Comment: Quantifying the Fraction of Missing Information for Hypothesis Testing in Statistical and Genetic Studies

I-Shou Chang, Chung-Hsing Chen, Li-Chu Chien and Chao A. Hsiung

Nicolae, Meng and Kong are to be congratulated on having treated an important practical problem in many scientific inquiries in which the investigator has chosen the testing procedure, but needs to know the impact of the missing data on the test in terms of the relative loss of information. To measure the relative information, they propose to compare how the observed-data likelihood deviates from flatness relative to the same deviation in the complete-data likelihood. Several measures of this deviation expressed by Bayesian method are explored and applied to the study of genetics and genomics. As noted in their paper, these measures are especially needed in small-sample problems with incomplete data.

We would like to explore the use of this type of measure in two examples to indicate its wide applicability and some computational issues. One concerns infectious disease data, which are usually highly dependent and incomplete; the investigators often need to decide if more data are needed, and in case they are, to know the type of data that is most desirable. The other concerns a test on the shape of a regression function; we will apply the Bayesian measure of relative information to select design points for collecting more data.

Because Bayesian tests are more tractable and natural than a frequentist approach in these two examples,

I-Shou Chang is Investigator, Institute of Cancer Research and Division of Biostatistics and Bioinformatics, National Health Research Institutes, 35 Keyan Road, Zhunan Town, Miaoli County 350, Taiwan (e-mail: ischang@nhri.org.tw). Chung-Hsing Chen is Postdoctor, Division of Biostatistics and Bioinformatics, National Health Research Institutes, 35 Keyan Road, Zhunan Town, Miaoli Country 350, Taiwan (e-mail: chchen@nhri.org.tw). Li-Chu Chien is Postdoctor, Division of Biostatistics and Bioinformatics, National Health Research Institutes, 35 Keyan Road, Zhunan Town, Miaoli County 350, Taiwan (e-mail: lcchien@nhri.org.tw). Chao A. Hsiung is Director, Division of Biostatistics and Bioinformatics, National Health Research Institutes, 35 Keyan Road, Zhunan Town, Miaoli County 350, Taiwan (e-mail: hsiung@nhri.org.tw).

we consider the following extensions of their (25) for the measure of relative information:

(BI3)
$$E_0\{\mathrm{Var}[\mathrm{lod}(\theta_0, \theta|Y_{\mathrm{ob}})|Y_{\mathrm{ob}}]\}$$
$$\cdot \left(E_0\left\{\mathrm{Var}[\mathrm{lod}(\theta_0, \theta|Y_{\mathrm{ob}})|Y_{\mathrm{ob}}] + \mathrm{Var}\left[\log\frac{P(Y_{\mathrm{co}}|Y_{\mathrm{ob}}, \theta)}{P(Y_{\mathrm{co}}|Y_{\mathrm{ob}}, \theta_0)}\Big|Y_{\mathrm{ob}}\right]\right\}\right)^{-1}$$

(BI4)
$$E_0\left\{\mathrm{Var}[\mathrm{lod}(\theta_0, \theta|Y_{\mathrm{ob}})|Y_{\mathrm{ob}}] \right.$$
$$\cdot \left(\mathrm{Var}[\mathrm{lod}(\theta_0, \theta|Y_{\mathrm{ob}})|Y_{\mathrm{ob}}] + \mathrm{Var}\left[\log\frac{P(Y_{\mathrm{co}}|Y_{\mathrm{ob}}, \theta)}{P(Y_{\mathrm{co}}|Y_{\mathrm{ob}}, \theta_0)}\Big|Y_{\mathrm{ob}}\right]\right)^{-1}\left.\right\}.$$

Here E_0 means average over θ_0 from the conditional posterior distribution on the null hypothesis. To shorten the presentation, we use only (BI3) in the following discussion.

1. INFECTIOUS DISEASE DATA

As discussed in Rhodes, Halloran and Longini (1996), there are several levels of information in the study of infectious disease data and it is of interest to decide the level of information in the study. We consider two levels of information in a simple model to illustrate the way that (BI3) may be used in this situation. Suppose there is a collection of disjoint households that suffer a transmissible disease and an individual can only be infected by members in the same household. We assume an S–I–R model; at any time point, each individual is in one of the three states: susceptible (S), infectious (I) or removed (R); a susceptible individual may become infectious and an infectious individual may become removed. Assume there are m people in one household. The transition of the health status of people in one household is described