Multiple Imputation with Propensity Score Models

Tobias Flörchinger

Department of Statistics, Humboldt University Berlin

Motivation

- For observational studies Rosenbaum and Rubin (1983) showed that the bias induced in the estimation of the average treatment effect (**ATE**) due to non-randomised treatment allocation $Z = 1$, can be removed with Propensity Score (**PS**) methods such as Inverse Probability of Treatment Weighting (**IPTW**).
- Non-randomised treatment allocation results in deviating distributions of the individuals' observed covariates in the treatment and control group which PS can balance.
- One major obstacle to overcome is the presence of partially observed covariates to make PS methods feasible. Here Multiple Imputation (**MI**) can offer a solution.
- Searching for an unbiased and inference valid estimator we compare the three MI pooling methods **MIte**, **MIps**, and **MIpar** in a replication of Leyrat et al. (2019) who challenged the claim of Mitra and Reiter (2016) that MIps outperforms MIte.

Propensity Score & IPTW Estimator

- If covariates and treatment allocation are not independent, a confounding bias occurs. The conditional independece assumption is violated $(Y^{Z=1}, Y^{Z=0}) \not\perp Z|X$.
- The PS denotes the individuals probability of receiving the treatment conditional on its covariates $P(Z = 1|X) = p(X)$. It is estimated with X by a logistic regression.
- The PS has a balancing effect, it establishes similar conditional distributions both in the treatment and control group such that $(Y^1, Y^0) \perp Z|p(X)$ recovers the ATE.
- The IPTW estimator is a PS method which aims at simulating a pseudo-population without confounding variables as in a randomized trial. This is achieved by weighting the individuals with their inverse PS, accordingly the group outcomes are given by:

$$\widehat{\mu}_1 = \left(\sum_{i=1}^{n} \frac{Y_i Z_i}{\widehat{e}_i}\right)\left(\sum_{i=1}^{n} \frac{Z_i}{\widehat{e}_i}\right)^{-1} \quad \widehat{\mu}_0 = \left(\sum_{i=1}^{n} \frac{Y_i(1-Z_i)}{1-\widehat{e}_i}\right)\left(\sum_{i=1}^{n} \frac{(1-Z_i)}{1-\widehat{e}_i}\right)^{-1} \tag{1}$$

where Z_i denotes the treatment indicator for individual i ($Z_i = 1$ if treated, 0 otherwise) $\widehat{e}_i$ and $1 - \widehat{e}_i$ the estimated PS and Y_i the individuals outcome.

MI Approaches

For covariate imputation, the R-package *mi* by Su et al. (2011) was used with $M = 10$. It is based on fully conditional specification and the iterations were set to 15. Bayesian GLMs (default) were used with the outcomes included for imputation as in Leyrat et al. (2019).

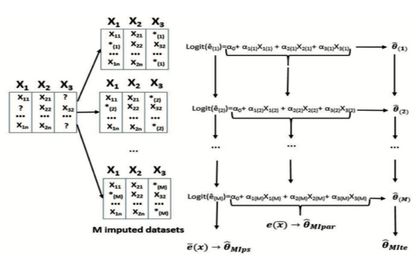

Figure 1: Illustration of the three MI approaches by Leyrat et al. (2019), p.8

- **MIte** averages M values of $\widehat{\mu}_1$ and respectively $\widehat{\mu}_0$
- **MIps** averages the M individual PS and computes $\widehat{\mu}_1$ and $\widehat{\mu}_0$ with these values.
- **MIpar** averages the M individual covariates and parameters of the logistic regressions and computes the individual PS, respectively $\widehat{\mu}_1$ and $\widehat{\mu}_0$ with these values.

Simulation Study

The simulation study follows Leyrat et al. (2019). Datasets of sample size $n = 1000$ were generated, in which the effect of the treatment on a fully observed binary outcome Y is investigated.

- Three covariates (X_1, X_2, X_3) are generated from a multivariate normal distribution, $\mathbf{X} \sim N_3(0, \Sigma)$, $\Sigma_{ii} = 1, \Sigma_{ij} = \rho$ for $i \neq j$. X_3 is dichotomised at a threshold of 0.
- The treatment assignment is specified by the covariates $\mathbf{X}$ in the following model
$$logit(p(Z = 1|X)) = -1.15 + 0.7x_1 + 0.6x_2 + 0.6x_3.$$
- The binary outcome variable is dependent on $\mathbf{X}$ and the treatment received Z, specified by
$$logit(p(Y = 1|Z, x)) = -1.5 + 0.5x_1 + 0.5x_2 + 0.3x_3 + \theta_e Z.$$
- A Missing at Random Mechanism is assumed in which the missingness of X_1 and X_3 is dependent of X_2, the treatment received Z and the outcome Y, specified by
$$logit(p(M_{1,3} = 1|Z, x_1, x_2, x_3, y) = \gamma_0 + z + x_2 + \gamma_y y.$$

The complete factorial design is displayed in Table 1. For all 8 scenarios 1000 samples were generated (Full). Afterwards missings were induced in 30% of data for X_1 and X_3.

