LECTURE 15

- Readings: Sections 7.1-7.3

Lecture outline

- Limit theorems:
  - Chebyshev inequality
  - Convergence in probability

Motivation

$X_1, \cdots, X_n$ i.i.d.,

$$M_n = \frac{X_1 + \cdots + X_n}{n} \qquad \text{(sample mean)}$$

What happens as $n \longrightarrow \infty$?

- Why bother?
- A tool: Chebyshev's inequality.
- Convergence "in probability".
- Convergence of M_n .

Chebyshev's Inequality

- Random variable X :

$$\sigma^2 = \int (x - \mathbf{E}[X])^2 f_X(x)\, dx$$

$$\sigma^2 \;\geq\; c^2 \mathbf{P}(|X - \mathbf{E}[X]| \geq c)$$

$$\mathbf{P}(|X - \mathbf{E}[X]| \geq c) \;\leq\; \frac{\sigma^2}{c^2}$$

$$\mathbf{P}(|X - \mathbf{E}[X]| \geq k\sigma) \;\leq\; \frac{1}{k^2}$$

Deterministic Limits: Review

- We have a: — <u>Sequence</u>: a_n
  — <u>Number</u>: a

- We say that a_n converges to a ,

  and write: $\lim_{n \to \infty} a_n = a$

- If (intuitively): "a_n eventually gets and stays (arbitrarily) close to a".

- If (rigorously): For every $\epsilon > 0$ there exists n_0, such that for all $n \geq n_0$, we have: $|a_n - a| \leq \epsilon$

Convergence "in probability"

- We have a sequence of random variables: Y_n
- We say that Y_n converges to a number a:
- If (intuitively):

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

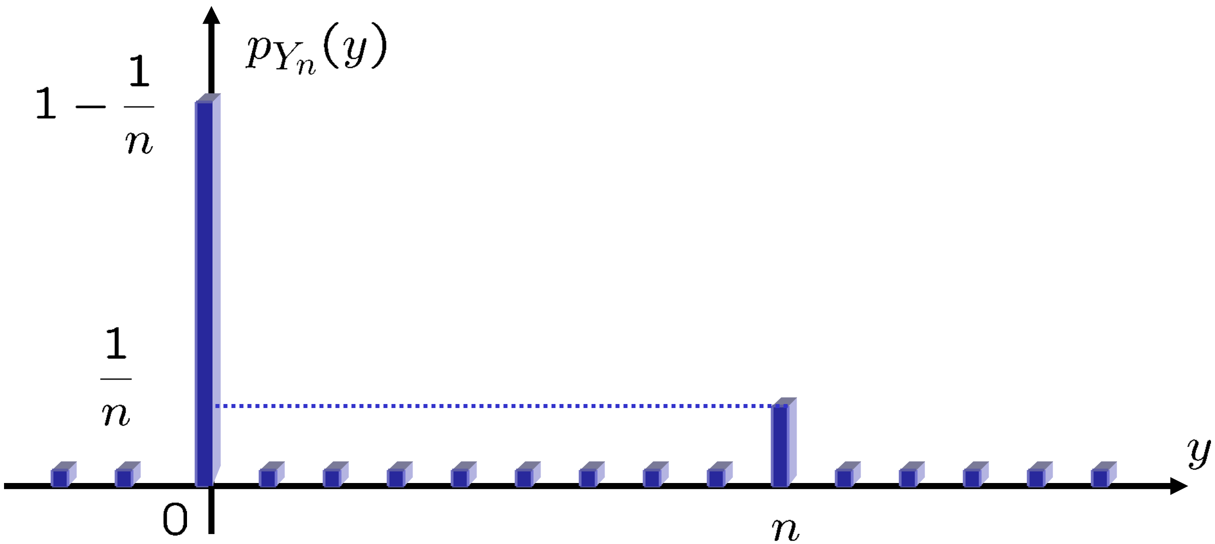
- If (rigorously):

  For every $\epsilon > 0$, we have:

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

Example

- Consider a sequence of random variables with the following sequence of PMFs:



- Does Y_n converge?

$$\lim_{n\to\infty} \mathbf{P}(|Y_n - 0| \geq \epsilon) = \lim_{n\to\infty} \frac{1}{n} = 0$$

- What is $\mathbf{E}[Y_n]$?

