Section 7: Identification and MLE

ARE 210

October 17, 2017

The section notes are available on the section Github at github.com/johnloeser/are210 in the "section7" folder.

1 Definitions

- **Identification**
  - Data $\mathbf{X}$, takes values in sample space $\mathcal{X}$, $\mathbf{X}$ has unknown distribution $P \in \mathbf{P}$ family of probability distributions on $\mathcal{X}$, typically assume
    1. Often $\mathbf{P} = \{P_\theta : \theta \in \Theta\}$
    2. $\mathbf{X} = \{X_i\}_{i=1}^n$ iid, so probability distribution of data $\times_{i=1}^n P$
  - A **parameter** is a mapping $\nu : \mathbf{P} \to \mathcal{N}$ (i.e. function of distribution of $\mathbf{X}$)
  - The **identified set** $\Theta(P) = \{\theta \in \Theta : \theta = \arg\max_{b \in \Theta} Q_0(b, P)\}$
    * More generally, $\nu(P)$, where we allow ν to be set valued
    * **Point identification** $\Leftrightarrow$ the identified set is a singleton
- **Estimation methods**
  - A typical problem:
    1. Show $\Theta(P)$ is a singleton ($P_{\theta_1} = P_{\theta_2} \Rightarrow \theta_1 = \theta_2$) or characterize the set
    2. Construct an estimator, often $\Theta(P_n)$, where P_n is the empirical distribution
    3. Derive its asymptotic properties using implicit function theorem, delta method, ...
  - **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)
    * $\widehat{\theta}_{MLE} = \arg\max_b \frac{1}{n}\sum_{i=1}^n \log p_b(y_i)$
    * **Invariance property of MLE**: Let $P_{\theta_0} \in \{P_\theta : \theta \in \Theta\}$, and let $\lambda_0 = h(\theta_0)$. Then $\widehat{\lambda}_{MLE} = h(\widehat{\theta}_{MLE})$.

2 Practice questions

1) Let $Y \sim \text{Bern}(p)$.

a) Construct the MLE of p, $\widehat{p}_{MLE}$.

b) What is the probability distribution of $\widehat{p}_{MLE}$? Note that this case is relatively unique in that the distribution of this estimator has a closed form solution.

c) Derive the asymptotic distribution of $\widehat{p}_{MLE}$.

2) (Problem set 4, Question 7) Consider the following model for an independently distributed sample $\{Y_i\}_{i=1}^n = \{Y_{i1}, Y_{i2}\}_{i=1}^n$ where each Y_{ij} is a binary valued random variable and $Y_{ij} \sim \text{Bernoulli}(\mu_{ij})$ for $i = 1, \ldots, n$. Suppose that (using the logit function, $\ell(\mu) = \log \frac{\mu}{1+\mu}$),

$$\ell(\mu_{ij}) = \beta_1 + \beta_2 z_{ij} + \theta_i$$

for a sequence of known constants $\{z_{i1}, z_{i2}\}_{i=1}^n$ and unknown constants $\{\theta_i\}_{i=1}^n$. Consider the statistic $T(Y_i) = Y_{i1} + Y_{i2}$ and show that the probability

$$P(Y_{i1} = a, Y_{i2} = b | T(Y_i) = 1)$$

does not depend upon θ_i (the θ_i are known as <u>fixed effects</u>).

3) (Midterm 2016, Question 7) Consider the following two-stage experiment. In the first stage we draw a random variable X with support $[0, 1]$. In the second stage we toss a coin thrice independently with the probability of heads in each toss given by the observed value of X. To fix ideas, let the density function of X be given by

$$f(x) = 6x(1-x)\mathbf{1}\{x \in [0, 1]\}$$

Let Y denote the number of heads in three independent tosses of a coin with probability x of heads (where $X = x$).

a) (4 points) What is the support of the vector of random variables (X, Y)? Does the joint distribution have any atoms?

b) (6 points) What is the density (mass) function $p_Y(y)$ of Y (i.e. the marginal distribution of Y)? Hint: Note that

$$\int_0^1 x^{a-1}(1-x)^{b-1}dx = \frac{(a-1)!(b-1)!}{(a+b-1)!}$$

for $a, b \in \mathbf{N}_+$. Compute $p_Y(0)$.

c) (2 points) What is $\mathbf{E}(Y|X)$?

d) (2 points) What is $\mathrm{Var}(Y|X)$?

e) (8 points) Use the variance equality

$$\mathrm{Var}(Y) = \mathbf{E}(\mathrm{Var}(Y|X)) + \mathrm{Var}(\mathbf{E}(Y|X))$$

to compute the variance of Y.