Frequentist Data Analysis

Prof. Pradeep Ravikumar
pradeepr@cs.cmu.edu

Recall: Model-based ML

- Learning: From data to model
  - A model explains how the data was "generated"
  - **E.g. given (symptoms, diseases) data, a model explains which symptoms arise from which diseases**

- Inference: From model to knowledge
  - Given the model, we can then answer questions relevant to us
  - **E.g. given (symptom, disease) model, given some symptoms, what is the disease?**

Model Learning: Data to Model

- What are some general principles in going from data to model?

- What are the guarantees of these methods?

LET US CONSIDER THE EXAMPLE OF A SIMPLE MODEL

Your first consulting job

- A billionaire from the suburbs of Seattle asks you a question:
  - He says: I have a coin, if I flip it, what’s the probability it will fall with the head up?
  - You say: Please flip it a few times:

Your first consulting job

- A billionaire from the suburbs of Seattle asks you a question:
  – He says: I have a coin, if I flip it, what's the probability it will fall with the head up?
  – You say: Please flip it a few times:

  – You say: The probability is **3/5**
  – **He says: Why???**
  – You say: Because… frequency of heads in all flips

Questions

- Why frequency of heads?

- How good is this estimation?
  - Would you be willing to bet money on your guess of the probability?
  - Why not?

Model-based Approach

- First we need a model that would explain the experimental data
- What is the experimental data?
- Coin Flips

Model

- First we need a model that would capture the experimental data
- What is the experimental data?
- Coin Flips

Model

- A model for coin flips
  - Bernoulli Distribution
- X is a random variable with Bernoulli distribution when:
  - X takes values in {0,1}
  - $P(X = 1) = p$
  - $P(X = 0) = 1 - p$
  - Where p in [0,1]

Model

- X is a random variable with Bernoulli distribution when:
  - X takes values in {0,1}
  - $P(X = 1) = p$, $P(X = 0) = 1 - p$, where p in [0,1]
  - X = 1 i.e. heads with probability p, and X = 0 i.e. tails with probability 1 – p
- And we draw **independent** samples that are **identically distributed** from same distribution
  - flip the same coin multiple times

Bernoulli distribution

Data, D = 

- P(Heads) = θ, P(Tails) = 1-θ

- Flips are **i.i.d.**:
  - **Independent** events
  - **Identically distributed** according to Bernoulli distribution

Choose θ that maximizes the probability of observed data

Review: Random Variables, Likelihood, Expectation

- Let us first briefly review some background concepts that are key to answering the question in previous slide
  - Random Variables
  - Likelihood
  - Expectation

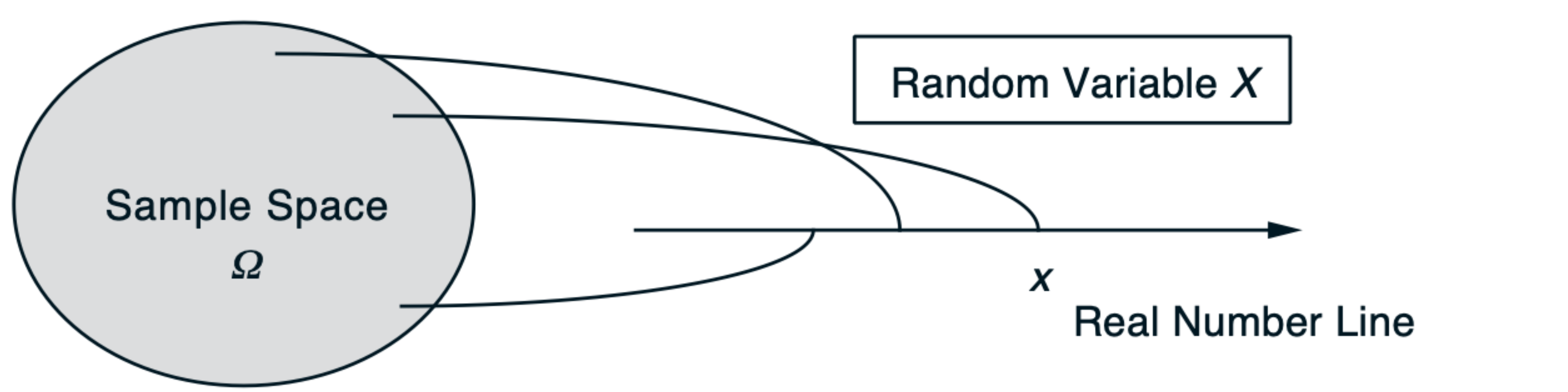
Random Variables

- Given an experiment, the outcomes could be numerical
    - roll of a four-side die
    - instrument readings
- Or they could be non-numerical
    - selection of people for a committee
- What if for any {numerical or non-numerical} outcomes, we assign numerical values to them anyway
    - Random Variables

