Stat210B: Theoretical Statistics **Lecture Date: February 27, 2007**

Lecture 12: Stochastic Equicontinuity and Chaining

Lecturer: Michael I. Jordan *Scribe: Guilherme V. Rocha*

1 Stochastic Equicontinuity

Definition 1. Stochastic Equicontinuity:
A collection of stochastic processes $\{Z_n(t)\}$ indexed by $t \in \mathcal{T}$ is said to be *stochastic equicontinuous at* $t_0 \in \mathcal{T}$ if $\forall \eta > 0, \forall \varepsilon > 0$, there exists a neighborhood U_{t_0} of t_0 such that:

$$\limsup_n \mathbb{P}\left(\sup_{t \in U} |Z_n(t) - Z_n(t_0)| > \eta\right) < \varepsilon$$

One application of stochastic equicontinuity will be in proving results of the kind:

Lemma 2. *Suppose $\{Z_n(t)\}$ is stochastically equicontinuous at $t_0 \in \mathcal{T}$. Let τ_n be a sequence of random elements of $\mathcal{T}$ known to satisfy:*

$$\tau_n \xrightarrow{P} t_0$$

It follows that:

$$Z_n(\tau_n) - Z_n(t_0) \xrightarrow{P} 0$$

Proof. Fix $\eta > 0$ and $\varepsilon > 0$. From the stochastic equicontinuity of Z_n, we know that there exists a neighborhood U of t_0 such that:

$$\limsup_n \mathbb{P}\left(\sup_{t \in U} |Z_n(t) - Z_n(t_0)| > \eta\right) < \frac{\varepsilon}{2}$$

Since $\tau_n \xrightarrow{p} t_0$, it follows that

$$\limsup_n \mathbb{P}\left(\tau_n \notin U\right) < \frac{\varepsilon}{2}$$

From the assumptions, we have that:

$$|Z_n(t) - Z_n(t_0)| > \eta \quad \Rightarrow \quad (\tau_n \notin U) \text{ OR } \left(\sup_{t \in U} |Z_n(t) - Z_n(t_0)| > \eta\right)$$

Now, using the union bound on the union of these two events yields the result. □

We now move on to results on asymptotic normality (AN) of M-estimators. We will cover the results in Pollard (1984) based on stochastic equicontinuity.

2 Chaining

We now develop chaining arguments leading to stochastic equicontinuity. Chaining arguments are based on building a multiresolution grid on the space of functions we are interested. We bound the fluctuations of the $Z(\cdot)$ process by controlling the fluctuations along short paths on the grid. To control the fluctuations along the path in the grid, we need to bound the covering integral defined below.

We are given:

- a stochastic process $\{Z(t) : t \in \mathcal{T}\}$;
- a semi-metric on $\mathcal{T} : d(s,t), s,t \in \mathcal{T}$;
- a pointwise exponential inequality: (e.g. the Hoeffding bound)

We want to find conditions on Z_t ensuring that the pointwise inequality can be upgraded to a uniform inequality. One important quantity in getting such results is the covering integral.

Definition 3.

$$J(\delta, d, T) \quad = \quad \int_0^\delta \left[2\log\left(\frac{N(u)^2}{u}\right)\right]^{1/2} du$$

where $N(\delta)$ is the smallest integer m such that there exist points $t_1, \ldots, t_m$ such that $\min_{1\le i\le m} d(t,t_i) \le \delta, \forall t \in T$. Here we implicitly restrict $t_1, t_2, ..., t_m$ to be points of T, which is different from the definition of covering number in earlier lectures.

Our next lemma establishes that boundedness of the covering integral $J(\delta, d, T)$ is sufficient to ensure that the difference between points close to one another are unlikely to be larger than a quantity related to the covering integral.

