CS109 Lecture 7 Worksheet: Variance and the Poisson

Spread of a Random Variable, and the Poisson Distribution

Summer 2026

Lecture 7 Worksheet: Variance and the Poisson

Topics: variance $\text{Var}(X) = E[(X-\mu)^2] = E[X^2] - (E[X])^2$, standard deviation, variance of the classic random variables, and the Poisson distribution $P(X=k) = \frac{e^{-\lambda}\lambda^k}{k!}$ (including its use as an approximation to the binomial).

How to use this sheet. Work in small groups. A TA and the instructor will circulate to help. We will go over selected solutions on the board. You do not need to finish every part; aim for understanding over speed. You may leave Poisson answers as expressions involving $e^{-\lambda}$ unless a decimal is requested; a quick `scipy.stats.poisson` check is suggested where useful.

Problem 1 (Review of Lecture 6): Expectation

Each person you date is a life partner with probability 0.2, independently, and you date until your first life partner. Let X be the number of people you date (including the match).

(a) Which classic random variable is X, and what is its parameter?

(b) What is $E[X]$?

(c) Separately, let $Y \sim \text{Bin}(500, 0.53)$ be Shaq's makes in 500 free throws. Use linearity / the binomial mean to find $E[Y]$.

Problem 2: Defining Variance

Three different random variables all have mean $E[X] = 3$, but different "spreads."

(a) In words, what does $\text{Var}(X) = E[(X-\mu)^2]$ measure, and why do we square the deviation $X-\mu$ rather than just averaging $X-\mu$?

(b) State the computational formula $\text{Var}(X) = E[X^2] - (E[X])^2$ and explain why standard deviation $\text{SD}(X) = \sqrt{\text{Var}(X)}$ is often the more interpretable quantity (think about units).

Problem 3: Variance of a Die Roll

Let X be the result of one roll of a fair six-sided die. You may use $E[X] = 3.5$ and $E[X^2] = \frac{91}{6}$ from last lecture.

(a) Compute $\text{Var}(X) = E[X^2] - (E[X])^2$.

(b) Compute the standard deviation $\text{SD}(X)$.

Problem 4: Variance of the Classic Random Variables

Use the formulas $\text{Var}(\text{Bern}(p)) = p(1-p)$ and $\text{Var}(\text{Bin}(n,p)) = np(1-p)$.

(a) A single free throw (made with probability 0.53) is a Bernoulli random variable. Find its variance.

(b) Shaq's makes in 500 free throws is $Y \sim \text{Bin}(500, 0.53)$. Find $\text{Var}(Y)$ and $\text{SD}(Y)$.

(c) For a fixed n, which value of p makes a binomial's variance **largest**? (Hint: maximize $p(1-p)$.) Briefly, why does that value give the most spread?

Problem 5: Meet the Poisson

The number of major earthquakes in a year is modeled as $X \sim \text{Poi}(\lambda = 2.79)$, with PMF $P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$.

(a) Write an expression for the probability of **exactly 3** major earthquakes next year, then give a decimal (a quick check: about 0.222).

(b) Write an expression for the probability of **zero** earthquakes next year.

(c) State $E[X]$ and $\text{Var}(X)$ for a Poisson. (For the Poisson these are equal; what are they here?)

Problem 6: Poisson Approximation to a Binomial (pset2: hash_table_poi)

A hash table has $n = 2000$ buckets, and $m = 10000$ strings are hashed in, each equally likely to land in any bucket. Let X be the number of strings that land in the **first** bucket.

(a) Exactly, $X \sim \text{Bin}(10000,\ 1/2000)$. Explain why this is well approximated by a Poisson, and find the rate λ for that Poisson.

(b) Using the Poisson approximation, write an expression for the probability that the first bucket has **8 or fewer** strings, $P(X \leq 8)$, and give a decimal to 4 places (a quick check: about 0.9319).

Problem 7: Web Server Load

A web server receives hits at an average rate of 2 per second. Let X be the number of hits in a given second, modeled as $X \sim \text{Poi}(2)$.

(a) Write an expression for $P(X < 5)$ (i.e. $P(X \leq 4)$) as a sum of Poisson terms.

(b) Give a decimal value (a quick check: about 0.947).

(c) The rate is "2 per second." If instead you asked about a **5-second** window, what rate λ would you use, and why must the rate's time unit match the question?

Challenge Problem (optional): DNA Data Storage

A strand stores $n = 10^4$ base pairs, each independently corrupted with probability $p = 10^{-6}$. Let X be the number of corrupted base pairs.

(a) Give the exact distribution of X, and the Poisson that approximates it (state $\lambda = np$).

(b) Find the probability that **at least one** base pair is corrupted, $P(X \geq 1)$, using the Poisson approximation and the complement. Why is the Poisson approximation especially convenient here compared with evaluating the binomial directly?