International Journal of Applied Engineering Research, ISSN 0973-4562 Volume 13, Number 23 (2018) pp. 16368–16375

Gradient Statistic: An option for conducting hypothesis testing in small sample size scenarios

Sébastien Lozano Forero, Vladimir Ballesteros Ballesteros, and Jorge Luis Nisperuza Toledo
Fundación Universitaria los Libertadores, Facultad de Ingeniería y Ciencias Básicas, Bogotá, Colombia.

Abstract: The gradient statistic, recently proposed in the literature, has gained attention on statistical practitioners due to the fact that it is a competitive alternative for the traditional statistics (likelihood ratio, Wald, Score) for performing hypothesis testing in parametric models with the special property of being easily computable. In this work, we present an exhaustive Monte Carlo simulation study with the four test statistics to assess its performance in a controlled scenario based on the exponential distribution. The obtained results suggest that the gradient test statistics is, indeed, a competitive option for the "Holy trinity" of statistical inference (Likelihood ratio, Score and Wald test statistics).

AMS Subject classification:
Keywords: hypothesis testing, gradient statistic, power of test and exponential distribution.

OVERVIEW

Initially, let us assume that, based in a parametric model, the following hypothesis testing should be performed:

$$\begin{cases} H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0, \\ H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0 \end{cases}$$

where $\boldsymbol{\theta}_0$ is a specified vector and $\boldsymbol{\theta}_0$ is the parametric vector indexing the initial parametric model.

The tests based on large samples approximations are often used in statistics, due to the fact that exact tests are not always available. These tests are called "first-order asymptotic", that is, they are based on critical values obtained from a known null limit distribution. A natural problem that arises is whether the first order approximation is adequate for the null distribution of the test statistic under consideration. The best known asymptotic test statistics whose reference distributions is chi-square are: likelihood ratio, score and Wald (often referred as the "holy trinity" of statistical inference). The statistics of these three tests are equivalent in large sample sizes and, in regular problems, converge under the null hypothesis H_0, to the distribution χ^2_q, where q is the number of restrictions imposed by H_0. In small samples, the first-order approximation may not be satisfactory and may lead to quite distorted null hypothesis rejection rates. The statistics involved and which will be considered are: Likelihood ratio, Wald and Score. The main idea is to test the hypothesis H_0. According to the notation used in Lemonte (2016), the Likelihood ratio, Wald and Score test statistics are, respectively, defined as:

$$\begin{aligned} S_{LR} &= 2[\ell(\hat{\boldsymbol{\theta}}) - \ell(\boldsymbol{\theta}_0)], \\ S_W &= (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T \mathbf{K}(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0), \\ S_R &= \mathbf{U}(\boldsymbol{\theta})^T \mathbf{K}(\boldsymbol{\theta}_0)^{-1} \mathbf{U}(\boldsymbol{\theta}_0) \end{aligned}$$

where $\hat{\boldsymbol{\theta}}$ is the maximum likelihood estimator of $\boldsymbol{\theta}$ which may be obtained from $\mathbf{U}(\hat{\theta}) = 0$ where $\mathbf{U}(\boldsymbol{\theta})$ is the score vector defined as

$$\mathbf{U}(\boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$$

and $\ell(\boldsymbol{\theta})$ is defined as the log-likelihood function, given by

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log f(x_i \mid \boldsymbol{\theta})$$

where $f(. \mid \boldsymbol{\theta})$ is the probability function of the random sample $(x_1, \cdots, x_n)^T$.

The Fischer information matrix is a way of measuring the amount of information a random variable contains over an unknown parameter vector and upon which the distribution depends. It is defined as

$$\mathbf{K}(\boldsymbol{\theta}) = E[\mathbf{U}(\boldsymbol{\theta})\mathbf{U}(\boldsymbol{\theta})^T] = -E\left(\frac{\partial \mathbf{U}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}\right)$$

Is now important to have tools to better address the problem of small-sample test statistics. A first attempt is given by a modified version of the Wald statistic, which is defined as

$$S_{WM} = (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T K(\boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$$

where a correction on the quantity that modifies the difference between the estimator and the parameter $\boldsymbol{\theta}_0$, is made. This is, the only difference between S_W and S_{WM} is the object to be evaluated in $\mathbf{K}(\boldsymbol{\theta})$

$$S_W \to \mathbf{K}(\hat{\theta}), \qquad S_{WM} \to \mathbf{K}(\boldsymbol{\theta}_0)$$

The test statistic S_{LR}, S_W, S_R e S_{WM} have an approximate chi-square distribution (central) with p degrees of freedom (χ^2_p) and, under null hypothesis $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$. we reject H_0 if the observed value of the test statistic exceeds the quantile $100(1-\alpha)\%$ of the distribution χ^2_p, where α is the nominal level of the test.

