Author: Lucas (Zheng) Zhang[1]
Empirical Process "Cheat" Sheet

1. *Weak Convergence*

1.1 Probability Space

$(\Omega, \mathcal{F}, P)$ is a probability space where: Ω denotes the outcome space, $\mathcal{F}$ is the σ-algebra on Ω, and P is a probability measure on $\mathcal{F}$. Example: $(\Omega, \mathcal{F}, P)$

1.2 Outer and Inner Expecation[2]

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $T : \Omega \to \mathbf{R}$. The outer expecation of T is defined as:
$\mathbb{E}^*[T] := \inf\{\mathbb{E}[U] : U \geq T,\ U : \Omega \to \mathbf{R} \text{ measurable},\ \mathbb{E}[U\mathbf{1}\{U \geq 0\}] < \infty \text{ or } \mathbb{E}[U\mathbf{1}\{U \leq 0\}] < \infty\}$,
and the inner expecation is $\mathbb{E}_*[T] = -\mathbb{E}^*[-T]$, where the expectation is defined as
$\mathbb{E}[U] := \int U(\omega) dP(\omega)$.

1.3 Outer and Inner Probability

$P^*(B) := \mathbb{E}^*[\mathbf{1}_B]; \quad P_*(B) := \mathbb{E}_*[\mathbf{1}_B]$

1.4 Notations

- $\mathbb{D}$: complete metric space. Eg. $(L^\infty(\mathbf{R}), \|\cdot\|_\infty)$ i.e. sup norm bounded functions on $\mathbf{R}$.
- $\mathcal{C}_b(\mathbb{D})$: $f : \mathbb{D} \to \mathbf{R}$ continuous and bounded.
- $L^\infty(T)$: $f : T \to \mathbf{R}$ measurable and bounded (i.e. $\sup_t |f(t)| < \infty$)

1.5 Tight Measure

A measure P is tight if $\forall\ \epsilon > 0$, $\exists$ compact K s.t. $P(K) \geq 1 - \epsilon$.

1.6 Borel Law

For a random variable[3] $X : (\Omega, \mathcal{F}, P) \to \mathbb{D}$, define the Borel law induced by P as $L := P \circ X^{-1}$, i.e. $\forall A \in \mathcal{B}(\mathbb{D})$, $L(A) := \int_A dL = P(X \in A) := P(\omega \in \Omega : X(w) \in A)$.

1.7 Weak Convergence

Let $(\Omega_n, \mathcal{F}_n, P_n)$ be a sequence of probability spaces and $X_n : \Omega_n \to \mathbb{D}$ be arbitrary (don't need to be measurable) maps. Then we say $X_n \xrightarrow{L} X$ if $\mathbb{E}^*[f(X_n)] \to \mathbb{E}[f(X)]$ for all $f \in \mathcal{C}_b(\mathbb{D})$.

1.8 Portmanteau TFAE:

(i) $X_n \xrightarrow{L} X$.
(ii) $\liminf P_*(X_n \in G) \geq P(X \in G)$ for all open G.
(iii) $\limsup P^*(X_n \in F) \leq P(X \in F)$ for all closed F.
(iv) $\lim P^*(X_n \in B) = \lim P_*(X_n \in B) = P(X \in B)$ for all $B \in \mathcal{B}(\mathbb{D})$ s.t. $P(X \in \partial B) = 0$

1.9 Continuous Mapping Theorem (CMT):

Let $g : \mathbb{D} \to \mathbb{F}$ be continuous at all $x \in \mathbb{D}_0 \subset \mathbb{D}$. If $X_n \xrightarrow{L} X$ and $P(X \in D_0) = 1$, then $g(X_n) \xrightarrow{L} g(X)$.

1.10 Important Example of CMT (the reason why we do Empirical Process):

Eg. Suppose $G_n, G : \mathbf{R} \to L^\infty(\mathbf{R})$ and $G_n \xrightarrow{L} G$. Classic example $G_n(t) = \frac{1}{\sqrt{n}} \sum(\mathbf{1}(X_i \leq t) - P(X \leq t))$. Since sup operation is continuous, CMT will then imply $\sup_t |G_n(t)| \xrightarrow{L} \sup_t |G(t)|$. If we know the distribution of G, then we can build a confidence interval.

