STAT 460/560 Class 22: Asymptotic Normality of Z-estimators with Nuisance Parameters

Ben Bloem-Reddy

Reading: Chapter 5.4, [van98].

Last time, we worked through an extended example in which we applied the abstract results for Z-estimators to the problem of non-linear least-squares regression. In many cases of practical importance, interest is in one parameter, say θ, but in order to get accurate estimates for θ we need to estimate another parameter(s), η. This situation arises, for example, in the estimation of causal effects, in which θ is a parameter that represents the causal effect of interest (say the effect of a treatment X on an outcome Y), and η corresponds to covariates and/or confounders Z that must be accounted for.

Theorem 22.1. *Let $\theta \in \mathbb{R}^k$, η be an element of a metric space. For each (θ, η), let $x \mapsto \psi_{\theta,\eta}(x)$ be a function taking values in $\mathbb{R}^k$ such that the class of functions $\{\psi_{\theta,\eta} \colon \|\theta - \theta_0\| < \delta, d(\eta, \eta_0) < \delta\}$ is sufficiently regular*[1] *for some $\delta > 0$. Assume that $P\psi_{\theta_0,\eta_0} = 0$, and that the maps $\theta \mapsto P\psi_{\theta,\eta}$ are differentiable at θ_0, uniformly in η in a neighborhood of η_0 with nonsingular derivative matrices $V_{\theta_0,\eta}$ such that $V_{\theta_0,\eta} \to V_{\theta_0,\eta_0}$. If $\sqrt{n}\hat{P}_n\psi_{\hat{\theta}_n,\hat{\eta}_n} = o_P(1)$ and $(\hat{\theta}_n, \hat{\eta}_n) \xrightarrow{p} (\theta_0, \eta_0)$, then*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = -V_{\theta_0,\eta_0}^{-1}\sqrt{n}P\psi_{\theta_0,\hat{\eta}_n} - V_{\theta_0,\eta_0}^{-1}\sqrt{n}\sum_{i=1}^{n}\psi_{\theta_0,\eta_0}(X_i) + o_P(1 + \sqrt{n}\|P\psi_{\theta_0,\hat{\eta}_n}\|) . \tag{22.1}$$

Proof. The proof largely follows the same structure as our proof from Class 20, but we need to take care to keep track of the nuisance parameters and their estimator.

The assumed regularity of $\psi_{\eta,\theta}$ allows us to use results from Chapter 19 that imply that

$$\sqrt{n}(\hat{P}_n\psi_{\hat{\theta}_n,\hat{\eta}_n} - P\psi_{\hat{\theta}_n,\hat{\eta}_n}) - \sqrt{n}(\hat{P}_n\psi_{\theta_0,\eta_0} - P\psi_{\theta_0,\eta_0}) \xrightarrow{p} 0 . \tag{22.2}$$

Let's look at the first term. By assumption, $\sqrt{n}\hat{P}_n\psi_{\hat{\theta}_n,\hat{\eta}_n} = o_P(1)$, so

$$\begin{aligned}
\sqrt{n}(\hat{P}_n\psi_{\hat{\theta}_n,\hat{\eta}_n} - P\psi_{\hat{\theta}_n,\hat{\eta}_n}) &= -\sqrt{n}P\psi_{\hat{\theta}_n,\hat{\eta}_n} + o_P(1) \\
&= \sqrt{n}(P\psi_{\theta_0,\hat{\eta}_n} - P\psi_{\hat{\theta}_n,\hat{\eta}_n}) - \sqrt{n}P\psi_{\theta_0,\hat{\eta}_n} + o_P(1) \\
&= \sqrt{n}V_{\theta_0,\hat{\eta}_n}(\theta_0 - \hat{\theta}_n) - \sqrt{n}P\psi_{\theta_0,\hat{\eta}_n} + o_P(1 + \sqrt{n}\|\theta_0 - \hat{\theta}_n\|) .
\end{aligned}$$