The proposal of Terrell (2002) is widely known as *Gradient Statistics* can be obtained as follows.

- Find a square matrix such that $\mathbf{L}^T\mathbf{L} = \mathbf{K}(\boldsymbol{\theta})$
- The test statistic S_R and S_{WM} might be re-written as:

$$S_R = [(\mathbf{L}^{-1})^T\mathbf{U}(\boldsymbol{\theta})]^T(\mathbf{L}^{-1})^T\mathbf{U}(\boldsymbol{\theta}_0)$$
$$S_{WM} = [\mathbf{L}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)]^T\mathbf{L}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$$

- Note that approximately

$$(\mathbf{L}^{-1})^T\mathbf{U}(\boldsymbol{\theta}_0) \sim N_p(\mathbf{0}, \boldsymbol{I_p})$$

$$\mathbf{L}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \sim N_p(\mathbf{0}, \boldsymbol{I_p})$$

where $\boldsymbol{I_p}$ is a $p \times p$ identity matrix.

Now, the inner product between $(\mathbf{L}^{-1})^T\mathbf{U}(\boldsymbol{\theta}_0)$ and $\mathbf{L}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ results in

$$\begin{aligned}[(\mathbf{L}^{-1})^T\mathbf{U}(\boldsymbol{\theta}_0)]^T\mathbf{L}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) &= \mathbf{U}(\boldsymbol{\theta}_0)^T\mathbf{L}^{-1}\mathbf{L}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \\ &= \mathbf{U}(\boldsymbol{\theta}_0)^T(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\end{aligned}$$

The proposal of Terrell (2002) is widely known as *Gradient Statistics* can be obtained, as a result of the previous computations, as follows:

Definition 1. The Gradient Statistic, S_T, to test the simple null hypothesis $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ against $H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ has the form

$$S_T = \mathbf{U}(\boldsymbol{\theta}_0)^T(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$$

Note that this statistic is derived from the Rao statistic and modified Wald statistic (see Lemonte (2010)). The following are some important properties of this new statistic

- Under H_0, S_T has an approximate distribution χ^2_p.
- S_T is very simple to calculate, not involving estimation of the information matrix nor the calculation of its inverse.
- It is not evident that S_T is nonnegative, but it must be asymptotically nonnegative.

COMPOSED NULL HYPOTHESIS

Now, we take a quick look at the case where the null and compound hypothesis. First, partition the vector $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_p)^T$ such that

$$\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T)^T.$$

where $\boldsymbol{\theta}_1 = (\theta_1, \cdots, \theta_q)^T$ e $\boldsymbol{\theta}_2 = (\theta_{q+1}, \cdots, \theta_p)^T$. Now, consider to test $H_0 : \boldsymbol{\theta}_2 = \boldsymbol{\theta}_{2,0}$ against $H_0 : \boldsymbol{\theta}_2 \neq \boldsymbol{\theta}_{2,0}$ em que $\boldsymbol{\theta}_{2,0}$ is a vector of known constants. To test this hypothesis, one can use the statistics S_{LR}, S_W, S_R, S_T given by

$$\begin{aligned} S_{LR} &= 2[\ell(\hat{\boldsymbol{\theta}}) - \ell(\tilde{\boldsymbol{\theta}})], \\ S_W &= (\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}})^T\mathbf{K}(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}) \\ S_R &= \mathbf{U}(\tilde{\boldsymbol{\theta}})^T\mathbf{K}(\tilde{\boldsymbol{\theta}})^{-1}\mathbf{U}(\tilde{\boldsymbol{\theta}}) \\ S_T &= \mathbf{U}(\tilde{\boldsymbol{\theta}})^T(\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}) \end{aligned}$$

where $\hat{\boldsymbol{\theta}}$ and $\tilde{\boldsymbol{\theta}}$ are the unrestricted maximum likelihood estimators (under H_1) and restricted (under H_0) of $\boldsymbol{\theta}$ respectively. All the statistics S_{LR}, S_W, S_R and S_T follow approximately χ^2_{p-q} distribution under H_0.

