REPORT DOCUMENTATION PAGE

Form Approved
OMB NO. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
December 1996

3. REPORT TYPE AND DATES COVERED
Technical

4. TITLE AND SUBTITLE
Likelihood Computations Without Bartlett Identities

5. FUNDING NUMBERS
DAAH04-95-1-0105

6. AUTHOR(S)
Per Aslak Mykland

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(ES)
University of Chicago
Chicago, Illinois 60637

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park,, NC 27709-2211

10. SPONSORING / MONITORING AGENCY REPORT NUMBER
ARO 31383.7-MA

11. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12 b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

ABSTRACT IN SUMMARY

19970210 248

14. SUBJECT TERMS

15. NUMBER IF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OR REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18
298-102

Likelihood Computations
without Bartlett Identities *

by

Per Aslak Mykland

TECHNICAL REPORT NO. 441

Department of Statistics
The University of Chicago
Chicago, Illinois 60637

December 1996

* This research was supported in part by National Science Foundation grant DMS 96-26266 and Army Research Office grant DAAH04-95-1-0105.

This manuscript was prepared using computer facilities supported in part by the National Science Foundation grants DMS 89-05292, and DMS 87-03942 awarded to the Department of Statistics at The University of Chicago, and by The University of Chicago Block Fund.

Some key words and phrases: Bartlett correction, Convergence of cumulants, Unconditional accuracy.

Likelihood Computations
without Bartlett Identities *

by

Per Aslak Mykland
The University of Chicago

Summary

The signed square root statistic R typically has cumulants on the form $\mathrm{cum}_p(R) = \delta_{2,p} + n^{-\frac{p}{2}} k_p + O(n^{-\frac{p+2}{2}})$. This paper shows how to compute k_p without invoking the Bartlett identities. As an application, we show how the family of alternatives influences the coverage accuracy of R, and in particular that a bad choice of family can lead to arbitrary undercoverage for confidence intervals based on R.

* This research was supported in part by National Science Foundation grant DMS 96-26266 and Army Research Office grant DAAH04-95-1-0105.

This manuscript was prepared using computer facilities supported in part by the National Science Foundation grants DMS 89-05292, and DMS 87-03942 awarded to the Department of Statistics at The University of Chicago, and by The University of Chicago Block Fund.

Some key words and phrases: Bartlett correction, Convergence of cumulants, Unconditional accuracy.

1. Introduction

The Bartlett identities are one of the most powerful tools available in likelihood theory. The general set of identities go back to Bartlett (1953 a, b), though earlier versions exist in the form of the so-called Wald identities. The first two identites were crucial in the early development of likelihood theory – Fisher, Neyman and Pearson, Cramer and Rao. Generalizations of the identites can be found on Skovgaard (1986), McCullagh (1987), and Mykland (1994, 1995b), among others.

Since their publication, these identities have been used for a variety of purposes, most notably the analysis of the higher order asymptotic behaviour of the likelihood ratio statistic and its square root R. This literature starts with Lawley (1956); contemporary research includes McCullagh (1984, 1987), McCullagh and Tibshirani (1990), DiCiccio and Romano (1989), DiCiccio, Hall and Romano (1991) and Andrews and Stafford (1993), to mention some. (There is also the parallel research done with saddlepoint type methods, as in Barndorff-Nielsen and Cox (1979, 1984) and Barndorff-Nielsen (1983, 1986, 1991), and the papers by Barndorff-Nielsen and Wood (1995), Jensen (1992, 1995, 1997) and Skovgaard (1990, 1996) discussed below). In addition to aiding the computation of coefficients in expansions, the identities can also be used to establish asymptotic normality and the existence of asymptotic expansions (as in the case of martingales in Mykland (1994, 1995b)).

There is, however, also a dark side to these identities. Computations are often long and tedious, and the answers can be hard to verify. This can to some extend be remedied with symbolic computation, but writing such programs is also no simple task. We believe that this is the general experience of research workers in this area, and we certainly speak from painful personal experience, having spent a month in the Spring of 1995 to show that $\mathrm{cum}_5(R) = O(n^{-\frac{5}{2}})$ using Taylor expansions and Bartlett identities (to laugh and/or cry with the author, have a look at the (unpublished) technical report no. 411 of the Department of Statistics at the University of Chicago).

As demonstrated in Mykland (1996), however, one can circumvent these identities to show that, in fact, $\mathrm{cum}_p(R) = O(n^{-\frac{p}{2}})$. This is done by deriving cumulant behaviour from the large deviation results of Barndorff-Nielsen and Wood (1995), Jensen (1992, 1995, 1997) and Skovgaard (1990, 1996).

The purpose of this paper is to show that one can take this further – that these large deviation techniques not only help for orders of convergence, they also help the computation of coefficients. Specifically, it will normally be the case that

$$\mathrm{cum}_p(R) = \delta_{2,p} + n^{-\frac{p}{2}} k_p + O(n^{-\frac{p+2}{2}}) \tag{1.1}$$

(see Wallace (1958), Bhattacharya and Ghosh (1978), Hall (1992), and Mykland (1996)), and we shall show how to find k_p. Similar methods seem capable of yielding higher order terms in the expansions of the form (1.1).

