta^u/\partial\mu^j\,\partial\mu^k|_{\mu=0},$$

$$A \equiv \bar{X}, \qquad B \equiv \Theta^T Q^{-1}\Theta A, \qquad A^* \equiv (A^{jk}),$$

$$U^u \equiv \tfrac{1}{2}\theta_{jk}^u(A - B)^j(A - B)^k \quad \text{and} \quad \Omega^{uj} \equiv \theta_{jk}^u\mu_l^k.$$

(We have redefined U.) Let θ_{jk} be the s-vector whose uth element is θ_{jk}^u. In this subsection we derive the following approximation to the empirical log-likelihood ratio l_θ, defined at (2.3): For s-vectors μ_1 of order $n^{-1/2}$,

$$\begin{aligned} n^{-1}l_\theta(\mu_1) &= (B - \mu_1)^T(B - \mu_1) + \tfrac{2}{3}\alpha^{jkl}(B - \mu_1)^j(B - \mu_1)^k(B - \mu_1)^l \\ &\quad -(B - \mu_1)^T A^*(B - \mu_1) \qquad (3.17)\\ &\quad + 2(B - \mu_1)^T\Theta^T Q^{-1}\{U + \Omega(A - B)\} + O_p(n^{-2}). \end{aligned}$$

Our starting point is the empirical log-likelihood ratio $l(\mu)$, defined at (2.1). Let μ be of order $n^{-1/2}$. It is readily seen that the solution t of (2.2) satisfies $t = A - \mu + \varepsilon_1$, where $\varepsilon_1 = O_p(n^{-1})$. Taylor-expanding the left-hand side of (2.2) as a function of t, substituting $t = A - \mu + \varepsilon_1$ and solving for ε_1, we conclude that

$$\varepsilon_1^j = \alpha^{jkl}(A - \mu)^k(A - \mu)^l - A^{jk}(A - \mu)^k + O_p(n^{-3/2}).$$

Putting $t = A - \mu + \varepsilon_1$, with this ε_1, into (2.1), and Taylor-expanding once more, we obtain

$$\begin{aligned} (3.18)\quad n^{-1}l(\mu) &= (A - \mu)^T(A - \mu) + \tfrac{2}{3}\alpha^{jkl}(A - \mu)^j(A - \mu)^k(A - \mu)^l \\ &\quad - A^{jk}(A - \mu)^j(A - \mu)^k + O_p(n^{-2}). \end{aligned}$$

It is not difficult to see that if $\hat{\mu}_1 = \hat{\mu}_1(\mu_1)$ is the value of μ which minimizes $l(\mu)$ subject to $\theta(\mu) = \theta(\mu_1)$, then $\hat{\mu}_1 = \mu_1 + A - B + \varepsilon$, where $\varepsilon = O_p(n^{-1})$. To find ε, note that

(3.19) $$n^{-1}l(\hat{\mu}_1) = a(\varepsilon) + (\text{terms not depending on } \varepsilon) + O_p(n^{-5/2}),$$

where

$$a(x) \equiv x^j x^j - 2x^j\{(B-\mu_1)^j + \alpha^{jkl}(B-\mu_1)^k(B-\mu_1)^l - A^{jk}(B-\mu_1)^k\}.$$

The latter function is exactly the term in x which arises on replacing μ by $\mu_1 + A - B + x$ on the right-hand side of (3.18), neglecting the term $O_p(n^{-2})$. Formula (3.19) is obtained on noting that the term of precise order n^{-2} in (3.18) has a form similar to that of the order $n^{-3/2}$ terms, and that it contains no terms in x of precise order n^{-2} when μ is replaced by $\mu_1 + A - B + x$ and x is of order n^{-1}.

Write $\hat{\theta}_1$ for $\theta(\hat{\mu}_1)$ and θ_1 for $\theta(\mu_1)$. We may show after a little Taylor expansion that for μ_1 of order $n^{-1/2}$, and with $\hat{\mu}_1 = \mu_1 + A - B + \varepsilon$, we have

(3.20)
$$\hat{\theta}_1 - \theta_1 = (\Theta + \Omega)(A - B + \varepsilon) + \tfrac{1}{2}\theta_{jk}(A-B)^j(A-B)^k + O_p(n^{-3/2}).$$

Therefore, up to terms of order $n^{-3/2}$, ε may be found by minimizing $a(\varepsilon)$ subject to

$$(\Theta + \Omega)(A - B + \varepsilon) + \tfrac{1}{2}\theta_{jk}(A-B)^j(A-B)^k = 0.$$

