

# The local power of the gradient test

Artur J. Lemonte · Silvia L. P. Ferrari

Received: 2 February 2010 / Revised: 28 June 2010 / Published online: 6 October 2010  
© The Institute of Statistical Mathematics, Tokyo 2010

**Abstract** The asymptotic expansion of the distribution of the gradient test statistic is derived for a composite hypothesis under a sequence of Pitman alternative hypotheses converging to the null hypothesis at rate  $n^{-1/2}$ ,  $n$  being the sample size. Comparisons of the local powers of the gradient, likelihood ratio, Wald and score tests reveal no uniform superiority property. The power performance of all four criteria in one-parameter exponential family is examined.

**Keywords** Asymptotic expansions · Chi-square distribution · Gradient test · Likelihood ratio test · Pitman alternative · Power function · Score test · Wald test

## 1 Introduction

The most commonly used large sample tests are the likelihood ratio (Wilks 1938), Wald (1943) and Rao score (1948) tests. Recently, Terrell (2002) proposed a new test statistic that shares the same first order asymptotic properties with the likelihood ratio (LR), Wald ( $W$ ) and Rao score ( $S_R$ ) statistics. The new statistic, referred to as the gradient statistic ( $S_T$ ), is markedly simple. In fact, Rao (2005) 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.” The present paper goes in this direction.

Let  $\mathbf{x} = (x_1, \dots, x_n)^\top$  be a random vector of  $n$  independent observations with probability density function  $\pi(\mathbf{x} | \boldsymbol{\theta})$  that depends on a  $p$ -dimensional vector of unknown parameters  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)^\top$ . Consider the problem of testing the composite null hypothesis  $\mathcal{H}_0 : \boldsymbol{\theta}_2 = \boldsymbol{\theta}_{20}$  against  $\mathcal{H}_1 : \boldsymbol{\theta}_2 \neq \boldsymbol{\theta}_{20}$ , where  $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$ ,

---

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

---

A. J. Lemonte · S. L. P. Ferrari (✉)  
Departamento de Estatística, Universidade de São Paulo, São Paulo, SP 05508-090, Brazil  
e-mail: silviaferrari.usp@gmail.com

$\theta_1 = (\theta_1, \dots, \theta_q)^\top$  and  $\theta_2 = (\theta_{q+1}, \dots, \theta_p)^\top$ ,  $\theta_{20}$  representing a  $(p - q)$ -dimensional fixed vector. Let  $\ell$  be the total log-likelihood function, i.e.  $\ell = \ell(\theta) = \sum_{l=1}^n \log \pi(x_l | \theta)$ . Let  $U(\theta) = \partial \ell / \partial \theta = (U_1(\theta)^\top, U_2(\theta)^\top)^\top$  be the corresponding total score function partitioned following the partition of  $\theta$ . The restricted and unrestricted maximum likelihood estimators of  $\theta$  are  $\hat{\theta} = (\hat{\theta}_1^\top, \hat{\theta}_2^\top)^\top$  and  $\tilde{\theta} = (\tilde{\theta}_1^\top, \tilde{\theta}_{20}^\top)^\top$ , respectively.

The gradient statistic for testing  $\mathcal{H}_0$  is

$$S_T = U(\tilde{\theta})^\top (\hat{\theta} - \tilde{\theta}). \tag{1}$$

Since  $U_1(\tilde{\theta}) = \mathbf{0}$ , the gradient statistic in (1) can be written as  $S_T = U_2(\tilde{\theta})^\top (\hat{\theta}_2 - \theta_{20})$ . Clearly,  $S_T$  has a very simple form and does not involve knowledge of the information matrix, neither expected nor observed, and no matrices, unlike  $W$  and  $S_R$ . Asymptotically,  $S_T$  has a central chi-square distribution with  $p - q$  degrees of freedom under  $\mathcal{H}_0$ . Terrell (2002) points out that the gradient statistic “is not transparently non-negative, even though it must be so asymptotically.” His Theorem 2 implies that if the log-likelihood function is concave and is differentiable at  $\tilde{\theta}$ , then  $S_T \geq 0$ .

In this paper, we derive the asymptotic distribution of the gradient statistic for a composite null hypothesis under a sequence of Pitman alternatives converging to the null hypothesis at a convergence rate  $n^{-1/2}$ . In other words, the sequence of alternative hypotheses is  $\mathcal{H}_{1n} : \theta_2 = \theta_{20} + n^{-1/2}\epsilon$ , where  $\epsilon = (\epsilon_{q+1}, \dots, \epsilon_p)^\top$ . Similar results for the likelihood ratio and Wald tests were obtained by Hayakawa (1975) and for the score test, by Harris and Peers (1980). Comparison of local power properties of the competing tests will be performed. Our results will be specialized to the case of the one-parameter exponential family. A brief discussion closes the paper.

