Section 8: Estimators and statistics

ARE 210

October 24, 2017

The section notes are available on the section Github at github.com/johnloeser/are210 in the "section8" folder.

Quick note - I won't be around next week Monday and Tuesday. We need to pick a two hour slot next week - either Wednesday 10-12 or Thursday 8-10. We may be constrained to Thursday due to room availability.

1 Definitions

- **Estimation**
  - **MLE**: $\{Y_i\}_{i=1}^n$ iid, $Y_i \sim P_\theta$ with density p_θ
    * $\theta = \arg\max_b \mathbf{E}_{P_\theta}[\log p_b(Y)]$ (proof using KLIC + Jensen's inequality)
    * $\hat{\theta}_{MLE} = \arg\max_b \frac{1}{n}\sum_{i=1}^n \log p_b(y_i)$
    * **Invariance property of MLE**: Let $P_{\theta_0} \in \{P_\theta : \theta \in \Theta\}$, and let $\lambda_0 = h(\theta_0)$. Then $\hat{\lambda}_{MLE} = h(\hat{\theta}_{MLE})$.
    * **CMLE**: $\hat{\theta}_{CMLE} = \arg\max_{b\in\Theta} \frac{1}{n}\sum_{i=1}^n \log p_b(y_i|z_i)$
  - **M-estimation**: Implicitly define $\theta(P) = \arg\min_{b\in\Theta} \mathbf{E}_P[q(X,b)]$, assume $\{X_i\}_{i=1}^n \sim$ iid P. Then $\hat{\theta}_M = \arg\min_{b\in\Theta} \frac{1}{n}\sum_{i=1}^n q(X_i, b)$
    * **NLLS**: $q(X_i, b) = (Y_i - \Phi(Z_i, b))^2$
    * **Quantile**: $q_Q(X_i, b) = (X_i - b)(\mathbf{1}\{Y > b\} - Q)$
  - **Method of moments**: $\{X_i\}_{i=1}^n$ iid, implicitly define $\mathbf{E}_P[m(X, \theta(P))] = 0$
    * Estimator: $\arg\min_b (\frac{1}{n}\sum_{i=1}^n m(x_i, b))^{\mathrm{T}} S (\frac{1}{n}\sum_{i=1}^n m(x_i, b))$
  - **Minimum distance**: $\hat{\theta}_n = \arg\min_b g_n(b)^{\mathrm{T}} S g_n(b)$
    * "Classical" minimum distance: $g_n(b) = \Pi_n - h(b)$
- **Statistics**
  - A function of the data $T(X)$ is a **statistic**
  - $X \sim P_{\theta_0}$, $\theta_0 \in \Theta$, $T(X)$ is **sufficient** for $\mathbf{P} = \{P_\theta : \theta \in \Theta\}$ (or sufficient for θ) if $P_\theta[X|T] = P[X|T]$
  - **Factorization theorem**: $\{f_\theta : \theta \in \Theta\}$, $T(X)$ statistic, T is sufficient $\Leftrightarrow$ $f(X, \theta) = g(T(X), \theta)h(X)$

- **Exponential family of distributions**:
  $p(x, \theta) = h(x) \exp(\sum_{j=1}^{K} \eta_j(\theta) T_j(x) - \beta(\theta))$
- T is a **minimally sufficient** statistic if and only if T is sufficient, and T is a function of any other sufficient statistic S
- S is **ancillary** for θ if the distribution of S does not depend on θ
- T is **complete** for θ if $\forall\, g$ such that $\mathbf{E}_\theta[g(T)] = 0 \;\forall\, \theta \in \Theta$, $g(T) = 0$ almost everywhere.
- X_i iid, $p(X, \theta)$ from the exponential family, then $T(X_1, \ldots, X_n) = (\sum_{i=1}^{n} T_1(X_i), \ldots, \sum_{i=1}^{n} T_J(X_i))$ is complete for θ if $\{\eta_1(\theta), \ldots, \eta_J(\theta)\}$ contains an open set in $\mathbf{R}^J$. By the factorization theorem, $T(X_1, \ldots, X_n)$ is always sufficient for θ.
- **Basu**: If T is complete and sufficient, then $T \perp$ every ancillary statistic.

2 Some useful bits

1. It's useful to think about sufficient statistics as capturing weakly more information than complete statistics, and the complete statistic with the most information is minimally sufficient, while a minimally sufficient statistic is complete. Ancillary statistics are roughly the parts of the data that are orthogonal to our sufficient statistics. All of this may or may not be strictly true but the intuition roughly holds.

3 Practice questions

1) (Midterm, Question 6) Let $Z = g(X)(Y - \mathbb{E}(Y|X))$ for some measurable function $g : \mathbb{R} \to \mathbb{R}$. Assume that all moments in question exist.

a) Compute $\mathbf{E}(Z|X)$.

b) Compute $V(Z)$

c) Now let Let $Z = \mathbf{E}(g(X))(Y - \mathbb{E}(Y|X))$. Compute $V(Z)$.

2) (Midterm, Question 7) Let $Y_i^* \sim N(\theta, 1)$ and let $Y_i = \mathbb{I}\{Y_i^* > 0\}$. Suppose we observe a an i.i.d. sample $\{Y_i\}_{i=1}^{n}$.

a) Let Φ^{-1} denote the inverse of the cumulative distribution function (CDF) of the $N(0, 1)$ distribution and let ϕ denote the probability density function of the $N(0, 1)$ distribution. Find the MLE for θ and denote it by $\hat{\theta}_n$.

b) Show whether $\hat{\theta}_n$ converges in probability to θ.

c) Derive the limiting distribution for (an appropriately normalized) $\hat{\theta}_n$

3) (Problem Set 4, Question 4) (**Identification in an Endogenous Non-Parametric Regression Model**) Suppose that we have a model given by

$$y_i = \mu(x_i) + \epsilon_i$$

and that we assume $\mathbb{E}(\epsilon_i|z_i) = 0$ for some variable z_i and the model is non-parametric in the sense that we place no restrictions on the form of the (unknown) function $\mu(\cdot)$. Consider taking expectations conditional on z to obtain

$$\int y f(y|z)\, dy = \int \mu(x) f(x|z)\, dx$$

The above is an example of an integral equation. Note that since we observe (y, x, z) the density functions $f(y|z)$ and $f(x|z)$ are identified. We will show that this integral equation will have a unique solution if the distribution of x conditional on z is complete (in z).

Suppose that there exists another function $\tilde{\mu}(\cdot)$ that satisfies the equation, then we must have

$$\int (\mu(x) - \tilde{\mu}(x)) f(x|z)\, dx = 0$$

In order for their to be a unique solution therefore we must have the following hold (setting $\delta(x) = \mu(x) - \tilde{\mu}(x)$)

$$\int \delta(x) f(x|z)\, dx = 0 \Rightarrow \delta(x) = 0 \tag{1}$$

for almost all z. Suppose that the distribution of x conditional on z is given by

$$f(x|z) = h(x) \exp(\eta(z)\tau(x) - B(z)) \tag{2}$$

Suppose further that the function $\tau(\cdot)$ is one-to-one.[1] Show that this implies (1) so a sufficient condition for the integral equation to have a unique solution is the distributional assumption (2) (Hint: Use the result on Completeness for Exponential Families. See Newey & Powell (2003) for an extension to cases where the conditional

[1]Formally we also need to assume that the support of $\mu(z)$ contains an open set and $h(x) > 0$.

distribution does not belong to an exponential family.)