27/10: L9 – TIGHTNESS CRITERION.

What techniques are available to show

$G_n \xrightarrow{\text{LWC}} (G, o)$? Analysis of BFS.

Some general metric space techniques?

Let X_n be random elements in a m. space (S, d).

We can talk of $X_n \xrightarrow{d} X$ or $\mathbb{P}_n \longrightarrow \mathbb{P}$

where $\mathbb{P}_n(\cdot) = \mathbb{P}(X_n \in \cdot)$.

How to show convergence of a deterministic sequence say $(x_n)_{n \geq 1}$?

– S.T. for any subsequence $\{n_k\}_{k\geq 1}$ $\exists$ a further subsequence $\{n_k'\}_{k \geq 1}$ $\ni$ $x_{n_k'} \to x$ as $n \to \infty$.

Equivalently

– S.T. any subsequence has a convergent subsequence (Relative compactness)

– S.T. subsequential limits are unique (Uniqueness).

Now we develop such a criterion for $\mathbb{P}_n$'s or X_n's.

LEMMA: If for any sequence $(n_k)_{k\geq 1}$, $\exists$ a further subsequence $n_{k(r)} \ni \mathbb{P}_{n_{k(r)}} \xrightarrow{d} \mathbb{P}$ then $\mathbb{P}_n \xrightarrow{d} \mathbb{P}$

(Here $\mathbb{P}_n$ & $\mathbb{P}$ are prob. measures.)

Proof: If not true, then for some $f \in C_b(S)$ & $\varepsilon > 0$ $\exists$ a subsequence $n_k \ni \left| \int f \, d\mathbb{P}_{n_k} - \int f \, d\mathbb{P} \right| > \varepsilon$.

a $\Rightarrow\Leftarrow$ to $\exists$ a subsequence of $n_k \ni \mathbb{P}_{n_{k(r)}} \xrightarrow{d} \mathbb{P}$. ∎

To prove $\mathbb{P}_n \xrightarrow{d} \mathbb{P}$, suffices to show two things.

(1) S.T. for any sequence $\mathbb{P}_{n_k}$ $\exists$ a converging subsequence $\mathbb{P}_{n_{k(r)}}$

(2) S.T. all subsequential limits are equal.

(1) — relative compactness. $\exists$ good criteria to show (1).

(2) — Often Ad-hoc.

Def: (Tightness): A family of prob. measures on (S,d) $(\mathbb{P}_i)_{i \in I}$ (I - index set, arbitrary) is said to be tight if for any $\varepsilon > 0$ $\exists$ a compact $K \subseteq S$

$$\ni \sup_{i \in I} \mathbb{P}_i(K^c) < \varepsilon.$$

Of course $\{\mathbb{P}\}$ is tight if $\exists$ a sequence of cpt sets $K_n \uparrow S$.

Then $\mathbb{P}(K_n) \uparrow \mathbb{P}(S) = 1 \Rightarrow \exists\, n \ni \mathbb{P}(K_n) \geq 1 - \varepsilon$

$\Updownarrow$

$\mathbb{P}(K_n^c) \leq \varepsilon$.

SELECTION THM: If $(\mathbb{P}_n)_{n \geq 1}$ is a tight sequence of prob. measures on (S,d), then $(\mathbb{P}_n)_{n\geq 1}$ is rel. compact i.e.,

$\forall$ subsequences n_k, $\exists$ a further subsequence $n_k' \ni$

$\mathbb{P}_{n_k'} \xrightarrow{d} \mathbb{P}$ for some $\mathbb{P}$. ($\mathbb{P}$ can depend on subsequence)

Cantor's Diagonalization: A countable set & $f_n : A \to \mathbb{R}$ $\forall\, n \geq 1$.

Then $\exists$ a subsequence $n_k \ni f_{n_k}(a) \longrightarrow f(a) \in [-\infty, \infty]$ $\forall\, a \in A$.

Proof: $(S=\mathbb{R})$. [Proof for General case, See DP THM 3.2 or wait for Stoch. Process course!]

Let A be a countable dense subset of $\mathbb{R}$. Consider the CDFs F_n for $\mathbb{P}_n$.

By diagonalization, $\exists$ a subsequence $n_k \ni F_{n_k}(a) \to F(a)\ \forall a \in A$.

