Bernoulli **32**(1), 2026, 81–95
https://doi.org/10.3150/25-BEJ1906

O. E. Barndorff-Nielsen's approximate conditional inference

NANCY REID[1,a] and HEATHER BATTEY[2,b]

[1]*Department of Statistical Sciences, University of Toronto, Toronto, Canada,* [a]*nancym.reid@utoronto.ca*
[2]*Department of Mathematics, Imperial College London, London, UK,* [b]*h.battey@imperial.ac.uk*

We review three interrelated contributions by Barndorff-Nielsen to statistical theory: his p^* approximation, modified profile likelihood function, and r^* approximation. The development was important for the foundations of inference, and the r^* approximation in particular has also proved useful in a range of applications.

Keywords: Ancillary statistic; asymptotic expansion; conditional inference; modified profile likelihood; large deviation

1. Introduction

The starting point for our paper is the p^* approximation, otherwise known as "the magical mystery formula". At the time of its origin the formula did indeed seem mysterious, at least partly due to Ole Barndorff-Nielsen's penchant for using highly abbreviated notation. In retrospect it is not quite so opaque, and emerges from direct conditional calculations. There were a few false starts in the early literature; for example, Barndorff-Nielsen (1980) presented two versions of p^*, one using observed, and the other expected, Fisher information; it quickly became apparent that the former was the more appropriate. A clear explanation of this point, and a careful proof of the validity of the p^* approximation, was provided in Skovgaard (1990a).

In spite of its apparent simplicity, the p^* approximation is not so simple to use in calculations, as will be explained below. However its definite integral is easier to calculate and this leads to the r^* approximation of Barndorff-Nielsen (1986a, 1990, 1991).

The motivating idea behind these approximations was Ole Barndorff-Nielsen's long-standing interest in conditional inference; see the paper by Michael Sørensen in this volume (Sørensen, 2026). This interest sparked his long and fruitful collaboration with David Cox, culminating in their book *Inference and Asymptotics* (Barndorff-Nielsen and Cox, 1994).

The idealized Fisherian programme of inference on a parameter θ of a parametric model $\{f(y;\theta) : \theta \in \mathbb{R}^{d_\theta}\}$ starts with a reduction of the observed data $y^o \in \mathbb{R}^n$ to a minimal sufficient statistic $s^o = s(y^o)$, where $s : \mathbb{R}^n \to \mathbb{R}^d$ and $d \geq d_\theta$ is the smallest such that s^o carries all the information on θ available in y^o. If $d = d_\theta$, inference is based on the distribution of the likelihood function $L(\theta) \propto f(Y;\theta)$, which depends on Y only through $S = s(Y)$ by definition. However if $d_\theta < d$, inference is based on a conditional likelihood function, obtained from a partition of S. This partition $S = (T, A)$ is such that T, or some one-to-one transformation of T, estimates θ and A has a distribution that is free of θ. By making the inference conditional on the observed value $A = a^o$, the information about θ lost through using T alone is recovered. More importantly, inference reflects the relevant information content of the data actually observed, as opposed to the average information content over hypothetical draws from the generating distribution.

The situations permitting a minimal sufficient reduction in dimension consist primarily of exponential-family models. While the popularity of many important examples owes partly to their having this form, scientific practice produces many plausible models that are irreducible in the manner

 © 2026 ISI/BS

just described. For these, it is natural to seek an approximate implementation of the idealized Fisherian blueprint. This relies on notions of approximate minimal sufficiency, approximate ancillarity, and a corresponding form of approximate reduction. The p^* approximation operationalizes this, drawing on ideas of small-sample distributional improvement for the likelihood ratio statistic introduced by Bartlett (1937), and inspired by further developments in conditional inference and approximation theory.

Bartlett (1936, 1937) pointed out that the maximum likelihood estimator of a parameter of interest can perform poorly when the number of nuisance parameters is large relative to the number of observations, and proposed a likelihood function based on a conditional or marginal distribution. This proposal was developed in a sequence of papers covering challenging regression-type situations (Cox, 1958a,b, 1968, 1970) and emphasizing the role of conditioning for the appropriate calibration of inferential statements. Daniels (1954) derived the saddlepoint expansion for the distribution of a sum of random variables, and Barndorff-Nielsen and Cox (1979) showed how Edgeworth and saddlepoint expansions could be useful for approximation of relevant conditional distributions in statistical problems.

With T taken as the maximum likelihood estimator and $\hat{\theta}$ an arbitrary evaluation point for the conditional density function of T given $A = a^o$, the p^* approximation to this conditional density function at $\hat{\theta}$ is, in the original notation of Barndorff-Nielsen (1980, 1983),

$$p^*(\hat{\theta} \mid a^o; \theta) = c|\hat{\jmath}|^{1/2} e^{\overline{\ell}}, \tag{1}$$

where $\overline{\ell} = \ell(\theta; \hat{\theta}, a^o) - \ell(\hat{\theta}; \hat{\theta}, a^o)$ is the normed log-likelihood function, $|\hat{\jmath}|$ is the determinant of the observed Fisher information matrix and c is a normalizing constant. Equation (1) is a higher-order approximation in the sense that its relative error is $O(n^{-3/2})$ in ordinary repeated sampling with sample size n from a continuous density, a considerable improvement over the $O(n^{-1/2})$ error from the limiting normal approximation.

The apparent simplicity of (1) hides some details revealed in the more explicit notation

$$p^*(\hat{\theta} \mid a^o; \theta) = c(\theta, a^o) \left| j\{\hat{\theta}; y(\hat{\theta}, a^o)\} \right|^{1/2} \frac{L\{\theta; y(\hat{\theta}, a^o)\}}{L\{\hat{\theta}; y(\hat{\theta}, a^o)\}}, \tag{2}$$

where L is the likelihood function and $y(\hat{\theta}, a^o)$ is any realisation of the observable random variables $Y = (Y_1, \ldots, Y_n)$ whose value is consistent with the implied minimal sufficient statistic $(\hat{\theta}, a^o)$. The random variables Y_i are each assumed independent, or independent conditionally on known covariates, from a continuous distribution that is known up to a parameter $\theta \in \mathbb{R}^{d_\theta}$. As part of the derivation of the p^* approximation, it emerges that the leading term in the normalizing constant is $(2\pi)^{-d_\theta/2}$, and more precisely,

$$c(\theta, a^o) = \frac{1}{(2\pi)^{d_\theta/2}} \left\{ 1 + \frac{c_1(\theta, a^o)}{n} + O(n^{-3/2}) \right\}. \tag{3}$$

Barndorff-Nielsen (1986b, Eq. (3.7)) provides an explicit expression for the $O(1)$ quantity c_1.

The right-hand side of (2) arose initially as the common approximation for two different model classes: transformation models and exponential-family models. These idealized cases are taken up in §2. In general models, Skovgaard (1990a) showed that

$$f_{T|A}(\hat{\theta} \mid a^o; \theta) = p^*(\hat{\theta} \mid a^o; \theta) \{1 + O(n^{-3/2})\},$$

under some regularity conditions on the model, provided that A is approximately ancillary to second order (Skovgaard, 1985, §4). Approximate ancillarity is discussed in more detail in Section 3.

2. Two idealized cases

Key to the development of p^* as a unifying construction was the observation by Barndorff-Nielsen (1980) that the same approximation emerges in two idealized cases with no nuisance parameters, as we now describe.