Factor	Values	Description
ρ	0.3 or 0.6	Correlation between the covariates
RR	1 or 2	Relative Risk (by adjusting θ_c)
γ_y	0 or -0.4	Association between outcome and missingness probability

Tab. 1: 2 x 2 x 2 factorial design of simulation study.

After imputation the treatment effect of interest is the relative risk given by $\widehat{RR} = \widehat{\mu}_1/\widehat{\mu}_0$ with (1). The performances of MIte, MIps and MIpar are evaluated comparing bias, coverage rates and standardized differences (measure of covariate balancing).

Results

The bias is obtained by the difference in the mean value of $\widehat{RR}$ over the 1000 samples and the corresponding true values of $RR = 1, 2$. For the coverages analytical variance were computed using estimators based on Williamson, Forbes, and White (2014) equation 4 that account for the consecutive estimation of $\widehat{e}$ and $\widehat{RR}$. The confidence intervals were determined with z-scores in the normalising transformation $\log(\widehat{RR})$ and mapped afterwards. Standardized differences were calculated with (2). The performance measures referring to full samples are set as benchmark. Figure 2 and 3 give an overview of the performance in terms of absolute bias and coverage rate and table 3 displays the standardized differences of X_1 in percent.

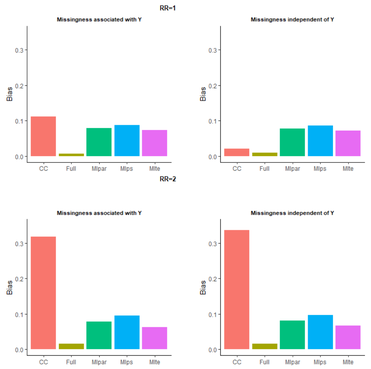

Figure 2: Absolute value of Bias for the methods compared for $\rho = 0.6$

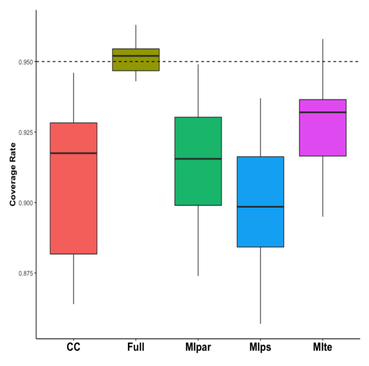

Figure 3: Coverage Rate of the 95% CI for the methods compared

- Among the MI approaches MIte consistently yields the smallest bias and coverage rates closest to the target of 95%. CC performances worst upon all methods.

Method	Full	CC	MIte	MIps	MIpar
Var	0.0113	0.0257	0.0100	0.0092	0.0096
emp.Var	0.0115	0.0280	0.0097	0.0095	0.0096

Tab. 2: Analytical and empirical Variances for $RR = 2$, $\rho = 0.3$, and $\gamma = -0.4$.

- Both variances coincide with CC being highly inefficient and MIte slightly supperior.

$$\text{SDiff} = \frac{100 \times |\bar{X}_1 - \bar{X}_0|}{\sqrt{\frac{\hat{s}_1^2 + \hat{s}_0^2}{2}}} \tag{2}$$

- Standardized differences below 10% indicate a suitable covariate balancing.

Method	Crude	Full	CC	MIte	MIps	MIpar
SDiff	81.52	3.92	10.40	3.50	16.70	16.02

Tab. 3: Standardized differences in X_1(in %) for the condition with $RR = 2$, $\rho = 0.3$, and $\gamma = -0.4$. Crude: data before deletion without IPTW-weighting; Full: data before deletion with IPTW-weighting; CC: Complete Cases.

Conclusion

- Our results replicate the relative advantages of the MIte approach found by Leyrat et al. (2019) over the MIps approach that Mitra and Reiter (2016) favor.
- MIte is less biased than both MIps and MIpar and its coverage rates also correspond more closely to the desired coverages. All three MI methods considerably outperformed estimation with complete cases especially in the conditions with a strong treatment effect $RR = 2$.
- The implicit conditioning on individual imputed data sets helps MIte achieve a much better covariance balancing than both MIps and MIpar achieve as indicated by the very low standardized differences.
- Overall, MIte is a plausible pooling strategy that naturally allows for the incorporation of between- and within-imputation variance in effect estimation using PS for IPTW.

References

Leyrat, C. et al. (2019). "Propensity score analysis with partially observed covariates: How should multiple imputation be used?" In: *Statistical methods in medical research* 28, pp. 3–19.

Mitra, Robin and Jerome P Reiter (2016). "A comparison of two methods of estimating propensity scores after multiple imputation". In: *Statistical methods in medical research* 25(1), pp. 188–204.

Rosenbaum, Paul R and Donald B Rubin (1983). "The central role of the propensity score in observational studies for causal effects". In: *Biometrika* 70(1), pp. 41–55.

Su, Yu-Sung et al. (2011). "Multiple imputation with diagnostics (mi) in R: Opening windows into the black box". In: *Journal of Statistical Software* 45(2), pp. 1–31.

Williamson, Elizabeth J, Andrew Forbes, and Ian R White (2014). "Variance reduction in randomised trials by inverse probability weighting using the propensity score". In: *Statistics in medicine* 33(5), pp. 721–737.