CS 316: Introduction to Deep Learning

Probability Overview
Week 3

Dr Abdul Samad

Lecture Outline

- Motivation
- Examples
- What is probability ?
- Basic probability theory
- Probability Mass Function (PMF)
- Probability Density Function (PDF)
- Common Probability Distributions
- Sampling Probability Distributions

Motivation

- **Machine Learning is a lot about probabilities.**
- Many times in machine learning what you want to do is, calculate a probability of something given some other factors.
- Examples:
  - Image Recognition
  - Classification
  - Sentimental Analysis
  - Generative Models

Image Recognition

- What is the probability that there is a cat in the image ?
- $\mathbf{P}(\mathrm{cat}|\mathrm{image}) = \mathbf{P}(\mathrm{cat}|\mathrm{pixel}_1, \mathrm{pixel}_2, \ldots, \mathrm{pixel}_n)$



Classification

- What is the probability that the patient is healthy ?
- $\mathbf{P}(\text{healthy}|\text{symptoms and history})$

Patient 1	
Age	29
Gender	Female
Height	169 cm
Weight	62 kg
Smoker	No
...	...
Heart rate	63
Blood pressure	120 90

Sentimental Analysis

- Is this a happy sentence?
- $\mathbf{P}(\text{happy}|\text{words in the sentence})$

the first cold shower
even the monkey seems to want
a little coat of straw

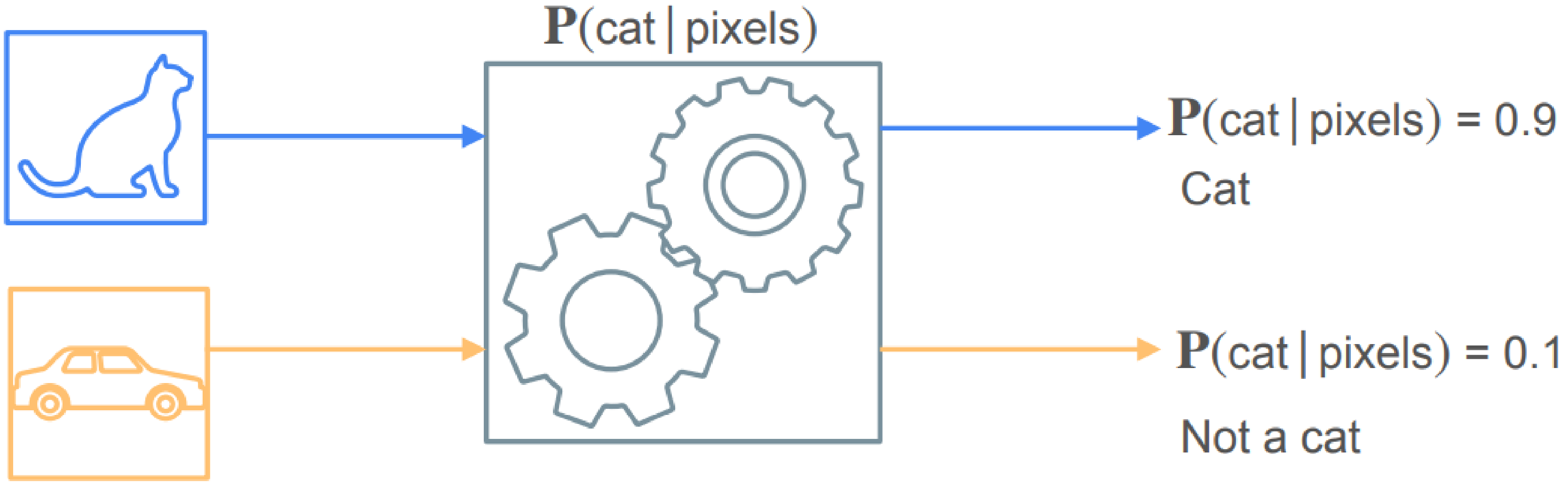
Matsuo Bashō

Generative Models

- Face generation
- Generate a group of pixels such that the resulting image looks like a human face.
- Generate images such that $\mathbf{P}(\text{face}|\text{pixels})$ is high.

