LECTURE X NOTES

1. Relevant Probability Background.

1.1. *Convergence in probability.* Last class we went through the epsilon, delta definition. A more intuitive definition is: A sequence of r.v.'s $X_n \xrightarrow{p} X$ if

$$P(|X_n - X| > \delta) \text{ as } n \to \infty \text{ for any } \delta > 0$$

1.2. *Convergence in distribution (weak convergence).* There are many different ways to define convergence in distribution. One useful definition is a convergence in expectation: A sequence of r.v.'s $X_n \xrightarrow{d} X$ if

$$E[g(X_n)] \to E[g(X)] \text{ for any bounded and continuous g}$$

When $X_n \in \mathbb{R}$, this is equivalent to the CDF definition:

$$Fn(t) \to F(t) \text{at all t s.t. F is continuous}$$

This definition has difficulties in the multivariate case in which there are many ways to define CDF's.

1.3. *Strength of convergence.* Convergence in probability is stronger than convergence in distribution, meaning that $X_n \xrightarrow{p} X \implies X_n \xrightarrow{d} X$, but the other implication does not hold. One notable case of the reverse implication is when a sequence of r.v.'s converge to a constant: $X_n \xrightarrow{d} C$.

1.4. *Notation.*

1.4.1. *Big O: Stocastic boundedness: .*

$$Xn = O_p(1) \text{ if } P(X_n > M) < \delta \text{ for any} \delta$$

.

In words, we can make the probability that this sequence goes above some bound M arbitrarily small by increasing n far enough.

1.4.2. *Small o: convergence in probability.*

$$Xn = o_p(1) \text{ if } X_n \xrightarrow{p} X$$

.

Furthermore,

$$X_n \sim O_p(B_n) \text{ if } \frac{X_n}{B_n} \sim O_p(1)$$

and,

$$X_n \sim o_p(B_n) \text{ if } \frac{X_n}{B_n} \sim o_p(1)$$

1.5. *Continuous Mapping Theorem.* Let X_n be a sequence of r.v.'s:

1. $X_n \xrightarrow{p} X \implies g(X_n) \xrightarrow{p} g(X)$
2. $X_n \xrightarrow{d} X \implies g(X_n) \xrightarrow{d} g(X)$, for any continuous function g.

1.6. *Slutsky's Theorem.* First you consider the joint convergence of two sequences of r.v., then apply continuous mapping theorem.

Let $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} c$. Then:

1. $X_n + Y_n \xrightarrow{d} X + c$
2. $X_n Y_n \xrightarrow{d} cX$

1.7. *Thm. Delta Method.* First is a description of the theorem, and then a more applicable corollary.

Let a_n be a nonneg sequence of scalars, X_n be a nonneg sequence of r.v., and c be a constant $\in \mathbb{R}$

If:

$$a_n(X_n - c) \xrightarrow{d} X$$

Then for any differentiable function g:

$$a_n(g(X_n) - g(c)) \xrightarrow{d} \nabla g(c)^T X$$

SKIPPED A BUNCH OF PROOF

1.7.1. *Corollary.* Let X_n be asymptotically normal:

$$\sqrt{n}(X_n - c) \xrightarrow{d} N(0, \Sigma)$$

$$\sqrt{n}(g(X_n) - g(c)) \xrightarrow{d} \nabla g(c)^T Z \text{ where Z is } N(0, \Sigma)$$

$$\text{or} \xrightarrow{d} Z \text{ where Z is } N(0, \nabla g(c)^T \Sigma \nabla g(c))$$

2. Large Sample Theory.

2.1. *Standard IID Setup.*

1. $\hat{\theta}_n = \delta(X_1, ..., X_n)$ is an estimator for the true θ^*. Because $\hat{\theta}_n$, depends on samples, it is random. δ is a fixed decision rule that does not depend on number of samples. (An example of δ would be "compute the MLE from the sample").
2. $X_i \sim IIDF_{\theta*}$

2.2. *Consistency.*

Consistency is an elementary criteria for the correctness of an estimator. Technically, consistency applies to a sequence of estimators (indexed by the sample number n), though some people say consistent estimators for short-hand. Informally, consistency means that the sequence of estimators converges to the right thing; it is asymptotically unbiased. Most statistics deals with only consistent estimators, though there are notable exceptions (which ones?)

2.2.1. *Example: Bernoulli Trials.*

Setup:

$$X_i \sim Bern(p*) \text{ IID}$$

We know the MLE:

$$\hat{p}_{mle} = \delta(X_1, ..., X_n) = \frac{1}{n}\sum X_i$$

By weak law of large numbers (WLLN), we know that the sample mean converges to the expected value of the r.v.:

$$\frac{1}{n}\sum Z_i \xrightarrow{p} p*$$

2.2.2. *Example: Sample Max of Uniform.*

Setup:

$$X_i \sim U(0, \theta*) \text{ IID}$$

MLE:

$$\hat{\theta}_{mle} = max(X_i)$$

$$|\hat{\theta}_{mle} - \theta *| = \theta^* - max(X_i)$$

We want to show, convergence in probability.

$$P(\theta^* - max(X_i) > \delta) \to 0$$

$$= P(\cap(X_i < \theta - \delta))$$
$$= \Pi P(X_i < \theta - \delta)$$
$$= (\frac{\theta d}{\theta})^n$$

which converges to 0.

COMMENT

2.3. *Uniform Law of Large Numbers (ULLN.*

2.4. *Proof of the Consistency of the MLE.* First, we will show that θ^* is the unique maximizer of $E_{\theta^*}[l_{X_i}(\theta)]$.

Proof:

$$\text{Want to show: } E_{\theta^*}[l_{X_i}(\theta)] < E_{\theta^*}[l_{X_i}(\theta^*)] \forall \theta \neq \theta^* \tag{2.1}$$
$$\equiv E_{\theta^*}[l_{X_i}(\theta)] - l_{X_i}(\theta^*)] < 0 \tag{2.2}$$
$$\equiv E_{\theta^*}[\log(\frac{f_\theta(X_i)}{f_{\theta^*}(X_i)})] < 0 \tag{2.3}$$
$$E_{\theta^*}[\log(\frac{f_\theta(X_i)}{f_{\theta^*}(X_i)})] < \log E_{\theta^*}[\frac{f_\theta(X_i)}{f_{\theta^*}(X_i)}] \tag{Jensen's Inequality}$$
$$= \log \int_{\mathcal{X}} \frac{f_\theta(X_i)}{f_{\theta^*}(X_i)} f_{\theta^*}(X_i) dX_i = \log \int_{\mathcal{X}} f_\theta(X_i) dX_i = \log 1 = 0 \tag{2.4}$$
$$\text{So, } E_{\theta^*}[\log(\frac{f_\theta(X_i)}{f_{\theta^*}(X_i)})] < 0 \text{ , as desired.} \tag{2.5}$$

Scribe 1
Scribe 2
Scribe 3
Scribe 3
November 3, 2015