A NOTE ON GOODNESS OF FIT AND ANCILLARITY

by

DAVID V. HINKLEY*
University of Minnesota

TR #320

*This research was supported by NSF Grant MCS 77-00959.

A NOTE ON GOODNESS OF FIT AND ANCILLARITY

by

DAVID V. HINKLEY

University of Minnesota
TR #320

Summary

The dual role of ancillary statistics in conditional inference and in goodness of fit tests is noted, with special reference to curved exponential families.

Key words: Ancillary statistic; likelihood; goodness of fit; exponential family

In the conditional approach to parametric statistical inference, when a sufficient statistic S includes an ancillary component A, inference is carried out via the conditional probability distribution of S given the value of A. Ancillary statistics, because they are distributed independently of the parameter, play a dual role as components of goodness of fit statistics. This was pointed out by R.A. Fisher (1928) in connexion with likelihood estimation for a multinomial linkage model.

Suppose that $x_1, \ldots, x_n$ are independently distributed according to a curved exponential family of densities

$$f_\theta(x) = g_{\lambda_\theta}(x) = \exp(\lambda_\theta^T x - \psi_\theta), \qquad (x \in R^k,\ \theta \in R^1) \qquad (1)$$

which is a one-dimensional subset of the unrestricted family of densities

$$g_\lambda(x) = \exp(\lambda^T x - \eta_\lambda) \qquad (x \in R^k,\ \lambda \in R^k). \qquad (2)$$

For both f and g the average $\bar{x} = n^{-1}\Sigma x_j$ is minimal sufficient. Let $\beta = E(x)$ and $\Sigma = \mathrm{var}(x)$, with added subscript θ when restricted to f_θ. Then maximum likelihood estimates are obtained by solving

$$\hat{\beta} = \bar{x}$$

$$\left[\frac{\partial\lambda_\theta}{\partial\theta}\right]^T_{\theta=\hat{\theta}} (\bar{x} - \beta_{\hat{\theta}}) = 0 \qquad (3)$$

for (2) and (1) respectively. Now write $\ell_\lambda = \log \prod_{j=1}^{n} g_\lambda(x_j)$, with $\ell_{\lambda_\theta} = \ell^*_\theta$ denoting the loglikelihood of θ for model (1). Then the likelihood ratio test statistic for testing the fit of model (1) within model (2) is

$$W = 2(\ell_{\hat{\lambda}} - \ell^*_{\hat{\theta}})$$

which may be shown to have asymptotic expansion

$$W = n(\bar{x} - \beta_{\hat{\theta}})^T \Sigma_{\hat{\theta}}^{-1}(\bar{x} - \beta_{\hat{\theta}}) + 0(n\|\bar{x} - \beta_{\hat{\theta}}\|^3). \tag{4}$$

The second term in (4) is asymptotically $0_p(n^{-\frac{1}{2}})$, while the first term, subject to (3), is asymptotically χ^2_{k-1} under model (1). See Aitchison and Silvey (1960). In the multinomial case, the quadratic form in (4) is the Pearson chi-square statistic.

In the theory of approximate conditional inference for θ, given model (1), the vector $\bar{x} - \beta_{\hat{\theta}}$ whose Mahalanobis length appears in (4) may be used to determine k-1 approximate ancillaries. Some details are given by Efron and Hinkley (1978). The dominant ancillary is determined by the observed information

$$I = \left[-\frac{d^2}{d\theta^2} \ell^*_\theta \right]_{\theta=\hat{\theta}},$$

and is

$$Q_1 = (1 - \frac{I}{\mathcal{I}_{\hat{\theta}}})/\gamma_{\hat{\theta}} \tag{5}$$

where $\mathcal{I}_\theta = E(-\frac{d^2}{d\theta^2} \ell^*_\theta)$ and $\gamma_\theta = \text{var}\,(\frac{d^2\ell^*_\theta}{d\theta^2} \,|\, \frac{d\ell^*_\theta}{d\theta})/\mathcal{I}^2_\theta$. The relevant conditional normal approximation for $\hat{\theta}$ has variance I^{-1}, as opposed to the unconditional variance $\mathcal{I}^{-1}_{\hat{\theta}}$. The statistic Q_1 measures the discrepancy between conditional and unconditional normal approximations, but only relative to the statistical curvature $\gamma_{\hat{\theta}}$.