In the following, we first discuss how Edgeworth and large deviation expansions hang together, and we state a result which gives the form of the generating function of the k_p's. k_3 and k_4 are given explicitly (k_1 and k_2 are previously known). A more rigorous development involving curved exponential families is given in Section 4. Meanwhile, in Section 3, we show how these results can be used to analyze the effect of the alternative on the null distribution of R, and how this affects the difference between nominal and actual coverage of confidence intervals.

It should be emphasized that we are not completely doing without the Bartlett identities. The coefficients come up again in the formulae we derive, see, *e.g.*, (2.9), and we invoke the identities themselves in Section 3. The Bartlett identities remain a powerful presence in likelihood theory, even if one can sometimes do without them.

2. The Main Formula

If an asymptotically normal statistic has density f_n and cumulant generating function K_n, the saddlepoint approximation has the form

$$f_n(n^{\frac{1}{2}}h) \;=\; \frac{1}{(2\pi K_n''(\hat{\tau}_n))^{\frac{1}{2}}} \exp(K_n(\hat{\tau}_n) - \hat{\tau}_n K_n'(\hat{\tau}_n))(1+o(1)), \tag{2.1}$$

where $K_n'(\hat{\tau}_n) = n^{\frac{1}{2}}h$. This goes back to Daniels (1952), see also Theorem 1 of Chaganty and Sethuraman (1985) and Theorem 1 of Mykland (1996). If we are dealing with the signed square root statistic R_n, whose cumulants are of the form (1.1), it is easy to Taylor expand (2.1) to get the following. First of all,

$$\xi(h) \;=\; \lim_{n\to\infty} \frac{f_n(n^{\frac{1}{2}}h)}{\phi(n^{\frac{1}{2}}h)} \tag{2.2}$$

exists – ϕ being the standard normal density. Also,

$$\xi(h) \;=\; \exp\left\{k_1 h + \frac{1}{2}k_2 h^2 + \frac{1}{3!}k_3 h^3 + \cdots\right\}. \tag{2.3}$$

Following the development in Section 4, we get the following formula for ξ.

THEOREM 1. *Let*

$$\widetilde{\ell}(\beta) = \lim\left(\ell_n(\beta) - \ell_n(\beta_0)\right)/n \tag{2.4}$$

and

$$J(\beta) = -\lim \ddot{\ell}_n(\beta)/n \tag{2.5}$$

where the limits are in probability under P_β. Set

$$h(\beta) = \sqrt{2} \quad \mathrm{sign}(\beta - \beta_0)\widetilde{\ell}(\beta)^{1/2}. \tag{2.6}$$

Then ξ (under P_{β_0}) is given by

$$\xi:\; h \rightarrow J^{1/2}\frac{\partial\beta}{\partial h}. \tag{2.7}$$

∎

In Appendix 1 we show that

$$\widetilde{\ell}(\beta) \approx \frac{1}{n}\sum_{p\geq 2}\frac{1}{p!}(\beta-\beta_0)^p \sum kb(q_1,...,q_k)\mathrm{cum}(\underbrace{\ell^{(1)},\ldots,\ell^{(1)}}_{q_1 \text{ times}},\ldots,\underbrace{\ell^{(k)},\ldots,\ell^{(k)}}_{q_k \text{ times}},\ldots), \tag{2.8}$$

where the second sum is over all $q_1, q_2, \ldots$ so that $q_1 + 2q_2 + \ldots + kq_k + \ldots = q$, and where the bs are the coefficients in the Bartlett identities, *i.e.*,

$$b(q_1,...,q_k) = \frac{q!}{\prod (k!)^{q_k} q_k!} \tag{2.9}$$

(see, *e.g.*, p. 159 of Barndorff-Nielsen and Cox (1989)). Similarly,

$$J(\beta) \approx -\frac{1}{n}\sum_{p\geq 2}\frac{1}{p!}(\beta-\beta_0)^p \sum \widetilde{b}(q_1,...,q_k)\mathrm{cum}(\underbrace{\ell^{(1)},\ldots,\ell^{(1)}}_{q_1 \text{ times}},\ldots,\underbrace{\ell^{(k)},\ldots,\ell^{(k)}}_{q_k \text{ times}},\ldots), \tag{2.10}$$

where

$$\widetilde{b}(q_1,\ldots,q_v) = b(q_1,\ldots,q_v)\frac{1}{\binom{p+2}{2}}\sum_{r=2}^{p+2}\binom{r}{2}q_r. \tag{2.11}$$

