Using an Approximate Bayesian Bootstrap to Multiply Impute Nonignorable Missing Data

Juned Siddique
The University of Chicago

Thomas R. Belin
UCLA

Reasons for Missing Data

Rubin (1976) Classifies the reasons for missing data as either *ignorable* or *nonignorable*.

In a dataset where variables X are fully observed and variables Y have missing values:

- Ignorable Missingness: Missing Y are only randomly different from observed Y when conditioning on X
- Nonignorable Missingness: Missing Y are systematically different from observed Y even when conditioning on X
- E.g. Missing Y are typically 20% larger than observed Y with the same values of X

Hot-deck Imputation

- Hot-deck imputation is an imputation technique that replaces missing values in a data set (donees) with observed values (donors)
- Most hot-deck imputation methods attempt to match donors and donees based on observed covariates
- *Predictive mean matching* (PMM), Rubin (1986), Little (1988)

Appoximate Bayesian Bootstrap

- Rubin (1987) suggests an Approximate Bayesian Bootstrap (ABB) to incorporate appropriate uncertainty into hot-deck procedures
- To incorporate an ABB into a hot-deck imputation model, the steps are:
  1. Draw a bootstrap sample of the observed values
  2. Impute missing values by drawing donors from this bootstrap sample
  3. For multiple imputation, repeat steps 1 and 2 m times, so that m bootstrap samples are drawn
- Applies an improper prior on the donor selection probabilities

Nonignorable ABB

- Rubin and Schenker (1991) discuss how an ABB can be modified to handle nonignorable missing data. Instead of drawing n_{obs} cases of Y_{obs} randomly with replacement (i.e. with equal probability), they suggest
- Draw n_{obs} cases of Y_{obs} with probability proportional to Y_{obs}^c so that the probability of selection for for $y_i \in Y_{obs}$ is

$$\frac{y_i^c}{\sum_{j=1}^{n_{obs}} y_j^c}. \tag{1}$$

- This skews the nonrespondents to have typically larger (when $c > 0$ and $y_j > 0$) values of Y than respondents.

Nonignorable ABB Terminology

We refer to ABBs where values of Y_{obs} are drawn with probability proportional to Y_{obs}^c where

- $c = -1$: Inverse-to-size ABB
- $c = 1$: Proportional-to-size ABB
- $c = 2$: Proportional-to-size-squared ABB
- $c = 3$: Proportional-to-size-cubed ABB

