Gradient statistic: higher-order asymptotics and Bartlett-type correction

Tiago M. Vargas, Silvia L.P. Ferrari, Artur J. Lemonte

Departamento de Estatística, Universidade de São Paulo, São Paulo/SP, Brazil

Abstract

We obtain an asymptotic expansion for the null distribution function of the gradient statistic for testing composite null hypotheses in the presence of nuisance parameters. The expansion is derived using a Bayesian route based on the shrinkage argument described in Ghosh and Mukerjee (1991). Using this expansion, we propose a Bartlett-type corrected gradient statistic with chi-square distribution up to an error of order $o(n^{-1})$ under the null hypothesis. Further, we also use the expansion to modify the percentage points of the large sample reference chi-square distribution. A small Monte Carlo experiment and various examples are presented and discussed.

Key-words: Asymptotic expansion; Bartlett-type correction; Bayesian route; Gradient statistic; Shrinkage argument.

1 Introduction

The most common hypothesis tests for large samples are the likelihood ratio (Wilks, 1938), the Wald (Wald, 1943), and the Rao score (Rao, 1948) tests. These tests are widely used in areas such as economics, biology, and engineering, among others, since exact tests are not always available. An alternative test uses the gradient statistic recently proposed by Terrell (2002). An advantage of the gradient statistic over the Wald and the score statistics is that it does not involve knowledge of the information matrix, neither expected nor observed. Additionally, the gradient statistic is quite simple to be computed. This has been emphasised by C.R. Rao (Rao, 2005), who wrote: ‘The suggestion by Terrell is attractive as it is simple to compute. It would be of interest to investigate the performance of the [gradient] statistic’.

Let $x_1, \ldots, x_n$ be a random sample of size $n$ with joint probability density function $f(\cdot; \theta)$, which depends on a $p$-dimensional vector of unknown parameters $\theta = (\theta_1, \ldots, \theta_p)^\top$. Let $\ell(\theta) = n^{-1} \sum_{i=1}^n \log f(x_i; \theta)$ and $U(\theta) = \partial\ell(\theta)/\partial\theta$ be the log-likelihood function and the score vector,
respectively; notice that, for convenience, both are divided by $n$. We wish to test the null hypothesis $H_0: \theta_1 = \theta_{10}$ against the two-sided alternative hypothesis $H_a: \theta_1 \neq \theta_{10}$, where $\theta_{10}$ is a fixed $q$-dimensional vector, $\theta_1 = (\theta_1, \ldots, \theta_q)\top$ and $\theta_2 = (\theta_{q+1}, \ldots, \theta_p)\top$. The partition in $\theta$ induces the corresponding partition in $U(\theta)$: $U(\theta) = (U_1(\theta)\top, U_2(\theta)\top)\top$. Let $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2)\top$ and $\tilde{\theta} = (\theta_{10}, \tilde{\theta}_2)\top$ be the unrestricted and the restricted (under $H_0$) maximum likelihood estimators of $\theta = (\theta_1^\top, \theta_2^\top)^\top$, respectively. The gradient statistic for testing $H_0$ is defined as

$$S = nU(\tilde{\theta})\top(\hat{\theta} - \tilde{\theta}),$$

and can also be written as $S = nU_1(\tilde{\theta})\top(\hat{\theta}_1 - \theta_{10})$, since $U_2(\tilde{\theta}) = 0$. Like the likelihood ratio, the Wald, and the score statistics, the gradient statistic has an asymptotic $\chi^2_q$ distribution under the null hypothesis, $q$ being the number of restrictions imposed by $H_0$.

Equation (1) is the inner product of the score vector evaluated at $H_0$ and the difference between the unrestricted and the restricted maximum likelihood estimators of $\theta$. Although the gradient statistic was derived by Terrell (2002) from the score and the Wald statistics, it is of a different nature. The score statistic measures the squared length of the score vector evaluated at $H_0$ using the metric given by the inverse of the Fisher information matrix, whereas the Wald statistic gives the squared distance between the unrestricted and the restricted maximum likelihood estimators of $\theta$ using the metric given by the Fisher information matrix. Moreover, both are quadratic forms. The gradient statistic, on the other hand, is not a quadratic form and measures the distance between the unrestricted and the restricted maximum likelihood estimators of $\theta$ from a different perspective. It measures the orthogonal projection of the score vector at $H_0$ on the vector $\hat{\theta} - \tilde{\theta}$.

Recently, the gradient test has been the subject of some research papers. In particular, Lemonte and Ferrari (2012a) obtained the local power of the gradient test under Pitman alternatives (a sequence of alternative hypotheses converging to the null hypothesis at the rate of $n^{-1/2}$). The authors compared the local power of the gradient test with those of the likelihood ratio, the Wald, and the score tests. They showed that none of the tests is uniformly more powerful than the others, and therefore, the gradient test is not only very simple to be calculated but it is also competitive with the others in terms of local power. Other recent works in which the gradient test is investigated are Lemonte (2011, 2012) and Lemonte and Ferrari (2011, 2012b,c).

The main result in Lemonte and Ferrari (2012a) regarding the local power of the gradient test up to an error of order $o(n^{-1/2})$ represents the first step in the study of higher order asymptotic properties of the gradient test. In the present paper, we wish to go further by focusing on deriving the second-order approximation to the null distribution of the gradient statistic. In other words, our aim is to obtain an asymptotic expansion for the cumulative distribution function of the gradient statistic under the null hypothesis up to an error of order $o(n^{-1})$.

The usual route for deriving expansions for the distribution of asymptotic chi-square test statistics involves multivariate Edgeworth series expansions. Although such a route has been followed by many authors, it is extremely lengthy and tedious (see, for example, Hayakawa, 1977; Harris, 1985).
Here, on the other hand, in order to derive an asymptotic expansion for the null distribution of the gradient statistic up to order $n^{-1}$, we follow a Bayesian route based on a shrinkage argument originally suggested by Ghosh and Mukerjee (1991) and described later in Mukerjee and Reid (2000). Although it uses a Bayesian approach, this technique can be used to solve frequentist problems, such as the derivation of Bartlett corrections and tail probabilities (Datta and Mukerjee, 2003).

Additionally, we obtain a Bartlett-type correction factor for the gradient statistic from the results in Cordeiro and Ferrari (1991). Under the null hypothesis, the corrected statistic is distributed as chi-square up to an error of order $o(n^{-1})$, while the uncorrected gradient statistic has a chi-square distribution up to an error of order $o(n^{-1/2})$; that is, the Bartlett-type correction factor makes the approximation error be reduced from $o(n^{-1/2})$ to $o(n^{-1})$. For a detailed survey on Bartlett and Bartlett-type corrections, the reader is referred to Cordeiro and Cribari-Neto (1996).

The paper unfolds as follows. In Section 2, we present our main results, namely an asymptotic expansion for the cumulative distribution function of the gradient statistic and its Bartlett-type correction. In Sections 3 and 4, we particularise our general results to one-parameter families and to families with two orthogonal parameters, respectively. A small Monte Carlo study is also presented in Section 4. Section 5 closes the paper with a brief discussion. Technical details are collected in two appendices.

2 The main result

First, let us introduce some notation. Let $D_j = \partial/\partial \theta_j$ ($j = 1, \ldots, p$) be the differential operator. We define $U_j = D_j \ell(\theta)$, $U_{jr} = D_j D_r \ell(\theta)$, $U_{jrs} = D_j D_r D_s \ell(\theta)$, and so on. We make the same assumptions, such as the regularity of the first four derivatives of $\ell(\theta)$ with respect to $\theta$ and the existence and uniqueness of the maximum likelihood estimator of $\theta$, as those fully outlined by Hayakawa (1977). Let $\kappa_{j,r} = E(U_j U_r)$, $\kappa_{j,r} = E(U_{j,r})$, $\kappa_{j,rs} = E(U_{j,rs})$, $\kappa_{j,rsu} = E(U_{j,rsu})$, $\kappa_{j,rs}$, $\kappa_{j,ru}$, $\kappa_{j,rs}$, $\kappa_{j,ru}$, $\kappa_{j,rs}$, $\kappa_{j,ru}$, etc., denote the cumulants of log-likelihood derivatives. The cumulants are not functionally independent, for instance, $\kappa_{j,r} = -\kappa_{j,r}$, $\kappa_{j,ra} + \kappa_{j,ra} = \kappa_{j,ra}^{(s)}$, $\kappa_{j,rsu} + \kappa_{j,rsu} = \kappa_{j,rsu}^{(j)}$, $\kappa_{j,rsu} + \kappa_{j,rsu} = \kappa_{j,rsu} + \kappa_{j,rsu} = \kappa_{j,rsu} + \kappa_{j,rsu} = \kappa_{j,rsu} + \kappa_{j,rsu}$, where $\kappa_{j,ra}^{(s)} = D_s \kappa_{j,r}$ and $\kappa_{j,rsu}^{(j)} = D_j D_r \kappa_{j,rsu}$, etc. Relations among them were first obtained by Bartlett (1953a,b). Further, let $K$ be the Fisher information matrix

