Local Level of the Wald Tests When the Information Is Zero

Livello locale dei test di Wald quando l'informazione è nulla

Matteo Bottai
Istituto Cnuce, CNR, Via Moruzzi 1, Pisa 56124
e-mail: matteo.bottai@cnuce.cnr.it

Riassunto: E' stato studiato il livello di test di Wald in modelli parametrici unidimensionali quando l'informazione di Fisher è nulla per un valore critico nello spazio dei parametri. In questo caso, la probabilità di commettere un errore di tipo I può essere molto elevata. Esistono, infatti, dei valori del parametro vicini al valore critico per cui il livello del test è non inferiore a 0.5 qualsiasi sia la distribuzione di riferimento utilizzata.

Keywords: Identifiability; Likelihood Ratio Test; Rate of Convergence; Uniform Coverage Probability

1 Introduction

Suppose that (Y, X) is a bivariate random vector. Assume that the marginal distribution of X is known and conditional on X, $Y = \exp(\theta X) - \theta X + \epsilon$, where ϵ is normally distributed random variable with mean zero and variance one. Note that the first derivative of the likelihood function with respect to the parameter θ is identically equal to zero when evaluated at $\theta = 0$. In general, suppose that $Y_1, ..., Y_n$ are n independent copies of a random variable Y with density $f(y;\theta)$ with respect to a dominating measure μ. Let $l(y;\theta)$ denote $\log f(y;\theta)$ and $l^{(j)}(y;\theta)$ its j^{th} derivative with respect to θ. Let $L_n(\theta)$ and $L_n^{(j)}(\theta)$ denote the sum of $l(y;\theta)$ and $l^{(j)}(y;\theta)$ respectively over the sample observations. Suppose that at the point θ^* in the parameter space,

$$l^{(1)}(Y;\theta^*) = 0 \tag{1}$$

almost everywhere. We are interested in studying the large sample properties of the test procedures the rejects for large values of the following statistics

$$\begin{aligned} W_n^{(1)}(\theta) &= -(\hat{\theta} - \theta)^2 L_n^{(2)}(\hat{\theta}) \\ W_n^{(2)}(\theta) &= n(\hat{\theta} - \theta)^2 i\left(\hat{\theta}\right) \end{aligned}$$

where $\hat{\theta}$ denotes the maximum likelihood estimator of θ, and $i(\theta) \equiv E_\theta[-l^{(2)}(Y;\theta)]$. In regular problems, that is when (1) does not hold, the test statistics $W_n^{(1)}(\theta)$ and $W_n^{(2)}(\theta)$, converge in distribution under θ to a χ_1^2 random variable and the convergence is locally uniform. That is, for all α,

$$\lim_{n\to\infty} \sup_{\theta} \Pr_\theta \left\{ W_n^{(j)}(\theta) \geq c_\alpha \right\} = \alpha$$

where $j = 1, 2$ and c_α is the α point of a χ_1^2 distribution. Furthermore, all tests are asymptotically equivalent (Cox and Hinkley, 1974). That is, under local (Pitman) parameters $\theta_n = \theta + an^{-1/2}$, where a is any non-zero constant, the test statistics converge in distribution to the same non-central chi-squared random variable. Thus test procedures that reject for large values of the statistics $W_n^{(1)}(\theta)$ and $W_n^{(2)}(\theta)$ have the same local power for detecting alternatives that differ from the null point by a quantity of order $O(n^{-1/2})$. Furthermore, the Wald tests are asymptotically equivalent to the test that rejects for large values of the likelihood ratio test statistic. As the likelihood ratio test, they tests are consistent. That is, test procedures that reject for large values of $W_n^{(1)}(\theta)$ and $W_n^{(2)}(\theta)$, have power for detecting a fixed alternative $\theta' \neq \theta$ that converges to one as the sample size converges to infinity.

Rotnitzky *et al.* (2000) pointed out that when the data are generated under parameter values in a shrinking neighborhood of θ^* the log likelihood has an unusual shape, it being bimodal and nearly symmetric around θ^* with probability going to 1/2 as n goes to infinity when the distance between the data generating θ and the critical point θ^* is of order $O\left(n^{-b}\right)$ with $b > 1/6$ and bimodal but asymmetric when this distance is of order $O\left(n^{-1/6}\right)$. Aside from the unusual shape of the likelihood based regions, they also noted that even though the likelihood ratio test statistics converged under θ^* to a random variable that is stochastically smaller than a chi-squared one distribution, its limiting law under some sequences of parameter values in shrinking neighborhoods of θ^* was indeed stochastically larger than the chi-squared one distribution.

