LECTURE 19
Limit theorems – I

- **Readings:** Sections 5.1-5.3; start Section 5.4

- $X_1, \ldots, X_n$ i.i.d.

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

  What happens as $n \to \infty$?

- Why bother?
- A tool: Chebyshev's inequality
- Convergence "in probability"
- Convergence of M_n (weak law of large numbers)

Chebyshev's inequality

- Random variable X (with finite mean μ and variance σ^2)

$$\begin{aligned}\sigma^2 &= \int (x-\mu)^2 f_X(x)\,dx \\ &\geq \int_{-\infty}^{-c} (x-\mu)^2 f_X(x)\,dx + \int_c^{\infty} (x-\mu)^2 f_X(x)\,dx \\ &\geq c^2 \cdot \mathbf{P}(|X-\mu| \geq c)\end{aligned}$$

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

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

Deterministic limits

- Sequence a_n
  Number a

- a_n converges to a

$$\lim_{n\to\infty} a_n = a$$

  "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 $|a_n - a| \leq \epsilon$.

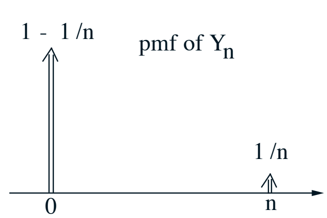
Convergence "in probability"

- Sequence of random variables Y_n

- converges in probability to a number a:
  "(almost all) of the PMF/PDF of Y_n, eventually gets concentrated (arbitrarily) close to a"

- For every $\epsilon > 0$,

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



Does Y_n converge?

Convergence of the sample mean

(Weak law of large numbers)

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

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

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

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

$$\mathbf{P}(|M_n - \mu| \geq \epsilon) \;\leq\; \frac{\text{Var}(M_n)}{\epsilon^2} \;=\; \frac{\sigma^2}{n\epsilon^2}$$

- M_n converges in probability to μ

The pollster's problem

- f: fraction of population that "..."
- 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
- Goal: 95% confidence of $\leq$1% error

$$\mathbf{P}(|M_n - f| \geq .01) \leq .05$$

- Use Chebyshev's inequality:

$$\mathbf{P}(|M_n - f| \geq .01) \;\leq\; \frac{\sigma^2_{M_n}}{(0.01)^2} = \frac{\sigma^2_x}{n(0.01)^2} \leq \frac{1}{4n(0.01)^2}$$

- If $n = 50,000$,
  then $\mathbf{P}(|M_n - f| \geq .01) \leq .05$
  (conservative)

Different scalings of M_n

- $X_1, \ldots, X_n$ i.i.d.
  finite variance σ^2
- Look at three variants of their sum:
- $S_n = X_1 + \cdots + X_n$ variance $n\sigma^2$
- $M_n = \frac{S_n}{n}$ variance σ^2/n
  converges "in probability" to $\mathbf{E}[X]$ (WLLN)
- $\frac{S_n}{\sqrt{n}}$ constant variance σ^2
  - Asymptotic shape?

The central limit theorem

- "Standardized" $S_n = X_1 + \cdots + X_n$:

$$Z_n = \frac{S_n - \mathbf{E}[S_n]}{\sigma_{S_n}} = \frac{S_n - n\mathbf{E}[X]}{\sqrt{n}\,\sigma}$$

  - zero mean
  - unit variance
- Let Z be a standard normal r.v.
  (zero mean, unit variance)
- **Theorem:** For every c:

$$\mathbf{P}(Z_n \leq c) \to \mathbf{P}(Z \leq c)$$

- $\mathbf{P}(Z \leq c)$ is the standard normal CDF, $\Phi(c)$, available from the normal tables