Final: Tue 21st Mar 11:30am-2:30pm
MS4000A.

Content: All of the course

Materials: Cheat sheet
+ basic calculator

Practice problems
on Canvas.

# Math 170E: Winter 2023

Lecture 24, Mon 13th Mar

The (weak) law of large numbers

**Proposition 5.10:**

Let $X_1, \ldots, X_n$ be discrete or continuous random variables.
Let $a_1, \ldots, a_n \in \mathbb{R}$ and let

$$Y = a_1 X_1 + \ldots + a_n X_n.$$

Then,

$$\mathbb{E}[Y] = \sum_{j=1}^{n} a_j \mathbb{E}[X_j].$$

$$\mathrm{var}(Y) = \sum_{j=1}^{n} \sum_{k=1}^{n} a_j a_k \mathrm{cov}(X_j, X_k). \quad --- (1).$$

In particular, if $X_1, \ldots, X_n$ are independent, then

$$\mathrm{var}(Y) = \sum_{j=1}^{n} a_j^2 \mathrm{var}(X_j) = a_1^2 \mathrm{var}(X_1) + \ldots + a_n^2 \mathrm{var}(X_n)$$

if $(X_j)$ are indep, then $\uparrow$ follows from $(1)$:

**Proof:** If $(X_j)$ are indep, $\mathrm{cov}(X_j, X_k) = \begin{cases} 0 & \text{if } j \neq k, \\ \mathrm{var}(X_j) & \text{if } j = k. \end{cases}$

① :

$$var(Y) = \mathbb{E}[Y^2] - \mathbb{E}[Y]^2$$

$$= \mathbb{E}\left[\left(\sum_{j=1}^{n} a_j X_j\right)\left(\sum_{k=1}^{n} a_k X_k\right)\right] - \left(\sum_{j=1}^{n} a_j \mathbb{E}[X_j]\right)^2.$$

$$= \mathbb{E}\left(\sum_{j=1}^{n}\sum_{k=1}^{n} a_j a_k X_j X_k\right] - \sum_{j=1}^{n}\sum_{n=1}^{n} a_j a_n \mathbb{E}[X_j]\,\mathbb{E}[X_k].$$

$$= \sum_{j=1}^{n}\sum_{n=1}^{n} a_j a_n \mathbb{E}[X_j X_k] - \sum_{j=1}^{n}\sum_{n=1}^{n} a_j a_n \mathbb{E}[X_j]\,\mathbb{E}[X_k]$$

$$= \sum_{j=1}^{n}\sum_{n=1}^{n} a_j a_k \left(\underbrace{\mathbb{E}[X_j X_k] - \mathbb{E}[X_j]\,\mathbb{E}[X_k]}_{cov(X_j, X_k).}\right).$$

**Definition 5.11:**

Let $X_1, \ldots, X_n$ be independent and identically distributed.
We define the sample sum

$$S_n = \sum_{j=1}^{n} X_j = \underbrace{X_1 + \ldots + X_n}$$

$\rightarrow$ $a_j = 1$ for all $j = 1, \ldots, n$.

and sample average

$$\overline{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j = \frac{1}{n} S_n.$$

$\rightarrow$ $a_j = \frac{1}{n}$ for all $j = 1, \ldots, n$.

**Example 6:**

- Let $X_1, \ldots, X_n$ be i.i.d random variables with mean $\mu$ and variance $\sigma^2$
- What are $\mathbb{E}[S_n]$, $\mathbb{E}[\overline{X}_n]$, $\mathrm{var}(S_n)$, and $\mathrm{var}(\overline{X}_n)$?

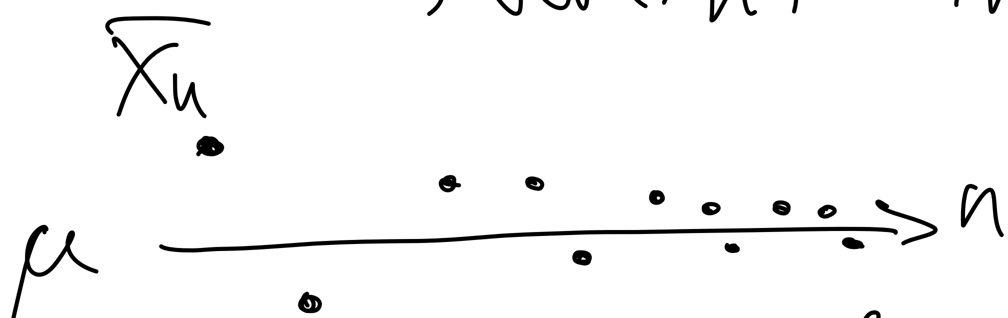
$$\mathbb{E}[S_n] = \sum_{j=1}^{n} \mathbb{E}[X_j] = \sum_{j=1}^{n} \mu = n\mu.$$

