Electronic Journal of Statistics
Vol. 7 (2013) 43–61
ISSN: 1935-7524
DOI: 10.1214/12-EJS763

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

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

Departamento de Estatística, Universidade de São Paulo, São Paulo/SP, Brazil
e-mail: tiagomoreiravargas@yahoo.com.br; silviaferrari.usp@gmail.com
arturlemonte@gmail.com

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 [10]. 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. Monte Carlo simulation experiments and various examples are presented and discussed.

Keywords and phrases: Asymptotic expansion, Bartlett-type correction, Bayesian route, gradient test, shrinkage argument.

Received November 2012.

Contents

1. Introduction

The most common statistical tests of a composite null hypothesis for large samples are the likelihood ratio [29], the Wald [28], and the Rao score [23] 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 [25]. 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 [24], 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 each x_i having probability density function $f(\cdot; \boldsymbol{\theta})$, which depends on a p-dimensional vector of unknown parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^\top$. Let $\ell(\boldsymbol{\theta}) = n^{-1} \sum_{i=1}^n \log f(x_i; \boldsymbol{\theta})$ and $\boldsymbol{U}(\boldsymbol{\theta}) = \partial \ell(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}$ 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 $\mathcal{H}_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_{10}$ against the two-sided alternative hypothesis $\mathcal{H}_a : \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_{10}$, where $\boldsymbol{\theta}_{10}$ is a fixed q-dimensional vector, $\boldsymbol{\theta}_1 = (\theta_1, \ldots, \theta_q)^\top$ and $\boldsymbol{\theta}_2 = (\theta_{q+1}, \ldots, \theta_p)^\top$. The partition in $\boldsymbol{\theta}$ induces the corresponding partition in $\boldsymbol{U}(\boldsymbol{\theta})$: $\boldsymbol{U}(\boldsymbol{\theta}) = (\boldsymbol{U}_1(\boldsymbol{\theta})^\top, \boldsymbol{U}_2(\boldsymbol{\theta})^\top)^\top$. Let $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\theta}}_1, \widehat{\boldsymbol{\theta}}_2)^\top$ and $\widetilde{\boldsymbol{\theta}} = (\boldsymbol{\theta}_{10}, \widetilde{\boldsymbol{\theta}}_2)^\top$ be the unrestricted and the restricted (under $\mathcal{H}_0$) maximum likelihood estimators of $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$, respectively. The gradient statistic for testing $\mathcal{H}_0$ is defined as

$$S = n \boldsymbol{U}(\widetilde{\boldsymbol{\theta}})^\top (\widehat{\boldsymbol{\theta}} - \widetilde{\boldsymbol{\theta}}). \tag{1}$$

It can also be expressed as $S = n \boldsymbol{U}_1(\widetilde{\boldsymbol{\theta}})^\top (\widehat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{10})$, since $\boldsymbol{U}_2(\widetilde{\boldsymbol{\theta}}) = \mathbf{0}$. Like the likelihood ratio, the Wald, and the score statistics, the gradient statistic has an asymptotic χ_q^2 distribution under the null hypothesis, q being the number of restrictions imposed by $\mathcal{H}_0$.

Although the gradient statistic was derived by [25] 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 $\mathcal{H}_0$ 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 $\boldsymbol{\theta}$ 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 $\boldsymbol{\theta}$ from a different perspective. It measures the orthogonal projection of the score vector at $\mathcal{H}_0$ on the vector $\widehat{\boldsymbol{\theta}} - \widetilde{\boldsymbol{\theta}}$. Unlike the score statistic and like the Wald statistic, the gradient statistic is not invariant under reparameterization of the model that preserves the parameter of interest.

Recently, the gradient test has been the subject of some research papers. In particular, [19] 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 $n^{-1/2}$); see also [20]. 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. Comparison among the local power of the classic tests and the gradient test in regression models can be found in [17, 18].

The main result in [19] 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, [12, 11]. 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 [10] and described later in [21]. 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 [9].

Additionally, we obtain a Bartlett-type correction factor for the gradient statistic from the results in [6]. 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 [7].

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. Monte Carlo simulation experiments are also presented in Section 4. Section 5 closes the paper with a brief discussion. Technical details are collected in two appendices.

2. Main results

First, let us introduce some notation. Let $D_j = \partial/\partial\theta_j$ $(j = 1, \ldots, p)$ be the differential operator. We define $U_j = D_j\ell(\boldsymbol{\theta})$, $U_{jr} = D_jD_r\ell(\boldsymbol{\theta})$, $U_{jrs} = D_jD_rD_s\ell(\boldsymbol{\theta})$, and $U_{jrsu} = D_jD_rD_sD_u\ell(\boldsymbol{\theta})$. We make the same assumptions, such as the regularity of the first four derivatives of $\ell(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ and the existence and uniqueness of the maximum likelihood estimator of $\boldsymbol{\theta}$, as those fully outlined by [12]. Let $\kappa_{jr} = E(U_{jr})$, $\kappa_{jrs} = E(U_{jrs})$, $\kappa_{jrsu} = E(U_{jrsu})$, $\kappa_{jr}^{(s)} = D_s\kappa_{jr}$, $\kappa_{jr}^{(su)} = D_sD_u\kappa_{jr}$, and $\kappa_{jrs}^{(u)} = D_u\kappa_{jrs}$. Note that all the κs are or order $O(1)$. Further, let $\boldsymbol{K}$ be the per observation Fisher information matrix

$$\boldsymbol{K} = -((\kappa_{jr})) = \begin{bmatrix} \boldsymbol{K}_{11} & \boldsymbol{K}_{12} \\ \boldsymbol{K}_{21} & \boldsymbol{K}_{22} \end{bmatrix},$$

with $\boldsymbol{K}^{-1} = -((\kappa^{jr}))$ denoting its inverse. Finally, define the matrices

$$\boldsymbol{A} = ((a^{jr})) = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{K}_{22}^{-1} \end{bmatrix}, \qquad \boldsymbol{M} = ((m^{jr})) = \boldsymbol{K}^{-1} - \boldsymbol{A}.$$

In what follows, we use the Einstein summation convention, where $\sum'$ denotes summation over all components of $\boldsymbol{\theta}$; that is, the indices j, r, s, k, l and u range over 1 to p. We now establish the following theorem.