This may be done using Lagrange multipliers, and produces

(3.21)
$$\varepsilon = (I - R)(B - \mu_1 + V) - \Theta^T Q^{-1}\{U + \Omega(A - B)\} + O_p(n^{-3/2}),$$

where $R = \Theta^T Q^{-1}\Theta$, $U^u \equiv \frac{1}{2}\theta^u_{jk}(A-B)^j(A-B)^k$ and

$$V^j \equiv \alpha^{jkl}(B-\mu_1)^k(B-\mu_1)^l - A^{jk}(B-\mu_1)^k.$$

At this point we note that the vector μ_1 may be restricted, as follows. Suppose we are given an s-vector μ_1 of order $n^{-1/2}$. Put $v_1 \equiv \mu_1$,

$$v_3 \equiv \tfrac{1}{2}\Theta^T Q^{-1}\theta_{jk}\{\mu_1^j\mu_1^k - (R\mu_1)^j(R\mu_1)^k\}$$

and $\mu_3 \equiv Rv_1 + v_3$. It follows after a little algebra that $\theta(\mu_3) = \theta(\mu_1) + O(n^{-3/2})$. Further refinement of μ_3 by adding a term of order $n^{-3/2}$ allows us to declare that for a vector $\mu_2 = Rv_1 + v_2$, where $v_2 = O(n^{-1})$, we have precisely $\theta(\mu_2) = \theta(\mu_1)$. In our work μ_1 serves only to index the value of $\theta(\mu_1)$, and so we may assume without loss of generality that $\mu_1 = Rv_1 + v_2$ where $v_1 = O(n^{-1/2})$ and $v_2 = O(n^{-1})$. For such a μ_1, noting that the matrix R is idempotent, we have

$$(I - R)(B - \mu_1) = (I - R)\{R(A - v_1) - v_2\} = (R - I)v_2 = O(n^{-1}),$$

whence

$$\begin{aligned}(B-\mu_1)^T(I-R)(B-\mu_1+V) &= \{(I-R)(B-\mu_1)\}^T(I-R)(B-\mu_1+V)\\ &= v_2^T(R-I)(v_2 - V) = O_p(n^{-2}).\end{aligned}$$

Therefore, taking $\mu = \hat{\mu}_1 = \mu_1 + A - B + \varepsilon$ in (3.18), noting (3.21) for ε and remembering that $\varepsilon = O_p(n^{-1})$, we deduce that $l(\hat{\mu}_1) = l_\theta(\mu_1)$ admits the formula announced at (3.17).

3.5. *Empirical log-likelihood as a function of* $\hat{\xi}$. Recall from Section 2 that

$$\hat{\xi} = \hat{\xi}(\mu_1) \equiv \left(Q^{1/2}\hat{Q}^{-1}Q^{1/2}\right)^{1/2} Q^{-1/2}\{\hat{\theta} - \theta(\mu_1)\}.$$

Let q be the cubic polynomial defined at (3.14). We show in this subsection that the log-likelihood expansion at (3.17) is equivalent to

$$n^{-1} l_\theta(\mu_1) = \hat{\xi}^u \hat{\xi}^u - 2q(\hat{\xi}) + O_p(n^{-2}). \tag{3.22}$$

Let Ω be as in Section 3.4, and note that, analogously to (3.8),

$$\begin{aligned}\hat{\theta} - \theta_1 &= \theta(\bar{X}) - \theta(\mu_1)\\ &= (\Theta + \Omega)(A - \mu_1) + \tfrac{1}{2}\theta_{jk}(A - \mu_1)^j (A - \mu_1)^k + O_p(n^{-3/2}).\end{aligned}$$

From this approximation and (3.6) and (3.7) we obtain

$$\begin{aligned}\hat{\xi}^T\hat{\xi} &= \left(\hat{\theta} - \theta_1\right)^T \hat{Q}^{-1}\left(\hat{\theta} - \theta_1\right)\\ &= \left(\hat{\theta} - \theta_1\right)^T \left(Q^{-1} - Q^{-1}\Delta Q^{-1}\right)\left(\hat{\theta} - \theta_1\right) + O_p(n^{-2})\\ &= (A - \mu_1)^T R (A - \mu_1) + (A - \mu_1)^T \Theta^T Q^{-1}\theta_{jk}(A - \mu_1)^j (A - \mu_1)^k \\ &\quad + 2(A - \mu_1)^T \Theta^T Q^{-1}\Omega(A - \mu_1)\\ &\quad - (A - \mu_1)^T \Theta^T Q^{-1}\Delta Q^{-1}\Theta(A - \mu_1) + O_p(n^{-2}),\end{aligned} \tag{3.23}$$

where $R \equiv \Theta^T Q^{-1}\Theta$. Recalling from the last paragraph of Section 3.4 that $\mu_1 = R v_1 + v_2$, where $v_1 = O(n^{-1/2})$ and $v_2 = O(n^{-1})$, we see after some algebra that

$$\begin{aligned}(A - \mu_1)^T R (A - \mu_1) &= (B - \mu_1)^T (B - \mu_1) + v_2^T (I - R) v_2\\ &= (B - \mu_1)^T (B - \mu_1) + O_p(n^{-2}).\end{aligned}$$

