A Review of Wu et al.'s Bayesian Model for Handling Nonignorably Missing Data

Rachel Tsong and Rebekah Yu

11/25/2019

Background and Motivation

Missing data are often a concern in longitudinal clinical trials. It is not uncommon for subjects enrolled in longitudinal trials to have intermittent missingness due to missed visits or permanent missingness due to drop-out. These missing data can be classified into three categories: missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). Data are considered to be MCAR if the missingness is independent of both observed and unobserved data. Data are MAR if the missingness is independent of the unobserved data conditioned on the observed data. MCAR and MAR are classified as "ignorable" mechanisms of missingness because the missing data mechanism does not need to be specified in the log-likelihood. However, MNAR data are when the probability of missingness depends on some unobserved variable, and this missing data mechanism is referred to as "nonignorable" because the missingness mechanism needs to be incorporated into the likelihood estimates. Inappropriate handling of missing data can lead to biased inferences, thus it is of great clinical significance to develop methods for handling missing data.

Bayesian methods are a useful approach for dealing with missing data because they consider missing data as random variables whose posterior distributions can be obtained by specifying priors and missing covariate distributions. Bayesian approaches have been shown to produce more reliable results compared to traditional missing data models such as multiple imputation. We reviewed a Bayesian model for nonignorably missing data with binary responses in a longitudinal study developed by Wu et al. (2018). Using the methods they developed, we used their method on a simulated dataset and assessed model performance.

The Proposed Models and Simulation

Longitudinal binary responses can be modeled by the following probit mixed-effects regression model for t=0, 1, 2, 3 time points where x_{i1} are baseline covariates, β_t^* are regression coefficients corresponding to the

treatment condition, ζ_k are the random effects with variance τ^*:

$$P(y_{it} = 1|z_i, x_{1i}, x_{2i}, \beta_t^*, \epsilon_t^*) = \Phi(\beta_{0t}^* + x_{i1}\beta_{1t}^* + x_{2i}\beta_{2t}^* + z_i\beta_{3t}^* + \epsilon_{it}^* \ (1)$$

We let x_{ij}~$N(0, 1)$, $x_{2i}|x_{1i}$~$Bernoulli(\frac{1}{1+e^{-0.2-0.2x_{1i}}})$, z_i~$Bernoulli(0.5)$, ϵ~$N(0, \frac{\sigma^2}{1+\sigma^2}\Sigma)$ where $\sigma^2 = 2$ and Σ is a 4x4 AR(1) correlation matrix with $\rho = 0.8$. The beta matrix is:

The missing data mechanism R_{it} was modeled R_{it}~$Bernoulli(P_{it})$ where P_{it} is:

$$logit(P_{it}) = \gamma_{0t} + x_{1i}\gamma_{1t} + x_{2i}\gamma_{2t} + z_i\gamma_{3t} + \sum R_{ij}\gamma_{4t} + y_{it-1}\gamma_{5t} + y_{it}\gamma_{6t}$$

The gamma matrix is:

References

Ma Z, Chen G. Bayesian methods for dealing with missing data problems. Journal of the Korean Statistical Society. 2018;47(3):297-313. doi:10.1016/j.jkss.2018.03.002.

Wu J, Ibrahim JG, Chen MH, Schifano ED, Fisher JD. Bayesian Modeling and Inference for Nonignorably Missing Longitudinal Binary Response Data with Applications to HIV Prevention Trials. Stat Sin. 2018;28:1929–1963. doi:10.5705/ss.202016.0319