G Estimation of Structural Nested Mean Models

Ashley I. Naimi, PhD

Outline

- Structural Nested Failure Time Models
- Structural Nested Mean Models
- G Estimation

Some Preliminaries

Structural Nested Failure Time Models

Structural nested models map a *conditional contrast* of potential outcomes to the treatment, within nested sub-groups of individuals defined by levels of A_1, Z_1, and A_0. To date, there are generally two types of structural nested models that have been used in the literature. Structural nested failure time models are used in the context of survival outcomes, and are usually written as:

$$T^{\overline{0}} = \int_0^T exp(\psi A_u)\mathrm{d}u$$

where $T^{\overline{0}}$ is the survival time that would have been observed under no exposure. In this case, $exp(-\psi)$ quantifies the *blip effect* of being exposed at any given point (say, time j) during follow-up, and then unexposed thereafter, relative to being unexposed at time j and thereafter. This blip effect is the fundamental unit quantified with structural nested models, but it is subtle so we will re-visit it later.

Structural nested failure time models are (as far as I can tell) the most commonly used structural nested models, but they are subject to an important and fundamental limitation: the need for artificial censoring (Joffe, Yang, and Feldman 2012). In simple terms, to get an unbiased estimate of ψ in the above SNM, one needs to remove a certain number of observed events from the sample. Removing these events, however, precludes the use of existing optimization algorithms that easily enable us to find estimates of ψ in a given dataset.[1]

Without commonly used optimization algorithms, researchers are forced to rely on a grid search method to solve for ψ, in which a range of potential ψ values are tested as potential estimates. However, using the grid search method forces researchers (for computational reasons) to specify relatively simple structural nested models, which effectively defeats the purpose of their use.

Structural nested models are *conditional* models (see chapter 2 of the handouts), in that we should have a parameter quantifying the exposure effect for each possible level of all relevant baseline and time-varying covariates. Unfortunately, in most applied settings, this

[1] For example, the Newton-Raphson algorithm is a commonly used optimization method that is employed any time one runs a generalized linear model, such as logistic or Poisson regression.

would lead to structural nested models with many parameters, which we would not be able to quantify because the grid-search method is too computationally demanding.

For this reason, structural nested failure time models are very difficult to use properly. The same is not true for structural nested mean models, which we will introduce next.

Structural Nested Mean Models

We are interested here in estimating the effect of treatment given at two time points on an outcome measured at the end of follow-up in a hypothetical study. Our example will follow the canonical HIV epidemiology study, where g-methods were first developed.[2] Treatment is measured at baseline (A_0) and once during follow up (A_1). The sole covariate is elevated HIV viral load ($Z = 1$ for those with > 200 copies/ml, $Z = 0$ otherwise), which is constant by design at baseline ($Z_0 = 1$) and measured once during follow up just prior to the second treatment (Z_1). The outcome is CD4 count measured at the end of follow up in units of cells/mm^3.

[2] We introduce the example here in a very quick and cursory way. More details will be provided in the sections on g computation (or g formula) and IP weighting.

The following figure depicts the data generating mechanism:

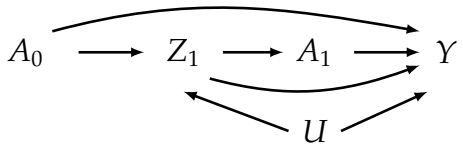

Figure 1: Causal diagram representing the relation between anti-retroviral treatment at time 0 (A_0), HIV viral load just prior to the second round of treatment (Z_1), anti-retroviral treatment status at time 1 (A_1), the CD4 count measured at the end of follow-up (Y), and an unmeasured common cause (U) of HIV viral load and CD4.

In our example, our structural nested model can be written as

$$E(Y^{a_0,a_1} - Y^{a_0,0} \mid A_0 = a_0, Z_1 = z_1, A_1 = a_1) = a_1(\psi_1 + \psi_2 a_0 + \psi_3 z_1 + \psi_4 a_0 z_1)$$
$$E(Y^{a_0,0} - Y^{0,0} \mid A_0 = a_0) = \psi_0 a_0$$

Note this model includes parameters for every possible level of the covariates: ψ_1 for the effect of a_1 among those with $a_0 = 0$ and $z_1 = 0$; ψ_2 for the effect of a_1 among those with $a_0 = 0$ and $z_1 = 0$; ψ_3 for the effect of a_1 among those with $a_0 = 0$ and $z_1 = 1$, and ψ_4 for the effect of a_1 among those with $z_1 = 1$ and $a_0 = 1$.

