Section 10: Large sample theory

ARE 210

November 7, 2017

The section notes are available on the section Github at github.com/johnloeser/are210 in the "section10" folder.

1 Definitions

- Assume we observe iid $W = \{W_i\}_{i=1}^n$, with $W_i \sim P$, interested in $\theta(P)$
  - $\hat{\theta}_n$ is a **consistent** estimator of θ if $\hat{\theta}_n \xrightarrow{p} \theta$
  - $\hat{\theta}_n$ is **asymptotically normal** if $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, V)$ for some V
  - If $V_n \xrightarrow{p} V$ as defined above, we say it's a **consistent variance estimator**
- Math tricks
  - ULLN (defined below) used repeatedly for convergence of functions, when we want to apply LLN but don't satisfy requirements
  - **MVT**: $h : \mathbf{R}^p \to \mathbf{R}^q$ continuously differentiable, then $h(b) = h(\theta) + \frac{\partial h(\tilde{b})}{\partial b}(b - \theta)$, where $\tilde{b} = \alpha b + (1-\alpha)\theta$ for $\alpha \in (0,1)$
  - Note that I'll confusingly use W to either denote the data or a random variable with distribution P
- **Extremum estimation**: $\hat{\theta}_n = \arg\max_{b \in \Theta} Q_n(W, b)$
  - **Consistency 1**: Assume (i) Θ is compact, (ii) Q_n is continuous in $b \in \Theta$, (iii) Q_n is measurable in W, (iv) $\exists\ Q_0(b)$ such that 1) $Q_0(b)$ is uniquely maximized at $\theta \in \Theta$, and 2) $\sup_{b \in \Theta} |Q_n(b, W) - Q_0(b)| \xrightarrow{p} 0$. Then $\hat{\theta}_n \xrightarrow{p} \theta$.
  - **Consistency 2**: Substitute (i) for $\theta \in \text{int}(\Theta)$ and Θ is convex, and (ii) for $Q_n(b, W)$ is concave in b
  - **Asymptotic normality**: Suppose (i) $\hat{\theta}_n \xrightarrow{p} \theta$, (ii) $\theta \in \text{int}(\Theta)$, (iii) $Q_n(b, W)$ twice continuously differentiable with respect to $b \in \mathcal{N}_\theta$, where $\mathcal{N}_\theta$ denotes a neighborhood of θ, (iv) $\sqrt{n}\frac{\partial}{\partial b}Q_n(b, W) \xrightarrow{d} N(0, \Sigma(\theta))$, (v) $\sup_{c \in \mathcal{N}_\theta} \left| \frac{\partial^2 Q_n(c, W)}{\partial b \partial b^{\mathrm{T}}} - H(c) \right| \xrightarrow{p} 0$, where $H(c) = \frac{\partial^2 Q_0(c)}{\partial b \partial b^{\mathrm{T}}}$. Then $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, V)$, where $V = H(\theta)^{-1}\Sigma(\theta)H(\theta)^{-1}$
- **M-estimation**: $\hat{\theta}_n = \arg\max_{b \in \Theta} \frac{1}{n}\sum_{i=1}^n q(W_i, b)$
  - **ULLN, Consistency 1**: (Used for proof of consistency) Suppose (i) Θ is compact, (ii) $q(W, b)$ is continuous in b, (iii) $q(W, b)$ is measurable in W, (iv) $\mathbf{E}[\sup_{b \in \Theta} |q(W, b)|] < \infty$. Then $\sup_{b \in \Theta} \left| \frac{1}{n}\sum_{i=1}^n q(W_i, b) - \mathbf{E}[q(W, b)] \right| \xrightarrow{p} 0$.

- **Consistency 2**: Assume (i) $\theta \in \text{int}(\Theta)$ and Θ is convex, (ii) $q(W, b)$ is concave in b, (iii) $q(W, b)$ is measurable in W, (iv) 1) $\mathbf{E}[q(W, b)]$ is uniquely maximized at θ and 2) $\mathbf{E}[|q(W, b)|] < \infty \ \forall \ b \in \Theta$. Then $\hat{\theta}_n \xrightarrow{p} \theta$.
- **Asymptotic normality**: Follow results for extremum estimation. Letting $S(W, b) = \frac{\partial q(W,b)}{\partial b}$ and $H(W, b) = \frac{\partial S(W,b)^T}{\partial b}$, $\Sigma(\theta) = V[S(W, \theta)]$ and $H(\theta) = \mathbf{E}[H(W, \theta)]$.
  - **NLLS**: Let $q((Y, Z), b) = -\frac{1}{2}(Y - \Psi(Z, b))^2$, and let $\epsilon = Y - \Psi(Z, \theta)$. Then $\Sigma(\theta) = \mathbf{E}[\epsilon^2 \frac{\partial \Psi}{\partial b} \frac{\partial \Psi}{\partial b}^{\mathrm{T}}]$ and $H(\theta) = -\mathbf{E}[\frac{\partial \Psi}{\partial b} \frac{\partial \Psi}{\partial b}^{\mathrm{T}}]$. We call $\mathbf{E}[\epsilon^2 | Z] = \sigma^2$ **conditional homoskedasticity**.

