Model-Based Approach to Causal Inference

Susanna Makela

March 2, 2017

This is a summary of what we talked about in class on March 1st and is based on Chapter 8 of Imbens and Rubin (2015) and Chapter 8 of Gelman et al. (2013).

Suppose we are in the context of randomized trial with a binary treatment. In that case, the vector of treatment assignments $\mathbf{W}$ is independent of the vectors of potential outcomes $(\mathbf{Y}(0), \mathbf{Y}(1))$. Our goal is to estimate a (super-) population treatment effect

$$\mathbb{E}\left[\tau(\mathbf{Y}(0), \mathbf{Y}(1)) \mid \mathcal{D}\right], \tag{1}$$

where $\mathcal{D}$ denotes the observed data $\mathcal{D} = (\mathbf{Y}_{obs}, \mathbf{W})$ for some sample of size n from the super-population so that

$$\mathbf{Y}_{obs} = (Y_{1,obs}, \ldots, Y_{n,obs}) \quad \text{and} \quad Y_{i,obs} = W_i Y_i(1) + (1 - W_i) Y_i(0). \tag{2}$$

One common choice for $\tau(\mathbf{Y}(0), \mathbf{Y}(1))$ is simply the difference

$$\tau(\mathbf{Y}(0), \mathbf{Y}(1)) = Y_i(1) - Y_i(0),$$

so that we can write the treatment effect as

$$\mathbb{E}\left[Y_i(1) - Y_i(0) \mid \mathcal{D}\right].$$

Let's expand our estimand (1) using iterated expectations:

$$\begin{aligned}
&\mathbb{E}[\tau(\mathbf{Y}(0), \mathbf{Y}(1)) \mid \mathcal{D}] \\
&\quad = \mathbb{E}[\tau(\mathbf{Y}(0), \mathbf{Y}(1)) \mid \mathbf{Y}_{obs}, \mathbf{W}] \\
&\quad = \mathbb{E}[\tilde{\tau}(\mathbf{Y}_{obs}, \mathbf{Y}_{mis}) \mid \mathbf{Y}_{obs}, \mathbf{W}]
\end{aligned}$$

where we've now rewritten the original treatment effect $\tau(\mathbf{Y}(0), \mathbf{Y}(1))$ in terms of the observed and missing data $\tilde{\tau}(\mathbf{Y}_{obs}, \mathbf{Y}_{mis})$, which we can do because of how we defined $\mathbf{Y}_{obs}$ in (2). This final expectation is the posterior expectation of the treatment effect. We therefore need the posterior distribution of $\tilde{\tau}$, which is the conditional distribution of $\tilde{\tau}$ given the observed data $(\mathbf{Y}_{obs}, \mathbf{W})$:

$$p(\tilde{\tau} \mid \mathbf{Y}_{obs}, \mathbf{W}). \tag{3}$$

How do we compute this quantity? Note that we can write (3) as a marginal distribution where we've integrated out $\mathbf{Y}_{mis}$:

$$\begin{aligned}
p(\tilde{\tau} \mid \mathbf{Y}_{obs}, \mathbf{W}) &= \int p(\tilde{\tau}, \mathbf{y}_{mis} \mid \mathbf{Y}_{obs}, \mathbf{W}) \, \mathrm{d}\mathbf{y}_{mis} \\
&= \int p(\tilde{\tau} \mid \mathbf{y}_{mis}, \mathbf{Y}_{obs}, \mathbf{W}) \, p(\mathbf{y}_{mis} \mid \mathbf{Y}_{obs}, \mathbf{W}) \, \mathrm{d}\mathbf{y}_{mis},
\end{aligned}$$

The first term is simple to understand: given values of $\mathbf{Y}_{obs}$ and $\mathbf{W}$ and draws of $\mathbf{Y}_{mis}$, we can calculate $\tilde{\tau}$, thereby generating a draw from $p(\tilde{\tau}|\mathbf{Y}_{mis}, \mathbf{Y}_{obs}, \mathbf{W})$. The second term in the integral is the posterior predictive distribution for $\mathbf{Y}_{mis}$. To obtain draws from this distribution, we first draw a value of θ from its posterior distribution given the observed data $(\mathbf{Y}_{obs}, \mathbf{W})$ and then draw a value of $\mathbf{Y}_{mis}$ from its posterior distribution given the observed data and θ. This relationship becomes clear when we write the second term, $p(\mathbf{y}_{mis} \mid \mathbf{Y}_{obs}, \mathbf{W})$, as a marginal distribution over θ:

$$\begin{aligned} p(\mathbf{y}_{mis} \mid \mathbf{Y}_{obs}, \mathbf{W}) &= \int p(\mathbf{y}_{mis}, \theta \mid \mathbf{Y}_{obs}, \mathbf{W}) \, \mathrm{d}\theta \\ &= \int p(\mathbf{y}_{mis} \mid \theta, \mathbf{Y}_{obs}, \mathbf{W}) p(\theta \mid \mathbf{Y}_{obs}, \mathbf{W}) \, \mathrm{d}\theta. \end{aligned}$$

The question now becomes how to obtain $p(\theta \mid \mathbf{Y}_{obs}, \mathbf{W})$, the posterior distribution of θ given the observed data $(\mathbf{Y}_{obs}, \mathbf{W})$. To calculate this, we can follow the ideas in Chapter 8 (specifically, pp 200-203 of Gelman et al. (2013)). Note that Gelman et al. (2013) uses the parameter ϕ, which governs the distribution of the inclusion vector $\mathbf{W}$, but since we are assuming a randomized trial, $p(\mathbf{W}) = \alpha$ for some constant α (generally 0.5). The posterior distribution of θ is then

$$\begin{aligned} p(\theta \mid \mathbf{Y}_{obs}, \mathbf{W}) &\propto p(\theta) p(\mathbf{Y}_{obs}, \mathbf{W} \mid \theta) \\ &= p(\theta) \int p(\mathbf{y}_{mis}, \mathbf{Y}_{obs}, \mathbf{W} \mid \theta) \, \mathrm{d}y_{mis} \\ &\propto p(\theta) \int p(\mathbf{y}_{mis}, \mathbf{Y}_{obs} \mid \theta) \, \mathrm{d}y_{mis} \\ &= p(\theta) p(\mathbf{Y}_{obs} \mid \theta) \end{aligned}$$

where the last line follows from the fact that $\mathbf{W}$ is independent of $(\mathbf{Y}(0), \mathbf{Y}(1))$ and thus of $(\mathbf{Y}_{mis}, \mathbf{Y}_{obs})$, and from folding the constant $p(\mathbf{W})$ into the denominator. The term $p(\mathbf{y}_{mis}, \mathbf{Y}_{obs} \mid \theta)$ is simply the complete-data likelihood, which we as the analyst specify.

In practice, we would do the following in Stan. First,

References

A. Gelman, J. Carlin, H. Stern, D. Dunson, A. Vehtari, and D. Rubin. *Bayesian Data Analysis, Third Edition (Chapman & Hall/CRC Texts in Statistical Science).* Chapman and Hall/CRC, London, third edition, Nov. 2013. ISBN 1439840954.

G. W. Imbens and D. B. Rubin. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction.* Cambridge University Press, New York, NY, USA, 2015. ISBN 0521885884, 9780521885881.