10-2017

Discussion of ”Dissecting multiple imputation from a multi-phase inference perspective: what happens when god's, imputer's and analyst's models are uncongenial?”

Shu Yang
North Carolina State University

Jae Kwang Kim
Iowa State University, jkim@iastate.edu

Discussion of ”Dissecting multiple imputation from a multi-phase inference perspective: what happens when god’s, imputer’s and analyst’s models are uncongenial?”

Abstract

We would like to first congratulate Drs. Xie and Meng on their excellent work on investigating the mystery of multiple imputation. Multiple imputation (MI) has been promoted as a general purpose estimation tool for missing data, but there are debates over its statistical validity in many practical situations. This article will certainly serve an important building block to address these debates from a multiphase inference perspective.

Disciplines

Statistical Methodology | Statistical Models

Comments

This article is part of a discussion from S. Yang and J.K. Kim (2017). Discussion of ”Dissecting multiple imputation from a multi-phase inference perspective: what happens when god’s, imputer’s and analyst’s models are uncongenial?” by Xie and Meng, *Statistica Sinica*, 27, 1485-1594. doi: 10.5705/ss.2014.067. Posted with permission.

Lehmann, E. L. (1959). *Testing Statistical Hypotheses.* John Wiley, New York.

Meng, X. L. and van Dyk, D. A. (1997). The EM Algorithm- An old folk song sung to a fast new tune (with discussion). *Journal of the Royal Statistical Society* B **59**, 511-567.

Neyman, J. (1934). On the two different aspects of the representative method. *Journal of the Royal Statistical Society* B **97**, 558-625.

Neyman, J. (1935). On the problem of confidence intervals. *Annals of Mathematical Statistics* **6**, 111-116.

Neyman, J. (1937). Outline of a theory of statistical estimation based on the classical theory of probability. *Philos. Trans. R. Soc. Lond.* A **236**, 333-380.

Neyman, J. (1941). Fiducial argument and the theory of confidence intervals. *Biometrika*, **32** 128-150.

Neyman, J. (1977). Frequentist probability and frequentist statistics. *Synthese* **36**, 97-131.

Pytkowski, W. (1932). *The Dependence of the Income in Small Farms upon Their Area, the Outlay and the Capital Invested in Cows*. Bibljoteka Pulawska, Warsaw.

Reid, C. (1982). *Neyman from Life*. Springer, New York.

Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of American Statistical Association* **91**, 473-489.

Small, C. G. and McLeish, D. L. (1994). *Hilbert Space Methods in Probability and Statistical Inference*. Wiley, New York.

Tu, X. M., Meng, X.-L. and Pagano, M. (1993). The AIDS epidemic: Estimating survival after AIDS diagnosis from surveillance data. *Journal of American Statistical Association* **88**, 26-36.

Department of Mathematics and Statistics, University of Guelph, Ontario, N1G 2W1, Canada
E-mail: tdesmond@uoguelph.ca

(Received June 2016; accepted June 2016)

DISCUSSION

BY X. XIE AND X. L. MENG

Shu Yang and Jae Kwang Kim

North Carolina State University and Iowa State University

1. Introduction

We would like to first congratulate Drs. Xie and Meng on their excellent work on investigating the mystery of multiple imputation. Multiple imputation (MI) has been promoted as a general purpose estimation tool for missing data, but there are debates over its statistical validity in many practical situations. This article will certainly serve an important building block to address these debates from a multi-phase inference perspective.

Multiple imputation was originally designed to handle missing data for public-released databases. The imputation process and subsequent analyses of the imputed datasets are separate. Therefore, this multi-phase inference features the possibility of uncongeniality. The authors focused on $m = \infty$ to avoid Monte Carlo error and introduced simple examples to highlight a number of key concepts. Specifically, we would like to discuss robustness, self-efficiency, confidence validity, and the links with the EM algorithm and fractional imputation.

