ABSTRACT
Imputation of data with general structures (e.g., data with continuous, binary, unordered categorical, and ordinal variables) is commonly performed with fully conditional specification (FCS) instead of joint modeling. A key drawback of FCS is that it does not invoke an appropriate data augmentation mechanism and as such convergence of the resulting Markov chain Monte Carlo procedure is not assured. Methods that use joint modeling lack these drawbacks but have not been efficiently implemented in data of general structures. We address these issues by developing a new method, General Efficient Regression-Based Imputation with Latent processes (GERBIL), that draws imputations from a latent joint multivariate normal model that underpins the generally structured data. This model is constructed using a sequence of flexible conditional linear models that enables the resulting procedure to be efficiently implemented on high dimensional datasets in practice. Simulations show that CMI performs well when compared to those that utilize FCS. Furthermore, the new method is computationally efficient in comparison to FCS procedures. The R package gerbil that implements the method is also discussed.