### 2 Notation and preliminaries

Our notation follows that of Hayakawa (1975, 1977). We introduce the following log-likelihood derivatives

$$y_r = n^{-1/2} \frac{\partial \ell}{\partial \theta_r}, \quad y_{rs} = n^{-1} \frac{\partial^2 \ell}{\partial \theta_r \partial \theta_s}, \quad y_{rst} = n^{-3/2} \frac{\partial^3 \ell}{\partial \theta_r \partial \theta_s \partial \theta_t},$$

their arrays  $\mathbf{y} = (y_1, \dots, y_p)^\top$ ,  $\mathbf{Y} = ((y_{rs}))$ ,  $\mathbf{Y}_{\dots} = ((y_{rst}))$ , the corresponding cumulants  $\kappa_{rs} = E(y_{rs})$ ,  $\kappa_{r,s} = E(y_r y_s)$ ,  $\kappa_{rst} = n^{1/2} E(y_{rst})$ ,  $\kappa_{r,st} = n^{1/2} E(y_r y_{st})$ ,  $\kappa_{r,s,t} = n^{1/2} E(y_r y_s y_t)$  and their arrays  $\mathbf{K} = ((\kappa_{r,s}))$ ,  $\mathbf{K}_{\dots} = ((\kappa_{rst}))$ ,  $\mathbf{K}_{\dots} = ((\kappa_{r,st}))$  and  $\mathbf{K}_{\dots} = ((\kappa_{r,s,t}))$ .

We make the same assumptions as in Hayakawa (1975). In particular, it is assumed that the  $\kappa$ 's are all  $O(1)$  and they are not functionally independent; for instance,  $\kappa_{r,s} = -\kappa_{rs}$ . Relations among them were first obtained by Bartlett (1953a,b). Also, it is assumed that  $\mathbf{Y}$  is non-singular and that  $\mathbf{K}$  is positive definite with inverse  $\mathbf{K}^{-1} = ((\kappa^{r,s}))$  say. For triple-suffix quantities we use the following summation notation

$$\mathbf{K}_{\dots} \circ \mathbf{a} \circ \mathbf{b} \circ \mathbf{c} = \sum_{r,s,t=1}^p \kappa_{rst} a_r b_s c_t, \quad \mathbf{K}_{\dots} \circ \mathbf{M} \circ \mathbf{b} = \sum_{r,s,t=1}^p \kappa_{r,st} m_{rs} b_t,$$

where  $\mathbf{M}$  is a  $p \times p$  matrix and  $\mathbf{a}$ ,  $\mathbf{b}$  and  $\mathbf{c}$  are  $p \times 1$  column vectors.

The partition  $\theta = (\theta_1^\top, \theta_2^\top)^\top$  induces the corresponding partitions:

$$Y = \begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix}, \quad K = \begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{bmatrix}, \quad K^{-1} = \begin{bmatrix} K^{11} & K^{12} \\ K^{21} & K^{22} \end{bmatrix},$$

$a = (a_1^\top, a_2^\top)^\top$ , etc. Also,

$$K_{2..} \circ a_2 \circ b \circ c = \sum_{r=q+1}^p \sum_{s,t=1}^p \kappa_{rst} a_r b_s c_t.$$

Using a procedure analogous to that of Hayakawa (1975), we can write the asymptotic expansion of  $S_T$  for the composite hypothesis up to order  $n^{-1/2}$  as

$$S_T = -(Zy + \xi)^\top Y(Zy + \xi) - \frac{1}{2\sqrt{n}} K_{...} \circ (Zy + \xi) \circ Y^{-1}y \circ Y^{-1}y \\ - \frac{1}{2\sqrt{n}} K_{...} \circ (Zy + \xi) \circ (Z_0y - \xi) \circ (Z_0y - \xi) + O_p(n^{-1}),$$

where  $Z = Y^{-1} - Z_0$ ,

$$Z_0 = \begin{bmatrix} Y_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad \xi = \begin{bmatrix} Y_{11}^{-1} & Y_{12} \\ -I_{p-q} & \end{bmatrix} \epsilon,$$

$I_{p-q}$  being the identity matrix of order  $p - q$ .