[1] This writing is solely based on the notes provided by Andres Santos and additional materials from *Weak Convergence and Empirical Processes* by Van Der Vaart, A. W., & Wellner, J. A. (1996). Only important results are presented and *most* proofs are omitted.
[2] Upper and Lower integral if familiar with real analysis.
[3] The measurability of X is implicit here.

1.11 Asymptotically Tight

Suppose $X_n : (\Omega_n, \mathcal{F}_n, P_n) \to (\mathbb{D}, d)$. Then we say $\{X_n\}$ is asymptotically tight if $\forall \epsilon > 0$, $\exists$ compact K s.t. $\liminf P_n(X_n \in K^\delta) \geq 1 - \epsilon \ \forall \delta > 0$, where $K^\delta := \{y \in \mathbb{D} : d(y, K) < \delta\}$.

1.12 Asymptotically Measurable

Suppose $X_n : (\Omega_n, \mathcal{F}_n, P_n) \to (\mathbb{D}, d)$. Then we say $\{X_n\}$ is asymptotically measurable if $\mathbb{E}^*[f(X_n)] - \mathbb{E}_*[f(X_n)] \to 0$ for all $f \in \mathcal{C}_b(\mathbb{D})$.

1.13 "Inheritance" Lemma[4]

(i) If $X_n \xrightarrow{L} X$, then X_n is asymptotically measurable. (Implicitly here X is measurable.)

(ii) If $X_n \xrightarrow{L} X$, then X_n is asymptotically tight if and only if X is tight.

Blanket assumption: whenever we write $X_n \xrightarrow{L} X$, we assume that X is measurable and tight.

1.14 Prohorov's Theorem

If $\{X_n\}$ is asymptotically tight and asymptotically measurable, then it has a subsequence $\{X_{n_j}\}$ that converges weakly to a tight Borel measure.

1.15 Marginal

If $X_n : \Omega \to L^\infty(T)$, then we say $X_n(t)$ is the marginal. (To facilitate understanding, for each $\omega \in \Omega$, $X_n(\omega) : T \to \mathbf{R} \in L^\infty(T)$, which is no longer random; on the other hand, for each $t \in T$, $X_n(t) : \Omega \to \mathbf{R}$ will be a random variable.)

1.16 L-infinity Lemma 1

Let $X_n : \Omega \to L^\infty(T)$ be asymptotically tight. Then it is asymptotically measurable if and only if $X_n(t)$ is asymptotically measurable $\forall t \in T$.

1.17 L-infinity Lemma 2

Let $X, Y : \Omega \to L^\infty(T)$ be tight Borel random variables. Then X, Y have the same distribution (Borel law) if and only if $X(t), Y(t)$ have the same distribution (Borel law) for all t.

1.18 Marginal + Tight Theorem

Let $X_n : \Omega_n \to L^\infty(T)$ be arbitrary. Then X_n converges in distribution to a tight limit if and only if X_n is asymptotically tight and the marginals $(X_n(t_1), \cdots, X_n(t_k))$ converges weakly to a limit for every finite subset $(t_1, \cdots, t_k)$.

1.19 Tight Verifying Theorem (Finite Approximation)

$X_n : \Omega_n \to L^\infty(T)$. $\{X_n\}$ is asymptotically tight if and only if $X_n(t) : \Omega_n \to \mathbf{R}$ is asymptotically tight $\forall t$ and $\forall \epsilon, \eta > 0$, $\exists$ a finite partition $T = \cup_{i=1}^k T_i$ such that

$$\limsup_{n\to\infty} P(\max_{1\leq i\leq k} \sup_{s,t\in T_i} |X_n(s) - X_n(t)| > \epsilon) < \eta$$

1.20 Asymptotic Uniform Equicontinuity

Suppose ρ is a semi-metric on T. A sequence $X_n : \Omega_n \to L^\infty(T)$ is asymptotically uniformly ρ-equicontinuous in probability if $\forall \epsilon, \eta > 0$, $\exists \delta > 0$ s.t.

$$\limsup_n P(\sup_{\rho(s,t)<\delta} |X_n(s) - X_n(t)| > \epsilon) < \eta$$

1.21 Tight Verifying Theorem (Asymptotic Uniform Equicontinuity)

$X_n : \Omega_n \to L^\infty(T)$ is asymptotically tight if and only if $X_n(t) : \Omega_n \to \mathbf{R}$ is asymptotically tight for all t and $\exists$ semimetric ρ on T s.t. (T, ρ) is totally bounded and X_n is asymptotically uniformly ρ-equicontinuous in probability.