In the first, there is no reduction in dimension by sufficiency, but there is an exact ancillary complement, so $Y \cong (T, A)$, T is of dimension d_θ, A is of dimension $n - d_\theta$ and

$$f_Y(y;\theta)dy = |J(\hat\theta, a)| f_Y\{y(\hat\theta, a);\theta\} d\hat\theta_1 \cdots d\hat\theta_{d_\theta} da_1 \cdots da_{n-d_\theta}.$$

The density function of T conditional on $A = a^o$ is

$$f_{T|A}(\hat\theta \mid a^o;\theta)d\hat\theta = \frac{f_Y\{y(\hat\theta, a^o);\theta\}|J(\hat\theta, a^o)|d\hat\theta}{\int_{a(y)=a^o} f_Y(y;\theta)dy} = \frac{f_Y\{y(\hat\theta, a^o);\theta\}|J(\hat\theta, a^o)|d\hat\theta}{\int f_Y\{y(t, a^o);\theta\}|J(t, a^o)|dt}, \tag{4}$$

and the denominator does not depend on θ by the definition of ancillarity.

The location model on $\mathbb{R}$ provides the most direct illustration; the density of Y_i is of the form $f_{Y_i}(y_i;\theta) = f_0(y_i - \theta)$, and

$$f_Y(y;\theta)dy = \prod_{i=1}^{n} f_0(y_i - \theta)dy_i = f(h(a) + \hat\theta - \theta) \prod_{i=1}^{n-1} f_0(a_i + \hat\theta - \theta)|J| da_i d\hat\theta;$$

here the completion of the information from an arbitrary data point y from the corresponding maximum likelihood estimate $\hat\theta$ is given by, say, $a = (a_1, \ldots, a_{n-1}) = (y_1 - \hat\theta, \ldots, y_{n-1} - \hat\theta)$, from which $y_i = a_i + \hat\theta$, $i = 1, \ldots, n-1$. The maximum likelihood estimator solves $\sum_{i=1}^{n} \nabla_\theta \log f(y_i - \hat\theta) = 0$, so $y_n = h(a) + \hat\theta$ for some function $h(\cdot)$, and the Jacobian of the transformation depends only on a. Because in the joint density function $\hat\theta$ and θ enter symmetrically, integration with respect to t in the denominator of (4) is equivalent to integration with respect to θ and

$$f_{T|A}(\hat\theta \mid a^o;\theta)d\hat\theta = \frac{L\{\theta; y(\hat\theta, a^o)\}d\hat\theta}{\int L\{t; y(\hat\theta, a^o)\}dt}. \tag{5}$$

The Laplace approximation to the integral in the denominator is

$$(2\pi)^{1/2}|j\{\hat\theta; y(\hat\theta, a^o)\}|^{-1/2} L\{\hat\theta; y(\hat\theta, a^o)\}. \tag{6}$$

By the symmetry of the location model and the definition of ancillarity, the denominator does not in fact depend on $\hat\theta$, but it is convenient to write it in this form as it shows the link with (2), and that the p^* approximation is exact after renormalization in this case. Incorporating the next term in the Laplace expansion verifies (3), with c_1 not depending on θ.

In the location model the distribution of any location-equivariant statistic T conditional on the corresponding ancillary A will lead to the same conditional inference for θ as that based on (5). Barndorff-Nielsen and Cox (1994, p. 36), for example, use $(y_{(n)}, a) := (y_{(n)}, y_{(1)} - y_{(n)}, \ldots, y_{(n-1)} - y_{(n)})$, where $\{y_{(1)}, \ldots, y_{(n)}\}$ are the order statistics.

The argument leading to (5) extends to any transformation model, which includes both location-scale and regression-scale models under standard regularity conditions (Barndorff-Nielsen, 1980).

In the second idealized case, the minimal sufficient statistic S has dimension $d = d_\theta$, in which case no ancillary complement is needed for inference. For example, the natural exponential family with $d = 1$

has density

$$f_Y(y;\theta)dy = \exp\Big\{\theta \sum_{i=1}^{n} y_i - n\kappa(\theta)\Big\} \prod_{i=1}^{n} f_0(y_i)dy_i;$$

the minimal sufficient statistic is $s(y) = \sum_i y_i$ and there is no ancillary statistic. We may consider as the basis for inference

$$f_S(s;\theta)ds = \exp\{\theta s - n\kappa(\theta)\}h_0(s)ds, \tag{7}$$

where

$$h_0(s) = \int_{y_1+\cdots+y_n=s} \prod_{i=1}^{n} f_0(y_i)dy_1, \ldots dy_n$$

is the density of S when $Y_1, \ldots, Y_n$ have the baseline density $f_0(y)$. This generalizes in an obvious way to canonical exponential families with canonical statistic $s(y) = \sum_{i=1}^{n} s(y_i)$ and canonical parameter $\theta \in \mathbb{R}^{d_\theta}$ leading to inference based on

$$f_S(s;\theta)ds = \exp\{\theta^{\mathrm{T}} s - n\kappa(\theta)\}h_0(s)ds, \tag{8}$$

with $h_0(s)$ appropriately redefined, where $ds = ds_1 \cdots ds_d$.

The saddlepoint approximation to the density of a sum S of independent, identically distributed random variables was derived by Daniels (1954) using the complex-variables method of steepest descents to approximately invert the characteristic function. Barndorff-Nielsen and Cox (1979) related this to the Edgeworth expansion, extended it to the multivariate setting, and provided several applications to approximate conditional inference. The multivariate version is derived in Barndorff-Nielsen and Cox (1989, §6.5).Writing $K_S(\tau) = \log \mathbb{E}\exp(\tau^{\mathrm{T}} S)$ for the cumulant generating function of $S = (s_1, \ldots, s_d) = (\sum s_1(y_i), \ldots, \sum s_d(y_i))$, the approximation is

$$f_S(s) = \frac{\exp\{K_S(\hat\tau) - \hat\tau^{\mathrm{T}} s\}}{(2\pi)^{d/2}|K_S''(\hat\tau)|^{1/2}}\{1 + O(n^{-1})\}, \tag{9}$$

where $K_S'(\hat\tau) = s$. The dependence on θ has been suppressed as this approximation is valid for any sum of independent random variables, provided that the moment-generating function exists in a neighbourhood of the origin.

In the exponential-family model (7), $K(\tau) = n\kappa(\theta + \tau) - n\kappa(\theta)$, from which it follows that $\hat\tau = \hat\theta - \theta$ and $K''(\hat\tau) = n\kappa''(\hat\theta)$; $K''(\cdot)$ and $\kappa''(\cdot)$ are $d \times d$ matrices. Further, $\ell(\theta; s) = \theta s - n\kappa(\theta)$, so the leading term of (9) gives the approximation

$$f_S(s;\theta)ds \doteq \frac{\exp\{\ell(\theta;s) - \ell(\hat\theta;s)\}}{(2\pi)^{d/2}|j(\hat\theta)|^{1/2}}ds,$$

and using $n\kappa''(\hat\theta)d\hat\theta = ds$ and the likelihood equation $n\kappa'(\hat\theta) = s$, we have

$$f_T(\hat\theta;\theta)d\hat\theta \doteq \frac{|j(\hat\theta)|^{1/2}}{(2\pi)^{d/2}}\frac{L(\theta;\hat\theta)}{L(\hat\theta;\hat\theta)}d\hat\theta,$$

which on renormalization is the p^* approximation (2). The simplicity of the transformation from s to $\hat\theta$ reflects the duality between the canonical parameter space $\Theta \subset \mathbb{R}^d$ and the parameter space $\mathscr{S}$ for $\eta = \mathbb{E}(S)$, which is also the sample space for S.