Since $\Theta(A - \mu_1) = \Theta(B - \mu_1)$ and $R\mu_1 = \mu_1 + O(n^{-1})$, then the sum of all terms on the right-hand side of (3.23) which involve θ_{jk}'s, equals

$$\begin{aligned}&(B - \mu_1)^T \Theta^T Q^{-1}\theta_{jk}\{(A - \mu_1)^j (A - \mu_1)^k\\ &\qquad\qquad + 2(A - \mu_1)^j \mu_1^k - 2A^j (B - \mu_1)^k\} + O_p(n^{-2})\\ &\quad = (B - \mu_1)^T \Theta^T Q^{-1}\theta_{jk}\left(A^j A^k - 2A^j B^k + 2A^j \mu_1^k - \mu_1^j \mu_1^k\right) + O_p(n^{-2})\\ &\quad = 2(B - \mu_1)^T \Theta^T Q^{-1}\{U + \Omega(A - B)\}\\ &\qquad - (B - \mu_1)^T \Theta^T Q^{-1}\theta_{jk}(B - \mu_1)^j (B - \mu_1)^k + O_p(n^{-2}),\end{aligned}$$

where U is as in (3.17). The only other contribution of order $n^{-3/2}$ on the right-hand side of (3.23) is

$$-(A-\mu_1)^T\Theta^TQ^{-1}\Theta A^*\Theta^TQ^{-1}\Theta(A-\mu_1)$$
$$=-(B-\mu_1)^TA^*(B-\mu_1)+O_p(n^{-2}).$$

Adding, we find that

$$Q\hat{\xi}^T\hat{\xi}=(B-\mu_1)^T(B-\mu_1)+2(B-\mu_1)^T\Theta^TQ^{-1}\{U+\Omega(A-B)\}$$
$$-(B-\mu_1)^T\Theta^TQ^{-1}\theta_{jk}(B-\mu_1)^j(B-\mu_1)^k \tag{3.24}$$
$$-(B-\mu_1)^TA^*(B-\mu_1)+O_p(n^{-2}).$$

Since

$$\hat{\xi}=Q^{-1/2}(\hat{\theta}-\theta)+O_p(n^{-1})=Q^{-1/2}\Theta(A-\mu_1)+O_p(n^{-1})$$
$$=Q^{-1/2}\Theta(B-\mu_1)+O_p(n^{-1}),$$

then

$$\Theta^TQ^{-1/2}\hat{\xi}=R(B-\mu_1)+O_p(n^{-1})=B-\mu_1+O_p(n^{-1}).$$

We may now deduce from (3.24) that the right-hand side of (3.17) equals

$$\hat{\xi}^T\hat{\xi}+\left\{\tfrac{2}{3}\alpha^{jkl}+\left(\Theta^TQ^{-1}\theta_{jk}\right)^l\right\}(B-\mu_1)^j(B-\mu_1)^k(B-\mu_1)^l+O_p(n^{-2})$$
$$=\hat{\xi}^u\hat{\xi}^u+\left\{\tfrac{2}{3}(Q^{-1/2}\Theta)^{uj}(Q^{-1/2}\Theta)^{vk}(Q^{-1/2}\Theta)^{wl}\alpha^{jkl}\right.$$
$$\left.+\tfrac{1}{3}(Q^{-1/2}\Theta)^{uj}(Q^{-1/2}\Theta)^{vk}\left(Q^{-1/2}\Theta_{jk}\right)^w[3]\right\}\hat{\xi}^u\hat{\xi}^v\hat{\xi}^w+O_p(n^{-2}),$$

which is identical to the right-hand side of (3.22).

3.6. *Completion.* Here we show that if ψ is the vector defined at (3.15), then twice the negative of the logarithm of the density of $\hat{\xi}_0+n^{-1}\psi$, expressed as a function of $\hat{\xi}(\mu_1)$, equals

$$l_\theta(\mu_1)+s\log(2\pi/n)+O_p(n^{-1}) \tag{3.25}$$

for vectors μ_1 of order $n^{-1/2}$. This verifies (3.1).

Recall from Section 3.3 that the density g of $n^{1/2}\hat{\xi}_0+n^{-1/2}\psi$ admits the Edgeworth expansion

$$g(y)=\phi(y)\{1+n^{-1/2}q(y)+O(n^{-1})\}.$$

Therefore the density of $\hat{\xi}_0+n^{-1}\psi$ is

$$n^{s/2}g(n^{1/2}y)=n^{s/2}(2\pi)^{-s/2}\phi(n^{1/2}y)\{1+n^{-1/2}q(n^{1/2}y)+O(n^{-1})\}$$
$$=(n/2\pi)^{s/2}\exp\{-\tfrac{1}{2}ny^Ty+nq(y)+O(n^{-1})\}.$$

Here we have taken y to be of order $n^{-1/2}$ and used the fact that $q(cy)=c^3q(y)$ for all scalars c. Hence, twice the negative of the logarithm of the density of

$\hat{\xi}_0 + n^{-1}\psi$ is

$$ny^T y - 2nq(y) + s\log(2\pi/n) + O(n^{-1}).$$

Taking $y = \hat{\xi}(\mu_1)$ and noting (3.22) for $l_\theta(\mu_1)$, we obtain (3.1) [and (3.25)].

4. Location adjustment.

4.1. *Introduction and summary.* We showed in Section 3 that the empirical likelihood region is close to the true likelihood region based on the statistic

$$\hat{\xi}_0 \equiv \left(Q^{1/2}\hat{Q}^{-1}Q^{1/2}\right)^{1/2} Q^{-1/2}\left(\hat{\theta} - \theta_0\right),$$

except that the empirical likelihood region is centered incorrectly. In the present section we shall prove that it is efficacious to empirically recenter the empirical likelihood region. Section 4.2 will demonstrate that an appropriately adjusted region is second-order correct for the "true" likelihood region based on the distribution of $\hat{\xi}_0$. Section 4.3 will show that Bartlett correction applies to the adjusted region in much the same manner that it is used for the nonadjusted region. This enables coverage error to be reduced by an order of magnitude, by applying a scale correction.