Theorem 1. *An asymptotic expansion for the null distribution of the gradient statistic for testing $\mathcal{H}_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_{10}$ against $\mathcal{H}_a : \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_{10}$ is*

$$\Pr(S \le x) = G_q(x) + \frac{1}{24n} \sum_{i=0}^{3} R_i G_{q+2i}(x) + o(n^{-1}), \tag{2}$$

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

$$\begin{aligned} A_1 = {} & 3 \sum{}' \kappa_{jrs}\kappa_{klu} \left\{ m^{jr} a^{lu} (m^{sk} + 2a^{sk}) + a^{jr} m^{sk} a^{lu} + 2m^{jk} a^{rl} a^{su} \right\} \\ & - 12 \sum{}' \kappa_{jr}^{(s)} \kappa_{kl}^{(u)} \left(\kappa^{sj}\kappa^{rk}\kappa^{lu} + a^{sj} a^{rk} a^{lu} + \kappa^{sk}\kappa^{lj}\kappa^{ru} + a^{sk} a^{lj} a^{ru} \right) \\ & - 6 \sum{}' \kappa_{jrs}\kappa_{kl}^{(u)} \Big\{ \left(a^{su} - \kappa^{su}\right)\left(\kappa^{jk}\kappa^{lr} - a^{jk}a^{lr}\right) + m^{jr} \Big(a^{sk}a^{lu} + \kappa^{sk}\kappa^{lu} \Big) \\ & \qquad + 2a^{rs}\left(\kappa^{jk}\kappa^{lu} - a^{jk}a^{lu}\right) + 2a^{rk}a^{ls}m^{ju} \Big\} \\ & + 6 \sum{}' \kappa_{jrsu} m^{jr} a^{su} - 6 \sum{}' \kappa_{jrs}^{(u)} \left\{ m^{jr}\left(a^{su} - \kappa^{su}\right) + 2m^{ju}a^{rs} \right\} \\ & + 12 \sum{}' \kappa_{rs}^{(ju)} \left(\kappa^{jr}\kappa^{su} - a^{jr}a^{su} \right), \end{aligned}$$

$$\begin{aligned} A_2 = {} & -3 \sum{}' \kappa_{jrs}\kappa_{klu} \Big\{ m^{jr} m^{sk} a^{lu} + m^{jr} a^{sk} m^{lu} + 2m^{jk} m^{rl} a^{su} \\ & \qquad + \frac{1}{4}\left(3m^{jr}m^{sk}m^{lu} + 2m^{jk}m^{rl}m^{su}\right) \Big\} \\ & + 6 \sum{}' \kappa_{jrs}\kappa_{kl}^{(u)} \left\{ m^{su}\left(\kappa^{jk}\kappa^{lr} - a^{jk}a^{lr}\right) + m^{jr}\left(\kappa^{sk}\kappa^{lu} - a^{sk}a^{lu}\right) \right\} \\ & + 6 \sum{}' \kappa_{jrs}^{(u)} m^{jr} m^{su} - 3 \sum{}' \kappa_{jrsu} m^{jr} m^{su}, \end{aligned}$$

$$A_3 = \frac{1}{4} \sum{}' \kappa_{jrs}\kappa_{klu} \left(3m^{jr}m^{sk}m^{lu} + 2m^{jk}m^{rl}m^{su}\right).$$

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 $q = p$, $\boldsymbol{A} = \mathbf{0}$ and $\boldsymbol{M} = \boldsymbol{K}^{-1}$. Therefore, an immediate consequence of Theorem 1 is the following corollary.

Corollary 1. *An asymptotic expansion for the null distribution of the gradient statistic for testing* $\mathcal{H}_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ *against* $\mathcal{H}_a : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ *is given by* (2) *with* $q = p$, $R_1 = 3A_3 - 2A_2 + A_1$, $R_2 = A_2 - 3A_3$, $R_3 = A_3$, $R_0 = -(R_1 + R_2 + R_3)$ *and the* A*'s are*

$$\begin{aligned} A_1 = &-12 {\sum}' \kappa_{jr}^{(s)} \kappa_{kl}^{(u)} \left(\kappa^{sj}\kappa^{rk}\kappa^{lu} + \kappa^{sk}\kappa^{lj}\kappa^{ru}\right) \\ &+ 6 {\sum}' \kappa_{jrs}\kappa_{kl}^{(u)} \left(\kappa^{su}\kappa^{jk}\kappa^{lr} + \kappa^{jr}\kappa^{sk}\kappa^{lu}\right) \\ &+ 12 {\sum}' \kappa_{rs}^{(ju)}\kappa^{jr}\kappa^{su} - 6 {\sum}' \kappa_{jrs}^{(u)}\kappa^{jr}\kappa^{su}, \end{aligned}$$

$$\begin{aligned} A_2 = &\frac{3}{4} {\sum}' \kappa_{jrs}\kappa_{klu} \left(3\kappa^{jr}\kappa^{sk}\kappa^{lu} + 2\kappa^{jk}\kappa^{rl}\kappa^{su}\right) \\ &- 6 {\sum}' \kappa_{jrs}\kappa_{kl}^{(u)} \left(\kappa^{su}\kappa^{jk}\kappa^{lr} + \kappa^{jr}\kappa^{sk}\kappa^{lu}\right) \\ &+ 6 {\sum}' \kappa_{jrs}^{(u)}\kappa^{jr}\kappa^{su} - 3 {\sum}' \kappa_{jrsu}\kappa^{jr}\kappa^{su}, \end{aligned}$$

$$A_3 = -\frac{1}{4} {\sum}' \kappa_{jrs}\kappa_{klu}\left(3\kappa^{jr}\kappa^{sk}\kappa^{lu} + 2\kappa^{jk}\kappa^{rl}\kappa^{su}\right).$$

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 [6] 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 [6], we have the following corollary.