The normal distribution with unit variance and unknown mean θ is both a one-parameter exponential family and a location model. Although there is no preliminary reduction by sufficiency invoked in the derivation of the location model version of p^*, (5), the result depends on the data only through the likelihood function, so by the factorization theorem only depends on the sufficient statistic, in this case $\bar{Y}$. The location form also implies that the saddlepoint approximation is exact for this model, as can be verified directly.

The simplest departure from the idealized cases is the curved exponential model, with density

$$f_Y(y;\theta)dy = \exp\{s(y)^{\mathrm{T}}\varphi(\theta) - \kappa(\varphi(\theta))\}\prod_{i=1}^{n} f_0(y_i)dy_i, \tag{10}$$

where $d_\theta < d$, and an exact or approximate ancillary complement A of dimension $d_A = d - d_\theta$ is required. Varying $\theta \in \Theta$ defines a d_θ-dimensional differentiable manifold Φ_Θ in Φ and the duality between Φ and $\mathscr{S}$ means that to Φ_Θ there corresponds a differentiable manifold $\mathscr{S}_\Theta$ in $\mathscr{S}$. This duality allows calculation of the Jacobian of the transformation; see Barndorff-Nielsen and Cox (1994, pp. 66–70).

3. Approximate ancillarity

The usual axiomatic definition of exact ancillarity for θ is that the distribution of A does not depend on θ; in which case

$$f_{T,A}(\hat{\theta}, a; \theta)d\hat{\theta}da \propto f_{T|A}(\hat{\theta} \mid a; \theta)f_A(a)d\hat{\theta}da,$$

and it is clear that no information about θ is lost by using the first factor for inference. Correspondingly, if $S = (T, A) \in \mathbb{R}^d$ and $d > d_\theta$, then conditioning on an exact ancillary statistic A recovers information in S that is lost by estimating θ by T.

When no exactly ancillary statistic exists, it is natural to seek an approximate ancillary statistic; a $d - d_\theta$-dimensional component of the minimal sufficient statistic that is approximately distribution constant with respect to changes in θ. The form of the p^* approximation in the two idealized cases above suggests that if we confine attention to approximations accurate to $O(n^{-3/2})$, it would suffice for A to be distribution constant at most to the same order. Several investigations of approximate ancillarity, in Efron and Hinkley (1978), Cox (1980), Hinkley (1980), Barndorff-Nielsen (1984), Skovgaard (1985) and McCullagh (1984) suggested that A need only be ancillary to $O(n^{-1})$. This was made more precise by Skovgaard (1990a): as part of the proof of the p^* approximation he established that conditioning on a second-order local ancillary statistic is enough to verify the relative error of the approximation to $O(n^{-3/2})$, in a $\sqrt{n}$-neighborhood of the maximum likelihood estimator, as well as its relative error of $O(n^{-1})$ in bounded neighborhoods of $\hat{\theta}$.

Moreover, the p^* approximation to the conditional density is essentially unique, in the sense that any second-order local ancillary statistic will lead to the same approximate inference; this is the approximate version of the result for location models that the distribution of any location-equivariant statistic, conditional on the appropriate complementing ancillary, leads to the same inference as that based on (5). This was anticipated by McCullagh (1984).

An approximate ancillary statistic for a curved exponential family with $d = 2, d_A = d_\theta = 1$ was derived in Efron and Hinkley (1978) as a standardized difference between the observed and expected Fisher information for θ; Skovgaard (1985) extended their definition to $d > 2$. Direct expansion of the log-likelihood function $\ell(\theta)$ is used to show that A follows a standard normal distribution with error of order $O(n^{-1/2})$. Since the leading-order approximate distribution does not depend on θ, this is typically

described as first-order ancillarity. However, if we consider only values of θ in a $\sqrt{n}$-neighbourhood of some fixed value θ_0 say, then A is second-order locally ancillary: $f_A(a;\theta) = f_A(a;\theta_0)\{1 + O(n^{-1})\}$. Cox (1980) introduced the notion of local ancillarity and constructed a locally ancillary statistic by finding a function of the log-likelihood derivatives whose moments are constant to the appropriate order; the result is closely related to the Efron-Hinkley ancillary.

We now briefly describe a general construction of an approximate ancillary statistic for curved exponential families due to Barndorff-Nielsen and Wood (1998). The model (10) can be naturally embedded in a d-dimensional full exponential family, with parameter $\varphi = (\varphi_1, \ldots, \varphi_d)$. The log-likelihood ratio statistic for assessing adequacy of the curved model $\varphi = \varphi(\theta)$ against the full model has a limiting $\chi^2_{d-d_\theta}$ distribution under standard regularity conditions. This statistic can be regarded as a sum of $d - d_\theta$ components, each of which assesses a nested model of one dimension lower. These components have asymptotically χ^2_1 distributions, so each can be converted to a standard Gaussian random variable by taking the signed square root. Together, these consistute a $d - d_\theta$ multivariate standard-normal random vector, which serves as the approximate ancillary. The detailed construction is somewhat more involved, as it is based on either the signed square root of the χ^2_1 statistic r, or its modification r^* described in Section 5 below. Barndorff-Nielsen and Wood (1998, §2.4) describe these two versions as unmodified, and modified, directed likelihoods, respectively.

The advantage of working with a curved exponential model is that a minimal sufficient statistic of dimension of the embedding model is well-defined, and the approximate ancillary statistic is a function of this minimal sufficient statistic. In a more general model, there may not be a reduction in dimension for the minimal sufficient statistic; the general location model provides an example of this, but the simplicity of its structure provides an exact solution as described at (5).

Suppose we have a sample of size n from an arbitrary one-parameter model which has no reduction in dimension by sufficiency. By Taylor expansion of the log-likelihood function about an arbitrary point θ_0,

$$f(y;\theta)dy \simeq f(y;\theta_0)\exp\Big\{\textstyle\sum_{j=1}^{p}(\theta-\theta_0)^j\ell^{(j)}(\theta_0;y)\Big\}dy, \quad (\theta \to \theta_0) \tag{11}$$

and this density approximation has the form of a curved exponential family, with p-dimensional sufficient statistic $S = \{\ell'(\theta_0;y), \ldots, \ell^{(p)}(\theta_0;y)\}$. This line of argument was suggested by Fisher (1934), and developed further by Cox (1980). In this approximating model, the embedding full exponential family is of fixed dimension and an approximate ancillary can be obtained using the arguments of Barndorff-Nielsen and Wood (1998) described above. In principle, this expansion could be extended to order n, but then the dimension of the implied ancillary complement would increase with n, making the asymptotic arguments more delicate. Skovgaard (1990b) developed this approach rigorously; see in particular the overview in his Chapter 2.

4. Nuisance parameters and higher-order conditional inference

4.1. Modified profile likelihood

Barndorff-Nielsen (1983) used the p^* approximation to suggest a modification to profile likelihood in models with nuisance parameters. For this we write $\theta = (\psi, \lambda)$, where $\psi \in \mathbb{R}^{d_\psi}$ is the parameter of interest, and $\lambda \in \mathbb{R}^{d_\theta - d_\psi}$ is the nuisance parameter. The profile log-likelihood function is

$$\ell_{\mathrm{p}}(\psi) = \ell(\psi, \hat{\lambda}_\psi), \tag{12}$$

where $\hat{\lambda}_\psi$ is the constrained maximum likelihood estimator, in regular models the solution in λ of

$$\partial \ell(\psi, \lambda; y)/\partial \lambda = 0. \tag{13}$$

For first-order approximations, statistical summaries can be derived treating the profile log-likelihood function as an ordinary log-likelihood function (Barndorff-Nielsen and Cox, 1994, §3.4). However, these approximations can be misleading in small samples, or even in moderately large samples when the dimension of λ is commensurably large, because no account is taken of errors in estimation of the nuisance parameter.

