The R package mice imputes incomplete multivariate data by chained equations.

Multiple imputation (Rubin 1987, 1996) is the method of choice for complex incomplete data problems. Missing data that occur in more than one variable presents a special challenge.

Two general approaches for imputing multivariate data have emerged: joint modeling (JM) and fully conditional specification (FCS), also known as multivariate imputation by chained equations (MICE).

**JM** involves specifying a multivariate distribution for the missing data, and drawing imputation from their conditional distributions by Markov chain Monte Carlo (MCMC) techniques. This methodology is attractive if the multivariate distribution is a reasonable description of the data. **FCS** specifies the multivariate imputation model on a variable-by-variable basis by a set of conditional densities, one for each incomplete variable. Starting from an initial imputation, FCS draws imputations by iterating over the conditional densities. A low number of iterations (say 10–20) is often sufficient.

FCS is attractive as an alternative to JM in cases where no suitable multivariate distribution can be found. The basic idea of FCS is already quite old, and has been proposed using a variety of names: stochastic relaxation (Kennickell 1991), variable-by-variable im- putation (Brand 1999), regression switching (van Buuren et al. 1999), sequential regressions (Raghunathan et al. 2001), ordered pseudo-Gibbs sampler (Heckerman et al. 2001), partially incompatible MCMC (Rubin 2003), iterated univariate imputation (Gelman 2004), MICE (van Buuren and Oudshoorn 2000; van Buuren and Groothuis-Oudshoorn 2011) and FCS (van Buuren 2007).

Features

This paper describes the R package mice 2.9 for multiple imputation: generating multiple imputation, analyzing imputed data, and for pooling analysis results. Specific features of the software are:

Columnwise specification of the imputation model (Section 3.2). Arbitrary patterns of missing data (Section 6.2). Passive imputation (Section 3.4). Subset selection of predictors (Section 3.3).

Support of arbitrary complete-data methods (Section 5.1). Support pooling various types of statistics (Section 5.3). Diagnostics of imputations (Section 4.5). Callable user-written imputation functions (Section 6.1).

