General Strategy for Error Expansion

BST 257: Theory and Methods for Causality II

Alex Levis, Fall 2021

Motivating example

Consider 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,

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

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, where

$$\gamma_W = \mathbb{E}_P[\boldsymbol{X}\boldsymbol{X}^T]^{-1}\mathbb{E}_P[\boldsymbol{X}W]$$

are the population least squares coefficients, for $W = A$ and $W = Y$. If we assume linear models for $\mathbb{E}_P(Y \mid \boldsymbol{X})$ and $\mathbb{E}_P(A \mid \boldsymbol{X})$, then we have $\beta(P) = \frac{\nu(P)}{\delta(P)}$.

In the problem, we took 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

$$\widehat{\gamma}_W = \mathbb{P}_n[\boldsymbol{X}\boldsymbol{X}^T]^{-1}\mathbb{P}_n[\boldsymbol{X}W]$$

are the sample least squares estimators, for $W = A$ and $W = Y$. We began by analyzing the numerator and denominator separately. Consider for concreteness the numerator functional $\nu(P)$ and its estimator $\widehat{\nu}$. In order to characterize the asymptotic behavior of $\sqrt{n}(\widehat{\nu} - \nu(P))$, we tried to write this as a sample average of iid mean zero terms — as we will soon see, such terms are called the influence function of the estimator $\widehat{\nu}$ — plus some asymptotically negligible quantity. This is useful because the central limit theorem and Slutsky's theorem can then be combined to easily obtain the asymptotic distribution.

With some ingenuity, we were able to show that

$$\sqrt{n}(\widehat{\nu} - \nu(P)) = \sqrt{n}\mathbb{P}_n[(Y - \gamma_Y^T \boldsymbol{X})(Z - \gamma_Z^T \boldsymbol{X}) - \nu(P)] + o_P(1).$$

But it was mysterious, at least to me, why the projection $\gamma_Z^T \boldsymbol{X} = \Pi[Z \mid \boldsymbol{X}]$ suddenly appeared in the influence function, despite $\widehat{\nu}$ not containing either γ_Z or $\widehat{\gamma}_Z$. We simply had to include it for the remainder term to be $o_P(1)$.

We will now discuss a general strategy — outlined to me by none other than Andrea — that is useful in many problems similar to this, where an estimator is a sample average of the observed data O but with an estimator of an unknown parameter plugged in for the truth. We will see that the appearance of $\Pi[Z \mid \boldsymbol{X}]$ can be thought of as the "price to pay" for plugging in $\widehat{\gamma}_Y$ instead of γ_Y in $\widehat{\nu}$. I hope you will find this approach as useful as I do!

General strategy

For any vector γ, let

$$U_\gamma(O) = (Y - \gamma^T \boldsymbol{X})Z.$$

Note that $\widehat{\nu} = \mathbb{P}_n[U_{\widehat{\gamma}_Y}]$ and $\nu(P) = \mathbb{E}_P[U_{\gamma_Y}]$. In general, we have the following expansion

$$\begin{aligned}
\sqrt{n}(\widehat{\nu} - \nu(P)) = \sqrt{n}\left\{\left(\mathbb{P}_n[U_{\widehat{\gamma}_Y}] - \mathbb{E}_P[U_{\widehat{\gamma}_Y}]\right) - \left(\mathbb{P}_n[U_{\gamma_Y}] - \mathbb{E}_P[U_{\gamma_Y}]\right)\right\} \quad (1)\\
+ \sqrt{n}\left\{\mathbb{P}_n[U_{\gamma_Y}] - \mathbb{E}_P[U_{\gamma_Y}]\right\} \quad (2)\\
+ \sqrt{n}\left\{\mathbb{E}_P[U_{\widehat{\gamma}_Y}] - \mathbb{E}_P[U_{\gamma_Y}]\right\}, \quad (3)
\end{aligned}$$

where

$$\mathbb{E}_P[U_\gamma] = \int_{\mathcal{O}} U_\gamma(o)\, dP(o),$$

i.e., γ is treated as a constant in the integral, even if it itself is random like $\widehat{\gamma}_Y$. Terms (1), (2), and (3) are typically handled separately, and have the following names/interpretations:

(1) The centered empirical process term $\mathbb{G}_n(U_{\widehat{\gamma}_Y} - U_{\gamma_Y})$, where for any function measurable real-valued function f of observed data O,

$$\mathbb{G}_n(f) = \sqrt{n}(\mathbb{P}_n[f] - \mathbb{E}_P[f]).$$

The goal in most problems is to show that this is $o_P(1)$. In fact, this is complicated in many problems, and requires ideas from empirical process theory, such as Donsker conditions, or may be dealt with via sample splitting.

(2) A well-behaved linear term $\sqrt{n}\mathbb{P}_n[U_{\gamma_Y} - \mathbb{E}_P(U_{\gamma_Y})] = \mathbb{G}_n(U_{\gamma_Y})$. As a scaled sample mean of mean zero quantities, it converges in distribution by the central limit theorem to $\mathcal{N}(0, \mathrm{Var}_P(U_{\gamma_Y}))$.

