LECTURE 22
THE CENTRAL LIMIT THEOREM

- $X_1, \ldots, X_n$ i.i.d.
  finite variance σ^2

- Look at three variants of their sum:

- $S_n = X_1 + \cdots + X_n$ variance $n\sigma^2$

- $M_n = \dfrac{S_n}{n}$ variance σ^2/n
  converges "in probability" to $\mathbf{E}[X]$ (WLLN)

- $\dfrac{S_n}{\sqrt{n}}$ 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:

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

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

What exactly does it say?

- CDF of Z_n converges to normal CDF
  - not a statement about convergence of PDFs or PMFs

Normal approximation

- Treat Z_n as if normal
  - also treat S_n as if normal

Can we use it when n is "moderate"?

- Yes, but no nice theorems to this effect

- Symmetry helps a lot

The pollster's problem using the CLT

- f: fraction of population that do XYZ

- ith person polled:

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

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

- Suppose we want:

$$\mathbf{P}(|M_n - f| \ge .01) \le .05$$

- Event of interest: $|M_n - f| \ge .01$

$$\left|\frac{X_1 + \cdots + X_n - nf}{n}\right| \ge .01$$

$$\left|\frac{X_1 + \cdots + X_n - nf}{\sqrt{n}\sigma}\right| \ge \frac{.01\sqrt{n}}{\sigma}$$

$$\begin{aligned} \mathbf{P}(|M_n - f| \ge .01) &\approx \mathbf{P}(|Z| \ge .01\sqrt{n}/\sigma) \\ &\le \mathbf{P}(|Z| \ge .02\sqrt{n}) \end{aligned}$$

Usefulness of the CLT

- only means and variances matter
- Much more accurate than Chebyshev's inequality
- Useful computational shortcut, even if we have a formula for the distribution of S_n
- Justification of models involving normal r.v.'s
  - Noise in electrical components
  - Motion of a particle suspended in a fluid (Brownian motion)

"Proof" of the CLT

- Assume for simplicity $\mathbf{E}[X] = 0$, $\sigma = 1$
- Need to show that

$$Z_n = \frac{X_1 + \cdots + X_n}{\sqrt{n}}$$

converges to standard normal.

$$M_{Z_n}(s) = \mathbf{E}[e^{sZ_n}] = \mathbf{E}\left[e^{(s/\sqrt{n})(X_1+\cdots+X_n)}\right]$$

$$\mathbf{E}[e^{sX/\sqrt{n}}] \approx 1 + \frac{s}{\sqrt{n}}\mathbf{E}[X] + \frac{s^2}{2n}\mathbf{E}[X^2]$$

$$M_{Z_n}(s) \approx \left(1 + \frac{s^2}{2n}\right)^n \longrightarrow e^{s^2/2}$$

which is the transform of standard normal