Economics 620, Lecture 9:

Asymptotics III: Maximum Likelihood Estimation

Jensen's Inequality: Suppose X is a random variable with $E(X) = \mu$, and f is a convex function. Then

$$E(f(X)) > f(E(X)).$$

This inequality will be used to get the consistency of the ML estimator.

Model and Assumptions

Let $p(x|\theta)$ be the probability density function of X given the parameter.

Consider a random sample of n observations and let

$$\ell(\theta|x_1, x_2, ..., x_n) = \sum_{i=1}^{n} \ln p(x_i|\theta)$$

be the log likelihood function.

Assume θ_0 is the true value and that $d \ln p/d\theta$ exists in an interval including θ_0, furthermore, make the assumptions:

Assumption 1:

$$\frac{d \ln p}{d\theta}; \frac{d^2 \ln p}{d\theta^2}; \frac{d^3 \ln p}{d\theta^3}$$

exist in an interval including θ_0.

Model and Assumptions2

Assumption 2:

$$E\left(\frac{p'}{p}|\theta\right) = 0; E\left(\frac{p''}{p}|\theta\right) = 0; E\left(\frac{p'^2}{p}|\theta\right) > 0$$

where $p' = dp/d\theta$ and $p'' = d^2p/d\theta^2$. These usually hold in the problems we will see.

Assumption 3:

$$\left|\frac{d^3 \ln p}{d\theta^3}\right| < M(x) \text{ where } E[M(x)] < K.$$

This is a technical assumption. It will control the expected error in Taylor expansions.

Consistency

We can get the consistency of the ML estimator immediately. We will use assumptions 1-3 to get the asymptotic normality of a consistent estimator in general and the ML estimator in particular.

Suppose $\bar{\theta}$ is an estimator for θ. We would like to require that the probability of $\bar{\theta}$ being close to the true value of θ (i.e., θ_0) should increase as the sample size increases.

Definition: As estimator $\bar{\theta}$ is said to be consistent for θ_0 if $\text{plim } \bar{\theta} = \theta_0$.

Consistency 2

Proposition: If $\ln p$ is differentiable, then the ML equation

$$\frac{d\ell}{d\theta} = 0$$

(*first order condition*) has a root with probability 1 which is consistent for θ, i.e., the ML estimator for θ is consistent.

Proof. Using Jensen's inequality for *concave* functions

$$E \ln \left[\frac{p(\theta_0 - \delta)}{p(\theta_0)} | \theta_0 \right] < 0; \; E \ln \left[\frac{p(\theta_0 + \delta)}{p(\theta_0)} | \theta_0 \right] < 0$$

where δ is a small number.

Consistency 3

The inequality is strict unless p does not depend on. To see this, note that

$$\begin{aligned} E \ln\left[\frac{p(\theta_0+\delta)}{p(\theta_0)}\right] &< \ln E\left[\frac{p(\theta_0+\delta)}{p(\theta_0)}\right] \\ &= \ln \int p(\theta_0+\delta)dx = \ln 1 = 0. \end{aligned}$$

Then noting the definition of $\ell(\theta)$ and using SLLN,

$$\lim\left(\frac{1}{n}[\ell(\theta_0 \pm \delta) - \ell(\theta_0)]\right) < 0$$

$\ell(\theta)$ has a local maximum at θ_0 in the limit.

Implying that the first order condition is satisfied at θ_0 in the limit.

Note that we have not shown that the MLE is a *global* max – this requires more conditions.

Asymptotic Normality of Consistent Estimators

Proposition: Let

$$-E\left[\frac{d^2 \ln p}{d\theta_0^2}\right] = E\left[\left(\frac{d \ln p}{d\theta_0}\right)^2\right] = i(\theta_0).$$

Let $\bar{\theta}$ be the consistent MLE estimator for θ. Then

$$\sqrt{n}\left[(\bar{\theta} - \theta_0)i(\theta_0) - \frac{1}{n}\frac{d\ell}{d\theta_0}\right] \to 0$$

in probability.