2. Robustness

The authors demonstrated the hidden robustness when the analyst assumes more than the imputer through a simple example in Section 2.2. In the missing data literature, two lines of research have focused on different parts of distributions: multiple imputation models the data distribution; inverse probability weighting and doubly robust estimation (Bang and Robins (2005); Kang and Schafer (2007)) model the response probability. To gain robustness, researchers have investigated combining inverse probability weighting and multiple imputation to improve robustness of estimation (Seaman et al. (2012); Han (2015)). The authors' theory for MI can be used to cover these phenomenons.

We would like to point out that robustness is generally achievable in many imputation methods. To illustrate the idea, consider the bivariate data $(x_i, y_i), i = 1, \cdots, N$, with y_i being subject to missingness. Without loss of generality, assume the first n $y's$ are observed and the other $N - n$ $y's$ are missing. Let $m(x; \beta)$ be the "working" model for $E(Y \mid x)$ and take $\widehat{y}_i = m(x_i; \widehat{\beta})$ as the imputed value for y_i, where $\widehat{\beta}$ satisfies $\sum_{i=1}^n \{y_i - m(x_i; \widehat{\beta})\} = 0$. In this case, the regression imputation estimator $\widehat{\theta}_I = N^{-1}\{\sum_{i=1}^n y_i + \sum_{i=n+1}^N \widehat{y}_i\}$ is algebraically equivalent to the two-phase regression estimator

$$\widehat{\theta}_{tp,reg} = N^{-1}\sum_{i=1}^{N}\widehat{y}_i + n^{-1}\sum_{i=1}^{n}(y_i - \widehat{y}_i).$$

Under MCAR, using the argument in Kim and Rao (2012), $\widehat{\theta}_I$ is asymptotically unbiased regardless of the choice of $m(x_i; \beta)$. If the response probability $\widehat{\pi}_i$ is available, then we can include $\widehat{\pi}_i^{-1}$ in X so that $\sum_{i=1}^n \widehat{\pi}_i^{-1}(y_i - \widehat{y}_i) = 0$ holds. Then, the regression imputation estimator is algebraically equivalent to

$$\widehat{\theta}_{tp,reg} = N^{-1}\sum_{i=1}^{N}\widehat{y}_i + N^{-1}\sum_{i=1}^{n}\widehat{\pi}^{-1}\left(y_i - \widehat{y}_i\right),$$

which is also asymptotically unbiased regardless of the choice of $m(x_i;\beta)$. Thus, as long as the column space of X includes $\widehat{\pi}_i^{-1}$, the resulting imputed estimator is doubly robust. This is essentially the main idea of doubly robust imputation as discussed in Kim and Haziza (2014).

3. Self-efficiency

We believe that self-efficiency is defined with respect to an analyst's model and the missing data mechanism. We agree that self-efficiency is indeed a weaker requirement than self-sufficiency, but is frequently violated in common practice for multi-purpose estimation. Even in the ideal case when the imputer and the analyst's models are congenial, the requirement for the complete-data estimator to be self-efficient is restrictive. We have examined several scenarios, which are fairly common in practice; however they fail this requirement.

Example 1. Consider a simple linear regression model $Y = \beta_0 + \beta_1 X + \epsilon$, where $\epsilon \sim N(0,\sigma^2)$, X is always observed, and Y is subject to missingness with MAR. Suppose the analyst is interested in estimating $\mu = E(Y)$ and $\eta = E\{I(Y < c)\}$, where c is a prespecified value. The complete-sample estimator solving $\sum_{i=1}^{n} Y_i - \mu = 0$ is self-efficient; however, the complete-sample estimator solving $\sum_{i=1}^{n} I(Y_i < c) - \eta = 0$ is not self-efficient.

Example 2. Consider the setup of Example 1 with $\beta_0 = 0$. Suppose the analyst is interested in estimating $\mu = E(Y)$ and consider the complete-sample estimator by solving $\sum_{i=1}^{n} Y_i - \mu = 0$. Yang and Kim (2016) claimed the Rubin's combining rule is not consistent in this case. There are two ways of viewing this in XM's framework: under the model $Y = \mu + \epsilon$, with $\epsilon \sim N(0,\sigma^2)$, the analyst's estimation procedure is self-efficient, but the model is not congenial with the imputer's model; under the model $Y = \beta_0 + \beta_1 X + \epsilon$, with $\epsilon \sim N(0,\sigma^2)$, the analyst' estimation procedure is not self-efficient.