Finding the expression for the function (2.7), therefore, is purely a matter of inverting the function $h \to \beta$, and then plugging it into $J(\beta)^{1/2}$ and also differentiating it. This is easily done by symbolic manipulation software; we have used Maple (Cher *et al.* 1991) to get the expressions (2.12) and (2.13) below. k_1 and k_2 are all well-documented in the literature, see, e.g., McCullagh (1987), p. 214. Here. we therefore give

$$\begin{aligned} k_3 = c_{11}^{-9/2}\Big[&-c_{111}c_{11}c_{22} + \frac{17}{4}c_{111}c_{11}c_{112} + \frac{7}{4}\,c_{111}c_{11}c_{1111} \\ &- \frac{125}{72}\,c_{111}^3 + c_{11}^2c_{23} - \frac{1}{2}\,c_{11}^2c_{113} - \frac{3}{2}\,c_{11}^2c_{1112} \\ &-\frac{3}{10}\,c_{11}^2c_{11111}\Big] \end{aligned} \tag{2.12}$$

and

$$\begin{aligned} k_4 = c_{11}^{-6}\Big[&\frac{45}{4}\,c_{11}^2c_{112}c_{1111} - \frac{23}{3}\,c_{111}c_{11}^2c_{23} - \frac{9}{8}\,c_{11}^2c_{22}c_{1111} \\ &- \frac{45}{4}\,c_{11}^2c_{22}c_{112} + \frac{1465}{144}\,c_{111}^4 - \frac{9}{2}\,c_{11}^2c_{22}^2 \\ &+ \frac{45}{4}\,c_{11}^2c_{112}^2 + \frac{33}{14}\,c_{11}^2c_{1111}^2 - 6\,c_{11}^3c_{24} - 8\,c_{11}^3c_{114} \\ &- \frac{11}{3}\,c_{11}^3c_{33} - 12\,c_{11}^3c_{1113} - \frac{51}{2}\,c_{11}^3c_{1122} \\ &- \frac{21}{2}\,c_{11}^3c_{11112} - \frac{13}{2}\,c_{11}^3c_{222} - \frac{3}{4}\,c_{11}^3c_{111111} \\ &+ \frac{113}{24}\,c_{111}^2c_{11}c_{22} - \frac{455}{12}\,c_{111}^2c_{11}c_{112} - \frac{341}{24}\,c_{111}^2c_{11}c_{1111} \\ &+ \frac{19}{3}\,c_{111}c_{11}^2c_{113} + 3\,c_{111}c_{11}^2c_{122} + 16\; + c_{111}c_{11}^2c_{1112} \\ &+3\,c_{111}c_{11}^2c_{11111} - 8\,c_{11}^3c_{123}\Big]\,, \end{aligned} \tag{2.13}$$

where $c_{q_1\ldots q_r} \simeq \mathrm{cum}(\ell^{(q_1)},\ldots,\ell^{(q_r)})/n$, and where we have adopted the convention from McCullagh (1987) of using a parametrization where $c_{1q} = 0$ for $q \geq 2$.

Note that in i.i.d. problems, the form of $\widetilde{\ell}$ and J are particularly straightforward:

$$\widetilde{\ell}(\beta) = E(\ell_1(\beta) - \ell_1(\beta_0)) \exp(\ell_1(\beta) - \ell_1(\beta_0)) \tag{2.14}$$

and

$$J(\beta) = -E\ddot{\ell}_1(\beta) \exp(\ell_1(\beta) - \ell_1(\beta_0)). \tag{2.15}$$

Finally, observe that the above viewpoint gives a new formula for an R^* statistic,

$$R^* = R + \frac{1}{R} \log \frac{U}{R}. \tag{2.16}$$

Various forms of U have been investigated, see, in particular Jensen (1992, 1997) and Skovgaard (1996). In view of the development in Jensen's papers, it is clear that (in curved exponential and analytic families), one can take

$$U = R/\xi(R/\sqrt{n}). \tag{2.17}$$

This has the right unconditional large deviation coverage up to $O(n^{-1})$, though, obviously, the conditional convergence properties are probably lost. This R^* is a function of R, and it can be seen as a large deviation version of Cornish-Fisher inversion.

3. The Accuracy of Confidence Intervals

One of the least studied phenomena of likelihood theory is the impact of the alternative on the coverage accuracy of confidence intervals. At first, this may seem like a contradiction in terms – coverage only concerns the behaviour of a statistic under the null hypothesis. The family of alternatives, however, sets up the likelihood function from which R is derived. Hence, different alternatives lead to different Rs, and hence to different behaviour under the null distribution.

From a traditional likelihood perspective, this may seem like a strange consideration, as the likelihood is determined by the actual family of alternatives. Recent years, however, have seen the increasing use of likelihoods that are designed to work under a multiplicity of null distributions, and such likelihoods need a pragmatic and sometimes deliberately wrong specification of the family of alternatives. Examples of this include the partial (Cox (1972, 1975), Wong (1986)), projective (McLeish and Small (1992) and dual (Mykland (1995a), Kong and Cox (1996)) likelihoods.

Since there are, therefore, several likelihoods that can go with the same alternative, it raises the question of how to compare them. The debate has been particularly acue in connection with empirical/dual likelihood, see Corcoran, Davison and Spady (1995) and Section 6 of Mykland (1996).