Now, there are natural questions regarding the performance of the statistics under study. The first fundamental fact to be solved is to establish when the gradient statistic can be considered in relation to the classic test statistics. We have the following facts

- The simple form of S_T, which in practice can be the simplest to calculate, is an interesting feature.
- In complex problems, not having to calculate, estimate and invert a Fischer information matrix is at an even more positive point.
- For example, problems in survival analysis in which there is censorship, the S_T statistic could be used without problems and thus would be an alternative to the likelihood ratio statistic

There is still the natural question of the will to compare the proposed statistics. We want to know if S_T is more, less or equally powerful than the other statistics. In order to answer, we will study the local power of the gradient test. According to Lemonte (2013), the following strategy was proposed

1. to present the asymptotic expansion of the local power function (up to order $n^-1/2$) of the gradient test under the sequence of alternative hypotheses
$$H_{1n} : \boldsymbol{\theta}_2 = \boldsymbol{\theta}_{2,0} + \frac{\epsilon}{\sqrt{n}}$$
where $\epsilon = \sqrt{n}(\boldsymbol{\theta}_2 - \boldsymbol{\theta}_{2,0}) = (\epsilon_{q+1}, \cdots, \epsilon_p)^T$
2. Make a local power study of the gradient test by comparing it with the local power of the likelihood ratio tests, Wald and score.

Some math involved due to Lemonte (2013):

- Derivatives of the log-likelihood function
$$y_r = n^{-1/2}\frac{\partial \ell(\boldsymbol{\theta})}{\partial \theta_r}, \qquad y_{rs} = n^{-1}\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \theta_r \theta_s},$$
$$y_{rst} = n^{-3/2}\frac{\partial^3 \ell(\boldsymbol{\theta})}{\partial \theta_r \partial \theta_s \partial \theta_t},$$
with $r, s, t = 1, \cdots, p$
- The arrays
$$\boldsymbol{y} = (y_1, \cdots, y_p)^T, \quad \boldsymbol{Y} = ((y_{rs})), \quad \boldsymbol{Y}_{...} = ((y_{rst})).$$
- Cumulants
$$\kappa_{rs} = E(y_{rs}), \quad \kappa_{r,s} = E(y_r y_s), \quad \kappa_{rst} = \sqrt{n}E(y_{rst}).$$
$$\kappa_{r,st} = \sqrt{n}E(y_r y_{st}) \qquad \kappa_{r,s,t} = \sqrt{n}E(y_r y_s y_t)$$
com $r, s, t = 1, \cdots, p$.
- The respective arrays
$$\boldsymbol{K} = ((\kappa_{r,s})) = -((\kappa_{rs})), \qquad , \boldsymbol{K}_{...} = ((\kappa_{rst})),$$
$$\boldsymbol{K}_{.,..} = ((\kappa_{r,st})), \qquad , \boldsymbol{K}_{.,.,.} = ((\kappa_{r,s,t})),$$
- For quantities with three indexes, the adopted notation would be:
$$\boldsymbol{K}_{...} \circ \boldsymbol{a} \circ \boldsymbol{b} \circ \boldsymbol{c} = \sum_{r,s,t=1}^{p} \kappa_{rst} a_r b_s c_t$$
$$\boldsymbol{K}_{...} \circ \boldsymbol{M} \circ \boldsymbol{b} = \sum_{r,s,t=1}^{p} \kappa_{r,st} m_{rs} b_t$$
where $\boldsymbol{M} = ((m_{rs}))$ is a matrix $p \times p$ and $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ are vectors $p-$ dimensional
- Define the matrices:
$$\boldsymbol{K} = \begin{bmatrix} \boldsymbol{K}_{11} & \boldsymbol{K}_{12} \\ \boldsymbol{K}_{21} & \boldsymbol{K}_{22} \end{bmatrix} \qquad \boldsymbol{A} = \begin{bmatrix} \boldsymbol{K}_{11}{}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}$$
$$\boldsymbol{M} = \boldsymbol{K}^{-1} - \boldsymbol{A}, \qquad \boldsymbol{K}_{22.1} = \boldsymbol{K}_{22} - \boldsymbol{K}_{21}\boldsymbol{K}_{11}{}^{-1}\boldsymbol{K}_{12}$$
- The vector
$$\boldsymbol{\epsilon}^* = \begin{bmatrix} \boldsymbol{K}_{11}{}^{-1}\boldsymbol{K}_{12} \\ \boldsymbol{I}_{p-q} \end{bmatrix} \boldsymbol{\epsilon}$$
where $\boldsymbol{I}_{p-q}$ is the identity matrix of order $p - q$.

