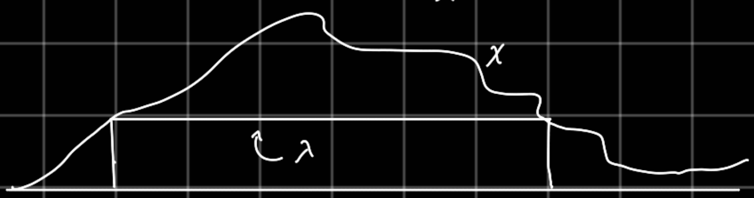
Lecture 8: MGFs, Bounds / Concentration Inequalities

(Markov, Chebyshev, Chernoff) $\boxed{\text{Passcode: Lewis}}$

Concentration Inequalities

- **Markov's Inequality**: If $X \geq 0$ is a r.v. then

$$P\{X \geq \lambda\} \leq \frac{\mathbb{E}[X]}{\lambda} \qquad \lambda \geq 0$$



$\lambda \mathbb{1}\{X \geq \lambda\}$

Example: (sometimes take fcn of X instead of X itself)

$$P\{\underbrace{|X - \mathbb{E}[X]| > \lambda}_{\text{"what's the probability that } X \text{ deviates from its mean"}}\} \leq \frac{\mathbb{E}[|X - \mathbb{E}X|]^k}{\lambda^k} \qquad \lambda > 0$$

λ^k ↳ decaying at a rate λ^k which is a much nicer decay than λ.

PF:

$$P\{|X - \mathbb{E}[X]|^k \geq \lambda^k\} \leq \frac{\mathbb{E}(|X - \mathbb{E}X]|^k}{\lambda^k}$$

- **Chebyshev's Inequality**

$$P\{|X - \mathbb{E}[X]| \geq \lambda\} \leq \frac{Var(X)}{\lambda^2}$$

↳ rooted in Markov's ineq. above (take $k=2$)

↳ want to take in more information about the moments to create a better bound → use MGF

Ex:

$$P\{X \geq \lambda\} \underset{t>0}{=} P\{e^{t\lambda} \geq e^{t\lambda}\} \underset{\text{Markov}}{\leq} \frac{\mathbb{E}e^{tX}}{e^{t\lambda}} = \frac{M_X(t)}{e^{t\lambda}} \quad \text{MGF}$$

↳ want to minimize over t to create the sharpest bound

Chernoff Bound:

$$P\{X \geq \lambda\} \leq \inf_{t>0} \frac{M_X(t)}{e^{\lambda t}}$$

Ex: (involving a Gaussian)

$$X \sim \underbrace{\mathcal{N}(0,1)}_{\text{std. Normal}} \qquad M_X(t) = \exp\left\{\frac{t^2}{2}\right\}$$

↳ applying Chernoff:

$1 - \Phi(\lambda) = \exp\left(-\frac{\lambda^2}{2} + o(\lambda^2)\right)$

$$\Pr\{X \geq \lambda\} \leq \min_{t>0} \exp\left\{\frac{t^2}{2} - t\lambda\right\} \overset{t=\lambda}{=} \underbrace{\exp\left\{-\frac{\lambda^2}{2}\right\}}_{} \quad ?$$

↳ no closed form for Gaussian CDF, so we use this to determine asymptotic behavior. (this is for complement)

↳ minimizing:

$$0 = \frac{d}{dt}\left(\frac{t^2}{2} - t\lambda\right) = t - \lambda \rightarrow t = \lambda$$

No longer need any non-negativity assumptions.

"Now this is where the probability starts" - Courtade

Thm:

Weak Law of Large Numbers

Let $X_1, X_2, \dots \overset{iid}{\sim}, X_n$. For any $\underbrace{\epsilon > 0}_{①\text{ Fix } \varepsilon}$ ← the order of qualifiers matters

② take the limit

$$\lim_{n\to\infty} P\left\{\left|\underbrace{\frac{1}{n}\sum_{i=1}^{n} X_i}_{\text{Empirical mean}} - \underbrace{\mathbb{E}[X]}_{\text{true mean}}\right| > \varepsilon\right\} = 0$$

"as n grows large, my empirical mean approaches my true mean"

↳ ie, if we perform identical experiments, and average outcomes, then that empirical average approaches true mean with overwhelming probability as the number of samples $\to \infty$.