We now describe the location adjustment. Let

$$\mathscr{R}(x) \equiv \{\theta(\mu_1) \colon l_\theta(\mu_1) \le x\}$$

be the nonadjusted empirical likelihood region. Recall from Section 3 that, to second order, empirical likelihood draws contours of the likelihood of $\hat{\xi}_0 + n^{-1}\psi$, where ψ is given by (3.15). If we replace unknowns by sample estimates in the formula for ψ, we obtain an estimate $\hat{\psi}$ satisfying $\hat{\psi} = \psi + O_p(n^{-1/2})$. Hence, $\hat{v} \equiv n^{-1}\hat{Q}^{1/2}\hat{\psi}$ has the property

$$\hat{v} = n^{-1}Q^{1/2}\psi + O_p(n^{-3/2}). \tag{4.1}$$

Throughout this section we take $\hat{v}$ to be a general s-vector satisfying (4.1). The location-adjusted empirical likelihood region is

$$\mathscr{R}_A(x) \equiv \mathscr{R}(x) + \hat{v} = \{\theta(\mu_1) + \hat{v} \colon l_\theta(\mu_1) \le x\}.$$

Define $H \equiv (Q^{1/2}\hat{Q}^{-1}Q^{1/2})^{1/2}Q^{-1/2}$.

4.2. *Second-order correctness of* $\mathscr{R}_A$. Let h denote the density of $n^{1/2}\hat{\xi}_0 = n^{1/2}\hat{H}(\hat{\theta} - \theta_0)$. The "true" likelihood-based confidence region for θ_0, founded on the distribution of $\hat{\xi}_0$, is

$$\mathscr{R}_T(x) \equiv \left\{\theta_1 \colon -2\log\left[(2\pi)^{s/2}h\left\{n^{1/2}\hat{H}\left(\hat{\theta} - \theta_1\right)\right\}\right] \le x\right\}. \tag{4.2}$$

In the present subsection we show that the boundary of this region is $O_p(n^{-3/2})$ away from that of the adjusted empirical likelihood region $\mathscr{R}_A(x)$.

Let $g(y) \equiv h(y - n^{-1/2}\psi)$ denote the density of $n^{1/2}\hat{\xi}_0 + n^{-1/2}\psi$. We showed in Section 3 that

$$\begin{aligned}\mathscr{R}(x) &\equiv \{\theta(\mu_1): l_\theta(\mu_1) \le x\} \\ &= \left\{\theta_1: -2\log\left[(2\pi)^{s/2} g\left\{n^{1/2}\hat{H}(\hat{\theta} - \theta_1)\right\}\right] + O_p(n^{-1}) \le x\right\} \\ &= \left\{\theta_1: -2\log\left[(2\pi)^{s/2} h\left\{n^{1/2}\hat{H}(\hat{\theta} - \theta_1) - n^{-1/2}\psi\right\}\right] + O_p(n^{-1}) \le x\right\} \\ &= \left\{\theta_1 - n^{-1}\hat{H}^{-1}\psi: -2\log\left[(2\pi)^{s/2} h\left\{n^{1/2}\hat{H}(\hat{\theta} - \theta_1)\right\}\right] + O_p(n^{-1}) \le x\right\}.\end{aligned}$$

Therefore, since

$$\hat{v} = n^{-1}Q^{1/2}\psi + O_p(n^{-3/2}) = n^{-1}\hat{H}^{-1}\psi + O_p(n^{-3/2}),$$

the boundary of $\mathscr{R}_A(x) \equiv \mathscr{R}(x) + \hat{v}$ is $O_p(n^{-3/2})$ away from that of the region

$$\left\{\theta_1: -2\log\left[(2\pi)^{s/2} h\left\{n^{1/2}\hat{H}(\hat{\theta} - \theta_1)\right\}\right] + O_p(n^{-1}) \le x\right\}.$$

The boundary of this region is $O_p(n^{-3/2})$ from that of $\mathscr{R}_T(x)$, the latter defined at (4.2).

4.3. *Effect of location adjustment on Bartlett correction.* We begin by describing Bartlett correction for the nonadjusted region, $\mathscr{R}(x)$. Let b be the constant defined by

$$E\{l_\theta(\mu_0)\} = s(1 + n^{-1}b) + O(n^{-2}),$$

and let $\hat{b}$ be a $\sqrt{n}$-consistent estimator of b obtained (for example) by replacing unknowns by sample estimates in the formula for b. The Bartlett-corrected confidence region $\mathscr{R}_B(x)$ is

$$\mathscr{R}_B(x) \equiv \left\{\theta(\mu_1): l_\theta(\mu_1) \le x(1 + n^{-1}\hat{b})\right\}.$$

Thus, Bartlett correction amounts to a scale correction applied to empirical likelihood. Now, the uncorrected region $\mathscr{R}(x)$ has nominal coverage $P(\chi_s^2 \le x)$ and coverage error $O(n^{-1})$, in the sense that

$$P\{\theta_0 \in \mathscr{R}(x)\} = P\left(\chi_s^2 \le x\right) + O(n^{-1}).$$

Bartlett correction reduces coverage error to $O(n^{-2})$:

$$P\{\theta_0 \in \mathscr{R}_B(x)\} = P\left(\chi_s^2 \le x\right) + O(n^{-2}).$$

See DiCiccio, Hall and Romano (1988b).