Intro to Random Variables

- An assignment of a value (number) to every possible outcome

- Mathematically: A function from the sample space Ω to the real numbers
  - discrete or continuous values



Random Variables

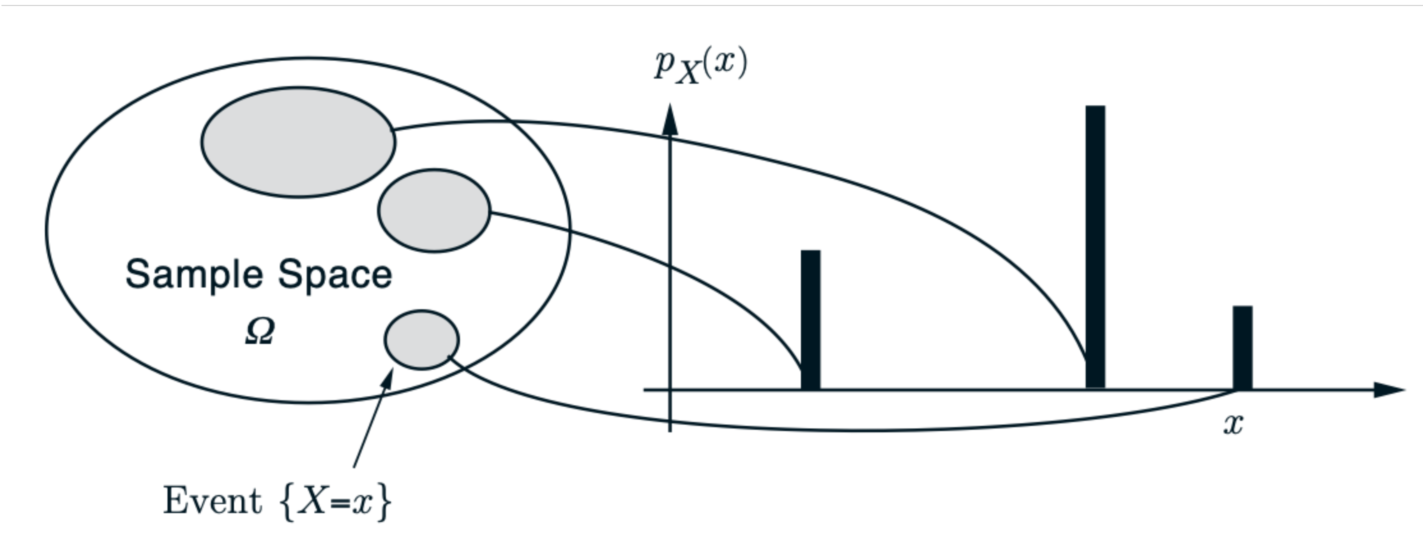
- Notation
  - Random variable X
  - Numerical Value x
- Examples
  - 1 Coin Toss. Define
    - $X(\text{Heads}) = 1$; $X(\text{Tails}) = 0$

Probability mass function (PMF)

- Notation:

$$\begin{aligned} p_X(x) &= \mathbf{P}(X = x) \\ &= \mathbf{P}(\{\omega \in \Omega \text{ s.t. } X(\omega) = x\}) \end{aligned}$$