MOMENT GENERATING FUNCTION

The moment generating function can be rewritten as
$$M(t) = (1 - 2t)^{\frac{1}{2}(p-q)} \exp\left(\frac{t}{1-2t}\boldsymbol{\epsilon}^\top \boldsymbol{K}_{22.1}^{\dagger}\boldsymbol{\epsilon}\right) \times \left[1 + \frac{1}{\sqrt{n}}\sum_{k=0}^{3} a_k(1-2t)^{-k}\right] + O(n^{-1}) \tag{1}$$
where $a_0 = -(a_1 + a_2 + a_3)$ and
$$\begin{aligned} a_1 = \frac{1}{4}\{&\boldsymbol{K}^\dagger \circ (\boldsymbol{K}^{-1})^\dagger \circ (\boldsymbol{\epsilon}^*)^\dagger - 2(\boldsymbol{K}_{...} + 2\boldsymbol{K}_{.,..})^\dagger \\ &\circ (\boldsymbol{\epsilon}^*)^\dagger \circ (\boldsymbol{\epsilon}^*)^\dagger \circ (\boldsymbol{\epsilon}^*)^\dagger \\ &- (4\boldsymbol{K}_{.,..} + 3\boldsymbol{K}_{...})^\dagger \circ \boldsymbol{A}^\dagger \circ (\boldsymbol{\epsilon}^*)^\dagger \\ &- 2(\boldsymbol{K}_{2..} + \boldsymbol{K}_{2,..})^\dagger \circ \boldsymbol{\epsilon} \circ (\boldsymbol{\epsilon}^*)^\dagger \circ (\boldsymbol{\epsilon}^*)^\dagger\} \end{aligned} \tag{2}$$
$$\begin{aligned} a_2 = -\frac{1}{4}\{&\boldsymbol{K}^\dagger \circ M^\dagger \circ (\boldsymbol{\epsilon}^*)^\dagger (\boldsymbol{K}_{...} + 2\boldsymbol{K}_{.,..})^\dagger \\ &\circ (\boldsymbol{\epsilon}^*)^\dagger \circ (\boldsymbol{\epsilon}^*)^\dagger \circ (\boldsymbol{\epsilon}^*)^\dagger\} \end{aligned} \tag{3}$$
$$a_3 = -\frac{1}{12}\boldsymbol{K}_{...}^\dagger \circ (\boldsymbol{\epsilon}^*)^\dagger \circ (\boldsymbol{\epsilon}^*)^\dagger \circ (\boldsymbol{\epsilon}^*)^\dagger \tag{4}$$
Referring the function $M(t)$, the following result is derived.

International Journal of Applied Engineering Research, ISSN 0973-4562 Volume 13, Number 23 (2018) pp. 16368–16375

MAIN RESULT

The following result is due to Lemonte (2013): The asymptotic expansion of the distribution of the gradient statistic (S_T) to a composite hypothesis under a sequence of local alternatives converging to the null hypothesis at a convergence rate of $n^{-1/2}$ is as follows:

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

where $G_{q,\lambda}(x)$ is the probability density function of a non-central chi-square random variable with q degrees of freedom and non-centrality parameter λ. Here $f = p - q$, $\lambda = \boldsymbol{\epsilon}^T \boldsymbol{K}_{22.1}^{\dagger} \boldsymbol{\epsilon}/2$ and the quantities a_k's were previously defined.

Corolary 2. The asymptotic expansion of the distribution of the gradient statistic S_T to a simple hypothesis under a sequence of local alternatives converging to the null hypothesis at a convergence rate of $n^{1/2}$ is as follows

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

where $\lambda = \boldsymbol{\epsilon}^T \boldsymbol{K}^{\dagger} \boldsymbol{\epsilon}/2$ e

$$a_0 = \frac{1}{6} \boldsymbol{K}^{\dagger}{}_{...}(\circ \boldsymbol{\epsilon})^3,$$

$$a_1 = -\frac{1}{4} \{\boldsymbol{K}\dagger_{...} \circ (\boldsymbol{K}^{-1})^{\dagger} \circ \boldsymbol{\epsilon} - 2\boldsymbol{K}^{\dagger}_{.,..}(\circ \boldsymbol{\epsilon})^3\}$$

$$a_2 = \frac{1}{4} \{\boldsymbol{K}^{\dagger}{}_{...} \circ (\boldsymbol{K}^{-1})^{\dagger} \circ \boldsymbol{\epsilon} - (\boldsymbol{K}_{...} + 2\boldsymbol{K}_{.,..})^{\dagger}(\circ \boldsymbol{\epsilon})^3\}$$

COMPARISON BETWEEN POWER OF THE CONSIDERED TEST

The main idea now is to be able to compare the podes of the tests that are being studied. The following facts held

1. Until the first order, the statistics S_{LR}, S_W, S_R and S_T have the same asymptotic properties under a null hypothesis H_0, or under an alternative local hypothesis.
2. Up to an error of order n^{-1}, the corresponding tests have the same size, however, their powers differ by the order term $n^{-1/2}$
3. Thus the powers of the different tests can be compared on the basis of the expansions of their power functions ignoring lesser terms of order than $n^{1/2}$