Here F is some fn. To complete the proof, we've to show that F is a CDF & convergence holds at all cty points of F.

$$\bar{F}(x) = \inf\{F(a) : a > x,\ a \in A\}$$

By construction, F is non-decreasing on A & so is $\bar{F}$ on $\mathbb{R}$.

$$\bar{F}(x) = \inf\{F(a) : x < a, a \in A\} = \inf_{y > x} \inf\{F(a) : a > y, a \in A\}$$
$$= \inf\{\bar{F}(y) : y > x\}$$

$\Rightarrow \bar{F}$ is right cts.

Tightness $\Rightarrow \exists\ a_1, a_2 \in A \ni \mathbb{P}_n((a_1, a_2]) \geq 1-\varepsilon \quad \forall n \geq 1$

$$\mathbb{P}_n((a_1, a_2]) = F_n(a_2) - F_n(a_1)$$

$\Rightarrow \quad \bar{F}(a_2) - \bar{F}(a_1) \geq 1-\varepsilon. \Rightarrow \quad \lim_{x\to\infty} \bar{F}(x) - \lim_{x \to -\infty} \bar{F}(x) = 1$

Since $\bar{F}$ non-decreasing, limits exist & $\bar{F} \in [0,1]$

$$\Rightarrow \lim_{x\to\infty} \bar{F}(x) = 1 = 1 - \lim_{x\to-\infty} \bar{F}(x).$$

$\Rightarrow \bar{F}$ is a CDF.

If x is a pt of cty of $\bar{F}$ then $\lim_{\substack{a \uparrow x \\ a \in A}} F(a) = \bar{F}(x)$ & $\lim_{\substack{b \downarrow x \\ b \in A}} F(b) = \bar{F}(x)$. — (1)

Let x be a cty pt of $\bar{F}$ & choose $a < x < b$, $a, b \in A$.

$$F(a) = \lim_{k\to\infty} F_{n_k}(a) \leq \underline{\lim}_{k\to\infty} F_{n_k}(x) \leq \overline{\lim}_{k\to\infty} F_{n_k}(x)$$

(non-decreasing of F_{n_k}'s)

$$\leq \overline{\lim}_{k\to\infty} F_{n_k}(b) = F(b).$$

(non-decreasing)

From (1), we get $\overline{\lim} F_{n_k}(x) = \underline{\lim} F_{n_k}(x) = \bar{F}(x)$ ✓ for $S = \mathbb{R}$.

General S (proof sketch):

From Stone-Weierstrass thm, $C_b(K)$ is separable for a cpt set K under supnorm.

$C_b(K) = C(K)$, cts fns on K.

$\mathbb{P}_n$ are tight $\Rightarrow$ $\forall\ r \geq 1$, $\exists$ cpt K_r $\ni$

$\mathbb{P}_n(K_r) > 1 - \frac{1}{r}$ $\forall\ n \geq 1$.

Let $C_r \subset C(K_r)$ be a countable dense subset.

By diagonalization, $\exists$ a subsequence n_k $\ni$

$$\int f\,d\mathbb{P}_{n_k} \longrightarrow \int f\,d\mathbb{P}^r \quad \forall\ f \in C_r.$$

Since C_r is dense, $\int f\,d\mathbb{P}_{n_k} \longrightarrow \int f\,d\mathbb{P}^r$ $\forall\ f \in C(K_r)$

$$\left[\ \left|\int f\,d\mathbb{P} - \int g\,d\mathbb{P}\right| \leq \int |f-g|\,d\mathbb{P} \leq \|f-g\|_\infty\ \right]$$

$\forall\ f \in C_b(S)$

$$\left|\int f\,d\mathbb{P}_{n_k} - \int_{K_r} f\,d\mathbb{P}_{n_k}\right| \leq \int_{K_r^c} f\,d\mathbb{P}_{n_k} \leq \|f\|_\infty \mathbb{P}_{n_k}(K_r^c) \leq \frac{\|f\|_\infty}{r}$$