[4] I made this name up. Basically the intuition here is that if $X_n \xrightarrow{L} X$, then X_n should inherits some properties from X.

2. *Empirical Process*

2.1 Empirical Measure

For a *random* sample $\{X_i\}_1^n$, the empirical measure P_n is defined as $P_n(C) := \frac{1}{n}\sum \delta_{X_i} = \frac{1}{n}\sum \mathbf{1}\{X_i \in C\}$. Intuitively, the measure puts mass $\frac{1}{n}$ at each point X_i.

2.2 Notations

- For a measure Q on $\mathcal{X}$ (domain of X_i), we define $Qf := \mathbb{E}_Q[f(x)] = \int_{\mathcal{X}} f dQ$.
- For empirical measure P_n, we have $P_n f = \mathbb{E}_{P_n}[f(x)] = \int_{\mathcal{X}} f dP_n = \frac{1}{n}\sum f(X_i)$.
- Let P be the distribution of X. For any function f, we have
$$G_n(f) := \sqrt{n}(P_n - P)f = \sqrt{n}[\frac{1}{n}\sum f(X_i) - \mathbb{E}_p[f(X_i)]] = \frac{1}{\sqrt{n}}\sum(f(X_i) - \mathbb{E}_p[f(X_i)])$$
that is, $G_n = \sqrt{n}(P_n - P)$. As the notation before, now $G_n(f)$ is the marginal and notice that $G_n(f) : \mathcal{X} \to \mathbf{R}$.
- $L^\infty(\mathcal{F}) := \{\varphi : \mathcal{F} \to \mathbf{R}, \|\varphi\|_{\mathcal{F}} := \sup_{f\in\mathcal{F}} |\varphi(f)| < \infty\}$.
- Eg. $\mathcal{F} = \{f(x) := g(x,\theta), \theta \in \Theta\}$; $\mathcal{F} = \{f(x) = \mathbf{1}\{x \le t\} : t \in \mathbf{R}\}$; $\mathcal{F} = \{f : \mathcal{X} \to \mathbf{R}, f \text{ monotonic}\}$
- $\{X_i\}$ not changing[5] $\implies$ G_n not changing. It is $\mathcal{F}$ that's changing.

2.3 Goal

- Marginals "easy": For any finitely many $(f_1, \cdots, f_k) \in \mathcal{F}$, show marginals $(G_n(f_1), \cdots, G_n(f_k)) \xrightarrow{L} (G(f_1), \cdots, G(f_k))$.
- Tightness "hard": Need to show asymptotic uniform equicontinuity in probability.

2.4 Big Picture

Let[6] $\mathcal{F}_\delta := \{f - g : f, g \in \mathcal{F}, \rho(f,g) < \delta\}$. If we can show
$$\mathbb{E}\Big[\sup_{\rho(f,g)<\delta}\Big|\frac{1}{\sqrt{n}}\sum(f(X_i)-\mathbb{E}[f(X_i)]) - \frac{1}{\sqrt{n}}\sum(g(X_i)-\mathbb{E}[g(X_i)])\Big|\Big] = \mathbb{E}\Big[\sup_{\rho(f,g)<\delta}|G_n(f-g)|\Big] = \mathbb{E}[\|G_n\|_{\mathcal{F}_\delta}] \le M(\delta)$$
where $M(\delta) \downarrow 0$ as $\delta \downarrow 0$, then Markov's Inequality, $P(\|G_n\|_{\mathcal{F}_\delta} > \epsilon) \le \mathbb{E}[\|G_n\|_{\mathcal{F}_\delta}]/\epsilon$, will take us home.

