Probability Theory

CS 70 Discussion 10A

Raymond Tsao

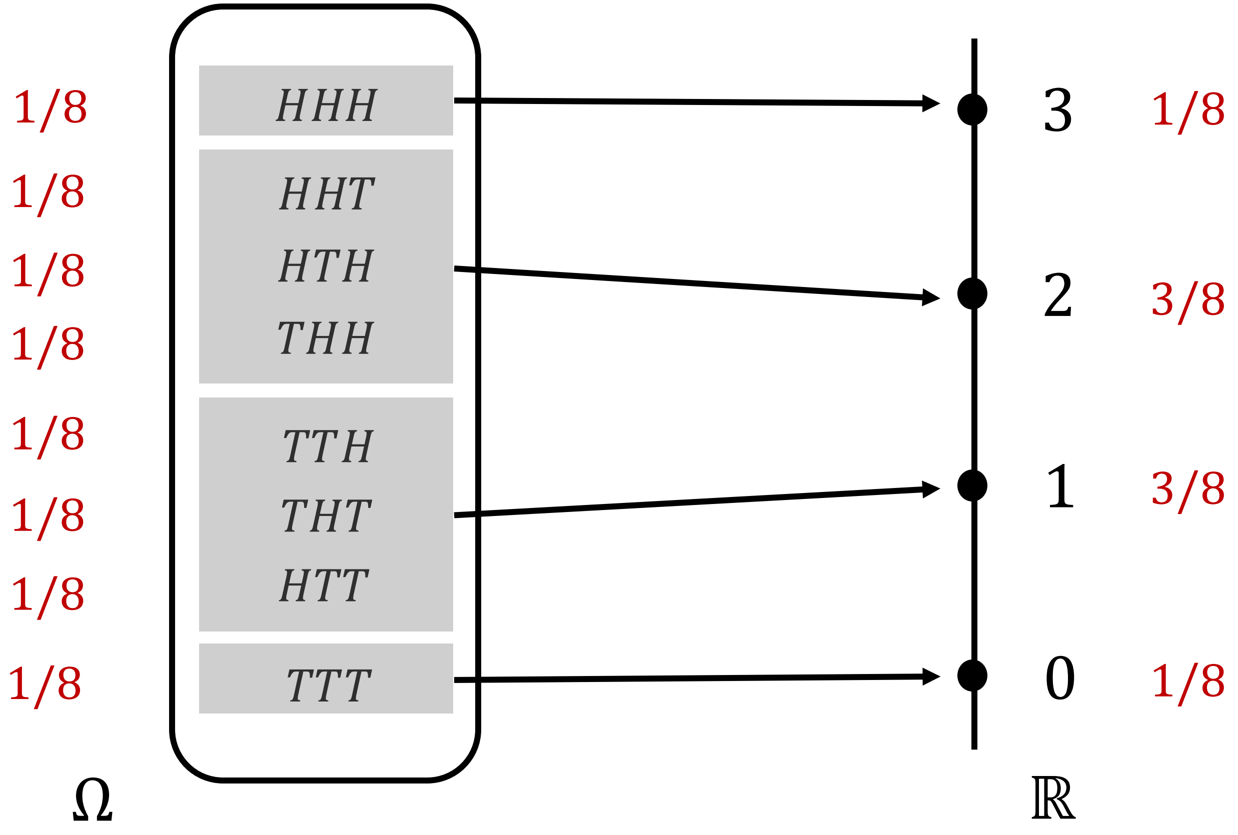
2025-04-04

Note: These slides are unofficial course materials. Please use the notes as the only single source of truth.

Random Variables

Recall: A probability space consists of a space space Ω and a probability function $\mathbb{P}$

A random variable: $X: \Omega \to \mathbb{R}$ induces a distribution over $\mathbb{R}$



Bernoulli v.s Binomial Distribution

Bernoulli distribution: simple model of one coin toss $X \in \{0, 1\}$

Parameters: $\mathbf{p}$ (success rate)

$$\mathbb{P}[X = 0] = 1 - p$$

$$\mathbb{P}[X = 1] = p$$

Binomial distribution: simple model of $\mathbf{n}$ coin toss $X \in \{0, 1\}$

Parameters: $\mathbf{p}$ (success rate), n (number of coin tosses)

$$\mathbb{P}[X = k] = \binom{n}{k} p^k (1 - p)^{n-k}$$

Problem 1: Head Count

(a) What is the probability that we get k heads (success)?