Lemma 4. ***Pollard (1984), Lemma 9***
Suppose that $J(\delta, d, T) < \infty$ and there exists D such that:

$$\mathbb{P}\left(|Z_n(s) - Z_n(t)| > \eta \cdot d(s,t)\right) \le 2\exp\left(-\frac{\eta^2}{2D^2}\right).$$

Then:

$$\mathbb{P}\left(|Z_n(s) - Z_n(t)| > 26DJ(\varepsilon, d, T) \text{ for some } s,t \in T \text{ with } d(s,t) < \varepsilon\right) < \varepsilon$$

Proof. Let $\delta_i = \frac{\varepsilon}{2^i}$ and define:

$$H(u) \triangleq \left(2\log\left(\frac{N(u)^2}{u}\right)\right)^{\frac{1}{2}}.$$

Now, construct a $2\delta_i$-net by following these steps:

1. Pick an arbitrary $t_1 \in T$;
2. For k going from 2 to $m_1 = N(\delta_1)$:
   (a) pick t_k such that $d(t_k, t_j) > 2\delta_1$ for all $j < k$;

3. Let $T_1 = \{t_1, \ldots, t_{m_1}\}$;
4. Pick t_{m_1+1} such that $d(t_{m_1}, t_j) > 2\delta_2$ for all $j < m_1 + 1$;
5. For k going from $m_1 + 2$ to $m_1 + m_2$ with $m_2 = N(\delta_2)$:
   (a) pick t_k such that $d(t_k, t_j) > 2\delta_2$ for all $j < k$;
6. Let $T_2 = \{t_{m_1}, \ldots, t_{m_1+m_2}\}$;
7. $\vdots$
8. Pick $t_{\sum_{j=1}^{l-1} m_j}$ such that $d(t_{m_1}, t_j) > 2\delta_l$ for all $j < (\sum_{j=1}^{l-1} m_j) + 1$;
9. For k going from $(\sum_{j=1}^{l-1} m_j) + 2$ to $(\sum_{j=1}^{l-1} m_j) + m_l$ with $m_l = N(\delta_l)$:
   (a) pick t_k such that $d(t_k, t_j) > 2\delta_2$ for all $j < k$;
10. Let $T_l = \{t_{(\sum_{j=1}^{l-1} m_j)+1}, \ldots, t_{(\sum_{j=1}^{l} m_j)}\}$;
11. $\vdots$
12. Let $T^* = \cup_i T_i$;

We now define the set A_i on which something "bad" happens at scale i, i.e., a set where the observed difference between $Z(s) - Z(t)$ is large for a pair of points on the grid at scale δ_i:

$$A_i \quad = \quad \{\omega \in \Omega : |Z(\omega, s) - Z(\omega, t)| > Dd(s,t)H(\delta_i) \text{ for some } s, t \in T_i\}.$$

Now, notice that A_i is the sum of at most $N(\delta_i)^2$ events whose probabilities can be controlled using the pointwise bound and conclude:

$$\mathbb{P}(A_i) \quad \leq \quad 2N(\delta_i)^2 \exp\left(-\frac{1}{2}H(\delta_i)^2\right) = 2\delta_i$$

It follows that:

$$\mathbb{P}(\cup_{i=1}^{\infty} A_i) \quad \leq \quad \sum \mathbb{P}(A_i) = 2\varepsilon$$

Now, we want to extend this result from the points in the grids T^* to the entire set T.

Now, let $s, t \in \mathcal{T}$ be such that $d(s,t) < \varepsilon$. Find n such that $\delta_n \leq d(s,t) \leq 2\delta_n$. Now link $s = s_{m+1}, s_m, \ldots, s_n$ such that $s_m \in T_m$ choosing the closest point at each step. By construction, $d(s_i, s_{i+1}) \leq 2\delta_i$.