We study the asymptotic level of test procedures based on the two commonly used forms of the Wald test statistics in identifiable one-dimensional models that have zero information at a parameter value in the parameter space. We show that tests based on the Wald test statistics have asymptotic type one error that can be dramatically large for null parameter values in a neighborhood of the critical point. Specifically, for tests that reject for large values of $W_n^{(1)}(\theta)$ and $W_n^{(2)}(\theta)$, there exist null points near the critical point such that the rejection probability of the tests converges to a value no smaller than 0.5 regardless of the reference distribution and of the nominal level of the test.

2 Asymptotic local level

In this section we show that when (1) holds, under some regularity conditions, the level of the two commonly used Wald test procedures of the null hypothesis $H_0 : \theta = \theta_0$ that reject for large values of $W_n^{(1)}(\theta_0)$ and $W_n^{(2)}(\theta_0)$ converge uniformly to a value greater than or equal to 1/2. That is,

$$\lim_{n\to\infty} \sup_\theta \Pr_\theta \left\{W_n^{(j)}(\theta) \geq \widetilde{c}_\alpha(\theta)\right\} \geq 0.5 \tag{2}$$

where $j = 1, 2$, $\widetilde{c}_\alpha(\theta)$ is the α point of the χ_1^2 if $\theta \neq \theta^*$ and $\widetilde{c}_\alpha(\theta^*)$ is a constant possibly different from $\widetilde{c}_\alpha(\theta)$, $\theta \neq \theta^*$. To do so consider the sequence $\theta_n = \theta^* + an^{-1/5}$, where a is any non-zero constant. Throughout this section $o_{p_n}(n^{-\alpha})$ and $O_{p_n}(n^{-\alpha})$ denote sequences of random variables such that when multiplied by n^α, as n goes to infinity, they converge to zero in probability and they are bounded in probability respectively under θ_n. Under some regularity conditions, Rotnitzky *et al.* (2000) showed that the maximum likelihood estimator of θ when $\theta = \theta_n$ takes the values $\widetilde{\theta}_1$ or $\widetilde{\theta}_2$ each with probability converging to 1/2 (throughout denoted as w.p.g.1/2) where $\widetilde{\theta}_j$ satisfies for $j = 1, 2$,

$$\widetilde{\theta}_j - \theta^* = (-1)^{j+1} \operatorname{sgn}(\theta_n - \theta^*) \left\{ (\theta_n - \theta^*)^2 + n^{-\frac{1}{2}} Z/I + o_{p_n}\left(n^{-\frac{1}{2}}\right) \right\}^{\frac{1}{2}}. \tag{3}$$

An easy calculation shows that (3) implies that

$$\begin{aligned} \widetilde{\theta}_1 &= \theta_n + n^{-\frac{3}{10}} a^{-1} Z/(2I) + o_{p_n}\left(n^{-\frac{3}{10}}\right) \\ \widetilde{\theta}_2 &= 2\theta^* - \theta_n - n^{-\frac{3}{10}} a^{-1} Z/(2I) + o_{p_n}\left(n^{-\frac{3}{10}}\right) \end{aligned} \tag{4}$$

In particular, this implies that

$$\left(\hat{\theta} - \theta_n\right)^2 = \begin{cases} \left(\widetilde{\theta}_1 - \theta^*\right)^2 = O_{p_n}\left(n^{-\frac{6}{10}}\right) \text{ w.p.g.1/2} \\ \left(\widetilde{\theta}_2 - \theta^*\right)^2 = 4(\theta^* - \theta_n)^2 + O_{p_n}\left(n^{-\frac{6}{10}}\right) \text{ w.p.g.1/2} \end{cases} \tag{5}$$

and $\left(\hat{\theta} - \theta_n\right)^3 = O_{p_n}\left(n^{-\frac{1}{2}}\right)$. Furthermore, it can be shown that $n^{-1} L_n^{(2)}\left(\hat{\theta}\right)$ is equal to

$$-\{I + o_{p_n}(1)\} \left\{ (\theta^* - \theta_n)^2 + 3(\theta^* - \theta_n)\left(\hat{\theta} - \theta_n\right) + \frac{3}{2}\left(\hat{\theta} - \theta_n\right)^2 \right\} + O_{p_n}\left(n^{-\frac{1}{2}}\right)$$

and $i\left(\hat{\theta}\right)$ is equal to

$$-\{I + o_{p_n}(1)\} \left\{ (\theta^* - \theta_n)^2 + 3(\theta^* - \theta_n)\left(\hat{\theta} - \theta_n\right) + \frac{3}{2}\left(\hat{\theta} - \theta_n\right)^2 \right\}$$

