STAT 460/560 Class 23: More on Nuisance Parameters and Orthogonality

Ben Bloem-Reddy

Today we'll look at a specific example of the general result we established last class. Before starting on that, let's recall a very useful property of conditional expectation. Suppose X and Y are $\mathbb{R}$-valued random variables, each with finite expectation. Recall that for any measurable function $x \mapsto f(x) \in \mathbb{R}$,

$$E[f(X)(Y - E[Y|X])] = E[f(X)E[(Y - E[Y|X])|X]] = 0 \ . \tag{23.1}$$

Heuristically, one can think of $E[Y|X]$ as the projection of Y onto the space of functions of X, and $Y - E[Y|X]$ as orthogonal to the functions of X. (This can be made precise using Hilbert space projections. See, for example, [Çin11].)

Hence, we can always write

$$Y = m(X) + \epsilon \ , \quad E[\epsilon|X] = 0 \ , \tag{23.2}$$

for some function m that we interpret as $m(X) = E[Y|X]$.

1. Partially linear regression model

Suppose that we're interested in estimating the effect of a treatment, X, on a health outcome, Y, but that we need to control for possible confounding with variables W. Moreover, assume that the data obeys the partially linear model (PLM),

$$Y = \theta_0 X + g_{\eta_0}(W) + \epsilon \ , \quad E[\epsilon|X, W] = 0 \ . \tag{23.3}$$

Activity 23.1. Show that

$$\psi_{\theta,\eta}(w, x, y) := \begin{pmatrix} x(y - \theta x - g_\eta(w)) \\ w(y - \theta x - g_\eta(w)) \end{pmatrix}$$

defines a Z-estimator, i.e., show that $P\psi_{\theta_0,\eta_0} = (0, 0)^\top$.

Solution: Since we're assuming the PLM, we see that the first component is

$$[P\psi_{\theta_0,\eta_0}]_1 = E[X(\theta_0 X + g_{\eta_0}(W) + \epsilon - \theta_0 X - g_{\eta_0}(W))] = E[X\epsilon] = 0 \ .$$

Similarly, the second component is easily shown to be $E[W\epsilon] = 0$.

In this situation, we don't really care about the value of η; we just want to estimate θ well. One can show that

$$P(\psi_{\theta_0,\eta_0}\psi_{\theta_0,\eta_0}^\top) = \begin{pmatrix} E[\epsilon^2 X^2] & E[\epsilon^2 XW] \\ E[\epsilon^2 XW] & E[\epsilon^2 W^2] \end{pmatrix} \ , \tag{23.4}$$

and assuming that at each w, the function $\eta \mapsto g_\eta(w)$ is differentiable (with derivative $\dot{g}_\eta(w)$) in a neighborhood of η_0,

$$V_{\theta_0,\eta_0} = - \begin{pmatrix} E[X^2] & E[\dot{g}_{\eta_0}(W)X] \\ E[XW] & E[\dot{g}_{\eta_0}(W)W] \end{pmatrix} \ ,$$

so that

$$-V_{\theta_0,\eta_0}^{-1} = \frac{1}{E[X^2]E[\dot{g}_{\eta_0}(W)W] - E[XW]E[\dot{g}_{\eta_0}(W)X]} \begin{pmatrix} E[\dot{g}_{\eta_0}(W)W] & -E[\dot{g}_{\eta_0}(W)X] \\ -E[XW] & E[X^2] \end{pmatrix} .$$

