

LECTURE 18: Inequalities, convergence, and the Weak Law of Large Numbers

- Inequalities
 - bound $\mathbf{P}(X \geq a)$ based on limited information about a distribution
 - Markov inequality (based on the mean)
 - Chebyshev inequality (based on the mean and variance)

- **WLLN:** X, X_1, \dots, X_n i.i.d.

$$\frac{X_1 + \dots + X_n}{n} \longrightarrow \mathbf{E}[X]$$

- application to polling
- Precise defn. of convergence
 - convergence “in probability”

The Markov inequality

- Use a bit of information about a distribution to learn something about probabilities of “extreme events”
- “If $X \geq 0$ and $\mathbf{E}[X]$ is small, then X is unlikely to be very large”

Markov inequality: If $X \geq 0$ and $a > 0$, then $\mathbf{P}(X \geq a) \leq \frac{\mathbf{E}[X]}{a}$.

$$Y = \begin{cases} 0, & \text{if } X < a \\ a, & \text{if } X \geq a \end{cases}$$

$$a \mathbf{P}(X \geq a) = \mathbf{E}[Y] \leq \mathbf{E}[X]$$

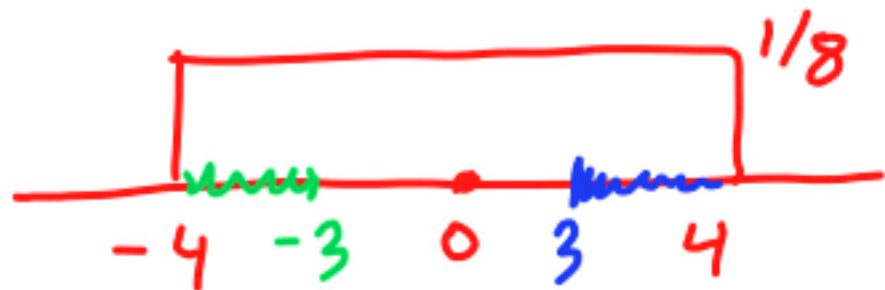
The Markov inequality

Markov inequality: If $X \geq 0$ and $a > 0$, then $P(X \geq a) \leq \frac{E[X]}{a}$

- Example:** X is Exponential($\lambda = 1$): $P(X \geq a) \leq \frac{1}{a}$



- Example:** X is Uniform $[-4, 4]$: $P(X \geq 3) \leq \frac{E[|X|]}{3} = \frac{2}{3}$



$$= \frac{1}{2} P(|X| \geq 3) \leq \frac{1}{3}$$



The Chebyshev inequality

- Random variable X , with finite mean μ and variance σ^2
- “If the variance is small, then X is unlikely to be too far from the mean”

Chebyshev inequality: $\mathbf{P}(|X - \mu| \geq c) \leq \frac{\sigma^2}{c^2}$

Markov inequality: If $X \geq 0$ and $a > 0$, then $\mathbf{P}(X \geq a) \leq \frac{\mathbf{E}[X]}{a}$

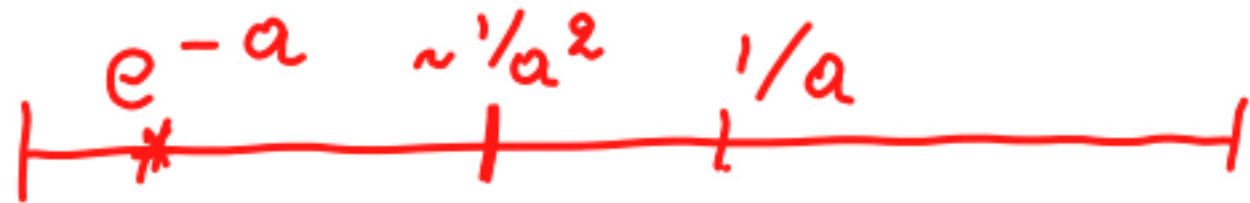
$$\mathbf{P}(|x - \mu| \geq c) = \mathbf{P}(\underbrace{(x - \mu)^2}_{\geq c^2} \geq c^2) \leq \frac{\mathbf{E}[(x - \mu)^2]}{c^2} = \frac{\sigma^2}{c^2}$$