(3) An asymptotic bias/drift term $\sqrt{n}\mathbb{E}_P[U_{\widehat{\gamma}_Y} - U_{\gamma_Y}]$. This can be interpreted as the average cost of using $\widehat{\gamma}_Y$ to estimate γ_Y. At times, such as in our motivating example, the contribution is non-negligible, and we will need to add this to the term in (2) to get the full story. We will see in the course that for some semiparametric theory-based estimators, this term can be $o_P(1)$, and in fact this can be seen as the motivation for such estimators!

More generally, a Taylor expansion of the following form is often helpful for handling this term:

$$\begin{aligned}
&\mathbb{E}_P[U_{\widehat{\gamma}_Y}] - \mathbb{E}_P[U_{\gamma_Y}]\\
&= \left(\frac{d}{d\gamma}\mathbb{E}_P[U_\gamma]\Big|_{\gamma=\gamma^*}\right)^T (\widehat{\gamma}_Y - \gamma_Y), \text{ for some } \gamma^* \text{ ``between'' } \widehat{\gamma}_Y \text{ and } \gamma_Y,\\
&= \left(\frac{d}{d\gamma}\mathbb{E}_P[U_\gamma]\Big|_{\gamma=\gamma_Y}\right)^T (\widehat{\gamma}_Y - \gamma_Y) + o_P(n^{-1/2}), \text{ since } \widehat{\gamma}_Y - \gamma_Y = O_P(n^{-1/2}),
\end{aligned}$$

by the continuous mapping theorem when $\mathbb{E}_P[U_\gamma]$ is continuously differentiable in γ.

In addition to providing some intuition for the breakdown of the estimation error, this paradigm is helpful from a calculation perspective; it takes the guesswork out of finding a nice asymptotic representation. We will see that we will obtain the same answer as we did by proceeding directly, but no ingenuity will be required.

Calculations in motivating example

We deal with the three terms individually. First, for (1):

$$\begin{aligned}
&\sqrt{n}\left\{\left(\mathbb{P}_n[U_{\widehat{\gamma}_Y}] - \mathbb{E}_P[U_{\widehat{\gamma}_Y}]\right) - \left(\mathbb{P}_n[U_{\gamma_Y}] - \mathbb{E}_P[U_{\gamma_Y}]\right)\right\} \\
&= \sqrt{n}\left\{\mathbb{P}_n\left[(\gamma_Y - \widehat{\gamma}_Y)^T \boldsymbol{X} Z\right] - \mathbb{E}_P\left[(\gamma_Y - \widehat{\gamma}_Y)^T \boldsymbol{X} Z\right]\right\} \\
&= -(\mathbb{P}_n[Z\boldsymbol{X}^T] - \mathbb{E}_P[Z\boldsymbol{X}^T]) \cdot \sqrt{n}(\widehat{\gamma}_Y - \gamma_Y) \\
&= o_P(1)O_P(1) \\
&= o_P(1).
\end{aligned}$$

Term (2) is straightforward:

$$\sqrt{n}\left\{\mathbb{P}_n[U_{\gamma_Y}] - \mathbb{E}_P[U_{\gamma_Y}]\right\} = \sqrt{n}\mathbb{P}_n\left\{(Y - \gamma_Y^T \boldsymbol{X})Z - \nu(P)\right\}.$$

Finally, for (3),

$$\begin{aligned}
&\sqrt{n}\mathbb{E}_P[U_{\widehat{\gamma}_Y} - U_{\gamma_Y}] \\
&= -\sqrt{n}(\widehat{\gamma}_Y - \gamma_Y)^T \mathbb{E}_P[\boldsymbol{X} Z].
\end{aligned}$$

But recall that the sample least squares coefficient for projecting Y on $\boldsymbol{X}$ is asymptotically linear:

$$\sqrt{n}(\widehat{\gamma}_Y - \gamma_Y) = \sqrt{n}\mathbb{E}_P[\boldsymbol{X}\boldsymbol{X}^T]^{-1}\mathbb{P}_n[\boldsymbol{X}(Y - \gamma_Y^T \boldsymbol{X})] + o_P(1).$$

Thus,

$$\begin{aligned}
&\sqrt{n}\mathbb{E}_P[U_{\widehat{\gamma}_Y} - U_{\gamma_Y}] \\
&= -\sqrt{n}\mathbb{P}_n[(Y - \gamma_Y^T \boldsymbol{X})\boldsymbol{X}^T]\mathbb{E}_P[\boldsymbol{X}\boldsymbol{X}^T]^{-1}\mathbb{E}_P[\boldsymbol{X} Z] + o_P(1). \\
&= -\sqrt{n}\mathbb{P}_n[(Y - \gamma_Y^T \boldsymbol{X})\boldsymbol{X}^T \gamma_Z] + o_P(1)
\end{aligned}$$

Adding (1), (2), and (3) together, we obtain

$$\sqrt{n}(\widehat{\nu} - \nu(P)) = \sqrt{n}\mathbb{P}_n[(Y - \gamma_Y^T \boldsymbol{X})(Z - \gamma_Z^T \boldsymbol{X}) - \nu(P)] + o_P(1),$$

exactly the same as before!