$$\mathbb{E}[\overline{X}_n] = \mathbb{E}\left[\tfrac{1}{n}S_n\right] = \tfrac{1}{n}\mathbb{E}[S_n] = \tfrac{1}{\cancel{n}}(\cancel{n}\mu) = \mu.$$

$$\mathrm{var}(S_n) \underset{\text{by independence}}{=} \sum_{j=1}^{n} \mathrm{cov}(X_j, X_j) = \sum_{j=1}^{n} \mathrm{var}(X_j) = n\sigma^2.$$

$$\mathrm{var}(\overline{X}_n) = \mathrm{var}\left(\tfrac{1}{n}S_n\right) = \tfrac{1}{n^2}\mathrm{var}(S_n) = \tfrac{1}{n^{\cancel{2}}}\cancel{n}\sigma^2 = \tfrac{\sigma^2}{n}.$$

$$\hookrightarrow \mathrm{var}(\overline{X}_n) = \sigma^2/n \to 0 \text{ as } n \to +\infty.$$




$$\longrightarrow \boxed{\overline{X}_n(\omega) \to \mu\,?}$$

<u>E.g.</u> $X_j \sim \mathrm{Bernoulli}(1/2) \leftarrow$ fair coins.

$$\mathbb{E}[\overline{X}_n] = \tfrac{1}{2}, \quad \mathrm{var}(\overline{X}_n) = \tfrac{1}{4n}.$$

- Let $X$ be a random variable, which we think of as modelling the outcome of an experiment.

- We defined $\mathbb{E}[X]$ as the "theoretical average outcome" of the experiment.

- Our goal in this lecture is to prove the (Weak) Law of large numbers

  *"The sample averages of an i.i.d. sequence of random variables converges to its expected value as the number of samples tends to infinity"*.

$\overline{X}_n(\omega) \hookrightarrow$ average given a particular outcome.

$\hookrightarrow$ " $\overline{X}_n(\omega) \longrightarrow \mathbb{E}[X]$ as $n \to +\infty$ "

$\omega \in \Omega$.

## Convergence of real numbers:

$x_n = 1/n$, $x_n = n$, $x_n = (-1)^n$.

Given a sequence of real numbers $(x_n)_{n=1}^{\infty}$ and a real number $x \in \mathbb{R}$, we say that $x_n$ **converges** to $x$ as $n \to \infty$, written

