EVERY MISSING NOT AT RANDOM MODEL HAS GOT A MISSING AT RANDOM COUNTERPART WITH EQUAL FIT

Geert Molenberghs (1,2), Michael G. Kenward (3), Geert Verbeke (2,1),
Caroline Beunckens (1) , and Cristina Sotto (1)

(1) Center for Statistics, Hasselt University, Diepenbeek, Belgium
(2) Biostatistical Centre, Katholieke Universiteit Leuven, Belgium
(3) Medical Statistics Unit, London School of Hygiene and Tropical Medicine, UK

Over the last decade a variety of models to analyze incomplete multivariate and longitudinal data have been proposed, many of which allowing for the missingness to be not at random (MNAR), in the sense that the unobserved measurements influence the process governing missingness, in addition to influences coming from observed measurements and/or covariates. The fundamental problems implied by such models, to which we refer as sensitivity to unverifiable modeling assumptions, has, in turn, sparked off various strands of research in what is now termed sensitivity analysis. The nature of sensitivity originates from the fact that an MNAR model is not fully verifiable from the data, rendering the empirical distinction between MNAR and random missingness (MAR), where only covariates and observed outcomes influence missingness, hard or even impossible, unless one is prepared to accept the posited MNAR model in an unquestioning way. In this paper, we show that the empirical distinction between MAR and MNAR is not possible, in the sense that each MNAR model fit to a set of observed data can be reproduced exactly by an MAR counterpart. Of course, such a pair of models will produce different predictions of the unobserved outcomes, given the observed ones. Theoretical considerations are supplemented with an illustration based on the Slovenian Public Opinion survey, analyzed before in the context of sensitivity analysis.

Missing data can be seen as latent variables. Such a view allows extension of our results to other forms of coarsening, such as grouping and censoring. In addition, the technology applies to random effects models, where a parametric form for the random effects can be replaced by certain other parametric (and non-parametric) form, without distorting the model's fit, latent classes, latent variables, etc.

Reference

Molenberghs, G., Beunckens, C., Sotto, C., and Kenward, M.G. (2007) Every missing not at random model has got a missing at random counterpart with equal fit. *Journal of the Royal Statistical Society*, *Series B*, **00**, 000--000.