Additional Nonignorable ABBs

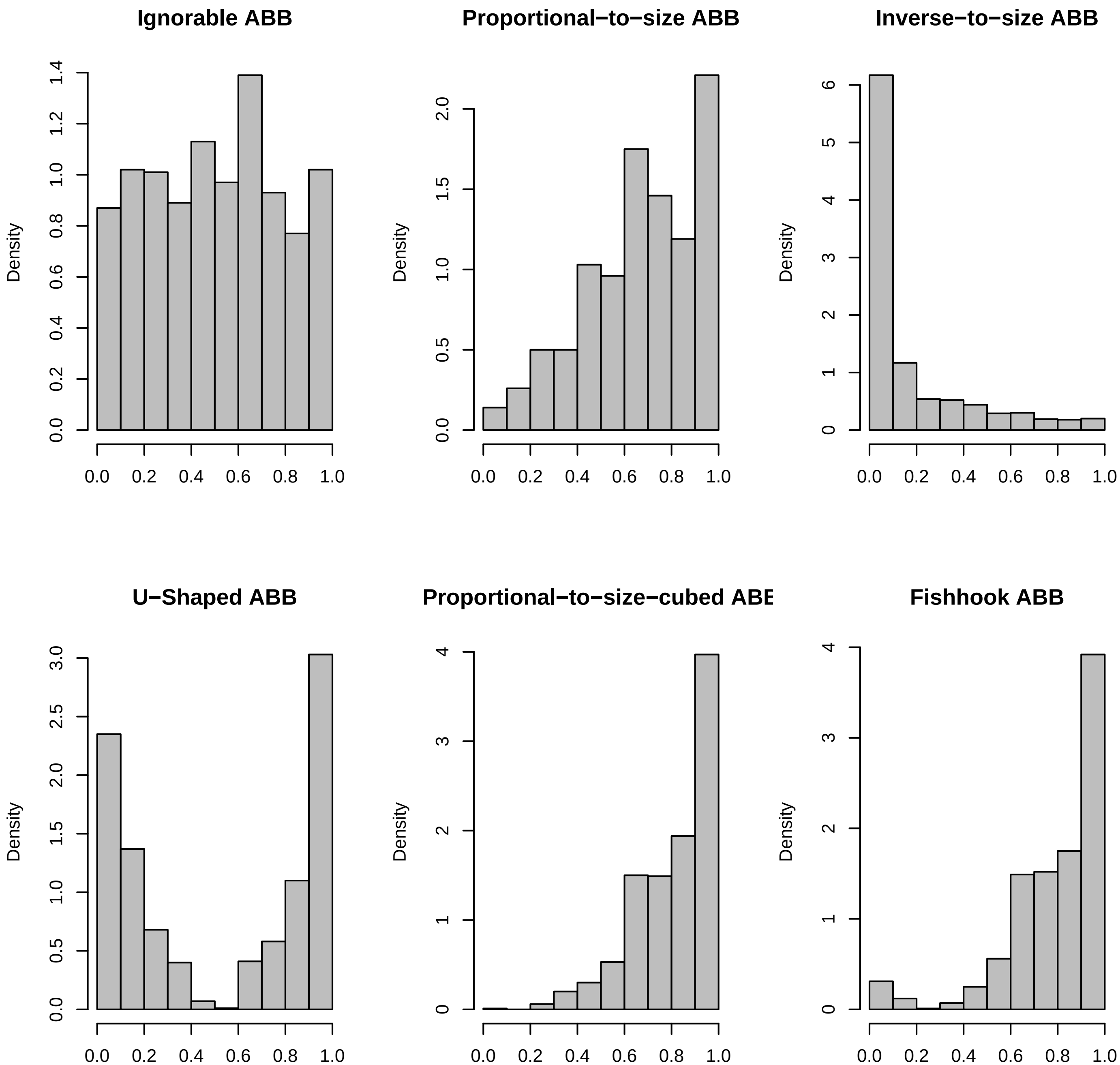
- A nonignorable ABB where n_{obs} cases of Y_{obs} are drawn with probability proportional to Y_{obs} centered around a quantile of Y_{obs}.
- E.g. center observed Y around its median. The implication of drawing with probability proportional to the distance from the median is to favor values for the non-respondents with either larger *or* smaller values than respondents with the same set of covariates (when $c > 0$).
- We refer to this approach as a "U-Shaped ABB" because observations in the extremes of the distribution of Y_{obs} have greater weight than observations in between that are close to the median

Additional Nonignorable ABBs cont'd

- We refer to the ABB that centers the donor sizes around the 1st quantile as a "Fishhook ABB," because this ABB mostly favors large values but retains a U-shaped pattern featuring a slight upturn in the weight given to the smallest observed values.

Nonignorable ABB Illustration

- Simulated $n = 1000$ observations from a uniform distribution
- Drew a ABB sample and recorded the number of times an observation was included in the sample
- Histogram using ABB weights
- Six different ABB types



Ignorable ABB
Proportional−to−size ABB
Inverse−to−size ABB
U−Shaped ABB
Proportional−to−size−cubed ABE
Fishhook ABB
Density

Prior Distributions

- Histograms can be considered prior distributions on the donor selection probabilities.
- Ignorable ABB is akin to a flat prior, and gives each donor equal prior probability of selection
- Other ABB types are informative priors, and give more or less weight to a donor based on the donor size.
- We will see that inferences are affected by the choice of ABB.

Motivating Example: The WECare Study

- The WECare study was a longitudinal depression treatment study of low income women
- Women were measured for depression every month for 6 months, every other month for 6 more months
- Here, we focus on the slope of the medication treatment over 6 months
- Missingness in months 1-6 ranges from 24% to 38%

Experiment Based on Motivating Example

- We multiple imputed the WECare data using the predictive mean matching hot-deck approach of Siddique and Belin (2008).
- A *closeness parameter* k determines the size of the donor pool: As $k \to \infty$ this procedure amounts to a nearest-neighbor hot-deck where the donor whose predicted mean is closest to the donee is always chosen. Conversely, when k equals 0, each donor has equal probability of selection, which is equivalent to a simple random hot-deck.

Experiment Cont'd

We imputed the WECare data many times with the following ABBs.