$$x_n \to x \quad \text{as} \quad n \to \infty$$

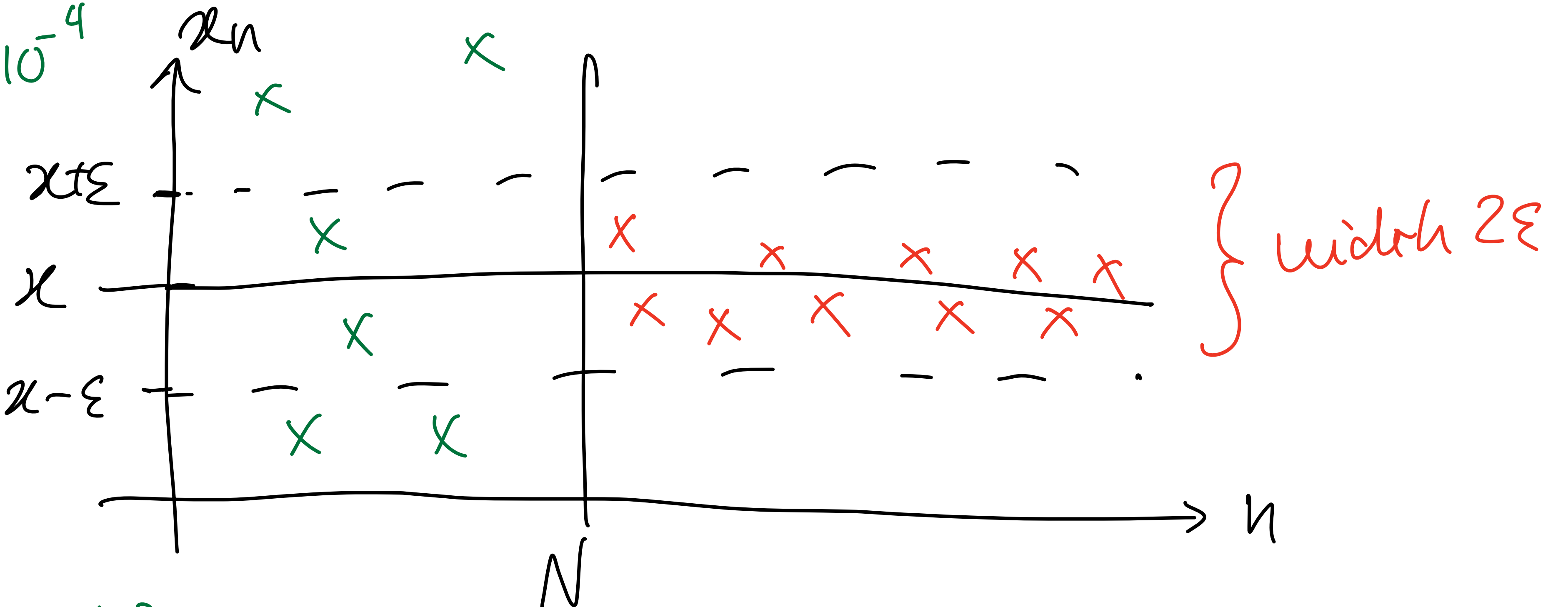
if, given any $\varepsilon > 0$, there exists some $N \geq 1$ such that for all $n \geq N$, we have

$$|x_n - x| < \varepsilon \quad \longrightarrow \quad x - \varepsilon < x_n < x + \varepsilon.$$

$\varepsilon = 10^{-2}$

$\varepsilon = 10^{-4}$

$\vdots$




$\varepsilon = 10^{-100}$

## Convergence of random variables:

- Let $(X_n)_{n=1}^{\infty}$ be a sequence of random variables and $X$ be another random variable

- There are many ways to say $X_n$ converges to $X$

- Convergence in probability: We say that

$$X_n \to X \quad \text{in probability: as} \quad n \to \infty$$

if for any $\varepsilon > 0$, we have

$$\mathbb{P}(|X - X_n| \geq \varepsilon) \to 0 \quad \text{as } n \to \infty$$

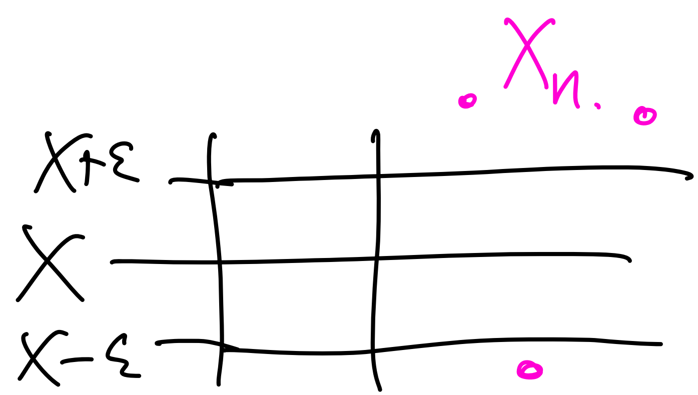
$\varepsilon = 10^{-2}$

$\varepsilon = 10^{-4}$

$\vdots$

$\varepsilon = 10^{-100}$

Roughly, this says that the prob. of a "rare event" $|X - X_n| \geq \varepsilon$ is very small as $n$ gets large



**Theorem 5.12: (The Weak Law of Large Numbers)**

Let $X_1, X_2, \ldots$ be an i.i.d. sequence of random variables with finite mean $\mu$. Then,

$$\overline{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j \to \mu \quad \text{in probability as } n \to \infty.$$

Strong $\Rightarrow$ Weak

$\not\Leftarrow$

~~Strong~~: almost sure convergence

we can find an event $\Sigma$, $\mathbb{P}(\Sigma) = 1$, for which

given any $\omega \in \Sigma$,

$\overline{X}_n(\omega) \to \mu$ as $n \to \infty$

**Proof:** We need some tools first.

Want to show

For any $\varepsilon > 0$, we have

$\mathbb{P}(|\overline{X}_n - \mu| \geq \varepsilon) \to 0$ as $n \to +\infty$.