Becuase the SNM quantifies a parameter for each possible level of the past time-varying covariates, it is a conditional model. We will note later that IP weighted MSMs cannot be used to estimate such parameters. Therefore, if one is interested in quantifying such interactions, one may need to use SNMs[3]. Indeed, the ability to explicitly quantify interactions between time-varying exposures and time-varying covariates is a major strength of structural nested models when effect modification is of interest (Robins and Hernán 2009).

[3] The g formula may also be used, though at the expense of heavy parametric modeling assumptions.

To simplify our exposition, we set $(\psi_3, \psi_4) = (0, 0)$ in our data example, allowing us to drop the $\psi_3 z_1$ and $\psi_4 a_0 z_1$ terms from the model. In effect, this renders our structural nested mean model equivalent to a semi-parametric marginal structural model. (In the Supplementary Material of Ashley I Naimi, Cole, and Kennedy (2016), we explain how marginal structural and structural nested models each relate to time-varying interactions in more detail.)

Intuitively, the structural nested model yields a comparison of counterfactual outcomes as depicted in the following Figure:

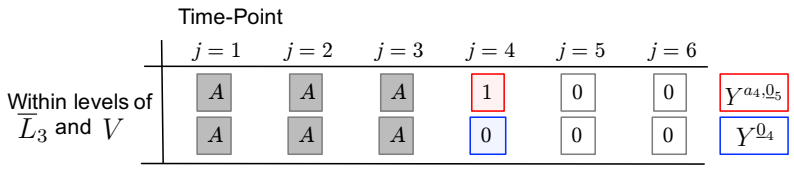

Figure 2: Diagram depicting the causal contrast defining our effect of interest in a given structural nested model.

We can now use g-estimation to estimate (ψ_0, ψ_1, ψ_2) in the above structural nested model. G-estimation is based on solving equations that directly result from conditional exchangeability (applied sequentially), combined with assumptions implied by the structural nested model. If, at each time point, the exposure is conditionally independent of the potential outcomes (sequential exchangeability) then the conditional covariance between the exposure and potential outcomes is zero (Vansteelandt and Joffe 2014) Formally, these conditional independence relations can be written as:

$$\begin{aligned} 0 &= \text{Cov}(Y^{a_0,0}, A_1 \mid Z_1, A_0) \\ &= \text{Cov}(Y^{0,0}, A_0) \end{aligned}$$

where $\text{Cov}(\cdot)$ is the well-known covariance formula (Wasserman 2006). These equalities are of little direct use for estimation, though, as they contain unobserved potential outcomes and are not yet functions of the parameters of interest. However, by counterfactual consistency and the structural nested model, we can replace these unknowns with quantities estimable from the data.

Specifically, the structural nested model, together with exchangeability and counterfactual consistency imply that we can replace the potential outcomes $Y^{a_0,0}$ and $Y^{0,0}$ in the above covariance formulas with their values implied by the structural nested model, yielding:

$$\begin{aligned} 0 &= \text{Cov}\{Y - A_1(\psi_1 + \psi_2 A_0), A_1 \mid Z_1, A_0\} \\ &= \text{Cov}\{Y - A_1(\psi_1 + \psi_2 A_0) - \psi_0 A_0, A_0\}. \end{aligned}$$

We provide an intuitive explanation for this substitution in the Supplementary Material of A I Naimi (2016). We also show how these covariance relations yield three equations that can be used to solve each of the unknowns in the above structural nested model (ψ_0, ψ_1, ψ_2).

Two of the three equations yield the following g estimators:

$$\begin{aligned} \hat{\psi}_{1_{GE}} &= \frac{\hat{E}[(1-A_0)Y\{A_1 - \hat{E}(A_1 \mid Z_1, A_0)\}]}{\hat{E}[(1-A_0)A_1\{A_1 - \hat{E}(A_1 \mid Z_1, A_0)\}]} \\ \hat{\psi}_{1_{GE}} + \hat{\psi}_{2_{GE}} &= \frac{\hat{E}[A_0 Y\{A_1 - \hat{E}(A_1 \mid Z_1, A_0)\}]}{\hat{E}[A_0 A_1\{A_1 - \hat{E}(A_1 \mid Z_1, A_0)\}]} \end{aligned}$$

Note that to solve these equations we need to model $E(A_1 \mid Z_1, A_0)$, which in practice we might assume can be correctly specified as the predicted values from a logistic model for A_1. In our simple setting, the correctness of this model is guaranteed by saturating it (i.e., conditioning the model on Z_1, A_0 and their interaction).

