Robustness of generalized likelihood ratio test for linear regression

Qinghua Zhang and Michèle Basseville

IRISA (INRIA & CNRS)
Rennes, France

Introduction

- Three approaches to parametric change detection:
  - Statistical test: $\zeta(\text{data}, \theta_0)$ small enough?
  - Repeated parameter estimation : $\hat{\theta} = \theta_0$?
  - Sum of squared errors: $\sum_k (\hat{y}_k(\theta_0) - y_k)^2$ small enough?
- Their relationship is studied in the case of linear regression

Framework for comparison: linear regression

$$y_k = \theta^T x_k + e_k$$

with $y_k \in \mathbb{R}$, $x_k \in \mathbb{R}^n$, $\theta \in \mathbb{R}^n$, $e_k \sim \mathcal{N}(0, \sigma^2)$.

Without loss of generality, assume the nominal value $\theta_0 = 0$.

In data window $k = 1, \dots, N$, define

$$\mathrm{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, \quad X = \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_N^T \end{bmatrix}, \quad \mathrm{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{bmatrix}$$

then

$$\mathrm{y} = X\theta + \mathrm{e}$$

Generalized likelihood ratio (GLR) test

The log-likelihood ratio

$$l_{\theta/0}(\mathrm{y}, X) = \log \frac{p_\theta(\mathrm{y}, X)}{p_0(\mathrm{y}, X)} = -\frac{1}{2\sigma^2}(\mathrm{y} - X\theta)^T(\mathrm{y} - X\theta) + \frac{1}{2\sigma^2}\mathrm{y}^T\mathrm{y}$$

The GLR test

$$\begin{aligned} t &= 2 \max_\theta l_{\theta/0}(\mathrm{y}, X) \\ &= \frac{1}{\sigma^2}\mathrm{y}^T X (X^T X)^{-1} X^T \mathrm{y} \end{aligned}$$

t is $\chi'^2(n, \lambda)$ distributed, $\lambda = \theta^T X^T X \theta$.

When X is almost rank deficient (poor excitation), direct computation of t may lead to large numerical errors. Advanced numerical computation is then necessary.

Advanced numerical computation of the GLR test

Perform the QR decomposition of X

$$X = QR$$
$$Q^T Q = I$$

then

$$t = \frac{1}{\sigma^2} y^T X (X^T X)^{-1} X^T y$$
$$= \frac{1}{\sigma^2} y^T QR(R^T Q^T QR)^{-1} R^T Q^T y$$
$$= \frac{1}{\sigma^2} \|Q^T y\|^2$$

By avoiding explicit matrix inversion, the advanced algorithm is robust against poor excitation (bad conditioning of $X^T X$).

Deficient excitation and poor excitation

When X is strictly rank deficient, the values of θ belonging to the null space of X are not detectable, the others are still detectable by the GLR test. In this case, some columns of Q can be deleted, reducing the number of degrees of freedom of the χ^2-distribution.

When X is almost rank deficient, some values of θ may not be detectable because of the noise. The detectability (related to the non-centrality parameter $\lambda = \theta^T X^T X \theta$) depends not only on the condition number of X, but also on the norm of θ.

In any case the GLR test can be computed with the same algorithm.

Relationship with parameter estimation

MLE (and LS estimate):

$$\hat{\theta} = (X^T X)^{-1} X^T y$$

When X is almost rank deficient, $\hat{\theta}$ may have a large numerical error.

When X is strictly rank deficient, $\hat{\theta}$ can be computed with pseudo-inverse, but the estimated value is somewhat arbitrary.

In both cases, it is not reliable to detect changes in θ by simply examining the components of $\hat{\theta}$.

Relationship with parameter estimation (contd.)

Estimation error $\tilde{\theta} = \hat{\theta} - \theta \sim \mathcal{N}(0, \Sigma)$, $\Sigma = \sigma^2 (X^T X)^{-1}$

Then $r = \hat{\theta}^T \Sigma^{-1} \hat{\theta}$ is equivalent to the GLR test. It can thus be used as an alternative computation of the GLR test.

In case of poor excitation, the advanced numerical algorithm is recommended to avoid matrix inversion.

In case of deficient excitation, $\hat{\theta}$ and r can be computed with pseudo-inverse. It is then mathematically equivalent to the advanced GLR algorithm with reduced Q, but numerically more expensive.

Comparison with sum of squared errors

The sum of squared errors $(\mathrm{y} - X\theta_0)^T(\mathrm{y} - X\theta_0) = \mathrm{y}^T\mathrm{y}$ (since $\theta_0 = 0$) is sometimes used to detect changes in θ.

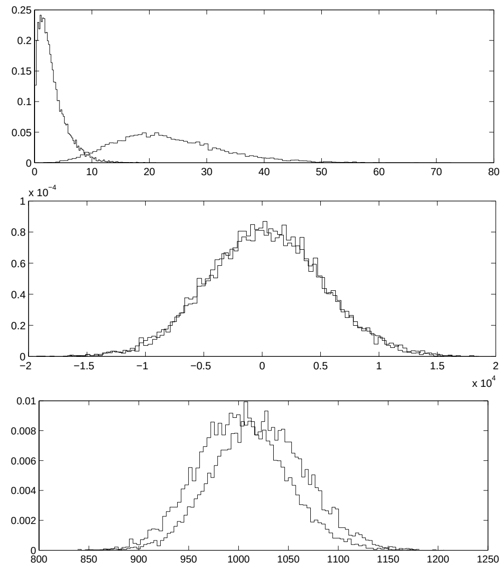
The normalized sum $s = \frac{1}{\sigma^2}\mathrm{y}^T\mathrm{y}$ also follows a χ^2-distribution, with non-centrality parameter $\theta^T X^T X\theta$ (like the GLR), but it has N degrees of freedom.

Typically $n << N$, thus the GLR test has a much lower variance!

Numerical example

Simulation with $n = 3$, $N = 1000$, $\theta_0 = 0$, $\theta = [0, 0, 0.2]^T$, condition number of $X = 1.99 \times 10^{10}$.

Histograms are made over 10000 random realizations.



$\theta_0 = 0$ and $\theta = [0, 0, 0.2]^T$

GLR test t

Third component of $\hat{\theta}$

Sum of squared errors s

Conclusion

- In case of poor excitation, change detection by screening estimated parameter components is not reliable.
- The advanced numerical algorithm for GLR test is robust to poor excitation.
- The sum of squared errors has a much larger variance than the GLR test, and is thus less reliable.