## Example: Monte Carlo integration

- Let $g : [0,1] \to \mathbb{R}$ be continuous $\longrightarrow \int_0^1 g(x)\,dx$.
- Let $X_1, X_2, \ldots$ be an i.i.d. sequence of Uniform($(0,1)$) random variables.
- What can we say about

$$\frac{1}{n}\sum_{j=1}^{n} g(X_j) \quad \text{as } n \to \infty?$$

$$\mathbb{E}[g(X_j)] = \int_0^1 g(x) f_{X_j}(x)\,dx = \int_0^1 g(x)\,dx.$$

As $(X_j)$ are i.i.d, we have $(g(X_j))$ are i.i.d and hence by the Weak law of large numbers

$$\frac{1}{n}\sum_{j=1}^{n} g(X_j) \longrightarrow \mathbb{E}[g(X_1)] = \int_0^1 g(x)\,dx$$

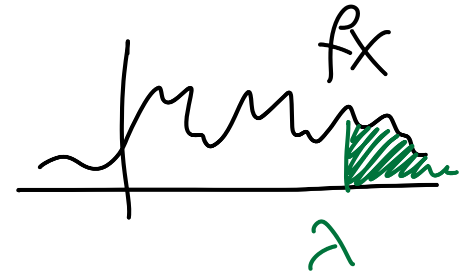
in probability as $n \to +\infty$

**Proposition 5.13: (Markov's inequality)**

Let $X$ be a non-negative random variable. Then, given $\lambda > 0$, we have

$X \geq 0$.

$$\mathbb{P}(X \geq \lambda) \leq \frac{\mathbb{E}[X]}{\lambda}.$$




**Proof:** Discrete case:

$$\mathbb{P}(X \geq \lambda) = \sum_{x \in S_X} p_X(x) \leq \sum_{\substack{x \in S_X \\ x \geq \lambda}} \frac{x}{\lambda} p_X(x)$$

$1 \leq \frac{x}{\lambda}$

$\rightarrow x \geq \lambda$

$\geq 0$ for all $x$

$$= \frac{1}{\lambda} \sum_{\substack{x \in S_X \\ x \geq \lambda}} \overbrace{x\, p_X(x)}^{\geq 0 \text{ for all } x}$$

$\hookrightarrow x \geq \lambda > 0.$

$$\leq \frac{1}{\lambda} \sum_{x \in S_X} x\, p_X(x) = \frac{\mathbb{E}[X]}{\lambda}.$$

**Example 6:**

- Let $X \sim \text{Binomial}(10, \frac{1}{2})$ $\longrightarrow$ $\mathbb{E}[X] = 10 \cdot \frac{1}{2} = 5$

- What estimate does Markov's inequality give for $\mathbb{P}(X \geq 6)$? $\lambda = 6$.

By Markov's ineq,

$$\mathbb{P}(X \geq 6) \leq \frac{\mathbb{E}[X]}{6} = \frac{5}{6} \approx 0.833.$$

$$\boxed{\mathbb{P}(X \geq 6) \approx 0.377.}$$

$X \sim \text{Bin}(10, 1/2)$

$\mathbb{E}[X^2] = \frac{5}{2} + 25$

Gen-Markov with $k=2$:

$$\mathbb{P}(X \geq 6) \leq \frac{\mathbb{E}[X^2]}{6^2} = \frac{55}{72} \approx 0.764.$$

## Proposition 5.14: (Generalised Markov's inequality)

Let $X$ be a non-negative random variable. Then, given $\lambda > 0$ and integer $k \geq 1$, we have

Cost: We need finite high moments.

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

faster decay than Markov's ineq.

**Proof:**

$$\mathbb{P}(X \geq \lambda) = \mathbb{P}(X^k \geq \lambda^k) \leq \frac{\mathbb{E}[X^k]}{\lambda^k}.$$

X non-negative.

Markov