$$K = ((\kappa_{j,r})) = -((\kappa_{j,r})) = \begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{bmatrix},$$

with $K^{-1} = ((\kappa_{j,r}^{(j)}))$ denoting its inverse. Finally, define the matrices

$$A = ((a^{jr})) = \begin{bmatrix} 0 & 0 \\ 0 & K_{22}^{-1} \end{bmatrix}, \quad M = ((m^{jr})) = K^{-1} - A.$$
In what follows, we use the Einstein summation convention, where $\sum'$ denotes summation over all components of $\theta$; that is, the indices $j$, $r$, $s$, $k$, $l$ and $u$ range over 1 to $p$. We now establish the following theorem.

**Theorem 1.** The asymptotic expansion for the null distribution of the gradient statistic for testing $\mathcal{H}_0 : \theta = \theta_{10}$ against $\mathcal{H}_a : \theta \neq \theta_{10}$ is

$$
\Pr(S \leq x) = G_q(x) + \frac{1}{24n} \sum_{i=0}^{3} R_i G_{q+i}(x) + o(n^{-1}),
$$

(2)

where $G_z(x)$ is the cumulative distribution function of a chi-square random variable with $z$ degrees of freedom, $R_1 = 3A_3 - 2A_2 + A_1$, $R_2 = A_2 - 3A_3$, $R_3 = A_3$, $R_0 = -(R_1 + R_2 + R_3)$.

$$
A_1 = 3 \sum' \kappa_{jrskl} a_{lu} \left\{ 3m^k a_{rs} + m^r (\kappa^s k + 2a^{sk}) \right\} \\
+ 6 \sum' \kappa_{jrsum} m^r a_{s} - 6 \sum' \left( \kappa_{jrzu} + \kappa_{jrsum} \right) \left( m^r \kappa^s u + 2m^u a_{rs} \right) \\
+ 6 \sum' (\kappa_{klu} + \kappa_{klu}) \left[ 2 \left( \kappa_{jrsk} + \kappa_{jrsk} \right) \left( \kappa^s j \kappa^r k \kappa^l u - a^{sk} a^{jr} a^{l} \right) - \kappa_{jrs} \left\{ \left( \kappa_{sz} + a_{sz} \right) \left( \kappa^{jsk} \kappa^{l} r - a^{jk} a^{ls} \right) \right\} \\
+ m^r (a_{sk} a_{lu} + \kappa^{sk} \kappa^{jr} \kappa^{s} a_{lu}) + 2a_{rz} (\kappa^{jrsk} \kappa^{l} u - a^{jk} a^{lu}) + 2a_{rz} a^{l} m^s a_{ju} \right\} \\
+ 12 \sum' \left( \kappa_{jrzu} + \kappa_{jrsu} + \kappa_{jzu} + \kappa_{jzu} + \kappa_{jzu} \right) \left( \kappa^{jrsk} u, r - a^{js} a^{ur} \right),
$$

$$
A_2 = -3 \sum' \kappa_{jrs} \left( \kappa_{klu} \left( m^r (m^u a_{lu} + \frac{3}{4} m^k m^l m^s a_{su} + \frac{1}{2} m^k m^l m^s a_{su} \right) \\
- 2 \left( \kappa_{klu} + \kappa_{klu} \right) \left\{ m^u (\kappa^{jsk} \kappa^{l} r - a^{jk} a^{lr}) + m^r (\kappa^{sk} \kappa^{l} u - a^{sk} a^{lu}) \right\} \right) \\
+ 3 \sum' \left( \kappa_{jrzu} + \kappa_{jrzu} \right) m^r m^s u,
$$

$$
A_3 = \frac{1}{12} \sum' \kappa_{jrskl} a_{kl} \left( 9m^j m^k a_{lu} a^l + 6m^j m^l m^l m^s u \right).
$$

**Proof.** The proof is presented in Appendix 1.

Basically, in order to prove Theorem 1 we follow a Bayesian route based on a shrinkage argument. This argument is described in Appendix 2.

If the null hypothesis is simple, we have $q = p$, $A = 0$ and $M = K^{-1}$. Therefore, an immediate consequence of Theorem 1 is the following corollary.

**Corollary 1.** The asymptotic expansion for the null distribution of the gradient statistic for testing $\mathcal{H}_0 : \theta = \theta_{0}$ against $\mathcal{H}_a : \theta \neq \theta_{0}$ is given by (2) with $q = p$, $R_1 = 3A_3 - 2A_2 + A_1$, $R_2 = A_2 - 3A_3$,
\[ R_3 = A_3, \quad R_0 = -(R_1 + R_2 + R_3) \quad \text{and the A's are} \quad A_3 = \sum' \kappa_{jrs} \kappa_{klu} \left( 9 \kappa^{j,r} \kappa^{s,k} \kappa^{l,u} + 6 \kappa^{j,k} \kappa^{r,l} \kappa^{s,u} \right) / 12, \]

\[
A_1 = -6 \sum' (\kappa_{jrs} + \kappa_{jrs,u}) \kappa^{j,r} \kappa^{s,u} \\
+ 6 \sum' (\kappa_{klu} + \kappa_{kl,u}) \left\{ 2(\kappa_{jrs} + \kappa_{jrs,u}) \left( \kappa^{s,j} \kappa^{r,k} \kappa^{l,u} + \kappa^{s,k} \kappa^{l,j} \kappa^{r,u} \right) \\
- \kappa_{jrs} \left( \kappa^{s,u} \kappa^{j,k} \kappa^{l,r} + \kappa^{j,r} \kappa^{s,k} \kappa^{l,u} \right) \right\} \\
+ 12 \sum' (\kappa_{jrs} + \kappa_{jrs,u} + \kappa_{jrs,r} + \kappa_{jrs,u} + \kappa_{jrs,rs}) \kappa^{j,s} \kappa^{r,u}, \\
\]

\[
A_2 = -3 \sum' \kappa_{jrs} \left\{ \kappa_{klu} \left( \frac{3}{4} \kappa^{j,r} \kappa^{s,k} \kappa^{l,u} + \frac{1}{2} \kappa^{j,k} \kappa^{r,l} \kappa^{s,u} \right) \\
- 2(\kappa_{klu} + \kappa_{kl,u}) \left( \kappa^{s,u} \kappa^{j,k} \kappa^{l,r} + \kappa^{j,r} \kappa^{s,k} \kappa^{l,u} \right) \right\} \\
+ 3 \sum' (\kappa_{jrs} + 2\kappa_{jrs,u}) \kappa^{j,r} \kappa^{s,u}. \\
\]

We are now able to present a Bartlett-type corrected gradient statistic. A Bartlett-type correction is a multiplying factor, which depends on the statistic itself, that results in a modified statistic that follows a chi-square distribution with approximation error of order less than \( n^{-1} \). Cordeiro and Ferrari (1991) obtained a general formula for a Bartlett-type correction for a wide class of statistics that have a chi-square distribution asymptotically. A special case is when the cumulative distribution function of the statistic can be written as \( (2) \), independently of the coefficients \( R_1, R_2, \) and \( R_3 \). Hence, from Theorem 1 and the results in Cordeiro and Ferrari (1991), we have the following corollary.

**Corollary 2.** The modified statistic

\[
S^* = S \{ 1 - (c + bS + aS^2) \}, \tag{3}
\]

where

\[
a = \frac{A_3}{12nq(q + 2)(q + 4)}, \quad b = \frac{A_2 - 2A_3}{12nq(q + 2)}, \quad c = \frac{A_1 - A_2 + A_3}{12nq},
\]

has a \( \chi^2_q \) distribution up to an error of order \( o(n^{-1}) \) under the null hypothesis.