In a departure from our previous notation, where we used T for the random variable and $\hat{\theta}$ for an arbitrary evaluation point of $f_T(\cdot)$, here we use $\hat{\theta}$ for both, with the context ensuring no ambiguity.

The modified profile log-likelihood function developed in Barndorff-Nielsen (1983) is

$$\ell_{\mathrm{mp}}(\psi) = \ell_{\mathrm{p}}(\psi) - \frac{1}{2}\log|j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)| + \log\left|\frac{d\hat{\lambda}}{d\hat{\lambda}_\psi}\right|, \tag{14}$$

where $j_{\lambda\lambda}(\psi, \lambda) = -\partial^2 \ell(\theta)/\partial\lambda\partial\lambda^{\mathrm{T}}$ is the nuisance-parameter submatrix of the observed Fisher information matrix, and in the final term it is assumed that $\hat{\lambda}_\psi = \hat{\lambda}_\psi(\psi, \hat{\psi}, \hat{\lambda}, a)$. The derivation of (14) is based on the use of two p^* approximations: one for the distribution of $\hat{\theta}$, as in (1), and a second version of (1) for the distribution of the constrained maximum likelihood estimator $\hat{\lambda}_\psi$ in the model with ψ considered fixed. The modified profile log-likelihood function is invariant to one-to-one transformations of the form $(\psi, \lambda) \mapsto (\psi, \chi(\psi, \lambda))$, or slightly more generally $(\psi, \lambda) \mapsto (\phi(\psi), \chi(\psi, \lambda))$; these are called interest-respecting reparametrizations.

In the nuisance parameter versions of the idealized cases discussed in Section 2, the modified profile likelihood function approximates the appropriate conditional or marginal density that would be used for inference about ψ (Barndorff-Nielsen, 1987, Barndorff-Nielsen and Cox, 1994, pp.265–7). The expression is, however, well-defined outside of these cases, although the computation of the final term may be difficult. Let

$$F(\hat{\lambda}_\psi, \hat{\lambda}) = \frac{\partial}{\partial \lambda}\ell\{\psi, \lambda; y(\hat{\psi}^o, \hat{\lambda}, a^o)\}\Big|_{\lambda=\hat{\lambda}_\psi}.$$

For any given ψ, the constrained maximum likelihood estimator $\hat{\lambda}_\psi$ solves $F(\hat{\lambda}_\psi, \hat{\lambda}) = 0$ and total differentiation of this equation shows that

$$\frac{d\hat{\lambda}}{d\hat{\lambda}_\psi} = -\left[\ell_{\lambda;\hat{\lambda}}(\psi, \hat{\lambda}_\psi)\right]^{-1} j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi), \tag{15}$$

where

$$\ell_{\lambda;\hat{\lambda}}(\psi, \lambda) = \frac{\partial^2 \ell(\psi, \lambda; \hat{\psi}^o, \hat{\lambda}, a^o)}{\partial\lambda\partial\hat{\lambda}^{\mathrm{T}}}$$

is a square matrix of mixed derivatives on the parameter space and sample space. In practice, (15) is evaluated either at $y^o = y(\hat{\psi}^o, \hat{\lambda}^o, a^o)$ or at $Y = y(\hat{\psi}, \hat{\lambda}, A)$ for establishing results like (16).

For inference about ψ, the modified profile likelihood function provides analogues of the Wald statistic and likelihood ratio statistic, using the approximations

$$\begin{aligned} \sqrt{n}(\tilde{\psi}_{\mathrm{mpl}} - \psi) &\doteq N_{d_\psi}\left(0, j_{\mathrm{mpl}}^{-1/2}(\tilde{\psi}_{\mathrm{mpl}})\right), \\ 2\{\ell_{\mathrm{mpl}}(\tilde{\psi}_{\mathrm{mpl}}) - \ell_{\mathrm{mpl}}(\psi)\} &\doteq \chi^2_{d_\psi}, \end{aligned} \tag{16}$$

where $\tilde{\psi}_{\text{mpl}}$ is the solution of $\ell'_{\text{mpl}}(\psi) = 0$. Although these are only first-order approximations, being based on the limiting distributions, in finite samples they tend to be more accurate than similar versions based on the profile likelihood function. Some insight into this can be obtained by studying Neyman-Scott problems, where $n = mq$ observations are taken in m clusters each of size q, each cluster having a different nuisance parameter (Barndorff-Nielsen, 1983, Sartori, 2003).

Cox and Reid (1987) used an approximate conditioning argument to propose an adjusted profile likelihood function

$$\ell_{\text{ap}}(\psi) = \ell_{\text{p}}(\psi) - \frac{1}{2}\log|j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi|, \tag{17}$$

i.e., ignoring the final term of (14). This expression is not invariant to interest-respecting reparametrizations, but the dependence on the parametrization was reduced by requiring that the nuisance parameter be orthogonal to the parameter of interest, with respect to the expected Fisher information metric, i.e. $\mathbb{E}_\theta\{-\partial^2\ell(\psi,\lambda)/\partial\psi\partial\lambda^{\text{T}}\} = 0$. Under orthogonality it can be shown that $\hat{\lambda}_\psi - \hat{\lambda} = O_p(n^{-1})$, for $\psi - \hat{\psi} = O_p(n^{-1/2})$, whereas if λ and ψ are not orthogonal, $\hat{\lambda}_\psi - \hat{\lambda} = O_p(n^{-1/2})$; lack of parametrization-invariance is of lower order than the terms retained. In full exponential families, where ψ is a component of the canonical parameter, it can be verified that $\hat{\lambda} = \hat{\lambda}_\psi$ if λ is the parameter that is orthogonal to ψ, in which case (17) is the same as modified profile likelihood. Barndorff-Nielsen and McCullagh (1993) study the differences between adjusted and modified profile likelihood in more detail, and provide examples where they are equal, even if the parameter of interest and the nuisance parameter are not orthogonal.

While the derivation of modified profile likelihood builds on the p^* approximation, for more accurate inference we might hope to have a p^* approximation for $\hat{\psi}$, instead of simply for $\hat{\theta}$. A version of this can be constructed, but in order that it be relevant for inference, it is necessary to consider more carefully the existence of conditional or marginal densities that eliminate the nuisance parameter, either exactly or approximately. The details are outlined in Barndorff-Nielsen and Cox (1994, §6.3). It turns out to be easier and more useful to calculate an approximation to the distribution function, described in the next section.

5. Related developments

5.1. Tail area approximations

For many purposes an approximation to the distribution function is more useful than an approximation to the density, as it provides both approximate p-values for testing a specified null hypotheses, and approximate confidence intervals at any desired level of confidence. The integration of the p^* approximation also provides further insight into the foundations of inference.

The approximation to the distribution function was developed in Barndorff-Nielsen (1986a) and Barndorff-Nielsen (1991), and came to be called the r^* approximation. It built on earlier work based on the saddlepoint approximation (Lugannani and Rice, 1980, Skovgaard, 1987, Daniels, 1987).