Similarly, build $\{t_n, \ldots, t_m, t_{m+1}\}$ for $t = t_{m+1}$. Now, using the triangular inequality:

$$|Z(s) - Z(t)| \leq |Z(s_n) - Z(t_n)| + \sum_{i=n}^{m} \Big[|Z(s_{i+1}) - Z(s_i)| + |Z(t_{i+1}) - Z(t_i)| \Big]$$

Now, on A_i^c, we have:

$$|Z(s_{i+1}) - Z(s_i)| \leq D \cdot d(s_{i+1}, s_i) \cdot H(\delta_{i+1}) \leq 2D \cdot \delta_{i+1} \cdot H(\delta_{i+1})$$

which substituting this back into the inequality above yields that on $(\cup_{i=1}^{\infty} A_i)^c$:

$$|Z(s) - Z(t)| \leq D \cdot d(s_n, t_n) \cdot H(\delta_n) + 2 \sum_{i=n}^{m} 2D\delta_i H(\delta_{i+1})$$

The distance along the chain is such that:

$$\begin{aligned} d(s_n, t_n) &\leq d(s,t) + \sum_{i=n}^{m} \Big(d(s_{i+1}, s_i) + d(t_{i+1}, t_i) \Big) \\ &\leq 2\delta_n + 4 \sum_{i=n}^{\infty} \delta_i \\ &\leq 10\delta_n \end{aligned}$$

As a result, since $\delta_i = 4(\delta_{i+1} - \delta_{i+2})$, we have on $(\cup_{i=1}^{\infty} A_i)^c$:

$$\begin{aligned} |Z(s) - Z(t)| &\leq 10D\delta_n H(\delta_n) + 4D \sum_{i=n}^{\infty} 4(\delta_{i+1} - \delta_{i+2}) H(\delta_i) \\ &\leq 10D\delta_n H(\delta_n) + 16D \sum_{i=n}^{\infty} \int \mathbf{I}(\delta_{i+2} \leq u \leq \delta_{i+1}) H(u) du \\ &\leq 10D\delta_n H(\delta_n) + 16D \cdot J(\delta_{n+1}) \\ &\leq 10D\delta_n H(\delta_n) + 16D \cdot J(\delta_{n+1}) \\ &\leq 26DJ\Big(d(s,t)\Big) \end{aligned}$$

□

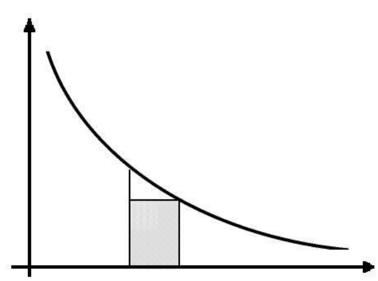

Figure 1: Pictorial argument for $(\delta_{i+1} - \delta_{i+2})H(\delta_{i+2}) \leq \int \mathbf{I}(\delta_{i+1} - \delta_{i+2})H(u)du$

3 Symmetrization, Equicontinuity and Chaining

Recall we constructed P_n^0 as a signed measure assigning mass $\pm n^{-1}$ to each of the observed points $\xi_1, \xi_2, \ldots, \xi_n$. We will now define rescaled version of P_n and P_n^0 as:

$$\begin{aligned} E_n f &= \sqrt{n} P_n f = \sqrt{n} \left[\frac{1}{n} \sum_{i=1}^{n} f(X_i) \right] \\ E_n^0 &= \sqrt{n} P_n^0 \end{aligned}$$

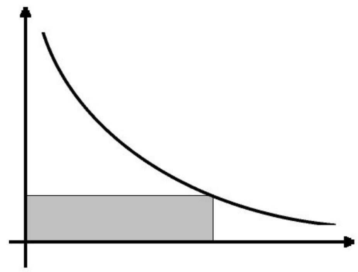

Figure 2: Pictorial argument for $\delta H(\delta) \leq \int_0^\delta H(u)du$

Let E denote a Brownian Bridge process and let $Ef = \int f(x)dE(x)$. We will be extending the theory to establish conditions on a class of functions $\mathcal{F}$ so convergence $E_n f = \sqrt{n}P_n f$ to Ef is obtained for every $f \in \mathcal{F}$.

Coming up next

In the coming classes, we will be covering:

- Pollard (1984), chapter 7
- Pollard (1984), Theorem 13
- Pollard (1984), Lemma 15
- Pollard (1984), Theorem 21: Use Stochastic Equicontinuity to get $E_n \xrightarrow{\mathcal{L}} E$

References

Pollard, D. (1984). *Convergence of Stochastic Processes.* Springer, New York.