2.5 Glivenko Cantelli and Donsker Definitions

- $\mathcal{F}$ is Glivenko Cantelli (ULLN) if $\|P_n - P\|_{\mathcal{F}} = \sup_{\mathcal{F}} |\frac{1}{n}\sum f(X_i) - \mathbb{E}[f(X)]| = o_p(1)$.
- $\mathcal{F}$ is Donsker (Functional CLT) if $G_n \xrightarrow{L} G$ on $L^\infty(\mathcal{F})$.

2.6 Entropy Numbers

- Covering number $N(\epsilon, \mathcal{F}, \|\cdot\|)$ is the smallest numer of ϵ balls under $\|\cdot\|$ that cover $\mathcal{F}$.
- Bracketing numbers $N_{[]}(\epsilon, \mathcal{F}, \|\cdot\|)$ is the smallest number of ϵ brackets that cover $\mathcal{F}$.
- Given functionals l and u, the bracket $[l, u] := \{f \in \mathcal{F} : l(x) \le f(x) \le u(x)\ \forall x\}$; and the ϵ-bracket is a bracket $[l, u]$ s.t. $\|u - l\| < \epsilon$.
- $N(\epsilon, \mathcal{F}, \|\cdot\|) \le N_{[]}(2\epsilon, \mathcal{F}, \|\cdot\|)$ since each 2ϵ-bracket is contained in an ϵ-ball.

2.7 Orlicz Norm

Let ψ be a nondecreasing convex function s.t. $\psi(0) = 0$ and let X be a random variable. Then the Orlicz norm $\|X\|_\psi := \inf\{C > 0 : \mathbb{E}[\psi(\frac{|X|}{C})]\}$.

- If $\psi(x) = x^p, p \ge 1$, then $\|X\|_\psi = \|X\|_p = \mathbb{E}[X^p]^{1/p}$.
- If $\psi(x) = e^{x^p} - 1, p \ge 1$, then $\psi^{-1}(m) = (\log(m+1))^{1/p}$, usually we use $p = 2$.

[5] Not changing in the sense that this set of random variable is fixed. They are still random.
[6] The choice of ρ will become clear later.

2.8 Finite Maximal Inequality

Let $\psi(x) = e^{x^2} - 1$ and suppose $P(|G(f_i)| > x) \leq Ke^{-Dx^2}$ for some K,D and for all f_i. Then if $C \geq (\frac{1+K}{D})^{1/2}$, we have $\mathbb{E}[\psi(\frac{|G(f_i)|}{C})] \leq 1$ and consequently $\mathbb{E}[\max_{1\leq i\leq m} |G(f_i)|] \leq C\sqrt{\log(m+1)}$.

2.9 Separable Process

Let G be a stochastic process on $\mathcal{F}$. Then G is separable if $\sup_{\mathcal{F}} |G(f)| = \sup_{\tilde{\mathcal{F}}} |G(f)|$ for $\tilde{\mathcal{F}}$ a countable subset of $\mathcal{F}$.

2.9 Sub-Gaussian Process

Let G be a stochastic process on $\mathcal{F}$ equipped with metric $d(f, g)$. Then G is sub-Gaussian if

$$P(|G(f) - G(g)| > x) \leq 2\epsilon^{-\frac{1}{2}x^2/d^2(f,g)}$$

for all $f, g \in \mathcal{F}$ and $x > 0$.

2.10 Maximal Inequality

Let G be a separable sub-Gaussian process on (totally bounded) $\mathcal{F}$, then

$$\mathbb{E}[\sup_{f,g\in\mathcal{F}} |G(f) - G(g)|] \leq K \int_0^{\mathrm{diam}(\mathcal{F})} \sqrt{\log N(\epsilon, \mathcal{F}, d)}d\epsilon$$

and

$$\mathbb{E}[\sup_{\mathcal{F}} |G(f)|] \leq \mathbb{E}[|G(f_0)|] + K \int_0^{\mathrm{diam}(\mathcal{F})} \sqrt{\log N(\epsilon, \mathcal{F}, d)}d\epsilon$$

for any $f_0 \in \mathcal{F}$.