$\Rightarrow$ $\exists$ a limit $I(f) := \lim_{k\to\infty} \int f\,d\mathbb{P}_{n_k}$

$$\Big[\ \because \left|\int f\,d\mathbb{P}_{n_k} - \int f\,d\mathbb{P}_{n_l}\right| \leq \left|\int f\,d\mathbb{P}_{n_k} - \int_{K_r} f\,d\mathbb{P}_{n_k}\right| + \left|\int_{K_r} f\,d\mathbb{P}_{n_k} - \int_{K_r} f\,d\mathbb{P}_{n_l}\right| + \left|\int_{K_r} f\,d\mathbb{P}_{n_l} - \int f\,d\mathbb{P}_{n_l}\right|$$

$$\leq \frac{2\|f\|_\infty}{r} + \left|\int_{K_r} f\,d\mathbb{P}_{n_k} - \int_{K_r} f\,d\mathbb{P}_{n_l}\right|\ \Big]$$

To complete the proof, we need to show that $\exists$ a $\mathbb{P}$ $\ni$ $I(f) = \int f\,d\mathbb{P}$. One can show that I is a pos lin. fnl & use Riesz representation thm $I(f \mathbb{1}_{K_r}) = \int_{K_r} f\,d\mathbb{P}^r$ ∎

Prohorov's theorem.

(S,d) is a Polish space. (X_n) is rel. compact $\Rightarrow$ $(X_n)_{n\geq 1}$ is tight.

THM (THM 2.6 of vdH-2).

Let G_n be a sequence of graphs on $[n]$ & O_n the random root chosen uniformly. If $(d^{(n)}_{O_n})$ is uniformly integrable then $(G_n, O_n)_{n\geq 1}$ is tight.

$\left(\lim_{K\to\infty} \overline{\lim_{n\to\infty}} \mathbb{E}\left[d^{(n)}_{O_n} \mathbb{1}[d^{(n)}_{O_n} \geq K]\right] = 0 \right)$

LEMMA (Compactness:)

Let $h: \mathbb{N} \to \mathbb{N}$ be an increasing fn.

Then $\mathcal{K}_h = \{ (G,o) : |B^G_r(o)| \leq h(r) \ \forall r \geq 1 \}$ is compact.

Proof: $\forall r \geq 1$, $\exists$ finitely many equivalence classes of rooted graphs $F_{r,1}, \ldots, F_{r,n_r}$ $\ni$ $|B^{F_{r,i}}_r(o)| \leq h(r)$ i.e., if $F \ni$ $|B^F_r(o)| \leq h(r)$ then $B^F_r(o) \cong F_{r,i}$ for some $1 \leq i \leq n_r$.

$\Rightarrow$ $\mathcal{K}_h \subseteq \bigcup_{i=1}^{n_r} \{ (G,o) : d_{\mathcal{G}_*}((G,o), F_{r,i}) \leq \frac{1}{r+1} \}$ ∎

Proof of thm: $\mathcal{K}_h^c = \bigcup_{r\geq 1} \{ (G,o) : |B^G_r(o)| > h(r) \}$

$\subseteq \bigcup_{r\geq 1} \{ (G,o) : \exists v \in B^G_r(o) \ni d_v > M_r \}$

$= \bigcup_{r\geq 1} F^M_r(o)$ for $M: \mathbb{N} \to \mathbb{N}$ increasing fn.

ETPT $\forall \varepsilon > 0$ & $\forall r \geq 1$ $\exists M < \infty \ni \sup_n \mathbb{P}((G_n, O_n) \in F^M_r) \leq \varepsilon$ — (1)

$\Rightarrow$ choose $\frac{\varepsilon}{2^r}$ for $r \geq 1$; $\exists M < \infty \ni \sup_n \mathbb{P}((G_n, O_n) \in F^M_r) \leq \frac{\varepsilon}{2^r}$

$\sup_n \mathbb{P}((G_n, O_n) \in \mathcal{K}^c_{h}) \leq \varepsilon \Rightarrow$ Tightness of (G_n, O_n).

Let $f(d) = \sup_n \mathbb{E}[d^{(n)}_{O_n} \mathbb{1}[d^{(n)}_{O_n} > d]]$

U.I. of $d^{(n)}_{O_n}$ $\Rightarrow$ $\lim_{d\to\infty} f(d) = 0$.

$m(G_n) = \mathbb{E}[d^{(n)}_{O_n}] \leq f(0) < \infty$; $m(G_n) = \frac{1}{n}\sum_{v\in[n]} d_{G_n}(v)$.

