Taylor Expansions and the Multivariate Delta Method

BST 257: Theory and Methods for Causality II
Alex Levis, Fall 2021

Consider again the IV functional problem we have been studying in lab: $O = (\boldsymbol{X}, Z, A, Y) \sim P$, and we assume that the conditional IV functional is constant,

$$\beta(P) = \frac{\text{Cov}_P(Y, Z \mid X)}{\text{Cov}_P(A, Z \mid X)} \equiv \frac{\mathbb{E}_P[\text{Cov}_P(Y, Z \mid X)]}{\mathbb{E}_P[\text{Cov}_P(A, Z \mid X)]}.$$

Let $\nu(P) = \mathbb{E}_P[(Y - \gamma_Y^T \boldsymbol{X})Z]$ and $\delta(P) = \mathbb{E}_P[(A - \gamma_A^T \boldsymbol{X})Z]$ for any P (we also are assuming $P[\text{Cov}_P(A, Z \mid X) > 0] = 1 \implies \delta(P) > 0$), where γ_A and γ_Y are the population least squares coefficients, for A and Y respectively. If we assume linear models for $\mathbb{E}_P(Y \mid \boldsymbol{X})$ and $\mathbb{E}_P(A \mid \boldsymbol{X})$, then we have $\beta(P) = \frac{\nu(P)}{\delta(P)}$.

We consider the outcome regression estimator $\widehat{\beta} = \frac{\mathbb{P}_n[(Y - \widehat{\gamma}_Y^T \boldsymbol{X})Z]}{\mathbb{P}_n[(A - \widehat{\gamma}_A^T \boldsymbol{X})Z]} = \frac{\widehat{\nu}}{\widehat{\delta}}$, where $\widehat{\gamma}_W$ are the sample least squares estimators, for A and Y respectively. Combining what we have learned so far:

$$\sqrt{n}\begin{pmatrix}\widehat{\nu} - \nu(P) \\ \widehat{\delta} - \delta(P)\end{pmatrix} = \sqrt{n}\mathbb{P}_n \begin{pmatrix}(Y - \gamma_Y^T \boldsymbol{X})(Z - \gamma_Z^T \boldsymbol{X}) - \nu(P) \\ (A - \gamma_A^T \boldsymbol{X})(Z - \gamma_Z^T \boldsymbol{X}) - \delta(P)\end{pmatrix} + \begin{pmatrix}o_P(1) \\ o_P(1)\end{pmatrix}.$$

How do we then characterize the limiting behavior of $\sqrt{n}(\widehat{\beta} - \beta(P)) = \sqrt{n}(\widehat{\nu}/\widehat{\delta} - \nu(P)/\delta(P))$? This is a classic example of where the multivariate delta method is useful. Let $g : \mathbb{R} \times (\mathbb{R} \setminus \{0\}) \to \mathbb{R}$ be given by $(u, v) \mapsto u/v$. Note that g is continuously differentiable on its domain with gradient

$$\nabla g(u, v) = \begin{bmatrix}\frac{1}{v} \\ -\frac{u}{v^2}\end{bmatrix} = v^{-1}\begin{bmatrix}1 \\ -\frac{u}{v}\end{bmatrix}.$$

by the mean value theorem, there exists some (ν^*, δ^*) on the line segment between $(\widehat{\nu}, \widehat{\delta})$ and $(\nu(P), \delta(P))$ such that

$$\begin{aligned}\sqrt{n}(\widehat{\beta} - \beta(P)) &= \sqrt{n}(g(\widehat{\nu}, \widehat{\delta}) - g(\nu(P), \delta(P))) \\ &= \sqrt{n}\nabla g(\nu^*, \delta^*)^T \begin{pmatrix}\widehat{\nu} - \nu(P) \\ \widehat{\delta} - \delta(P)\end{pmatrix}\end{aligned}$$

Since ∇g is continuous on its domain, the continuous mapping theorem yields

$$\nabla g(\nu^*, \delta^*) - \nabla g(\nu(P), \delta(P)) = \begin{pmatrix}o_P(1) \\ o_P(1)\end{pmatrix}.$$

It follows that

$$\sqrt{n}(\widehat{\beta} - \beta(P)) = \sqrt{n}\nabla g(\nu(P), \delta(P))^T \mathbb{P}_n \begin{pmatrix}(Y - \gamma_Y^T \boldsymbol{X})(Z - \gamma_Z^T \boldsymbol{X}) - \nu(P) \\ (A - \gamma_A^T \boldsymbol{X})(Z - \gamma_Z^T \boldsymbol{X}) - \delta(P)\end{pmatrix} + o_P(1).$$

Since $\nabla g(\nu(P), \delta(P)) = \delta(P)^{-1}\begin{pmatrix}1 \\ -\beta(P)\end{pmatrix}$, we obtain $\sqrt{n}(\widehat{\beta} - \beta(P)) = \sqrt{n}\mathbb{P}_n[\beta_P^1] + o_P(1)$, where

$$\beta_P^1(O) = \frac{1}{\delta(P)}\{(Y - \gamma_Y^T \boldsymbol{X})(Z - \gamma_Z^T \boldsymbol{X}) - \beta(P)(A - \gamma_A^T \boldsymbol{X})(Z - \gamma_Z^T \boldsymbol{X})\} + o_P(1).$$

In particular, $\sqrt{n}(\widehat{\beta} - \beta(P)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \text{Var}_P[\beta_P^1(O)])$, by Slutsky's and the central limit theorem!