Image generated by a StyleGAN

Machine Learning Model



What is Probability ?

- As image resolution decreases, it becomes challenging to tell cat and dog apart.
- Probability gives a formal way of reasoning about our level of uncertainty.

Fig. 2.6.1 Images of varying resolutions (10×10, 20×20, 40×40, 80×80, and 160×160 pixels).

Key Terms

- Outcome When something happens at random there are several potential outcomes. Exactly one of the outcomes occur.
- Sample space S contains all the outcomes.
- Event An event is defined to be some collection of outcomes.
- Empty set ϕ contains no outcomes. Sample Space contains all the outcomes.

Example

- E.g. roll a six sided die.
- The sample space S contains all the outcomes.

$$S = \{1,2,3,4,5,6\}$$

- Let A denote the event that the die roll is odd.

$$A = \{1,3,5\}$$

- Let B denote the event that the die roll is equal 2 or higher.

$$B = \{2,3,4,5,6\}$$

Axioms of Probability

- For any event A, its probability is never negative, i.e.,

$$P(A) \geq 0$$

- Probability of the entire sample space is 1, i.e.,

$$P(S) = 1$$

- For any countable sequence of events $A_1, A_2, \cdots$ that are mutually exclusive the probability that any happens is equal to the sum of their individual probabilities,

$$P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} A_i$$

Independence of Events

- Events A and B are called independent if

$$P(A \cap B) = P(A) \cdot P(B)$$

- A collection of 3 events A, B, C is called mutually independent if all four of the following conditions are satisfied.

$$P(A \cap B) = P(A) \cdot P(B)$$
$$P(A \cap C) = P(A) \cdot P(C)$$
$$P(B \cap C) = P(B) \cdot P(C)$$
$$P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)$$

Conditional Probability

- Conditional probability is the probability of an event occurring, given that another event has already occurred.
- E.g. Given that the test is positive, what is the probability the person has Covid.
- Let A denote the event of interest and B the event that we know has occurred. Then the probability of event A given B is computed as follows:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \text{ where } P(B) > 0$$

Bayes Theorem

- The Bayes' theorem is expressed in the following formula:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

where $P(A|B)$ is the probability of event A occurring, given event B has occurred

$P(B|A)$ is the probability of event B occurring, given event A has occurred,

$P(A)$ is the probability of event A occurring,

$P(B)$ is the probability of event A occurring

Example

Ali has two bags. Bag 1 has 7 red and 2 blue balls and Bag II has 5 red and 9 blue balls. Ali draws a ball at random and it turns out to be red. Determine the probability that the ball was from Bag I

Example

Ali has two bags. Bag 1 has 7 red and 4 blue balls and Bag II has 5 red and 9 blue balls. Ali draws a ball at random and it turns out to be red. Determine the probability that the ball was from Bag I

- Let A denote the event that the ball is picked from Bag I and B denote the event that the ball picked has the color red.

