Lecture 14: Empirical Processes

Isabella Zhu

1 April 2025

We are officially in part 2 of the class! Yippee! Here's the content that will be covered:

- Empirical process theory
- Metric entropy and chaining
- Non-parametric estimation
- Statistical lower bounds

§1 Empirical Process Theory

Setup: some family of functions $f \in F$ and $\{X_i\}_{i=1}^n \sim i.i.d.$ Consider random variable

$$\frac{1}{n}\sum_{i=1}^n f(x_i) - \mathbb{E}[f(X)]$$

We can say that this is zero-mean and by LLN converges to 0. What empirical process theory does is we instead look at the collection of random variables

$$\left\{ Z_f = \frac{1}{n}\sum_{i=1}^n f(x_i) - \mathbb{E}[f(X)] \mid f \in F \right\}$$

One question we might ask is: does a *uniform* law of large numbers hold? i.e. does

$$\sup_{f \in F} \left\{ Z_f = \frac{1}{n}\sum_{i=1}^n f(x_i) - \mathbb{E}[f(X)] \mid f \in F \right\}$$

converge to 0?

Example 1.1

We have matrix $\mathbb{X} \in \mathbb{R}^{n\times d}$, each $X_i \in \mathbb{R}^d$ drawn i.i.d, $i = 1, \dots n$.

(a) Estimate $\mu^* = \mathbb{E}[X] \in \mathbb{R}^d$. One natural estimator is the empirical mean. We can write $||\hat{\mu} - \mu^*||_2$ as

$$\sup_{||a||_2=1} a^T(\hat{\mu} - \mu^*)$$

so our class of functions is

$$F = \{f_a(x) = a^T x \mid ||a||_2 = 1\}$$

(b) Estimate the covariance matrix $\Sigma = \mathbb{E}[XX^T]$. One natural estimator is the empirical $\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^n X_i X_i^T$. We can write $||\hat{\Sigma} - \Sigma^*||_{op}$ so that

$$||\hat{\Sigma} - \Sigma^*||_{op} = \sup_{||a||_2=1} a^T(\hat{\Sigma} - \Sigma^*)a = \sup_{||a||_2=1} \left\{ \frac{1}{n}\sum_{i=1}^n (a^T X_i)^2 - \mathbb{E}[(a^T X)^2] \right\}$$

so family of functions is $f_a(X_i) = (a^T X_i)^2$.

(c) We have scalar random variables $X_i \in \mathbb{R}$. Let $F(t) = \mathbb{P}(x \leq t)$ be the CDF. We can estimate with the empirical CDF $\hat{F}(t)$. We care about $||\hat{F} - F||_\infty = \sup_{t\in\mathbb{R}} |\hat{F}(t) - F(t)|$.

Remark 1.2. The takeaway here is that a lot of things can be expressed as the solution of an optimization problem.

§2 Statistical Functionals

Many interesting objects are functionals of the CDF. Basically,

$$F \to \gamma(F) \in \mathbb{R}$$

a function mapping a CDF to a real number.

Example 2.1

Some examples of statistical functionals are

1. $\gamma(F) = \int_{-\infty}^{\infty} (F(t) - F_0(t))^2 dt$
2. $\gamma_q(F) = \inf_\alpha \{F(\alpha) \geq q\}$ (quantile).
3. $\mathbb{E}[X] = \int x dF(x)$.

§2.1 Plug-In Approach

The plugin approach is to use estimate

$$\gamma(\hat{F}) = \text{plug-in estimator}$$

If γ is Lipschitz, for example, we can bound

$$|\gamma(F) - \gamma(G)| \leq L||F - G||_\infty$$

§3 Glivenko-Cantelli

Going back to empirical CDFs, we have

$$||\hat{F} - F||_\infty = \sup_{t \in \mathbb{R}} \left| \frac{1}{n} \sum_{i=1}^n 1(X_i \leq t) - \mathbb{E}[1(X \leq t)] \right|$$

Define the function class to be

$$F = \{f_t \mid t \in \mathbb{R}\}, \;\; f_t(x) = \begin{cases} 1 & \text{if } x \geq t \\ 0 & \text{otherwise} \end{cases}$$