2.11 Consequence of Maximal Inequality

Apply Maximal Inequality on $\mathcal{F}_\delta$, we get

$$P(\sup_{f\in\mathcal{F}_\delta} |G_n(f)| > \epsilon) \leq \frac{1}{\epsilon}\mathbb{E}[\sup_{f\in\mathcal{F}_\delta} |G_n(f)|] \leq \frac{1}{\epsilon}(\mathbb{E}[G_n(f_0)] + K \int_0^{\mathrm{diam}(\mathcal{F}_\delta)} \sqrt{\log N(\epsilon, \mathcal{F}_\delta, d)}d\epsilon)$$

Then take $f_0 = 0$, and since $\mathrm{diam}(\mathcal{F}_\delta) = 2\delta$, we get

$$\limsup P(\sup_{f\in\mathcal{F}_\delta} |G_n(f)| > \epsilon) \leq \frac{1}{\epsilon}\mathbb{E}[\sup_{f,g\in\mathcal{F}_\delta} |G_n(f) - G_n(g)|] \leq \frac{K}{\epsilon} \int_0^{2\delta} \sqrt{\log N(\epsilon, \mathcal{F}_\delta, d)}d\epsilon \xrightarrow{\delta\to 0} 0$$

2.12 Hoeffding's Inequality

Let $a = (a_1, \cdots, a_n) \in \mathbf{R}^n$ be constant and $\epsilon_1, \cdots, \epsilon_n$ be independent Rademacher random variables (i.e. $P(\epsilon_i = 1) = P(\epsilon_i = -1) = \frac{1}{2}$). Then

$$P(|\sum_{i=1}^n \epsilon_i a_i| > x) \leq 2\epsilon^{-\frac{1}{2}\frac{x^2}{\|a\|^2}}$$

where $\|a\| = (\sum_{i=1}^n a_i^2)^{1/2}$ is the usual Euclidean norm.

2.13 Important Hoeffding's Inequality Example

For any f, g and fixed sample $\{x_i\}_1^n$, define $a_i = \frac{1}{\sqrt{n}}(f(x_i) - g(x_i))$ and hend $d_n(f, g)^2 := \|a\|^2 = \frac{1}{n}\sum_{i=1}^n (f(x_i) - g(x_i))^2$. Then we have

$$P\Big(|\sum_{i=1}^n \frac{\epsilon_i}{\sqrt{n}}(f(x_i) - g(x_i))| > x \Big| \{x_i\}_1^n\Big) \leq 2e^{-\frac{1}{2}\frac{x^2}{d_n^2(f,g)}}$$

2.14 Symmetrization Lemma

For every nondecreasing and convex $\Phi : \mathbf{R} \to \mathbf{R}$ and a class of measurable functions $\mathcal{F}$, we have

$$\mathbb{E}^*[\Phi(\|P_n - P\|_{\mathcal{F}})] \leq \mathbb{E}^*[\Phi(2\|P_n^0\|_{\mathcal{F}})]$$

where $(P_n - P)(f) = \frac{1}{n}\sum_{i=1}^n f(X_i) - \mathbb{E}[f(X_i)]$ and $P_n^0(f) := \frac{1}{n}\sum_{i=1}^n \epsilon_i f(X_i)$. The common choice is to choose $\Phi(x) = x$.

