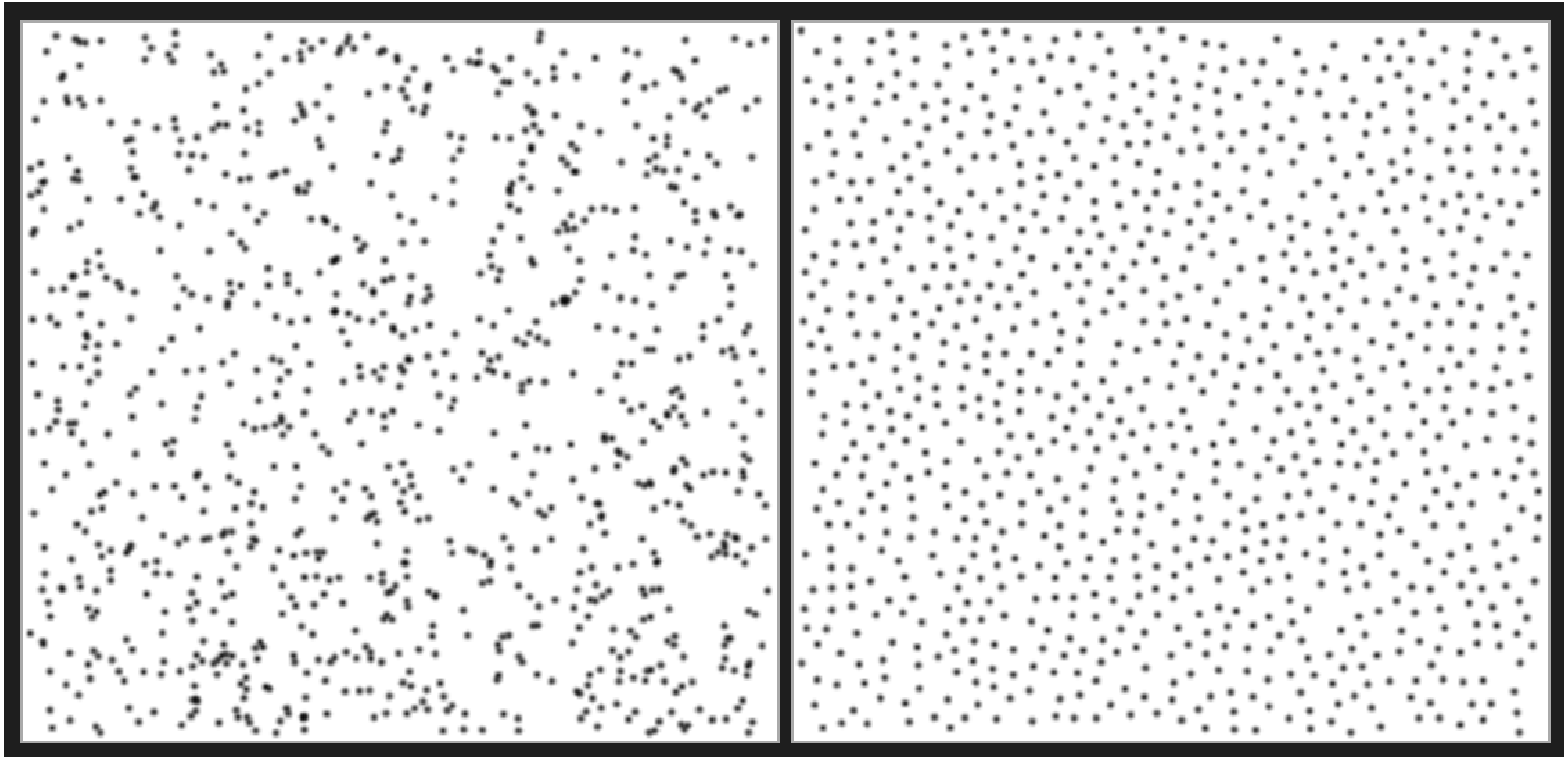
Reinforcement 3

Chris Piech
CS109, Stanford University

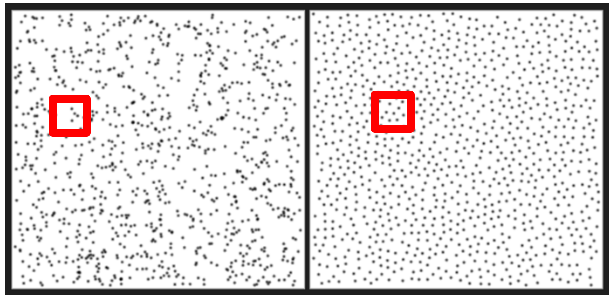
Which one is Poisson? Justify with Math…