We can now use a multivariate Edgeworth Type A series expansion of the joint density function of  $y$  and  $Y$  up to order  $n^{-1/2}$  (Peers 1971), which has the form

$$f_1 = f_0 \left[ 1 + \frac{1}{6\sqrt{n}} (K_{....} \circ K^{-1}y \circ K^{-1}y \circ K^{-1}y - 3K_{....} \circ K^{-1} \circ K^{-1}y) \right. \\ \left. - \frac{1}{\sqrt{n}} K_{....} \circ K^{-1}y \circ D \right] + O(n^{-1}),$$

where

$$f_0 = (2\pi)^{-p/2} |K|^{-1/2} \exp \left\{ -\frac{1}{2} y^\top K^{-1} y \right\} \prod_{r,s=1}^p \delta(y_{rs} - \kappa_{rs}),$$

$D = ((d_{bc}))$ ,  $d_{bc} = \delta'(y_{bc} - \kappa_{bc}) / \delta(y_{bc} - \kappa_{bc})$ , with  $\delta(\cdot)$  being the Dirac delta function (Bracewell 1999), to obtain the moment generating function of  $S_T$ ,  $M(t)$  say.

From  $f_1$  and the asymptotic expansion of  $S_T$  up to order  $n^{-1/2}$ , we arrive, after long algebra, at

$$M(t) = (1 - 2t)^{-\frac{1}{2}(p-q)} \exp \left( \frac{t}{1 - 2t} \epsilon^\top K_{22.1} \epsilon \right) \\ \times \left[ 1 + \frac{1}{\sqrt{n}} (A_1 d + A_2 d^2 + A_3 d^3) \right] + O(n^{-1}),$$

where  $d = 2t/(1 - 2t)$ ,  $\mathbf{K}_{22.1} = \mathbf{K}_{22} - \mathbf{K}_{21}\mathbf{K}_{11}^{-1}\mathbf{K}_{12}$ ,  $A_1 = -(\mathbf{K}_{...} \circ \mathbf{K}^{-1} \circ \boldsymbol{\epsilon}^* + 4\mathbf{K}_{...} \circ \mathbf{A} \circ \boldsymbol{\epsilon}^* + \mathbf{K}_{...} \circ \mathbf{A} \circ \boldsymbol{\epsilon}^* + \mathbf{K}_{...} \circ \boldsymbol{\epsilon}^* \circ \boldsymbol{\epsilon}^* \circ \boldsymbol{\epsilon}^*)/4$ ,  $A_2 = -(\mathbf{K}_{...} \circ \mathbf{K}^{-1} \circ \boldsymbol{\epsilon}^* - \mathbf{K}_{...} \circ \mathbf{A} \circ \boldsymbol{\epsilon}^* - 2\mathbf{K}_{...} \circ \boldsymbol{\epsilon}^* \circ \boldsymbol{\epsilon}^* \circ \boldsymbol{\epsilon}^*)/4$ ,  $A_3 = -\mathbf{K}_{...} \circ \boldsymbol{\epsilon}^* \circ \boldsymbol{\epsilon}^* \circ \boldsymbol{\epsilon}^*/12$ ,

$$\boldsymbol{\epsilon}^* = \begin{bmatrix} \mathbf{K}_{11}^{-1} & \mathbf{K}_{12} \\ -\mathbf{I}_{p-q} & \mathbf{0} \end{bmatrix} \boldsymbol{\epsilon}, \quad \mathbf{A} = \begin{bmatrix} \mathbf{K}_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

When  $n \rightarrow \infty$ ,  $M(t) \rightarrow (1 - 2t)^{-(p-q)/2} \exp\{2t\lambda/(1 - 2t)\}$ , where  $\lambda = \boldsymbol{\epsilon}^\top \mathbf{K}_{22.1} \boldsymbol{\epsilon}/2$ , and hence the limiting distribution of  $S_T$  is a non-central chi-square distribution with  $p - q$  degrees of freedom and non-centrality parameter  $\lambda$ . Under  $\mathcal{H}_0$ , i.e. when  $\boldsymbol{\epsilon} = \mathbf{0}$ ,  $M(t) = (1 - 2t)^{-(p-q)/2} + O(n^{-1})$  and, as expected,  $S_T$  has a central chi-square distribution with  $p - q$  degrees of freedom up to an error of order  $n^{-1}$ . Also, from  $M(t)$  we may obtain the first three moments of  $S_T$  up to order  $n^{-1/2}$  as  $\mu'_1(S_T) = p - q + \lambda + 2A_1/\sqrt{n}$ ,  $\mu_2(S_T) = 2(p - q + 2\lambda) + 8(A_1 + A_2)/\sqrt{n}$  and  $\mu_3(S_T) = 8(p - q + 3\lambda) + 6(A_1 + 2A_2 + A_3)/\sqrt{n}$ .