Corollary 2. *The modified statistic*

$$S^* = S\left\{1 - \left(c + bS + aS^2\right)\right\}, \tag{3}$$

where

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

has a χ_q^2 *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 χ^2 distribution than the distribution of the uncorrected gradient statistic.

Instead of modifying the test statistic as in (3), we may modify the reference χ^2 distribution using the inverse expansion formula in [13]. To be specific, let γ be the desired level of the test, and $x_{1-\gamma}$ be the $1 - \gamma$ percentile of the χ^2 limiting distribution of the test statistic. From expansion (2), we have the following corollary.

Corollary 3. *An asymptotic expansion for the $1-\gamma$ percentile of S takes the form*

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

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 χ_q^2 distribution; (ii) by referring the gradient statistic S to the approximate cumulative distribution function (2); (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 mean, the variance, and the third central moment, up to order n^{-1} under the null hypothesis, of the gradient statistic are*

$$\mu_1'(S) = q + \frac{A_1}{12n}, \qquad \mu_2(S) = 2q + \frac{A_1 + A_2}{3n},$$
$$\mu_3(S) = 8q + \frac{2(A_1 + 2A_2 + A_3)}{n},$$

respectively.

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 ϕ. The interest lies in testing the null hypothesis $\mathcal{H}_0 : \phi = \phi_0$ against $\mathcal{H}_a : \phi \neq \phi_0$, where ϕ_0 is a fixed value. Let $\kappa_{\phi\phi} = E(\partial^2\ell(\phi)/\partial\phi^2)$, $\kappa_{\phi\phi\phi} = E(\partial^3\ell(\phi)/\partial\phi^3)$, $\kappa_{\phi\phi\phi\phi} = E(\partial^4\ell(\phi)/\partial\phi^4)$, $\kappa_{\phi\phi}^{(\phi)} = \partial\kappa_{\phi\phi}/\partial\phi$, $\kappa_{\phi\phi\phi}^{(\phi)} = \partial\kappa_{\phi\phi\phi}/\partial\phi$, and $\kappa_{\phi\phi}^{(\phi\phi)} = \partial^2\kappa_{\phi\phi}/\partial\phi^2$. The gradient statistic for testing $\mathcal{H}_0$ is $S = nU(\phi_0)(\widehat{\phi} - \phi_0)$, where $\widehat{\phi}$ is the maximum likelihood estimator of ϕ. Here, A_1, A_2, and A_3 given in Corollary 1 reduce to

$$A_1 = \frac{6\kappa_{\phi\phi}(2\kappa_{\phi\phi}^{(\phi\phi)} - \kappa_{\phi\phi\phi}^{(\phi)}) + 12\kappa_{\phi\phi}^{(\phi)}(\kappa_{\phi\phi\phi} - 2\kappa_{\phi\phi}^{(\phi)})}{\kappa_{\phi\phi}^3}, \quad (5)$$

$$A_2 = \frac{12\kappa_{\phi\phi}(2\kappa_{\phi\phi\phi}^{(\phi)} - 3\kappa_{\phi\phi\phi\phi}) + 3\kappa_{\phi\phi\phi}(5\kappa_{\phi\phi\phi} - 16\kappa_{\phi\phi}^{(\phi)})}{4\kappa_{\phi\phi}^3}, \tag{6}$$

$$A_3 = -\frac{5\kappa_{\phi\phi\phi}^2}{4\kappa_{\phi\phi}^3}. \tag{7}$$

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

$$f(x;\phi) = \frac{1}{\phi}\mathrm{e}^{-x/\phi}, \qquad x > 0, \qquad \phi > 0.$$

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 4 only in terms of order less than n^{-1}. The Bartlett-type corrected gradient statistic obtained from Corollary 3 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 ϕ in the parameter space, where $\beta(\phi) = \xi'(\phi)/(\xi(\phi)\alpha'(\phi))$. Here, primes denote derivatives with respect to ϕ. For instance, $\beta' = \beta'(\phi) = \mathrm{d}\beta(\phi)/\mathrm{d}\phi$. 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 $S = n(\phi_0 - \widehat{\phi})\alpha'(\phi_0)(\beta(\phi_0) + \bar{d})$, where $\bar{d} = n^{-1}\sum_{i=1}^n d(x_i)$. From (5), (6), and (7), we can write

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

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

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

We now present some special cases.

