Qüestiió - V. 7, n.º 4 (desembre 1983) pp. 527-529

ASYMPTOTIC THEORY: SOME RECENT DEVELOPMENTS

DAVID R. COX
IMPERIAL COLLEGE OF SCIENCE AND TECHNOLOGY
LONDON

A review is given of recent work on asymptotic theory leading to a recommendation to use likelihood ratio rests with, where available, a Bartlett adjustment factor.

Keywords: ANCILLARY STATISTIC; LIKELIHOOD RATIO TEST; SCORE TEST; MAXIMUM LIKELIHOOD; WALD TEST; BARTLETT FACTOR; INFORMATION.

1. INTRODUCTION.

The object of this paper is to review, so far as possible without going into technical details, some recent ideas and results in the asymptotic theory of inference in parametric statistical models. That is, we suppose that the n×1 observational vector y is the observed value of a random variable Y having a probability density $f_Y(y;\theta)$ specified except for the unknown parameter θ of dimension d_θ ; if only certain components ψ of θ of dimension d_ψ are of interest, we write $\theta = (\psi,\lambda)$, where λ is a nuisance parameter.

The information about θ or ψ that can be obtained from y is usually best expressed via a nested collection of confidence regions for the parameter. For convenience of exposition, however, we discuss the complementary and mathematically equivalent problem of testing the null hypothesis $\theta = \theta_o$ or $\psi = \psi_o$, where θ_o, ψ_o are fixed and arbitrary.

While, especially when $d_\psi = 1$, this problem has a unique definitive solution for many interesting and practically important cases, nevertheless these are in another sense very special and for most of the more complex analyses arising in applied statistics reliance is made on approximate theory, such as the asymptotic normality of maximum likelihood estimates. In a few cases this reliance is a matter of computational convenience - rather than necessity, the 'exact' solution being available in principle but being hard to evaluate.

These approximations are derived via the limit theorems of probability theory by supposing that $n \to \infty$. It must be stressed, however, that this limiting notion has no physical significance, being a technical device to generate approximations, whose adequacy has been in principle to be judged.

2. COMPETITIVE LIKELIHOOD-BASED PROCEDURES.

Many of the key ideas are best ilustrated by the simplest special case of a scalar parameter $d_\theta = 1$. Write

$$\ell(\theta;\, Y) = \log f_Y\,(Y;\, \theta) \tag{1}$$

for the log likelihood, $\theta = \hat\theta(y)$ for the maximum likelihood estimate, assumed unique, and

$$j(\theta;y) = -\frac{\partial^2 \ell(\theta;y)}{\partial\theta^2}\ , \quad j(\theta) = E\{j(\theta;Y)\}\ , \tag{2}$$

the observed and expected information; the expectation in $j(\theta)$ is taken with respect to $f_Y(y;\theta)$.

To test $\theta = \theta_o$, we may consider

(i) Wald type statistics of the form

$$(\hat\theta - \theta_o)v^{-\frac{1}{2}} \quad \text{or } W_e = (\hat\theta - \theta_o)^2/v\ , \tag{3}$$

- David R. Cox - Imperial College of Science and Technology - Department of Mathematics - Huxley Building Queen's Gate. London SW7 2BZ - London, England.

- Article rebut el Novembre del 1983.

- Aquest treball es una ponencia invitada al "First Catalan International Symposium on Statistics" a Barcelona del 23 al 24 de Setembre de 1983.

where

$$v^{-1} = j(\theta_o;y) \text{ or } j(\hat{\theta};y) \text{ or } j(\theta_o) \text{ or } j(\hat{\theta}). \qquad (4)$$

(ii) Score or Rao type statistics in which

$$u(y;\theta_o) = \left[\frac{\partial \ell(\theta;y)}{\partial\theta}\right]_{\theta=\theta_o} \qquad (5)$$

and the test statistic is

$$u(y;\theta_o)v^{\frac{1}{2}} \text{ or } W_u = u^2(y;\theta_o) v, \qquad (6)$$

where usually v^{-1} is taken to be $j(\theta_o)$, the exact variance of $u(Y;\theta_o)$ at $\theta = \theta_o$.

(iii) The (maximum) likelihood ratio statistic

$$\text{sgn } (\hat{\theta}- \theta_o)W^{\frac{1}{2}} \text{ or } W = 2\{\ell (\hat{\theta};y) - \ell(\theta_o;y)\}. \qquad (7)$$

The statistics in (3), (6) and (7) are asymptotically equivalent and the W forms have asymptotically a χ^2_1 distribution when $\theta = \theta_o$. All generalize directly to d dimensional θ and to problems with nuisance parameters. While often the various tests give virtually identical results, this is not always so, and therefore the choice between the forms has to be considered.