Example 3. Consider a log linear regression model, $\log Y = X^T\beta + \epsilon$, where $\epsilon \sim N(0,\sigma^2)$. This model is especially useful for economic data that have skewed populations where the assumption of a normal distribution is unlikely to hold. Under this model, the analyst's complete-sample estimator of $\mu = E(Y)$ solving $\sum_{i=1}^{n} Y_i - \mu = 0$ is not self-efficient. This example is discussed in Yang and Kim (2015).

4. Confidence Validity Versus Type 2 Error

The authors suggest constructing a conservative variance estimator $2T_\infty$ for which the multiple imputation procedure has confidence validity. Our concern is how useful "confidence validity" is. Being conservative can protect Type 1 error, but how about Type 2 error? We can have a situation where the statistical power of the test based on MI is so low that it is better not to perform MI at all. To illustrate this point, we performed a simple simulation study. In the simulation, $B = 2,000$ Monte Carlo samples of size $n = 1,000$ were independently generated from

$$y_i = -1.5 + \beta_1 x_i + e_i, \tag{4.1}$$

where $\beta_1 \in \{0.05, 0.1, 0.15\}$, $x_i \sim N(2,1)$, $e_i \sim N(0, 1.04)$, and x_i and e_i are independent. Variable x_i is always observed but the probability π_i that y_i responds follows $\text{logit}(\pi_i) = -1 + 0.5x_i$.

For each realized sample, we computed two estimators: the Complete-Case (CC) method that only uses the complete cases for the regression analysis and the MI estimator with $m = 100$. The imputer's and analyst's models are correctly specified as (4.1). The prior for the parameters is a flat prior.

From the imputed data, we computed the 95% confidence intervals for β_1. For the MI estimator, we used the conservative method $2T_\infty$. Table 1 shows that the MI method loses quite a bit of power compared to the CC method. While the point estimators are essentially the same in both methods, variance estimator in MI is positively biased and the test based on MI is less powerful.

Table 1. Results of power estimates for testing $H_0 : \beta_1 = 0$ based on $B = 2,000$ simulated datasets. CC: the complete-case estimator; MI: the multiple imputation estimator with the conservative variance estimator.

	CC	MI
$\beta_1 = 0.05$	0.2	0.04
$\beta_1 = 0.10$	0.56	0.28
$\beta_3 = 0.15$	0.90	0.66

5. Links with EM Algorithm and Fractional Imputation

The theoretical setup in Section 4 in XM's article serves as a general platform that links several important techniques, such as the EM algorithm (Dempster, Laird and Rubin (1977)), Data Augmentation (Tanner and Wong (1987)), and Fractional Imputation (Kim (2011); Yang and Kim (2015)). MI was originally motivated in a Bayesian prospective, but its frequentist properties have been studied by a number of researchers via the Bernstein-von Mises theorem. See for example, Robins and Wang (2000); Yang and Kim (2016). Following the authors' notation, $\bar{\theta}_\infty$ is the solution to

$$E\{S^A(Z_{com};\theta^A) \mid Z_{obs};\widehat{\theta}^I_{obs}\} = 0. \tag{5.1}$$

Here, $S^A(Z_{com};\theta^A)$ is not necessarily the score function, rather, it is the estimating function that defines the parameter. That is, θ is defined through $E\{S^A(Z_{com};\theta^A)\} = 0$. If $S^A(Z_{com};\theta^A)$ is chosen to be the score function, the method is equivalent to the EM algorithm.