How to compute a PMF

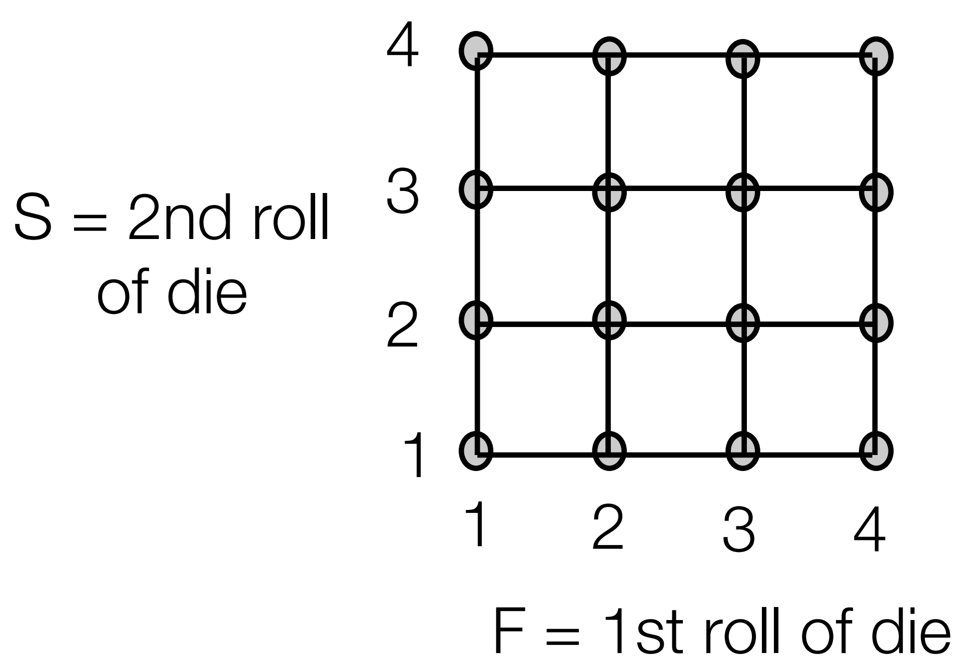


- collect all possible outcomes for which X is equal to x
- add their probabilities
- repeat for all x

Example I

- Two independent throws of a fair tetrahedral die

  F: outcome of first throw
  S: outcome of second throw
  X = min(F , S)

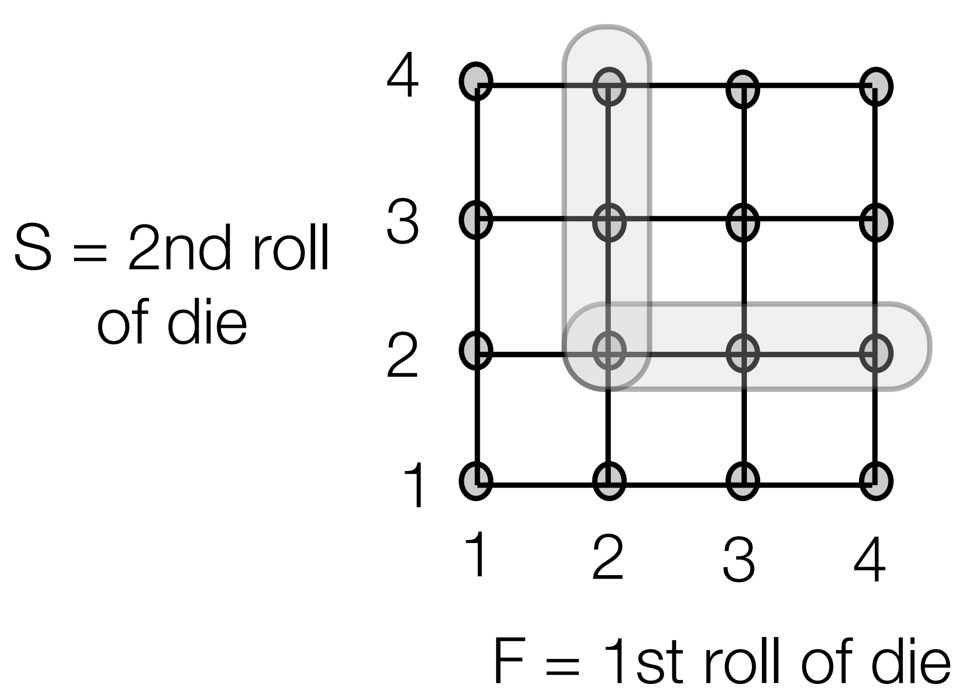


$p_X(2) =$

Example I

- Two independent throws of a fair tetrahedral die

  F: outcome of first throw
  S: outcome of second throw
  X = min(F , S)