The local power function of the tests that use statistics is defined S_{LR}, S_W, S_R e S_T as

$$\Pi_i = 1 - P(S_i \leq x_\gamma) = P(S_i > x_\gamma), \quad i = LR, W, R, T$$

where x_γ represents the proper quantile of the χ^2_{p-q} distribution for a choose nominal level γ and

$$P(S_i \geq x_\gamma) = G_{p-q,\lambda}(x_\gamma) + \frac{1}{\sqrt{n}} \sum_{k=0}^{3} a_{ik} G_{p-q+2k,\lambda}(x_\gamma) + O(n^{-1})$$

SIMULATION RESULTS

Until this point, the different test statistics have been presented and, as mentioned before, there may be differences in the powers of the four tests in small sample sizes, which is the reason the present simulation is made, to make a comparison between the powers of the tests that are generally used (Likelihood ratio, Score and Wald) and the statistical gradient for testing as well. Simulations are made via Montecarlo for the exponential distribution function in the following form:

$$f_X(x \mid \theta) = \theta e^{-x\theta}$$

with $\theta > 0$ and $x > 0$. The corresponding log-likelihood function is given by :

$$\Pi_{i=1}^n f_X(x_i \mid \theta) = \Pi_{i=1}^n \theta e^{-x_i\theta} = \theta^n e^{\theta \sum_{i=1}^n x_i}$$

We have that

$$\begin{aligned} \ell(\theta) &= \log(\Pi_{i=1}^n f_X(x_i \mid \theta)) = \log\left(\theta^n e^{\theta \sum_{i=1}^n x_i}\right) \\ &= -n \log(\theta) + \theta \sum_{i=1}^n x_i \end{aligned} \tag{5}$$

To obtain the maximum likelihood estimator of θ, we make

$$\mathbf{U}(\theta) = \frac{\partial \ell(\theta)}{\partial \theta} = \frac{-n}{\theta} + \sum_{i=1}^n x_i = 0$$

this way, we obtain the following estimator

$$\hat{\theta} = \frac{1}{\bar{X}}$$

Since some of the test statistics depend on the Fisher information matrix, this corresponds to:

$$\frac{n}{t} - \sum_{i=0}^n x_i$$

International Journal of Applied Engineering Research, ISSN 0973-4562 Volume 13, Number 23 (2018) pp. 16368–16375

where n is the sample size.

We fixed the number of Monte Carlo replicas in 10,000. In each Monte Carlo simulation, samples for different sample sizes (10,30,50 and 100) and different levels of significance

$$(\alpha = 0.1, \alpha = 0.05, \alpha = 0.01)$$

were generated in order to have the value of the parameter $\theta = 1$ as the value of the real parameter and for the generation of the samples.

To calculate the power of the test at each step the sample was generated with the actual value θ and the hypothesis $\theta = \theta_0$ against $\theta \neq 0$ was tested. For different values of θ at the end the percentage of times that was rejected test will be the measure of the power in each case.

RESULTS

For a sample size of 10 and for values lower than 1, the gradient statistic was found to be more potent, even than the likelihood ratio test statistic. Although the differences in powers are not very large and all of them tend to have the same behavior from the same point. These behaviors were observed at the three levels of significance, these results can be seen in the table 13 and in the figures 1, 2, 3.

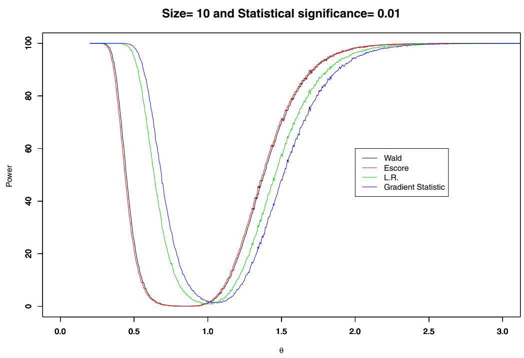

FIGURE 1. Power of the considered test statistics with $n = 10$ and $\alpha = 0.01$

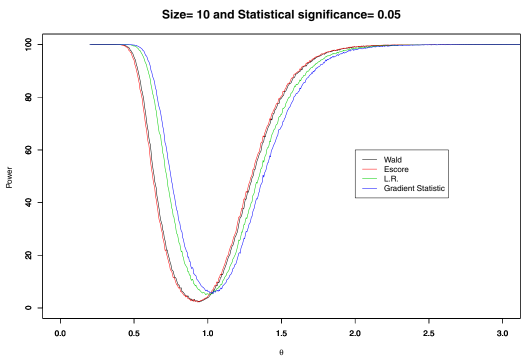

FIGURE 2. Power of the considered test statistics with $n = 10$ and $\alpha = 0.05$