Fractional imputation is another effective imputation tool for general-purpose estimation with its advantage of not requiring the congeniality condition. With $m = \infty$, the fractional imputation estimator of θ^A is also the solution to (5.1), where $\widehat{\theta}^I_{obs}$ is a consistent estimator of θ^I in the imputation model. Rubin's approach of multiple imputation conducts separate analyses and then combining them, whereas fractional imputation creates a single weighted imputed dataset for analysis. To investigate the asymptotic variance of $\bar{\theta}^A_\infty$, we can view $\bar{\theta}^A_\infty = \bar{\theta}^A_\infty(\widehat{\theta}^I_{obs})$ and apply Taylor linearization:

$$\begin{aligned}\bar{\theta}^A_\infty(\widehat{\theta}^I_{obs}) &\cong \bar{\theta}^A_\infty(\theta^I_0) + E(\frac{\partial\bar{\theta}^A_\infty}{\partial\theta^I})(\widehat{\theta}^I_{obs} - \theta^I_0)\\ &\cong \bar{\theta}^A_\infty(\theta^I_0) - E(\frac{\partial\bar{\theta}^A_\infty}{\partial\theta^I})E\{\frac{\partial S^I(Z_{obs};\theta^I_0)}{\partial\theta^I}\}^{-1}S^I(Z_{obs};\theta^I_0),\end{aligned}$$

where $\widehat{\theta}^I_{obs}$ is the solution to $S^I(Z_{obs};\theta^I) = 0$. Thus, the variance of $\bar{\theta}^A_\infty(\widehat{\theta}^I_{obs})$ is approximated by the variance of $\bar{\theta}^A_\infty(\theta^I_0) - BS^I(Z_{obs};\theta^I_0)$, where $B = E(\partial\bar{\theta}^A_\infty/\partial\theta^I)$ $E\{\partial S^I(Z_{obs};\theta^I_0)/\partial\theta^I\}^{-1}$. This is the standard linearization method for imputation variance estimation, as discussed by Clayton et al. (1998), Robins and Wang (2000), Kim (2011), and Yang and Kim (2015). Resampling method will also provide valid variance estimation. Therefore, the fractionally imputed dataset coupled with replicated resampling weights provide another basis for consistent inference for multi-purpose usage. Of course, this may come at the price of a larger data storage space and more complex analysis.

6. Concluding Remarks

We conclude by thanking XM for their enlightening article, and we appreciate the opportunity to offer our viewpoints on this interesting problem. We look forward to their responses to our major points regarding robustness, self-efficiency, confidence validity, and the links with fractional imputation.

References

Bang, H. and Robins, J. M. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics* **61**, 962–973.

Clayton, D., Spiegelhalter, D., Dunn, G. and Pickles, A. (1998). Analysis of longitudinal binary data from multiphase sampling. *J. R. Stat. Soc. Ser. B.* **60**, 71–87.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Ser. B.* **39**, 1–38.

Han, P. (2015). Combining inverse probability weighting and multiple imputation to improve robustness of estimation. *Scandinavian J. Stat.* **43**, 246–260.

Kang, J. D. Y. and Schafer, J. L. (2007). Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statistic. Sci.* **22**, 523–539.

Kim, J. K. (2011). Parametric fractional imputation for missing data analysis. *Biometrika* **98**, 119–132.

Kim, J. K. and Haziza, D. (2014). Doubly robust inference with missing survey data. *Statistica Sinica* **24**, 375–394.

Kim, J. K. and Rao, J. N. K. (2012). Combining data from two independent surveys: a model-assisted approach. *Biometrika* **99**, 85–100.

Robins, J. M. and Wang, N. (2000). Inference for imputation estimators. *Biometrika* **87**, 113–124.

Seaman, S. R., White, I. R., Copas, A. J. and Li, L. (2012). Combining multiple imputation and inverse-probability weighting. *Biometrics* **68**, 129–137.

Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *J. Amer. Statist. Assoc.* **82**, 528–540.

Yang, S. and Kim, J. K. (2015). Fractional imputation in survey sampling: A comparative review. *Statistical Science* **31**, 415–432.

Yang, S. and Kim, J. K. (2016). A note on multiple imputation for method of moments estimation. *Biometrika* **103**, 244–251.

Department of Statistics, North Carolina State University, Raleigh, NC 27695, USA.

E-mail: syang24@ncsu.edu

Department of Statistics, Iowa State University, Ames, IA 50011, USA.

E-mail: jkim@iastate.edu

(Received March 2016; accepted April 2016)

DISCUSSION

Roderick Little and Tingting Zhou

University of Michigan

Xie and Meng's paper is a theoretical tour de force, providing further insight