Any sequence with exactly k heads (so $20 - k$ tails) occurs with probability:

$$\left(\frac{2}{5}\right)^k \left(\frac{3}{5}\right)^{20-k}$$

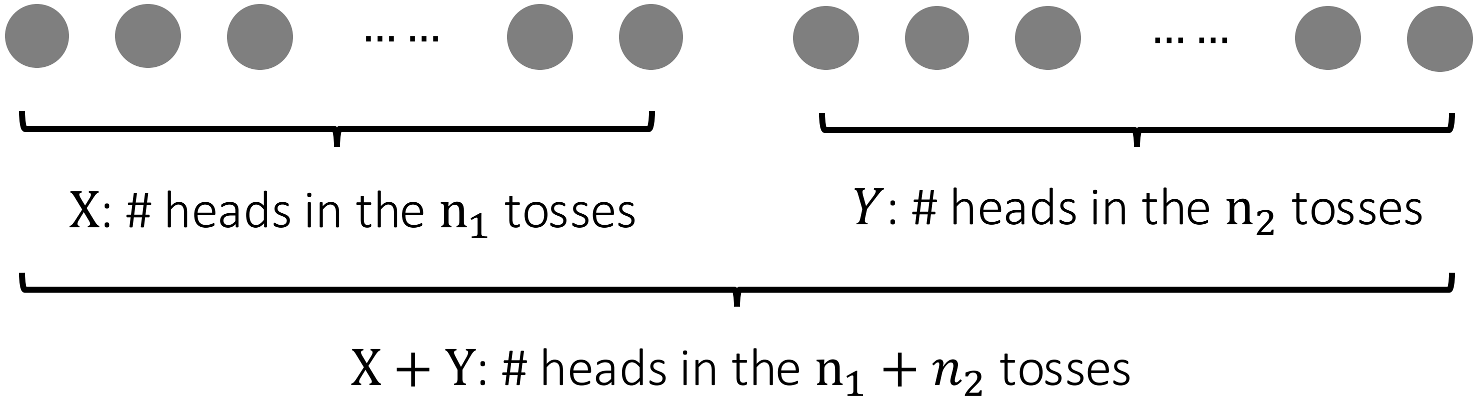
How many such sequences are there?

How many ways we can permute a sequence with k heads and $20 - k$ tails?

$$\mathbb{P}[X = k] = \binom{20}{k} \left(\frac{2}{5}\right)^k \left(\frac{3}{5}\right)^{20-k}$$

Problem 1: Head Count

(e) If $X \sim Bin(n_1, p), Y \sim Bin(n_2, p)$, what is the distribution of $X + Y$?



We expect $X + Y \sim Bin(n_1 + n_2, p)$! ← # heads in the $n_1 + n_2$ tosses

How to find the distribution of X?

- Way 1: Find a formula for $\mathbb{P}[X = k]$
- Way 2: Find a formula for $\mathbb{P}[X \leq k]$ $\quad \mathbb{P}[X = k] = \mathbb{P}[X \leq k] - \mathbb{P}[X \leq k - 1]$

Problem 1: Head Count

(e) Let $\mathrm{Z} = X + Y$

Independence

$$\mathbb{P}[Z = k] = \mathbb{P}[\mathrm{X} + \mathrm{Y} = \mathrm{k}]$$

$$= \sum_{i=0}^{k} \mathbb{P}[X = i, Y = k - i] = \sum_{i=0}^{k} \mathbb{P}[X = i]\mathbb{P}[Y = k - i]$$

$$= \sum_{i=0}^{k} \binom{n_1}{i} p^i (1-p)^{n-i} \cdot \binom{n_2}{k-i} p^{k-i} (1-p)^{n-(k-i)}$$

$$= \binom{n_1 + n_2}{k} p^k (1-p)^{n_1+n_2-k}$$

Problem 2: Geometric

Binomial distribution: Given n coin tosses, probability of k successes

Geometric distribution: Number of coin tosses until we get one success

Parameters: p (success rate)

$$\mathbb{P}[X = k] = (1-p)^{k-1}p$$

(a) If $\mathrm{p} = 3/4$

F F F F …. F F F F S

$\mathrm{k} - 1$ fails

$$\mathbb{P}[X = k] = \left(\frac{1}{4}\right)^{k-1} \cdot \frac{3}{4}$$

Problem 2: Geometric

(c) $\mathbb{P}[X > k]$

$$\mathbb{P}[X > k] = \sum_{i=k+1}^{\infty} \mathbb{P}[X = i]$$

$$= \sum_{i=k+1}^{\infty} \left(\frac{1}{4}\right)^{i-1} \cdot \frac{3}{4}$$

$$= \left(\frac{1}{4}\right)^{\mathrm{k}}$$

Need more than k toss to get one successes $\Longleftrightarrow$ First k tosses must be fails