The Chebyshev inequality

Chebyshev inequality: $P(|X - \mu| \geq c) \leq \frac{\sigma^2}{c^2}$

$$P(|X - \mu| \geq k\sigma) \leq \frac{\sigma^2}{k^2 \sigma^2} = \frac{1}{k^2} \quad k=3 \quad \leq \frac{1}{9}$$

- Example:** X is Exponential($\lambda = 1$): $P(X \geq a) \leq \frac{1}{a}$ (Markov)



$$P(X \geq a) = P(X-1 \geq a-1) \leq P(|X-1| \geq a-1) \leq \frac{1}{(a-1)^2} \sim \frac{1}{a^2}$$

The Weak Law of Large Numbers (WLLN)

- X_1, X_2, \dots i.i.d.; finite mean μ and variance σ^2

Sample mean: $M_n = \frac{X_1 + \dots + X_n}{n}$

$$\mu = E[X_i]$$

- $E[M_n] = \frac{E[X_1 + \dots + X_n]}{n} = \frac{n\mu}{n} = \mu$

- $\text{Var}(M_n) = \frac{\text{Var}(X_1 + \dots + X_n)}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$

$$P(|M_n - \mu| \geq \epsilon) \leq \frac{\text{Var}(M_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \xrightarrow{n \rightarrow \infty} 0 \quad (\text{fixed } \epsilon > 0)$$

WLLN: For $\epsilon > 0$, $P(|M_n - \mu| \geq \epsilon) = P\left(\left|\frac{X_1 + \dots + X_n}{n} - \mu\right| \geq \epsilon\right) \rightarrow 0$, as $n \rightarrow \infty$

Interpreting the WLLN

$$M_n = (X_1 + \cdots + X_n)/n$$

WLLN: For $\epsilon > 0$, $\mathbf{P}\left(|M_n - \mu| \geq \epsilon\right) = \mathbf{P}\left(\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| \geq \epsilon\right) \rightarrow 0$, as $n \rightarrow \infty$

- One experiment
 - many measurements $X_i = \mu + W_i$
 - W_i : measurement noise; $\mathbf{E}[W_i] = 0$; independent W_i
 - **sample mean** M_n is unlikely to be far off from **true mean** μ
- Many independent repetitions of the same experiment
 - event A , with $p = \mathbf{P}(A)$
 - X_i : indicator of event A
 - the sample mean M_n is the **empirical frequency** of event A

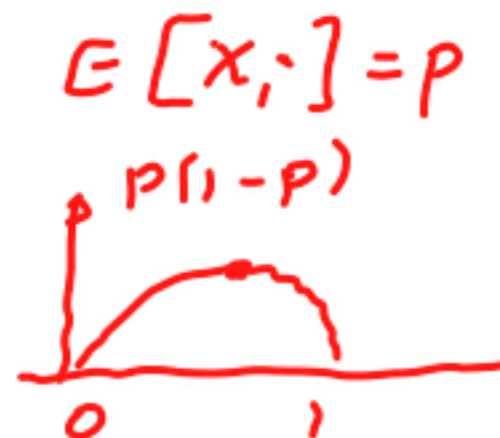
$$X_i = \begin{cases} 1, & \text{if } A \text{ occurs} \\ 0, & \text{o.w.} \end{cases} \quad \mathbf{E}[X_i] = p$$

