

LECTURE 19
Limit theorems – I

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

- X_1, \dots, X_n i.i.d.

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

What happens as $n \rightarrow \infty$?

- 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

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

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

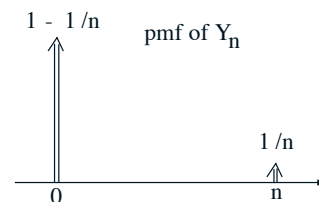
- For every $\epsilon > 0$, there exists n_0 , such that for every $n \geq n_0$, we have $|a_n - a| \leq \epsilon$.

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 $\epsilon > 0$,

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



Does Y_n converge?

Convergence of the sample mean

(Weak law of large numbers)

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

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

- $E[M_n] =$

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

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

- M_n converges in probability to μ

The pollster's problem

- f : fraction of population that "..."
- i th (randomly selected) person polled:

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

- $M_n = (X_1 + \dots + X_n)/n$
fraction of "yes" in our sample
- Goal: 95% confidence of $\leq 1\%$ error

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

- Use Chebyshev's inequality:

$$\begin{aligned} P(|M_n - f| \geq .01) &\leq \frac{\sigma_{M_n}^2}{(0.01)^2} \\ &= \frac{\sigma_x^2}{n(0.01)^2} \leq \frac{1}{4n(0.01)^2} \end{aligned}$$

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

Different scalings of M_n

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

The central limit theorem

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

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

- zero mean
- unit variance
- Let Z be a standard normal r.v.
(zero mean, unit variance)
- **Theorem:** For every c :
$$P(Z_n \leq c) \rightarrow P(Z \leq c)$$
- $P(Z \leq c)$ is the standard normal CDF,
 $\Phi(c)$, available from the normal tables

MIT OpenCourseWare
<http://ocw.mit.edu>

6.041 / 6.431 Probabilistic Systems Analysis and Applied Probability
Fall 2010

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.