The debate has mainly been one the accuracy of possible procedures. This is because the dual and true likelihoods have the same power to first order in contiguous neighborhoods (Mykland (1995a), Section 5), and hence, typically, also to second order (Bickel, Chibisov, and van Zwet (1981)). And also because to third order, though the two do not have the same power, sometimes one does better, sometimes the other (Lazar and Mykland (1997)).

The formulae in the previous section permit us to characterize the impact of the alternative on accuracy. Consider the following setup: we are looking at log likelihoods ℓ having the same score $\dot{\ell}$, but where we can otherwise vary $\ddot{\ell}$, $\dddot{\ell}$, and so on, as we see fit. ℓ being a log likelihood implies that $\mathrm{var}(\dot{\ell}) + \mathrm{E}(\ddot{\ell}) = 0$, so first and second order efficiency only depends on $\dot{\ell}$. The only restriction we impose is that $\mathrm{cov}(\dot{\ell}, \ddot{\ell}) = \mathrm{cov}(\dot{\ell}, \ddot{\ell}) = \cdots = 0$ (as in McCullagh (1987), Chapter 7), since this can be done by a reparametrization which does not alter the statistic R.

The expansion for the density f_n of the signed square root R is

$$f_n(r) = \phi(r)\left\{1 + (n^{-\frac{1}{2}}k_1 + n^{-\frac{3}{2}}k_1')h_1(r) + \frac{1}{2!}\, n^{-1}k_2h_2(r) + \frac{1}{3!}\, n^{-\frac{3}{2}}k_3h_3(r) + O((r^4+1)n^{-2})\right\}, \tag{3.1}$$

where $h_q(r)$ is the q'th Hermite polynomial. (3.1) doubles as an Edgeworth and a large deviation expansion (cf. Chaganty and Sethuraman (1985)). In the same notation as (2.12)–(2.13), we have that

$$k_1 \;=\; -c_{11}^{-3/2}c_{111}/3!, \tag{3.2}$$

$$k_1' \;= \overline{k}_1' - c_{11}^{-2}\left[\frac{1}{4}c_{23} + \frac{1}{4}c_{1112} + \frac{1}{12}c_{113}\right] - c_{11}^{-3}c_{111}\left[\frac{17}{24}c_{112} + \frac{9}{16}c_{22}\right] \tag{3.3}$$

and

$$k_2 \;=\; c_{11}^{-2}\left[\frac{1}{4}c_{22} - \frac{1}{2}c_{112} - \frac{1}{4}c_{1111}\right] \;+\; \frac{7}{18}c_{11}^{-3}c_{111}^2, \tag{3.4}$$

where $\overline{k}_1'$ is the corresponding quantity for an exponential family with the same score. k_1 and k_2 comes from p. 214 of McCullagh (1987); k_1' is derived in Appendix 2. Note, incidentally, that the

k_q can depend on n to the extent that the c's do. One could also, obviously, expand the c's in orders of n, but that would only deepen the messiness of expressions.

It is clear from this that the convergence error at the $n^{-\frac{1}{2}}$ level is fixed by the score $\dot{\ell}$, the n^{-1} behaviour depends on the score and $\ddot{\ell}$, the $n^{-\frac{3}{2}}$ behaviour on $\dot{\ell}$, $\ddot{\ell}$ and $\dddot{\ell}$, and so on. In itself, not particularly surprising.

What is surprising, however, is that there is a radical difference between what can go wrong at the n^{-1} level and the $n^{-\frac{3}{2}}$ level. We shall argue below that a bad choice of $\ddot{\ell}$ can result in arbitrary overcoverage, but limited undercoverage. On the other hand, a bad choice of $\dddot{\ell}$ can also lead to unlimited undercoverage. The latter is, obviously, particularly dangerous.

The thing is, that k_2 is quadratic in $\ddot{\ell}$, with positive sign in front of the square term. Set $\ddot{\ell}_{lf} = [\dot{\ell}, \dot{\ell}] + a\dot{\ell} - 2\text{var}(\dot{\ell})$, where $[\dot{\ell}, \dot{\ell}]$ is the observed (optional) quadratic variation of $\dot{\ell}$, and where $a = -\text{cov}(\dot{\ell}, [\dot{\ell}, \dot{\ell}])/\text{var}(\dot{\ell})$. Suppose that $\ddot{\ell} = \ddot{\ell}_{lf} + m + R$, where m is a martingale orthogonal to $\dot{\ell}$, and R is $O_p(1)$ and asymptotically independent of $\dot{\ell}$, $[\dot{\ell}, \dot{\ell}]$ and m. This will be the case in most regular situations; the independent case is obvious; for Markov chains, see p. 448 of Jacod and Shiryaev (1987); for mixing sums, see Ch. 5 of Hall and Heyde (1980), or also Jacod and Shiryaev (1987). By the Bartlett identities for martingales (Mykland (1994),

$$\text{cov}(\ddot{\ell}_{lf}, m) = \text{cum}(\dot{\ell}, \dot{\ell}, m), \tag{3.5}$$

and hence

$$k_2 = k_{2,lf} + \frac{1}{4}c_{11}^{-2}\frac{1}{n}\text{var}(m) + o(1), \tag{3.6}$$

where $k_{2,lf}$ is the value of k_2 when $\ddot{\ell}_{lf}$ is taken as the second derivative of ℓ. Thus,

$$k_2 \geq k_{2,lf} + o(1), \tag{3.7}$$

establishing our claim about limited undercoverage at this level.