As we show in the Supplementary Material, implementing these equations in software can be easily done using either an instrumental variables (i.e., two-stage least squares) estimator, or ordinary least squares.

Once the above parameters are estimated, the next step is to subtract the effect of A_1 and $A_1 A_0$ from Y to obtain $\widetilde{Y} = Y - \hat{\psi}_{1_{GE}} A_1 - \hat{\psi}_{2_{GE}} A_1 A_0$. We can then solve for the last parameter using a sample

version of the third g estimation equality, yielding our final estimator and completing the procedure:

$$\hat{\psi}_{0_{GE}} = \frac{\hat{E}[\widetilde{Y}\{A_0 - \hat{E}(A_0)\}]}{\hat{E}[A_0\{A_0 - \hat{E}(A_0)\}]}.$$

Again the above estimator can be implemented using an instrumental variable or ordinary least squares estimator. Here, we demonstrate how to do this using the least squares approach:

```
# arrange into wide data
a0 <- c(0, 0, 0, 0, 1, 1, 1, 1)
z1 <- c(0, 0, 1, 1, 0, 0, 1, 1)
a1 <- c(0, 1, 0, 1, 0, 1, 0, 1)
y <- c(87.29, 112.11, 119.65, 144.84, 105.28, 130.18, 137.72, 162.83)
N <- c(209271, 93779, 60654, 136293, 134781, 60789, 93903, 210530)
D <- NULL
for (i in 1:8) {
    d <- data.frame(cbind(rep(a0[i], N[i]), rep(z1[i], N[i]), rep(a1[i], N[i]), rep(y[i], N[i])))
    D <- rbind(D, d)
}
nrow(D)
```

```
## [1] 1000000
```

```
names(D) <- c("a0", "z1", "a1", "y")

# model the exposure at the last time point and create residuals
D$rA1 <- D$a1 - glm(a1 ~ z1 + a0 + z1:a0, data = D, family = binomial("logit"))$fitted.values
D$rA1a0 <- D$rA1 * D$a0
stage1 <- coef(lm(y ~ -1 + rA1 + rA1a0, data = D))
# the effect of the second expsoure is:
stage1[1]
```

```
##      rA1
## 24.96551
```

```
# transform the outcome
D$y_tilde <- D$y - stage1[1] * D$a1 - stage1[2] * D$a1 * D$a0
```

```r
# model the exposure at the first time point and create residuals
D$rA0 <- D$a0 - glm(a0 ~ 1, data = D, family = binomial("logit"))$fitted.values
# estimate the second exposure effect
stage2 <- coef(lm(y_tilde ~ -1 + rA0, data = D))
# the effect of the first expsoure is:
stage2
```

```
##      rA0
## 24.98026
```

Implementing this procedure in our example data, we obtain $[\psi_{0_{GE}} = 25.0, \psi_{1_{GE}} = 25.0, \psi_{2_{GE}} = 0.0]$, thus yielding $\psi_{GE} = 50.0$.

The potential outcome under no treatment can be thought of as a given subject's baseline prognosis: in our setting, individuals with poor baseline prognosis will have low CD_4 levels, no matter what their treatment status may be. In the absence of confounding or selection bias, one expects this baseline prognosis to be independent of treatment status. G estimation exploits this independence by assuming no uncontrolled confounding (conditional on measured confounders), and assigning values to $\hat{\psi}_{GE}$ that render the potential outcomes independent of the exposure. However, assigning the correct values to $\hat{\psi}_{GE}$ depends on there being no confounding or selection bias.

References

Joffe, Marshall M., Wei Peter Yang, and Harold Feldman. 2012. "G-Estimation and Artificial Censoring: Problems, Challenges, and Applications." *Biometrics* 68 (1): 275–86.

Naimi, A I. 2016. "The Counterfactual Implications of Fundamental Cause Theory." *Curr Epidemiol Reports* In Press.

Naimi, Ashley I, Stephen R Cole, and Edward H Kennedy. 2016. "An Introduction to G Methods." *Int J Epidemiol* In Press.

Robins, James M, and Miguel Á Hernán. 2009. "Estimation of the Causal Effects of Time-Varying Exposures." In *Advances in Longitudinal Data Analysis*, edited by G Fitzmaurice, M. Davidian, G Verbeke, and G Molenberghs, 553–99. Boca Raton, FL: Chapman & Hall.

Vansteelandt, Stijn, and Marshall Joffe. 2014. "Structural Nested Models and G-Estimation: The Partially Realized Promise." *Statist. Sci.* 29 (4): 707–31.

Wasserman, Larry. 2006. *All of Nonparametric Statistics*. New York; London: Springer.