3. BASIS FOR COMPARISON: QUALITATIVE CONSIDERATIONS.

The following points are essentially qualitative;

(i) while use of $\hat{\theta} - \theta_o$ and W_e leads to a simple expression of conclusions and to elliptical confidence regions for θ, the answers are not invariant under reparameterization. Indeed W_e may lead to logically nonsensical answers;

(ii) only W leads to even qualitatively reasonable answers if the likelihood function has multiple maxima, or a supremum at ∞;

(iii) the test statistic W and those other statistics using observed rather than expected information have forms independent of stopping rule, although their distribution will typically depend somewhat on that rule;

(iv) if there are ancillary, i.e. conditioning statistics, for the problem, the comments (iii) apply;

(v) when a complex model is augmented in various ways to test its adequacy, the score is computationally the most convenient.

4. SIMPLICITY OF DISTRIBUTIONAL FORM.

It is a convenience if the distributional properties of the test statistic have, to a close approximation, a simple form. Bartlett M.S./1/ noted in two particular cases that if

$$E(W) = d_\psi \ (1 + b/n) + o(1/n)$$

then the first three cumulants of

$$W' = W(1 + b/n)^{-1} \qquad (8)$$

agree with those of $\chi^2_{d_\psi}$ with error $o(1/n)$. This suggests that the modified statistic W' has to the indicated order the chi-squared distribution.

We call $1 + b/n$ a Bartlett adjustment. Its study has a long history. Recently Barndorff-Nielsen and Cox /3/ , drawing on a distributional result of Barndorff-Nielsen /2/ , have given a relatively simple and general proof that the distributional result holds conditionally on ancillary statistics and also unconditionally, and have also given an alternative indirect method of evaluating b. A slightly more complicated version (McCullagh, /6/ applies to the signed statistic.

5. FURTHER CONSIDERATIONS: ANCILLARITY.

It is widely although not universally agreed that if a theory of statistical inference for the interpretation of unique sets of data is to be based on the frequency theory of probability via notions of confidence coefficients and the like, then the arguments used should be made adequately conditional on the observed values of so-called ancillary statistics. The precise formulation of this notion raises some major questions, however.

A consequence es that where approximate (asymptotic) arguments are used, approximately ancillary statistics should be used for conditioning. A formal notion of second-order local - ancillarity was introduced by Cox /4/ when d = 1. McCullagh /6/ has clarified and extended these results to confirm that when there are no nuisance parameters, W, or its signed version, is in effect the unique statistic for inference and is distributed independently of the approximately ancillary statistic; see also Efron and Hinkley /5/ and Sprott and Viveros /7/.

6. CONSORING AND STOPPING RULES.

Bayesian methods using proper rather than improper prior distributions and pure likelihood methods do not depend on uninformative stopping rules or on the uninformative censoring encountered in the analysis of failure data. For a broad but not universal class of such rules, the asymptotic χ^2 distribution will apply to W and the dependence of the Bartlett adjustment b on the stopping or censoring rule has been analyzed in some special cases (Barndorff-Nielsen and Cox, 1984b).

7. NUISANCE PARAMETERS.

To an appreciable extent the above arguments apply when a nuisance parameter λ is present, provided that d_λ, the dimensionality of λ, is small. (It is well known that there can be major problems with methods related to maximum likelihood when d_λ is large). The uniqueness results of Section 5 do not in general apply and there is need for further development of the above ideas to cover that, the most important, case. It is likely that this will be via notions of conditional, marginal and partial likelihood.

8. REFERENCES.

/1/ BARTLETT, M.S., (1937). Proc. Roy. Soc. A 160, 268-282.

/2/ BARNDORFF-NIELSEN, O. (1983). Biometrika 70, 343-366.

/3/ BARNDORFF-NIELSEN, O. & COX, D.R.,(1984a). J.R. Statist. Soc. B, to appear. (1984b). Submitted.

/4/ COX, D.R. (1981). Biometrika 67, 279-292.

/5/ EFRON, B. & HINKLEY, D. V. (1978). Biometrika 65, 457-487.

/6/ McCULLAGH, P. (1984). Biometrika, to appear.

/7/ SPROTT, D. A. and VIVEROS, R. (1984). Submitted.