The factor \( 1 - (c + bS + aS^2) \) in (3) can be regarded as a Bartlett-type correction factor for the gradient statistic in such a way that the null distribution of \( S^* \) is better approximated by the reference \( \chi^2 \) distribution than the distribution of the uncorrected gradient statistic.

Instead of modifying the test statistic as in (3), we may modify the reference \( \chi^2 \) distribution using the inverse expansion formula in Hill and Davis (1968). To be specific, let \( \gamma \) be the desired level of the test, and \( x_{1-\gamma} \) be the \( 1 - \gamma \) percentile of the \( \chi^2 \) limiting distribution of the test statistic. From expansion (2), we have the following corollary.
Corollary 3. The asymptotic expansion for the $1 - \gamma$ percentile of $S$ to order $n^{-1}$ takes the form

$$z_{1-\gamma} = x_{1-\gamma} + \frac{1}{12n} \left[ \frac{A_3 x_{1-\gamma}}{q(q+2)(q+4)} \left\{ x_{1-\gamma}^2 + (q+4)x_{1-\gamma} + (q+2)(q+4) \right\} ight. \\
+ \frac{x_{1-\gamma}(x_{1-\gamma} + q + 2)}{q(q+2)} (A_2 - 3A_3) + \frac{x_{1-\gamma}}{q} (3A_3 - 2A_2 + A_1) \right], \tag{4}$$

where $\Pr(\chi_q^2 \geq x_{1-\gamma}) = \gamma$.

In general, equations (3) and (4) depend on unknown parameters. In this case, we can replace these unknown parameters by their maximum likelihood estimates obtained under $\mathcal{H}_0$. It should be noticed that the improved gradient test of the null hypothesis $\mathcal{H}_0$ may be performed in three ways: (i) by referring the corrected statistic $S^*$ in (3) to the $\chi^2_q$ distribution; (ii) by referring the gradient statistic $S$ to the approximate cumulative distribution function $\chi^2_q$; (iii) by comparing $S$ with the modified upper percentile in (4). These three procedures are equivalent to order $n^{-1}$.

Finally, the three moments, up to order $n^{-1}$ under the null hypothesis, of the gradient statistic are presented in the following corollary.

Corollary 4. The first three moments, up to order $n^{-1}$ under the null hypothesis, of the gradient statistics are

$$\mu'_1(S) = q + \frac{A_1}{12n}, \quad \mu_2(S) = 2q + \frac{A_1 + A_2}{3n},$$
$$\mu_3(S) = 8q + \frac{2(A_1 + 2A_2 + A_3)}{n}.$$

In the next sections, we consider some applications of the general results derived in this section in two special cases: a one-parameter model and a two-parameter model under orthogonality of parameters.

3 The one-parameter case

We initially assume that the model is indexed by a scalar unknown parameter, say $\phi$. The interest lies in testing the null hypothesis $\mathcal{H}_0 : \phi = \phi_0$ against $\mathcal{H}_a : \phi \neq \phi_0$, where $\phi_0$ is a fixed value. Let $\kappa_{(\phi)} = E(\partial^2 \ell(\phi)/\partial \phi^2)$, $\kappa_{(\phi)(\phi)} = E(\partial^3 \ell(\phi)/\partial \phi^3)$, $\kappa_{(\phi)(\phi)(\phi)} = E(\partial^4 \ell(\phi)/\partial \phi^4)$, $\kappa_{(\phi)} = \partial \kappa_{(\phi)} / \partial \phi$, $\kappa_{(\phi)(\phi)} = \partial^2 \kappa_{(\phi)} / \partial \phi^2$, and $\kappa_{(\phi)(\phi)(\phi)} = \partial^3 \kappa_{(\phi)} / \partial \phi^3$. The gradient statistic for testing $\mathcal{H}_0$ is $S = nU(\phi_0)(\hat{\phi} - \phi_0)$, where $\hat{\phi}$ is the maximum likelihood estimator of $\phi$. Here, $A_1$, $A_2$, and $A_3$ given in Corollary 1 reduce to

$$A_1 = \frac{6\kappa_{(\phi)}(2\kappa_{(\phi)} - \kappa_{(\phi)(\phi)}) + 12\kappa_{(\phi)}(\kappa_{(\phi)(\phi)} - 2\kappa_{(\phi)}^3)}{\kappa_{(\phi)}^3}, \tag{5}$$
$$A_2 = \frac{12\kappa_{(\phi)}(2\kappa_{(\phi)} - 3\kappa_{(\phi)(\phi)}) + 3\kappa_{(\phi)(\phi)}(5\kappa_{(\phi)(\phi)} - 16\kappa_{(\phi)}^3)}{4\kappa_{(\phi)}^3}, \tag{6}$$
\[ A_3 = -\frac{5\kappa_{\phi\phi\phi}^2}{4\kappa_{\phi\phi}^3}. \] (7)

We now present some examples.

**Example 1.** *(Exponential distribution)*

Let \( x_1, \ldots, x_n \) be a random sample of an exponential distribution with density

\[ f(x; \phi) = \frac{1}{\phi} e^{-x/\phi}, \quad x > 0, \quad \phi > 0. \]

Here, \( \kappa_{\phi\phi} = -\phi^{-2}, \kappa_{\phi\phi\phi} = 4\phi^{-3}, \) and \( \kappa_{\phi\phi\phi\phi} = -18\phi^{-4}. \) The gradient statistic assumes the form

\[ S = n(\bar{x} - \phi_0)^2/\phi_0^2, \]

where \( \bar{x} = n^{-1} \sum_{i=1}^n x_i, \) which equals the score statistic. It is easy to see that \( A_1 = 0, A_2 = 18, \) and \( A_3 = 20. \) The first three moments (up to order \( n^{-1} \)) of \( S \) are \( \mu_1(S) = 1, \mu_2(S) = 2 + 6/n, \) and \( \mu_3(S) = 8 + 112/n. \) A partial verification of our results can be accomplished by comparing the exact moments of \( S \) with the approximate moments given above. Since \( n\bar{X} \) has a gamma distribution with parameters \( n \) and \( 1/(n\phi), \) it can be shown that the first three exact moments of \( S \) are \( 1, 2 + 6/n, \) and \( 8 + 112/n + 120/n^2, \) respectively. These moments differ from the approximate moments obtained from Corollary \( \text{[4]} \) only in terms of order less than \( n^{-1}. \) The Bartlett-type corrected gradient statistic obtained from Corollary \( \text{[5]} \) is \( S^* = S\{1 - (3 - 11S + 2S^2)/(18n)\} \).

**Example 2.** *(One-parameter exponential family)*

Let \( x_1, \ldots, x_n \) be a random sample of size \( n \) in which each \( x_i \) has a distribution in the one-parameter exponential family with density

\[ f(x; \phi) = \frac{1}{\xi(\phi)} \exp\{-\alpha(\phi)d(x) + v(x)\}, \]

where \( \alpha(\cdot), v(\cdot), d(\cdot), \) and \( \xi(\cdot) \) are known functions. Also, \( \alpha(\cdot) \) and \( \xi(\cdot) \) are assumed to have first three continuous derivatives, with \( \xi(\cdot) > 0, \) \( \alpha'(\phi), \) and \( \beta'(\phi) \) being different from zero for all \( \phi \) in the parameter space, where \( \beta(\phi) = \xi'(\phi)/\{\xi(\phi)\alpha'(\phi)\}. \) Here, primes denote derivatives with respect to \( \phi. \) For instance, \( \beta' = \beta'(\phi) = d\beta(\phi)/d\phi. \) It can be shown that \( \kappa_{\phi\phi} = -\alpha'/\beta', \kappa_{\phi\phi\phi} = -(2\alpha''\beta' + \alpha'\beta'') \) and \( \kappa_{\phi\phi\phi\phi} = 3\alpha''\beta'' - 3\alpha''\beta' - \alpha'\beta'''. \) The gradient statistic takes the form

\[ S = n(\phi_0 - \hat{\phi})\alpha'(\phi_0)(\beta(\phi_0) + \bar{d}), \]

where \( \bar{d} = n^{-1} \sum_{i=1}^n d(x_i). \) From (5), (6), and (7), we can write

\[ A_1 = \frac{6}{\alpha'\beta'} \left\{ 2 \left( \frac{\beta''}{\beta'} \right)^2 + \frac{\alpha''\beta''}{\alpha'\beta'} - \frac{\beta''}{\beta'} \right\}, \]

\[ A_2 = \frac{3}{\alpha'\beta'} \left[ \frac{n}{\beta'} \left( \frac{4\alpha''}{\alpha'} - \frac{\beta''}{4\beta'} \right) + 3 \left\{ \frac{\alpha''}{\alpha'} \right\}^2 + \left( \frac{\beta''}{\beta'} \right)^2 - \left( \frac{\alpha''}{\alpha'} - \frac{\beta''}{\beta'} \right) \right], \]

\[ A_3 = \frac{5}{\alpha'\beta'} \left( \frac{\alpha''}{\alpha'} + \frac{\beta''}{2\beta'} \right)^2. \]