The pollster's problem

- p : fraction of population that will vote "yes" in a referendum

- i th (randomly selected) person polled:
uniformly, independently

$$X_i = \begin{cases} 1, & \text{if yes,} \\ 0, & \text{if no.} \end{cases}$$



- $M_n = (X_1 + \dots + X_n)/n$: fraction of "yes" in our sample

- Would like "small error," e.g.: $|M_n - p| < 0.01$

- Try $n = 10,000$

- $P(|M_{10,000} - p| \geq \underline{0.01}) \leq \frac{\sigma^2}{n \epsilon^2} = \frac{p(1-p)}{10^4 \cdot 10^{-4}} \leq \frac{1}{4}$ *← want $\leq 5\%$*

$$\frac{1/4}{n 10^{-4}} \leq \frac{5}{10^2} \iff n \geq \frac{10^6}{20} = 50,000 \quad \leftarrow \text{will suffice}$$

Convergence “in probability”

WLLN: For any $\epsilon > 0$, $\mathbf{P}(|M_n - \mu| \geq \epsilon) \rightarrow 0$, as $n \rightarrow \infty$

- Would like to say that “ M_n converges to μ ”
- Need to define the word “converges”
- Sequence of random variables Y_n ; not necessarily independent

$M_n \xrightarrow[n \rightarrow \infty]{i.p.} \mu$

Definition: A sequence Y_n converges in probability to a number a if:

for any $\epsilon > 0$, $\lim_{n \rightarrow \infty} \mathbf{P}(|Y_n - a| \geq \epsilon) = 0$

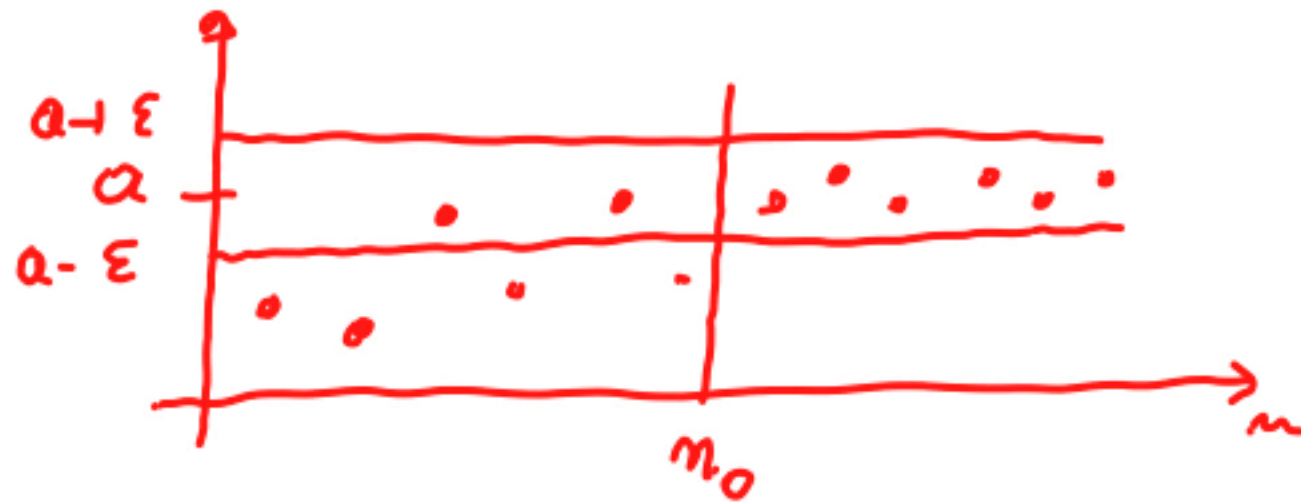
Understanding convergence “in probability”

- Ordinary convergence

- Sequence a_n ; number a

$$a_n \rightarrow a$$

“ a_n eventually gets and stays (arbitrarily) close to a ”