Suppose first that $\theta \in \mathbb{R}$, and consider evaluating

$$\int_{-\infty}^{\hat{\theta}^o} p^*(\hat{\theta} \mid a^o; \theta)d\hat{\theta} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\hat{\theta}^o} \exp\{\ell(\theta;\hat{\theta},a^o) - \ell(\hat{\theta};\hat{\theta},a^o)\}|j(\hat{\theta};\hat{\theta},a^o)|^{1/2}d\hat{\theta}; \tag{18}$$

this provides a one-sided p-value for testing θ, and as a function of θ recovers the so-called confidence distribution of Cox (1958c). Let $-r^2/2 = \ell(\theta;\hat{\theta},a^o) - \ell(\hat{\theta};\hat{\theta},a^o)$. The change of variables from $\hat{\theta}$ to r

is

$$-r\,dr = \{\ell_{;\hat\theta}(\theta;\hat\theta,a^o) - \ell_{\hat\theta;}(\hat\theta;\hat\theta,a^o) - \ell_{;\hat\theta}(\hat\theta;\hat\theta,a^o)\}d\hat\theta = \{\ell_{;\hat\theta}(\theta;\hat\theta,a^o) - \ell_{;\hat\theta}(\hat\theta;\hat\theta,a^o)\}d\hat\theta,$$

where

$$\ell_{;\hat\theta}(\theta;\hat\theta,a^o) = \frac{\partial\ell(\theta;\hat\theta,a^o)}{\partial\hat\theta}, \quad \ell_{\hat\theta;}(\hat\theta;\hat\theta,a^o) = \left.\frac{\partial\ell(\theta;\hat\theta,a^o)}{\partial\theta}\right|_{\theta=\hat\theta}. \tag{19}$$

The first expression in (19) is a sample-space derivative of the log-likelihood function, and its inclusion is inevitable if we want to evaluate the cumulative distribution function of R. The second is the score function evaluated at the maximum likelihood estimate, hence equal to zero. Thus

$$\int_{-\infty}^{\hat\theta^o} p^*(\hat\theta \mid a^o;\theta)d\hat\theta = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{r^o}\frac{r\exp(-r^2/2)}{u(r)}dr, \tag{20}$$

with

$$u(r) = u(\hat\theta;\theta) = \{\ell_{;\hat\theta}(\hat\theta;\hat\theta,a^o) - \ell_{;\hat\theta}(\theta;\hat\theta,a^o)\}|j(\hat\theta;\hat\theta,a^o)|^{-1/2}. \tag{21}$$

Barndorff-Nielsen (1990) described $u(r)$ as a type of dual score statistic. The upper limit in (20), $r^o = r(\hat\theta^o;\theta)$, is defined by $r^{o2}/2 = \ell(\hat\theta^o;\hat\theta^o,a^o) - \ell(\theta;\hat\theta^o,a^o)$; the sign attached to the square root is that of $u(r)$. The final step is to express the integrand of (20) as $\exp(-r^{*2}/2)$, with

$$r^* = r + \frac{1}{r}\log\left(\frac{u(r)}{r}\right), \tag{22}$$

and to verify that $|dr/dr^*| = 1 + O(n^{-1})$, obtaining the r^* approximation

$$\mathrm{pr}(\hat\theta \le \hat\theta^o \mid a^o;\theta) \doteq \Phi\big(r^*(\hat\theta^o;\theta)\big), \tag{23}$$

where $\Phi(\cdot)$ is the standard normal distribution function. The standard normal distribution is an approximation to both the conditional distribution of r^* given $A = a^o$ and the unconditional distribution. The implication is that an analytical form for the ancillary statistic is unnecessary to apply (22); all that is required is a evaluation of the partial derivative in (19), or, in the more general notation in (2),

$$\frac{\partial\ell(\theta;y(\hat\theta,a^o))}{\partial y}\frac{\partial y(\hat\theta,a^o)}{\partial\hat\theta}. \tag{24}$$

In two special cases r^* reduces to standard quantities. If the log-likelihood function is that for the canonical exponential model with $d_\theta = 1$, as at (7),

$$\ell(\theta;s) = s\theta - \kappa(\theta),$$

and no ancillary complement is needed. Also $s = n\kappa(\hat\theta)$, from which we have

$$\ell_{;\hat\theta}(\theta;s) = \theta|ds/d\hat\theta| = \theta\kappa''(\hat\theta),$$

and $u = (\hat\theta - \theta)j^{1/2}(\hat\theta)$, the Wald statistic for assessing the value θ; it is asymptotically equivalent to the log-likelihood root, r. If the log-likelihood function is of location-model form, then $y = y(\hat\theta,a)$ with $a = (a_1,\ldots,a_n) = (y_1 - \hat\theta,\ldots,y_n - \hat\theta)$, and $\ell(\theta;y(\hat\theta,a)) = \log f(a + \hat\theta - \theta)$, and $u = \ell'(\theta)j^{-1/2}(\hat\theta)$ is the standardized score statistic for assessing the value θ.

The dependence of the expression r^*, and hence the approximation $\Phi(r^*)$ on the sample-space derivative is in marked contrast to the approximation of a Bayesian posterior. A Laplace approximation to the posterior density is

$$\pi(\theta \mid y) \doteq \frac{|j(\hat\theta)|^{1/2}}{\sqrt{2\pi}} \frac{\pi(\theta)}{\pi(\hat\theta)} \frac{L(\theta)}{L(\hat\theta)},$$

and the integral of this can be approximated using the same transformation $-r^2/2 = \ell(\theta) - \ell(\hat\theta)$, but now the change of variable from θ to r involves only $\partial\ell(\theta; y)/\partial\theta$. One conclusion from this is that the approximate posterior distribution function cannot agree with the r^* approximation to higher order than $O(n^{-1/2})$, unless a term related to the sample space derivative in (21) is incorporated into the prior.

If $\theta \in \mathbb{R}^{d_\theta}$, we assume as in §4 that $\theta = (\psi, \lambda)$, with $\psi \in \mathbb{R}$ the parameter of interest. The development of the r^* approximation proceeds similarly to that outlined above. In Barndorff-Nielsen (1991) and Barndorff-Nielsen and Cox (1994, §6.6) this is achieved by using the p^* approximation twice, once to approximate the density of $\hat\theta = (\hat\psi, \hat\lambda)$ and again to approximate the density of $\hat\lambda_\psi$, the constrained maximum likelihood estimator in the model that treats ψ as fixed. The result is an r^* approximation of the same form as (22) and (23), with

$$u(r) = \frac{|\ell_{;\hat\theta}(\hat\theta) - \ell_{;\hat\theta}(\hat\theta_\psi) \quad \ell_{\lambda;\hat\theta}(\hat\theta_\psi)|}{|\ell_{\theta;\hat\theta}(\hat\theta)|} \left\{ \frac{|j(\hat\theta)|}{|j_{\lambda\lambda}(\hat\theta_\psi)|} \right\}^{1/2}, \tag{25}$$

where $\hat\theta_\psi = (\psi, \hat\lambda_\psi)$. The matrix in the numerator of the first factor in u is $d_\theta \times d_\theta$. Its first column is analogous to the expression for u in the scalar parameter case (21). The remaining $d_\theta - 1$ columns are composed of mixed nuisance-parameter derivatives. Evaluation of the determinant then essentially takes the dual score statistic in the first column and subtracts the regression on these nuisance parameter derivatives. The expressions involving the full Fisher information $j(\hat\theta)$ and the sub-matrix for the nuisance parameter $j_{\lambda\lambda}(\hat\theta_\psi)$ are the volume elements for the full and constrained maximum likelihood estimators.

