MATH 180A: INTRO TO PROBABILITY
(FOR DATA SCIENCE)

www.math.ucsd.edu/~tkemp/180A

Today: § 4.1 – 4.3

Next: § 4.3 – 4.4

Midterm grades released; regrade requests Wed 10/30 8am–11pm

Lab 3 grades released; regrade requests Thu 10/31 8am–11pm

HW4 due Friday by 11:59pm

Lab 4 due next Wednesday (Nov 6) by 11:59pm

Poisson Approximation of Binomial

Fix $\lambda > 0$. For each n, let $T_n \sim \mathrm{Bin}(n, \lambda/n)$.
For any fixed $k \in \mathbb{N}$,

$$\lim_{n\to\infty} \mathbb{P}(T_n = k) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

In particular, for $a \le b$ positive numbers,

$$\lim_{n\to\infty} \mathbb{P}(a \le T_n \le b) = \sum_{\substack{a \le k \le b \\ k \in \mathbb{N}}} e^{-\lambda} \frac{\lambda^k}{k!}.$$

Normal Approximation of Binomial

Fix $p \in (0,1)$. For each n, let $S_n \sim \mathrm{Bin}(n,p)$.
For any fixed $a \le b$,

$$\lim_{n\to\infty} \mathbb{P}\left(a \le \frac{S_n - np}{\sqrt{np(1-p)}} \le b \right) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx.$$

Question:

A fair 20-sided die is tossed 400 times. We want to calculate the probability that a 13 came up at least 25 times. We should use:

(a) Poisson Approximation.
(b) Normal Approximation.
(c) Either.
(d) Neither.

Eg. A fair die is rolled 720 times. What is the probability that exactly 113 sixes come up?

S = # sixes $\sim \text{Bin}(720, \frac{1}{6})$. $\mathbb{P}(S=113)$

Normal Approximation:

$\mathbb{E}(S) = 720 \cdot \frac{1}{6} = 120 \qquad \text{Var}(S) = 720 \cdot \frac{1}{6} \cdot \frac{5}{6} = 100$

$\mathbb{P}(S=113)$

Fun Example: Roll a fair die. If it comes up 1 or 2, take 2 steps.
If it comes up 3 or more, take 3 steps.

Repeat. After 450 rolls, how likely is it you've taken more than 1185 steps?

What is $\mathbb{E}$?

Precise definition:

Intuition:

Binomial $S_n \sim \mathrm{Bin}(n,p)$ $\quad \mathbb{E}(S_n) =$

Theorem (Law of Large Numbers for Bernoulli Trials)

Let $X_1, X_2, \dots, X_n$ be independent Bernoulli trials with success probability p. Then

4.2

If $S_n \sim \text{Bin}(n,p)$ (p fixed), and $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(\left| \frac{S_n}{n} - p \right| > \varepsilon \right) = 0 .$$

Proof:

Example

Flip a fair coin n times. How does

$$\mathbb{P}\left(\frac{\#\text{Heads}}{n} \geq 50.01\%\right)$$

behave as $n \to \infty$?

Suppose after 10,000 flips, there are 5,001 Heads. Should we doubt that the coin is really fair?

What if, after 1,000,000 flips, there are 500,100 Heads. Now how confident should we be that the coin is really fair?

Confidence Intervals

4.3

Suppose we have a coin that is biased by some unknown amount;

$$X \sim \mathrm{Ber}(p) \quad \text{unknown } p\,!$$

How can we figure out what p is?

Use the law of large numbers: $p = \lim_{n\to\infty} \frac{S_n}{n}$

We can't actually wait around for $n \to \infty$. Instead, we estimate

$$p \approx \hat{p} := \frac{S_n}{n} \quad \text{for some large } n.$$

The question is: how good an estimate is this for given n? Or, turning it around: how big must you take n to get an estimate of a certain accuracy?

A Maximum Likelihood Estimate

Want to find n large enough that (with $\hat{p} = S_n/n$)

$$\mathbb{P}(|\hat{p} - p| < \varepsilon) \geq \text{(high probability)}$$

$\uparrow$ chosen tolerance

$$\mathbb{P}(|\hat{p} - p| < \varepsilon) \approx 2\,\Phi\left(\varepsilon\sqrt{n}\big/\sqrt{p(1-p)}\right) - 1 .$$

Conclusion: $\mathbb{P}(|\hat{p} - p| < \varepsilon) \underset{(\approx)}{\geq} 2\,\Phi(\qquad) - 1 .$

Example (of the Beast): How many times should we flip a coin, biased an unknown amount p, so that the estimate $\hat{p} = S_n/n$ is within a tolerance of 0.05 of the true value p, with probability $\geq 99\%$?