Comment

Ronald W. Butler

I welcome this paper by Professor Bjørnstad because it calls attention to the important practical problem of prediction which, of course, has been of scientific interest long before the subject of statistical science itself. This paper is also timely because I believe there will be much future interest in this area driven by the current concern for quality control. Indeed, most of the stated goals and objectives in the quality control area are predictive aims to which the predictive methodology herein might be more appropriately and profitably applied.

The first portion of Bjørnstad's article summarizes efforts to produce a likelihood-based approach to predictive inference. Lauritzen (1974), Hinkley (1979) and Butler (1986) all conditioned on sufficient statistics in order to judge the compatibility of future observable values with the data. Such conditional inference methodology is consistent with Fisher's (1973) use of conditioning for parametric inference in two-by-two tables. Section 1 below elaborates on this comparison and also motivates conditional predictive likelihood (denoted L_c by Bjørnstad) for discrete data. In addition, various profile-based predictive likelihoods can in turn be motivated as saddlepoint approximations to these conditional predictive likelihoods (see Butler, 1989).

Section 4 of Bjørnstad's article introduces new material on predictive likelihood assessment. These assessment procedures are based on the accuracy of certain unconditional coverage probabilities which I do not believe are either relevant or useful for assessing and choosing among the various predictive likelihood recipes. Section 2 below discusses an assessment procedure based on the accuracy of conditional coverages given the appropriate ancillary statistics. Since the original motivation for predictive likelihood is founded on ideas of conditional inference, it seems fitting and indeed more meaningful (to me at least) that assessment should be conditional as in Barnard (1986), Butler (1989) and in Section 2 below.

1. CONDITIONAL PREDICTIVE LIKELIHOOD

The conditional predictive likelihood recipe

$$L_c(z \mid y) = \frac{f(y, z; \theta)}{f(r(y, z); \theta)} \tag{1}$$

Ronald W. Butler is Associate Professor, Department of Statistics, Colorado State University, Fort Collins, Colorado 80523.

first appeared in Hinkley (1979) in a rather disguised form that precluded its general usage. Subsequently it was written in the form (1) in Butler (1986, page 4) and suggested therein for use with discrete data only. Motivation for (1) was also provided in the same article (page 3) which I now expand upon.

The Bayesian analyst uses the marginal distribution of the data for model criticism (Box, 1980), i.e.,

$$f(y) = \int f(y \mid \theta) f(\theta)\, d\theta, \tag{2}$$

where $f(\theta)$ denotes a prior distribution. In prioritizing a generic value z of future observable Z, the Bayesian analyst also includes the value z with the data y and criticizes the model with

$$f(y, z) = \int f(y, z \mid \theta) f(\theta)\, d\theta, \tag{3}$$

which is proportional in z to $f(z \mid y)$, the Bayesian predictive density. Criticism in (3) is not concerned with whether the theta associated with the distribution of Y is the same as that of the distribution of Z given $Y = y$; these have been assumed to be the same. What is being criticized is the level of agreement between the generically assumed value z and the observed data y.

From a likelihood perspective model criticism generally proceeds by conditioning the data on a minimal sufficient statistic $r(y)$ for the parameter (see Cox and Hinkley, 1974, pages 37–38), i.e.,

$$f(y \mid r(y)) = \frac{f(y; \theta)}{f(r(y); \theta)}. \tag{4}$$

In prioritizing a generic value z of a future observable we incorporate it into the data as does the Bayesian analyst and use

$$f(y, z \mid r(y, z)) = L_c(z \mid y), \tag{5}$$

or conditional predictive likelihood, to assess the compatibility of the value z with data y.

We can illustrate these principles using a simple example in which $y = (x_1, \cdots, x_n)$ is assumed to be an iid sample of Bernoulli (θ) trials. Conditioning y on $r(y) = \sum x_i = r$ leads to a uniform distribution over all $\binom{n}{r}$ subsets or configurations of the r successes (ones) and $n - r$ failures (zeros). Suppose we are concerned that $\theta = \text{pr}\{\text{success}\}$ might be increasing with trial number and wish to measure such concern with data $y = (0, 1, 0, 0, 1, 1)$. Then among $\binom{6}{3} = 20$ configurations, we count those which are at least as