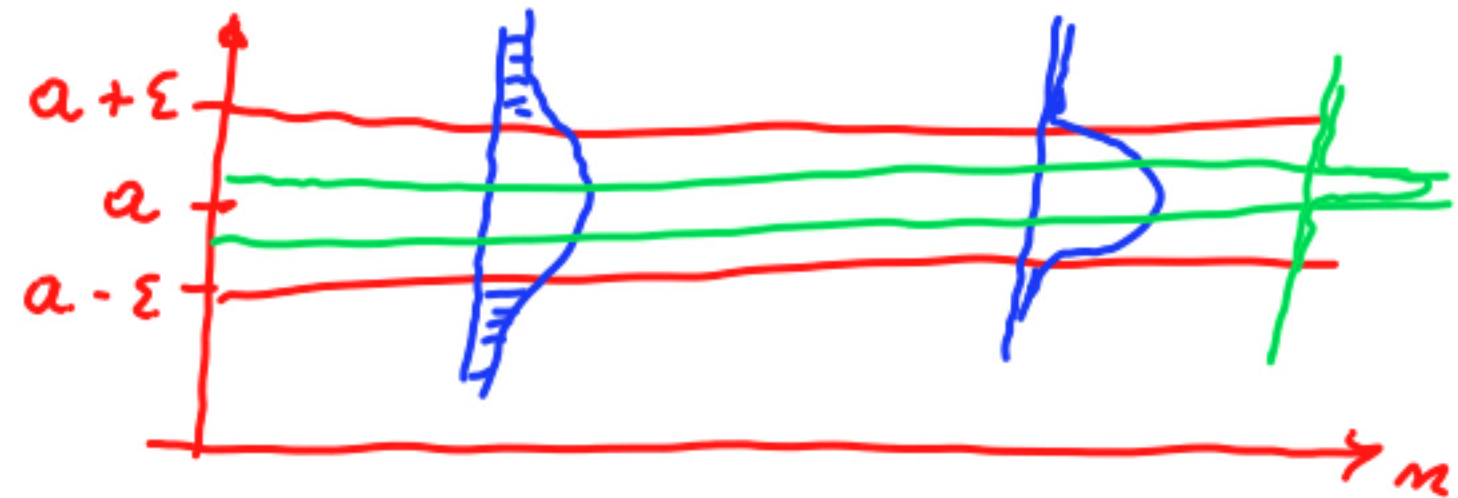
- For every $\epsilon > 0$, there exists n_0 , such that for every $n \geq n_0$, we have $|a_n - a| \leq \epsilon$

- Convergence in probability

- Sequence Y_n ; number a

$$Y_n \rightarrow a$$

- for any $\epsilon > 0$, $\mathbf{P}(|Y_n - a| \geq \epsilon) \rightarrow 0$



“(almost all) of the PMF/PDF of Y_n eventually gets concentrated (arbitrarily) close to a ”