$m(G_n) \geq 1$ for $(G_n, o_n) \in \mathcal{G}_{\neq}$ (assume (G_n, o_n) has no isolated nodes)

Define $G_n^* \in \mathcal{G}_*$ $\ni$

$$\mathbb{P}(G_n^* = (G_n, v)) = \frac{d_{G_n}(v)}{m(G_n)} \mathbb{P}((G_n, o_n) = (G_n, v)) = \frac{d_{G_n}(v) \setminus n}{m(G_n)}$$

Biasing the prob. measure by the degree.

$$\mathbb{P}((G_n, o_n) = \cdot) \leq m(G_n) \mathbb{P}(G_n^* = \cdot) \leq f(\cdot) \mathbb{P}(G_n^* = \cdot) \quad -(2)$$

$\Rightarrow$ if $(G_n^*)_{n \geq 1}$ is tight then so is (G_n, o_n).

Further $\mathbb{P}(o_n^* = v) = \dfrac{d_{G_n}(v)}{n\, m(G_n)} = \dfrac{d_{G_n}(v)}{\sum_{u \in [n]} d_{G_n}(u)}$

$$\Rightarrow \mathbb{E}\left[d_{o_n^*}^{(n)} \mathbb{1}[d_{o_n^*} \geq K]\right] = \sum_{v \in [n]} d_{G_n}(v) \mathbb{1}[d_{G_n}(v) \geq K] \mathbb{P}(o_n^* = v)$$

$$= \frac{1}{n} \sum_{v \in [n]} \frac{d_{G_n}(v)^2 \mathbb{1}[d_{G_n}(v) \geq K]}{m(G_n)}$$

$$= \frac{\mathbb{E}\left[(d_{o_n}^{(n)})^2 \mathbb{1}[d_{o_n}^{(n)} \geq K]\right]}{\mathbb{E}[d_{o_n}^{(n)}]}$$

(Ex)

$\Rightarrow d_{o_n^*}^{(n)}$ is U.I.

From (2) ETPT, G_n^* is tight. T.S.T, G_n^* satisfies (1)

$\mathbb{P}(o_n^* = \cdot)$ – Stationary measure for SRW on G_n.

$\Rightarrow X_n$ = Unif. neighbour of $o_n^* \stackrel{d}{=} o_n^*$

$$\mathbb{P}(X_n = v) = \sum_{u \in N_v} \underbrace{\frac{d_{G_n}(u)}{n\, m(G_n)} \times \frac{1}{d_{G_n}(u)}}_{\mathbb{P}(o_n^* = u) \times \mathbb{P}(X_n = v \mid o_n^* = u)} = \frac{d_{G_n}(v)}{n\, m(G_n)}$$

$$\Rightarrow (G_n, o_n^*) \stackrel{d}{=} (G_n, X_n)$$

$$\mathbb{P}((G_n, o_n^*) \in F_{r+1}^M) \leq \mathbb{P}(d_{o_n^*}^{(n)} > d)$$
$$+ \mathbb{E}\left[\mathbb{1}[d_{o_n^*}^{(n)} < d] \, \mathbb{P}[(G_n, o_n^*) \in F_{r+1}^M \mid d_{o_n}^{(n)}]\right]$$

By U.I., $\exists d$ $\ni \forall n$

$$\leq \frac{\varepsilon}{2} + \mathbb{E}\left[d\, \mathbb{1}[d_{o_n^*}^{(n)} < d]\, \mathbb{P}((G_n, X_n) \in F_r^M \mid d_{o_n}^{(n)})\right]$$

$$\left\{ (G_n, o_n^*) \in F_{r+1}^M \implies d_{G_n}(v) > M \text{ for some } v \in B_{r+1}(o_n^*) \right.$$

$$\left. \implies d_{G_n}(v) > M \text{ for some } v \in B_r(u) \text{ where } u \sim o_n^* \right]$$

$$\leq \frac{\varepsilon}{2} + d\, \mathbb{P}((G_n, X_n) \in F_r^M)$$

$$= \frac{\varepsilon}{2} + d\, \mathbb{P}((G_n, o_n^*) \in F_r^M) \qquad [(G_n, X_n) \stackrel{d}{=} (G_n, o_n^*)]$$

Proof complete by induction.