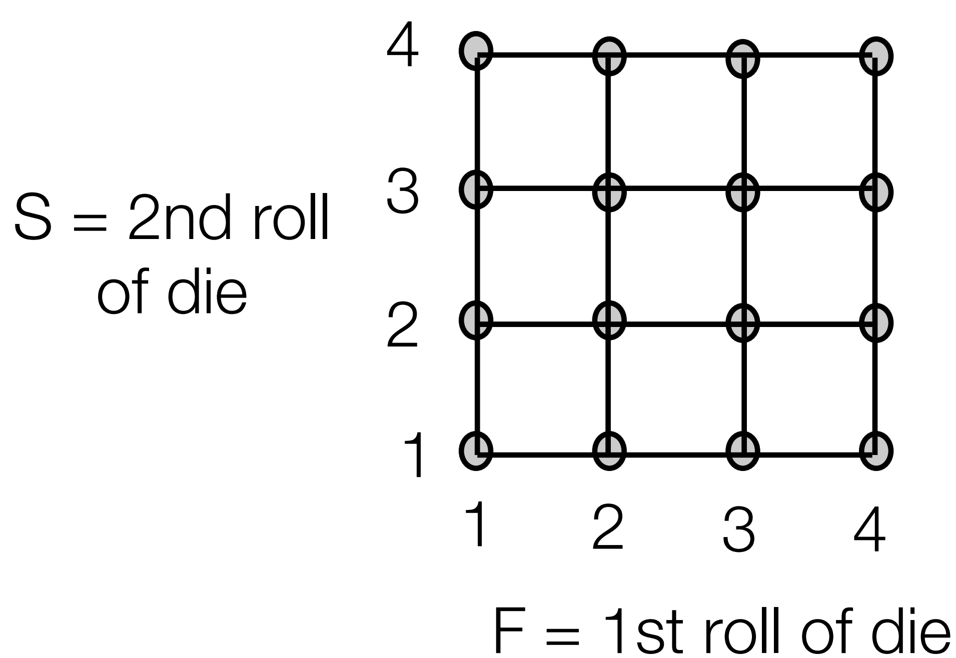


$$p_X(2) = 5/16$$

Example II

- Two independent throws of a fair tetrahedral die

  F: outcome of first throw
  S: outcome of second throw
  X = max(F , S)

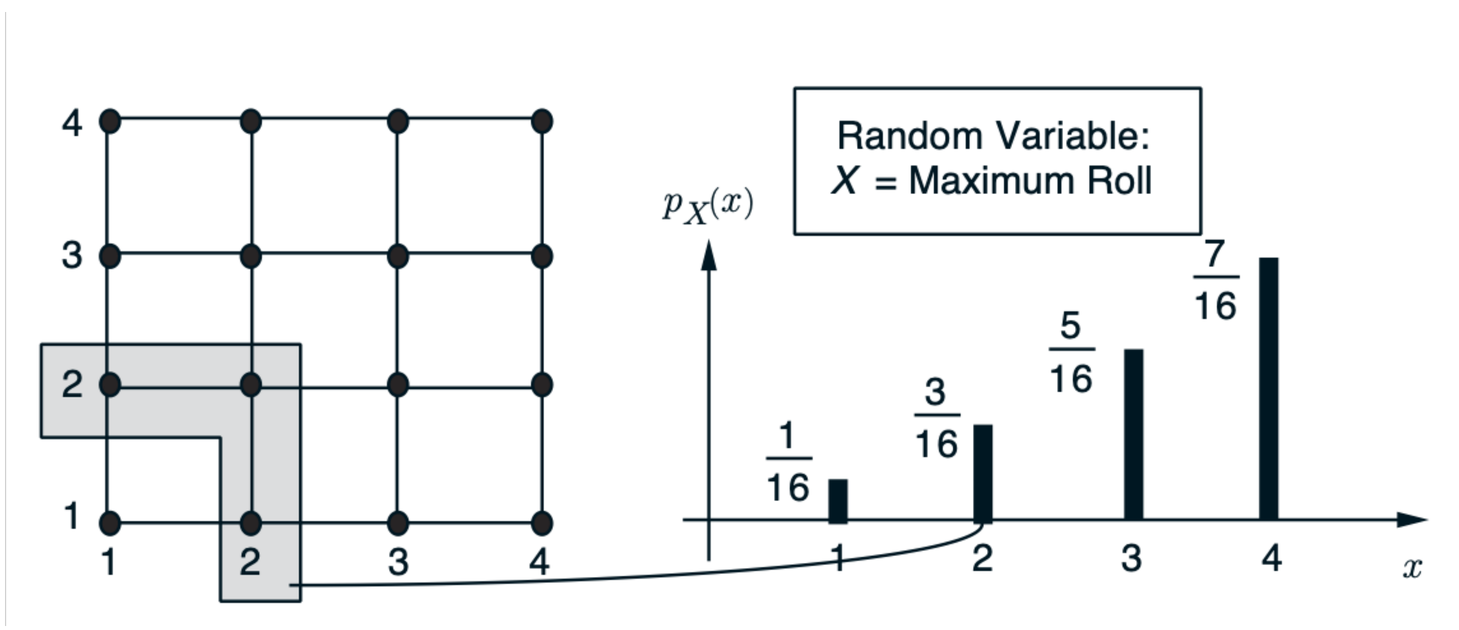


$p_X(2) =$

Example II

- Two independent throws of a fair tetrahedral die

  F: outcome of first throw
  S: outcome of second throw
  X = max(F , S)



Example III

- **Example:** X=number of coin tosses until first head
  - assume independent tosses, $\mathbf{P}(H) = p > 0$

    $$p_X(k) \ = \ \mathbf{P}(X = k)$$

Example III

- **Example:** X=number of coin tosses until first head
  - assume independent tosses, $\mathbf{P}(H) = p > 0$

$$\begin{aligned} p_X(k) &= \mathbf{P}(X = k) \\ &= \mathbf{P}(TT \cdots TH) \end{aligned}$$