We now present some special cases.
1. Normal ($\phi > 0, \mu \in \mathbb{R}, x \in \mathbb{R}$):

- $\mu$ known: $\alpha(\phi) = 1/(2\phi), \xi(\phi) = \phi^{1/2}, d(x) = (x-\mu)^2, \text{and } v(x) = -\log(2\pi)/2$. We have $A_1 = 0, A_2 = 36, \text{and } A_3 = 40$. The first three moments of $S$ up to order $n^{-1}$ are $\mu'_1(S) = 1, \mu_2(S) = 2(1 + 6/n), \text{and } \mu_3(S) = 8(1 + 29/n)$. The Bartlett-corrected gradient statistic is $S^* = S\{1 - (1 - 11S/3 + 2S^2/3)/(3n)\}$.

- $\phi$ known: $\alpha(\mu) = -\mu/\phi, \xi(\mu) = \exp(\mu^2/2\phi), d(x) = x, \text{and } v(x) = -x^2/2 - \log(2\pi\phi)/2$. Here, $A_1 = A_2 = A_3 = 0$, as expected.

2. Inverse normal ($\phi > 0, \mu > 0, x > 0$):

- $\mu$ known: $\alpha(\phi) = \phi, \xi(\phi) = 1/\phi^{1/2}, \text{and } v(x) = -\log(2\pi x^3)/2$. Here, $A_1 = 24, A_2 = 30, \text{and } A_3 = 10$, and the three first moments of $S$ are $\mu'_1(S) = 1 + 2/n, \mu_2(S) = 2 + 18/n, \text{and } \mu_3(S) = 8 + 188/n$. The Bartlett-corrected gradient statistic takes the form $S^* = S\{1 - (S + 2)(S + 3)/(18n)\}$.

- $\phi$ known: $\alpha(\mu) = \phi(2\mu^2), \xi(\phi) = \exp(-\phi/\mu), d(x) = x, \text{and } v(x) = -\phi/(2x^2) + \log(2\pi x^3)/2$. We have $A_1 = 0, A_2 = 45\mu/\phi$. The first three approximate moments of $S$ are $\mu'_1(S) = 1, \mu_2(S) = 2 + 15\mu/(n\phi), \text{and } \mu_3(S) = 8 + 270\mu/(n\phi)$. Also, $S^* = S\{1 - S/(S - 5)/(4n\phi)\}$.

3. Gamma ($k$ known, $k > 0, \phi > 0, x > 0$): $\alpha(\phi) = \phi, \xi(\phi) = \phi^{-k}, d(x) = x, \text{and } v(x) = (k - 1)\log x - \log \Gamma(k), \text{where } \Gamma(\cdot) \text{ denotes the gamma function}$. We have $A_1 = 12/k, A_2 = 15/k, A_3 = 5/k, \text{and } \mu'_1(S) = 1 + 1/(nk), \mu_2(S) = 2 + 9/(nk), \text{and } \mu_3(S) = 8 + 94/(nk)$. Also, $S^* = S\{1 - (S + 2)(S + 3)/(36nk)\}$.

4. Truncated extreme value ($\phi > 0, x > 0$): $\alpha(\phi) = 1/\phi, \xi(\phi) = \phi, d(x) = \exp(x - 1)$, and $v(x) = x$. We have $A_1 = 0, A_2 = 12, A_3 = 20, \mu'_1(S) = 1, \mu_2(S) = 2 + 4/n, \mu_3(S) = 8 + 88/n$, and $S^* = S\{1 - (12 - 15S + 2S^2)/(18n)\}$.

5. Pareto ($\phi > 0, k > 0, k \text{ known, } x > k$): $\alpha(\phi) = 1 + \phi, \xi(\phi) = (\phi k^\phi)^{-1}, \text{and } v(x) = 0$. Here, $A_1 = 12, A_2 = 15, A_3 = 5, \mu'_1(S) = 1 + 1/n, \mu_2(S) = 2 + 9/n, \mu_3(S) = 8 + 94/n$, and $S^* = S\{1 - (S + 2)(S + 3)/(36n)\}$.

6. Power ($\theta > 0, \phi > 0, \theta \text{ known, } x > \phi$): $\alpha(\phi) = 1 - \phi, \xi(\phi) = \phi^{-\theta}\phi$, and $v(x) = 0$. The $A_1$'s, the first three approximate moments, and the Bartlett-type corrected statistic coincide with those obtained for the Pareto distribution.

7. Laplace ($\theta > 0, k \in \mathbb{R}, k \text{ known, } x \in \mathbb{R}$): $\alpha(\theta) = \theta^{-1}, \xi(\theta) = 2\theta, d(x) = |x-k|, \text{and } v(x) = 0$. We have $A_1 = 0, A_2 = 18, A_3 = 20, \mu'_1(S) = 1, \mu_2(S) = 2 + 6/n, \mu_3(S) = 8 + 112/n$, and $S^* = S\{1 - (3 - 11S + 2S^2)/(18n)\}$.
4 Models with two orthogonal parameters

The two-parameter families of distributions under orthogonality of the parameters (Cox and Reid, 1987), say $\phi$ and $\beta$, will be the subject of this section. The null hypothesis under test is $H_0: \phi = \phi_0$, where $\phi_0$ is a fixed value, and $\beta$ acts as a nuisance parameter. The orthogonality between $\phi$ and $\beta$ leads to considerable simplification in the formulas of $A_1$, $A_2$, and $A_3$. Here, $\kappa_{\phi,\beta} = E(\partial^3 \ell(\theta)/\partial \beta \partial \phi^2)$, $\kappa'_{\phi,\beta} = \partial \kappa_{\phi,\beta}/\partial \beta$, etc. After some algebra, we have

$$A_1 = A_1\phi + A_{1\phi,\beta}, \quad A_2 = A_2\phi + A_{2\phi,\beta}, \quad A_3 = -\frac{5\kappa_{\phi,\phi}^2}{4\kappa_{\phi,\phi}^2}, \quad (8)$$

where $A_1\phi$ and $A_2\phi$ are equal to $A_1$ and $A_2$ given in (5) and (6), respectively, and

$$A_{1\phi,\beta} = \frac{3\left\{4\kappa_{\phi,\phi,\beta}\kappa_{\phi,\phi}^{(3)} + \kappa_{\phi,\phi,\beta}(4\kappa_{\phi,\phi,\phi}^{(2)} - \kappa_{\phi,\phi,\phi})\right\}}{\kappa_{\phi,\phi}^2\kappa_{\beta,\beta}} + \frac{6\left(\kappa_{\phi,\phi,\beta,\beta} - 2\kappa_{\phi,\phi,\beta} - 2\kappa_{\phi,\phi,\phi}^{(2)}\right)}{\kappa_{\phi,\phi}\kappa_{\beta,\beta}} \kappa_{\phi,\phi,\beta}^2,$$

$$A_{2\phi,\beta} = \frac{3\left(3\kappa_{\phi,\phi,\beta}\kappa_{\phi,\phi,\phi}^2 + \kappa_{\phi,\phi,\beta}^2\right)}{\kappa_{\phi,\phi}^2\kappa_{\beta,\beta}}.$$

The expressions for $A_{1\phi,\beta}$ and $A_{2\phi,\beta}$ in (8) can be regarded as the additional contribution introduced in the expansion of the cumulative distribution function of the gradient statistic owing to the fact that $\beta$ is unknown and has to be estimated from the data. In the following, we present some examples.