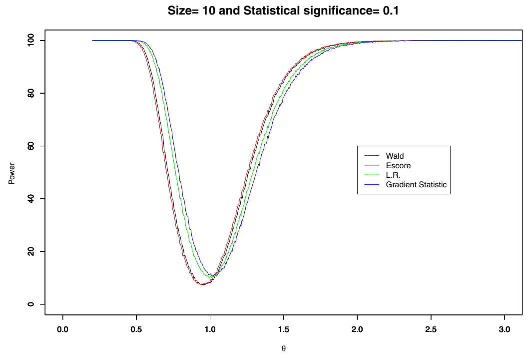

FIGURE 3. Power of the considered test statistics with $n = 10$ and $\alpha = 0.1$

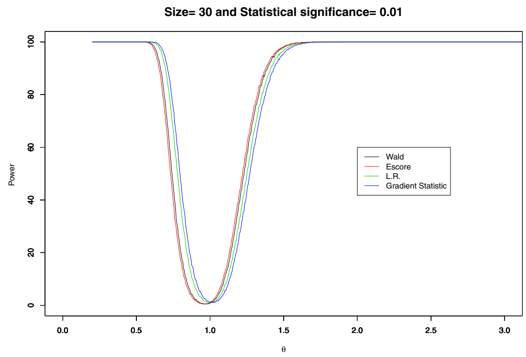

FIGURE 4. Power of the considered test statistics with $n = 30$ and $\alpha = 0.01$

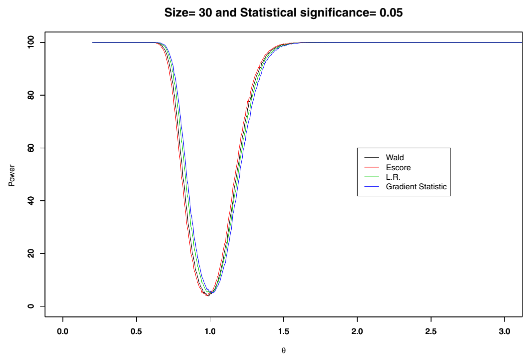

FIGURE 5. Power of the considered test statistics with $n = 30$ and $\alpha = 0.05$

For a sample size of 30 and for all values of θ_0, the statistics tend to have the same behavior, the three present values close to their power values and in this case the curves

International Journal of Applied Engineering Research, ISSN 0973-4562 Volume 13, Number 23 (2018) pp. 16368–16375

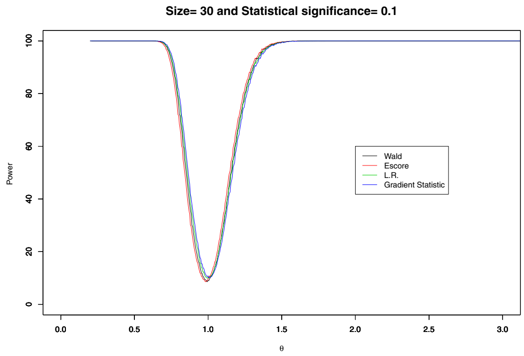

FIGURE 6. Power of the considered test statistics with $n = 30$ and $\alpha = 0.1$

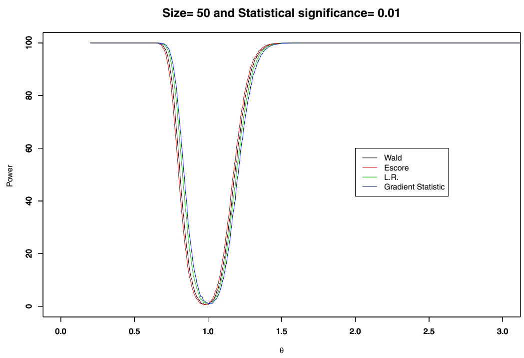

FIGURE 7. Power of the considered test statistics with $n = 50$ and $\alpha = 0.01$

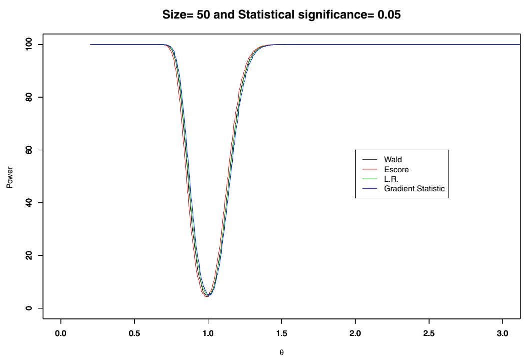

FIGURE 8. Power of the considered test statistics with $n = 50$ and $\alpha = 0.05$