Assuming further that $E[\epsilon^2|X,W] = \sigma^2$, the marginal asymptotic variance of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ is

$$\sigma_{\theta_0}^2(\eta_0) = \sigma^2 \frac{E[X^2]E[\dot{g}_{\eta_0}(W)W]^2 + E[W^2]E[\dot{g}_{\eta_0}(W)X]^2 - 2E[XW]E[\dot{g}_{\eta_0}(W)W]E[\dot{g}_{\eta_0}(W)X]}{(E[X^2]E[\dot{g}_{\eta_0}(W)W] - E[XW]E[\dot{g}_{\eta_0}(W)X])^2} \tag{23.5}$$

$$= \sigma^2 \frac{E[\dot{g}_{\eta_0}(W)W]}{E[X^2]E[\dot{g}_{\eta_0}(W)W] - E[XW]E[\dot{g}_{\eta_0}(W)X]} + \sigma^2 \frac{E[\dot{g}_{\eta_0}(W)X](E[W^2]E[\dot{g}_{\eta_0}(W)X] - E[XW]E[\dot{g}_{\eta_0}(W)W])}{(E[X^2]E[\dot{g}_{\eta_0}(W)W] - E[XW]E[\dot{g}_{\eta_0}(W)X])^2} \tag{23.6}$$

Exercise 23.1. Derive the expressions above for $P(\psi_{\theta_0,\eta_0}\psi_{\theta_0,\eta_0}^\top)$, V_{θ_0,η_0}, and V_{θ_0,η_0}^{-1}. Use them to obtain $\sigma_{\theta_0}^2$.

Exercise 23.2. Show that if $E[X|W] = \beta_0 W$ and $g_\eta(W) = \eta W$ then

$$\sigma_{\theta_0}^2(\eta_0) = \frac{\sigma^2}{E[X^2] - \beta_0^2 E[W^2]} .$$

2. Orthogonal estimation of the PLM

All of the above is rather involved, and we get a complicated expression for the asymptotic variance $\sigma_{\theta_0}^2$. We also pay a price for estimating η_0, though it's not clear how much without more explicit assumptions.

Instead, let's estimate the conditional expectations $h_{\gamma_0}(W) = E[Y|W]$ and $m_{\beta_0}(W) = E[X|W]$, and consider the function

$$\psi_{\theta,\gamma,\beta}(w,x,y) = (y - h_\gamma(w) - \theta(x - m_\beta(w)))(x - m_\beta(w)) . \tag{23.7}$$

Where does this come from? Observe that in the PLM,

$$\begin{aligned} E[Y|W] = h_{\gamma_0}(W) &= \theta_0 E[X|W] + E[g_{\eta_0}(W)|W] + E[\epsilon|W] \\ &= \theta_0 m_{\beta_0}(W) + g_{\eta_0}(W) + 0 , \end{aligned}$$

so that

$$Y - h_{\gamma_0}(W) = \theta_0(X - m_{\beta_0}(W)) + \epsilon . \tag{23.8}$$

Rewriting as "residuals," $\tilde{Y} = Y - h_{\gamma_0}(W)$ and $\tilde{X} = X - m_{\beta_0}(W)$, we have

$$\tilde{Y} = \theta_0 \tilde{X} + \epsilon ,$$

which can be estimated using the function $\tilde{\psi}_\theta(\tilde{x},\tilde{y}) = \tilde{x}(\tilde{y} - \theta\tilde{x})$. This is equivalent to (23.7).

Back to (23.7), we see that

$$P\psi_{\theta,\gamma,\beta} = E[((\theta_0 - \theta)X + g_{\eta_0}(W) - h_\gamma(W) + \theta m_\beta(W))(X - m_\beta(W))] \tag{23.9}$$

Activity 23.2. Show that $P\psi_{\theta_0,\gamma_0,\beta_0} = 0$.

Solution: Using the expression above,

$$P\psi_{\theta_0,\gamma_0,\beta_0} = E[(g_{\eta_0}(W) - h_{\gamma_0}(W) + \theta_0 m_{\beta_0}(W))(X - m_{\beta_0}(W))] \ .$$

Since the terms in the first parentheses are all functions of W, and $m_{\beta_0}(W) = E[X|W]$, this is equal to zero by (23.1).