1. Normal ($\phi > 0$, $\mu \in \mathbb{R}$, $x \in \mathbb{R}$):
   - μ known: $\alpha(\phi) = 1/(2\phi)$, $\xi(\phi) = \phi^{1/2}$, $d(x) = (x-\mu)^2$, and $v(x) = -\log(2\pi)/2$. We have $A_1 = 0$, $A_2 = 36$, and $A_3 = 40$. The first three moments of S up to order n^{-1} are $\mu_1'(S) = 1$, $\mu_2(S) = 2(1+6/n)$, and $\mu_3(S) = 8(1+29/n)$. The Bartlett-corrected gradient statistic is $S^* = S\{1-(1-11S/3+2S^2/3)/(3n)\}$.
   - ϕ known: $\alpha(\mu) = -\mu/\phi$, $\xi(\mu) = \exp(\mu^2/2\phi)$, $d(x) = x$, and $v(x) = -x^2/2 - \log(2\pi\phi)/2$. Here, $A_1 = A_2 = A_3 = 0$, as expected.
2. Inverse normal ($\phi > 0$, $\mu > 0$, $x > 0$):
   - μ known: $\alpha(\phi) = \phi$, $\xi(\phi) = 1/\phi^{1/2}$, $d(x) = (x-\mu)^2/(2\mu^2 x)$, and $v(x) = -\log(2\pi x^3)/2$. Here, $A_1 = 24$, $A_2 = 30$, and $A_3 = 10$, and the three first moments of S are $\mu_1'(S) = 1+2/n$, $\mu_2(S) = 2+18/n$, and $\mu_3(S) = 8+188/n$. The Bartlett-corrected gradient statistic takes the form $S^* = S\{1-(S+2)(S+3)/(18n)\}$.
   - ϕ known: $\alpha(\mu) = \phi/(2\mu^2)$, $\xi(\phi) = \exp(-\phi/\mu)$, $d(x) = x$, and $v(x) = -\phi/(2x^2) + \log(2\pi x^3)/2$. We have $A_1 = 0$ and $A_2 = A_3 = 45\mu/\phi$. The first three approximate moments of S are $\mu_1'(S) = 1$, $\mu_2(S) = 2+15\mu/(n\phi)$, and $\mu_3(S) = 8+270\mu/(n\phi)$. Also, $S^* = S\{1-\mu S(S-5)/(4n\phi)\}$.
3. 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 $S^* = S\{1-(12-15S+2S^2)/(18n)\}$.
4. Pareto ($\phi > 0$, $k > 0$, k known, $x > k$): $\alpha(\phi) = 1+\phi$, $\xi(\phi) = (\phi k^\phi)^{-1}$, 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 $S^* = S\{1-(S+2)(S+3)/(36n)\}$.
5. Power ($\theta > 0$, $\phi > 0$, θ known, $x > \theta$): $\alpha(\phi) = 1-\phi$, $\xi(\phi) = \phi^{-1}\theta^\phi$, and $v(x) = 0$. The A's, the first three approximate moments, and the Bartlett-type corrected statistic coincide with those obtained for the Pareto distribution.
6. Laplace ($\theta > 0$, $k \in \mathbb{R}$, k known, $x \in \mathbb{R}$): $\alpha(\theta) = \theta^{-1}$, $\zeta(\theta) = 2\theta$, $d(x) = |x-k|$, 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 [8], say ϕ and β, will be the subject of this section. The null hypothesis under test is $\mathcal{H}_0 : \phi = \phi_0$, where ϕ_0 is a fixed value, and β acts as a nuisance parameter. The orthogonality between ϕ and β leads to considerable simplification in the formulas of A_1, A_2, and A_3. Here, $\kappa_{\phi\phi\beta} = E(\partial^3\ell(\boldsymbol{\theta})/\partial\beta\partial\phi^2)$, $\kappa_{\phi\phi\beta}^{(\beta)} = \partial\kappa_{\phi\phi\beta}/\partial\beta$, etc. After some algebra, we have

$$A_1 = A_{1\phi} + A_{1\phi\beta}, \qquad A_2 = A_{2\phi} + A_{2\phi\beta}, \qquad A_3 = -\frac{5\kappa_{\phi\phi\phi}^2}{4\kappa_{\phi\phi}^3}, \tag{8}$$

where $A_{1\phi}$ and $A_{2\phi}$ are equal to A_1 and A_2 given in (5) and (6), respectively, and

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

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

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 β 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 for testing $\mathcal{H}_0 : \phi = \phi_0$ can be written in the form

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

where $T_1 = n(\bar{x} - \phi_0)^2$, $T_2 = \sum_{i=1}^n (x_i - \bar{x})^2$, and $\bar{x} = n^{-1}\sum_{i=1}^n x_i$. Under the null hypothesis, T_1/β and T_2/β are independent with distributions χ_1^2 and χ_{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 exact 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. *(Two-parameter Birnbaum–Saunders distribution)*

The two-parameter Birnbaum–Saunders distribution was proposed by [3] and has cumulative distribution function in the form $G(x) = \Phi(v)$, with $x > 0$, where $v = \phi^{-1}\rho(x/\beta)$, $\rho(z) = z^{1/2} - z^{-1/2}$, 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 $\mathcal{H}_0 : \phi = \phi_0$ against the alternative hypothesis $\mathcal{H}_a : \phi \neq \phi_0$, where ϕ_0 is a known positive constant. The gradient statistic to test $\mathcal{H}_0$ is

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

where $\bar{s} = (n\widetilde{\beta})^{-1}\sum_{i=1}^n x_i$, $\bar{r} = \widetilde{\beta}n^{-1}\sum_{i=1}^n x_i^{-1}$, and $\widetilde{\beta}$ is the maximum likelihood estimator of β obtained under $\mathcal{H}_0$. 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 $h(\phi) = \phi(\pi/2)^{1/2} - \pi \mathrm{e}^{2/\phi^2}\{1 - \Phi(2/\phi)\}$. After some algebra, we obtain $A_{1\phi} = -3$, $A_{2\phi} = 69/8$, $A_{2\phi\beta} = -45(2+\phi^2)/[2\{1 + \phi(2\pi)^{-1/2}h(\phi)\}]$, $A_3 = 125/8$, and

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

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 scale parameter β. Next, we shall present a small Monte Carlo simulation regarding the test of the null hypothesis $\mathcal{H}_0 : \phi = 1$.

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 test 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.

Next, we set $n = 10$ and consider the first-order approximation (χ_1^2 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