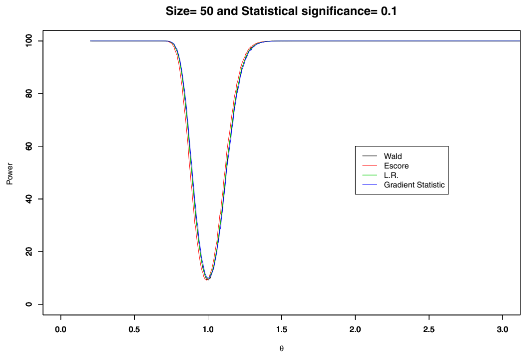

FIGURE 9. Power of the considered test statistics with $n = 50$ and $\alpha = 0.1$

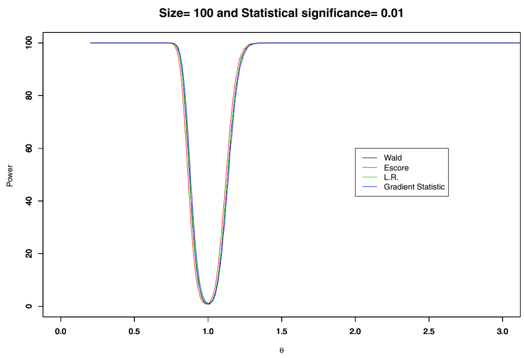

FIGURE 10. Power of the considered test statistics with $n = 100$ and $\alpha = 0.01$

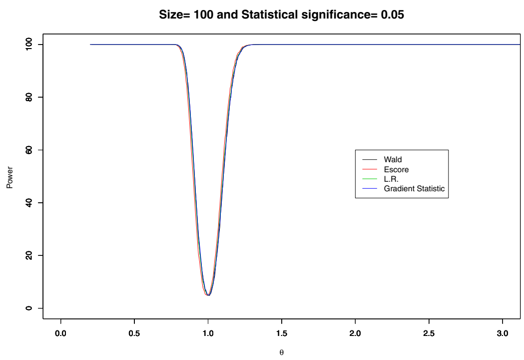

FIGURE 11. Power of the considered test statistics with $n = 100$ and $\alpha = 0.05$

are faster than in the case The results of this study are shown in the table below and in the graphs in the following table: 14 and figures 4, 5, 6.

For sample sizes of 50 and 100, the power curves are much faster growing, that is to say that the tests are much more sensitive to small changes in the value of θ real and θ_0, besides we see that in these large sample sizes, the differences between the four tests are really minimal, so it makes no difference in how much power in the exponential distribution use any of the four tests. This can be observed in tables 15 and 16, and figures 7,8,9, 10,11,12

In summary, we conclude, via Monte Carlo simulations, that for small sample sizes differing in the power of the test is influenced by the true value of the parameter and that on average the behavior of the four tests exposed is the same, in addition we see that for large sample sizes the

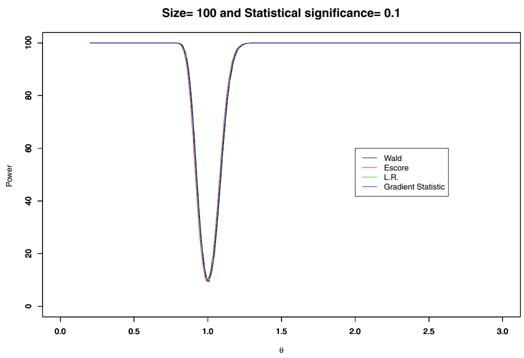

FIGURE 12. Power of the considered test statistics with $n = 100$ and $\alpha = 0.1$

Size 10					
Significance level	Value of θ	Wald	Score	RV	Gradient
10%	0.1	100	100	100	100
	0.4	100	100	100	100
	0.7	71.5	71.5	78.7	81.8
	1	30.6	30.6	25.1	24.1
	1.3	98.3	98.3	97.9	97.2
	1.6	100	100	100	100
	1.9	100	100	100	100
	2.2	100	100	100	100
5%	0.1	100	100	100	100
	0.4	100	100	100	100
	0.7	52.8	52.8	66.5	70.6
	1	24.7	24.7	19.7	16.6
	1.3	97.3	97.3	95.3	94.5
	1.6	100	100	100	100
	1.9	100	100	100	100
	2.2	100	100	100	100
1%	0.1	100	100	100	100
	0.4	100	100	100	100
	0.7	16.4	16.4	38	46.5
	1	10	10	6.3	5
	1.3	93.4	93.4	89.4	86
	1.6	100	100	100	99.9
	1.9	100	100	100	100
	2.2	100	100	100	100