Thus, by (4) and (5) we obtain that w.p.g.1/2

$$\begin{aligned} n^{-1} L_n^{(2)}\left(\hat{\theta}\right) &= -13\{I + o_{p_n}(1)\}(\theta^* - \theta_n)^2 + O_{p_n}\left(n^{-\frac{1}{2}}\right) \\ &= -13 I a^2 n^{-\frac{2}{5}} + o_{p_n}\left(n^{-\frac{2}{5}}\right) \end{aligned}$$

and similarly

$$i\left(\hat{\theta}\right) = -13 I a^2 n^{-\frac{2}{5}} + o_{p_n}\left(n^{-\frac{2}{5}}\right)$$

The latter implies that w.p.g.1/2 $W_n^{(1)}(\theta_n) = 52 I a^4 n^{\frac{1}{5}} + o_{p_n}\left(n^{\frac{1}{5}}\right)$ and $W_n^{(2)}(\theta_n) = 52 I a^4 n^{\frac{1}{5}} + o_{p_n}\left(n^{\frac{1}{5}}\right)$ which converge to $+\infty$ as n goes to infinity. Thus, the tests that reject when $W_n^{(1)}(\theta)$ and $W_n^{(2)}(\theta)$ are greater than a fixed constant, regardless of the value of the constant, have levels that converge to a value no smaller than 1/2. Equation (2) for $j = 1, 2$ now follows from

$$\sup_{\theta} \Pr_{\theta}\left\{W_n^{(j)}(\theta) \geq \widetilde{c}_{\alpha}(\theta)\right\} \geq \Pr_{\theta_n}\left\{W_n^{(j)}(\theta_n) \geq \widetilde{c}_{\alpha}(\theta_n)\right\},$$

and hence

$$\sup_{\theta} \Pr_{\theta} \left\{ W_n^{(j)}(\theta) \geq \widetilde{c}_{\alpha}(\theta) \right\} \geq \lim_{n \to \infty} \Pr_{\theta_n} \left\{ W_n^{(j)}(\theta_n) \geq \widetilde{c}_{\alpha}(\theta_n) \right\} \geq 1/2.$$

Interestingly, it can be shown that $W_n^{(1)}(\theta^*)$ or $W_n^{(2)}(\theta^*)$ converge in law under θ^* to $4I^{-1}Z^2 I(Z > 0)$ and $6I^{-1}Z^2 I(Z > 0)$. That is, the asymptotic distributions of $W_n^{(1)}(\theta^*)$ and of $W_n^{(2)}(\theta^*)$ are equal to the law of the mixtures of the constant 0 and $4\chi_1^2$ and of 0 and $6\chi_1^2$ respectively with mixing probabilities equal to 1/2. A simple calculation shows that the probability that $W_n^{(1)}(\theta^*)$ is greater than 95th percentile of the χ_1^2 distribution converges under θ^* to 0.16. This limit is equal to 0.21 for $W_n^{(2)}(\theta^*)$. Thus, use of the χ_1^2 as the reference distribution for determining the rejection region of the test procedure based on $W_n^{(1)}(\theta^*)$ or $W_n^{(2)}(\theta^*)$ yields to an overly liberal test.

3 Example

For the non linear regression example presented above, we calculated the observed rejection proportions of the tests that reject for values of the Wald statistics greater than the 95th percentile of the χ_1^2 among 3,000 repetitions of samples with sizes 10,000, 1000 and 100, over a grid of parameter values about $\theta^* = 0$, with $X \sim$ Uniform(-1,2) and $X \sim$ Uniform(-1,1). The observed rejection proportion confirmed the results derived above. That is they were about 0.16 and 0.21 for $W_n^{(1)}(\theta^*)$ and $W_n^{(2)}(\theta^*)$, and were about 0.5 at values at a distance from θ^* of order $an^{-1/5}$ for both test statistics. (Due to typographical constraints, tables or figures are not included.)

References

COX, D.R. AND HINKLEY, D.V. (1974). *Theoretical Statistics*. Chapman & Hall, London.

ROTNITZKY, A., COX, D.R., BOTTAI, M., ROBINS, J. (2000). Likelihood-based inference with singular information. *Bernoulli* **6**(2), 243-284.