Geometric PMF

- **Example:** X=number of coin tosses until first head
  - assume independent tosses, $\mathbf{P}(H) = p > 0$

$$
\begin{aligned}
p_X(k) &= \mathbf{P}(X = k) \\
&= \mathbf{P}(TT \cdots TH) \\
&= (1-p)^{k-1}p, \qquad k = 1, 2, \ldots
\end{aligned}
$$

  - **geometric** PMF

Example IV

- X: number of heads in n independent coin tosses
- $\mathbf{P}(H) = p$
- Let $n = 4$

  $p_X(2) \ =$

Example IV

- X: number of heads in n independent coin tosses
- $\mathbf{P}(H) = p$
- Let $n = 4$

$$
\begin{aligned}
p_X(2) \;=\; & \mathbf{P}(HHTT) + \mathbf{P}(HTHT) + \mathbf{P}(HTTH) \\
& + \mathbf{P}(THHT) + \mathbf{P}(THTH) + \mathbf{P}(TTHH)
\end{aligned}
$$

Binomial PMF

- X: number of heads in n independent coin tosses
- $\mathbf{P}(H) = p$
- Let $n = 4$

$$\begin{aligned} p_X(2) &= \mathbf{P}(HHTT) + \mathbf{P}(HTHT) + \mathbf{P}(HTTH) \\ &\quad + \mathbf{P}(THHT) + \mathbf{P}(THTH) + \mathbf{P}(TTHH) \\ &= 6p^2(1-p)^2 \\ &= \binom{4}{2} p^2 (1-p)^2 \end{aligned}$$

In general:

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, 1, \ldots, n$$

Expectation

- Definition:

$$\mathbf{E}[X] = \sum_x x p_X(x)$$

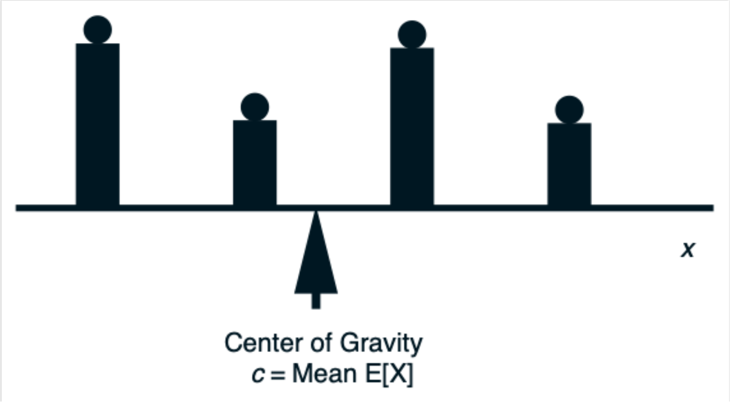
- Interpretations:
  - Center of gravity of PMF

Expectation

- Definition:

$$\mathbf{E}[X] = \sum_x x p_X(x)$$

- Interpretations:
  - Center of gravity of PMF



Given a bar with a weight $p_X(x)$ placed at each point x with $p_X(x) > 0$, the center of gravity c is the point at which the sum of the torques from the weights to its left are equal to the sum of the torques from the weights to its right.

$$\sum_x (x - c) p_X(x) = 0, \qquad \text{or} \quad c = \sum_x x p_X(x),$$

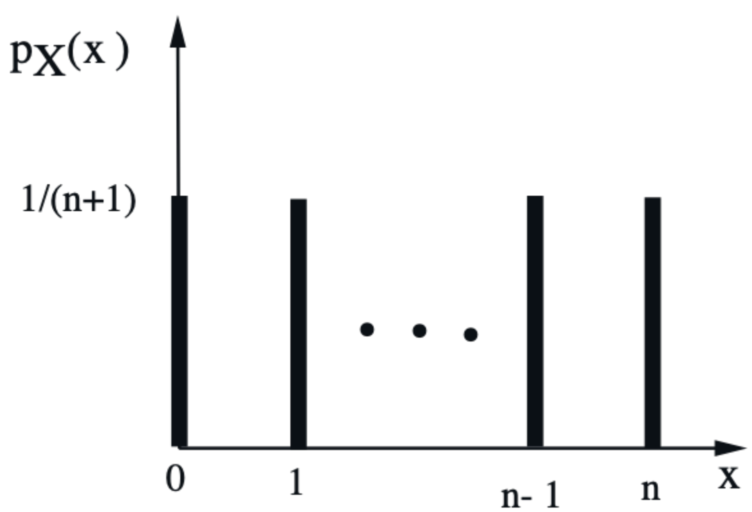
Expectation

- Definition:

$$\mathbf{E}[X] = \sum_x x p_X(x)$$

- Interpretations:
  - Center of gravity of PMF
  - Average in large number of repetitions of the experiment (to be substantiated later in this course)