- **MLE**: $\hat{\theta}_n = \arg\max_{b \in \Theta} \frac{1}{n} \sum_{i=1}^{n} \log p(W_i, b)$
  - **Identification**: Sufficient that θ uniquely maximizes the log likelihood
  - **Consistency**: See conditions for M-estimation or GMM (using moment condition $\mathbf{E}_\theta \left[\frac{\partial \log p(W_i, \theta)}{\partial b} \right] = 0$)
  - **Asymptotic normality**: Follow results for M-estimation. Letting $I(\theta)$ be the Fisher Information Matrix (variance of the score), $\Sigma(\theta) = I(\theta)$ and $H(\theta) = I(\theta)$, so $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, I(\theta)^{-1})$
    - Ronald Fisher thought $\Rightarrow$ MLE is asymptotically efficient and achieves the Cramer-Rao lower bound, so you'll be forgiven for thinking this also

- **GMM**: $\hat{\theta}_n = \arg\max_{b \in \Theta} -\frac{1}{2}[\frac{1}{n} \sum_{i=1}^{n} m(W_i, b)]^{\mathrm{T}} S_n(W) [\frac{1}{n} \sum_{i=1}^{n} m(W_i, b)]$
  - **Identification**: Let $Q_0(b) = -\frac{1}{2}\mathbf{E}[m(W, b)]^{\mathrm{T}} S \mathbf{E}[m(W, b)]$. Suppose $\mathbf{E}[m(W, b)] = 0 \Leftrightarrow b = \theta$. Then if S is strictly positive definite, $\theta = \arg\max_b Q_0(b)$.
  - **Consistency 1**: Suppose (i) Θ is a compact subset of $\mathbf{R}^d$, (ii) $m(W, b)$ is continuous in b, (iii) $m(W, b)$ is measurable in W, and (iv) $S_n(W) \xrightarrow{p} S$, where S is symmetric positive definite. Suppose further (1) assumptions made above for identification, and (2) $\mathbf{E}[\sup_{b \in \Theta} |m(W, b)|] < \infty$. Then $\hat{\theta}_n \xrightarrow{p} \theta$.
  - **Consistency 2**: Replace (i), (ii), and (2) with $Q_n(b)$ is concave and $\mathbf{E}[m(W, b)]$ exists and is finite
  - **Asymptotic normality**: Let $m_n(b) = \sum_{i=1}^{n} m(W_i, b)$, $M_n(b) = \frac{\partial m_n(b)}{\partial b}$, and $M(b) = \mathbf{E}[\frac{\partial m(W,b)}{\partial b}]$. Then assuming 1) $M_n(\hat{\theta}_n) W_n M_n(\tilde{b}_n)'$ is invertible, for $\tilde{b}_n = \alpha \hat{\theta}_n + (1 - \alpha)\theta$ for $\alpha \in [0, 1]$, 2) $M_n(\hat{\theta}_n) \xrightarrow{p} M(\theta)$, and 3) assumptions made above for consistency, $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, V)$ where $V = (M(\theta) S M(\theta)^{\mathrm{T}})^{-1} M(\theta) S V[m(W, \theta)] S M(\theta)^{\mathrm{T}} (M(\theta) S M(\theta)^{\mathrm{T}})^{-1}$

2 Some useful bits

1. This is the fun part! Get excited!
2. Each find-the-estimator problem will generally follow three steps - 1) show consistency, 2) show asymptotic normality, and 3) show you've got a consistent estimator of the variance. Typically we'll work with M-estimators, MLE, or GMM estimators, so you'll be able to recognize the class of estimator and appeal to existing results for each step. You'll often need to apply the MVT and a ULLN to show the variance estimator is consistent.

3 Practice questions

1) Derive the Fisher Information matrix of a $N(\mu, \sigma^2)$ random variable. Construct an unbiased 95% confidence interval for $\widehat{\mu}_{MLE}$ using $\widehat{\sigma}_{MLE}$. What is its asymptotic variance? How can we interpret it asymptotically?

2) (Problem 1, PS5) Consider the parametric model $\{p\,(x, \theta) : \theta > 0\}$ where

$$p\,(x, \theta) = \theta x^{\theta - 1} \quad x \in (0, 1)$$

a. Suppose we observe an i.i.d. sample from this density. Find the Maximum Likelihood estimator of θ and calculate the Fisher Information.

b. Show whether the the MLE is consistent for θ.

c. Derive the limiting distribution of the MLE.

d. Find a Method of Moments estimator for θ and discuss its consistency.

e. Does there exist a UMVUE for θ? If so, does it attain the Cramer-Rao lower bound?

3) Prove the asymptotic normality of the GMM estimator.