Activity 23.3. Assume that $\beta \mapsto m_\beta(w)$ has derivative $\dot{m}_\beta(w)$ for each w, and similarly for $\gamma \mapsto h_\gamma(w)$. Moreover, assume that we can interchange differentiation and expectation in (23.9).

Show that $\frac{\partial}{\partial\beta} P\psi_{\theta_0,\gamma_0,\beta_0} = \frac{\partial}{\partial\gamma} P\psi_{\theta_0,\gamma_0,\beta_0} = 0$.

Show that $\frac{\partial}{\partial\theta} P\psi_{\theta_0,\gamma_0,\beta_0} = -E[(X - m_{\beta_0}(W))^2]$.

Show that $P(\psi^2_{\theta_0,\gamma_0,\beta_0}) = E[\epsilon^2(X - m_{\beta_0}(W))^2]$.

Solution: We find that

$$\frac{\partial}{\partial\beta} P\psi_{\theta_0,\gamma_0,\beta_0} = E[\theta_0 \dot{m}_{\beta_0}(W)(X - m_{\beta_0}(W))] - E[(g_{\eta_0}(W) - h_{\gamma_0}(W) + \theta_0 m_{\beta_0}(W))\dot{m}_{\beta_0}(W)] \ .$$

The first term is zero by (23.7), but the second term appears problematic. However, by rearranging (23.8), we see that

$$\begin{aligned} Y &= h_{\gamma_0}(W) + \theta_0(X - m_{\beta_0}(W)) + \epsilon = \theta_0 X + h_{\gamma_0}(W) - \theta_0 m_{\beta_0}(W) + \epsilon \\ &= \theta_0 X + g_{\eta_0}(W) + \epsilon \ . \end{aligned}$$

Conditional expectations are unique up to null sets, so it must be that $g_{\eta_0}(W) = h_{\gamma_0}(W) - \theta_0 m_{\beta_0}(W)$ almost surely. Making this substitution in the derivative above, we find that

$$\frac{\partial}{\partial\beta} P\psi_{\theta_0,\gamma_0,\beta_0} = -E[(g_{\eta_0}(W) - h_{\gamma_0}(W) + \theta_0 m_{\beta_0}(W))\dot{m}_{\beta_0}(W)] = 0 \ .$$

Similarly,

$$\frac{\partial}{\partial\gamma} P\psi_{\theta_0,\gamma_0,\beta_0} = -E[\dot{h}_{\gamma_0}(W)(X - m_{\beta_0}(W))] = 0 \ .$$

Moreover,

$$\frac{\partial}{\partial\theta} P\psi_{\theta_0,\gamma_0,\beta_0} = -E[(X - m_{\beta_0}(W))^2] \ .$$

Finally, some algebra yields $P(\psi^2_{\theta_0,\gamma_0,\beta_0}) = E[\epsilon^2(X - m_{\beta_0}(W))^2]$.

Theorem 22.1 from last class then tells us that as long as $\sqrt{n}((\hat{\gamma}_n, \hat{\beta}_n) - (\gamma_0, \beta_0)) = O_P(1)$ then

$$\begin{aligned} \sqrt{n}(\hat{\theta}_n - \theta_0) &= \frac{1}{E[(X - m_{\beta_0}(W))^2]} \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_{\theta_0,\gamma_0,\beta_0}(W_i, X_i, Y_i) + o_P(1) \\ &\rightsquigarrow \mathcal{N}\left(0, \frac{E[\epsilon^2(X - m_{\beta_0}(W))^2]}{E[(X - m_{\beta_0}(W))^2]^2}\right) \ . \end{aligned}$$

If we also assume that $E[\epsilon^2|X,W] = \sigma^2$ then this becomes

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \rightsquigarrow \mathcal{N}\left(0, \frac{\sigma^2}{E[(X - m_{\beta_0}(W))^2]}\right) .$$

References

[Çin11] E. Çinlar. *Probability and Stochastics*. Springer New York, 2011.