Example 3. (Normal distribution)

Let $x_1, \ldots, x_n$ be a random sample from a normal distribution $N(\phi, \beta)$. The gradient statistic can be written in the form

$$S = n\frac{T_1/T_2}{1 + T_1/T_2},$$

where $T_1 = n(\bar{x} - \phi_0)^2$ and $T_2 = \sum_{i=1}^n (x_i - \bar{x})^2$, where $\bar{x} = n^{-1}\sum_{i=1}^n x_i$. Under the null hypothesis, $T_1/\beta$ and $T_2/\beta$ are independent with distributions $\chi_1^2$ and $\chi_{n-1}^2$, respectively. It can be shown that $n^{-1}S$ has a beta distribution with parameters $1/2$ and $(n - 1)/2$. The first three exact moments of $S$ are $1$, $2(n - 1)/(n + 2)$, and $8(n - 1)(n - 2)/\{(n + 2)(n + 4)\}$, respectively. Here, $A_1 = A_3 = 0$ and $A_2 = -18$. The first three approximate moments of $S$ are $\mu'(S) = 1$, $\mu_2(S) = 2 - 6/n$, and $\mu_3(S) = 8 - 72/n$. These moments differ from the approximate moments only by terms of order less than $n^{-1}$. The Bartlett-type corrected gradient statistic is $S^* = S\{(1 - (3 - S)/(2n)\}.$

Example 4. (Bivariate two-parameter exponential distribution)

Let $x_{11}, \ldots, x_{1n_1}$ and $x_{21}, \ldots, x_{2n_2}$ be two independent random samples from exponential distributions with means $\mu$ and $\phi\mu$, respectively. It can be shown that $\phi$ and $\beta = \mu\phi^{1/2}$ are globally
orthogonal. The parameter of interest is \( \phi \) – the ratio of the means – and the interest lies in testing \( H_0 : \phi = 1 \), which is equivalent to the equality of the two population means, against \( H_a : \phi \neq 1 \). We consider the balanced case \( (n_1 = n_2 = n/2, n \geq 2 \text{ even}) \). Let \( \bar{x}_1 \) and \( \bar{x}_2 \) be the sample means. The log-likelihood function can be written as

\[
\ell(\phi, \beta) = -\log \beta - \frac{\bar{x}_1}{2\beta \phi^{1/2}} - \frac{\bar{x}_2}{2\beta \phi^{1/2}}.
\]

The gradient statistic for testing \( H_0 \) takes the form

\[
S = \frac{n(\bar{x}_1 - \bar{x}_2)^2}{4\bar{x}_1 x},
\]

where \( \bar{x} = (\bar{x}_1 + \bar{x}_2)/2 \). The cumulants of log-likelihood derivatives are \( \kappa_{\phi\phi} = -1/(4\phi^2) \), \( \kappa_{\phi\beta} = 3/(4\phi^3) \), \( \kappa_{\beta\beta\beta} = -45/(16\phi^4) \), \( \kappa_{\beta\beta} = -1/\beta^2 \), \( \kappa_{\phi\beta} = 4/\beta^3 \), \( \kappa_{\phi\phi} = 0 \), \( \kappa_{\phi\beta} = 1/(4\beta^2) \), and \( \kappa_{\beta\beta\phi} = -1/(2\beta^2\phi^2) \). From (8), we have \( A_1 = 24 \), \( A_2 = 63 \), and \( A_3 = 45 \). The corrected gradient statistic becomes \( S^* = S\{1 - (S - 1)(S - 2)/(4n)\} \).

**Example 5.** (Two-parameter Birnbaum–Saunders distribution)

The two-parameter Birnbaum–Saunders distribution was proposed by Birnbaum and Saunders (1969) and has cumulative distribution function in the form \( G(x) = \Phi(v) \), with \( x > 0 \), where \( v = \phi^{-1}\rho(x/\bar{\phi}), \rho(z) = z^{1/2} - z^{-1/2}, \) and \( \Phi(\cdot) \) is the standard normal cumulative distribution function; \( \phi > 0 \) and \( \beta > 0 \) are the shape and scale parameters, respectively. We wish to test \( H_0 : \phi = \phi_0 \) against the alternative hypothesis \( H_a : \phi \neq \phi_0 \), where \( \phi_0 \) is a known positive constant. The gradient statistic to test \( H_0 \) is

\[
S = \frac{n(\bar{\phi} - \phi_0)}{\phi_0^3}\{s + r - (2 + \phi_0^2)\},
\]

where \( \bar{s} = (n\bar{\phi})^{-1}\sum_{i=1}^{n} x_i \), \( \bar{r} = \bar{\beta}n^{-1}\sum_{i=1}^{n} x_i^{-1} \), and \( \bar{\beta} \) is the maximum likelihood estimator of \( \beta \) obtained under \( H_0 \). We have \( \kappa_{\phi\phi} = -2/\phi^2, \kappa_{\phi\beta} = 0, \) and \( \kappa_{\beta\beta} = -\{1 + \phi(2\pi)^{-1/2}h(\phi)\}/(\phi^2\beta^2) \), where \( h(\phi) = \phi(\pi/2)^{1/2} - \pi e^{2/\phi^2}\{1 - \Phi(2/\phi)\} \). After some algebra, we obtain \( A_{1\phi} = -3, A_{2\phi} = 69/8, A_{2\beta\beta} = -45(2 + \phi^2)/[2\{1 + \phi(2\pi)^{-1/2}h(\phi)\}], A_3 = 125/8, \) and

\[
A_{1\phi\beta} = \frac{9 - 15\phi^2/2}{1 + \phi(2\pi)^{-1/2}h(\phi)} - \frac{3(\phi^2 + 2)}{2\{1 + \phi(2\pi)^{-1/2}h(\phi)\}^2}\left\{ -4(1 + \phi^2) + \frac{2(4 + \phi^2)h(\phi)}{\phi\sqrt{2\pi}} \right\}.
\]

Since the necessary quantities to obtain the \( A \)'s were derived, a Bartlett-corrected gradient statistic may be obtained from Corollary[2]. It is interesting to note that the \( A \)'s do not depend on the unknown scalar parameter \( \beta \). Next, we shall present a small Monte Carlo simulation regarding the test of the null hypothesis \( H_0 : \phi = 1 \).

The simulations were performed by setting \( \beta = 1 \) and sample sizes ranging from 5 to 22 observations. All results are based on 10,000 replications. The size distortions (i.e. estimated minus nominal sizes) for the 5% nominal level of the gradient statistic and its Bartlett-corrected version for
different sample sizes are plotted in Figure 1(a). It is clear from this figure that the Bartlett-corrected test displays smaller size distortions than the original gradient test.

Finally, we set \( n = 10 \) and consider the first-order approximation (\( \chi^2_1 \) distribution) for the distribution of the gradient statistic and the expansion obtained in this paper. Figure 1(b) presents the curves. The difference between the curves is evident from this figure, and hence, the \( \chi^2_1 \) distribution may not be a good approximation for the null distribution of the gradient statistic in testing the null hypothesis \( H_0 : \phi = 1 \) for the two-parameter Birnbaum–Saunders model if the sample is small.

![Figure 1](image)

Figure 1: (a) Size distortion of the gradient test (solid) and the Bartlett-corrected gradient test (dashes); (b) first-order approximation (solid) and expansion to order \( n^{-1} \) (dashes) of the null cumulative distribution function of the gradient statistic.

5 Discussion

Lemonte and Ferrari (2012a) showed that the gradient test can be an interesting alternative to the classic large-sample tests, namely the likelihood ratio, the Wald, and the Rao score tests, since none is uniformly superior to the others in terms of second-order local power. Additionally, as remarked before, the gradient statistic does not require to obtain, estimate, or invert an information matrix, unlike the Wald and the Rao score statistics. Its formal simplicity is always an attraction.

The exact null distribution of the gradient statistic is usually unknown and the test relies upon an asymptotic approximation. The chi-square distribution is used as a large-sample approximation to the true null distribution of this statistic. However, for small sample sizes, the chi-square distribution may be a poor approximation to the true null distribution; that is, the asymptotic approximation may deliver inaccurate inference. In order to overcome this shortcoming, an alternative strategy is to use a
higher-order asymptotic theory.

The asymptotic expansion up to order $n^{-1}$ for the null distribution function of the gradient statistic was derived in this paper. A Bayesian route based on the shrinkage argument (Ghosh and Mukerjee, 1991; Mukerjee and Reid, 2000) proved to be extremely useful in this context. The expansion is very general in the sense that the null hypothesis can be composite in the presence of nuisance parameters. We show that the coefficients which define this expansion depend on the joint cumulants of log likelihood derivatives for the full data. Unfortunately, these coefficients are very difficult to interpret in generality.