Same number of points…

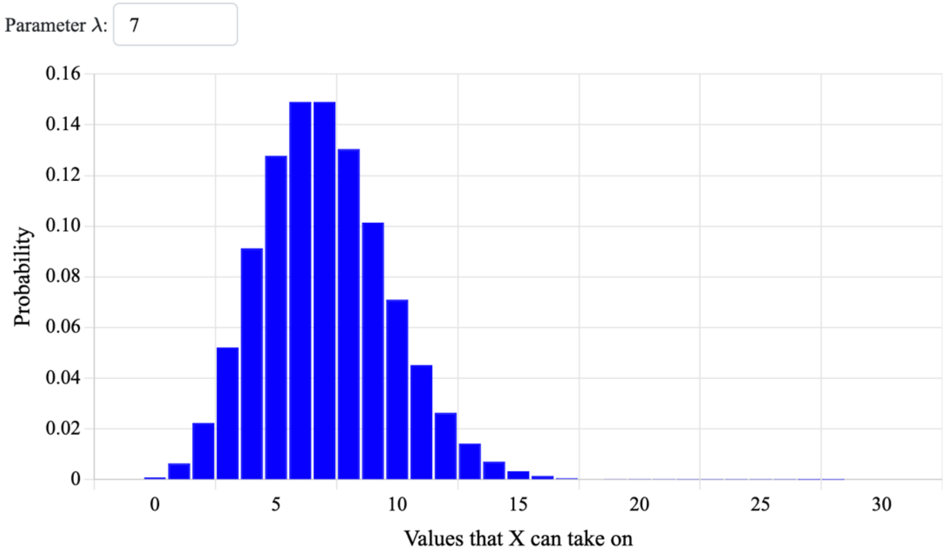
Which one is Poisson? Sampling Solution

Squares that are 1x1

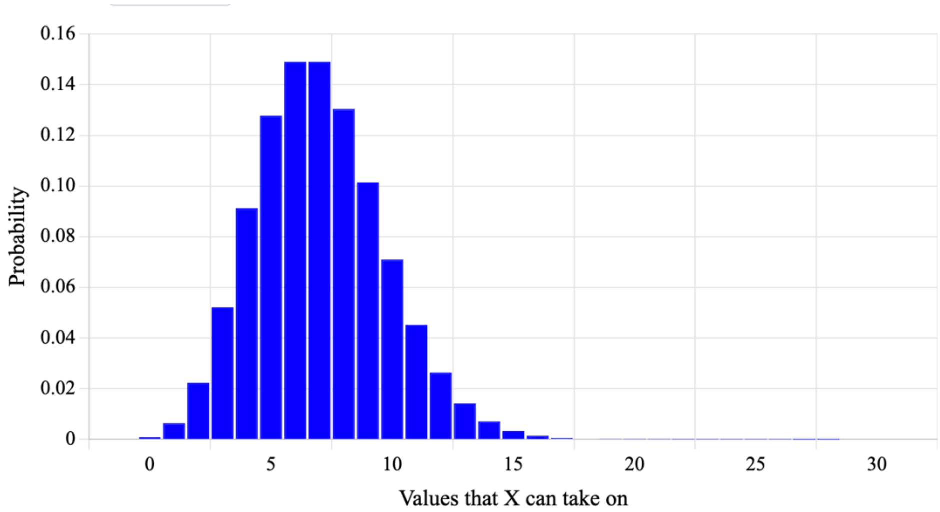


True rate = 7 points per unit size

Theory $P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$

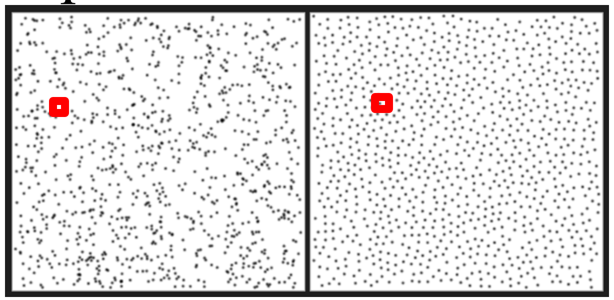


Observed (for both)