### 3 Main result

The moment generating function of  $S_T$  in a neighborhood of  $\boldsymbol{\theta}_2 = \boldsymbol{\theta}_{20}$  can be written, after some algebra, as

$$M(t) = (1 - 2t)^{-\frac{1}{2}(p-q)} \exp\left(\frac{t}{1 - 2t} \boldsymbol{\epsilon}^\top \mathbf{K}_{22.1}^\dagger \boldsymbol{\epsilon}\right) \times \left[1 + \frac{1}{\sqrt{n}} \sum_{k=0}^3 a_k (1 - 2t)^{-k}\right] + O(n^{-1}),$$

where

$$\begin{aligned} a_1 &= \frac{1}{4} \left\{ \mathbf{K}_{...}^\dagger \circ (\mathbf{K}^{-1})^\dagger \circ (\boldsymbol{\epsilon}^*)^\dagger - (4\mathbf{K}_{...} + 3\mathbf{K}_{...})^\dagger \circ \mathbf{A}^\dagger \circ (\boldsymbol{\epsilon}^*)^\dagger \right. \\ &\quad - 2(\mathbf{K}_{...} + 2\mathbf{K}_{...})^\dagger \circ (\boldsymbol{\epsilon}^*)^\dagger \circ (\boldsymbol{\epsilon}^*)^\dagger \circ (\boldsymbol{\epsilon}^*)^\dagger \\ &\quad \left. - 2(\mathbf{K}_{2..} + \mathbf{K}_{2,..})^\dagger \circ \boldsymbol{\epsilon} \circ (\boldsymbol{\epsilon}^*)^\dagger \circ (\boldsymbol{\epsilon}^*)^\dagger \right\}, \\ a_2 &= -\frac{1}{4} \left\{ \mathbf{K}_{...}^\dagger \circ (\mathbf{K}^{-1} - \mathbf{A})^\dagger \circ (\boldsymbol{\epsilon}^*)^\dagger \right. \\ &\quad \left. - (\mathbf{K}_{...} + 2\mathbf{K}_{...})^\dagger \circ (\boldsymbol{\epsilon}^*)^\dagger \circ (\boldsymbol{\epsilon}^*)^\dagger \circ (\boldsymbol{\epsilon}^*)^\dagger \right\}, \\ a_3 &= -\frac{1}{12} \mathbf{K}_{...}^\dagger \circ (\boldsymbol{\epsilon}^*)^\dagger \circ (\boldsymbol{\epsilon}^*)^\dagger \circ (\boldsymbol{\epsilon}^*)^\dagger, \end{aligned} \tag{2}$$

and  $a_0 = -(a_1 + a_2 + a_3)$ . The symbol “ $\dagger$ ” denotes evaluation at  $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_{20}^\top)^\top$ . Inverting  $M(t)$ , we arrive at the following theorem, our main result.

**Theorem 1** *The asymptotic expansion of the distribution of the gradient statistic for testing a composite hypothesis under a sequence of local alternatives converging to the null hypothesis at rate  $n^{-1/2}$  is*

$$\Pr(S_T \leq x) = G_{f,\lambda}(x) + \frac{1}{\sqrt{n}} \sum_{k=0}^3 a_k G_{f+2k,\lambda}(x) + O(n^{-1}), \tag{3}$$

where  $G_{m,\lambda}(x)$  is the cumulative distribution function of a non-central chi-square variate with  $m$  degrees of freedom and non-centrality parameter  $\lambda$ . Here,  $f = p - q$ ,  $\lambda = \epsilon^\top \mathbf{K}_{22.1}^\dagger \epsilon / 2$  and the  $a_k$ 's are given in (2).

If  $q = 0$ , the null hypothesis is simple,  $\epsilon^* = -\epsilon$  and  $\mathbf{A} = \mathbf{0}$ . Therefore, an immediate consequence of Theorem 1 is the following corollary.

**Corollary 1** *The asymptotic expansion of the distribution of the gradient statistic for testing a simple hypothesis under a sequence of local alternatives converging to the null hypothesis at rate  $n^{-1/2}$  is given by (3) with  $f = p$ ,  $\lambda = \epsilon^\top \mathbf{K}^\dagger \epsilon / 2$ ,  $a_0 = \mathbf{K}_{\dots}^\dagger \circ \epsilon \circ \epsilon \circ \epsilon \circ \epsilon / 6$ ,  $a_1 = -\{\mathbf{K}_{\dots}^\dagger \circ (\mathbf{K}^{-1})^\dagger \circ \epsilon - 2\mathbf{K}_{\dots}^\dagger \circ \epsilon \circ \epsilon \circ \epsilon\} / 4$ ,  $a_2 = \{\mathbf{K}_{\dots}^\dagger \circ (\mathbf{K}^{-1})^\dagger \circ \epsilon - (\mathbf{K}_{\dots} + 2\mathbf{K}_{\dots})^\dagger \circ \epsilon \circ \epsilon \circ \epsilon\} / 4$  and  $a_3 = \mathbf{K}_{\dots}^\dagger \circ \epsilon \circ \epsilon \circ \epsilon \circ \epsilon / 12$ .*

