Test (2007) 16: 445–447
DOI 10.1007/s11749-007-0081-7

DISCUSSION

Comments on: Nonparametric inference with generalized likelihood ratio tests

Peter J. Bickel

Published online: 6 November 2007

Abstract This is a very interesting paper reviewing the technique for testing semi-parametric hypotheses using GLR tests. I'd like to supplement Fan and Jiang's review with some cautions and a somewhat different point of view.

1 The Wilks phenomenon

Rigorous results for smooth parametric models, see, for example, Bickel and Doksum (2005, Chap. 6) do say that, if $\hat{\theta}, \hat{\eta}$ or, equivalently, $(\hat{\theta}_{\hat{\eta}}, \hat{\eta})$ are MLE's, then $2(\ell(\hat{\theta}, \hat{\eta}) - \ell(\theta_0, \hat{\eta}_{\theta_0})) \Longrightarrow \chi_d^2$, where d is the dimension of Θ. But if $\hat{\eta}$ is not the MLE this result may fail to hold. In particular it will fail if, in the case of θ, η real, $E_{(\theta_0,\eta_0)} \frac{\partial \ell}{\partial \theta}(X_1, \theta_0, \eta_0) \cdot \frac{\partial \ell}{\partial \eta}(X_1, \theta_0, \eta_0) \neq 0$. More generally, if θ and η are infinite dimensional, the requirement is that the tangent spaces at (θ_0, η_0) of the models with $\theta = \theta_0$ kept fixed, $\underset{\eta}{\overset{\circ}{\mathcal{P}}}$, and $\eta = \eta_0$ kept fixed, $\underset{\theta}{\overset{\circ}{\mathcal{P}}}$, are orthogonal in $L_2(P_{(\theta_0,\eta_0)})$—see Bickel et al. (1993). All of Fan and Jiang's examples satisfy this condition—appropriately generalized to the general dependent case—see Bickel and Kwon (2001). Murphy's does not but the estimator $(\hat{\theta}, \hat{\eta})$ that she uses is efficient, i.e., behaves like the MLE in nice parametric situations. We give a heuristic argument below why the Wilks phenomenon can only be expected if $\hat{\eta}$ is efficient or the tangent spaces are orthogonal.

This comment refers to the invited paper available at: http://dx.doi.org/10.1007/s11749-007-0080-8.

P.J. Bickel (✉)
Department of Statistics, University of California at Berkeley, 367 Evans Hall,
Berkeley, CA 94710-3860, USA
e-mail: bickel@stat.berkeley.edu

2 Power issues

A fact that Fan and Jiang allude to is that omnibus tests like the GLR test have very little power in any particular direction. The fact that they achieve minmax power rates for various smoothness classes is little comfort. Bickel et al. (2006), in view of this failure, espouse a different point of view. They show how to construct tests which have power at the $n^{-\frac{1}{2}}$ scale against selected subclasses of alternatives which are viewed as most important, but still achieve consistency against all alternatives—i.e., do not miss extremely strong evidence against the hypothesis which is inconsistent with one's prior views of the important alternatives. Of course, the power received is necessarily very small—see Lehmann and Romano (2005, pp. 617–621) for a nice explanation, and these tests are not minimax although I believe minimax versions of these which also exhibit power in a limited set of directions can be constructed. They do not exhibit the Wilks type of phenomenon but critical values can be set using bootstraps in the way Fan and Jiang describe.

3 Wilks phenomenon heuristic calculation

Here is a heuristic calculation when θ, η are both one-dimensional.

Using

$$\frac{\partial \ell}{\partial \theta}(\hat{\theta}_{\hat{\eta}}, \hat{\eta}) = 0 = \frac{\partial \ell}{\partial \eta}(\theta_0, \hat{\eta}_0),$$

we obtain after some manipulation that, under H,

$$2\Lambda \simeq n\big[(\hat{\theta}_{\hat{\eta}} - \theta_0)^2 I_{11} - I_{22}(\hat{\eta} - \eta_0)^2\big], \tag{1}$$

where $I = \begin{pmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{pmatrix}$ is the Fisher information at (θ_0, η_0). Further,

$$n^{-\frac{1}{2}} \frac{\partial \ell}{\partial \theta}(\theta, \hat{\eta}) \simeq I_{12}(\hat{\theta}_{\hat{\eta}} - \theta_0), \tag{2}$$

$$n^{-\frac{1}{2}} \frac{\partial \ell}{\partial \eta}(\theta_0, \eta_0) \simeq I_{22}(\hat{\eta}_0 - \eta_0). \tag{3}$$

Finally,

$$n^{\frac{1}{2}}(\hat{\eta} - \eta_0) \cong n^{\frac{1}{2}}(\hat{\eta}_0 - \eta_0) + \Delta,$$

where Δ is asymptotically $N(0, \tau^2)$ with $\tau^2 = 0$ iff $I_{12} = 0$. After some algebra and formally taking expectations we arrive at

$$E(2\Lambda) \simeq 1 - \frac{I_{12}^2}{I_{11} I_{22}} - \tau^2 \left(\frac{I_{12}^2}{I_{11}} - I_{22} \right). \tag{4}$$

We conclude from (4) that $E(2\Lambda) \simeq 1$ iff either $I_{12} = 0$ or

$$\tau^2 = \frac{I_{12}^2}{I_{11} I_{22}} \left(\frac{-I_{12}^2 + I_{11} I_{22}}{I_{11}} \right)^{-1} = I^{22} - I_{22}^{-1}, \tag{5}$$

where $\|I^{ij}\| \equiv I^{-1}$. But (5) holds iff $\hat{\eta}$ is efficient, i.e., $(\hat{\theta}_{\hat{\eta}}, \hat{\eta})$ are equivalent to the MLE in regular situations. Note that in general the limit of $E(2\Lambda)$ here is a sum of weighted independent χ_1^2 variables. If Θ is infinite dimensional, failure of the Wilks' phenomenon means that $E(2\Lambda)/(2*)\mathrm{Var}(2\Lambda)$ does not converge to 1.

References

Bickel PJ, Doksum KA (2005) Mathematical statistics: basic ideas and selected topics, vol 1. Prentice Hall, Englewood Cliffs

Bickel PJ, Kwon J (2001) Inference for semiparametric models: some questions and an answer (with discussion). Stat Sin 11:863–960

Bickel PJ, Klaassen C, Ritov Y, Wellner J (1993) Efficient and adaptive estimation in semiparametric models. John Hopkins Press, Baltimore

Bickel PJ, Ritov Y, Stoker T (2006) Tailor-made tests for goodness of fit for semiparametric hypotheses. Ann Stat 34:721–741

Lehmann EL, Romano J (2005) Testing Statistical Hypotheses, 3rd edn. Springer, New York