$$P(A) = \frac{1}{2}$$

$$P(B|A) = \frac{7}{11}$$

$$P(B) = \frac{1}{2} \cdot \frac{7}{11} + \frac{1}{2} \cdot \frac{5}{14}$$

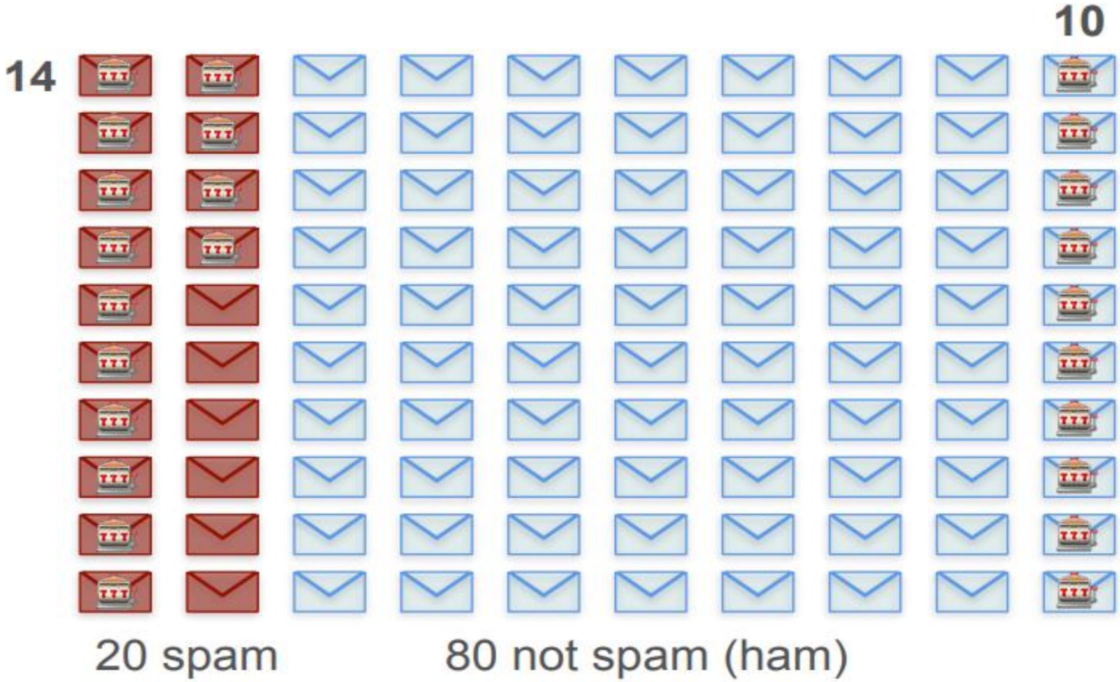
$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

$$P(A|B) = \frac{\frac{7}{11}\frac{1}{2}}{\frac{7}{11}\frac{1}{2} + \frac{5}{14}\frac{1}{2}} = 0.64$$

Example - Spam Classification

What is the probability that an email containing lottery is a spam?

$\mathbf{P}(\mathrm{spam}|\mathrm{lottery})$



Spam Classification - Intuitive Solution

$\mathbf{P}(\text{spam} \mid \text{lottery})$

$$\mathbf{P}(A \mid B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)}$$

14

24 emails

containing lottery

$$\mathbf{P}(\text{spam} \mid \text{lottery}) = \frac{\text{spam and lottery}}{\text{all lottery}}$$

$$= \frac{14}{24}$$

$$= \frac{7}{12} = 0.583$$

10

Spam Classification – Bayes Theorem

$$\mathbf{P}(\text{spam} \mid \text{lottery})$$

$$\mathbf{P}(A \mid B) = \frac{\mathbf{P}(A) \cdot \mathbf{P}(B \mid A)}{\mathbf{P}(A) \cdot \mathbf{P}(B \mid A) + \mathbf{P}(A') \cdot \mathbf{P}(B \mid A')}$$

A: Email is spam B: Email contains lottery

$$\mathbf{P}(\text{spam} \mid \text{lottery}) = \frac{\mathbf{P}(\text{spam}) \cdot \mathbf{P}(\text{lottery} \mid \text{spam})}{\mathbf{P}(\text{spam}) \cdot \mathbf{P}(\text{lottery} \mid \text{spam}) + \mathbf{P}(\text{not spam}) \cdot \mathbf{P}(\text{lottery} \mid \text{not spam})}$$

