LECTURE 20
THE CENTRAL LIMIT THEOREM

- Readings: Section 5.4
- $X_1, \ldots, X_n$ i.i.d., finite variance σ^2
- "Standardized" $S_n = X_1 + \cdots + X_n$:

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

  – $\mathbf{E}[Z_n] = 0, \qquad \text{var}(Z_n) = 1$

- 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

Usefulness

- universal; only means, variances matter
- accurate computational shortcut
- justification of normal models

What exactly does it say?

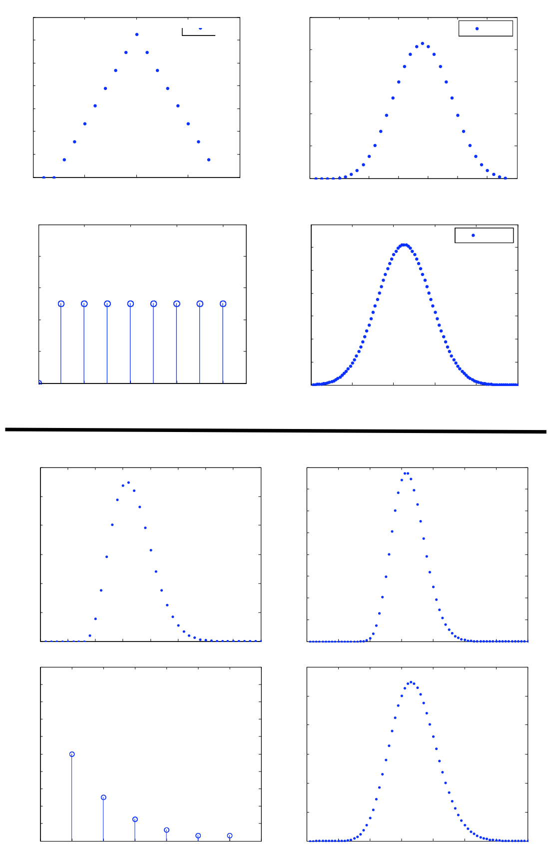
- CDF of Z_n converges to normal CDF
  – not a statement about convergence of PDFs or PMFs

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 "..."
- ith (randomly selected) person polled:

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

- $M_n = (X_1 + \cdots + X_n)/n$
- Suppose we want:

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

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

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

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

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

Apply to binomial

- Fix p, where $0 < p < 1$
- X_i: Bernoulli(p)
- $S_n = X_1 + \cdots + X_n$: Binomial(n, p)
  - mean np, variance $np(1-p)$
- CDF of $\dfrac{S_n - np}{\sqrt{np(1-p)}} \longrightarrow$ standard normal

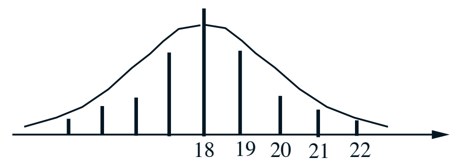
Example

- $n = 36$, $p = 0.5$; find $\mathbf{P}(S_n \le 21)$
- Exact answer:

$$\sum_{k=0}^{21} \binom{36}{k} \left(\frac{1}{2}\right)^{36} = 0.8785$$

The 1/2 correction for binomial approximation

- $\mathbf{P}(S_n \le 21) = \mathbf{P}(S_n < 22)$, because S_n is integer
- Compromise: consider $\mathbf{P}(S_n \le 21.5)$



De Moivre–Laplace CLT (for binomial)

- When the 1/2 correction is used, CLT can also approximate the binomial p.m.f. (not just the binomial CDF)

$$\mathbf{P}(S_n = 19) = \mathbf{P}(18.5 \le S_n \le 19.5)$$

$$18.5 \le S_n \le 19.5 \quad \Longleftrightarrow$$

$$\frac{18.5 - 18}{3} \le \frac{S_n - 18}{3} \le \frac{19.5 - 18}{3} \quad \Longleftrightarrow$$

$$0.17 \le Z_n \le 0.5$$

$$\begin{aligned}
\mathbf{P}(S_n = 19) &\approx \mathbf{P}(0.17 \le Z \le 0.5) \\
&= \mathbf{P}(Z \le 0.5) - \mathbf{P}(Z \le 0.17) \\
&= 0.6915 - 0.5675 \\
&= 0.124
\end{aligned}$$

- Exact answer:

$$\binom{36}{19} \left(\frac{1}{2}\right)^{36} = 0.1251$$

Poisson vs. normal approximations of the binomial

- Poisson arrivals during unit interval equals: sum of n (independent) Poisson arrivals during n intervals of length $1/n$
  - Let $n \to \infty$, apply CLT (??)
  - Poisson=normal (????)
- Binomial(n, p)
  - p fixed, $n \to \infty$: normal
  - np fixed, $n \to \infty$, $p \to 0$: Poisson
- $p = 1/100$, $n = 100$: Poisson
- $p = 1/10$, $n = 500$: normal