- Example: Uniform on $0, 1, \ldots, n$



$$\mathbf{E}[X] = 0\times\frac{1}{n+1} + 1\times\frac{1}{n+1} + \cdots + n\times\frac{1}{n+1} =$$

Example

- You toss a fair coin multiple times, and you count the number of coin tosses till the first head appears. If this number is n, you receive 2^n dollars.
    - What is the expected amount that you will receive that you receive?
    - (How much would you be willing to play this game?)

Example

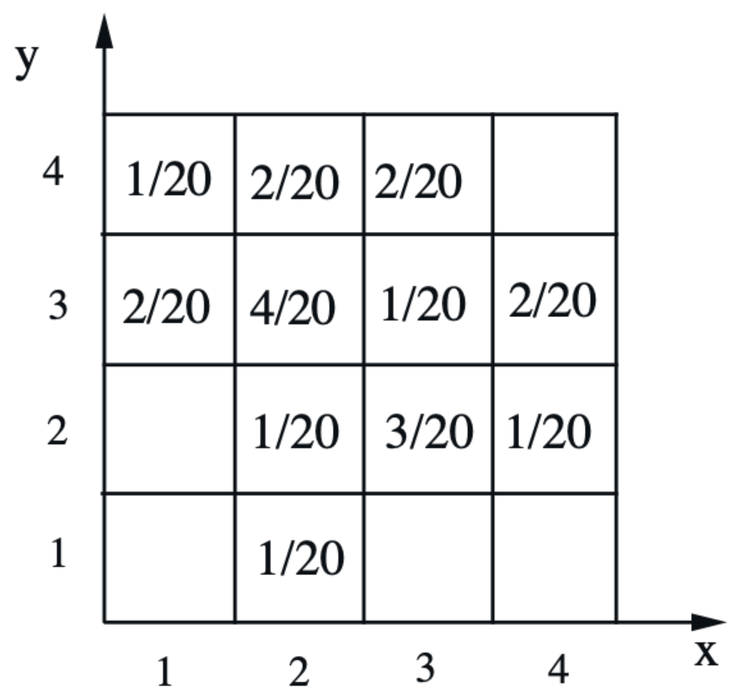
- You toss independently a fair coin, and you count the number of coin tosses till the first head appears. If this number is n, you receive 2^n dollars.
    - What is the expected amount that you will receive that you receive?
    - (How much would you be willing to play this game?)

$$\mathbf{E}[X] = \sum_{k=1}^{\infty} 2^k \cdot 2^{-k} = \sum_{k=1}^{\infty} 1 = \infty.$$

Would you be willing to pay any amount to play this game?

Joint PMFs

- $p_{X,Y}(x, y) = \mathbf{P}(X = x \text{ and } Y = y)$



- $\sum_x \sum_y p_{X,Y}(x, y) = 1$

Joint PMFs :: Marginals

- $p_{X,Y}(x,y) = \mathbf{P}(X = x \text{ and } Y = y)$

y \ x	1	2	3	4
4	1/20	2/20	2/20	
3	2/20	4/20	1/20	2/20
2		1/20	3/20	1/20
1		1/20		

- $p_X(x) = \sum_y p_{X,Y}(x,y)$

Joint PMFs :: Conditional PMFs

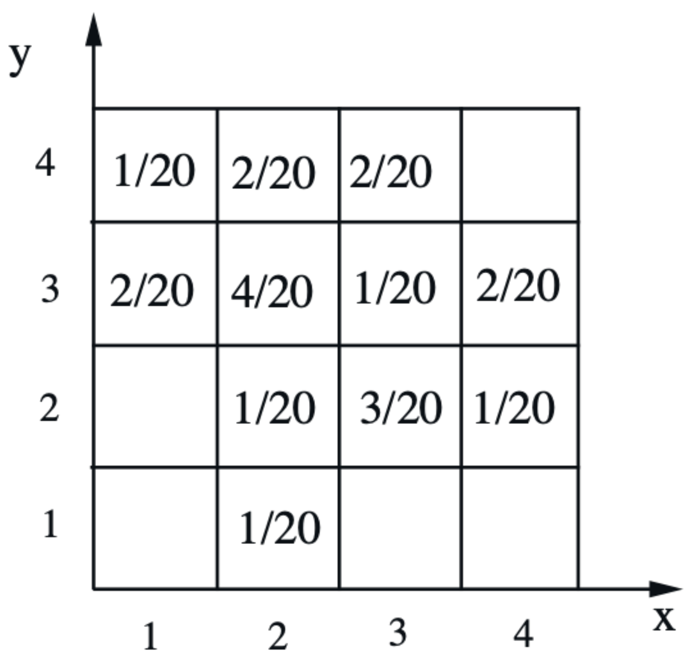
- $p_{X,Y}(x,y) = \mathbf{P}(X = x \text{ and } Y = y)$