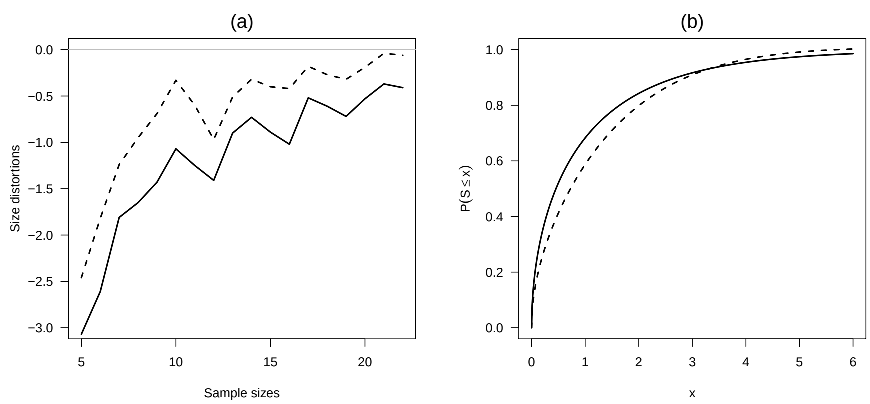

FIG 1. *(a) Size distortion of the gradient test (solid) and the Bartlett-corrected gradient test (dashes), (b) first-order approximation (solid) and expansion to order* n^{-1} *(dashes) of the null cumulative distribution function of the gradient statistic; Birnbaum–Saunders distribution.*

is evident from this figure, and hence, the χ_1^2 distribution may not be a good approximation for the null distribution of the gradient statistic in testing the null hypothesis $\mathcal{H}_0 : \phi = 1$ for the two-parameter Birnbaum–Saunders model if the sample is small.

Example 5. *(Gamma distribution)*

Let $x_1, \ldots, x_n$ be a random sample from a gamma distribution with mean β and coefficient of variation $\phi^{1/2}$. Here, we consider the problem of testing the null hypothesis $\mathcal{H}_0 : (\beta, \phi) = (\beta_0, \phi_0)$, where β_0 and ϕ_0 are fixed positive values. Note that the null hypothesis is simple, and the A's can be obtained from Corollary 1 with $q = p = 2$. After some algebra, we obtain

$$A_1 = 6(d_1 - d_2 + d_3 - 2d_4), \quad A_2 = \frac{18}{\phi} + \frac{3}{4}(d_1 + 2d_2 + 4d_3 - 11d_4),$$

$$A_3 = \frac{20}{\phi} - \frac{1}{4}(9d_1 - 6d_2 + 5d_4),$$

where $d_1 = 1/[\phi(1-\phi\psi')]$, $d_2 = (1+\phi^2\psi'')/[\phi(1-\phi\psi')^2]$, $d_3 = (2-\phi^3\psi''')/[\phi(1-\phi\psi')^2]$, and $d_4 = (1+\phi^2\psi'')^2/[\phi(1-\phi\psi')^3]$, with $\psi = \psi(\phi) = \Gamma'(\phi)/\Gamma(\phi)$, $\Gamma'(\phi) = \mathrm{d}\Gamma(\phi)/\mathrm{d}\phi$, $\psi' = \mathrm{d}\psi/\mathrm{d}\phi$, $\psi'' = \mathrm{d}^2\psi/\mathrm{d}\phi^2$, and $\psi''' = \mathrm{d}^3\psi/\mathrm{d}\phi^3$, and $\Gamma(\cdot)$ represents the gamma function.

We now present results from a simulation study with 10,000 replications. The null hypothesis is $\mathcal{H}_0 : \beta = \phi = 1$. Note that $\mathcal{H}_0$ means that the data come from an exponential distribution with unity mean. Figure 2(a) presents a plot of size distortions for the 10% nominal level of the gradient test and its Bartlett-corrected version for different sample sizes. It can be noticed that the gradient test is oversized and its corrected version is clearly less size distorted.

Finally, we set $n = 10$ and plot the first-order and the second-order approximations for the distribution of the gradient statistic. Visual inspection of

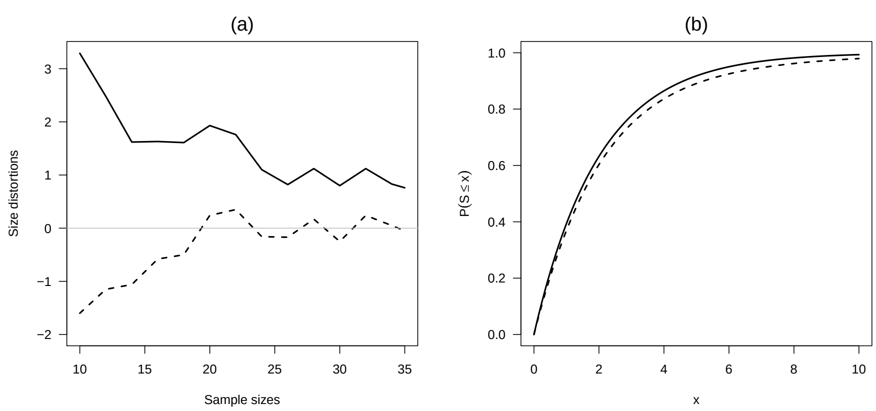

FIG 2. *(a) Size distortion of the gradient test (solid) and the Bartlett-corrected gradient test (dashes), (b) first-order approximation (solid) and expansion to order n^{-1} (dashes) of the null cumulative distribution function of the gradient statistic; gamma distribution.*

Figure 2(b) reveals that the first-order χ_2^2 approximation can be inaccurate in small samples.

5. Discussion

Lemonte and Ferrari [19] 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 one 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 [10, 21] 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. It can be used to investigate how closely the asymptotic distribution approximates the (unknown) true distribution of the gradient statistic.

Cordeiro and Ferrari [6] 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 [16] 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 [30, 14, 26, 27, 1, 15], and [22].