### 4 Power comparisons between the rival tests

To first order  $S_T$ , LR,  $W$  and  $S_R$  have the same asymptotic distributional properties under either the null or local alternative hypotheses. Up to an error of order  $n^{-1}$  the corresponding criteria have the same size but their powers differ in the  $n^{-1/2}$  term. The power performance of the different tests may then be compared based on the expansions of their power functions ignoring terms of order less than  $n^{-1/2}$ . Harris and Peers (1980) presented a study of local power, up to order  $n^{-1/2}$ , for the likelihood ratio, Wald and score tests. They showed that none of the criteria is uniformly better than the others.

Let  $S_i$  ( $i = 1, 2, 3, 4$ ) be, respectively, the likelihood ratio, Wald, score and gradient statistics. We can write their local powers as  $\Pi_i = 1 - \Pr(S_i \leq x) = \Pr(S_i > x)$ , where

$$\Pr(S_i \leq x) = G_{p-q,\lambda}(x) + \frac{1}{\sqrt{n}} \sum_{k=0}^3 a_{ik} G_{p-q+2k,\lambda}(x) + O(n^{-1}).$$

The coefficients that define the local powers of the likelihood ratio and Wald tests are given in Hayakawa (1975), those corresponding to the score and gradient tests are given in Harris and Peers (1980) and in (2), respectively. All of them are complicated functions of joint cumulants of log-likelihood derivatives but we can draw the following general conclusions:

- all the four tests are locally biased;
- if  $\mathbf{K}_{\dots} = \mathbf{0}$ , the likelihood ratio, Wald and gradient tests have identical local powers;
- if  $\mathbf{K}_{\dots} = 2\mathbf{K}_{\dots}$ , the score and gradient tests have identical local powers.

Further classifications are possible for appropriate subspaces of the parameter space; see, for instance, [Harris and Peers \(1980\)](#) and [Hayakawa and Puri \(1985\)](#). Therefore, there is no uniform superiority of one test with respect to the others. Hence, the gradient test, which is very simple to compute as pointed out by [Rao \(2005\)](#), is an attractive alternative to the likelihood ratio, Wald and score tests.

### 5 One-parameter exponential family

Let  $\mathbf{x} = (x_1, \dots, x_n)^\top$  be a random sample of size  $n$ , with each  $x_l$  having probability density function  $\pi(x; \theta) = \exp\{t(x; \theta)\}$ , where  $\theta$  is a scalar parameter. To test  $\mathcal{H}_0 : \theta = \theta_0$ , where  $\theta_0$  is a fixed known constant, the likelihood ratio, Wald, score and gradient statistics are, respectively,

$$S_1 = 2 \sum_{l=1}^n \{t(x_l; \hat{\theta}) - t(x_l; \theta_0)\}, \quad S_2 = n(\hat{\theta} - \theta_0)^2 K(\hat{\theta}),$$

$$S_3 = \frac{(\sum_{l=1}^n t'(x_l; \theta_0))^2}{nK(\theta_0)}, \quad S_4 = (\hat{\theta} - \theta_0) \sum_{l=1}^n t'(x_l; \theta_0),$$

where  $\hat{\theta}$  is the maximum likelihood estimator of  $\theta$  and  $K = K(\theta)$  denotes the Fisher information for a single observation. Under  $\mathcal{H}_0$  all the four statistics have a central chi-square distribution with one degree of freedom asymptotically.

Now, let  $\kappa_{\theta\theta} = E\{t''(x; \theta)\}$ ,  $\kappa_{\theta\theta\theta} = E\{t'''(x; \theta)\}$ ,  $\kappa_{\theta\theta,\theta} = E\{t''(x; \theta)t'(x; \theta)\}$ ,  $\kappa^{\theta,\theta} = -\kappa_{\theta\theta}^{-1}$ , etc, where primes denote derivatives with respect to  $\theta$ ; for instance  $t''(x; \theta) = d^2 t(x; \theta)/d\theta^2$ . The asymptotic expansion of the distribution of the gradient statistic for the null hypothesis  $\mathcal{H}_0 : \theta = \theta_0$  under the sequence of local alternatives  $\mathcal{H}_{1n} : \theta = \theta_0 + n^{-1/2}\epsilon$  is given by (3) with  $f = 1$ ,  $\lambda = K^\dagger \epsilon^2/2$ ,

