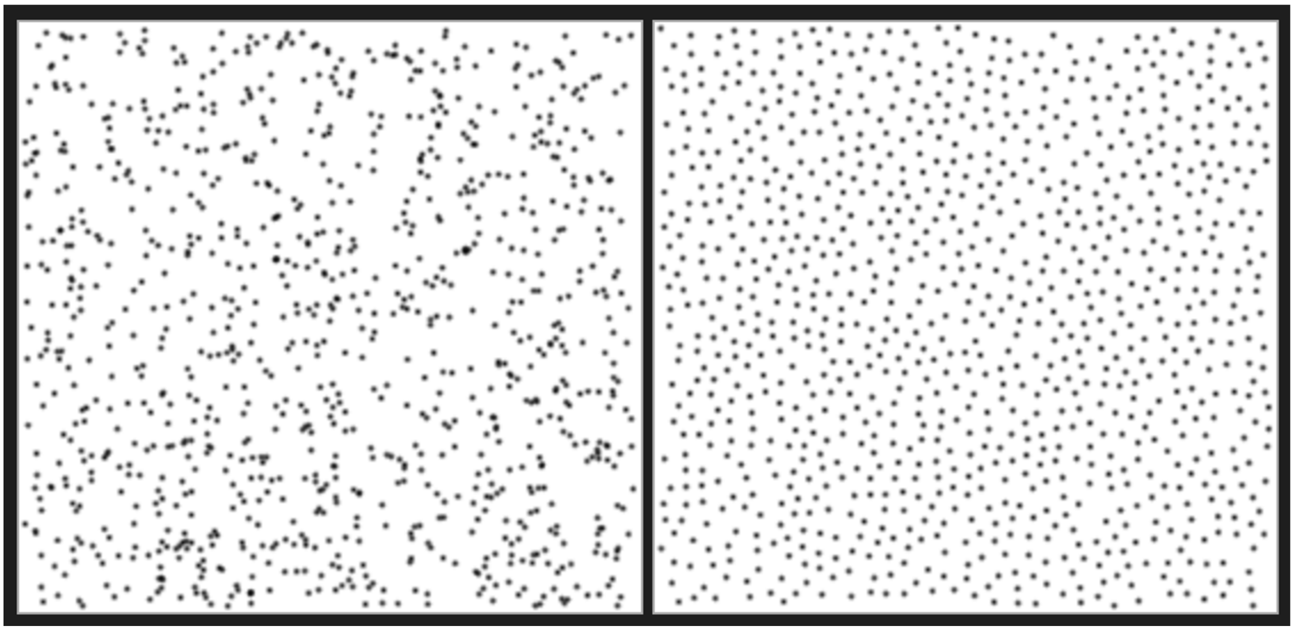
Chris Piech
CS 109

Review #3
November 21, 2025

Review 3

1. **Which one is Poisson?**



Justify your answer.

2. **Linear Regression p-value**

You have a set of 50 data points (x, y) where $x \in \mathbb{R}$ and $y \in \mathbb{R}$.

You fit a linear regression model which gives you a fit

$$\hat{y} = \theta_0 + \theta_1 \cdot x$$

You are surprised to see that $\theta_1 = 2$. You want to claim that this value of θ_1 is significantly high. What is the probability of seeing a value of $\theta_1 \geq 2$, given there is no relationship between x and y? Estimate your answer using sampling. Provide pseudo-code:

3. **Magical Puppy Drug**

A magical-cure-drug for puppies has gone through clinical trial. It worked for 5 puppies but did not work for 1. Prior to the study you had a uniform belief for the probability the drug would work.

If you now give the drug to 10 new puppies, what is the probability that it works exactly 8 out of the 10 times?

4. **What an Ideal Classifier Looks Like**

   Suppose we have data where Y is a **medical record** and X is a **summary** that was written from that record. We want to measure how much information the summary X contains about the medical record Y. Last class I gave you a full proof that if we assume we have access to an *ideal classifier* that identifies which medical record corresponds to a given summary, then the classifiers average log-probability of being correct forms a good approximation to mutual information. Here is the most critical part of the proof.

   For each summary X, we construct a list of N candidate medical records: $(Y_1, Y_2, \ldots, Y_N)$. Exactly one of these (call it Y_I) is the *true* medical record. The other $N-1$ candidates are randomly drawn distractors from the marginal $P(Y)$, independently of X. The index $I \in \{1, \ldots, N\}$ is the correct answer. We compute the true conditional probability $P(I = i \mid x, y_1, \ldots, y_N)$.

$$P(I = i \mid x, y_1, \ldots, y_N) = \frac{P(x, y_1, \ldots, y_N, i)}{P(x, y_1, \ldots, y_N)}.$$

   Because the distractors Y_j are drawn independently from $P(Y)$ and independently of X,

$$P(x, y_1, \ldots, y_N, i) = \frac{1}{N} P(x, y_i) \prod_{j \neq i} P(y_j).$$

   The denominator is a sum over all possible positions of the true record:

$$P(x, y_1, \ldots, y_N) = \sum_{k=1}^{N} \frac{1}{N} P(x, y_k) \prod_{j \neq k} P(y_j).$$

   Canceling the common factors $\frac{1}{N} \prod_j P(y_j)$,

$$\begin{aligned} P(I = i \mid x, y_1, \ldots, y_N) &= \frac{P(x, y_i)/P(y_i)}{\sum_{k=1}^{N} P(x, y_k)/P(y_k)} \\ &= \frac{P(y_i \mid x) P(y_i)}{\sum_{k=1}^{N} P(y_k \mid x)/P(y_k)} \end{aligned}$$

5. **(Optional) Exponential Backoff Total Expected Run Time**

   In Problem Set 2, question 11 we solved a problem for the expected run time given we knew the number of collisions. But in reality, we won't know in advance how many collisions will happen! Now that you have learned more in CS109 you are ready for the challenge problem: if the algorithm doesnt know in advance how many collisions will happenwhats the overall expected time until success?