Combining this with (22.2) (which implies that $\sqrt{n}(\hat{P}_n\psi_{\hat{\theta}_n,\hat{\eta}_n} - P\psi_{\hat{\theta}_n,\hat{\eta}_n}) = \sqrt{n}(\hat{P}_n\psi_{\theta_0,\eta_0} - P\psi_{\theta_0,\eta_0}) + o_P(1)$), we have

$$\sqrt{n}(\hat{P}_n\psi_{\theta_0,\eta_0} - P\psi_{\theta_0,\eta_0}) + o_P(1) = \sqrt{n}V_{\theta_0,\hat{\eta}_n}(\theta_0 - \hat{\theta}_n) - \sqrt{n}P\psi_{\theta_0,\hat{\eta}_n} + o_P(1 + \sqrt{n}\|\theta_0 - \hat{\theta}_n\|) ,$$

and therefore

$$\begin{aligned}
\sqrt{n}(\hat{\theta}_n - \theta_0) &= -V_{\theta_0,\hat{\eta}_n}^{-1}\sqrt{n}P\psi_{\theta_0,\hat{\eta}_n} - V_{\theta_0,\hat{\eta}_n}^{-1}\sqrt{n}(\hat{P}_n\psi_{\theta_0,\eta_0} - P\psi_{\theta_0,\eta_0}) + o_P(1 + \sqrt{n}\|\theta_0 - \hat{\theta}_n\|) \\
&= -V_{\theta_0,\eta_0}^{-1}\sqrt{n}P\psi_{\theta_0,\hat{\eta}_n} - V_{\theta_0,\eta_0}^{-1}\sqrt{n}\hat{P}_n\psi_{\theta_0,\eta_0} + o_P(1 + \sqrt{n}\|\theta_0 - \hat{\theta}_n\|) \\
&\quad + \underbrace{(V_{\theta_0,\eta_0}^{-1} - V_{\theta_0,\hat{\eta}_n}^{-1})}_{=o_P(1)}(\sqrt{n}P\psi_{\theta_0,\hat{\eta}_n} + \underbrace{\sqrt{n}\hat{P}_n\psi_{\theta_0,\eta_0}}_{O_P(1)}) .
\end{aligned}$$

[1] Namely, it is Donsker (see van der Vaart [van98], Chapter 19) and $P\|\psi_{\theta,\eta} - \psi_{\theta_0,\eta_0}\|^2 \to 0$ as $(\theta, \eta) \to (\theta_0, \eta_0)$.

Finally, the result follows by showing that

$$\sqrt{n}\|\theta_0 - \hat{\theta}_n\|(1 + o_P(1)) \le O_P(1) + \sqrt{n} P\psi_{\theta_0,\hat{\eta}_n} \ . \tag{22.3}$$

□

Exercise 22.1. Show that (22.3) holds.

Compared to the asymptotic normality of Z-estimators without nuisance parameters, we typically pay the cost of extra variance from estimating η. Even in the nice case in which $\sqrt{n}\|P\psi_{\theta_0,\hat{\eta}_n}\| = O_P(1)$, and $\sqrt{n}(\hat{\eta}_n - \eta_0) \rightsquigarrow \mathcal{N}(0, \Sigma_{\eta_0})$, and $\eta \mapsto P\psi_{\theta_0,\eta}$ is differentiable at η_0 with derivative W_{θ_0,η_0}, the limiting distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ is

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = -V_{\theta,\eta}^{-1} W_{\theta_0,\eta_0} \sqrt{n}(\hat{\eta}_n - \eta_0) - V_{\theta,\eta}^{-1}\sqrt{n}\sum_{i=1}^{n} \psi_{\theta_0,\eta_0}(X_i) + o_P(1 + \sqrt{n}\|P\psi_{\theta_0,\hat{\eta}_n}\|) \tag{22.4}$$

$$\rightsquigarrow \mathcal{N}\left(0, V_{\theta,\eta}^{-1} W_{\theta_0,\eta_0} \Sigma_{\eta_0} W_{\theta_0,\eta_0}^{\top} (V_{\theta,\eta}^{-1})^{\top} + V_{\theta,\eta}^{-1} P(\psi_{\theta_0,\eta_0}\psi_{\theta_0,\eta_0}^{\top})(V_{\theta,\eta}^{-1})^{\top}\right) \ , \tag{22.5}$$

which amounts to additional variance coming from the estimation of $\hat{\eta}$.

However, if we're clever and/or lucky, we can set things up so that $W_{\theta_0,\eta_0} = 0$.

Activity 22.1. Suppose that $X_i \in \mathbb{R}$, with density symmetric about θ_0, so that $p(x-\theta_0) = p(-(x-\theta_0))$. Let $\psi(\bullet)$ be an odd function that satisfies the necessary assumptions above, with $\hat{\theta}_n$ obtained from estimation equations

$$\frac{1}{n}\sum_{i=1}^{n} \psi\left(\frac{X_i - \theta}{\hat{\eta}_n}\right) \ ,$$

and $\hat{\eta}_n$ is a $\sqrt{n}$-consistent estimator of the scale of the data.

Show that the "additional asymptotic variance" from estimating η is zero.

References

[van98] A. W. van der Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1998.