This can be rewritten as the empirical process

$$||\hat{F} - F||_\infty = \sup_{t \in \mathbb{R}} \left| \frac{1}{n} \sum_{i=1}^n 1(X_i \leq t) - \mathbb{E}[1(X \leq t)] \right| = \sup_{t \in \mathbb{R}} \left| \frac{1}{n} \sum_{i=1}^n f_t(X_i) - \mathbb{E}[f_t(X_i)] \right|$$

Theorem 3.1
(Glivenko-Cantelli) Given i.i.d X_i, $i = 1, \ldots, n$,

$$\mathbb{P}\left(||\hat{F} - F||_\infty \geq 8\sqrt{\log(n+1)/n} + \delta\right) \leq e^{-2n\delta^2}$$

Lemma 3.2
We have

$$\mathbb{P}(||\hat{F} - F||_\infty \geq \mathbb{E}[||\hat{F} - F||] + \delta) \leq e^{-2n\delta^2}$$

Proof. We can think of $||\hat{F} - F||_\infty$ as a function Z, where $Z = Z(X_1, \ldots X_n)$. We use *bounded differences*

$$|Z(X_1, \ldots X_j, \ldots X_n) - Z(X_1, \ldots X_j', \ldots X_n)| \leq \frac{1}{n}$$

which is true (just reason about $\hat{F}$). Then we apply McDiarmid's and we're done.

Lemma 3.3
We have

$$\mathbb{E}[||\hat{F} - F||_\infty] \leq 8\sqrt{\log(n+1)/n}$$

Proof. Let $(X'_1, \dots X'_n)$ be a ghost sample independent of the original sample.

$$\begin{aligned}\mathbb{E}[||\hat{F} - F||_\infty] &= \mathbb{E}\left[\sup_{t\in\mathbb{R}}\left|\frac{1}{n}\sum_{i=1}^n f_t(X_i) - \mathbb{E}_{X'}[f(X'_i)]\right|\right] \\ &\le \mathbb{E}_X\mathbb{E}_{X'}\left[\sup_{t\in\mathbb{R}}\left|\frac{1}{n}\sum_{i=1}^n (f_t(X_i) - f_t(X'_i))\right|\right] \\ &= \mathbb{E}_X\mathbb{E}_{X'}\mathbb{E}_\epsilon \sup_{t\in\mathbb{R}}\left|\frac{1}{n}\sum_{i=1}^n \epsilon_i(f_t(X_i) - f_t(X'_i))\right|\end{aligned}$$

where $\epsilon_i \in \{-1, 1\}$ is a Rademacher random variable. We can do this because of symmetry.

Note that we can move the $\mathbb{E}_{X'}[f(X'_i)]$ outside because of convexity (need to check that adding the sup is still convex).

Now we can apply the triangle inequality to get

$$\begin{aligned}\mathbb{E}||\hat{F} - F||_\infty &\le \mathbb{E}_{X,\epsilon}\sup_{t\in\mathbb{R}}\left|\frac{1}{n}\sum_{i=1}^n \epsilon f_t(X_i)\right| + \mathbb{E}_{X',\epsilon}\sup_{t\in\mathbb{R}}\left|\frac{1}{n}\sum_{i=1}^n \epsilon f_t(X'_i)\right| \\ &\le 2\mathbb{E}_X\left[\mathbb{E}_\epsilon\left[\sup_{t\in\mathbb{R}}\left|\frac{1}{n}\sum_{i=1}^n \epsilon_i f_t(X_i)\right|\right]\right]\end{aligned}$$

Now we use the indicator function structure, where we condition on the X_is. In fact, this is a finite maximum in disguise. The number of choices is at most $n+1$. Now we have a bound summing over n i.i.d bounded variables. We get the above equal to

$$\mathbb{E}_\epsilon \max_{a\in A}\left|\frac{1}{n}\sum_{i=1}^n \epsilon_i a_i\right| \le 4\sqrt{\log(n+1)/n}$$

by results on subgaussian finite maxima.