If the model is an exponential family with canonical parameter $\theta = (\psi, \lambda)$, then it can be shown that

$$u(r) = (\hat\psi - \psi) j_{\mathrm{p}}^{1/2}(\hat\psi) \rho(\hat\psi, \psi),$$

and if the model is a location family with location parameter $\theta = (\psi, \lambda)$ then

$$u(r) = \ell_{\mathrm{p}}'(\psi) j_{\mathrm{p}}^{-1/2}(\hat\psi) \rho^{-1}(\hat\psi, \psi).$$

Here $\ell_{\mathrm{p}}(\psi) = \ell(\psi, \hat\lambda_\psi)$ is the profile log-likelihood function, $j_{\mathrm{p}}(\psi) = -\ell_{\mathrm{p}}''(\psi)$ is the associated Fisher information, and

$$\rho(\psi, \hat\psi) = \frac{|j_{\lambda\lambda}(\hat\psi, \hat\lambda)|^{1/2}}{|j_{\lambda\lambda}(\psi, \hat\lambda_\psi)|^{1/2}}. \tag{26}$$

This latter factor makes an adjustment to the Wald or score statistic to reflect uncertainty in estimation of the nuisance parameter λ. The similarity of these expressions, and a discussion of the adjustment term from r to r^* as composed of a nuisance component and an information component is given in Pierce and Peters (1992) and Barndorff-Nielsen and Cox (1994, §6.6).

As noted at (24), it is not necessary to have an explicit expression for the approximate ancillary statistic, as long as the partial derivatives in the expression for u can be evaluated. Fraser (1990) used this observation to define instead a surface tangent to the sample space contours defined by fixing

$A = a^o$, and derived these contours from a local location model (Fraser, 1964). This approach to the r^* approximation is reviewed in Davison and Reid (2024); several examples of its application are provided in Brazzale, Davison and Reid (2007).

Various simplifications of the expression for $u(r)$ have been proposed in the literature. The version due to Barndorff-Nielsen and Chamberlin (1994) is described in Barndorff-Nielsen and Cox (1994, §8.4). The approximation to $u(r)$ developed by Skovgaard (1996) also leads to a version of r^* with relative error $O(n^{-1})$ instead of $O(n^{-3/2})$, while avoiding sample-space derivatives. Skovgaard's approximation is also valid in large deviation regions, i.e. $\hat{\psi} - \psi = O(1)$, not just normal deviation regions $\hat{\psi} - \psi = O(n^{-1/2})$. Pierce and Bellio (2017) provide a software implementation of Skovgaard's approximation, with several examples. Comparison to the approach of Fraser (1990) is discussed in Reid and Fraser (2010). Estimation of the moments needed to implement Skovgaard's approximation was proposed in Severini (1999).

5.2. Inference for vector parameters

For inference about a vector parameter, the likelihood ratio statistic

$$w(\theta) = 2\{\ell(\hat{\theta}) - \ell(\theta)\}, \tag{27}$$

has a limiting $\chi^2_{d_\theta}$ distribution, and the relative error of the approximation is

$$f_W(w;\theta) = q_{d_\theta}(w)\{1 + O(n^{-1})\},$$

where $q_d(\cdot)$ is the density of the χ^2_d distribution. The apparent improvement from $O(n^{-1/2})$ to $O(n^{-1})$ is slightly misleading, as it is an artifact of the two-sided nature of the approximation. The nuisance parameter version

$$W(\psi) = 2\{\ell_{\mathrm{p}}(\hat{\psi}) - \ell_{\mathrm{p}}(\psi)\}, \tag{28}$$

similarly is approximated by the $\chi^2_{d_\psi}$ density to $O(n^{-1})$.

As usual, $\ell(\theta) = \ell(\theta; \hat{\theta}, a)$ so the conditional distribution of w given $a = a^o$ can be derived from the p^* approximation to the distribution of $\hat{\theta}$, offering the possibility of improving the χ^2 approximation. This was developed in Barndorff-Nielsen and Cox (1984). The result can be expressed via Bartlett correction: writing

$$\mathbb{E}_\theta\{W(\theta)\} = d_\theta\{1 + B(\theta)/n + O(n^{-2})\},$$

and

$$\widetilde{W}(\theta) = W(\theta)/\{1 + B(\theta)/n\}, \tag{29}$$

the result is

$$f_{\widetilde{W}|A}(\widetilde{w};\theta \mid a^o) = q_{d_\theta}(\widetilde{w})\{1 + O(n^{-3/2})\}; \tag{30}$$

i.e. the χ^2 approximation is improved by one order of magnitude on rescaling w by the leading term of its expected value. The error of approximation is unchanged if $B(\theta)$ is estimated by $B(\hat{\theta})$, or, in the case of the profile version (28), by $B(\psi, \hat{\lambda}_\psi)$. Barndorff-Nielsen and Hall (1988) show that the relative error is in fact $O(n^{-2})$, whether or not $B(\theta)$ is estimated.

The derivation of (30) from p^* leads to a direct relationship between $B(\theta)$ and the normalizing constant in (2). Explicit expressions for $B(\theta)$ are available in terms of the moments of various log-likelihood derivatives, but these are somewhat cumbersome to use in practice. A more direct approach

is to estimate the expected value of W using the parametric bootstrap, as suggested by Bickel and Ghosh (1990). Skovgaard (2001) proposed two alternative versions to $\widetilde{W}$, motivated by the approximation to R^* that he developed in Skovgaard (1996).

The statistics $W(\psi_0)$, $\widetilde{W}(\psi_0)$, and Skovgaard's version all provide omnibus tests of the hypothesis $H_0 : \psi = \psi_0$, against all alternatives $\psi \neq \psi_0$. Directional tests of H_0 were developed from p^* approximations in Davison et al. (2014) and Fraser, Reid and Sartori (2016), building on Skovgaard (1988). The directional approach essentially creates a one-parameter sub-hypothesis about the length of the parameter of interest, and uses a version of the r^* approximation to compute a directional p-value. Potential limitations of this approach are noted in Jensen (2021).

6. Conclusion

Ole Barndorff-Nielsen made many contributions to the theory of inference; we have emphasized here three inferential objects that are most closely associated with him: p^*, r^*, and ℓ_{mp}. Among these, the most important practical tool is the r^* approximation, an inferential quantity for which conditioning on ancillary statistics and adjusting for nuisance parameters is more or less automatic, and computation is relatively straightforward. In addition to his own many contributions on these topics, his work inspired colleagues in Denmark and around the world to develop a wide range of related ideas, and as is clear in his autobiographical memoir (Barndorff-Nielsen, 2018), he was inspired in return.

We have inevitably omitted many of these contributions that helped to advance this field. The surveys by Skovgaard (2001) and Reid (2003) give complementary overviews of the main ideas. McCullagh (1987) continues to be a key resource for a more detailed treatment.

Some important extensions that we have omitted include theoretical advances in the geometry of models (Barndorff-Nielsen and Blaesild, 1993, Barndorff-Nielsen and Jupp, 1997, Blaesild, 1994, Jupp, 2024), development of approximation theory for more complex data structures, including time series and stochastic processes (Jensen, 1987, 1995, Sweeting, 1992), and many results linking higher order approximations to parametric bootstrap resampling (DiCiccio and Efron, 1996, Lee and Young, 2005, 2016, DiCiccio, Kuffner and Young, 2017).

