Stat210A: Theoretical Statistics **Lecture Date: October 16, 2014**

Asymptotics

Lecturer: Michael I. Jordan *Scribe: Rebecca Roelofs*

1 Consistency of the Maximum Likelihood Estimator

This is a brief introduction to some concepts that will appear in later lectures. We will try to use asymptotics to prove that the maximum likelihood estimator is consistent. We will use the notion of consistency in probability (there is an almost sure definition of consistency, but we will not use that here). The MLE is consistent if

$$\hat{\theta}_n \xrightarrow{P} \theta$$

We will also look at a classical property of the MLE, which is

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow{d} N\left(0, \frac{1}{I(\theta)}\right)$$

More details on this will follow later. We will use Empirical Process Theory to prove this. Empirical Process Theory is the study of these random processes and their convergence properties. For example, in Empirical Process Theory, we might ask: does this sequence of functions converge in the supremum norm to a limiting function? Examples from this literatue include Brownian motion and Donsker's Theorems. We will see this next semester.

The functions that we will look at have the notation $\overline{W}_n{}'(\theta)$. Typically, these functions will be a sum of independent random functions of θ. When we evaluate $\overline{W}_n{}'(\tilde{\theta})$, we are evaluating a random function at a random point, which is where empirical processes arise.

$$\sqrt{n}(\hat{\theta}_n - \theta) = \frac{\sqrt{n}\overline{W}'_n(\theta)}{-\overline{W}_n{}''(\tilde{\theta}_n)}$$

The $\sqrt{n}$ scaling will take that number and turn it into a distribution. Downstairs will converge to a number, not a distribution. (We would need the full empirical process theory if the downstairs thing converged to a distribution).

2 Weak Law for Random Functions

Assume $X_1, X_2, \ldots$ are i.i.d..

Ket K be a compact set. One definition of compactness is that all sequences have a convergent subsequence. In Euclidean space, compactness means bounded and closed. The more modern way of thinking of this is that you can cover the set with a finite number of open balls.

Let $W_i(t) = h(t, X_i),\ t \in K$. Assume $h(t,x)$ is continuous in t, for all x. This then implies that $W_1, W_2, \dots$ are random functions taking values in $C(K)$, the set of continuous functions.

Remark 1. $C(K)$ is a linear space. We can check that it satisfies the properties of a vector space. If we want to talk about topology and convergence ideas, we must have a complete vector space. To talk about completeness we must define a norm. We will complete $C(K)$ using the suprenum norm.

Definition 2 (Supremum Norm)**.** For $w \in C(K)$, the supremum norm of w is defined as

$$||w||_\infty = \sup_{t \in K} |w(t)|.$$

Convergence means $||w_n - w||_\infty \to 0$.

If a set is dense and countable, the space is called separable. In classical Empirical Process theory, some results go out the window if you don't have separability, but then people got around this.

Lemma 3. *[keener's p. 152] Let W be a random function in $C(k)$. Define $\mu(t) = E(W(t)),\ t \in K$ (the mean function). If $E||W||_\infty < \infty$, then μ is continous. Also, $\sup_{t \in K} E \sup_{s:||s-t||<\epsilon} |W(s) - W(t)| \to 0$ as $\epsilon \to 0$.*

Stochastic Equicontinuity is equivalent to the supremum of the supremum statement.

Proof. Let $t_n \to t$ such that $t_n, t \in K$. For each realization of W, $W(t_n)$ will converge to $W(t)$. The $W(t_n)$ are dominated by $||W||_\infty$. By dominated convergence, we have

$$\mu(t_n) = EW(t_N) \overset{\text{DC}}{\to} EW(t) = \mu(t)$$

Let $M_\epsilon(t) = \sup_{s:||s-t||<\epsilon} |W(s) - W(\theta)|$. Let $\lambda_\epsilon(t) = EM_\epsilon(t)$. Since W is continuous, this implies $M_\epsilon(t)$ is continuous. (This is not an obvious step). So $M_\epsilon(t)$ has an upper bound of the form $M_\epsilon(t) \le 2||W||_\infty$. This follows from the triangle inequality and the definition of $M_\epsilon(t)$. Again, by dominating convergence, this implies that λ_ϵ is continuous. Moreover, we can say something pointwise by noting that $M_\epsilon(t) \to 0$ for $t \in K$ as $\epsilon \to 0$. Therefore, dominating convergence will imply that $\lambda_\epsilon(t) \to 0$ as $\epsilon \to 0$.

Typically, we would now use the hammer that is general Empirical Process theory. However, since we have continuous functions, we can use a topological result called Dini's Theorem.

