CS109: Probability for Computer Scientists

Lecture 15 Worksheet — Central Limit Theorem (TPS)

Feb 11, 2026

1 Quick Check: Are These i.i.d.?

For each case below, decide whether the random variables are i.i.d. and why.

a) $X \sim \text{Ber}(0.5)$, $Y \sim \text{Ber}(0.5)$, and X, Y are independent.
Solution: Yes, i.i.d. They are independent and have the same distribution.

b) $X \sim \text{Ber}(0.3)$, $Y \sim \text{Ber}(0.7)$, and X, Y are independent.
Solution: Not i.i.d. They are independent, but not identically distributed.

c) $X \sim \text{Ber}(p)$ and $Y = 1 - X$.
Solution: Not i.i.d. Y has the same Bernoulli form only if $p = 0.5$, but in general X, Y are dependent since knowing X determines Y.

2 Convolution Intuition: Probability of a Tie

Let X be your score and Y your opponent's score in a game. Assume X and Y are independent and discrete with known PMFs $P(X = x)$ and $P(Y = y)$.

a) Write an expression for $P(X = Y)$ in terms of the PMFs.
Solution: $P(X = Y) = \sum_x P(X = x)P(Y = x)$ (independence lets us multiply).

b) Using the same setup, write an expression for $P(X + Y = 10)$.
Solution: $P(X + Y = 10) = \sum_i P(X = i)P(Y = 10 - i)$.

3 Wildlife Disease Outbreak (Exact Distribution)

Population A has infections $A \sim \text{Bin}(5, 0.1)$ and population B has infections $B \sim \text{Bin}(8, 0.5)$, with A, B independent.

a) Write a formula for $P(A + B = k)$. What are the possible values of k?
Solution:

$$P(A + B = k) = \sum_{i=\max(0,k-8)}^{\min(5,k)} P(A = i)P(B = k - i),$$

with

$$P(A = i) = \binom{5}{i}(0.1)^i(0.9)^{5-i}, \quad P(B = k - i) = \binom{8}{k-i}(0.5)^{k-i}(0.5)^{8-(k-i)}.$$

Possible values: $k = 0, 1, \ldots, 13$.

4 Linear Transform (Bug Hunt)

Let $X \sim \mathcal{N}(\mu, \sigma^2)$ and $Y = X + X = 2X$. Here are two ways students might think to reason about Y:

Version A: Thinking of Y as a linear transform $Y = 2X \Longrightarrow Y \sim \mathcal{N}(2\mu, 4\sigma^2)$.

Version B: Thinking of Y as a sum of independent normals

$$Y = X + X \Longrightarrow Y \sim \mathcal{N}(\mu + \mu, \sigma^2 + \sigma^2) = \mathcal{N}(2\mu, 2\sigma^2).$$

Notice these two different framings lead to different variances. Which one of these framings is the correct one, and why?
Solution: Version A is correct. Version B incorrectly treats the two X's as independent copies. But X is not independent of itself. So $\text{Var}(X + X) = \text{Var}(2X) = 4\sigma^2$, not $2\sigma^2$.

5 ELO Model and Subtracting Gaussians

In an ELO-style model, team abilities are sampled as

$$A_W \sim \mathcal{N}(S_W, 200^2), \qquad A_O \sim \mathcal{N}(S_O, 200^2),$$

independently. The Warriors win if $A_W > A_O$.

a) Could you come up with a way to define $D = A_W - A_O$?
Solution: Define $D = A_W - A_O$. Since independent normals subtract to a normal,

$$D \sim \mathcal{N}\big(S_W - S_O,\ 200^2 + 200^2\big) = \mathcal{N}(S_W - S_O,\ 80000).$$

Then

$$P(\text{Warriors win}) = P(A_W > A_O) = P(D > 0) = \Phi\left(\frac{S_W - S_O}{200\sqrt{2}}\right).$$

6 Virus Infections Revisited

Let $A \sim \text{Bin}(50, 0.1)$, and $B \sim \text{Bin}(100, 0.4)$ where A and B are independent. We want $P(A + B \geq 40)$. I want to solve this without using a for loop.

a) What is the insight you would use to get you started solving this problem?
Solution: Use normal approximation (CLT / binomial-to-normal) for each binomial, then sum normals.

b) Solve it here.
Solution:

$$E[A] = 50(0.1) = 5, \quad \text{Var}(A) = 50(0.1)(0.9) = 4.5$$
$$E[B] = 100(0.4) = 40, \quad \text{Var}(B) = 100(0.4)(0.6) = 24.$$

So

$$W = A + B \approx \mathcal{N}(45,\ 28.5).$$

With continuity correction:

$$P(A + B \geq 40) \approx P(W \geq 39.5).$$
$$z = \frac{39.5 - 45}{\sqrt{28.5}} \approx -1.03.$$
$$P(W \geq 39.5) = 1 - \Phi(-1.03) = \Phi(1.03) \approx 0.848.$$

Approximate answer: $\boxed{0.85}$.

7 Sum of 100 Dice Rolls

Let $X_1, \dots, X_{100}$ be i.i.d. fair six-sided dice and $S = \sum_{i=1}^{100} X_i$.

a) How could you use the CLT to approximate S?
Solution: For one die:

$$E[X_i] = 3.5, \qquad \mathrm{Var}(X_i) = \frac{35}{12}.$$

So

$$E[S] = 100(3.5) = 350, \qquad \mathrm{Var}(S) = 100 \cdot \frac{35}{12} = \frac{3500}{12}.$$

By CLT,

$$S \approx \mathcal{N}\left(350, \frac{3500}{12}\right).$$

Then approximate any probability on S using this normal (with continuity correction for integer thresholds).

8 Dice Game

You roll 10 fair six-sided dice such that $X = \sum_{i=1}^{10} X_i$. You win if $X \leq 25$ or $X \geq 45$.

a) Use CLT to approximate $P(X \leq 25 \text{ or } X \geq 45)$.
Solution: For one die: $E[X_i] = 3.5,\ \mathrm{Var}(X_i) = 35/12$. So

$$E[X] = 35, \qquad \mathrm{Var}(X) = 10 \cdot \frac{35}{12} = \frac{350}{12}, \quad \sigma \approx 5.40.$$

CLT: $X \approx \mathcal{N}(35, 350/12)$.

Continuity correction:

$$P(X \leq 25 \text{ or } X \geq 45) \approx P(X \leq 25.5) + P(X \geq 44.5).$$

$$z_1 = \frac{25.5 - 35}{5.40} \approx -1.76, \qquad z_2 = \frac{44.5 - 35}{5.40} \approx 1.76.$$

Thus

$$\approx \Phi(-1.76) + \big(1 - \Phi(1.76)\big) \approx 0.039 + 0.039 = 0.078.$$

Approximate answer: $\boxed{0.079 \text{ (about } 7.9\%)}$.