Our purpose in the present subsection is to show that Bartlett correction works just as well for location-adjusted regions $\mathscr{R}_A(x)$, albeit with a different constant b. The proof of this result is most easily obtained by working with the signed root log-likelihood ratio, which is an s-vector $s_\theta(\mu_1)$ having the property

$$s_\theta(\mu_1)^T s_\theta(\mu_1) = l_\theta(\mu_1).$$

We may deduce from (3.14) and (3.22) that

$$n^{-1/2}s_\theta^u(\mu_1) = \hat{\xi}^u + \tfrac{1}{6}\lambda^{uvw}\hat{\xi}^v\hat{\xi}^w + O_p(n^{-3/2}),$$

where λ^{uvw} is defined at (3.22). Longer expansions are given by DiCiccio, Hall and Romano (1988b).

Let μ_0 denote the true value of μ. Define $\nabla s_\theta(\mu_0)$ to be the $s \times r$ matrix whose (u, j)th element is

$$\partial s_\theta^u/\partial \mu^j|_{\mu=\mu_0}.$$

Let b_A be the constant defined by

$$E\{s_\theta(\mu_0) - n^{1/2}\nabla s_\theta(\mu_0)\Theta^T Q^{-1}\hat{v}\}^T\{s_\theta(\mu_0) - n^{1/2}\nabla s_\theta(\mu_0)\Theta^T Q^{-1}\hat{v}\}$$
$$= s(1 + n^{-1}b_A) + O(n^{-2}).$$

Write $\hat{b}_A$ for a $\sqrt{n}$-consistent estimate of b_A, and let

$$\mathscr{R}_{AB}(x) \equiv \{\theta(\mu_1) + \hat{v}: l_\theta(\mu_1) \le x(1 + n^{-1}\hat{b}_A)\}$$

denote the location-adjusted, Bartlett-corrected confidence region. We shall prove that $\mathscr{R}_{AB}(x)$ has nominal coverage $P(\chi_s^2 \le x)$ with error $O(n^{-3/2})$. A longer argument, using oddness and evenness properties of polynomials in Edgeworth expansions, will show that the error is actually $O(n^{-2})$. See Barndorff-Nielsen and Hall (1988).

Let $\tilde{\mu}_0 = \mu_0 - \Theta^T Q^{-1}\hat{v} + O_p(n^{-2})$ be a vector such that $\theta(\mu_0) = \theta(\tilde{\mu}_0) + \hat{v}$. Then the coverage probability of $\mathscr{R}_{AB}(x)$ equals

$$P\{\mu_0 \in \mathscr{R}_{AB}(x)\} = P\{l_\theta(\tilde{\mu}_0) \le x(1 + n^{-1}\hat{b}_A)\}.$$

Now, $l_\theta(\tilde{\mu}_0) = s_\theta(\tilde{\mu}_0)^T s_\theta(\tilde{\mu}_0)$, and since $\hat{v} = O_p(n^{-1})$,

$$s_\theta(\tilde{\mu}_0) = s_\theta(\mu_0) + \nabla s_\theta(\mu_0)(\tilde{\mu}_0 - \mu_0) + O_p(n^{-3/2})$$
$$= s_\theta(\mu_0) - \nabla s_\theta(\mu_0)\Theta^T Q^{-1}\hat{v} + O_p(n^{-3/2}).$$

In consequence, defining

$$S \equiv s_\theta(\mu_0) - \nabla s_\theta(\mu_0)\Theta^T Q^{-1}\hat{v},$$

we have

$$P\{\mu_0 \in \mathscr{R}_{AB}(x)\} = P\{S^T S \le x(1 + n^{-1}\hat{b}_A)\} + O(n^{-3/2})$$
$$= P\{S^T S \le x(1 + n^{-1}b_A)\} + O(n^{-3/2})$$
$$= P\{S^T S \le xE(S^T S)/s\} + O(n^{-3/2}).$$

To prove that

$$P\{S^T S \le xE(S^T S)/s\} = P(\chi_s^2 \le x) + O(n^{-3/2}),$$

it suffices to show that all fourth-order cumulants of the s-vector S equal $O(n^{-3/2})$. [To appreciate why the latter result is sufficient, see Section 2 of Barndorff-Nielsen and Cox (1984).] DiCiccio, Hall and Romano (1988b) demonstrate that all fourth-order cumulants of $s_\theta(\mu_0)$ are of order n^{-2}. To extend this result to S, note that in summation notation,

$$S = s_\theta(\mu_0) + n^{-1/2}\alpha + n^{-1/2}\beta_j \bar{Z}^j + O_p(n^{-3/2}),$$

where α, β_j are fixed s-vectors and $\bar{Z}$ is a mean (of length depending on the number of unknowns estimated in $\hat{v}$) having zero expectation. Since fourth-order cumulants are location-invariant then fourth-order cumulants of S are the same as their counterparts for

$$T \equiv s_\theta(\mu_0) + n^{-1/2}\beta_j \bar{Z}^j,$$

up to terms of order $n^{-3/2}$ (actually, order n^{-2}).