The coefficients in the $n^{-3/2}$ term, however, tell a different story. In both k_3 and k_1', $\dddot{\ell}$ enters linearly. If we focus on k_3, let $\dot{\ell}$ and $\ddot{\ell}$ be given, and consider a zero mean martingale m, orthogonal to $\dot{\ell}$, so that

$$\text{cov}(\ddot{\ell}, m) - \frac{1}{2}\text{cum}(\dot{\ell}, \dot{\ell}, m) = \nu n + o(n), \tag{3.8}$$

where $\nu \neq 0$. Replace the original $\dddot{\ell}$ by $\dddot{\ell}_\alpha = \dddot{\ell} + \alpha m$. The new $\dddot{\ell}_\alpha$ satisfies the third Bartlett identity (and is hence a valid third derivative of ℓ), and also $\operatorname{cov}(\dddot{\ell}_\alpha, \dot{\ell}) = 0$. In this setup,

$$k_{3,\alpha} = k_3 + \alpha c_{11}^{-5/2} \nu + o(1), \tag{3.9}$$

which can take on any value. In other words, both under- and overcoverage is potentially unbounded at this level.

4. Curved Exponential Families

For a more rigorous development, consider a curved exponential family

$$\ell_n(\beta) = \ell_n(\beta_0) + (\beta - \beta_0)\dot{\ell}_n(\beta_0) + \frac{1}{2}(\beta - \beta_1)^2 \ddot{\ell}(\beta_0) + \dots \tag{4.1}$$

of order p (i.e., terms of order $p+1$ and higher are nonrandom). We shall consider R for testing $H_0 : \beta = \beta_0$. Suppose that there is a valid saddlepoint approximation to the density of the vector $(\dot{\ell}_n(\beta), \dots, \ell_n^{(p)}(\beta))$. One can then proceed as follows.

Begin by fixing $\beta_1 \neq \beta_0$. Then reparametrize the family as in Section 7.2.3 (p. 204–207) of McCullagh (1987) to make $\operatorname{cov}_{\beta_1}(\dot{\ell}(\beta_1), \ell^{(q)}(\beta_1)) = 0$ for $2 \leq q \leq p$. It is clear from McCullagh that this is accomplished by using parameter ϕ, given by $\phi_1 = \beta$, and

$$\begin{aligned} \phi(\beta) - \phi_1 &= \beta - \beta_1 + \frac{1}{2}(\beta - \beta_1)^2 \frac{\operatorname{cov}_{\beta_1}(\dot{\ell}(\beta_1), \ddot{\ell}(\beta_1))}{\operatorname{var}_{\beta_1}(\dot{\ell}(\beta_1))} + \dots \\ &= \frac{E_{\beta_1}\dot{\ell}(\beta_1)(\ell(\beta) - \ell(\beta_1))}{\operatorname{var}_{\beta_1}(\dot{\ell}(\beta_1))}. \end{aligned} \tag{4.2}$$

Hence

$$\begin{aligned} \phi_0 - \phi_1 &= \phi(\beta_0) - \phi_1 \\ &= \frac{\frac{\partial}{\partial \beta} E_{\beta_0} g(\ell(\beta) - \ell(\beta_0))|_{\beta = \beta_1}}{E_{\beta_1}\ddot{\ell}(\beta_1)} \\ &\approx -\frac{\tilde{\dot{\ell}}(\beta_1)}{J(\beta_1)} \end{aligned} \tag{4.3}$$

as $n \to \infty$ under P_β. Here $g(x) = (x-1)e^x$, which can be replaced by $g(x) = xe^x$ since $E_{\beta_0} \exp(\ell(\beta_1) - \ell(\beta_0)) = 1$. In the new parametrization, the null hypothesis is $\phi = \phi_0$.

Now embed $\ell_n(\beta) - \ell_n(\beta_0)$ in a full exponential family. In the notation of Jensen (1997) (which we shall be using in the following), $\overline{T}_q = \ell^{(q)}/q!n$. Note that we do not require the $\overline{T}_q$'s to be means, only that the saddlepoint approximation hold.

Our larger family is then (in the new parametrization)

$$\ell_n(\phi_0) + \theta_1 \dot{\ell}_n(\phi_0) + \ldots + \theta_p \frac{1}{p!} \ell_n^{(p)}(\phi_0).$$

A reparametrization of the θ's is given by

$$\theta_1 = \phi$$

and

$$\theta_\ell = \phi b_\ell \quad \text{for} \quad \ell \geq 2 \tag{4.4}$$

(in Jensen's notation, ϕ is β_0 and b_ℓ is β_ℓ). A corresponding sequence of null hypotheses is

$$H_0^{(1)}: \ \phi = \phi_0$$

and

$$H_0^{(\ell)}: \ b_\ell = 1 \quad \text{for} \quad \ell \geq 2. \tag{4.5}$$

Hence, $H_0^{(1)}$ is our original null hypothesis.