$$a_0 = \frac{\kappa_{\theta\theta\theta}^\dagger \epsilon^3}{6}, \quad a_1 = -\frac{\kappa_{\theta\theta\theta}^\dagger (\kappa^{\theta,\theta})^\dagger \epsilon - 2\kappa_{\theta,\theta\theta}^\dagger \epsilon^3}{4},$$

$$a_2 = \frac{\kappa_{\theta\theta\theta}^\dagger (\kappa^{\theta,\theta})^\dagger \epsilon - (\kappa_{\theta\theta\theta} + 2\kappa_{\theta,\theta\theta})^\dagger \epsilon^3}{4}, \quad a_3 = \frac{\kappa_{\theta\theta\theta}^\dagger \epsilon^3}{12}.$$

We now specialize to the case where  $\pi(x; \theta)$  belongs to the one-parameter exponential family. Let  $t(x; \theta) = -\log \zeta(\theta) - \alpha(\theta)d(x) + v(x)$ , where  $\alpha(\cdot)$ ,  $\zeta(\cdot)$ ,  $d(\cdot)$  and  $v(\cdot)$  are known functions. Also,  $\alpha(\cdot)$  and  $\zeta(\cdot)$  are assumed to have first three continuous derivatives, with  $\zeta(\cdot) > 0$ ,  $\alpha'(\theta)$  and  $\beta'(\theta)$  being different from zero for all  $\theta$  in the parameter space, where  $\beta(\theta) = \zeta'(\theta)/\{\zeta(\theta)\alpha'(\theta)\}$ . Since  $K = \alpha'(\theta)\beta'(\theta)$ ,  $\sum_{l=1}^n t(x_l; \theta) = -n\{\log \zeta(\theta) + \alpha(\theta)\bar{d} - \bar{v}\}$ ,  $\sum_{l=1}^n t'(x_l; \theta) = -n\alpha'(\theta)\{\beta(\theta) + \bar{d}\}$ , with  $\bar{d} = \sum_{l=1}^n d(x_l)/n$  and  $\bar{v} = \sum_{l=1}^n v(x_l)/n$ , we have

$$S_1 = 2n \left[ \log \left\{ \frac{\zeta(\theta_0)}{\zeta(\hat{\theta})} \right\} + \{\alpha(\theta_0) - \alpha(\hat{\theta})\}\bar{d} \right], \quad S_2 = n(\hat{\theta} - \theta_0)^2 \alpha'(\hat{\theta})\beta'(\hat{\theta}),$$

$$S_3 = \frac{n\alpha'(\theta_0)\{\beta(\theta_0) + \bar{d}\}^2}{\beta'(\theta_0)}, \quad S_4 = n(\theta_0 - \hat{\theta})\alpha'(\theta_0)\{\beta(\theta_0) + \bar{d}\}.$$

Let  $\alpha' = \alpha'(\theta)$ ,  $\alpha'' = \alpha''(\theta)$ ,  $\beta' = \beta'(\theta)$  and  $\beta'' = \beta''(\theta)$ . It can be shown that  $\kappa_{\theta\theta} = -\alpha'\beta'$ ,  $\kappa_{\theta\theta\theta} = -(2\alpha''\beta' + \alpha'\beta'')$ ,  $\kappa_{\theta,\theta\theta} = \alpha''\beta'$ ,  $\kappa_{\theta,\theta,\theta} = \alpha'\beta'' - \alpha''\beta'$ . The coefficients that define the local powers of the tests that use  $S_1, S_2, S_3$  and  $S_4$  are

$$\begin{aligned} a_{10} = a_{20} = a_{30} = -a_{23} = 2a_{43} &= -\frac{(2\alpha''\beta' + \alpha'\beta'')\epsilon^3}{6}, & a_{11} &= \frac{\alpha''\beta'\epsilon^3}{2}, \\ a_{12} = a_{33} = -a_{40} &= \frac{(\alpha'\beta'' - \alpha''\beta')\epsilon^3}{6}, & a_{31} &= \frac{\alpha''\beta'\epsilon^3}{2} - \frac{(\alpha'\beta'' - \alpha''\beta')\epsilon}{2\alpha'\beta'}, \\ a_{21} = -a_{22} &= \frac{\alpha''\beta'\epsilon^3}{2} - \frac{(2\alpha''\beta' + \alpha'\beta'')\epsilon}{2\alpha'\beta'}, & a_{32} &= \frac{(\alpha'\beta'' - \alpha''\beta')\epsilon}{2\alpha'\beta'}, & a_{13} &= 0, \\ a_{41} &= \frac{\alpha''\beta'\epsilon^3}{2} + \frac{(2\alpha''\beta' + \alpha'\beta'')\epsilon}{4\alpha'\beta'}, & a_{42} &= \frac{\alpha'\beta''\epsilon^3}{4} - \frac{(2\alpha''\beta' + \alpha'\beta'')\epsilon}{4\alpha'\beta'}. \end{aligned}$$