The fourth-order cumulant of T of type (u_1, u_2, u_3, u_4) is

$$\begin{aligned}\kappa^{u_1u_2u_3u_4}(T) = {} & E(T^{u_1}T^{u_2}T^{u_3}T^{u_4}) - E(T^{u_1}T^{u_2})E(T^{u_3}T^{u_4})[3] \\ & - E(T^{u_1})E(T^{u_2}T^{u_3}T^{u_4})[4] + 2E(T^{u_1})E(T^{u_2})E(T^{u_3}T^{u_4})[6] \\ & - 6E(T^{u_1})E(T^{u_2})E(T^{u_3})E(T^{u_4}),\end{aligned}$$

where a digit in square brackets indicates the number of times the previous term should be included for distinct rearrangements of its superscripts. Write $T = U + V$ where

$$U \equiv s_\theta(\mu_0), \qquad V \equiv n^{-1/2}\beta_j \bar{Z}^j.$$

Since $V = O_p(n^{-1})$, $E(V) = 0$ and $E(U) = O(n^{-1/2})$, then

$$\begin{aligned}\kappa^{u_1u_2u_3u_4}(T) = {} & \kappa^{u_1u_2u_3u_4}(U) + E(U^{u_1}U^{u_2}U^{u_3}V^{u_4})[4] \\ & - E(U^{u_1}U^{u_2})E(U^{u_3}V^{u_4})[12] + O(n^{-3/2}).\end{aligned} \tag{4.3}$$

Let γ_j be s-vectors such that $U = n^{1/2}\gamma_j \bar{Z}^j + O_p(n^{-1/2})$; if necessary, lengthen the vector $\bar{Z}$ to ensure this representation. Then

$$\begin{aligned} & E(U^{u_1}U^{u_2}U^{u_3}V^{u_4})[4] - E(U^{u_1}U^{u_2})E(U^{u_3}V^{u_4})[12] \\ & \quad = n\gamma_j^{u_1}\gamma_k^{u_2}\gamma_l^{u_3}\beta_m^{u_4}[4]\{E(\bar{Z}^j\bar{Z}^k\bar{Z}^l\bar{Z}^m) \\ & \qquad\qquad - E(\bar{Z}^j\bar{Z}^k)E(\bar{Z}^l\bar{Z}^m)\}[3] + O(n^{-3/2}) \\ & \quad = O(n^{-3/2}),\end{aligned} \tag{4.4}$$

since

$$E(\bar{Z}^j\bar{Z}^k\bar{Z}^l\bar{Z}^m) = n^{-3}(n-1)E(Z^jZ^k)E(Z^lZ^m)[3] + n^{-3}E(Z^jZ^kZ^lZ^m)$$

and $E(\bar{Z}^j\bar{Z}^k) = n^{-1}E(Z^jZ^k)$. Combining results (4.3) and (4.4), and remembering that

$$\kappa^{u_1u_2u_3u_4}(U) = O(n^{-2}),$$

we deduce that

$$\kappa^{u_1u_2u_3u_4}(T) = O(n^{-3/2}),$$

as had to be proved.

5. Relations between $\hat{\xi}_0$ and $\hat{\eta}_0$.

5.1. *Introduction and summary.* Recall from Section 2 that

$$\hat{\xi}_0 \equiv \left(Q^{1/2}\hat{Q}^{-1}Q^{1/2}\right)^{1/2}Q^{-1/2}\left(\hat{\theta} - \theta_0\right) \quad \text{and} \quad \hat{\eta}_0 \equiv \hat{Q}^{-1/2}\left(\hat{\theta} - \theta_0\right).$$

Since 1×1 matrices commute, then $\hat{\xi}_0$ and $\hat{\eta}_0$ are identical in the case of $s = 1$ dimension. However, in $s \geq 2$ dimensions the question of relationship between $\hat{\xi}_0$ and $\hat{\eta}_0$ is more delicate. Our purpose in the present section is to describe this relationship.

First, we point out similarities between $\hat{\xi}_0$ and $\hat{\eta}_0$. Observe that $\hat{\xi}_0$ and $\hat{\eta}_0$ both equal $Q^{-1/2}\Theta(\overline{X} - \mu_0) + O_p(n^{-1})$, and that $n^{1/2}Q^{-1/2}\Theta(\overline{X} - \mu_0)$ is asymptotically s-variate normal $N(0, I)$. Therefore $\hat{\xi}_0$ and $\hat{\eta}_0$ always agree to first order. The *distributions* of $\hat{\xi}_0 - E(\hat{\xi}_0)$ and $\hat{\eta}_0 - E(\hat{\eta}_0)$ agree to second order. Indeed, we shall prove in Section 5.2 that Edgeworth expansions of the densities of $n^{1/2}(\hat{\xi}_0 - E\hat{\xi}_0)$ and $n^{1/2}(\hat{\eta}_0 - E\hat{\eta}_0)$ are identical up to and including terms of order $n^{-1/2}$. In consequence, the expansion displayed at (3.2) applies equally to both densities.