From the general expansion derived in this paper and using results in [6], we also obtained a Bartlett-type correction factor for the gradient statistic. The advantage of the corrected statistic over its uncorrected counterpart it that it is better approximated by the chi-square distribution. Unfortunately, the corrected gradient statistic is more difficult to obtain. 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, v, and w range over 1 to p and the indices j', r', s', u', v', 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 $\lambda_{jr} = -\psi_{jr} = -\{D_j D_r \ell(\boldsymbol{\theta})\}_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}$, $\psi_{jrs} = \{D_j D_r D_s \ell(\boldsymbol{\theta})\}_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}$, $\psi_{jrsu} = \{D_j D_r D_s D_u \ell(\boldsymbol{\theta})\}_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}$, etc. The matrix $\boldsymbol{\Lambda} = ((\lambda_{jr}))$ is the observed information matrix evaluated at $\widehat{\boldsymbol{\theta}}$. The partition of $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$ induces the partitions

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

where $\boldsymbol{\Lambda}^{-1}$ is the inverse of $\boldsymbol{\Lambda}$. Let ${\boldsymbol{\Lambda}^{11}}^{-1} = ((\lambda_{1w'j'}))$, $\sigma^{jr} = \lambda^{jr} - \lambda^{jw'}\lambda_{1w'j'}\lambda^{j'r}$, $\tau^{jj'} = \lambda^{jw'}\lambda_{1w'j'}$, $\sigma^{(1)}_{suvw} = \sigma^{su}\sigma^{vw}[3]$, $\lambda^{(1)}_{j'r's'u'} = \lambda^{j'r'}\lambda^{s'u'}[3]$, and $\lambda^{(2)}_{j'r's'u'v'w'} = \lambda^{j'r'}\lambda^{s'u'}\lambda^{v'w'}[15]$, where $[\cdot]$ denotes a summation with the number in brackets indicating the number of terms obtained by permutation of indices. For instance, $\sigma^{su}\sigma^{vw}[3] = \sigma^{su}\sigma^{vw} + \sigma^{sv}\sigma^{uw} + \sigma^{sw}\sigma^{uv}$. Let $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_q)^\top = n^{1/2}(\boldsymbol{\theta}_1 - \widehat{\boldsymbol{\theta}}_1)$, $\Psi^{(1)}_{j'} = \psi_{jrs}\sigma^{rs}\tau^{jj'}/2$, $\Psi^{(3)}_{j'r's'} = \psi_{jrs}\tau^{jj'}\tau^{rr'}\tau^{ss'}/6$,

$$\Psi^{(4)}_{j'r's'u'} = \frac{1}{24}\{\psi_{jrsu} + \sigma^{vw}(2\psi_{jrs}\psi_{uvw} + 3\psi_{jrv}\psi_{suw})\}\tau^{jj'}\tau^{rr'}\tau^{ss'}\tau^{uu'}.$$

Lemma 1. *An asymptotic expansion under the null hypothesis for the gradient statistic (1) is*

$$S = \boldsymbol{\epsilon}^\top {\boldsymbol{\Lambda}^{11}}^{-1}\boldsymbol{\epsilon} - \frac{3}{\sqrt{n}}\Psi^{(3)}_{j'r's'}\epsilon_{j'}\epsilon_{r'}\epsilon_{s'} - \frac{4}{n}\left(\Psi^{(4)}_{j'r's'u'} - \Psi^{(3)}_{j'r's'}\Psi^{(1)}_{u'}\right)\epsilon_{j'}\epsilon_{r'}\epsilon_{s'}\epsilon_{u'} + o_p(n^{-1}). \tag{9}$$

Proof. Using a procedure analogous to that of [5], the result holds. ☐

Let $\pi = \pi(\boldsymbol{\theta})$ be a prior density for $\boldsymbol{\theta}$, $\pi_j = D_j\pi(\boldsymbol{\theta})$, $\pi_{jr} = D_jD_r\pi(\boldsymbol{\theta})$, $\widehat{\pi} = \pi(\widehat{\boldsymbol{\theta}})$, $\widehat{\pi}_j = \pi_j(\widehat{\boldsymbol{\theta}})$, $\widehat{\pi}_{jr} = \pi_{jr}(\widehat{\boldsymbol{\theta}})$,

$$\Psi^{(2)}_{j'r'} = \left\{\frac{\widehat{\pi}_{jr}}{2\widehat{\pi}} + \frac{1}{4}\psi_{jrsu}\sigma^{su} + \frac{1}{24}\big(2\psi_{jrs}\psi_{uvw} + 3\psi_{jsu}\psi_{rvw}\big)\sigma^{(1)}_{suvw}\right\}\tau^{jj'}\tau^{rr'},$$

$$\Gamma^{(1)}_{j'} = \Psi^{(1)}_{j'} + \frac{\widehat{\pi}_j}{\widehat{\pi}}\tau^{jj'}, \qquad \Gamma^{(2)}_{j'r'} = \Psi^{(2)}_{j'r'} + \frac{1}{2\widehat{\pi}}\big(\psi_{jrs}\widehat{\pi}_u + \psi_{jsu}\widehat{\pi}_r\big)\sigma^{su}\tau^{jj'}\tau^{rr'},$$

$$\Gamma^{(4)}_{j'r's'u'} = \Psi^{(4)}_{j'r's'u'} + \frac{\widehat{\pi}_u}{6\widehat{\pi}}\psi_{jrs}\tau^{jj'}\tau^{rr'}\tau^{ss'}\tau^{uu'}.$$

From [10], [4] derive an expansion up to order n^{-1} for the marginal posterior density of $\boldsymbol{\epsilon}$, which takes the form

$$\begin{aligned}\pi_{post}(\boldsymbol{\epsilon}) = \phi_q(\boldsymbol{\epsilon};\boldsymbol{\Lambda}^{11})\Bigg[1 &+ \frac{1}{\sqrt{n}}\big(\Gamma^{(1)}_{j'}\epsilon_{j'} + \Gamma^{(3)}_{j'r's'}\epsilon_{j'}\epsilon_{r'}\epsilon_{s'}\big)\\ &+ \frac{1}{n}\Big\{\Gamma^{(2)}_{j'r'}\big(\epsilon_{j'}\epsilon_{r'} - \lambda^{j'r'}\big) + \Gamma^{(4)}_{j'r's'u'}\big(\epsilon_{j'}\epsilon_{r'}\epsilon_{s'}\epsilon_{u'} - \lambda^{(1)}_{j'r's'u'}\big)\\ &+ \frac{1}{2}\Psi^{(3)}_{j'r's'}\Psi^{(3)}_{u'v'w'}\big(\epsilon_{j'}\epsilon_{r'}\epsilon_{s'}\epsilon_{u'}\epsilon_{v'}\epsilon_{w'} - \lambda^{(2)}_{j'r's'u'v'w'}\big)\Big\}\Bigg] + o(n^{-1}),\end{aligned} \tag{10}$$

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

We now follow the Bayesian route described in [21]; see Appendix 2.