Which one is Poisson? Sampling Solution

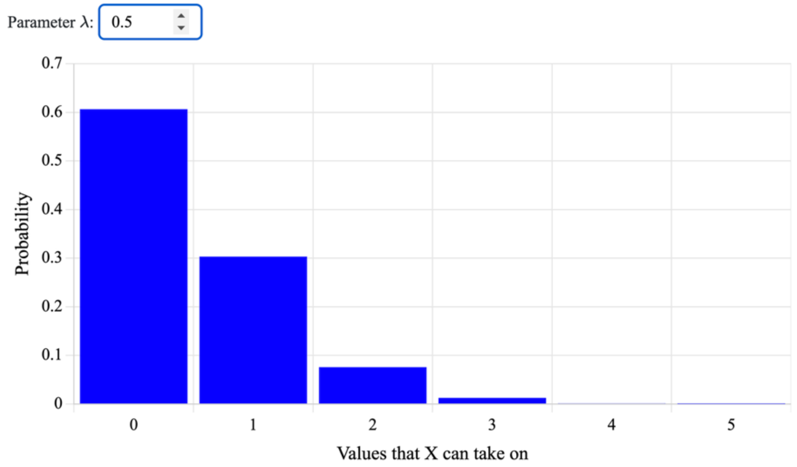
Squares that are .01 x .01

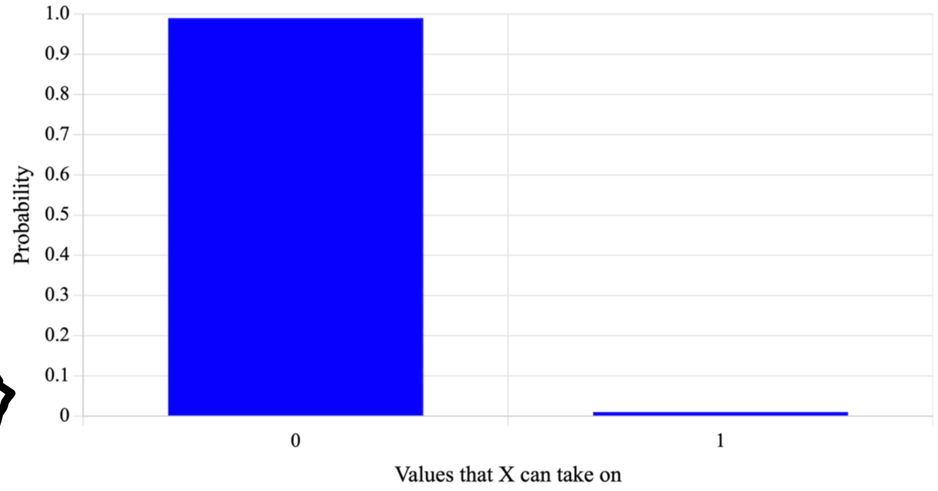


True rate = 7 points per unit size

Theory $P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$

Observed (for right)





Bonus, calculate the KL divergence between these two distributions

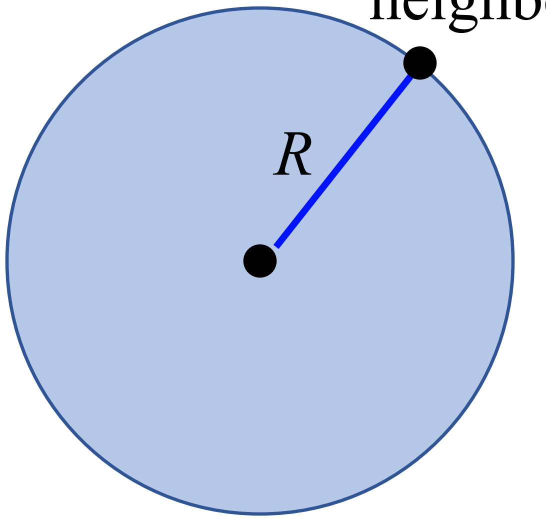
Which one is Poisson? Nearest Neighbor Distance