Next, we indicate the differences between $\hat{\xi}_0$ and $\hat{\eta}_0$. Here it is convenient to focus on a special case, and we choose that where $r = s = 2$ and $\theta(\mu) \equiv \mu$. For this circumstance we shall derive in Section 5.3 an approximation to $\hat{Q}^{-1/2}$. Section 5.4 will apply that result to prove that the differences $E(\hat{\xi}_0 - \hat{\eta}_0)$ and $\hat{\xi}_0 - \hat{\eta}_0 - E(\hat{\xi}_0 - \hat{\eta}_0)$ are both genuinely of order n^{-1}. That is to say, both quantities differ in terms of second order, despite the fact that distributions of $\hat{\xi}_0 - E(\hat{\xi}_0)$ and $\hat{\eta}_0 - E(\hat{\eta}_0)$ agree to second order.

Bootstrap likelihood confidence regions are usually based on $\hat{\eta}_0$, and we know from Sections 3 and 4 that empirical likelihood regions are in effect based on $\hat{\xi}_0$. In view of the second-order differences between $\hat{\xi}_0$ and $\hat{\eta}_0$ there will often be, in $s \geq 2$ dimensions and for a given sample, second-order differences between bootstrap likelihood and empirical likelihood confidence regions, even when the latter are adjusted for location.

5.2. *Second-order agreement between distributions of $\hat{\xi}_0 - E(\hat{\xi}_0)$ and $\hat{\eta}_0 - E(\hat{\eta}_0)$.* In this subsection we show that distributions of $\hat{\xi}_0 - E(\hat{\xi}_0)$ and $\hat{\eta}_0 - E(\hat{\eta}_0)$ agree in terms of second order. Now, both these centered vectors have zero first-order cumulants, and both have variance matrices expressible as $n^{-1}I + O(n^{-2})$. Therefore, our claim will be verified if we show that third-order cumulants of $\hat{\xi}_0$ and $\hat{\eta}_0$ are identical up to and including terms of order n^{-2}.

Recall from (3.6) and (3.7) that

$$\hat{Q} = Q + \Delta + O_p(n^{-1}),$$

where Δ is an $s \times s$ matrix satisfying $\Delta = O_p(n^{-1/2})$. Therefore

$$\left(Q^{1/2}\hat{Q}^{-1}Q^{1/2}\right)^{1/2}Q^{-1/2} = Q^{-1/2} - \tfrac{1}{2}Q^{-1/2}\Delta Q^{-1} + O_p(n^{-1})$$

and

$$\hat{Q}^{-1/2} = Q^{-1/2} + \varepsilon + O_p(n^{-1}),$$

where we may take ε to be the unique $s \times s$ symmetric matrix satisfying

$$Q^{1/2}\varepsilon + \varepsilon Q^{1/2} = -Q^{-1/2}\Delta Q^{-1/2}.$$

Put

$$W \equiv \tfrac{1}{2}(\varepsilon - Q^{1/2}\varepsilon Q^{-1/2})\Theta A.$$

It may be shown that third-order cumulants of $\hat{\eta}_0$ and $\hat{\xi}_0$ are related by the formula

$$\begin{aligned} &E\{(\hat{\eta}_0 - E\hat{\eta}_0)^u(\hat{\eta}_0 - E\hat{\eta}_0)^v(\hat{\eta}_0 - E\hat{\eta}_0)^w\} \\ &\quad = E\{(\hat{\xi}_0 - E\hat{\xi}_0)^u(\hat{\xi}_0 - E\hat{\xi}_0)^v(\hat{\xi}_0 - E\hat{\xi}_0)^w\} \\ &\qquad + E\{(Q^{-1/2}\Theta A)^u(Q^{-1/2}\Theta A)^v(W - EW)^w\}[3] + O(n^{-3}). \end{aligned}$$

Hence, our claim follows from

$$E\{(Q^{-1/2}\Theta A)^u(Q^{-1/2}\Theta A)^v(W - EW)^w\}[3] = O(n^{-3}),$$

which may be demonstrated by routine algebra.

5.3. *Approximation to* $\hat{Q}^{-1/2}$. In this subsection and the next we take $r = s = 2$. Given a, b, c, α, β, γ, put

$$\begin{aligned} u &= (b^4 + c^4 + b^2c^2 + ab^3)\alpha + c^2(a^2 + b^2 - c^2 + 3ab)\beta \\ &\quad - 2\{b^2c(2a + b) + ac^3\}\gamma, \\ v &= c^2(a^2 + b^2 - c^2 + 3ab)\alpha + (a^4 + c^4 + a^2c^2 + a^3b)\beta \\ &\quad - 2\{a^2c(a + 2b) + bc^3\}\gamma, \\ w &= -c(ac^2 + 2ab^2 + b^3)\alpha - c(bc^2 + 2a^2b + a^3)\beta \\ &\quad + 2\{a^2b^2 + c^2(a^2 + b^2 + ab)\}\gamma, \end{aligned}$$

$$P = \begin{bmatrix} a & c \\ c & b \end{bmatrix}, \quad \Delta = \begin{bmatrix} \alpha & \gamma \\ \gamma & \beta \end{bmatrix}, \tag{5.1}$$

$$\varepsilon = -\tfrac{1}{2}(ab - c^2)^{-3}(a + b)^{-1}\begin{bmatrix} u & w \\ w & v \end{bmatrix},$$