y \ x	1	2	3	4
4	1/20	2/20	2/20	
3	2/20	4/20	1/20	2/20
2		1/20	3/20	1/20
1		1/20		

- $p_{X|Y}(x \mid y) = \mathbf{P}(X = x \mid Y = y) = \dfrac{p_{X,Y}(x,y)}{p_Y(y)}$

Joint PMFs :: Conditional PMFs

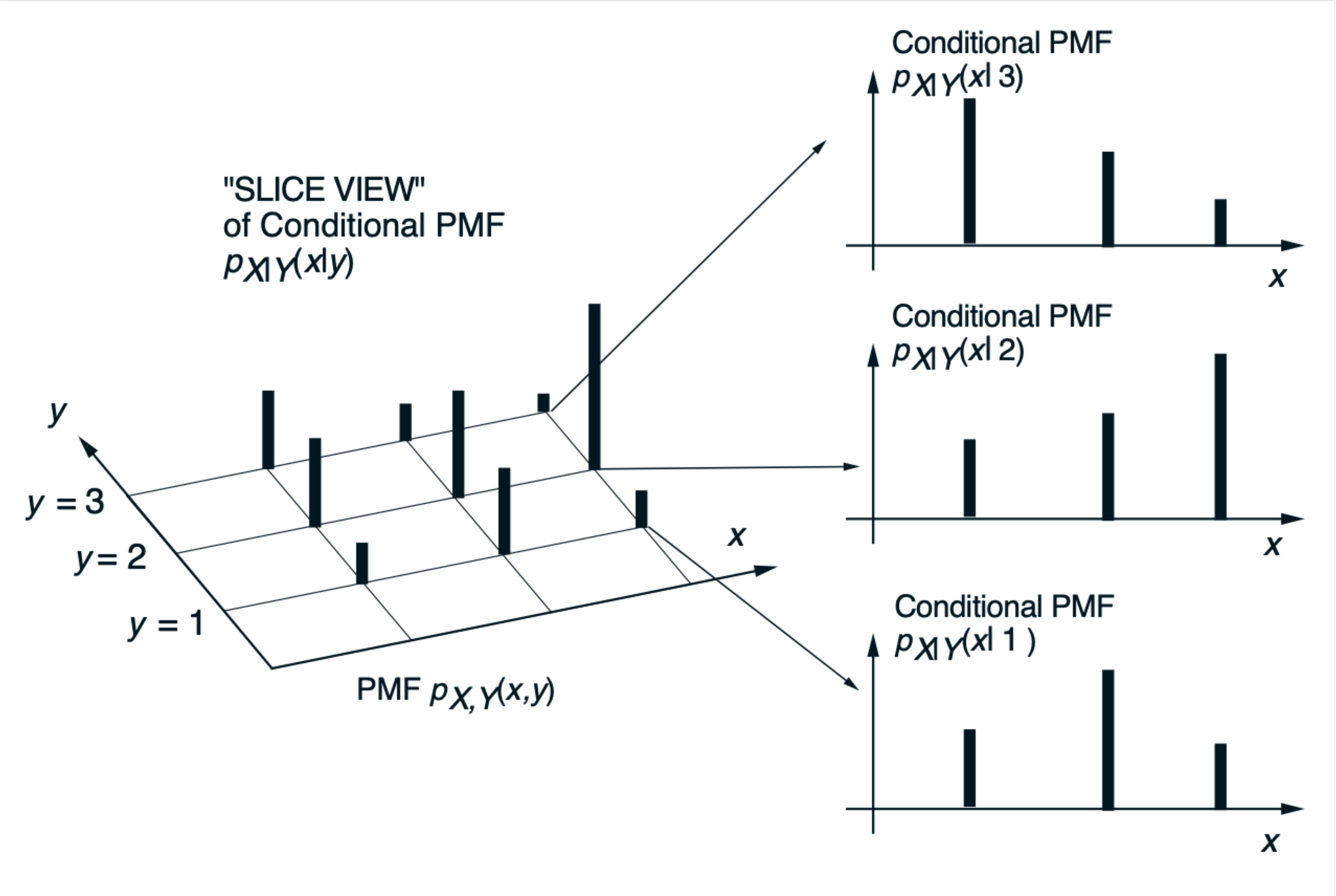
- $p_{X,Y}(x,y) = \mathbf{P}(X = x \text{ and } Y = y)$