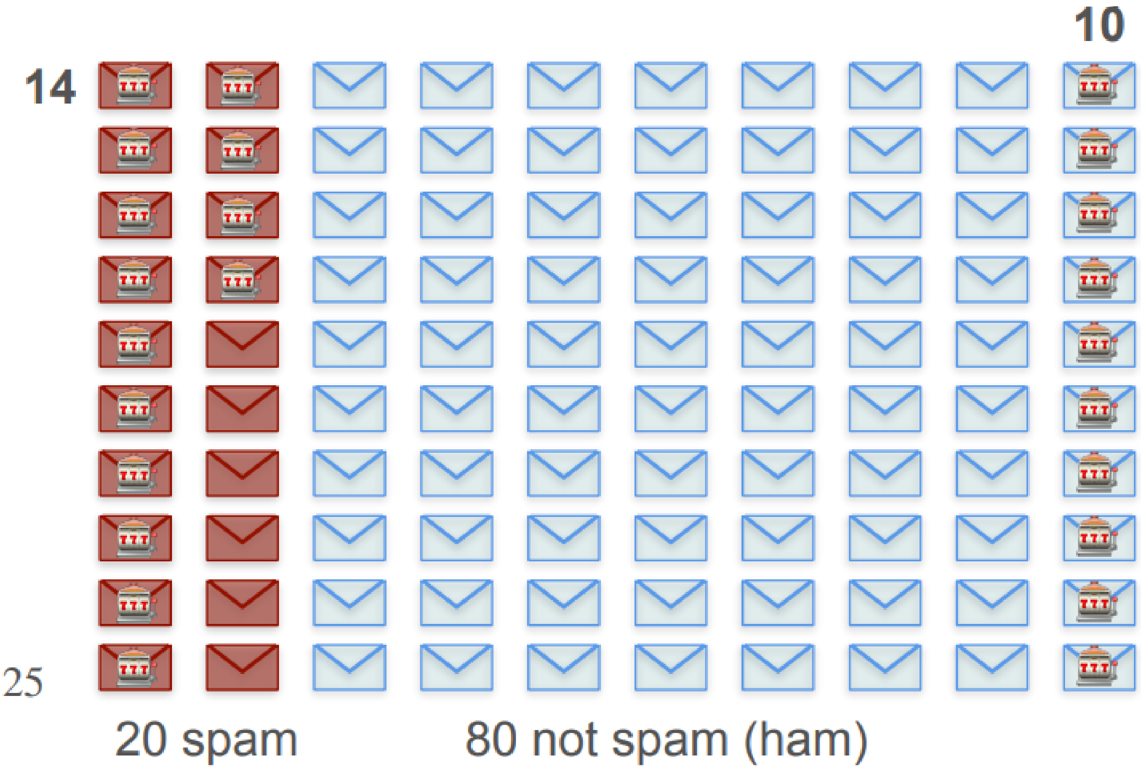
Spam Classification – Bayes Theorem

$$P(\text{spam}) = \frac{20}{100} = 0.2$$

$$P(\text{not spam}) = \frac{80}{100} = 0.8$$

$$P(\text{lottery} \mid \text{spam}) = \frac{14}{20} = 0.7$$

$$P(\text{lottery} \mid \text{not spam}) = \frac{10}{80} = 0.125$$



Spam Classification – Bayes Theorem

$$\mathbf{P}(\text{spam}) = 0.2 \qquad \mathbf{P}(\text{lottery} \mid \text{spam}) = 0.7$$

$$\mathbf{P}(\text{not spam}) = 0.8 \qquad \mathbf{P}(\text{lottery} \mid \text{not spam}) = 0.125$$

$$\mathbf{P}(\text{spam} \mid \text{lottery}) = \frac{\mathbf{P}(\text{spam}) \cdot \mathbf{P}(\text{lottery} \mid \text{spam})}{\mathbf{P}(\text{spam}) \cdot \mathbf{P}(\text{lottery} \mid \text{spam}) + \mathbf{P}(\text{not spam}) \cdot \mathbf{P}(\text{lottery} \mid \text{not spam})}$$

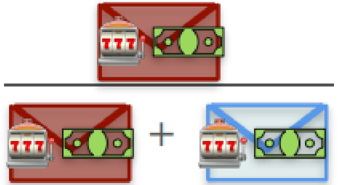
$$\mathbf{P}(\text{spam} \mid \text{lottery}) = \frac{0.2 \times 0.7}{(0.2 \times 0.7) + (0.8 \times 0.125)} = 0.583$$

Spam Classification – Naïve Bayes Theorem

What is the probability that an email containing lottery and winning is a spam?