If  $\alpha(\theta) = \theta$ ,  $\pi(x; \theta)$  corresponds to a one-parameter natural exponential family. In this case,  $\alpha' = 1$ ,  $\alpha'' = 0$  and the  $a$ 's simplify considerably.

We now present some analytical comparisons among the local powers of the four tests for a number of distributions within the one-parameter exponential family. Let  $\Pi_i$  and  $\Pi_j$  be the power functions, up to order  $n^{-1/2}$ , of the tests that use the statistics  $S_i$  and  $S_j$ , respectively, with  $i \neq j$  and  $i, j = 1, 2, 3, 4$ . We have,

$$\Pi_i - \Pi_j = \frac{1}{\sqrt{n}} \sum_{k=0}^3 (a_{jk} - a_{ik}) G_{1+2k,\lambda}(x). \tag{4}$$

It is well known that

$$G_{m,\lambda}(x) - G_{m+2,\lambda}(x) = 2g_{m+2,\lambda}(x), \tag{5}$$

where  $g_{v,\lambda}(x)$  is the probability density function of a non-central chi-square random variable with  $v$  degrees of freedom and non-centrality parameter  $\lambda$ . From (4) and (5), we can state the following comparison among the powers of the four tests. Here, we assume that  $\theta > \theta^{(0)}$ ; opposite inequalities hold if  $\theta < \theta^{(0)}$ .

1. Normal ( $\theta > 0, -\infty \leq \mu \leq \infty$  and  $x \in \mathbb{R}$ ):
  - $\mu$  known:  $\alpha(\theta) = (2\theta)^{-1}$ ,  $\zeta(\theta) = \theta^{1/2}$ ,  $d(x) = (x - \mu)^2$  and  $v(x) = -\{\log(2\pi)\}/2$ ,  $\Pi_4 > \Pi_3 > \Pi_1 > \Pi_2$ .
  - $\theta$  known:  $\alpha(\mu) = -\mu/\theta$ ,  $\zeta(\mu) = \exp\{\mu^2/(2\theta)\}$ ,  $d(x) = x$  and  $v(x) = -\{x^2 + \log(2\pi\theta)\}/2$ ,  $\Pi_1 = \Pi_2 = \Pi_3 = \Pi_4$ .
2. Inverse normal ( $\theta > 0, \mu > 0$  and  $x > 0$ ):
  - $\mu$  known:  $\alpha(\theta) = \theta$ ,  $\zeta(\theta) = \theta^{-1/2}$ ,  $d(x) = (x - \mu)^2/(2\mu^2x)$  and  $v(x) = -\{\log(2\pi x^3)\}/2$ ,  $\Pi_1 > \Pi_4 > \Pi_2 = \Pi_3$ .
  - $\theta$  known:  $\alpha(\mu) = \theta/(2\mu^2)$ ,  $\zeta(\mu) = \exp\{-\theta/\mu\}$ ,  $d(x) = x$  and  $v(x) = -\{\theta/(2x) - \log(\theta/(2\pi x^3))\}/2$ ,  $\Pi_4 > \Pi_3 > \Pi_1 > \Pi_2$ .
3. Gamma ( $k > 0, k$  known,  $\theta > 0$  and  $x > 0$ ):  $\alpha(\theta) = \theta$ ,  $\zeta(\theta) = \theta^{-k}$ ,  $d(x) = x$  and  $v(x) = (k - 1) \log(x) - \log\{\Gamma(k)\}$ ,  $\Gamma(\cdot)$  is the gamma function,  $\Pi_4 > \Pi_1 > \Pi_2 = \Pi_3$ .