- $p_{X|Y}(x \mid y) = \mathbf{P}(X = x \mid Y = y) = \dfrac{p_{X,Y}(x,y)}{p_Y(y)}$

- $\sum_x p_{X|Y}(x \mid y) =$

Conditional PMFs



- $p_{X|Y}(x \mid y) = \mathbf{P}(X = x \mid Y = y) = \dfrac{p_{X,Y}(x,y)}{p_Y(y)}$

Recap Example III

- Two coins are simultaneously tossed until one of them comes up a head and the other a tail. The first coin comes up with head with probability p, and the second with probability q.

    - Find the PMF of the number of tosses.

Example

- Two coins are simultaneously tossed until one of them comes up a head and the other a tail. The first coin comes up with head with probability p, and the second with probability q.
  - Find the PMF of the number of tosses.
    - X = number of coin tosses, is geometric with prob. $P(\{HT, TH\}) = p(1-q) + q(1-p)$

Example

- Two coins are simultaneously tossed until one of them comes up a head and the other a tail. The first coin comes up with head with probability p, and the second with probability q.
    - Find the PMF of the number of tosses.
    - What is the probability that the last toss of the first coin is a head?

Example

- Two coins are simultaneously tossed until one of them comes up a head and the other a tail. The first coin comes up with head with probability p, and the second with probability q.
  - Find the PMF of the number of tosses.
  - What is the probability that the last toss of the first coin is a head?

$$\mathbf{P}\big(HT \mid \{HT, TH\}\big) = \frac{p(1-q)}{p(1-q) + (1-q)p}.$$

Joint, Marginal, Conditional

$$p_X(x) = \mathbf{P}(X = x)$$

$$p_{X,Y}(x, y) = \mathbf{P}(X = x, Y = y)$$

$$p_{X|Y}(x \mid y) = \mathbf{P}(X = x \mid Y = y)$$

$$p_X(x) = \sum_y p_{X,Y}(x, y)$$

$$p_{X,Y}(x, y) = p_X(x) p_{Y|X}(y \mid x)$$

Multiple Random Variables

$$p_{X,Y,Z}(x, y, z) = p_X(x)p_{Y|X}(y \mid x)p_{Z|X,Y}(z \mid x, y) \qquad \ldots \text{ Multiplication Rule}$$

Independent Random Variables

$$p_{X,Y,Z}(x,y,z) = p_X(x)p_{Y|X}(y \mid x)p_{Z|X,Y}(z \mid x,y) \qquad \ldots \text{ Multiplication Rule}$$

- Random variables X, Y, Z are independent if:

  $$p_{X,Y,Z}(x,y,z) = p_X(x) \cdot p_Y(y) \cdot p_Z(z)$$

  for all x, y, z

Independent Random Variables

$$p_{X,Y,Z}(x,y,z) = p_X(x)p_{Y|X}(y \mid x)p_{Z|X,Y}(z \mid x,y) \qquad \ldots \text{ Multiplication Rule}$$

- Random variables X, Y, Z are independent if:

  $$p_{X,Y,Z}(x,y,z) = p_X(x) \cdot p_Y(y) \cdot p_Z(z)$$

  for all x, y, z

EXAMPLE:

y \ x	1	2	3	4
4	1/20	2/20	2/20	
3	2/20	4/20	1/20	2/20
2		1/20	3/20	1/20
1		1/20		

- Independent?

Expectation

$$\mathbf{E}[X] = \sum_x x p_X(x)$$

$$\mathbf{E}[g(X,Y)] = \sum_x \sum_y g(x,y) p_{X,Y}(x,y)$$

- In general: $\mathbf{E}[g(X,Y)] \neq g\Big(\mathbf{E}[X], \mathbf{E}[Y]\Big)$

- $\mathbf{E}[\alpha X + \beta] = \alpha \mathbf{E}[X] + \beta$

- $\mathbf{E}[X + Y + Z] = \mathbf{E}[X] + \mathbf{E}[Y] + \mathbf{E}[Z]$

- If X, Y are independent:
  - $\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[Y]$
  - $\mathbf{E}[g(X)h(Y)] = \mathbf{E}[g(X)] \cdot \mathbf{E}[h(Y)]$

Bernoulli distribution

Data, D = 

- P(Heads) = θ, P(Tails) = 1-θ

- Flips are **i.i.d.**:
  - **Independent** events
  - **Identically distributed** according to Bernoulli distribution

Choose θ that maximizes the probability of observed data