1. An Ignorable ABB that treats donors as a priori equally likely
2. An Inverse-to-size ABB
3. A Proportional-to-size ABB
4. A Proportional-to-size-cubed ABB
5. A U-Shaped ABB
6. A Fishhook ABB
7. A Mixture ABB where a different ABB is implemented for each of the 5 imputed data sets.

Each ABB was imputed using closeness parameter values 0-10 (we only show for values of 2)

Mixture ABB

- The mixture ABB implements a different ABB for each dataset.
- Motivation is to average over the uncertainty regarding the missing data mechanism
- We will see that this approach provides better coverage than other ABB approaches
- For the WECare Experiment, the following ABBs were used in the mixture ABB
  1. An Inverse-to-size ABB (c=-1)
  2. An Ignorable ABB (c=0)
  3. A Proportional-to-size ABB (c=1)
  4. A Proportional-to-size-squared ABB (c=2)
  5. A Proportional-to-size-cubed ABB (c=3)

Table of Medication Intervention Slopes by ABB Type based on one replication with a closeness parameter value of 2

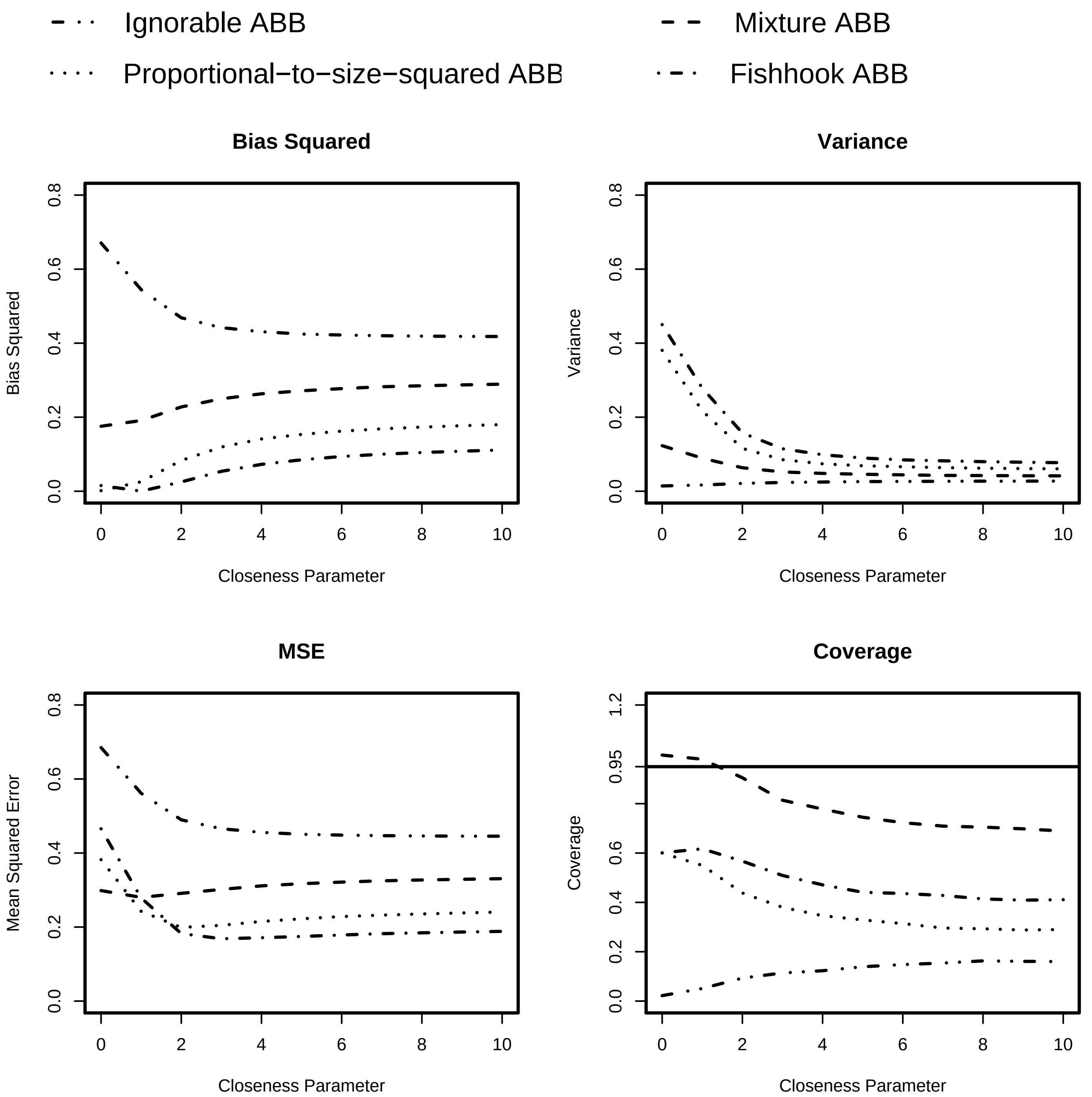
ABB Type	Estimate	Std. Error	t value	$Pr(> \|t\|)$
Ignorable	-2.18	0.47	-4.66	<0.0001
Inverse-to-size	-2.79	0.48	-5.86	<0.0001
Proportional-to-size	-1.83	0.45	-4.05	0.0002
Proportional-to-size-cubed	-1.42	0.43	-3.30	0.0012
U-Shaped	-1.22	0.74	-1.65	0.1319
Fishhook	-1.34	0.52	-2.61	0.0141
Mixture	-2.13	0.70	-3.04	0.0156

