Lecture #3

- How to think about Hoeffding? For a fixed ε, we want t big enough so $\exp(-t\varepsilon) << 1$ but not so large that $\prod_{i=1}^{n} \exp[t^2 |b_i - a_i|^2/8]$ is large.

- This result is especially powerful to apply to empirical averages. In this case, we are interested in

$$\mathbb{P}\Big(\underbrace{\frac{1}{n}\sum_{i=1}^{n} \bar{X}_i}_{\text{average of } n \text{ i.i.d. samples}} > \varepsilon\Big) = \mathbb{P}\Big(\underbrace{\sum_{i=1}^{n} \bar{X}_i \geq n\varepsilon}_{\text{Run Hoeffding}}\Big)$$

$$\leq \underbrace{\exp(-n\varepsilon t)}_{\substack{\to 0 \\ \text{as } n\to\infty \text{, and} \\ \text{fast!}}} \underbrace{\prod_{i=1}^{n} \exp\left(t^2[b_i - a_i]^2/8\right)}_{\text{constant w.r.t. } n}$$

- So, from Hoeffding we immediately get the weak law of large numbers (WLLN) for any bounded r.v..

Proof of Hoeffding: We will use Markov's inequality and the boundedness of X_i.

Applying Markov for t fixed gives:

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i \geq \varepsilon\right) = \mathbb{P}\left(t\sum_{i=1}^{n} X_i \geq t\varepsilon\right)$$

$$= \mathbb{P}\left(\exp\left[t\sum_{i=1}^{n} X_i\right] \geq \exp\left[t\varepsilon\right]\right)$$

Markov

$$\leq \frac{1}{\exp[t\varepsilon]} \cdot \mathbb{E}\left[\exp\left[t\sum_{i=1}^{n} X_i\right]\right]$$

$$= \exp(-t\varepsilon) \cdot \mathbb{E}\left[\prod_{i=1}^{n} \exp(tX_i)\right]$$

iid

$$= \exp(-t\varepsilon) \prod_{i=1}^{n} \mathbb{E}\left(\exp(tX_i)\right).$$

So, it suffices to show $\mathbb{E}\left(\exp(tX_i)\right) \leq \exp\left(t^2[b_i-a_i]^2/8\right)$, for all i.

Since $a_i \leq X_i \leq b_i$, write $X_i = \lambda b_i + [1-\lambda]a_i$, where λ is the r.v.

$$\lambda = \frac{X_i - a_i}{b_i - a_i}.$$

Note $\mathbb{E}[\lambda] = \dfrac{\mathbb{E}[X_i] - a_i}{b_i - a_i} = \dfrac{-a_i}{b_i - a_i}$

Then by convexity, $\exp[tX_i] = \exp[t\lambda b_i + [1-\lambda]a_i]$

$$\leq \lambda \exp[tb_i] + [1-\lambda]\exp[ta_i]$$

$$\Rightarrow \mathbb{E}\left[\exp\left[tX_i\right]\right] \le \mathbb{E}[\lambda]\cdot\exp[tb_i] + \mathbb{E}[1-\lambda]\cdot\exp[ta_i]$$

$$= \frac{-a_i}{b_i-a_i}\cdot\exp(tb_i) + \frac{b_i}{b_i-a_i}\cdot\exp[ta_i]. \quad (\star)$$

To analyze $(\star)$, introduce the auxilliary function

$$g(u) = -\gamma u + \log\left[1-\gamma+\gamma\exp(u)\right], \quad \gamma = \frac{-a_i}{b_i-a_i}.$$

Letting $u = t[b_i-a_i]$, we see $\mathbb{E}\left[\exp[tX_i]\right] \le \exp(g(u))$.

To finish, we do some calculus on u. Notice:

- $g(0) = 0 + \log[1] = 0$
- $g'(u) = -\gamma + \dfrac{\gamma\exp(u)}{1-\gamma+\gamma\exp(u)} \Rightarrow g'(0) = -\gamma + \dfrac{\gamma}{1} = 0$
- $g''(u) = \gamma\dfrac{d}{du}\left[\dfrac{\exp(u)}{1-\gamma+\gamma\exp(u)}\right] = \gamma\cdot\dfrac{\exp(u)\cdot[1-\gamma+\gamma\exp(u)] - \exp(u)[\gamma\exp(u)]}{(1-\gamma+\gamma\exp(u))^2}$,

from which some algebra yields $g''(u) \le \frac{1}{4}$ for all $u \ge 0$.

- Hence, Taylor expanding $g(u)$ around $u=0$ yields

$$g(u) = g(0) + g'(0)\cdot u + g''(\xi)\cdot \frac{u^2}{2} \quad \text{for some } \xi \in (0,u]$$

$$\leq \frac{u^2}{8}$$

defn. of u

$$= \frac{t^2[b_i - a_i]^2}{8}. \quad \blacksquare$$

- Much of statistics is build around laws of large numbers and central limit theorems. We will need notions of convergence of random variables to make proper sense of them.

Defn: Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of r.v. Let X be a r.v.

(1.) We say $\{X_n\}_{n=1}^{\infty}$ converge to X in probability, written $X_n \xrightarrow{\mathbb{P}} X$, if $\forall \varepsilon > 0$, $\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0$.

(2.) We say $\{X_n\}_{n=1}^{\infty}$ converges to X in L^2, written $X_n \xrightarrow{L^2} X$, if $\lim_{n\to\infty} \mathbb{E}|X_n - X|^2 = 0$.

(3.) Let $\{F_n\}_{n=1}^{\infty}$ and F be the c.d.f.s of $\{X_n\}_{n=1}^{\infty}$ and X, respectively

We say $\{X_n\}_{n=1}^{\infty}$ converges to X in distribution, written $X_n \rightsquigarrow X$,
if $\lim_{n\to\infty} F_n(t) = F(t)$ for all t s.t. F is continuous.