Problem 2: Geometric

(e) Memoryless property: $\mathbb{P}[X > k | X > m]$

$$\mathbb{P}[X > k | X > m] = \frac{\mathbb{P}[X > k, X > m]}{\mathbb{P}[X > m]}$$

$$= \frac{\mathbb{P}[X > k]}{\mathbb{P}[X > m]}$$

$$= \left(\frac{1}{4}\right)^{\mathrm{k} - m}$$

Problem 2: Geometric

(f) Find the distribution of $Z = \min\{X, Y\}$

$$\mathbb{P}[Z \geq k] = \mathbb{P}[\min\{X, Y\} \geq k]$$

$$= \mathbb{P}[X \geq k, Y \geq k] \quad = \mathbb{P}[X \geq k]\mathbb{P}[Y \geq k]$$

$$= (1-p)^{k-1}(1-q)^{k-1} = \big(1-(p+q-pq)\big)^{k-1}$$

$$\mathbb{P}[Z = k] = \mathbb{P}[Z \geq k] - \mathbb{P}[Z \geq k+1]$$

$$= \big(1-(p+q-pq)\big)^{k-1}(p+q-pq)$$

Extension: Negative Binomial Distribution

Geometric distribution: $X =$ Number of trials before observing first success

Negative binomial distribution: $X =$ Number of trials before observing r success

F F S F …. F S F F S

$\mathrm{k}-1$ trials contain:

- $\mathrm{r}-1$ success
- $k-r$ fails

rth success

$$\mathbb{P}[X=n]=\binom{n-1}{r-1}(1-p)^{n-r}p^r$$

Conclusion

Problem type 1:

Fix # of trials n, count # of successes k

Bernoulli distribution

$n = 1$ ↓

Binomial distribution

$$\mathbb{P}[X = k] = \binom{n}{k} p^k (1-p)^{n-k}$$

Poisson distribution: approximates

Problem type 2:

Fix # of success r, count # of trials n

Geometric distribution

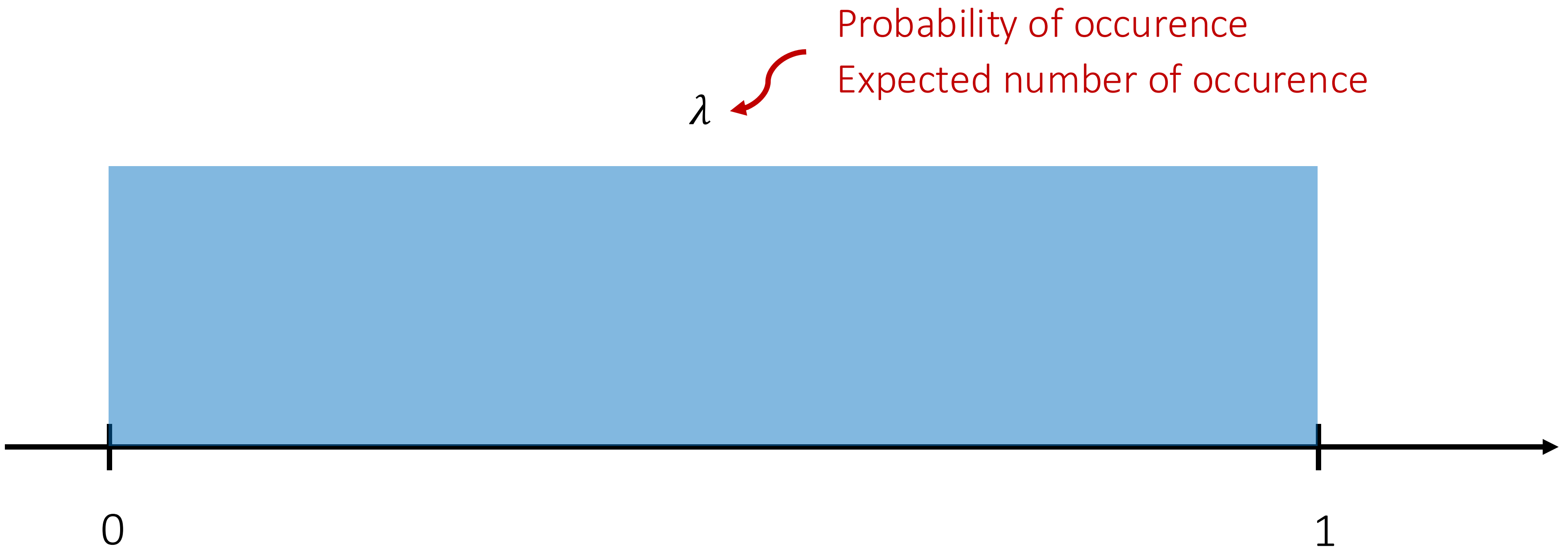
$r = 1$ ↓

Negative Binomial distribution

$$\mathbb{P}[X = n] = \binom{n-1}{r-1} p^r (1-p)^{n-r}$$

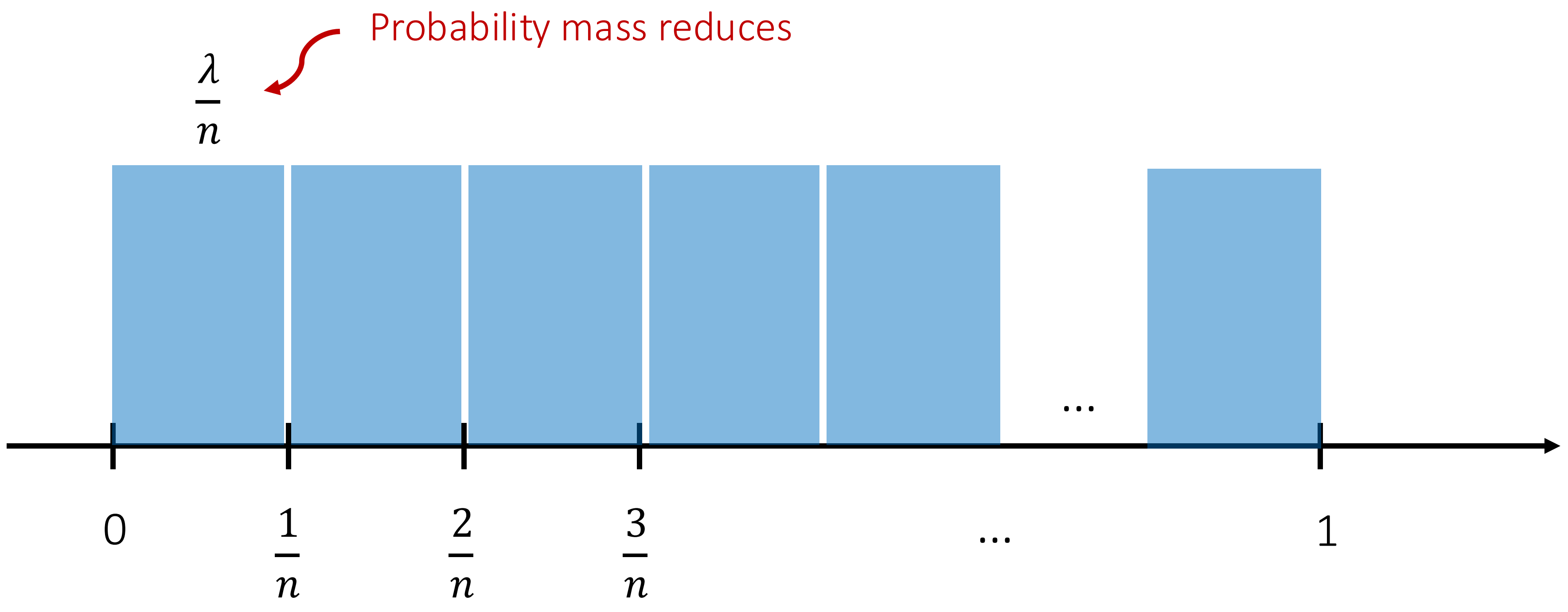
Poisson distribution as approximation of Binomial

Suppose we want to count the number of occurrence (X) over $[0, 1]$



Poisson distribution as approximation of Binomial

Discretize into n bins, then $X \sim Bin\left(n, \frac{\lambda}{n}\right)$



Poisson distribution as approximation of Binomial

Discretize into n bins, then $X \sim Bin\left(n, \frac{\lambda}{n}\right)$

Limit as $n \to \infty$: $X \to Poisson(\lambda)$

$$\mathbb{P}[X = k] = \frac{e^{-\lambda}\lambda^k}{k!}$$

Law of rare events: If n large, p small, and $np = \lambda$, then $X \sim Bin(n, p)$ can be approximated with $Poisson(\lambda)$

Need to be careful with the dimensions!

λ Unit: Number of occurence

Problem 3: Shuttles and Taxis At Airport

(a) Rate of occurrence for shuttle $\lambda_1 = 1/20$ (1 per 20 minutes)

Rate of occurrence for taxis $\lambda_2 = 1/10$ (1 per 10 minutes)

Number of occurrence of shuttle in $[0, 20]$ Need to do unit conversion!

$$X \sim Poisson(1) \qquad \frac{1\ (shuttle)}{20\ (minutes)} \cdot 20\ (minutes) = 1(shuttle)$$

Number of occurrence of taxis in $[0, 20]$

$$Y \sim Poisson(2) \qquad \frac{1\ (taxis)}{10\ (minutes)} \cdot 20\ (minutes) = 2(taxis)$$

Number of occurrence of vehicles in $[0, 20]$

$$X + Y \sim Poisson(3) \qquad 1(shuttle) + 2(taxis) = 3(vehicles)$$