Step 1. The approximate posterior characteristic function of S is

$$M_\pi(t) = E_\pi\{\exp(\xi S)\} = \int \exp(\xi S)\pi_{post}(\boldsymbol{\epsilon})\mathrm{d}\boldsymbol{\epsilon},$$

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

$$\begin{aligned}\exp(\xi S)\pi_{post}(\boldsymbol{\epsilon}) = (1-2\xi)^{-q/2}\phi_q\left(\boldsymbol{\epsilon};\frac{\boldsymbol{\Lambda}^{11}}{1-2\xi}\right)\times&\\ \Bigg[1 + \frac{1}{\sqrt{n}}\Big\{(1-3\xi)\Psi^{(3)}_{j'r's'}\epsilon_{j'}\epsilon_{r'}\epsilon_{s'} + \Gamma^{(1)}_{j'}\epsilon'_j\Big\}&\\ + \frac{1}{n}\Bigg[\frac{1}{2}\Psi^{(3)}_{j'r's'}\Psi^{(3)}_{u'v'w'}\Big\{\frac{1}{9}(1-3\xi)^2\epsilon_{j'}\epsilon_{r'}\epsilon_{s'}\epsilon_{u'}\epsilon_{v'}\epsilon_{w'} - \lambda^{(2)}_{j'r's'u'v'w'}\Big\}&\\ - \Big[\xi\big\{4\Psi^{(4)}_{j'r's'u'} + \Psi^{(3)}_{j'r's'}\big(3\Gamma^{(1)}_{u'} - 4\Psi^{(1)}_{u'}\big)\big\} - \Gamma^{(4)}_{j'r's'u'}\Big]\epsilon_{j'}\epsilon_{r'}\epsilon_{s'}\epsilon_{u'}&\\ + \Gamma^{(2)}_{j'r'}\big(\epsilon_{j'}\epsilon_{r'} - \lambda^{j'r'}\big) - \Gamma^{(4)}_{j'r's'u'}\lambda^{(1)}_{j'r's'u'}\Bigg]\Bigg] + o_p(n^{-1}).&\end{aligned}$$

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

$$M_\pi(t) = (1-2\xi)^{-q/2}\left\{1+\frac{1}{n}\sum_{i=0}^{3} H_i(1-2\xi)^{-i}\right\} + o_p(n^{-1}), \qquad (11)$$

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

$$H_1 = \frac{9}{8}\Psi^{(3)}_{j'r's'}\Psi^{(3)}_{u'v'w'}\lambda^{(2)}_{j'r's'u'v'w'} + \Gamma^{(2)}_{j'r'}\lambda^{j'r'} + \lambda^{(1)}_{j'r's'u'}\left\{2\big(\Psi^{(4)}_{j'r's'u'} - \Psi^{(3)}_{j'r's'}\Psi^{(1)}_{u'}\big) + \frac{3}{2}\Psi^{(3)}_{j'r's'}\Gamma^{(1)}_{u'}\right\},$$

$$H_2 = -\frac{3}{4}\Psi^{(3)}_{j'r's'}\Psi^{(3)}_{u'v'w'}\lambda^{(2)}_{j'r's'u'v'w'} + \lambda^{(1)}_{j'r's'u'}\left\{\Gamma^{(4)}_{j'r's'u'} - 2\big(\Psi^{(4)}_{j'r's'u'} - \Psi^{(3)}_{j'r's'}\Psi^{(1)}_{u'}\big) - \frac{3}{2}\Psi^{(3)}_{j'r's'}\Gamma^{(1)}_{u'}\right\},$$

$$H_3 = \frac{1}{8}\Psi^{(3)}_{j'r's'}\Psi^{(3)}_{u'v'w'}\lambda^{(2)}_{j'r's'u'v'w'}.$$

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

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

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

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

where $\bar{J}_0 = -(\bar{J}_1+\bar{J}_2+\bar{J}_3)$,

$$\begin{aligned}\bar{J}_1 = {}& \frac{1}{32}\kappa_{jrs}\kappa_{uvw}\big(9m^{jr}m^{su}m^{vw} + 6m^{ju}m^{rv}m^{sw}\big) + \frac{1}{4}\kappa_{jrsu}m^{jr}m^{su} \\ & + \frac{1}{4}\kappa_{jrv}\kappa_{suw}a^{vw}\big(m^{jr}m^{su}[3]\big) + \frac{1}{8}\kappa_{jrs}\kappa_{uvw}a^{vw}(m^{jr}m^{su}[3]) \\ & + \frac{3}{4}\kappa_{jrs}m^{jr}m^{su}\frac{\bar{\pi}_u}{\bar{\pi}} + \frac{1}{2}m^{jr}\left\{\frac{\bar{\pi}_{jr}}{\bar{\pi}} + \frac{1}{2}\kappa_{jrsu}a^{su}\right. \\ & \left. + \frac{1}{12}(2\kappa_{jrs}\kappa_{uvw} + 3\kappa_{jsu}\kappa_{rvw})(a^{su}a^{vw}[3]) + \frac{\bar{\pi}_u}{\bar{\pi}}\kappa_{jrs}a^{su} + \frac{\bar{\pi}_r}{\bar{\pi}}\kappa_{jsu}a^{su}\right\},\end{aligned}$$