Asymptotic Normality of Consistent Estimators 2

Proof: From the first order condition, we get the following expansion:

$$0 = \frac{d\ell}{d\theta_0} + (\bar{\theta} - \theta_0)\frac{d\ell}{d\theta_0^2} + \frac{(\bar{\theta} - \theta_0)^2}{2}\frac{d^3\ell}{d\theta_0^3}$$

$$\Rightarrow \sqrt{n}(\bar{\theta} - \theta) = \frac{-\frac{1}{\sqrt{n}}\frac{d\ell}{d\theta_0}}{\frac{1}{n}\left[\frac{d^2\ell}{d\theta_0^2} + \frac{(\bar{\theta}-\theta_0)}{2}\frac{d^3\ell}{d\theta_0^3}\right]}$$

Taking the probability limit we note that the first expression in the denominator converges to $-i(\theta_0)$ and the second expression in the denominator converges to 0. (*Why*?)

Asymptotic Normality of Consistent Estimators 3

Proposition: (*Asymptotic Normality*)

$$\sqrt{n}(\bar{\theta} - \theta_0) \to N(0, i(\theta_0)^{-1})$$

Proof. Note that $\sqrt{n}(\bar{\theta} - \theta_0) i(\theta_0)$ has the same asymptotic distribution as

$$\frac{1}{\sqrt{n}} \frac{d\ell}{d\theta_0} = \sqrt{n} \left(\frac{1}{n} \sum \frac{d \ln p}{d\theta_0} \right).$$

We know that

$$E \left[\frac{d \ln p}{d\theta_0} \right] = 0$$

since

$$\int p(x|\theta_0) dx = 1 \Rightarrow \int p' dx = 0 = E \left[\frac{d \ln p}{d\theta_0} \right].$$

Asymptotic Normality of Consistent Estimators 4

Note that differentiating $\int p \; d \ln p \; dx = 0$ again implies that

$$\int p \left(\frac{p'}{p} \right) d \ln p \; dx + \int p \; d^2 \ln p \; dx = 0.$$

The first term is just the variance of $d \ln p / d\theta_0$ and the second expression is $-i(\theta_0)$. Thus,

$$V \left[\frac{d \ln p}{d\theta_0} \right] = i(\theta_0).$$

Now we use the Central Limit Theorem for

$$\sqrt{n} \left(\frac{1}{n} \sum \frac{d \ln p}{d\theta_0} \right) \text{ with } \begin{aligned} E \left[\frac{d \ln p}{d\theta_0} \right] &= 0, \\ V \left[\frac{d \ln p}{d\theta_0} \right] &= i(\theta_0) \end{aligned}$$

Asymptotic Normality of Consistent Estimators 5

to obtain

$$\sqrt{n}(\bar{\theta}-\theta_0)i(\theta_0) \rightarrow N(0, i(\theta_0)).$$

Thus (using $z \sim N(0, \sum) \Rightarrow Az \sim N(0, A \sum A'))$

$$\sqrt{n}(\bar{\theta}-\theta_0) \rightarrow N(0, i(\theta_0)^{-1})$$

Basic result: Approximate the distribution of

$$(\bar{\theta}-\theta_0) \text{ by } N\left(0, \frac{i(\theta_0)^{-1}}{n}\right).$$

Of course, $i(\theta_0)^{-1}$ is consistently estimated by $i(\bar{\theta})^{-1}$ under our assumptions. (*Why*?)

Applications

1. This will give the exact distribution in estmating a normal mean. Check this.

2. Consider a regression model with $Ey = X\beta$, $Vy = \sigma^2 I$ and $y \sim$ normal. Check that the asymptotic distribution of $\hat{\beta}$ is equal to its exact distribution.

Miscellaneous Useful Results

Consistency of continuous functions of ML estimators:

Suppose $\hat{\theta}$ is the ML estimator.

Recall that plim $\hat{\theta} = \theta_0 \Rightarrow$ plim $g(\hat{\theta}) = g(\theta_0)$.