Some properties

- Suppose that $X_n \rightarrow a$, $Y_n \rightarrow b$, in probability

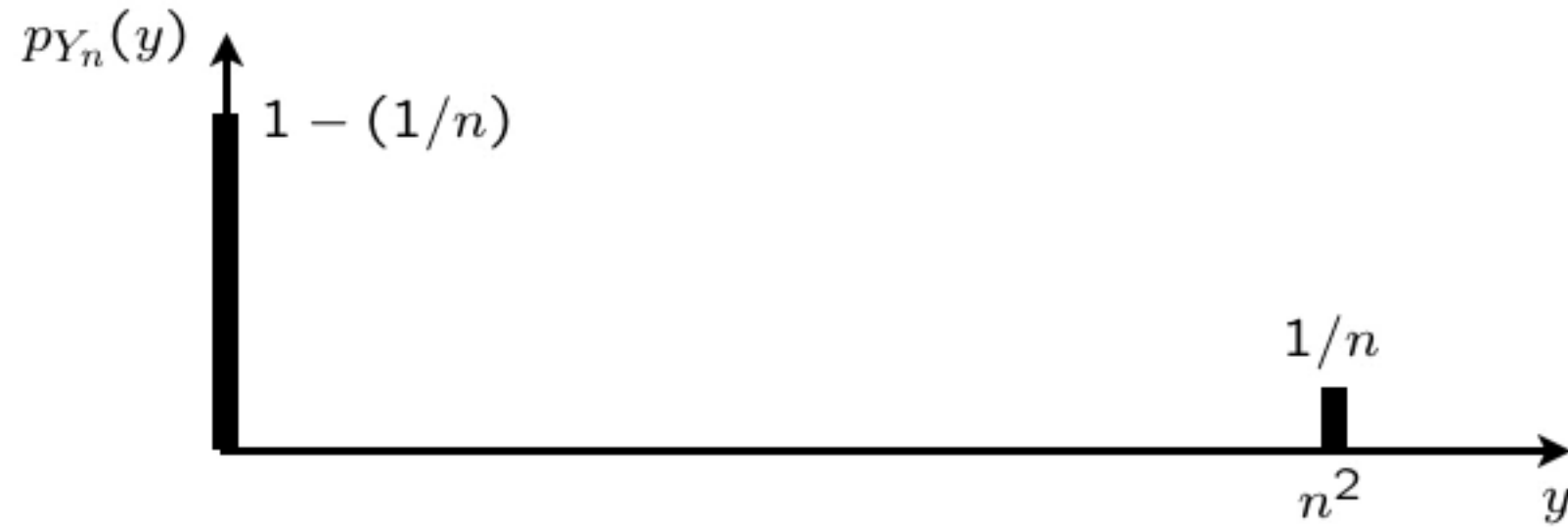
- If g is continuous, then $g(X_n) \rightarrow g(a)$

$$X_n^2 \rightarrow a^2$$

- $X_n + Y_n \rightarrow a + b$

- **But:** $E[X_n]$ need not converge to a

Convergence in probability examples



$$Y_n \xrightarrow[n \rightarrow \infty]{i.p.} 0.$$

$$\varepsilon > 0 \quad P(|Y_n - 0| \geq \varepsilon) = 1/n \xrightarrow[n \rightarrow \infty]{} 0$$

$$E[Y_n] = n^2 \cdot \frac{1}{n} = n \xrightarrow[n \rightarrow \infty]{} \infty$$

- convergence in probability does **not** imply convergence of expectations

Convergence in probability examples

- X_i : i.i.d., uniform on $[0, 1]$
- $Y_n = \min\{X_1, \dots, X_n\}$



$$Y_{n+1} \leq Y_n$$

$$\mathbf{P}(|Y_n - 0| \geq \epsilon) = \mathbf{P}(Y_n \geq \epsilon).$$

$$\epsilon > 0$$

$$= \mathbf{P}(X_1 \geq \epsilon, \dots, X_n \geq \epsilon)$$

$$Y_n \xrightarrow[n \rightarrow \infty]{i.p.} 0$$

$$\epsilon > 1$$

$$= \mathbf{P}(X_1 \geq \epsilon) \cdots \mathbf{P}(X_n \geq \epsilon)$$

$$\epsilon \leq 1$$

$$= (1 - \epsilon)^n \xrightarrow[n \rightarrow \infty]{} 0$$

Related topics

- Better bounds/approximations on tail probabilities

$$\leq \sim \frac{1}{n}$$

- Markov and Chebyshev inequalities

- Chernoff bound $P(|M_n - \mu| \geq a) \leq e^{-n \underbrace{h(a)}_{>0}}$

- Central limit theorem " $M_n \sim N(\mu, \sigma^2/n)$ "

- Different types of convergence

- Convergence in probability

- Convergence "with probability 1" $P\left(\left\{\omega : Y_n(\omega) \xrightarrow{n \rightarrow \infty} Y(\omega)\right\}\right) = 1$

- Strong law of large numbers $M_n \xrightarrow[n \rightarrow \infty]{\text{w.p.1}} \mu$

- Convergence of a sequence of distributions (CDFs) to a limiting CDF