For general k, further approximate ancillaries $Q_2, \ldots, Q_{k-1}$ may be constructed from $\bar{x} - \beta_{\hat{\theta}}$ in such a way as to be asymptotically independent N(0,1) variates. This corresponds to a partition of the dominant part of the goodness of fit statistic W, by (4); that is,

$$W \sim n(\bar{x} - \beta_{\hat{\theta}})^T \Sigma_{\hat{\theta}}^{-1}(\bar{x} - \beta_{\hat{\theta}}) = \sum_{j=1}^{k-1} Q_j^2.$$

The larger is W, the more discrepancy is likely between conditional and unconditional inference based on $\hat{\theta}$. (The two agree exactly only if $\bar{x} = \beta_{\hat{\theta}}$, when W = 0.) The magnitude of the discrepancy depends on invariants such as γ, as is seen in (5). It may sometimes be informative to effect this decomposition of W, for example splitting off Q_1^2 and testing the remainder. The next component Q_2 presumably affects bias of $\hat{\theta}$, as well as skewness of distribution, being a standardized form of the third derivative of $\ell_{\hat{\theta}}^*$; further theoretical analysis is required to determine the exact form and effects of Q_2, Q_3, ...

As a simple example, consider Fisher's (1928) multinomial model with reduced cell probabilities

$$\beta_\theta = \tfrac{1}{4}(2 + \theta,\ 2 - 2\theta,\ \theta).$$

Here k = 2, $\hat{\theta} = .0357$, $\gamma_{\hat{\theta}}^2 = 5.643n^{-1}$, n = 3839 and $q_1^2 = 2.013$. The data do not deviate significantly from the model, but with the high curvature there might be a sizeable difference between I and $\mathcal{J}_{\hat{\theta}}$. In fact $I/\mathcal{J}_{\hat{\theta}} = 0.938$, so that only a 3% error would be made in computing standard error for $\hat{\theta}$ as $1/\sqrt{\mathcal{J}_{\hat{\theta}}}$. Had q_1^2 been the same at n = 250, the error would have been 11%. Thus modest lack of fit can indicate appreciable difference between conditional and unconditional analysis.

Note that the goodness of fit of a specific value θ is measured by

$$2(\ell_{\hat{\lambda}} - \ell^*_{\hat{\theta}}) = W + 2(\ell^*_{\hat{\hat{\theta}}} - \ell^*_{\hat{\theta}}), \tag{6}$$

with W as in (4), which is a decomposition into asymptotically independent χ^2_{k-1} and χ^2_1 variates. This suggests the approximate conditional validity of the likelihood ratio method of setting confidence limits for θ; that is

$$\mathrm{pr}\{2(\ell^*_{\hat{\theta}} - \ell^*_{\theta}) \geq c \,|\, Q_1, \ldots, Q_{k-1}\} \sim \mathrm{pr}(\chi^2_1 \geq c).$$

Easterling (1976) has suggested setting confidence limits for θ from (6), in which case a larger W would result in narrower confidence limits. This is unwise from the conditional viewpoint, because the value of Q_1 leading to large W might be positive, corresponding to $I < \mathcal{J}_{\hat{\theta}}$, in which case wider confidence limits would be appropriate.

This research was supported by NSF Grant MCS 77-00959

References

AITCHISON, J. and SILVEY, S.D. (1960). Maximum-likelihood estimation procedures and associated tests of significance. J.R. Statist. Soc., B, 22, 154-71.

EFRON, B. and HINKLEY, D.V. (1978). Assessing the accuracy of the maximum likelihood estimator: observed versus expected Fisher information. Biometrika, 65 (to appear).

EASTERLING, R.G. (1976). Goodness of fit and parameter estimation. Technometrics, 18, 1-9.

FISHER, R.A. (1928). On a property connecting the chi-square measure of discrepancy with the method of maximum likelihood. Atti del Congresso Internazionale dei Matematici, Bologna, 6, 95-100.