FIGURE 13. Behaviour of the considered test statistics with $n = 10$

differences in powers of the tests are almost imperceptible

Size 30					
Significance level	Value of θ	Wald	Score	RV	Gradient
10%	0.1	100	100	100	100
	0.4	100	100	100	100
	0.7	71.5	71.5	78.7	81.8
	1	30.6	30.6	25.1	24.1
	1.3	98.3	98.3	97.9	97.2
	1.6	100	100	100	100
	1.9	100	100	100	100
	2.2	100	100	100	100
5%	0.1	100	100	100	100
	0.4	100	100	100	100
	0.7	52.8	52.8	66.5	70.6
	1	24.7	24.7	19.7	16.6
	1.3	97.3	97.3	95.3	94.5
	1.6	100	100	100	100
	1.9	100	100	100	100
	2.2	100	100	100	100
1%	0.1	100	100	100	100
	0.4	100	100	100	100
	0.7	16.4	16.4	38	46.5
	1	10	10	6.3	5
	1.3	93.4	93.4	89.4	86
	1.6	100	100	100	99.9
	1.9	100	100	100	100
	2.2	100	100	100	100

FIGURE 14. Behaviour of the considered test statistics with $n = 30$

and by both asymptotically these tests are equivalent. As the gradient statistic is much simpler in its calculation it is an attractive alternative since it is equally powerful for the three tests that are commonly used and does not need a complex evaluation as to its calculation.

CONCLUSIONS

- All four considered test statistics are locally unbiased
- If $\boldsymbol{K}_{\dots} = \mathbf{0}$ the likelihood ratio, Wald and gradient tests have identical local power properties.
- If $\boldsymbol{K}_{\dots} = 2\boldsymbol{K}_{.,..}$ the Score and gradient test statistics have identical local power properties.

Size 50					
Significance level	Value of θ	Wald	Score	RV	Gradient
10%	0.1	100	100	100	100
	0.4	100	100	100	100
	0.7	91.7	91.7	94.1	95.1
	1	43.3	43.3	39.3	37.6
	1.3	100	100	99.7	99.7
	1.6	100	100	100	100
	1.9	100	100	100	100
	2.2	100	100	100	100
5%	0.1	100	100	100	100
	0.4	100	100	100	100
	0.7	84.5	84.5	89.5	91.6
	1	34.3	34.3	29.6	27.3
	1.3	99.5	99.5	99.4	99.4
	1.6	100	100	100	100
	1.9	100	100	100	100
	2.2	100	100	100	100
1%	0.1	100	100	100	100
	0.4	100	100	100	100
	0.7	48.5	48.5	66	72.4
	1	16.4	16.4	11	8.4
	1.3	99	99	98.8	98.1
	1.6	100	100	100	100
	1.9	100	100	100	100
	2.2	100	100	100	100

FIGURE 15. Behaviour of the considered test statistics with $n = 50$

Size 100					
Significance level	Value of θ	Wald	Score	RV	Gradient
10%	0.1	100	100	100	100
	0.4	100	100	100	100
	0.7	99.7	99.7	99.7	99.7
	1	61.4	61.4	58.2	55.9
	1.3	100	100	100	100
	1.6	100	100	100	100
	1.9	100	100	100	100
	2.2	100	100	100	100
5%	0.1	100	100	100	100
	0.4	100	100	100	100
	0.7	99.4	99.4	99.6	99.6
	1	54.5	54.5	50.4	48.3
	1.3	100	100	100	100
	1.6	100	100	100	100
	1.9	100	100	100	100
	2.2	100	100	100	100
1%	0.1	100	100	100	100
	0.4	100	100	100	100
	0.7	92.9	92.9	96.4	97.6
	1	35.4	35.4	30.2	26.8
	1.3	100	100	100	100
	1.6	100	100	100	100
	1.9	100	100	100	100
	2.2	100	100	100	100

FIGURE 16. Behaviour of the considered test statistics with $n = 100$

- There is no uniform superiority of one test statistic over others.

REFERENCES

[1] Harris, P. & Peers, H.W. (1980). *The local power of the efficient score test statistic*. *Biometrika* **67**, 525–529.

[2] Hayakawa, T. (1975). *The likelihood ratio criterion for a composite hypothesis under a local alternative*. *Biometrika* **62**, 451–460.

[3] Peers, H.W. (1971). *Likelihood ratio and associated test criteria*. *Biometrika* **58**, 577–587.

[4] Rao, C.R. (2005). *Likelihood ratio and associated test criteria*. *Biometrika* **58**, 577–587.

[5] Lemonte, A (2010). *Estatistica gradiente e refinamento de mtodos assintticos no modelo de regression Birnbaum-Saunders*. *Tese de doutorado* USP, Brazil.

[6] Lemonte, A (2016). *The Gradient Statistic*. *Academic Press*.

[7] 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.