Comment

George A. Barnard

I broadly agree with much of what Good says, although I am sometimes unhappy with his choice of terms. For instance, his reference to maximum likelihood seems to suggest that the method amounts to treating the most likely value of the parameter as if it were certainly the true value. Unfortunately there are some who make this mistake.

It is perhaps unfair, when an author has made quite clear what his topic is to be, to complain that it should have been something else. But I must express regret that Good deferred to a final short paragraph his remarks on the role of statisticians as summarizers of data. Because the objective, efficient summarization of data seems to me to be the most important function of the statistician as such. The function to which Good devotes a great part of his paper, the "cheminement de la pensee," in Emile Meyerson's phrase, or "good thinking," as Good so aptly calls it, is one in which the statistician functions along with many other specialists, and where his role is not the primary one. Besides, as I hope to indicate, the theme of the Bayes/non-Bayes compromise finds an excellent expression in connection with data summarization.

It is of the essence of a summarization of data that it should embody what the data have to say on the topic of interest, omitting only material that is irrelevant, and adding only material accepted as true by all prospective readers of the summary. In the vast majority of cases where continuous measurements are involved, the topic of interest and what is accepted as true by all prospective readers can be expressed in terms of parameters and a specific function $\mathbf{p}(x, \theta)$ of the observations and the parameters, taken to have a known distribution. The summarization procedure consists in transforming $\mathbf{p}$, by a 1–1 transformation, to $(\mathbf{q}, \mathbf{a})$, where $\mathbf{q}$ depends as much as possible on the parameters alone, and $\mathbf{a}$ depends as much as possible on the observations alone. In so far as $\mathbf{a}$ does not involve the parameters its value becomes known when the observations are known. If we then imagine learning the data by first being told the value of $\mathbf{a}$, then, after an interval, being told the values of any functions of the data entering into $\mathbf{q}$, our position just before the interval is similar to what it was at the beginning, with a statistical model now specified by the function $\mathbf{q}$, with its probability function $f(\mathbf{q})$ derived by conditioning on the known value of $\mathbf{a}$. Then learning the values of the data functions involved in $\mathbf{q}$ is like learning the values of the original observations. The data can be efficiently summarized by specifying $\mathbf{q}$, its distribution $f(\mathbf{q})$ and the values of the observational functions entering into $\mathbf{q}$. The value of $\mathbf{a}$ merely tells us the value of a quantity of known distribution, not dependent on any of the parameters; it is irrelevant information.

In some cases a complete separation can be made, so that $\mathbf{q}$ is a function of parameters only, while $\mathbf{a}$ is a function of data only. This is the "full Bayesian case," where the inference consists simply of the posterior distribution of $\mathbf{q}$. The information taken as known in the statistical model has been combined with the data in a fully efficient manner and expressed in the posterior distribution. Cases are, however, rare where a statistical model allowing such treatment can be taken as accepted by all potential readers. More commonly a partial separation is all that can be achieved.

A typical intermediate case is one in which unknown scale and location parameters are involved, with a large number of observations x_i. The function $\mathbf{p}$ then has components $p_i = (x_i - \lambda)/\sigma$, where (λ, σ) are the unknown parameters. If $(\bar{x}, s)$ denotes any convenient location-scale pair of functions of the sample (such, for example, as the sample mean and the sample estimated standard error of the mean), $\mathbf{q}$ may be taken to be $(t, z) = ((\bar{x} - \lambda)/s, (\ln s - \ln \sigma))$, while $\mathbf{a}$ has components $(x_i - \bar{x})/s$. If the conditional density of (t, z) is $f(t, z)$, the data are then objectively and efficiently summarized by specifying (t, z), $f(t, z)$ and the observed values of $(\bar{x}, s)$.

If (t_0, z_0) denotes the result of substituting the observed values of $(\bar{x}, s)$ in (t, z), then $f(t_0, z_0)$ provides the likelihood function of (λ, σ) on the basis of the data. Any reader who can supplement the statistical model with his personal prior for (λ, σ) may combine this with $f(t_0, z_0)$ to derive his posterior for the parameters. If we think of the readers of the summary, some will have their own priors, whereas others may not be prepared to so commit themselves. The specification of (t, z), $f(t, z)$ and the observed values of (x, s) then can be thought of as a Bayes/non-Bayes compromise.

Other compromises will be needed in practice. It will be unwise to suppose the distribution of $\mathbf{p}$ to be exactly known, so a "model adjustment parameter"

George A. Barnard is Professor Emeritus at the University of Essex. His mailing address is: Mill House, 54 Hurst Green, Brightlingsea, Colchester, Essex C07 0EH, England.