Convergence of the Sample Mean

$X_1, \cdots, X_n$ i.i.d., (finite mean μ and variance σ^2)

$$M_n = \frac{X_1 + \cdots + X_n}{n}$$

- **Mean:** $\mathbf{E}[M_n] = \mu$
- **Variance:** $\mathsf{Var}(M_n) = \frac{\sigma^2}{n}$
- **Chebyshev**: $\mathbf{P}(|M_n - \mathbf{E}[M_n]| \geq \epsilon) \leq \frac{\mathsf{Var}(M_n)}{\epsilon^2}$
- **Limit:** $\mathbf{P}(|M_n - \mu| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2}$

The Pollster's Problem

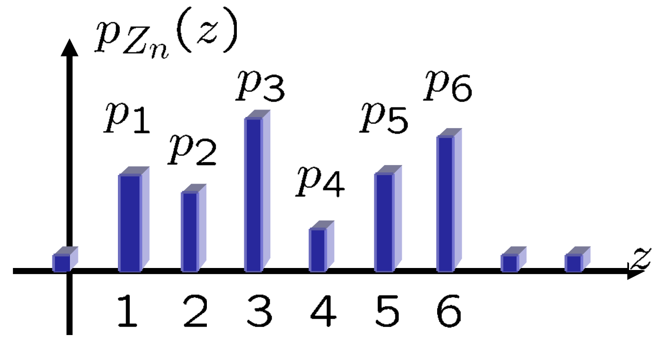
- f : fraction of population that do "...........".
- i^{th} person polled: $X_i = \begin{cases} 1 & \text{If "Yes".} \\ 0 & \text{If "No".} \end{cases}$
- $M_n = \frac{X_1 + \cdots + X_n}{n}$: fraction of "Yes" in our sample.
- Suppose we want: $\mathbf{P}(|M_n - f| \geq .01) \leq .05$
- Chebyshev: $\mathbf{P}(|M_n - \mu_x| \geq \epsilon) \leq \frac{\sigma_x^2}{n\epsilon^2}$

  But we have : $\mu_x = f \qquad \sigma_x^2 = f(1-f) \leq \frac{1}{4}$
- Thus: $\mathbf{P}(|M_n - f| \geq .01) \leq \frac{1}{.0004n}$
- So, let $n > 50,000$ (conservative).

Die Experiment (1)

- Unfair die, with probability of face $i = p_i$.
- Independent throws: $Z_n =$ Value of n^{th} throw.

  Thus, Z_n are i.i.d. with PMF:



- Define: $X_n = \begin{cases} 1 & \text{If } Z_n = i. \\ 0 & \text{Otherwise.} \end{cases}$
- Let: $M_n = \frac{X_1 + \cdots + X_n}{n}$ "frequency of face i"

Die Experiment (2)

- X_n is Bernoulli with probability p_i, thus:

$$\mu_x = p_i \qquad \sigma_x^2 = p_i(1-p_i) \leq \frac{1}{4}$$

Then: $\mathbf{E}[M_n] = p_i \qquad \text{Var}(M_n) = \frac{\sigma_x^2}{n} \leq \frac{1}{4n}$

- Chebyshev: $\mathbf{P}(|M_n - p_i| \geq \epsilon) \leq \dfrac{1}{4n\epsilon^2}$
- It follows that: $\lim_{n\to\infty} \mathbf{P}(|M_n - p_i| \geq \epsilon) = 0$
- Therefore, the sample frequency of each face converges "in probability" to the probability of that face.
- This allows us to do "simulations".