Cordeiro and Ferrari (1991) showed that, quite generally, continuous statistics having a chi-square distribution asymptotically can be modified by a suitable correction term that makes the modified statistic have chi-square distribution to order $n^{-1}$. Their work can be viewed as an extension of Bartlett corrections to the likelihood ratio statistic (Lawley, 1956) to other statistics having a chi-square distribution asymptotically. The correction term comes from the coefficients of the $O(n^{-1})$ term in the expansion of the cumulative distribution function of the test statistic in such a way that it becomes better approximated by the reference chi-square distribution. It is known as the Bartlett-type correction. It is well known that Bartlett and Bartlett-type corrections have become a widely used method for improving the large-sample chi-square approximation to the null distribution of the likelihood ratio and Rao score statistics, respectively. In recent years there has been a renewed interest in Bartlett factors and several papers have been published giving expressions for computing these corrections for special models. Some references are Zucker et al. (2000), Lagos and Morettin (2004), Tu et al. (2005), van Giersbergen (2009), Bai (2009), Lagos et al. (2010), and Noma (2011).

From the general expansion derived in this paper and using results in Cordeiro and Ferrari (1991), we also obtained a Bartlett-type correction factor for the gradient statistic. Our results are very general and not tied to special classes of models. They allow the parameter vector to be multidimensional and are valid regardless of whether nuisance parameters are present or not. Additionally, as the coefficients in the expansion, and consequently in the Bartlett-type correction factor, are written as functions of cumulants of log-likelihood derivatives, they can be obtained for all the classes of parametric models for which those cumulants can be determined. Therefore, applications of our general results in several parametric models, such as the generalised linear models and extensions, can be studied in future research.

**Acknowledgments**

We gratefully acknowledge grants from FAPESP and CNPq (Brazil).
Appendix 1

Proof of Theorem 1

Except when indicated, the indices $j$, $r$, $s$, $u$, and $w$ range over 1 to $p$ and the indices $j'$, $r'$, $s'$, $u'$, and $w'$ range over 1 to $q$. Also, an array index repeated as both a superscript and a subscript indicates an implied summation over the appropriate range. Let $\Lambda_{jr} = -\psi_{jr} = -\{D_j D_r \ell(\theta)\}_{\theta = \hat{\theta}}$, $\psi_{jrs} = \{D_j D_r D_s \ell(\theta)\}_{\theta = \hat{\theta}}$, $\psi_{jrsu} = \{D_j D_r D_s D_u \ell(\theta)\}_{\theta = \hat{\theta}}$, etc. The matrix $\Lambda = ((\lambda_{jr}))$ is the observed information matrix evaluated at $\hat{\theta}$. The partition of $\theta = (\theta_1^\top, \theta_2^\top)^\top$ induces the partition

$$\Lambda = ((\lambda_{jr})) = \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{bmatrix}, \quad \Lambda^{-1} = ((\lambda^{jr})) = \begin{bmatrix} \Lambda_{11}^{-1} & \Lambda_{12}^{-1} \\ \Lambda_{21}^{-1} & \Lambda_{22}^{-1} \end{bmatrix},$$

where $\Lambda^{-1}$ is the inverse of $\Lambda$. Let $\Lambda_{11}^{-1} = ((\lambda_{uw}^{jr})), \sigma_{j}^{jr} = \lambda_{jr}^{wd} - \lambda_{jw}^{rd} \lambda_{uw}^{jr} \lambda_{ww}^{jr}$, $\tau^{jj'} = \lambda_{jj'}^{wd} \lambda_{ww}^{jr}$, $\sigma_{suw}^{suw} = \sigma_{su}^{su} \sigma_{uw}^{uw}$, $\lambda_{jrs'}^{uw} = \lambda_{jrs'}^{uw} [3]$, and $\lambda_{jrs'}^{uw} = \lambda_{jrs'}^{uw} \lambda_{uw}^{jr} \lambda_{w}^{jr}$ $[15]$, where $[\cdot]$ denotes a summation with the number of indices obtained by permutation of indices. For instance, $\sigma_{su}^{su} \sigma_{uw}^{uw} [3] = \sigma_{su}^{su} \sigma_{uw}^{uw} + \sigma_{su}^{su} \sigma_{uw}^{uw}$. Let $e = (e_1, \ldots, e_q)^T = n^{1/2}(\theta_1 - \hat{\theta}_1)$, $\Psi_j^{(4)} = \psi_{jrs} \sigma_{ss}^{rr} \tau^{jj'} / 2, \Psi_j^{(4)} = \psi_{jrs} \sigma_{ss}^{rr} \tau^{jj'} / 6$, $\Psi_{jrs}^{(3)}$.

Lemma 1. An asymptotic expansion under the null hypothesis for the gradient statistic $11$ is

$$S = e^\top \Lambda^{-1} e - \frac{3}{\sqrt{n}} \psi_{jrs}^{(3)} \epsilon_j^{ss} e_j e_{s'} e_{s'} - 4 \frac{\psi_{jrs}^{(4)} \epsilon_j^{ss} e_{s'} e_{s'} + 3 \psi_{jrs}^{(3)} \epsilon_j^{ss} e_{s'} e_{s'}}{\sqrt{n}} \epsilon_j^{ss} e_{s'} e_{s'} + o_p(n^{-1}). \quad (9)$$

Proof. Using a procedure analogous to that of Chang and Mukerjee (2011), the result holds. \hfill \square

Let $\pi = \pi(\theta)$ be a prior density for $\theta$, $\pi_j = D_j \pi(\theta)$, $\pi_{jr} = D_j D_r \pi(\theta)$, $\tilde{\pi} = \pi(\hat{\theta})$, $\tilde{\pi}_j = \pi_j(\hat{\theta})$, $\tilde{\pi}_{jr} = \pi_{jr}(\hat{\theta})$,

$$\psi_{jrs'}^{(2)} = \left\{ \frac{\tilde{\pi}_j}{2\pi} + \frac{1}{4} \psi_{jrsu}^{(3)} \epsilon_j^{ss} e_j e_{s'} e_{s'} + \frac{1}{24} \left( 2\psi_{jrs} \psi_{uwu} + 3\psi_{jrs} \psi_{sur} \right) \sigma_{suw}^{(1)} \right\} \tau^{jj'} \tau^{rr'},$$

$$\Gamma_j^{(1)} = \psi_{jrs'}^{(1)} + \frac{\tilde{\pi}_j}{2\pi} \tau^{jj'}, \quad \Gamma_j^{(2)} = \psi_{jrs'}^{(2)} + \frac{1}{2\pi} \left( \psi_{jrs} \tilde{\pi}_u + \psi_{jrs} \tilde{\pi}_r \right) \sigma_{suw}^{(1)} \tau^{jj'} \tau^{rr'},$$

$$\Gamma_j^{(4)} = \psi_{jrs'}^{(4)} + \frac{\tilde{\pi}_u}{6\pi} \psi_{jrs} \tau^{jj'} \tau^{rr'} \tau^{ss'} \tau^{uu'}.$$ From Ghosh and Mukerjee (1991), Chang and Mukerjee (2010) derive an expansion up to order $n^{-1}$ for the marginal posterior density of $\epsilon$, which takes the form

$$\pi_{post}(\epsilon) = \phi_q(\epsilon; \Lambda^{11}) \left[ 1 + \frac{1}{\sqrt{n}} \left( \Gamma_j^{(1)} \epsilon_j + \Gamma_j^{(3)} \epsilon_j e_j e_{s'} e_{s'} \epsilon_{s'} e_{s'} + \frac{1}{n} \left( \Gamma_j^{(2)} \epsilon_j e_j e_s - \lambda_j^{(1)} \epsilon_j e_j e_s e_{s'} - \lambda_j^{(2)} \epsilon_j e_j e_s e_{s'} e_{s'} - \lambda_j^{(3)} \epsilon_j e_j e_s e_{s'} e_{s'} e_{s'} - \lambda_j^{(4)} \epsilon_j e_j e_s e_{s'} e_{s'} e_{s'} e_{s'} \right) \right) + o(n^{-1}), \quad (10)$$

13
where \( \phi_q(z; \Sigma) \) denotes the density of the \( q \)-variate normal distribution with mean 0 and covariance matrix \( \Sigma \).

We now follow the Bayesian route described in [Mukerjee and Reid (2000); see Appendix 2.