A family of random variables $\{X_i\}_{i\in I}$ is called *uniformly integrable (u.i.)* if for any given $\epsilon > 0$, there exists A large enough so that

$$\sup_{i\in I} \mathbb{E}\left(|X_i| \mathbf{1}[|X_i| > A]\right) < \epsilon.$$

Show that $\{X_i\}_{i\in I}$ is u.i. iff $\sup_{i\in I} \mathbb{E}[|X_i|] < \infty$ (i.e., uniformly bounded in L^1) and also that for any $\epsilon > 0$, there exists $\delta > 0$ such that for any measurable A with $\mathbb{P}(A) < \delta$ implies that $\sup_{i\in I} \mathbb{E}[|X_i| \mathbf{1}_A]] < \epsilon$.

If degree of the root has bounded pth moments for p > 1 then it is uniformly integrable.

$$\mathbb{E}\left[|X|\, \mathbb{1}[|X| \geq A]\right] \leq \mathbb{E}[|X|^p]^{1/p}\, \mathbb{P}(|X| \geq A)^{1/q}$$

$$\frac{1}{p} + \frac{1}{q} = 1.$$

- If $d_{o_n}^{(n)}$ has bdd p^{th} moments for some $p > 1$, then (G_n, o_n) is tight i.e., has convergent subsequences.

$$\mathbb{E}[d_{o_n}^{(n)\,p}] = \sum k^p\, p_{G_n}(k) \; ; \; p_{G_n}(k) = \frac{\#\{v : d_v = k\}}{n}$$

often $p_{G_n}(k) \to p_G(k)$ & $\mathbb{E}[X_n^p] \to \sum k^p\, p_G(k)$. $\to$ long.

Defn: (estimable functions). Let $\mathcal{G}$ be the set of fin. unlabeled graphs & $\mathcal{H} \subseteq \mathcal{G}$. A fn $f : \mathcal{G} \to \mathbb{R}$ is ESTIMABLE over $\mathcal{H}$ if for any $G_n \in \mathcal{H}$ $\ni G_n \xrightarrow{LWd} (G, o)$ we have that $f(G_n) \to \mathbb{E}\hat{f}(G, o)$ for some $\hat{f} : \mathcal{G}_* \to \mathbb{R}$.

Eg. $\varphi : \mathcal{G}_* \to \mathbb{R}$ bdd cts.

& $f(G) = \frac{1}{|V|} \sum_{v \in V} \varphi(G, v)$

$$f(G_n) = \mathbb{E}\varphi(G_n, o_n) \to \mathbb{E}\varphi(G, o)$$

$\Rightarrow$ f is estimable over $\mathcal{G}$

Eg. $f(G) = \frac{1}{|V|} \sum_{\substack{S \subseteq V \\ |S| = p}} \mathbb{1}\{G(S) \cong H\}$, H - transitive graph on p vertices.

$G(S)$ - G restricted to the subset S.

$$= \frac{1}{p|V|} \sum_{v \in V} \varphi(G, v)$$

$\varphi(G, v)$ = # of subgraphs rooted at v & isomorphic to H^* where $H^* = (H, o)$ for an arbitrary vertex o

[Since H is transitive, we can choose any vertex $o \in H$ as its root].

= # of subgraphs isomorphic to H & containing v.

If H is connected then $\varphi(G, v)$ is cts. (Ex)

$$f(G_n) = \frac{1}{p}\mathbb{E}[\varphi(G_n, o_n)]$$

(Ex). $\to \frac{1}{p}\mathbb{E}[\varphi(G, o)]$ if $\{\varphi(G_n, o_n)\}_{n \geq 1}$ is U.I.

$$\mathcal{G}_d = \{ G \in \mathcal{G} : \max_{v \in V} d_G(v) \le d \}$$

$\Rightarrow$ $\varphi(G, v)$ is bounded $\forall$ $G \in \mathcal{G}_d$

$\Rightarrow$ $f(G_n) \longrightarrow \frac{1}{p} \mathbb{E}[\varphi(G, o)]$ $\forall$ $G_n \subset \mathcal{G}_d$.

$\Rightarrow$ $f(G_n)$ is estimable over $\mathcal{G}_d$.

Eg. $f(G) = \frac{1}{|V|} \log c_k(G)$, $c_k(G) = \#$ proper k-colorings of G with $k > 2d$.

is estimable over $\mathcal{G}_d$.

$f(G) = \frac{1}{|V|} \log t(G)$ $t(G) = \#$ spanning trees of G.

is estimable over conn. graphs.

Ex: What if H isn't transitive?