4. Truncated extreme value ( $\theta > 0$  and  $x > 0$ ):  $\alpha(\theta) = \theta^{-1}$ ,  $\zeta(\theta) = \theta$ ,  $d(x) = \exp(x) - 1$  and  $v(x) = x$ ,  $\Pi_4 > \Pi_3 > \Pi_1 > \Pi_2$ .
5. Pareto ( $\theta > 0$ ,  $k > 0$ ,  $k$  known and  $x > k$ ):  $\alpha(\theta) = 1 + \theta$ ,  $\zeta(\theta) = (\theta k^\theta)^{-1}$ ,  $d(x) = \log(x)$  and  $v(x) = 0$ ,  $\Pi_4 > \Pi_1 > \Pi_2 = \Pi_3$ .
6. Laplace ( $\theta > 0$ ,  $-\infty < k < \infty$ ,  $k$  known and  $x > 0$ ):  $\alpha(\theta) = \theta^{-1}$ ,  $\zeta(\theta) = 2\theta$ ,  $d(x) = |x - k|$  and  $v(x) = 0$ ,  $\Pi_4 > \Pi_3 > \Pi_1 > \Pi_2$ .
7. Power ( $\theta > 0$ ,  $\phi > 0$ ,  $\phi$  known and  $x > \phi$ ):  $\alpha(\theta) = 1 - \theta$ ,  $\zeta(\theta) = \theta^{-1}\phi^\theta$ ,  $d(x) = \log(x)$  and  $v(x) = 0$ ,  $\Pi_4 > \Pi_1 > \Pi_2 = \Pi_3$ .

## 6 Discussion

The gradient test can be an interesting alternative to the classic large-sample tests, namely the likelihood ratio, Wald and Rao score tests. It is competitive with the other three tests since none is uniformly superior to the others in terms of second order local power as we showed. Unlike the Wald and the score statistics, the gradient statistic does not require to obtain, estimate or invert an information matrix, which can be an advantage in complex problems.

Theorem 3 in Terrell (2002) points to another important feature of the gradient test. It suggests that we can, in general, improve the approximation of the distribution of the gradient statistic by a chi-square distribution under the null hypothesis by using a less biased estimator to  $\theta$ . It is well known that the maximum likelihood estimator can be bias-corrected using Cox and Snell (1968) results or the approach proposed by Firth (1993). The effect of replacing the maximum likelihood estimator by its bias-corrected versions will be studied in future research. Note that, unlike LR and  $S_R$ , the gradient statistic is not invariant under non-linear reparameterizations, as is the case of  $W$ . However, we can improve its performance, under the null hypothesis, by choosing a parameterization under which the maximum likelihood estimator is nearly unbiased.

Our results are quite general, and can be specified to important classes of statistical models, such as the generalised linear models. Local power comparisons of the three usual large-sample tests in generalised linear models are presented by Cordeiro et al. (1994) and Ferrari et al. (1997). The extension of their studies to include the gradient test will be reported elsewhere.

As a final remark, the power comparisons performed in the present paper consider the four tests in their original form, i.e. they are not corrected to achieve local unbiasedness; see Rao and Mukerjee (1997) and references therein for this alternative approach. In fact, this approach can be explored in future work for the gradient test.

## References

- 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.
- Bracewell, R. (1999). *The Fourier transform and its applications* (3rd ed.). New York: McGraw-Hill.
- Cordeiro, G. M., Botter, D. A., Ferrari, S. L. P. (1994). Nonnull asymptotic distributions of three classic criteria in generalised linear models. *Biometrika*, 81, 709–720.



- Cox, D. R., Snell, E. J. (1968). A general definition of residuals (with discussion). *Journal of the Royal Statistical Society B*, 30, 248–275.
- Ferrari, S. L. P., Botter, D. A., Cribari-Neto, F. (1997). Local power of three classic criteria in generalised linear models with unknown dispersion. *Biometrika*, 84, 482–485.
- Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, 80, 27–38.
- Harris, P., Peers, H. W. (1980). The local power of the efficient score test statistic. *Biometrika*, 67, 525–529.
- Hayakawa, T. (1975). The likelihood ratio criterion for a composite hypothesis under a local alternative. *Biometrika*, 62, 451–460.
- Hayakawa, T. (1977). The likelihood ratio criterion and the asymptotic expansion of its distribution. *Annals of the Institute of Statistical Mathematics*, 29, 359–378.
- Hayakawa, T., Puri, M. L. (1985). Asymptotic expansions of the distributions of some test statistics. *Annals of the Institute of Statistical Mathematics*, 37, 95–108.
- Peers, H. W. (1971). Likelihood ratio and associated test criteria. *Biometrika*, 58, 577–587.
- 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 N. Balakrishnan, N. Kannan, H. N. Nagaraja (Eds.), *Advances in ranking and selection, multiple comparisons, and reliability*. Boston: Birkhauser.
- Rao, C. R., Mukerjee, R. (1997). Comparison of LR, score, and Wald tests in a non-iid setting. *Journal of Multivariate Analysis*, 60, 99–110.
- Terrell, G. R. (2002). The gradient statistic. *Computing Science and Statistics*, 34, 206–215.
- 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.