Some non-standard applications of r^* include non-regular models (Ritz and Skovgaard, 2005, Elkantassi et al., 2023), an application to satellite conjunction analysis (Elkantassi and Davison, 2022), hypothesis testing for Gaussian graphical models (DiCaterina, Reid and Sartori, 2025), and directional tests in linear models (McCormack et al., 2019).

It was a pleasure and an honour to review a small portion of Ole Barndorff-Nielsen's contributions to stochastics in science, and to have the opportunity to refer to his memoir (Barndorff-Nielsen, 2018), which we highly recommend, not least for its reflection of his warm personality and infectious enthusiasm for science.

Acknowledgments

We thank Mark Podolskij for the invitation to contribute to this volume, and the scientific committee of the Ole E. Barndorff-Nielsen memorial conference, held in Aarhus in May 2024, for their invitation to speak. Material more closely resembling that talk, and emphasising different aspects to those detailed in this more extensive discussion, can be found in Battey (2024).

Funding

The first author was supported by Natural Sciences and Engineering Research Council of Canada Grant RGPIN-2020-05897. The second author was supported by a UK Engineering and Physical Sciences Research Fellowship EP/T01864X/1.

References

Barndorff-Nielsen, O.E. (1980). Conditionality resolutions. *Biometrika* **67** 293–310. MR0581727 https://doi.org/10.1093/biomet/67.2.293

Barndorff-Nielsen, O.E. (1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrika* **70** 343–365. MR0712023 https://doi.org/10.1093/biomet/70.2.343

Barndorff-Nielsen, O.E. (1984). On conditionality resolution and the likelihood ratio for curved exponential models. *Scand. J. Stat.* **11** 157–170. MR0767240

Barndorff-Nielsen, O.E. (1986a). Inference on full or partial parameters based on the standardized signed log likelihood ratio. *Biometrika* **73** 307–322. MR0855891 https://doi.org/10.2307/2336207

Barndorff-Nielsen, O.E. (1986b). Likelihood and observed geometries. *Ann. Statist.* **14** 856–873. MR0856794 https://doi.org/10.1214/aos/1176350038

Barndorff-Nielsen, O.E. (1987). Contribution to discussion of "Parameter orthogonality and approximate conditional inference". *J. Roy. Statist. Soc. Ser. B, Methodol.* **49** 18–20.

Barndorff-Nielsen, O.E. (1990). Approximate interval probabilities. *J. Roy. Statist. Soc. Ser. B, Methodol.* **52** 485–496. MR1086800

Barndorff-Nielsen, O.E. (1991). Modified signed log-likelihood ratio. *Biometrika* **78** 557–563. MR1130923 https://doi.org/10.1093/biomet/78.3.557

Barndorff-Nielsen, O.E. (2018). *Stochastics in Science: Some Autobiographical Notes*. Højbjerg: Forlaget Morgenstjerne.

Barndorff-Nielsen, O.E. and Blaesild, P. (1993). Orthogeodesic models. *Ann. Statist.* **21** 1018–1039. MR1232530 https://doi.org/10.1214/aos/1176349162

Barndorff-Nielsen, O.E. and Chamberlin, S.R. (1994). Stable and invariant adjusted directed likelihoods. *Biometrika* **81** 485–499. MR1311092 https://doi.org/10.1093/biomet/81.3.485

Barndorff-Nielsen, O.E. and Cox, D.R. (1979). Edgeworth and saddlepoint approximations with statistical applications (with discussion). *J. Roy. Statist. Soc. Ser. B, Methodol.* **41** 279–312. MR0557595

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, Methodol.* **46** 483–495. MR0790634

Barndorff-Nielsen, O.E. and Cox, D.R. (1989). *Asymptotic Techniques for Use in Statistics*. London: Chapman & Hall. MR1010226 https://doi.org/10.1007/978-1-4899-3424-6

Barndorff-Nielsen, O.E. and Cox, D.R. (1994). *Inference and Asymptotics*. London: Chapman & Hall. MR1317097 https://doi.org/10.1007/978-1-4899-3210-5

Barndorff-Nielsen, O.E. and Hall, P. (1988). On the level-error after Bartlett adjustment of the likelihood ratio statistic. *Biometrika* **75** 374–378. MR0946056 https://doi.org/10.1093/biomet/75.2.374

Barndorff-Nielsen, O.E. and Jupp, P.E. (1997). Statistics, yokes and symplectic geometry. *Ann. Fac. Sci. Toulouse Math.* **6** 389–427. MR1610891

Barndorff-Nielsen, O.E. and McCullagh, P. (1993). A note on the relation between modified profile likelihood and the Cox-Reid adjusted profile likelihood. *Biometrika* **80** 321–328. MR1243507 https://doi.org/10.1093/biomet/80.2.321

Barndorff-Nielsen, O.E. and Wood, A.T.A. (1998). On large deviations and choice of ancillary for p^* and r^*. *Bernoulli* **4** 35–63. MR1611871 https://doi.org/10.2307/3318531

Bartlett, M.S. (1936). Statistical information and properties of sufficiency. *Proc. R. Soc. Lond. A* **154** 124–137. MR0024103

Bartlett, M.S. (1937). Properties of sufficiency and statistical tests. *Proc. R. Soc. Lond. A* **160** 268–282. MR0024103

Battey, H.S. (2024). Maximal co-ancillarity and maximal co-sufficiency. *Inf. Geom.* **7** 355–369.

Bickel, P.J. and Ghosh, J.K. (1990). A decomposition for the likelihood ratio statistic and the Bartlett correction—a Bayesian argument. *Ann. Statist.* **18** 1070–1090. MR1062699 https://doi.org/10.1214/aos/1176347740

Blaesild, P. (1994). Maximum likelihood estimation in exponential orthogeodesic models. *Ann. Inst. Statist. Math.* **46** 43–55. MR1272746 https://doi.org/10.1007/BF00773591

Brazzale, A.R., Davison, A.C. and Reid, N. (2007). *Applied Asymptotics: Case Studies in Small-Sample Statistics*. Cambridge: Cambridge University Press. MR2342742 https://doi.org/10.1017/CBO9780511611131

Cox, D.R. (1958a). The regression analysis of binary sequences. *J. Roy. Statist. Soc. Ser. B, Methodol.* **20** 215–242. MR0099097

Cox, D.R. (1958b). Two further applications of a model for binary regression. *Biometrika* **45** 562–565.

Cox, D.R. (1958c). Some problems connected with statistical inference. *Ann. Math. Stat.* **29** 357–372. MR0094890 https://doi.org/10.1214/aoms/1177706618

Cox, D.R. (1968). Notes on some aspects of regression analysis (with discussion). *J. Roy. Statist. Soc., Ser. A, Statist. Soc.* **131** 265–279.

Cox, D.R. (1970). *The Analysis of Binary Data*. London: Methuen. MR0282453

Cox, D.R. (1980). Local ancillarity. *Biometrika* **67** 279–286. MR0581725 https://doi.org/10.1093/biomet/67.2.279

Cox, D.R. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference (with discussion). *J. Roy. Statist. Soc. Ser. B, Methodol.* **49** 1–39. MR0893334

Daniels, H.E. (1954). Saddlepoint approximations in statistics. *Ann. Math. Stat.* **25** 631–650. MR0066602 https://doi.org/10.1214/aoms/1177728652

Daniels, H.E. (1987). Tail probability approximations. *Int. Stat. Rev.* **55** 37–48. MR0962940 https://doi.org/10.2307/1403269

Davison, A.C. and Reid, N. (2024). The tangent exponential model. In *Handbook of Bayesian, Fiducial, and Frequentist Inference* (J.O. Berger, X.L. Meng, N. Reid and M. Xie, eds.) 210–237. Boca Raton: Chapman and Hall/CRC.