**Step 1.** The approximate posterior characteristic function of \( S \) is

\[
M_\pi(t) = E_\pi \{ \exp(\xi S) \} = \int \exp(\xi S) \pi_{post}(\epsilon) d\epsilon,
\]

where \( \xi = i t \) with \( i = (-1)^{1/2} \). From Lemma [2] and after some algebra, we can write

\[
\exp(\xi S) \pi_{post}(\epsilon) = (1 - 2\xi)^{-q/2} \phi_q \left( \epsilon, \frac{\Lambda^{11}}{1 - 2\xi} \right) \left[ 1 + \frac{1}{\sqrt{n}} \left\{ (1 - 3\xi) \Psi_{j',r',s'}^{(3)} \epsilon_j' \epsilon_{j'} \epsilon_r \epsilon_s + \Gamma^{(1)}_{j',r'} \right\} \\
+ \frac{1}{n} \left( \frac{1}{2} \Psi_{j',r',s'}^{(3)} \Psi_{u',w'}^{(3)} \left[ \frac{1}{9} (1 - 3\xi)^2 \epsilon_j' \epsilon_r \epsilon_s \epsilon_u \epsilon_{u'} \epsilon_{w'} - \lambda_{j',r',s'}^{(2)} \right] \\
- \left[ \xi \left( 4 \Psi_{j',r',s'}^{(4)} + \Psi_{j',r',s'}^{(3)} \left( 3 \Gamma_{u'}^{(1)} - 4 \Psi_{u'}^{(1)} \right) \right) - \Gamma^{(4)}_{j',r',s'} \right] \epsilon_j' \epsilon_r \epsilon_s \epsilon_u \right] \\
+ \Gamma_{j',r'}^{(2)} (\epsilon_{j'} \epsilon_r - \lambda_{j'}^{(2)}) - \Gamma^{(4)}_{j',r',s'} \lambda_{j',r',s'}^{(1)} \right] + o_p(n^{-1}).
\]

Now, by writing \( \xi = -\frac{1}{2} (1 - 2\xi) + \frac{1}{4}, \xi^2 = \frac{1}{4} (1 - 2\xi)^2 - \frac{1}{2} (1 - 2\xi) + \frac{1}{4}, \) and assuming that \( \theta \) is in the interior of the support of \( \pi \), we obtain after some algebra

\[
M_{\tilde{\pi}}(t) = (1 - 2\xi)^{-q/2} \left\{ 1 + \frac{1}{n} \sum_{i=0}^{3} H_i (1 - 2\xi)^{-1} \right\} + o_p(n^{-1}),
\]

where \( H_0 = -(H_1 + H_2 + H_3) \),

\[
H_1 = \frac{9}{8} \Psi_{j',r',s'}^{(3)} \Psi_{u',w'}^{(3)} \lambda_{j',r',s'}^{(2)} \Gamma_{j',r',s'}^{(2)} \\
+ \lambda_{j',r',s'}^{(1)} \left\{ 2 \Psi_{j',r',s'}^{(4)} - \Psi_{j',r',s'}^{(3)} \Psi_{u'}^{(1)} + \frac{3}{2} \Psi_{j',r',s'}^{(3)} \Gamma_{u'}^{(1)} \right\},
\]

\[
H_2 = -\frac{3}{4} \Psi_{j',r',s'}^{(3)} \Psi_{u',w'}^{(3)} \lambda_{j',r',s'}^{(2)} \Gamma_{j',r',s'}^{(2)} \\
+ \lambda_{j',r',s'}^{(1)} \left\{ \Gamma_{j',r',s'}^{(4)} - 2 \Psi_{j',r',s'}^{(4)} - \Psi_{j',r',s'}^{(3)} \Psi_{u'}^{(1)} - \frac{3}{2} \Psi_{j',r',s'}^{(3)} \Gamma_{u'}^{(1)} \right\},
\]

\[
H_3 = \frac{1}{8} \Psi_{j',r',s'}^{(3)} \Psi_{u',w'}^{(3)} \lambda_{j',r',s'}^{(2)} \Gamma_{j',r',s'}^{(2)}.
\]

**Step 2.** Let \( \tilde{\pi}(\cdot) \) be an auxiliary prior density for \( \theta \) satisfying the conditions in [Bickel and Ghosh (1990)]. We now obtain an approximate posterior characteristic function of \( S \) under the prior \( \tilde{\pi}(\cdot) \), say \( M_{\tilde{\pi}}(t) \). From (11), we have

\[
M_{\tilde{\pi}}(t) = (1 - 2\xi)^{-q/2} \left\{ 1 + \frac{1}{n} \sum_{i=0}^{3} \tilde{H}_i (1 - 2\xi)^{-1} \right\} + o_p(n^{-1}),
\]

where \( \tilde{H}_i \) denotes the counterpart of \( H_i \) obtained by replacing \( \pi(\cdot) \) with \( \tilde{\pi}(\cdot) \). After some algebra, we have

\[
\Delta(\theta) = E_\theta(M_{\tilde{\pi}}) = (1 - 2\xi)^{-q/2} \left\{ 1 + \frac{1}{n} \sum_{i=0}^{3} \tilde{\bar{H}}(1 - 2\xi)^{-1} \right\} + o(n^{-1}),
\]

14
where $J_0 = -(J_1 + J_2 + J_3)$,

$$J_1 = \frac{1}{32} \kappa_{jrs} \kappa_{uvw} (9m^j r_m^{su} m^{vw} + 6m^j u_m^r m^{sw}) + \frac{1}{4} (\kappa_{jrsu} + 3\kappa_{jrv} \kappa_{suw} a^{uw}) m^j r_m^{su}$$

$$+ \frac{3}{8} \kappa_{jrs} \kappa_{uvw} a^{uw} m^j r_m^{su} + \frac{3}{4} \kappa_{jrs} m^j r m^u_a \frac{\bar{\pi}_u}{\pi}$$

$$+ \frac{1}{2} m^j r \left\{ \frac{\bar{\pi}_j}{\pi} + \frac{1}{2} \kappa_{jrsu} a^{su} + \frac{1}{4} (2\kappa_{jrs} \kappa_{uvw} + 3\kappa_{jrsu} \kappa_{vuw} a^{uw}) \right\},$$

$$J_2 = -\frac{1}{48} \kappa_{jrs} \kappa_{uvw} (9m^j r_m^{su} m^{vw} + 6m^j u_m^r m^{sw}) - \frac{1}{4} (\kappa_{jrsu} + 3\kappa_{jrv} \kappa_{suw} a^{uw}) m^j r_m^{su}$$

$$- \frac{3}{8} \kappa_{jrs} \kappa_{uvw} a^{uw} m^j r_m^{su} - \frac{3}{4} \kappa_{jrs} m^j r m^u_a \frac{\bar{\pi}_u}{\pi}$$

$$+ 3m^j r m^s u \left[ \frac{1}{24} (\kappa_{jrsu} + (2\kappa_{jrs} \kappa_{uvw} + 3\kappa_{jrv} \kappa_{suw}) a^{uw}) + \frac{1}{6} \kappa_{jrs} \frac{\bar{\pi}_u}{\pi} \right],$$

$$J_3 = \frac{1}{288} \kappa_{jrs} \kappa_{uvw} (9m^j r_m^{su} m^{vw} + 6m^j u_m^r m^{sw}).$$

