from the ancillary statistic A_a (as well as the model) while $\hat{\theta}$, a complimentary portion of the data, is being used to assess the accuracy of these coverage functions. In parametric inference the roles of these statistics are reversed in that the ancillary statistic A_a assesses the accuracy of $\hat{\theta}$ in determining the true model. Practical examples are needed to bear out the sensibility of basing recipe choice on coverages at and near $\hat{\theta}$.

Many practical models such as generalized linear models do not admit exact ancillary A_a upon which to condition. In such instances we must find approximate ancillaries as has been done in Hinkley (1980) and Barndorff-Nielsen (1980, 1983).

I do not agree with Bjørnstad's suggestion that $\text{pr}\{\mathbf{Z} \in I_{.9}(\mathbf{Y}); \theta\}$ as an unconditional probability can be used to meaningfully assess the various recipes. Also measuring the worth of an interval (or its associated recipe) by its guarantee of 90% coverage, $\inf_\theta \text{pr}\{C_\theta(\mathbf{Y}) \geq .9\}$ where $C_\theta(y) = \text{pr}\{\mathbf{Z} \in I_{.9}(y) \mid y; \theta\}$, amounts to a worst case scenario assessment. This could be a very unrepresentative assessment measure to use as a basis for recipe choice.

ADDITIONAL REFERENCES

BARNARD, G. A. (1985). A coherent view of statistical inference. Technical Report, Dept. Statistics, Univ. Waterloo, Waterloo, Ontario, Canada.

BOX, G. E. P. (1980). Sampling and Bayes' inference in scientific modelling and robustness (with discussion). *J. Roy. Statist. Soc. Ser. A* **143** 383–430.

BURRIDGE, J. (1986). Discussion of "Predictive likelihood inference with applications" by R. W. Butler. *J. Roy. Statist. Soc. Ser. B* **48** 29–30.

COX, D. R. (1986). Discussion of "Predictive likelihood inference with applications" by R. W. Butler. *J. Roy. Statist. Soc. Ser. B* **48** 27.

COX, D. R. and HINKLEY, D. V. (1974). *Theoretical Statistics.* Chapman and Hall, New York.

EFRON, B. and HINKLEY, D. V. (1978). Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information (with discussion). *Biometrika* **65** 457–487.

FISHER, R. A. (1934). Two properties of mathematical likelihood. *Proc. Roy. Soc. Ser. A* **144** 285–307.

FISHER, R. A. (1935). The fiducial argument in statistical inferences. *Ann. Eugen.* **6** 391–398.

FISHER, R. A. (1973). *Statistical Methods and Scientific Inference,* 3rd ed. Hafner, New York.

HINKLEY, D. V. (1980). Likelihood as approximate pivotal distribution. *Biometrika* **67** 287–292.

Comment

Tom Leonard, Kam-Wah Tsui and John S. J. Hsu

Professor Bjørnstad is to be congratulated on an excellent review of an important area. Previous statistical practice largely referred to point predictions and estimated standard errors when predicting future observations from current data. When analyzing time series, contingency tables or nonlinear regression models, it is often thought necessary to refer to asymptotics, even to obtain an approximate standard error. However, methods are now available permitting precise predictions based upon finite samples. Moreover, the applied statistician can refer to an entire predictive likelihood or density or probability mass function, summarizing the information in the data about any future observation. This broadens the type of nonlinear model, with several parameters, which may yield useful predictions. These predictions can now be expressed in terms of probability statements, thus enhancing their interpretability, e.g., for noisy data sets.

Tom Leonard is Associate Professor, Kam-Wah Tsui is Professor and John S. J. Hsu is a graduate student, Department of Statistics, University of Wisconsin-Madison, Madison, Wisconsin 53706.

Let $p(\mathbf{y} \mid \boldsymbol{\theta})$ denote our density (or probability mass function) for an $n \times 1$ vector $\mathbf{y}$ of current observations, given a $p \times 1$ vector $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_p)^T$ of unknown parameters, and $p(\mathbf{z} \mid \boldsymbol{\theta})$ represent the corresponding density for an independent $m \times 1$ vector $\mathbf{z}$ of future observations. If $\pi(\boldsymbol{\theta})$ is the prior density of $\boldsymbol{\theta}$, for $\boldsymbol{\theta}$ lying in the parameter space Θ, then the predictive distribution

$$(1) \qquad p(\mathbf{z} \mid \mathbf{y}) = \int_\Theta p(\mathbf{z} \mid \boldsymbol{\theta}) \pi(\boldsymbol{\theta} \mid \mathbf{y}) \, d\boldsymbol{\theta}$$

of $\mathbf{z}$ given $\mathbf{y}$ is also representable in the form

$$(2) \qquad p(\mathbf{z} \mid \mathbf{y}) = \frac{p(\mathbf{z} \mid \boldsymbol{\theta}) \pi(\boldsymbol{\theta} \mid \mathbf{y})}{\pi(\boldsymbol{\theta} \mid \mathbf{y}, \mathbf{z})}, \quad \boldsymbol{\theta} \in \Theta.$$

Here we have

$$(3) \qquad \pi(\boldsymbol{\theta} \mid \mathbf{y}) \propto \pi(\boldsymbol{\theta}) p(\mathbf{y} \mid \boldsymbol{\theta}), \quad \boldsymbol{\theta} \in \Theta,$$

denoting the *posterior density* of $\boldsymbol{\theta}$, given $\mathbf{y}$, and

$$(4) \qquad \pi(\boldsymbol{\theta} \mid \mathbf{y}, \mathbf{z}) \propto p(\mathbf{z} \mid \boldsymbol{\theta}) \pi(\boldsymbol{\theta} \mid \mathbf{y}), \quad \boldsymbol{\theta} \in \Theta,$$

denoting the *postposterior density* of $\boldsymbol{\theta}$, given $\mathbf{y}$ and $\mathbf{z}$.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *Statistical Science.*
www.jstor.org