Email contains ‘lottery’ and ‘winning’

$\mathbf{P}(\text{spam} \mid \text{lottery \& winning}) =$



$$\frac{\text{\# Spam emails with 'lottery' and 'winning'}}{\text{\# Emails with 'lottery' and 'winning'}}$$

?

$$\mathbf{P}(\text{spam} \mid \text{lottery \& winning}) = \frac{\mathbf{P}(\text{spam}) \cdot \mathbf{P}(\text{lottery \& winning} \mid \text{spam})}{\mathbf{P}(\text{spam}) \cdot \mathbf{P}(\text{lottery \& winning} \mid \text{spam}) + \mathbf{P}(\text{ham}) \cdot \mathbf{P}(\text{lottery \& winning} \mid \text{ham})}$$

Spam Classification – Naïve Bayes Theorem

What is the probability that an email containing lottery and winning is a spam?

Naive assumption

The appearances of 'lottery' and 'winning' are independent

$$\mathbf{P}(A \cap B) = \mathbf{P}(A) \cdot \mathbf{P}(B)$$

$$\mathbf{P}(\text{spam} \mid \text{lottery \& winning}) = \frac{\mathbf{P}(\text{spam}) \cdot \mathbf{P}(\text{lottery} \mid \text{spam}) \cdot \mathbf{P}(\text{winning} \mid \text{spam})}{\mathbf{P}(\text{spam}) \cdot \mathbf{P}(\text{lottery} \mid \text{spam}) \cdot \mathbf{P}(\text{winning} \mid \text{spam}) + \mathbf{P}(\text{ham}) \cdot \mathbf{P}(\text{lottery} \mid \text{ham}) \cdot \mathbf{P}(\text{winning} \mid \text{ham})}$$

Spam Classification – Naïve Bayes Theorem

What is the probability that an email containing the words $w_1, w_2, \ldots, w_n$ a spam?

Naive assumption

The appearances of the words $w_1, w_2, \ldots, w_n$ are independent

$$\mathbf{P}(\text{spam} \mid w_1, \ldots, w_n) = \frac{\mathbf{P}(\text{spam}) \cdot \mathbf{P}(w_1 \mid \text{spam}) \cdots \mathbf{P}(w_n \mid \text{spam})}{\mathbf{P}(\text{spam}) \cdot \mathbf{P}(w_1 \mid \text{spam}) \cdots \mathbf{P}(w_n \mid \text{spam}) + \mathbf{P}(\text{ham}) \cdot \mathbf{P}(w_1 \mid \text{ham}) \cdots \mathbf{P}(w_n \mid \text{ham})}$$

Probability mass function (PMF) of a discrete r.v

- It is the "probability law" or "probability distribution" of X.
- If we fix some *x* , then "X = *x*" is an event

$$p_X(x) = P(X = x) = P(\omega \in \Omega \text{ s.t. } X(\omega) = x)$$

- Properties:

$$p_X(x) \geq 0$$

$$\sum_x p_X(x) = 1$$

Expectation, Variance, Standard Deviation

$$E[X] = \sum_x x \cdot p_X(X = x)$$

$$Var[X] = E[(X - \mu)^2] = \sum_x p_X(X = x)(X - \mu)^2$$

$$Var[X] = E[X^2] - (E[X])^2$$

$$S.D(X) = \sqrt{Var(X)}$$

Example

Let X be a random variable which denotes the toss of a single unfair die. The following table is a PMF of X.