$$\bar{J}_2 = -\frac{1}{48}\kappa_{jrs}\kappa_{uvw}\big(9m^{jr}m^{su}m^{vw} + 6m^{ju}m^{rv}m^{sw}\big) - \frac{1}{4}\kappa_{jrsu}m^{jr}m^{su}$$
$$- \frac{1}{4}\kappa_{jrv}\kappa_{suw}a^{vw}\big(m^{jr}m^{su}[3]\big) - \frac{1}{8}\kappa_{jrs}\kappa_{uvw}a^{vw}(m^{jr}m^{su}[3])$$
$$- \frac{3}{4}\kappa_{jrs}m^{jr}m^{su}\frac{\bar{\pi}_u}{\bar{\pi}} + \frac{1}{8}\left(\kappa_{jrsu} + \frac{1}{6}\kappa_{jrs}\frac{\bar{\pi}_u}{\bar{\pi}}\right)m^{jr}m^{su}$$
$$+ \frac{1}{24}\big(2\kappa_{jrs}\kappa_{uvw} + 3\kappa_{jrv}\kappa_{suw}\big)a^{vw}(m^{jr}m^{su}[3]),$$

$$\bar{J}_3 = \frac{1}{288}\kappa_{jrs}\kappa_{uvw}\big(9m^{jr}m^{su}m^{vw} + 6m^{ju}m^{rv}m^{sw}\big).$$

Step 3. We now compute

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

by integrating the $\bar{J}$'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 $\boldsymbol{\theta}$, we arrive at

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

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

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

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

$$A_1 = 12D_jD_rm^{jr} - 6D_u(\kappa_{jrs}m^{jr}m^{su}) - 12D_u\big(\kappa_{jrs}m^{jr}a^{su}\big)$$
$$- 12D_r\big(\kappa_{jsu}m^{jr}a^{su}\big) + 6\kappa_{jrsu}m^{jr}a^{su} + \kappa_{jrs}\kappa_{uvw}a^{vw}\big(m^{jr}m^{su}[3]\big)$$
$$+ (2\kappa_{jrs}\kappa_{uvw} + 3\kappa_{jsu}\kappa_{rvw})m^{jr}\big(a^{su}a^{vw}[3]\big),$$

$$A_2 = 6D_u\big(\kappa_{jrs}m^{jr}m^{su}\big) - \big(\kappa_{jrs}\kappa_{uvw} + 3\kappa_{jrv}\kappa_{suw}\big)a^{vw}\big(m^{jr}m^{su}[3]\big)$$
$$- 3\kappa_{jrsu}m^{jr}m^{su} - \frac{3}{4}\kappa_{jrs}\kappa_{uvw}\big(3m^{jr}m^{su}m^{vw} + 2m^{ju}m^{rv}m^{sw}\big),$$

$$A_3 = \frac{1}{4}\kappa_{jrs}\kappa_{uvw}\big(3m^{jr}m^{su}m^{vw} + 2m^{ju}m^{rv}m^{sw}\big).$$

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 1.

Appendix 2

The Shrinkage Argument

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

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

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

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

A detailed justification can be found in [21].

References

[1] Bai, P. (2009). Sphericity test in a GMANOVA-MANOVA model with normal error. *Journal of Multivariate Analysis* **100**, 2305–2312. MR2560371

[2] 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. MR1062699

[3] Birnbaum, Z.W., Saunders, S.C. (1969). A new family of life distributions. *Journal of Applied Probability* **6**, 319–327. MR0253493

[4] 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. MR2734243

[5] 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.

[6] Cordeiro, G.M., Ferrari, S.L.P. (1991). A modified score test statistic having chi-squared distribuition to order n^{-1}. *Biometrika* **78**, 573–582. MR1130925

[7] Cordeiro, G.M., Cribari-Neto, F. (1996). On Bartlett and Bartlett-type corrections. *Econometric Reviews* **15**, 339–367. MR1423902

[8] Cox, D.R., Reid, N. (1987). Parameter orthogonality and approximate conditional inference (with discussion). *Journal of the Royal Statistical Society B* **40**, 1–39. MR0893334

[9] Datta, G.S., Mukerjee, R. (2003). *Probability Matching Priors: Higher Order Asymptoptics*. Springer-Verlag: New York. MR2053794

[10] Ghosh, J.K., Mukerjee, R. (1991). Characterization of priors under wich Bayesian and frequentist Bartlett corrections are equivalent in

the multiparameter case. *Journal of Multivariate Analysis* **38**, 385–393. MR1131727

[11] Harris, P. (1985). An asymptotic expansion for the null distribution of the efficient score statistic. *Biometrika* **72**, 653–659. MR0817580

[12] Hayakawa, T. (1977). The likelihood ratio criterion and the asymptotic expansion of its distribution. *Annals of the Institute of Statistical Mathematics* **29**, 359–378. MR0474588

[13] Hill, G.W., Davis, A.W. (1968). Generalized asymptotic expansions of Cornish–Fisher type. *The Annals of Mathematical Statistics* **39**, 1264-73. MR0226762

[14] 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. MR2042112

[15] 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. MR2719695

[16] Lawley, D. (1956). A general method for approximating to the distribution of likelihood ratio criteria. *Biometrika* **43**, 295–303. MR0082237

[17] Lemonte, A.J. (2011). Local power of some tests in exponential family nonlinear models. *Journal of Statistical Planning and Inference* **141**, 1981–1989. MR2763226

[18] 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. MR2879762

[19] 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. MR2878911

[20] 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.

[21] Mukerjee, R., Reid, N. (2000). On the Bayesian approach for frequentist computations. *Brazilian Journal of Probability and Statistics* **14**, 159–166. MR1860054

[22] Noma, H. (2011). Confidence intervals for a random-effects meta-analysis based on Bartlett-type corrections. *Statistics in Medicine* **30**, 3304–3312. MR2861615

[23] Rao, C.R. (1948). Large sample tests of statistical hypotheses concerning several parameters with applications to problens of estimation. *Proceedings of the Cambridge Philosophical Society* **44**, 50–57. MR0024111

[24] 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. MR2111498

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

[26] 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. MR2283012

[27] van Giersbergen, N.P.A. (2009). Bartlett correction in the stable AR(1) model with intercept and trend. *Econometric Theory* **25**, 857–872. MR2507537

[28] 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. MR0012401

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

[30] 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. MR1796295