data. I have encountered many incomplete multivariate datasets where the "ideal" imputation model has far more parameters than the observed data can estimate; simulating imputations using Bayesian methods and standard noninformative priors simply does not work. When this happens, the imputer may either (i) trim the model by omitting less-crucial variables or restricting the parameter space, or (ii) stabilize the inference by applying a mildly informative prior distribution. The first option may be easier and less controversial, but the second may be more satisfying from an inferential point of view. Choosing an informative prior distribution can be made more automatic and less subjective by allowing some aspects of the prior to be determined by the data, in the spirit of empirical Bayes. A discussion of model trimming for imputation of a large, multipurpose sample survey is given by Schafer, Khare and Ezzati-Rice (1993). An example of a mild, data-determined informative prior for categorical data is given by Clogg et al. (1991). For continuous data, one can often apply a data-determined prior similar to that used in ridge regression. Several analyses of incomplete data sets using informative, data-determined priors will appear in Schafer (1994).

THE NUMBER OF IMPUTATIONS

In practice, a small number of imputations is usually adequate when the fraction of missing information about the estimand is small to moderate. In advance of the analysis, however, it is difficult to know what the fraction of missing information will be. The estimate of this fraction given by Rubin (1987, pages 93–94) can be quite noisy, particularly for small m. For this reason, Meng's suggestion that imputers make available a generous number of imputations (say $m = 30$) is wise, even if most analysts will use only a smaller subset of them for any particular inference. Once 30 or more imputations are made available, however, I suspect that analysts will eventually gravitate toward using all of them rather than just a subset. Otherwise, questions about the objectivity of published analyses (Did they really select their imputations at random?) will naturally arise. Moreover, the analysts themselves will probably want to look at more imputations than they really need. When working with a small number of imputations, there is always a gnawing question in the back of my mind: What will happen if I add just a few more? I have performed analyses in which an effect looks statistically significant (p-value less than 0.05) with $m = 5$, but the significance disappears for $m \geq 10$. When generating imputations for personal use, I have a strong temptation to use a larger-than-necessary value of m just to remove as much random variation as possible from the final summary statistics. I suspect that many analysts, like myself, would have strong desire for the results of their analyses to be essentially deterministic and reproducible by another analyst working with the same observed data. When multiply imputed data files are released to the public, the complete set of m imputations—however large m is—will tend to develop an air of authenticity and objectivity that arbitrary subsets will not have.

Comment

Chris Skinner

Meng's paper provides both a response to Fay's (1991, 1992) specific critique of multiple imputation as a method of variance estimation, and also a general case for multiple imputation as a method of both point and interval estimation. My comments will address these two aspects separately.

Fay (1991, 1992) presented examples where variance estimators based on multiple imputation could be inconsistent. Doctor Meng's framework, in particular the introduction of the concept of "uncongenial" to apply to differences between an imputation model and an analysis procedure, is I think very helpful for understanding such examples. One of Fay's examples is essentially that in Section 3.1. Meng's analysis agrees with Fay's in finding that, even though the imputation model may be correct and the analysis procedure may be sensible, the multiple-imputation variance estimator may be inconsistent. Meng argues, however, both for this specific example and in the Main Result more generally, that under reasonable conditions multiple-imputation intervals will be conservative and their width will be bounded by the width of confidence intervals based on corresponding incomplete-data procedures.

Chris Skinner is Professor, Department of Social Statistics, University of Southampton, Southampton S017 1BJ, United Kingdom.