Let R be the distance to the nearest neighbor of a point

$$P(R > r) = 1 - P(\text{Zero points in radius } R)$$
$$= 1 - P(X = 0)$$

Let X be the number of points in radius R. $X \sim \text{Poi}(\lambda = 7 \cdot \pi \cdot r^2)$

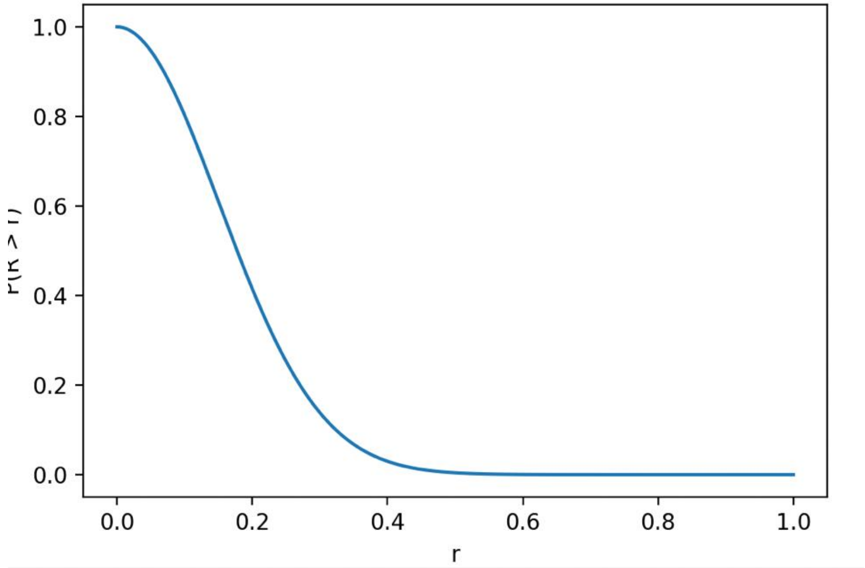
$$= 1 - \frac{(7\pi r^2)^0 e^{-7\pi r^2}}{0!}$$
$$= 1 - e^{7\pi r^2}$$



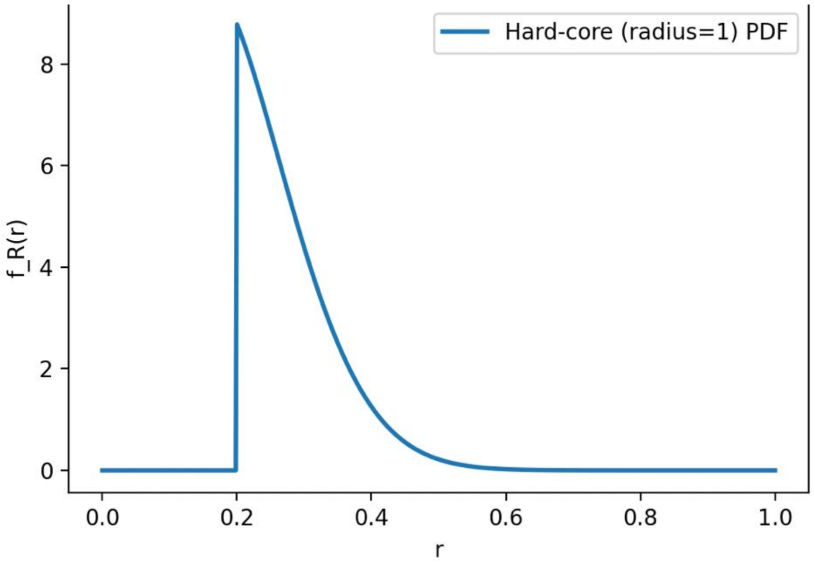
Which one is Poisson? Nearest Neighbor Distance

$$P(R > r) = 1 - e^{7\pi r^2}$$

Theory and picture on the left:



Picture on the right:



Linear Regression P-Value

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:

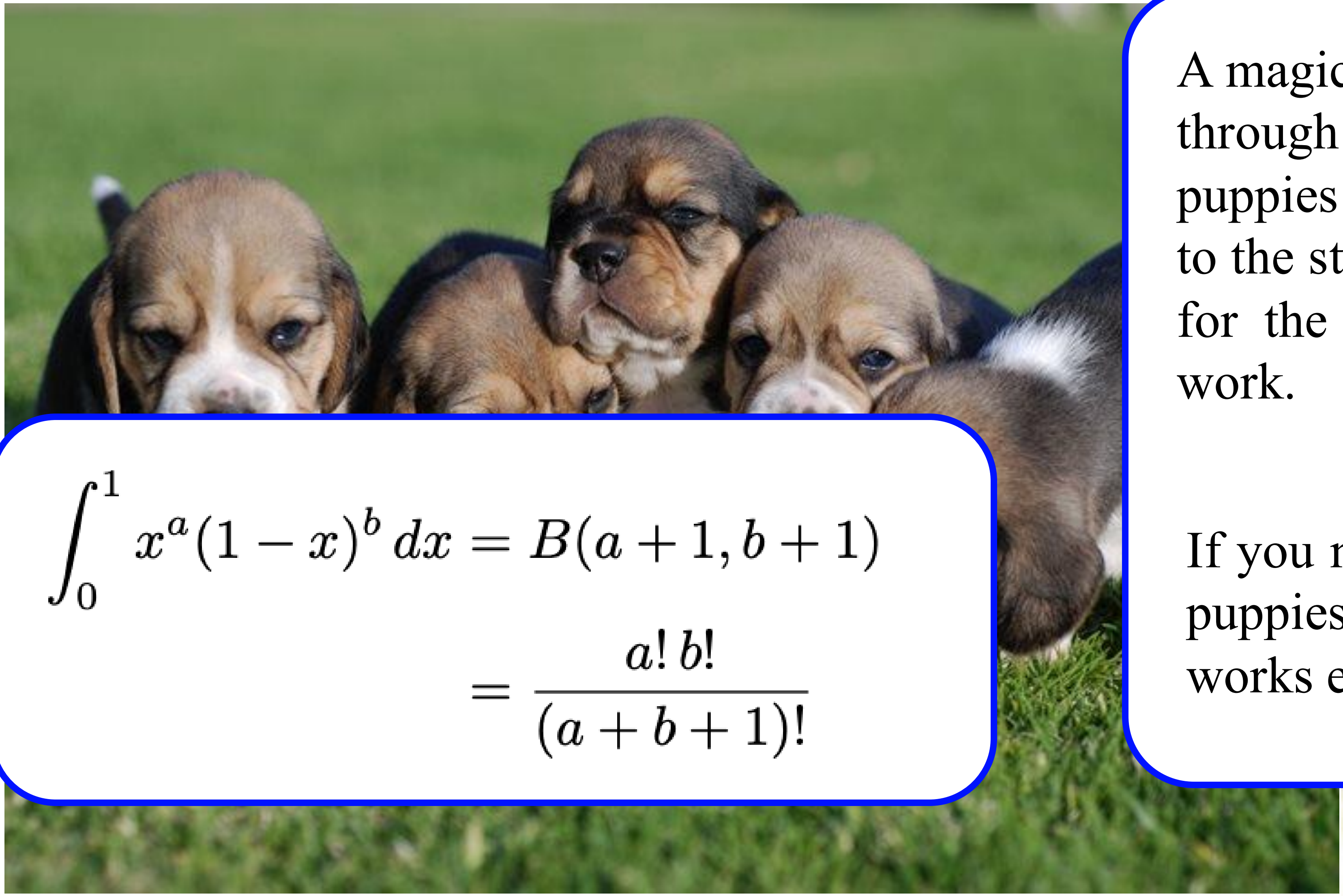
Linear Regression P-Value

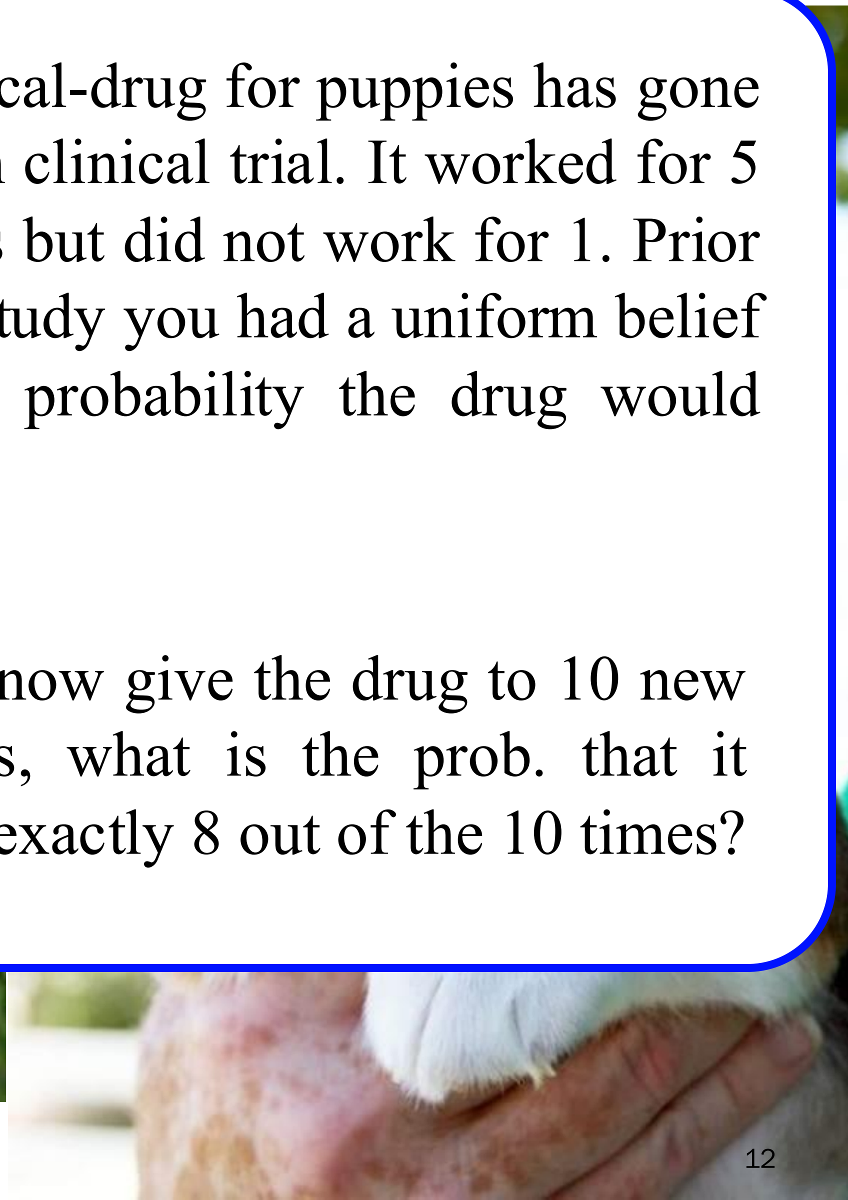
```python
def bootstrapping(x, y, num_bootstrap=10000, threshold=2.0):
    N = len(x)
    count_extreme = 0
    for _ in range(num_bootstrap):
        # resample independently
        x_boot = random.choices(x, k=N)
        y_boot = random.choices(y, k=N)

        # pseudo code
        model = LinearRegression().fit(x_boot, y_boot)
        slope = model.get_theta_1()

        if slope >= threshold:
            count_extreme += 1

    return count_extreme / num_bootstrap
```

Magical Puppy Drugs

Magical Puppy Drugs

Magical Puppy Drugs





A magical-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 prob. that it works exactly 8 out of the 10 times?

$$\int_0^1 x^a (1-x)^b \, dx = B(a+1, b+1)$$

$$= \frac{a!\, b!}{(a+b+1)!}$$

Info NCE

Many questions want mutual information scores

InfoNCE Proof

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

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 $\frac{1}{N} \prod_j P(y_j)$,

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