Proof:

① Fix $\varepsilon > 0$ $\quad M_n := \frac{1}{n}\sum_{i=1}^{n} X_i$

↳ by linearity of expectation:

$$\mathbb{E}[M_n] = \mathbb{E}[X]$$

X_i's iid ⇒ uncorrelated

$$P\{|M_n - \mathbb{E}[X]| \geq \varepsilon\} \underset{\text{Chebyshev}}{\leq} \frac{\mathrm{Var}(M_n)}{\varepsilon^2} = \frac{1}{n^2} \cdot \frac{\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right)}{\varepsilon^2}$$

$$= \frac{\mathrm{Var}(X)}{n\,\varepsilon^2} \rightarrow 0 \text{ as } n \rightarrow \infty$$

Frequentist view of probability: If χ models outcome of random experiment, then $P\{\chi \in B\}$ is the frequency at which repeated identical experiments would have outcomes taking values in B.

↳ ie the $P\{$sun will come up tmrw$\} = 1$ by frequentist POV

Axiomatic View of Prob:

If $\chi_1, \chi_2, \chi_3 \dots$ are iid experiments, then

$$\underbrace{\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{\chi_i \in B\}}}_{\text{empirical frequency at which } \chi_i\text{'s take values in } B.} \to P\{\chi \in B\} \quad \text{"in probability"}$$

Punchline: WLLN reconciles the frequentist approach w/ the Axiomatic approach

Example: An eecs126 student completes a sequence of problems.

$$\text{Reader score on problem } i = \underbrace{\theta_i}_{\text{score you deserve}} + z_i \qquad z_1, z_2, \dots \overset{\text{iid}}{\sim} \text{Unif}(-2,2)$$

$$\text{Professor score at end of course} = \frac{1}{n}\sum_{i=1}^{n}(\theta_i + z_i)$$

$$= \underbrace{\frac{1}{n}\sum_{i=1}^{n}\theta_i}_{\text{score you deserve}} + \underbrace{\frac{1}{n}\sum_{i=1}^{n} z_i}_{\to 0 \text{ with overwhelming probability}}$$

↳ Law of Large Numbers leads to idea of Convergence in probability

↳ motivates idea of modes of convergence

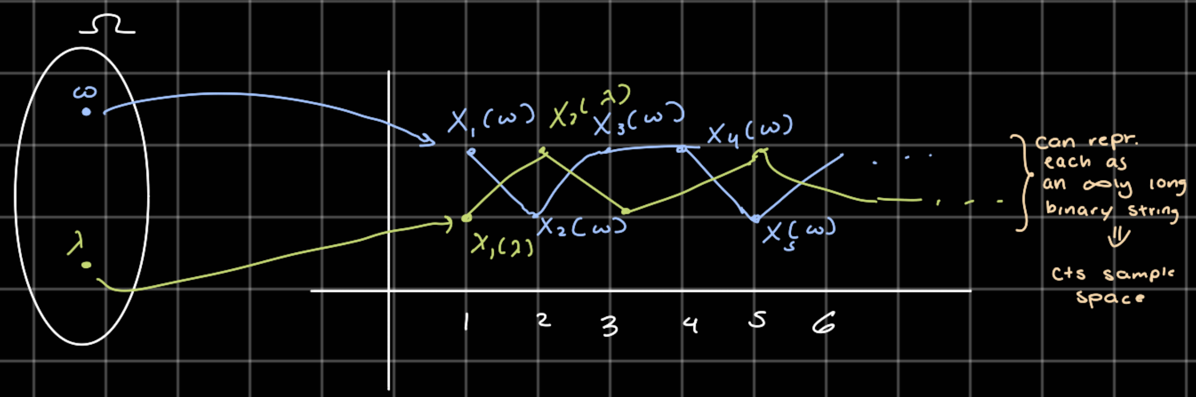
Random / Stochastic Process

- sequence of random variables on a common probability space

Ex: $\chi_1, \chi_2, \dots \overset{\text{iid}}{\sim} \chi$ ← this is a stochastic process

↳ can't be modeled on a discrete sample space

→ must be continuous



Ex: $\underbrace{M_1, M_2, M_3}$ ...
$\hookrightarrow$ empirical means

$$M_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Modes of Convergence: How random processes converge

- Weak Law of Large #'s: $\underbrace{M_n \to \mathbb{E}[X]}_{\text{can't write this by itself bc } X \text{ is a random variable}}$ in probability

Banach - Tarski