$Q = P^2$. If Δ is a 2×2 matrix of order $n^{-1/2}$ such that $\hat{Q} = Q + \Delta + O_p(n^{-1})$, then it may be shown after tedious algebra that $\hat{Q}^{-1/2} = Q^{-1/2} + \varepsilon + O_p(n^{-1})$.

5.4. *Second-order differences between* $\hat{\xi}_0$ *and* $\hat{\eta}_0$. In this subsection we derive a formula for $\hat{\xi}_0 - \hat{\eta}_0$ and use it to show that $E(\hat{\xi}_0 - \hat{\eta}_0)$ and $\hat{\xi}_0 - \hat{\eta}_0 - E(\hat{\xi}_0 - \hat{\eta}_0)$ contain terms of order n^{-1}.

Assume for the sake of definiteness that $r = s = 2$, $\theta(\mu) \equiv \mu$, the true mean of X is $\mu_0 = 0$ and the variance of X is $\Sigma = Q = P^2$, where P is as in Section 5.3

with $a = 2$, $b = c = 1$. Then $\hat{Q} = Q + \Delta + O_p(n^{-1})$, where Δ is given at (5.1) with

$$\alpha \equiv n^{-1} \sum_i \left(X_i^1 X_i^1 - 2 \right), \qquad \beta \equiv n^{-1} \sum_i \left(X_i^2 X_i^2 - 1 \right)$$

and

$$Y \equiv n^{-1} \sum_i \left(X_i^1 X_i^2 - 1 \right).$$

We may deduce from Section 5.3 that, in the notation of (5.1),

$$u = 5\alpha + 10\beta - 14\gamma, \qquad v = 10\alpha + 29\beta - 34\gamma \quad \text{and} \quad w = -7\alpha - 17\beta + 22\gamma.$$

From this we may deduce that

$$\hat{\xi}_0 - \hat{\eta}_0 = \left\{ \left(Q^{1/2} \hat{Q}^{-1} Q^{1/2} \right)^{1/2} Q^{-1/2} - \hat{Q}^{-1/2} \right\} \overline{X}$$

$$= \frac{1}{6} \begin{bmatrix} -\alpha + \beta + \gamma & 2\alpha - 8\beta + 4\gamma \\ -\alpha + \beta + \gamma & \alpha + 11\beta - 7\gamma \end{bmatrix} \begin{bmatrix} \overline{X}^1 \\ \overline{X}^2 \end{bmatrix} + O_p\left(n^{-3/2}\right).$$

Since each of $\overline{X}^1$, $\overline{X}^2$, α, β and γ is of precise order $n^{-1/2}$ with zero mean, and since $nE(\alpha \overline{X}^1) = \alpha^{111}$, $nE(\beta \overline{X}^2) = \alpha^{222}$, $nE(\beta \overline{X}^1) = nE(\gamma \overline{X}^2) = \alpha^{122}$ and $nE(\gamma \overline{X}^1) = nE(\alpha \overline{X}^2) = \alpha^{112}$, then in general $E(\hat{\xi}_0 - \hat{\eta}_0)$ is asymptotic to a nonzero vector multiplied by n^{-1}, and $\hat{\xi}_0 - \hat{\eta}_0 - E(\hat{\xi}_0 - \hat{\eta}_0)$ is of precise order n^{-1}.

Acknowledgments. My ideas on empirical likelihood developed during many stimulating conversations with Tom DiCiccio, Art Owen and Joe Romano. I am grateful to all three for their time. The paper has also benefited from constructive suggestions by two referees and the Associate Editor.

REFERENCES

Barndorff-Nielsen, O. E. and Cox, D. R. (1984). Bartlett adjustments to the likelihood ratio statistic and the distribution of the maximum likelihood estimator. *J. Roy. Statist. Soc. Ser. B* **46** 483–495.

Barndorff-Nielsen, O. E. and Hall, P. (1988). On the level-error after Bartlett adjustment of the likelihood ratio statistic. *Biometrika* **75** 374–378.

DiCiccio, T. J., Hall, P. and Romano, J. P. (1988a). Comparison of parametric and empirical likelihood functions. *Biometrika*, to appear.

DiCiccio, T. J., Hall, P. and Romano, J. P. (1988b). Bartlett correction for empirical likelihood. Unpublished.

Hall, P. (1987). On the bootstrap and likelihood-based confidence regions. *Biometrika* **74** 481–493.

James, G. S. (1955). Cumulants of a transformed variable. *Biometrika* **42** 529–531.

James, G. S. (1958). On moments and cumulants of systems of statistics. *Sankhyā* **20** 1–30.

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

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

Department of Statistics
Australian National University
G.P.O. Box 4, Canberra, A.C.T. 2601
Australia