Simulation Study

- Bivariate normal-lognormal (Y_1, Y_2) data was simulated where $Y_1 \sim N(1.0, 0.13)$, $Y_2 \sim LN(1.7, 4.7)$, and $\text{corr}(Y_1, Y_2) = 0.54$. ($Y_2 = e^{Z_2}$ where $Z_2 \sim N(0, 1)$)
- 50% nonignorable missingness was applied to the lognormal variable (Y_2) so that larger values tended to be missing
- $N = 100$
- Hot-deck of Siddique and Belin (2008) was used to create five multiply imputed data sets
- Bias, variance, MSE, and coverage were calculated for the mean of Y_2
- 1000 replications

Simulation Study Cont'd

- Data were imputed using closeness parameter values between 0-10
- Investigate the effect of donor pool size on ABB performance
- Four different ABBs
  1. Nonignorable ABB
  2. Proportional-to-size-squared ABB
  3. Mixture ABB
  4. Fishhook ABB


Ignorable ABB
Mixture ABB
Proportional−to−size−squared ABB
Fishhook ABB
Bias Squared
Variance
MSE
Coverage
Bias Squared
Variance
Mean Squared Error
Coverage
Closeness Parameter
0.0
0.2
0.4
0.6
0.8
0.95
1.2
0
2
4
6
8
10

Implementation Guidelines

- Nonignorable ABB proved to be effective, but not all performed equally well
- Mixture ABB outperformed those that used the same ABB for each imputed dataset
- Committing to one ABB when the missing data mechanism is unknown may provide overly precise estimates
- Similarly, avoid strong nonignorability assumptions. I.e. ABB that chooses donors with probability proportional to donor size raised to 100 will always choose largest donor value=single imputation.

Implementation Guidelines Cont'd

- Recommendation is to use Mixture ABB approach
- Accounts for appropriate uncertainly
- Nominal coverage in simulation study
- Large closeness parameter values reduce effectiveness of Nonignorable ABB, use small values: 1 or 2
- Mixture ABB can be designed to favor larger values: $c = -1, 0, 1, 2, 3$, smaller values: $c = -3, -2, -1, 0, 1$ or just to add additional uncertainty: $c = -2, -1, 0, 1, 2$
- Can also incorporate U-Shaped and Fishhook ABBs if deemed plausible mechanisms

End

Extra Slides

Computation Details

For all nonignorable ABBs described above, when values of Y_{obs} are less than or equal to 0, the values of Y_{obs} need to be transformed to ensure that the selection probabilities in the nonignorable ABB are positive and (in the case where Y_{obs} are drawn with probability proportional to Y_{obs}^c, $c > 0$) that the selection probability for $y_i \in Y_{obs}$ is greater than the selection probability for $y_j \in Y_{obs}$ when $y_i > y_j$. Define α and β to be the smallest and second smallest values of Y_{obs} respectively where $\alpha \neq \beta$. Transform $y_i \in Y_{obs}$ using $y_i + |\alpha| + |\alpha - \beta|$. Then Equation 1 is rewritten as

$$\frac{(y_i + |\alpha| + |\alpha - \beta|)^c}{\sum_{j=1}^{n_{obs}} (y_j + |\alpha| + |\alpha - \beta|)^c}. \qquad (2)$$

Transformation only for calculating the selection probabilities. The original values of Y_{obs} are used for imputation.