Let the $\overline{u}_\ell$'s be chosen as in Section 3 of Jensen (1997). In view of Section 2 of the same paper, the joint density of $(R_1, R_{L,2}, \ldots, R_{L,p})$ is, in a large deviation region,

$$\frac{1}{(2\pi)^{p/2}} \frac{h_1}{\overline{u}_1} \exp\left(-\frac{1}{2} r_1^2 - \frac{1}{2} \sum_{i=2}^{p} r_{\ell,i}^2\right) \left\{1 + O(n^{-1})\right\}. \tag{4.6}$$

Note that $R_1 = R$. By using Skorokhod embedding, it therefore follows that, under P_{β_1},

$$\frac{f_{\beta_0}(R \mid R_{L,2}, \ldots, R_{L,p})}{\phi(R)} = \frac{R_1}{\overline{U}_1} \ \left\{1 + O_p(n^{-1})\right\}. \tag{4.7}$$

Clearly, $R_1/\sqrt{n} = h(\beta_1)(1+O_p(n^{-1/2}))$, where h is given in (2.6). Hence, if we can show that

$$\overline{U}_1/\sqrt{n} = (\phi_1 - \phi_0)J(\beta_1)^{1/2}\left\{1+O_p(n^{-1/2})\right\}, \tag{4.8}$$

it follows from (4.3), and by averaging over $(R_{L,2},\ldots,R_{L,p})$, that

$$\begin{aligned}\frac{f_{\beta_0}(R)}{\phi(R)} &= \frac{J(\beta_1)^{1/2}h(\beta_1)}{\dot{\tilde{\ell}}(\beta_1)}\left\{1+O_p(n^{-1/2})\right\} \\ &= J^{1/2}(\beta_1)\frac{\partial\beta}{\partial h}(h)\left\{1+O_p(n^{-1/2})\right\}\end{aligned} \tag{4.9}$$

under P_{β_1}. Again by Skorokhod embedding, we get

THEOREM 2. *Under the above assumptions,*

$$f_{\beta_0,R}(r) = \phi(r)J^{1/2}(\beta)\frac{\partial\beta}{\partial h}(h)\left\{1+O(n^{-1/2})\right\} \tag{4.10}$$

in a large deviation region $|h| \leq c$, *with* $h = r\sqrt{n}$. ∎

Theorem 1 is an immediate corollary.

It remains to show (4.8). In Jensen's (1997) notation,

$$\hat{\theta}^\ell - \hat{\theta}^{\ell-1} = \left(\hat{\beta}_1^\ell - \hat{\beta}_1^{\ell-1}, \hat{\theta}_2^\ell - \hat{\theta}_2^{\ell-1}, \ldots, \hat{\theta}_{\ell-1}^\ell - \hat{\theta}_{\ell-1}^{\ell-1}, \hat{\theta}_\ell^\ell - (\hat{\beta}_1^{\ell-1})^\ell, (\hat{\beta}_1^\ell) - (\hat{\beta}_1^{\ell-1})^{\ell+1}, \ldots, (\hat{\beta}_1^\ell)^p - (\hat{\beta}_1^{\ell-1})^p\right) \tag{4.11}$$

where $\hat{\theta}_\ell - (\hat{\beta}_1^{\ell-1})^\ell$ is the term in the ℓ'th column. Note that $(\beta_1^\ell)^k$ is $\beta_1^{\ell-1}$ raised to power k, which is the only instance of power notation in (4.11).

Since $\mathrm{corr}(\overline{T}_1, \overline{T}_\ell) \simeq 0$, (note that this is where the reparametrization above is used), $\hat{\theta}_\ell^\ell - (\hat{\beta}_1^{\ell-1})^\ell$ is $O_p(n^{-1/2})$ but not $o_p(1)$. On the other hand, for $\ell \geq 2$, $\hat{\beta}_1^\ell - \hat{\beta}_1^{\ell-1} = O_p(n^{-1})$. Hence, for $\ell \geq 2$,

$$\hat{\theta}^\ell - \hat{\theta}^{\ell-1} = (0, \hat{\theta}_2^\ell - \hat{\theta}_2^{\ell-1}, \ldots, \hat{\theta}_\ell^\ell - (\hat{\beta}_1^{\ell-1})^\ell, 0, \ldots, 0) + O_p(n^{-1}). \tag{4.12}$$

Hence the determinants in equation (7) in Jensen (1997) can be evaluated by multiplying the diagonal, and so (5.8) follows. Note that in the above argument, if $\overline{T}_\ell$ is zero, one just deletes line and column ℓ and makes the appropriate modification to the next column. This does not affect the result.

