

STATEMENT OF RESEARCH INTERESTS

My work spans several major areas of biostatistics and is driven by an interest in developing statistically sound solutions to real-world problems. My methodological research interests can be categorized into two broad areas: analysis of incomplete longitudinal data and stochastic simulation.

As far as missing data are concerned, I focus on multiple imputation (MI) inferences where missing observations are replaced by a set of plausible values that are drawn from a predictive distribution. Once multiple versions of the completed data sets are obtained, one can proceed with analyzing them with standard complete-data methods, and incorporating the results into a single inferential summary. As a result, with MI, uncertainty due to missing data is formally taken into account in the modeling process. I have been working on MI-based methodologies over the years. In particular, I explored sensitivity issues of pattern-mixture models in a longitudinal setting via conducting MI under linear mixed-effects models (Demirtas and Schafer, 2003). Subsequently, I developed a new class of hierarchical Bayesian pattern-mixture models for nonignorable missing continuous longitudinal data (Demirtas 2005a, 2005b, 2004b). In addition to parametric conditional models, I have been interested in semiparametric marginal approaches such as weighted estimating equations (Demirtas, 2004c). One future research direction will be exploring innovative MI techniques for binary, nominal, and ordinal longitudinal or clustered data within the mixed-effects modeling framework through a saturated multinomial model that allows for arbitrarily complex associations.

From stochastic simulation point of view, I am deeply interested in random number generation (Demirtas, 2004d, 2005c, 2006), and modern computational statistics topics such as EM-related algorithms, MCMC (data augmentation, Gibbs sampler), density estimation, bootstrapping, smoothing, Bayesian computing and non-linear optimization to the extent they apply to missing-data problems. I have a few published or accepted papers on simulation-based inferences within the MI paradigm (Demirtas, 2004a; Demirtas et al., 2007; Demirtas, Freels and Yucel, 2007; Demirtas, 2007a, 2007b; Demirtas and Hedeker, 2008). I plan to maintain a substantial degree of exposure to simulation notion in the context of incomplete data in the future. I find the idea of simulation --where we describe a real phenomenon by generating an environment within which stochastic process under consideration is assumed to operate--exciting. Creating mirror images of the perceived truth, and iteratively refining and re-defining the truth have always been thrilling to me.

Being trained at a math/stat department, and currently working in a public-health oriented biostatistics department, give me a good opportunity for striking a delicate balance of technique and judgment. In other words, methodological advancements guided by applied context of the problem ideally suits to my background and current position.

My intermediate research goal is relating random number generation/statistical distribution theory and MI, which I believe are closely connected (Demirtas and Hedeker, 2007). The fundamental step in parametric MI is filling in the missing data by drawing from the conditional distribution of the missing data given the observed data. This usually entails positing a model for the data and using it to derive this conditional distribution. For continuous data, multivariate normality among the variables has been perceived as a convenient assumption since the conditional distribution of the missing data given the observed data is then also multivariate normal. Recently, extending the practice of MI from normality to more general classes of densities has begun to receive attention. Considering the restrictive nature of the normality assumption, employing a distributional setup that spans a broader spectrum of symmetry-elongation behavior in the imputation process may provide a reasonable way to handle non-Gaussian continuous data. In this regard, generalized distributions are sensible alternatives because of the ability of accommodating a variety of distributional shapes. Imputation under generalized distributions such as Tukey's classes, the Burr family, the Johnson Family, the Pearson family, generalized lambda and Beta families, and Fleishman polynomials, that include many standard densities as exact or approximate special cases, appears to have significant potential to capture real-data trends. I have two unpublished works that are currently under review on MI under Fleishman's power polynomials and the generalized lambda distribution. I believe that this line of research has prospects to be very fruitful. On a related note, one can take advantage of well-studied computational and conceptual merits of Gaussian imputation model, yet still adequately address non-normal features (skewness, heavy tails, flat regions, multimodality, skip patterns, etc.) of data by probability integral transformation that makes a connection between quantile functions. This promising possible solution has not been investigated from an imputation standpoint. I would like to direct a major portion of my intellectual energy to this, along with richer families of parametric distributions that I mention above.

Thinking beyond the realm of normality also has potential of leading to substantial advances in modeling non-normal random effects, especially in longitudinal settings. Estimating random effects nonparametrically, and subsequently expressing the empirical cumulative distribution function in terms of normal quantile function through probability integral transformation might be a viable approach. Non-normal MI models and random effects can be studied simultaneously considering their shared characteristics, and that is what I plan to do in the near future.

On the collaborative front, my concentration will be on genetic epidemiology. I have been working on genome-wide association studies to separate the environmental and genetic factors in developing complex diseases with researchers from different medical disciplines. These projects produced a few publications in prestigious medical journals such as New England Journal of Medicine and Lancet. I plan to continue to work on these areas.

Note: References are in accordance with my CV (Pages 4 and 5).