Dini's Theorem allows us to take a pointwise statement and turn it into a uniform statement. We need the set to be compact for this to work.

Using Dini will imply that $\sup_{t \in K} \lambda_\epsilon(t) \to 0$. □

Theorem 4 (Dini's Theorem)**.** *Suppose $f_1 \ge f_2 \ge \dots$ are positive functions in $C(K)$. If $f_n(x) \to 0, \forall x \in K$, then $\sup_{x \in K} f_n(K) \to 0$.*

The proof of Dini's theorem uses the fact that the set is compact so there must be a finite subcover. Then it applies the Union Bound.

Again, Dini's theorem is specialized to continuous functions. In more general Empirical Process theory, we would need to use a counting argument.

Theorem 5 (Uniform Weak Law)**.** *[keener's p. 153] Let $W, W_1, W_2, \dots$ be i.i.d. random functions in $C(K)$, K compact, with mean μ and $E||W||_\infty < \infty$. Let $\overline{W}_n = \frac{1}{n}\sum_{i=1}^n W_i$. This implies $||\overline{W}_n - \mu||_\infty \xrightarrow{P} 0$.*

This is the Weak Law of Large numbers for functions. (You should compare the statement of this theorem to the Weak Law for Random Variables.)

Proof. Let $\epsilon > 0$. Let $M_{\delta,j}(t) = \sup_{s:||s-t||<\delta} |W_j(s) - W_j(t)|$. Let $\lambda_\delta(t) = EM_{\delta,j}(t)$. From the second assertion of Lemma 3, we can choose δ such that

$$\lambda_\delta(t) = E \sup_{s:||s-t||<\delta} |W(t) - W(s)| < \epsilon$$

Thus, if $||s - t|| \leq \delta$,

$$\begin{aligned} |\mu(t) - \mu(s)| &= |E(W(t) - W(s))| && (1) \\ &\leq E|W(t) - W(s)| && (2) \\ &\leq \epsilon && (3) \end{aligned}$$

Now use compactness. Let $B_\delta(t) = \{s : ||s - t|| \leq \delta\}$. Since K is compact, there exists a finite subcover of K. Call this finite subcover $O_i = B_\delta(t_i), i = 1, \ldots, m$. (These finite subcovers are typically referred to meshes or grids on a space.) Then

$$||\overline{W}_n - \mu||_\infty = \max_{i=1,\ldots,m} \sup_{t \in O_i} |\overline{W}_N(t) - \mu(t)|.$$

The next step is to add and subtract certain quantities so we can apply the triangle inequality.

The Wak Law of Large Numbers will imply that

$$\overline{M}_{\delta,n}(t) \xrightarrow{P} \lambda_\delta(t_i) < \epsilon$$

We then have

$$||\overline{W}_n - \mu||_\infty < 2\epsilon + \max_i |\overline{M}_{\delta,n}(t_i) - \lambda_\delta(t_i)| + \max_i |\overline{W}_n(t_i) - \mu(t_i)|$$

Choose n large enough such that $\max_i |\overline{M}_{\delta,n}(t_i) - \lambda_\delta(t_i)| < \epsilon/2$. Using convergence in probability, we can bound the third term. We will bound both of the second terms by $\epsilon/2m$. Then when we take the max, we will use the union bound, which gives us that each term is bounded by $\epsilon/2$.

So,

$$P(||\overline{W}_n - \mu||_\infty) > 3\epsilon) \to 0$$

$\square$

Note that the definition of union bound is $P(\cup A_i) \leq \sum_i P(A_i)$.

Now, we want to get distributions on objects that optimize functions.

Theorem 6. *[keener's p. 154] Let G_n, $n \geq 1$, be random functions in $C(K)$, K compact, and suppose that $||G_n - g||_\infty \xrightarrow{P} 0$, for g a nonrandom, continuous function. (Note that g must be continous because we will not allow discontinous limits.)*

- *If t_n^* are random variables converging in probability to a constant $t^* \in K$, i.e. $t_n^* \xrightarrow{P}$, then $G_n(t_n) \xrightarrow{p} g(t^*)$.*
- *If g achieves its maximum at a unique value t^*, and if t_n are random variables maximizing G_n, that is $G_n(t_n) = \sup_{t\in K} G_n(t)$, Then $t_n \xrightarrow{P} t^*$.*
- *If $K \subset R$, and $g(t) = 0$ has a unique root t^*, and if t_n are R.V.s solving $G_n(t_n) = 0$, then $t_n \xrightarrow{p} t^*$.*

(Read the proof of Theorem 9.4 in keener's book, we will see this next Tuesday).