

SEMIPARAMETRIC INFERENCE AND G-ESTIMATION

Consider the simple structural (causal) specification given by

$$\mathbb{E}[Y(z) - Y(0)|X] \equiv \mathbb{E}[Y(z) - Y(0)] = z\psi_0.$$

This states that, compared to the ‘baseline’ case of $z = 0$, exposure at level z leads to a change $z\psi_0$ in the expected (potential) outcome, and that this quantity does not depend on confounders X .

The model can be considered as one that specifies that

$$Y(z) = \mu_0(X) + z\psi_0 + \epsilon$$

where $\mu_0(X) = \mathbb{E}[Y(0)]$.

If $\mu_0(X) = \mu_0(X; \beta_0)$ is proposed as a parametric submodel for this nuisance component, we consider the probability model based on the semiparametric regression formulation

$$f_{X,Y,Z}(x, y, z) = f_\epsilon(y - \mu_0(x; \beta) - z\psi)f_{X,Z}(x, z)$$

and focus on the semiparametric submodel for the $f_\epsilon(\cdot)$ component

$$f_\epsilon(\epsilon) \equiv f_\epsilon(\epsilon; \sigma_0) \equiv \text{Normal}(0, \sigma_0^2).$$

Then the score function for this submodel is

$$\begin{pmatrix} S_\beta \\ S_\psi \end{pmatrix} = \frac{1}{\sigma_0^2} \begin{pmatrix} \mu_{0\beta}(X; \beta_0) \\ Z \end{pmatrix} (Y - \mu_0(X; \beta_0) - Z\psi_0) = \frac{1}{\sigma_0^2} \begin{pmatrix} \mu_{0\beta}(X; \beta_0) \\ Z \end{pmatrix} \epsilon$$

say.

The nuisance tangent space corresponds to the first component, and can be seen to take the form

$$\{h(X)\epsilon : h(X) \text{ arbitrary}\}$$

where $h(X)$ has the same dimension as β . We then need to project S_ψ onto this tangent space: the projection is the quantity $h_0(X)\epsilon$, where we must have

$$\mathbb{E}[(Z\epsilon - h_0(X)\epsilon)^\top h(X)\epsilon] = 0 \quad \forall h(X)$$

By iterated expectation, we therefore must have

$$\mathbb{E}[(e(X)\epsilon - h_0(X)\epsilon)^\top h(X)\epsilon] = 0 \quad \forall h(X)$$

implying that

$$h_0(X) \equiv e(X)$$