MARGINAL STRUCTURAL MODELS

Marginal structural models are based on an extension of the weighting approach.

"Marginal" because they model the marginal means $\mathrm{E}(Y_t(\bar{z}_t))$ (or $E(Y_t(\bar{z}_t) \mid W)$), where W is a set of baseline covariates.

"Structural" because they model potential outcomes e.g.

$$E(Y_t(\bar{z}_t)) = \gamma_0 + \gamma_1 \sum_{\ell=1}^{t-1} 1(Z_\ell = 1) + \gamma_2 1(Z_t = 1).$$

MARGINAL STRUCTURAL MODELS

Recall when treatment assignment Z is strongly ignorable given covariates $\mathbf{X}_1$:

$$E\left(\frac{(ZY)}{\Pr(Z = 1 \mid \mathbf{X}_1)}\right) = E(Y(1))$$

$$E\left(\frac{(1 - Z)Y}{\Pr(Z = 0 \mid \mathbf{X}_1)}\right) = E(Y(0))$$

The weighting approach exploits this by adjusting the distribution of the covariates to be the same among treated and untreated units.

MARGINAL STRUCTURAL MODELS

Consider a subject in the treatment group with covariates $\mathbf{X}_1$ and probability $\Pr(Z = 1 \mid \mathbf{X}_1) = .1$. That subject represents $1/10$ of the subjects in the population with covariates $\mathbf{X}_1$. To make that subject representative of the population of units with all covariates $\mathbf{X}_1$, it is necessary to endow that subject with a weight of 10.

Similarly, units with the same values of the covariates $\mathbf{X}_1$ in the control group have probability .9, and therefore receive a weight of $10/9$.

Weighting creates a pseudo-population in which both treated and control subjects have the same distribution of covariates. Averaging the weighted outcomes in the pseudo-population treatment group then gives $E(Y(1))$, and averaging the weighted outcomes in the control group gives $E(Y(0))$.

MARGINAL STRUCTURAL MODELS

In practice we do not observe the population. In an observational study, the assignment probabilities are not known and must be estimated.

For a continuous outcome Y, estimates of $E(Y(1))$ and $E(Y(0))$ can be computed using weighted least squares regression: $Y = \alpha_0 + \alpha_1 Z + \varepsilon$, where units with $Z = 1$ are weighted inversely to $\hat{\Pr}(Z = 1 \mid \mathbf{X}_1)$ and units with $Z = 0$ inversely to $1 - \hat{\Pr}(Z = 1 \mid \mathbf{X}_1)$. To take into account the fact that the weights were estimated, robust standard errors should be used.

For a dichotomous outcome, logistic regression or probit regression might be used instead.

MARGINAL STRUCTURAL MODELS

We have used so called unstabilized weights. Had we used stabilized weights $\frac{\Pr(Z=1)}{\Pr(Z=1|\mathbf{X}_1)}$ and $\frac{\Pr(Z=0)}{\Pr(Z=0|\mathbf{X}_1)}$, the same results would be obtained since the model for Y is saturated.

It is usually necessary to model the outcomes using unsaturated models. Stabilized weights then lead to smaller confidence intervals for estimated parameters.

The results above extend to the longitudinal case under the ignorability and positivity conditions:

$$E(Y_t(\bar{z}_t)) = E(\prod_{\ell=1}^{t} \frac{1(Z_\ell = z_\ell)}{\Pr(Z_\ell = z_\ell \mid \bar{Z}_{\ell-1} = \bar{z}_{\ell-1}, \bar{\mathbf{X}}_\ell)} Y_t),$$

where $1(Z_\ell = z_\ell) = 1$ if $Z_\ell = z_\ell$, 0 otherwise.

MARGINAL STRUCTURAL MODELS

Robins and Hernan (2009) call the weights $\prod_{\ell=1}^{t}(\Pr(Z_\ell = z_\ell \mid \bar{Z}_{\ell-1} = \bar{z}_{\ell-1}, \bar{\mathbf{X}}_\ell))^{-1}$ unstabilized. The stabilized weights are $\prod_{\ell=1}^{t} \frac{\Pr(Z_\ell = z_\ell \mid \bar{Z}_{\ell-1} = \bar{z}_{\ell-1})}{\Pr(Z_\ell = z_\ell \mid \bar{Z}_{\ell-1} = \bar{z}_{\ell-1}, \bar{\mathbf{X}}_\ell)}$.

In order to estimate $E(Y_1(z_1))$, weight inversely by an estimate of the stabilized or unstabilized weight.

At the next step, in order to estimate $E(Y_2(z_1, z_2))$, the units observed in period 2 with $Z_1 = z_1$ and $Z_2 = z_2$ must be reweighted so that observations with covariates $\mathbf{X}_1$ and $\mathbf{X}_2$ have the same frequency as observations with the same values of the covariates, but with $Z_1 = z_1^*, Z_2 = z_2^*$, etc.

MARGINAL STRUCTURAL MODELS

As before, it is necessary to estimate the weights and, as t increases, it may be necessary to model the outcome as well. e.g.

$$E(Y_t) = \alpha + \beta_1 \sum_{\ell=1}^{t-1} 1(Z_\ell = 1) + \beta_2 1(Z_{t-1} = 1).$$

Concerns about misspecification for the case $T = 1$ apply here as well, but more forcefully, as the opportunities for misspecification increase.

Baseline covariates $\mathbf{W} \subset \mathbf{X_1}$ can also be included in the marginal structural model. Hernan and Robins (2018) recommend using modified stabilized weights $\prod_{\ell=1}^{t} \frac{\Pr(Z_\ell = z_\ell | \bar{Z}_{\ell-1} = \bar{z}_{\ell-1}, \mathbf{W})}{\Pr(Z_\ell = z_\ell | \bar{Z}_{\ell-1} = \bar{z}_{\ell-1}, \bar{\mathbf{X}}_\ell)}$. These lead to smaller standard errors.

MARGINAL STRUCTURAL MODELS

Both the g-formula and IPTW can be used to estimate $E(Y_t(\bar{z}_t) - Y_t(\bar{z}_t^*))$. It is useful to look at the advantages and disadvantages associated with each approach.

Using the g-formula requires modeling $E(Y_t \mid \bar{z}_t, \bar{\mathbf{x}}_t)$ for all $\bar{\mathbf{x}}_t$ realized under the sequence $\mathbf{z}_{t-1}$ and the probability functions $f(\mathbf{x}_\ell \mid \bar{\mathbf{x}}_{\ell-1}, \bar{z}_{\ell-1})$ for $\ell = 1, ..., t$.

Using IPTW requires modeling the assignment probabilities, conditional on past covariates and assignments. Although it appears that IPTW does not require modeling outcomes, remember that as t increases, the number of treatment regimens increases exponentially, and it will often be necessary to parsimoniously model the outcomes.

MARGINAL STRUCTURAL MODELS

In a sequentially randomized experiment, the treatment probabilities are known. Using IPTW, it is only necessary to model the outcomes. IPTW would thus be preferred since using the g-formula would also require modeling the distribution of the time varying confounders.

In an observational study, however, the assignment probabilities are unknown and must be modeled. In this case, misspecification of the model for the assignment process can lead to very biased estimates, as can misspecification of the outcome model. Bang and Robins (2005) present a doubly robust estimator.

In an observational study, a good strategy might be to use both approaches and see if the estimates obtained agree. While such agreement does not actually indicate that either or both approaches have led to a good estimate, the agreement does offer some grounds for reassurance. However, the analysis is predicated on the sequential randomization assumption.