2.15 Another Big Picture

- Step 1: Need to show tightness: for δ small,
$$\limsup_n P\Big(\sup_{d(f,g)<\delta} |G_n(f) - G_n(g)| > \epsilon\Big) < \eta$$
- Step 2: Let $\mathcal{F}_\delta := \{f - g : f, g \in \mathcal{F} \text{ and } d(f,g) < \delta\}$, then above is equal to
$$\limsup_n P\Big(\sup_{f\in\mathcal{F}_\delta} |G_n(f)| > \epsilon\Big) < \eta$$
- Step 3: By Markov's Inequality, we have
$$P\Big(\sup_{f\in\mathcal{F}_\delta} |G_n(f)| > \epsilon\Big) \le \frac{1}{\epsilon}\mathbb{E}[\sup_{f\in\mathcal{F}_\delta} |G_n(f)|]$$
- Step 4: By Symmetrization, we bound LHS by a sub-Gaussian process
$$\mathbb{E}[\sup_{f\in\mathcal{F}_\delta} |G_n(f)|] \le 2\mathbb{E}[\sup_{f\in\mathcal{F}_\delta} |\frac{1}{\sqrt{n}}\sum_{i=1}^n \epsilon_i f(X_i)|]$$
- Step 5: By Maximal Inequality, (conditional on the random sample $\{X_i\}$ we have
$$\mathbb{E}_\epsilon[\sup_{f\in\mathcal{F}_\delta} |\frac{1}{\sqrt{n}}\sum_{i=1}^n \epsilon_i f(X_i)|] \le \int_0^{\text{diam}(\mathcal{F}_\delta)} \sqrt{\log N(\epsilon, \mathcal{F}_\delta, L_2(P_n))}d\epsilon$$
Then using law of iterated expecatation (over $\{X_i\}$), we get
$$\mathbb{E}[\sup_{f\in\mathcal{F}_\delta} |\frac{1}{\sqrt{n}}\sum_{i=1}^n \epsilon_i f(X_i)|] \le \mathbb{E}[\int_0^{\text{diam}(\mathcal{F}_\delta)} \sqrt{\log N(\epsilon, \mathcal{F}_\delta, L_2(P_n))}d\epsilon]$$
where $L_2(P_n)(h) := \sqrt{\frac{1}{n}\sum_{i=1}^n h^2(X_i)}$ for $h \in \mathcal{F}_\delta$.

2.16 Glivenko-Cantelli (Bracketing Numbers)

Let $\mathcal{F}$ be a class of measurable functions such that $\mathbb{N}_{[]}(\epsilon, \mathcal{F}, L_1(P)) < \infty$ for all $\epsilon > 0$. Then $\mathcal{F}$ is Glivenko-Cantelli.

2.17 Glivenko-Cantelli (Covering Numbers)

Let $\mathcal{F}$ be P-measurable with envelope F s.t. $\mathbb{E}[F(X)] < \infty$. If $\log N(\epsilon, \mathcal{F}_M, L_1(P_n)) = o_p(n)$ for every $\delta, M > 0$, then $\mathcal{F}$ is Glivenko-Cantelli.

- $\mathcal{F}$ has envelope F if $|f(x)| \le F(x)$ for all x and all $f \in \mathcal{F}$.
- $\mathcal{F}_M := \{f(x)\mathbf{1}\{F(x) \le M\} : f \in \mathcal{F}\}$
- A class $\mathcal{F}$ is P-measurable if $\sup_{\mathcal{F}} \left|\frac{1}{n}\sum_{i=1}^n f(X_i)\epsilon_i\right|$ is measurable.
- $\|f - g\|_{L_1(P)} = \mathbb{E}[|f(X) - g(X)|]$; $\|f - g\|_{L_1(P_n)} = \mathbb{E}_{P_n}[|f(X) - g(X)|] = \frac{1}{n}\sum_{i=1}^n |f(X_i) - g(X_i)|$.

2.18 Donsker (Bracketing Numbers)

Let $\mathcal{F}$ be a class of measurable functions with envelope F s.t. $\mathbb{E}[F^2(X)] < \infty$. If $\int_0^\infty \sqrt{\log N_{[]}(\epsilon, \mathcal{F}, L_2(P))} < \infty$, then $\mathcal{F}$ is Donsker.

2.19 Donsker (Covering Numbers)

Let $\mathcal{F}$ be a class of measurable functions satisfying the uniform entropy condition. If $\mathcal{F}_\delta$ and $\mathcal{F}_\infty^2$ are P-measurable and $\mathbb{E}[F^2(X)] < \infty$, then $\mathcal{F}$ is Donsker.

- $\|f\|_{P,2} := \|f\|_{L_2(P)} = (\mathbb{E}_P[f^2(X)])^{1/2}$
- If $\mathcal{F}$ has envelope F, the uniform entropy condition is $\int_0^\infty \sup_Q \sqrt{\log N(\epsilon\|F\|_{Q,2}, \mathcal{F}, \|\cdot\|_{Q,2})}d\epsilon < \infty$ where sup is taken over all discrete measure Q.
- $\mathcal{F}_\infty = \{f - g : f, g \in \mathcal{F}\}$; $\mathcal{F}_\infty^2 = \{(f-g)^2 : f, g \in \mathcal{F}\}$.