Davison, A.C., Fraser, D.A.S., Reid, N. and Sartori, N. (2014). Accurate directional inference for vector parameters in linear exponential families. *J. Amer. Statist. Assoc.* **109** 302–314. MR3180565 https://doi.org/10.1080/01621459.2013.839451

DiCaterina, C., Reid, N. and Sartori, N. (2025). Directional tests in Gaussian graphical models. *Statist. Sinica* **35** 361–387. https://doi.org/10.5705/ss.202022.0394. MR4853618 https://doi.org/10.5705/ss.202022.0394

DiCiccio, T.J. and Efron, B. (1996). Bootstrap confidence intervals. *Statist. Sci.* **11** 189–228. MR1436647 https://doi.org/10.1214/ss/1032280214

DiCiccio, T.J., Kuffner, T.A. and Young, G.A. (2017). The formal relationship between analytic and bootstrap approaches to parametric inference. *J. Statist. Plann. Inference* **191** 81–87. MR3679110 https://doi.org/10.1016/j.jspi.2017.05.007

Efron, B. and Hinkley, D.V. (1978). Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information. *Biometrika* **65** 457–482. MR0521817 https://doi.org/10.1093/biomet/65.3.457

Elkantassi, S. and Davison, A.C. (2022). Space oddity? A statistical formulation of conjunction assessment. *J. Guid. Control Dynamics* **45** 2258–2274.

Elkantassi, S., Bellio, R., Brazzale, A.R. and Davison, A.C. (2023). Improved inference for a boundary parameter. *Canad. J. Statist.* **51** 780–799. MR4635242 https://doi.org/10.1002/cjs.11791

Fisher, R.A. (1934). Two new properties of mathematical likelihood. *Proc. R. Soc. Lond. A* **144** 285–307.

Fraser, D.A.S. (1964). Local conditional sufficiency. *J. Roy. Statist. Soc. Ser. B, Methodol.* **26** 52–62. MR0169327

Fraser, D.A.S. (1990). Tail probabilities from observed likelihoods. *Biometrika* **77** 65–76. MR1049409 https://doi.org/10.1093/biomet/77.1.65

Fraser, D.A.S., Reid, N. and Sartori, N. (2016). Accurate directional inference for vector parameters. *Biometrika* **103** 625–639. MR3551788 https://doi.org/10.1093/biomet/asw022

Hinkley, D.V. (1980). Likelihood as approximate pivotal. *Biometrika* **67** 287–292. MR0581726 https://doi.org/10.1093/biomet/67.2.287

Jensen, J.L. (1987). On asymptotic expansions in non-ergodic models. *Scand. J. Stat.* **14** 305–318. MR0943291

Jensen, J.L. (1995). *Saddlepoint Approximations*. Oxford: Oxford University Press. MR1354837

Jensen, J.L. (2021). On the use of saddlepoint approximations in high dimensional inference. *Sankhyā* **83-A** 379–392. MR4227217 https://doi.org/10.1007/s13171-019-00188-x

Jupp, P.E. (2024). A parameterisation-invariant modification of the score test. *Inf. Geom.* **7** 429–439. MR4682880 https://doi.org/10.1007/s41884-023-00101-4

Lee, S.M.S. and Young, G.A. (2005). Parametric bootstrapping with nuisance parameters. *Statist. Probab. Lett.* **71** 143–153. MR2126770 https://doi.org/10.1016/j.spl.2004.10.026

Lee, S.M.S. and Young, G.A. (2016). Distribution of likelihood-based p-values under a local alternative hypothesis. *Biometrika* **103** 641–652. MR3551789 https://doi.org/10.1093/biomet/asw021

Lugannani, R. and Rice, S. (1980). Saddlepoint approximation for the distribution of the sum of independent random variables. *Adv. Appl. Probab.* **12** 475–490. MR0569438 https://doi.org/10.2307/1426607

McCormack, A., Reid, N., Sartori, N. and Theivendran, S.A. (2019). A directional look at F-tests. *Canad. J. Statist.* **47** 619–627. MR4035792 https://doi.org/10.1002/cjs.11515

McCullagh, P. (1984). Local sufficiency. *Biometrika* **71** 233–244. MR0767151 https://doi.org/10.1093/biomet/71.2.233

McCullagh, P. (1987). *Tensor Methods in Statistics* (2nd Edition Dover 2018). London: Chapman & Hall. MR0907286

Pierce, D.A. and Bellio, R. (2017). Modern likelihood-frequentist inference. *Int. Stat. Rev.* **85** 519–541. MR3723615 https://doi.org/10.1111/insr.12232

Pierce, D.A. and Peters, D. (1992). Practical use of higher order asymptotics for multiparameter exponential families (with discussion). *J. Roy. Statist. Soc. Ser. B, Methodol.* **54** 701–738. MR1185218

Reid, N. (2003). Asymptotics and the theory of inference. *Ann. Statist.* **31** 1695–1731. MR2036388 https://doi.org/10.1214/aos/1074290325

Reid, N. and Fraser, D.A.S. (2010). Mean loglikelihood and higher-order approximations. *Biometrika* **97** 159–170. MR2594424 https://doi.org/10.1093/biomet/asq001

Ritz, C. and Skovgaard, I.M. (2005). Likelihood ratio tests in curved exponential families with nuisance parameters present only under the alternative. *Biometrika* **92** 507–517. MR2202642 https://doi.org/10.1093/biomet/92.3.507

Sartori, N. (2003). Modified profile likelihoods in models with stratum nuisance parameters. *Biometrika* **90** 533–549. MR2006833 https://doi.org/10.1093/biomet/90.3.533

Severini, T.A. (1999). An empirical adjustment to the likelihood ratio statistic. *Biometrika* **86** 235–347. MR1705371 https://doi.org/10.1093/biomet/86.2.235

Skovgaard, I.M. (1985). A second-order investigation of asymptotic ancillarity. *Ann. Statist.* **13** 534–551. MR0790555 https://doi.org/10.1214/aos/1176349537

Skovgaard, I.M. (1987). Saddlepoint expansions for conditional distributions. *J. Appl. Probab.* **24** 875–887. MR0913828 https://doi.org/10.2307/3214212

Skovgaard, I.M. (1988). Saddlepoint expansions for directional test probabilities. *J. Roy. Statist. Soc. Ser. B, Methodol.* **50** 269–280. MR0964180

Skovgaard, I.M. (1990a). On the density of minimum contrast estimators. *Ann. Statist.* **18** 779–789. MR1056336 https://doi.org/10.1214/aos/1176347625

Skovgaard, I.M. (1990b). *Analytic Statistical Models*. *IMS Monograph Series*. Hayward, CA. MR1218396

Skovgaard, I.M. (1996). An explicit large-deviation approximation to one-parameter tests. *Bernoulli* **2** 145–165. MR1410135 https://doi.org/10.2307/3318548

Skovgaard, I.M. (2001). Likelihood asymptotics. *Scand. J. Stat.* **28** 3–32. MR1844348 https://doi.org/10.1111/1467-9469.00223

Sørensen, M. (2026). Ole Barndorff-Nielsen: Sand, wind, and inference. Bernoulli.

Sweeting, T.J. (1992). Asymptotic ancillarity and conditional inference for stochastic processes. *Ann. Statist.* **20** 580–589. MR1150364 https://doi.org/10.1214/aos/1176348542

Received June 2025