Shared-parameter Models and Missingness at Random

An Creemers[1], Niel Hens[1], Marc Aerts[1], Geert Molenberghs[1,2], Geert Verbeke[2,1], Michael G. Kenward[3]

[1] Center for Statistics, Universiteit Hasselt, B-3590 Diepenbeek, Belgium
[2] Biostatistical Centre, Katholieke Universiteit Leuven, B-3000 Leuven, Belgium
[3] Medical Statistics Unit, London School of Hygiene and Tropical Medicine, London WC1E7HT, UK

When data are incomplete, models are often cataloged according to one of three modeling frameworks to which they belong: selection models (SeM), pattern-mixture models (PMM), and shared-parameter models (SPM). At the same time, the missing data mechanism is conveniently classified as provided by Rubin (1976): missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). Whereas MCAR is naturally simple, in the sense that measurement and missingness processes are fully independent, and hence easy to describe in every modeling framework, this is less the case for MAR. The conventional definition (Rubin 1976) is cast in the SeM framework. Molenberghs *et al* (1998) provided a characterization for PMM.
In this paper, MAR is characterized for the SPM framework too, based upon a general, appealing definition of a broad class of SPM. The obtained characterization of MAR in the general SPM framework results in a complex formula. Therefor, a specific sub-class of SPM models is considered, satisfying a sufficient condition for the SPM to be MAR. This sufficient condition states that the random effects driving the observed measurements and/or the missing-data process, do not influence the missing measurements, given the observed ones. In other words, all information about the missing measurements, apart from covariates, stems from the observed measurements only.
Providing some examples, it is shown that not all situations satisfying the characterization of MAR in the general shared-parameter family, are included in the considered MAR sub-family. However, these situations will necessarily be more *ad hoc* and less intuitive than these included in the sub-family. Further, particular implications for time-ordered longitudinal data subject to dropout are studied. It is indicated how SPM can be constrained such that dropout at a given point in time can depend on current and past, but not on future measurements, in analogy to the result by Kenward, Molenberghs, and Thijs (2003) for the PMM family. While a natural requirement, it is less easily imposed in the PMM and SPM frameworks than in the SeM case.
Finally, some of the models proposed are illustrated using a clinical trial in toenail dermatophyte onychomycosis.

References

Kenward, M.G., Molenberghs, G., and Thijs, H. (2003). Pattern-mixture models with proper time dependence. *Biometrika* **90**, 53–71.

Molenberghs, G., Michiels, B., Kenward, M.G., and Diggle, P.J. (1998). Monotone missing data and pattern-mixture models. *Statistica Neerlandica* **52**, 153–161.

Rubin, D.B. (1976). Inference and missing data. *Biometrika* **63**, 581–592.