ACKNOWLEDGEMENTS

The author would like to thank Ole Barndorff-Nielsen, Jens Jensen. Peter McCullagh, Nicole Lazar, Ib Skovgaard, Trevor Sweeting and Andy Wood for extremely useful discussions, and Mitzi Nakatsuka for typing part of the paper.

APPENDIX 1

Obviously, if $g(x) = xe^x$,

$$\begin{aligned}\widetilde{\ell}(\beta) &\simeq \frac{1}{n} E_\beta \left(\ell(\beta) - \ell(\beta_0)\right) \\ &= \frac{1}{n} E_{\beta_0} g\left(\ell(\beta) - \ell(\beta_0)\right),\end{aligned} \tag{A1.1}$$

so that the p'th derivative is

$$\widetilde{\ell}^{(p)}(\beta) \simeq \frac{1}{n} E_{\beta_0} \sum_{q_1+2q_2+\cdots+vqv=p} g^{(v)}(0) c(q_1, \ldots, q_v) \ell(\beta_0)^{q_1} \cdots \ell^{(v)}(\beta_0)^{q_v} \tag{A1.2}$$

which yields (2.5) since $g^{(v)}(0) = v$ and since moments can be replaced by cumulants in the above. The latter can either be seen by direct computation, or by observing that the right hand side of (A1.1) is $O(1)$.

To find J, note that

$$\frac{\partial^p}{\partial \beta^p} \ddot{\ell}(\beta) \exp(\ell(\beta) - \ell(\beta_0)) = \sum_{r=0}^{p} \binom{p}{r} \ell^{(r+2)} \sum_{q_1+2q_2+\cdots+vq_v=p-r} c(q_1, \ldots, q_v) (\dot{\ell})^{q_1} \cdots (\ell^{(v)})^{q_v} \tag{A1.3}$$

which gives (2.7) for the same reasons as used above, and because

$$\widetilde{c}(q_1, \ldots, q_v) = \sum_{r=0}^{p} \binom{p}{r} c(q_1, \ldots, q_{r+1}, q_{r+2} - 1, q_{r+3}, \ldots, q_v), \tag{A1.4}$$

which gives (2.12) by direct computation, using (2.6).

Appendix 2

This is the calculation of (3.3). The expectation is calculated by the Taylor expansion method as we need an additional term in $E(R_n)$ to that provided by Theorem 1. In principle, the technology used to derive Theorem 1 can also be used to get higher order terms (like k_1'), but since we are just looking for an expectation, this is a little like shooting sparrows with cannons. If one were looking for higher order terms in higher oder cumulants, an extension of Theorem 1 would greatly ease one's life.

The derivation is here done in the multivariate setting. We shall use the notation in McCullagh (1987). In the univariate case $R = W_1\sqrt{\nu^{1,1}}$. By adding the $O_p(n^{-3/2})$ term to the stochastic expansion (7.15) on p. 214 in McCullagh (1987), one gets that

$$\begin{aligned} W_r = \overline{W_r} &+ n^{-1/2} Z_{rs} Z^s/2 \\ &+ n^{-1}\{Z_{rst}Z^sZ^t/3! + (\nu_{rstu} + \nu_{r,s,t,u})Z^sZ^tZ^u/4! \\ &\quad + 3Z_{rs}Z^{st}Z_t/8 + 5Z_{rs}Z_tZ_u\nu^{stu}/12\} \\ &+ n^{-3/2}\{(\nu_{rstuv} + \nu_{r,s,t,u,v})Z^sZ^tZ^uZ^v/5! \\ &+ 11(\nu_{rstu} + \nu_{r,s,t,u})\nu^{s,v}\nu_{vwx}Z^tZ^uZ^wZ^x/144 + U\} \\ &+ O_p(n^{-2}) \end{aligned} \tag{A2.1}$$

where $\overline{W}_r$ is the corresponding quantity for an exponential family alternative with the same score as the original alternative, and where U is a generic term referring to a linear combination of products of 4 Z's, at least one and at most three of which is a higher order derivative. It follows that

$$\begin{aligned} EW_r = E\overline{W}_r & \\ &+ n^{-3/2}\{\nu^{s,i}\nu^{t,j}\nu_{rst,i,j}/3! + (\nu_{rstu} + \nu_{r,s,t,u})\nu^{s,i}\nu^{t,j}\nu^{u,k}\nu_{i,j,k}/4! \\ &\quad + 3\nu^{s,i}\nu^{t,j}\nu_{rs,ij,t}/8 + 5\nu_{rs,t,u}\nu^{stu}/12 \\ &\quad + 3(\nu_{rstuv} + \nu_{r,s,t,u,v})\nu^{s,t}\nu^{u,v}/5! + 11(\nu_{rstu} + \nu_{r,s,t,u})\nu^{s,v}\nu_{vwx}\nu^{t,u}\nu^{w,x}/48\} \\ &+ O(n^{-5/2}) \end{aligned} \tag{A2.2}$$

Simplifying to univariate notation gives (3.3).

