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

The Markov inequality

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

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

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

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

The Chebyshev inequality

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

$$\mathbf{P}\Big(|X-\mu|\geq k\sigma\Big) \;\leq$$

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

The Weak Law of Large Numbers (WLLN)

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

  Sample mean: $M_n = \dfrac{X_1 + \cdots + X_n}{n}$

- $\mathbf{E}[M_n] =$

- $\text{Var}(M_n) =$

$\mathbf{P}(|M_n - \mu| \geq \epsilon) \;\; \leq$

WLLN: For $\epsilon > 0$, $\mathbf{P}\Big(|M_n - \mu| \geq \epsilon\Big) = \mathbf{P}\Big(\Big|\dfrac{X_1 + \cdots + X_n}{n} - \mu\Big| \geq \epsilon\Big) \to 0$, 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
  - empirical frequency is unlikely to be far of from true probability p

The pollster's problem

- p: fraction of population that will vote "yes" in a referendum
- ith (randomly selected) person polled: $X_i = \begin{cases} 1, & \text{if yes,} \\ 0, & \text{if no.} \end{cases}$
- $M_n = (X_1 + \cdots + X_n)/n$: fraction of "yes" in our sample
- Would like "small error," e.g.: $|M_n - p| < 0.01$
- Try $n = 10,000$
- $\mathbf{P}\big(|M_{10,000} - p| \geq 0.01\big) \leq$

Convergence "in probability"

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

- 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

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"

- Convergence in probability
  - Sequence Y_n; number a

    $Y_n \to a$

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

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

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

Some properties

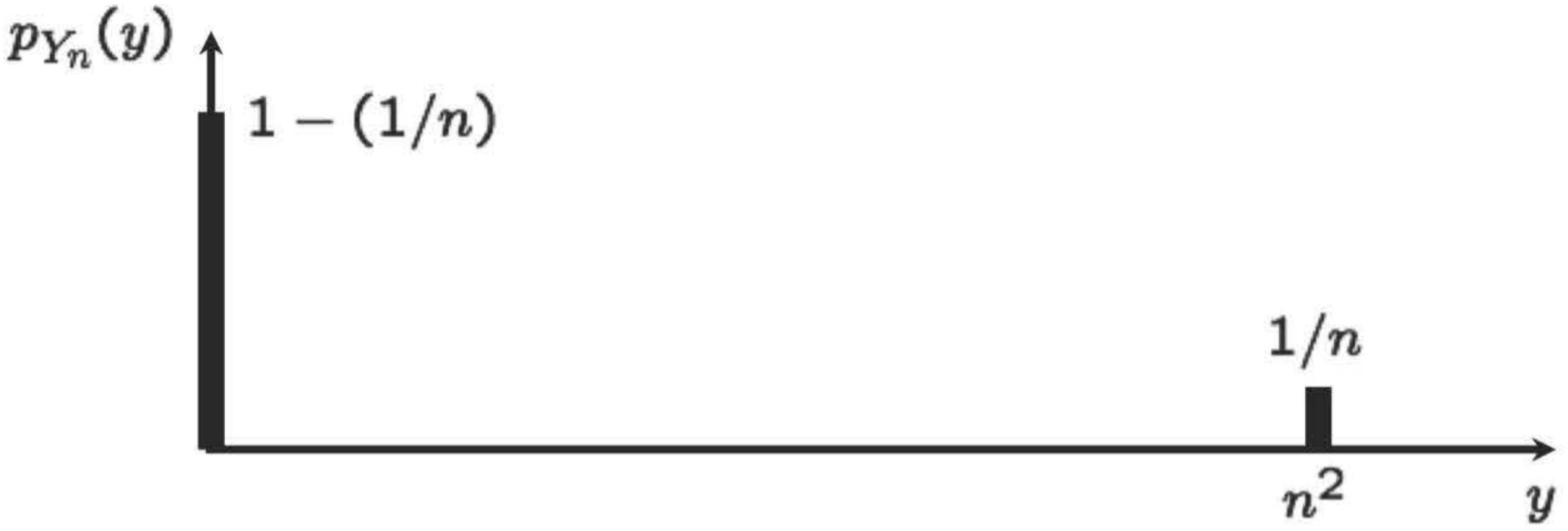
- Suppose that $X_n \to a$, $Y_n \to b$, in probability

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

- $X_n + Y_n \to a + b$

- **But:** $\mathbf{E}[X_n]$ need not converge to a

Convergence in probability examples



$$\mathbf{E}[Y_n] =$$

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

Convergence in probability examples

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

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

Related topics

- Better bounds/approximations on tail probabilities
  - Markov and Chebyshev inequalities
  - Chernoff bound
  - Central limit theorem
- Different types of convergence
  - Convergence in probability
  - Convergence “with probability 1”
  - Strong law of large numbers
  - Convergence of a sequence of distributions (CDFs) to a limiting CDF