**Step 3.** We now compute

$$\int \Delta(\theta) \bar{\pi}(\theta) d\theta = (1 - 2\xi)^{-q/2} \left\{ 1 + \frac{1}{n} \sum_{i=0}^{3} (1 - 2\xi)^{-i} \int \bar{J}_i(\theta) d\theta \right\} + o(n^{-1}),$$

by integrating the $J$’s with respect to $\bar{\pi}$. After integrating each term that depends on the prior distributions and by allowing $\bar{\pi}(\cdot)$ to converge weakly to the degenerate prior at the true value of $\theta$, we arrive at

$$E_\theta \{ \exp(\xi S) \} = (1 - 2\xi)^{-q/2} \left\{ 1 + \frac{1}{n} \sum_{i=0}^{3} \bar{A}_i(1 - 2\xi)^{-i} \right\} + o(n^{-1}),$$

where the $\bar{A}$’s are functions of cumulants of log-likelihood derivatives. By writing $d = 2\xi/(1 - 2\xi)$ and using the fact that $\sum_{i=0}^{3} \bar{A}_i = 0$, we arrive at

$$M(t) = (1 - 2\xi)^{-q/2} \left\{ 1 + \frac{1}{24n} (A_1 d + A_2 d^2 + A_3 d^3) \right\} + o(n^{-1}), \quad (12)$$

with $A_1 = 24(\bar{A}_1 + 2\bar{A}_2 + 3\bar{A}_3)$, $A_2 = 24(\bar{A}_2 + 3\bar{A}_3)$, and $A_3 = 24\bar{A}_3$. We can write

$$A_1 = 12D_J D_r m^j r m^j r m^s u - 6D_u (\kappa_{jrs} m^j r a^{su}) - 12D_u (\kappa_{jrs} m^j r m^r a^{su})$$

$$+ 6\kappa_{jrs} a^{su} + 3\kappa_{jrs} \kappa_{uvw} (m^j r m^s u a^{uw} + 2m^j r a^{su} a^{uw}) + 9\kappa_{jrsu} \kappa_{vuw} m^r a^{su} a^{uw},$$

$$A_2 = 6D_u (\kappa_{jrs} m^j r m^s u) - 3\kappa_{jrs} \kappa_{uvw} (m^j r m^s u a^{uw} + \frac{3}{4} m^j r m^s u m^{sw} + \frac{1}{2} m^j u m^r m^{sw})$$

$$- 3\kappa_{jrsu} m^j r m^s u - 9\kappa_{jrv} \kappa_{suw} m^j r m^s u a^{uw},$$

$$A_3 = \frac{1}{12} \kappa_{jrs} \kappa_{uvw} (9m^j r m^s u m^{uw} + 6m^j u m^r m^{sw}).$$

Inverting $M(t)$ in (12) and interchanging the indices in a suitable manner, after some algebra, we arrive at the expression for $A_1$, $A_2$, and $A_3$ as given in Theorem.
Appendix 2

The Shrinkage Argument

Let $x = (x_1, \ldots, x_n)^\top$ be a random vector with density $f(\cdot, \theta)$, where $\theta \in \Theta$ is a $p$-dimensional parameter and $\Theta \subseteq \mathbb{R}^p$ is an open subset of the Euclidean space. Let $Q(\cdot, \theta)$ be a measurable function. Assume that $Q$ is continuous for all $\theta$ and that its expectation exists. A Bayesian route for obtaining $E_{\theta}\{Q(\cdot, \theta)\}$ based on a shrinkage argument involves the three steps described below.

Step 1. Obtain $E_{\pi}\{Q(\theta, X)|X = x\}$, the posterior expectation of $Q$ under the prior $\pi(\cdot)$ for $\theta$.

Step 2. Find $E_{\theta}[E_{\pi}\{Q(\theta, X)|X = x\}] = \Delta(\theta)$, for $\theta \in \text{int}_s(\pi)$, where $\text{int}_s(\pi)$ denotes the interior of the support of $\pi$.

Step 3. Integrate $\Delta(\theta)$ with respect to $\pi(\cdot)$ and allow $\pi(\cdot)$ to converge weakly to the degenerate prior at $\theta$, where $\theta \in \text{int}_s(\pi)$. This yields $E_{\theta}\{Q(X, \theta)\}$.

A detailed justification can be found in Mukerjee and Reid (2000).

References

Bai, P. (2009). Sphericity test in a GMANOVA-MANOVA model with normal error. *Journal of Multivariate Analysis* **100**, 2305–2312.

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.

Bickel, P.J., Ghosh, J.K. (1990). A decomposition for the likelihood ratio statistic and the Bartlett correction - a Bayesian argument. *Annals of Statistics* **18**, 1070–1090.

Birnbaum, Z.W., Saunders, S.C. (1969). A new family of life distributions. *Journal of Applied Probability* **6**, 319–327.

Chang, H.I., Mukerjee, R. (2010). Highest posterior density regions with approximate frequentist validity: the role of data-dependent priors. *Statistics and Probability Letters* **80**, 1791–1797.

Chang, H.I., Mukerjee, R. (2011). Data-dependent probability matching priors for likelihood ratio and adjusted likelihood ratio statistics. *Statistics*. In press, DOI:10.1080/02331888.2011.587880.

Cordeiro, G.M., Ferrari, S.L.P. (1991). A modified score test statistic having chi-squared distribution to order $n^{-1}$. *Biometrika* **78**, 573–582.

Cordeiro, G.M., Cribari-Neto, F. (1996). On Bartlett and Bartlett-type corrections. *Econometric Reviews* **15**, 339–367.
Cox, D.R., Reid, N. (1987). Parameter orthogonality and approximate conditional inference (with discussion). *Journal of the Royal Statistical Society B* **40**, 1–39.

Datta, G.S., Mukerjee, R. (2003). *Probability Matching Priors: Higher Order Asymptotics*. Springer-Verlag: New York.

Ghosh, J.K., Mukerjee, R. (1991). Characterization of priors under which Bayesian and frequentist Bartlett corrections are equivalent in the multiparameter case. *Journal of Multivariate Analysis* **38**, 385–393.

Harris, P. (1985). An asymptotic expansion for the null distribution of the efficient score statistic. *Biometrika* **72**, 653–659.

Hayakawa, T. (1977). The likelihood ratio criterion and the asymptotic expansion of its distribution. *Annals of the Institute of Statistical Mathematics* **29**, 359–378.

Hill, G.W., Davis, A.W. (1968). Generalized asymptotic expansions of Cornish–Fisher type. *The Annals of Mathematical Statistics* **39**, 1264-73.

Lagos, B.M., Morettin, P.A. (2004). Improvement of the likelihood ratio test statistic in ARMA models. *Journal of Time Series Analysis* **25**, 83–101.

Lagos, B.M., Morettin, P.A., Barroso, L.P. (2010). Some corrections of the score test statistic for gaussian ARMA models *Brazilian Journal of Probability and Statistics* **24**, 434–456.

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

Lemonte, A.J. (2011). Local power of some tests in exponential family nonlinear models. *Journal of Statistical Planning and Inference* **141**, 1981–1989.

Lemonte, A.J. (2012). Local power properties of some asymptotic tests in symmetric linear regression models. *Journal of Statistical Planning and Inference* **142**, 1178–1188.

Lemonte, A.J., Ferrari, S.L.P. (2011). Size and power properties of some tests in the Birnbaum–Saunders regression model. *Computational Statistics and Data Analysis* **55**, 1109–1117.

Lemonte, A.J., Ferrari, S.L.P. (2012a). The local power of the gradient test. *Annals of the Institute of Statistical Mathematics* **64**, 373–381.

Lemonte, A.J., Ferrari, S.L.P. (2012b). A note on the local power of the LR, Wald, score and gradient tests. *Electronic Journal of Statistics* **6**, 421–434.

Lemonte, A.J., Ferrari, S.L.P. (2012c). Local power and size properties of the LR, Wald, score and gradient tests in dispersion models. *Statistical Methodology* **9**, 537–554.
Mukerjee, R., Reid, N. (2000). On the Bayesian approach for frequentist computations. *Brazilian Journal of Probability and Statistics* **14**, 159–166.

Noma, H. (2011). Confidence intervals for a random-effects meta-analysis based on Bartlett-type corrections. *Statistics in Medicine* **30**, 3304–3312.

Rao, C.R. (1948). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Proceedings of the Cambridge Philosophical Society* **44**, 50–57.

Rao, C.R. (2005). Score test: historical review and recent developments. In *Advances in Ranking and Selection, Multiple Comparisons, and Reliability*, N. Balakrishnan, N. Kannan and H. N. Nagaraja, eds. Birkhuser, Boston.

Terrell, G.R. (2002). The gradient statistic. *Computing Science and Statistics* **34**, 206–215.

Tu, D., Chen, J., Shi, P., Wu, Y. (2005). A Bartlett type correction for Rao’s score test in Cox regression model. *Sankhya* **67**, 722–735.

van Giersbergen, N.P.A. (2009). Bartlett correction in the stable AR(1) model with intercept and trend. *Econometric Theory* **25**, 857–872.

Wald, A. (1943). Tests of statistical hypothesis concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society* **54**, 426–482.

Wilks, S.S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypothesis. *Annals of Mathematical Statistics* **9**, 60–62.

Zucker, D.M., Lieberman, O., Manor, O. (2000). Improved small sample inference in the mixed linear model: Bartlett correction and adjusted likelihood. *Journal of the Royal Statistical Society B* **62**, 827–838.