(Choice of parametrization is irrelevant in this regard.)

Note: Do not use the ambiguous term “asymptotically unbiased” estimators.

Why do we use ML estimators?

Under our assumptions which provide lots of smoothness, ML estimators are asymptotically efficient - attaining (asymptotically) the Cramer-Rao lower bound on variance. (*What is the relation to the Gauss-Markov property*?)

Proposition: (*Cramer-Rao bound for unbiased estimators.*) Let p be the density function.

Suppose θ^* is an unbiased estimator of θ_0. Then

$$V(\theta^*) \geq \left[V\left(\frac{d \ln p}{d\theta_0}\right)\right]^{-1} = i(\theta_0)^{-1}.$$

Cramer-Rao Bound

Proof: Note that

$$E(\theta^*) = \theta_0 = \int \theta^* p dx$$

from unbiasedness. Note that θ^* is a function of x but not θ_0.

Differentiating the above equality with respect to θ_0, we get

$$\begin{aligned} 1 &= \int \theta^* p' dx = \int \theta^* \left(\frac{p'}{p}\right) p dx \\ &= E\left[\theta^* \left(\frac{p'}{p}\right)\right] = E\left[\theta^* \frac{d \ln p}{d\theta_0}\right] \\ &= cov\left[\theta^*, \frac{d \ln p}{d\theta_0}\right] \end{aligned}$$

The Cauchy-Schwartz inequality implies that

$$[cov(X, Y)]^2 < V(X)V(Y).$$

Thus

$$\begin{aligned}\left[cov\left[\theta^*, \frac{d \ln p}{d\theta_0}\right]\right]^2 &= 1 \leq V(\theta^*)V\left(\frac{d \ln p}{d\theta_0}\right) \\ &\Rightarrow V(\theta^*) \geq \left[V\left(\frac{d \ln p}{d\theta_0}\right)\right]\end{aligned}$$

Note:

$$E\left(\frac{d^2 \ln p}{d\theta^2}\right) = -E\left[\left(\frac{d \ln p}{d\theta}\right)^2\right] = -V\left(\frac{d \ln p}{d\theta}\right)$$

Since

$$\begin{aligned}\frac{d^2 \ln p}{d\theta^2} &= \frac{d}{d\theta}\left(\frac{p'}{p}\right) = \frac{pp'' - (p')^2}{p^2} = \frac{p''}{p} - \left(\frac{p'}{p}\right)^2 \\ &\Rightarrow E\left(\frac{d^2 \ln p}{d\theta^2}\right) = -E\left[\left(\frac{d \ln p}{d\theta}\right)^2\right]\end{aligned}$$

(*why*?)

Thus we have an expression for the variance of the first derivative on $\ln p$ in term of the second derivative - a property we have seen before.

Linear Model:

The assumption of "fixed in repeated samples" is rarely useful in economics. The basic assumption is that the distribution of X satisfies

$$p \lim \left[\frac{X'X}{n} \right] = Q$$

where Q is positive definite, *and* does not depend on parameters.

Our density of observables is $p(y, x)$; usually, we assume that this is $p(y|x)p(x)$ and focus on the first factor. (*Why is this restrictive*?)

Then the ML estimator depends on the conditional distribution.

It is useful to go through the asymptotics applied to the linear model.

Recall that $\hat{\beta} = \beta + (X'X)^{-1}X\varepsilon = \beta + [X'X/n]^{-1}[X/n]\varepsilon$.

If $p\lim\ [X'X/n] = Q$ and $p\lim[X'\epsilon/n] = 0$, then $p\lim \hat{\beta} = \beta$ (i.e., $\hat{\beta}$ is a consistent estimator of β). Recall that if also

$$n^{1/2}[X'\varepsilon/n] \xrightarrow{D} N(0, \sigma^2 Q), \text{ then}$$
$$n^{1/2}[\hat{\beta} - \beta] \xrightarrow{D} N(0, \sigma^2 Q^{-1}).$$

