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: $X, X_1, \ldots, X_n$ i.i.d.

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

  - 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 $\mathbf{P}(X \geq a) \leq \dfrac{\mathbf{E}[X]}{a}$.

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

$$a\,P(X \geq a) = E[Y] \leq E[X]$$

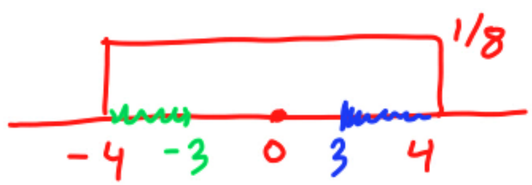
The Markov inequality

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

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

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

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



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: $\mathbf{P}\big(|X-\mu| \geq c\big) \leq \dfrac{\sigma^2}{c^2}$

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

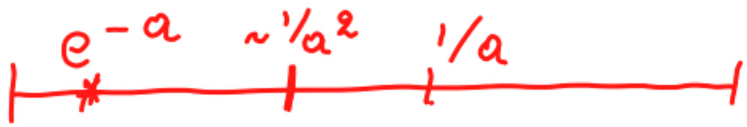
$$P(|X-\mu| \geq c) = P\left((X-\mu)^2 \geq c^2\right) \leq \frac{E\left[(X-\mu)^2\right]}{c^2} = \frac{\sigma^2}{c^2}$$

The Chebyshev inequality

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

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

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



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

The Weak Law of Large Numbers (WLLN)

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

  Sample mean: $M_n = \dfrac{X_1 + \cdots + X_n}{n}$ $\qquad \mu = E[X_i]$

- $\mathbf{E}[M_n] = \dfrac{E[X_1 + \cdots + X_n]}{n} = \dfrac{n\mu}{n} = \mu$

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

  $\mathbf{P}(|M_n - \mu| \geq \epsilon) \;\leq\; \dfrac{\text{var}(M_n)}{\varepsilon^2} = \dfrac{\sigma^2}{n\varepsilon^2} \xrightarrow[n\to\infty]{} 0$ $\quad$ (fixed $\varepsilon > 0$)

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

Interpreting the WLLN

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

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

- 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

$X_i = 1$, if A occurs
0, o.w.

$E[X_i] = p$

The pollster's problem

- p: fraction of population that will vote "yes" in a referendum
- ith (randomly selected) person polled: $X_i = \begin{cases} 1, & \text{if yes,} \\ 0, & \text{if no.} \end{cases}$

uniformly, independently

$E[X_i] = p$

- $M_n = (X_1 + \cdots + X_n)/n$: fraction of "yes" in our sample
- Would like "small error," e.g.: $|M_n - p| < 0.01$
- Try $n = 10{,}000$

Sample size

- $\mathbf{P}\big(|M_{10,000} - p| \geq 0.01\big) \leq \dfrac{\sigma^2}{n\varepsilon^2} = \dfrac{p(1-p)}{10^4 \cdot 10^{-4}} \leq \dfrac{1}{4}$ ← want $\leq 5\%$

Accuracy

Probability of Large error

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

Convergence "in probability"

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

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

- 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

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

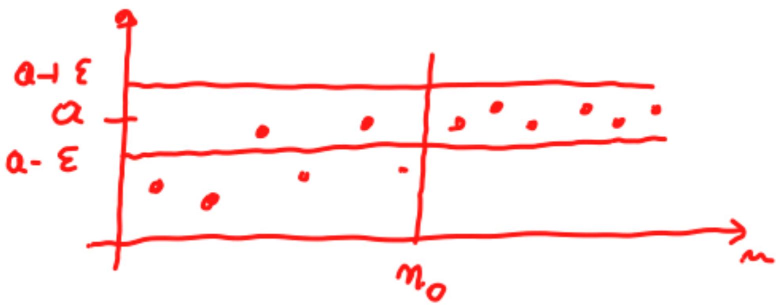
for any $\epsilon > 0$, $\lim_{n\to\infty} \mathbf{P}\Big(|Y_n - a| \geq \epsilon\Big) = 0$

Understanding convergence "in probability"

- Ordinary convergence
  - Sequence a_n; number a

    $a_n \to a$

  "a_n eventually gets and stays (arbitrarily) close to a"

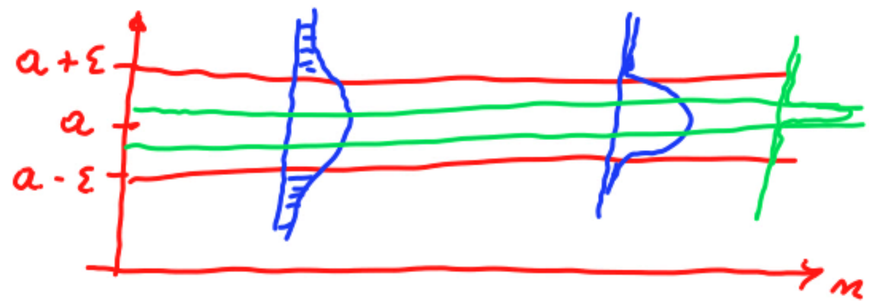


- For every $\epsilon > 0$, there exists n_0, such that for every $n \geq n_0$, we have $|a_n - a| \leq \epsilon$

- Convergence in probability
  - Sequence Y_n; number a

    $Y_n \to a$

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



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

Some properties

- Suppose that $X_n \to a$, $Y_n \to b$, in probability
- If g is continuous, then $g(X_n) \to g(a)$ $\quad X_n^2 \to a^2$
- $X_n + Y_n \to a + b$
- **But:** $\mathbf{E}[X_n]$ need not converge to a

Convergence in probability examples

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

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

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

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

Convergence in probability examples

- X_i: i.i.d., uniform on $[0,1]$

- $Y_n = \min\{X_1, \ldots, X_n\}$

$Y_{n+1} \le Y_n$

$$\mathbf{P}\Big(|Y_n - 0| \ge \epsilon\Big) = P(Y_n \ge \varepsilon)$$

$\varepsilon > 0$

$$= P(X_1 \ge \varepsilon, \ldots, X_n \ge \varepsilon)$$

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

$\varepsilon > 1$

$$= P(X_1 \ge \varepsilon) \cdots P(X_n \ge \varepsilon)$$

$\varepsilon \le 1$

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

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" $P(\{w : Y_n(w) \xrightarrow[n\to\infty]{} Y(w)\}) = 1$
  - Strong law of large numbers $M_n \xrightarrow[n\to\infty]{wp1} \mu$
  - Convergence of a sequence of distributions (CDFs) to a limiting CDF