References

Andrews, D. F., and Stafford, J. E. (1993). Tools for the symbolic computation of asymptotic expansions. *J. Roy. Statist. Soc.* **55** 613- 627.

Bartlett, M.S., 1953a, Approximate confidence intervals, *Biometrika* **40** 12–19.

Bartlett, M.S., 1953b, Approximate confidence intervals. II. More than one unknown parameter, *Biometrika* **40** 306–317.

Barndorff-Nielsen, O. E. (1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrika* **70** 343–365.

Barndorff-Nielsen, O. E. (1986). Inference on full or partial parameters based on the standardized signed log likelihood ratio. *Biometrika* **73** 307–322.

Barndorff-Nielsen, O. E. (1991). Modified signed log likelihood ratio. *Biometrika* **78**, 557–63.

Barndorff-Nielsen, O. E., and Cox, D. R. (1979). Edgeworth and saddle-point approximations with statistical applications (with discussion). *J. Roy. Statist. Soc. B* **41** 279–312.

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. B* **46** 483–495.

Barndorff-Nielsen, O. E., and Cox, D. R. (1989). *Asymptotic Techniques for use in Statistics* (Chapman and Hall, London).

Barndorff-Nielsen, O. E., and Wood, A. T. A. (1995). On large deviations and choice of ancillary for p^* and the modified directed likelihood. Preprint.

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

Bickel, P.J., Chibisov, D.M. and van Zwet, W.R. (1981). On efficiency of first and second order. *Int. St. Rvw.* **49**, 169–175.

Chaganty, N. R., and Sethuraman, J. (1985). Large deviation local limit theorems for arbitrary sequences of random variables. *Ann. Probab.* **13** 97–114.

Corcoran, S.A., Davison. A.C. and Spady, R.H. (1995). Reliable inference from empirical likelihoods'. Preprint.

Cox, D.R. (1972). Regression models and life tables (with discussion), *J. Roy. Statist. Soc. B* **34** 187–220.

Cox, D.R. (1975). Partial likelihood, *Biometrika* **62** 269–276.

Daniels, H. E. (1952). Saddlepoint approximations in statistics. *Ann. Math. Statist.* **25** 631–650.

DiCiccio, T.J., Hall, P., and Romano, J.P. (1991). Empirical likelihood is Bartlett-correctable, *Ann. Statist.* **19** 1053–1061.

DiCiccio, T.J., and Romano, J.P. (1989). On adjustments based on the signed root of the empirical likelihood ratio statistic, *Biometrika* **76** 447–456.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion* (Springer-Verlag, New York).

Jacod, J., and Shiryaev, A.N. (1987). *Limit Theorems for Stochastic Processes* (Springer-Verlag).

Jensen, J. L. (1992). The modified signed likelihood statistic and saddlepoint approximations. *Biometrika* **79** 693–703

Jensen, J. L. (1995). *Saddlepoint Approximations in Statistics* (Oxford University Press, Oxford).

Jensen, J. L. (1997). A simple derivation of a natural large deviation modified likelihood statistic. To appear in *Scand. J. Statist.*

Kong, A. and Cox, N.J. (1996). From efficient nonparametric tests for linkage analysis to semiparametric models and lodscores. Technical report no. 435, Department of Statistics, University of Chicago.

Lawley, D. N. (1956). A general method for approximating the distribution of likelihood ratio criteria, *Biometrika* **43** 295–303.

Lazar, N., and Mykland. P.A. (1997). An evaluation of the power and conditionality properties of empirical likelihood (in preparation).

McCullagh, P. (1984). Local sufficiency. *Biometrika* **71** 233–244.

McCullagh, P. (1987). *Tensor Methods in Statistics* (Chapman and Hall, London).

McCullagh, P., and Tibshirani, R. (1990). A simple method for the adjustment of profile likelihoods. *J. Roy. Statist. Soc. B* **52** 325-344.

McLeish, D.L., and Small, C.G. (1992). A projected likelihood function for semiparametric models, *Biometrika* **79** 93–102.

Mykland, P.A. (1994). Bartlett type identities for martingales. *Ann. Statist.* **22** 21–38.

Mykland, P.A. (1995a). Dual likelihood. *Ann. Statist.* **23** 396–421.

Mykland, P.A. (1995b). Embedding and asymptotic expansions for martingales, *Probab. Theory Rel. Fields* **103** 475-492.

Mykland, P.A. (1996). The accuracy of likelihood, Technical report no. 420, Department of Statistics, University of Chicago.

Skovgaard, Ib (1986). A note on the differentiation of cumulants of log likelihood derivatives. *Internat. Statist. Rev.* **54** 29–32.

Skovgaard, Ib (1990). On the density of minimum contrast estimators. *Ann. Statist.* **18** 779–789.

Skovgaard, Ib (1996). A general large deviation approximation to one-parameter tests. Bernoulli **2** 145-165.

Wallace, D,L. (1958). Asymptotic approximations to distributions. *Ann. Math. Statist.* **29** 635–654.

Wong, W.H., 1986, Theory of partial likelihood, *Ann. Statist.* 14 88–123.