X	1	2	3	4	5	6
P(X = x)	0.1	0.1	0.1	0.1	0.1	0.5

Compute E[X] and Var[X]

Example

Let X be a random variable which denotes the toss of a single unfair die. The following table is a PMF of X.

X	1	2	3	4	5	6
P(X = x)	0.1	0.1	0.1	0.1	0.1	0.5

Compute E[X] and Var[X]

$$E[X] = \sum_x x \cdot p_X(x)$$

$$E[X] = 0.1 + 0.2 + 0.3 + 0.4 + 0.5 + 3 = 4.5$$

$$Var[X] = E[X^2] - (E[X])^2$$

$$E[X^2] = 0.1 + 0.4 + 0.9 + 1.6 + 2.5 + 18 = 23.5$$

$$Var[X] = E[X^2] - (E[X])^2 = 23.5 - 4.5^2 = 3.25$$

Joint PMF and Marginal PMF

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

$$\sum_{x} \sum_{y} p_{(X,Y)}(x, y) = 1$$

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

$$p_Y(y) = \sum_{y} p_{(X,Y)}(x, y)$$

Independence and Conditional PMF

$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$$

$$p_Y(y)p_{X|Y}(x|y) = p_{X,Y}(x,y)$$

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

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

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

Example

Consider two random variables X and Y with the joint PMF given in the table below.

	Y = 2	Y = 4	Y = 5
X = 1	1/12	1/24	1/24
X = 2	1/6	1/12	1/8
X = 3	1/4	1/8	1/12

a) Find P(X<= 2, Y <=4)
b) Find the marginal PMFs of X and Y
c) Find P (Y = 2 | X = 1)
d) Are X and Y independent ?

Example

Consider two random variables X and Y with the joint PMF given in the table below.

	Y = 2	Y = 4	Y = 5
X = 1	1/12	1/24	1/24
X = 2	1/6	1/12	1/8
X = 3	1/4	1/8	1/12

a) Find P(X<= 2, Y <=4)
b) Find the marginal PMFs of X and Y
c) Find P (Y = 2 | X = 1)
d) Are X and Y independent ?

$$a) P(X \leq 2, Y \leq 4) = \frac{1}{12} + \frac{1}{24} + \frac{1}{6} + \frac{1}{12} = \frac{3}{8}$$

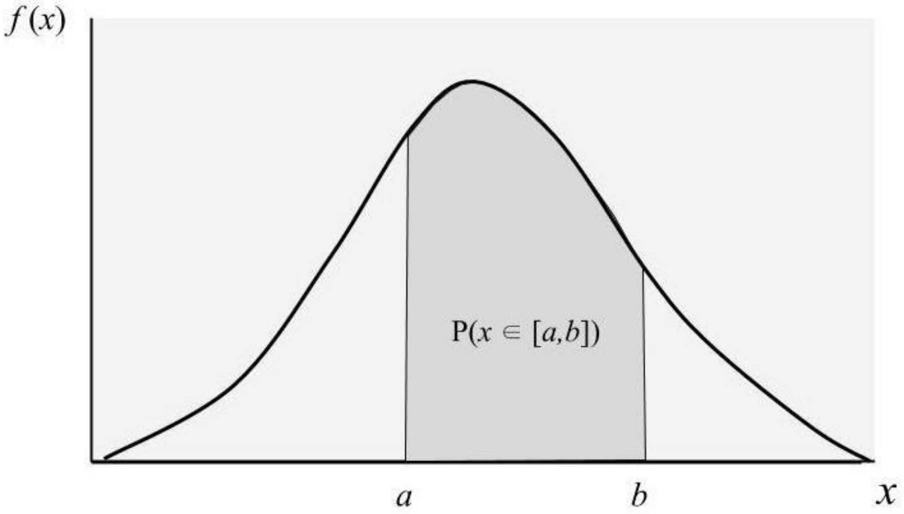
Y	2	4	5
P(Y=y)	1/2	1/4	1/4

X	1	2	3
P(X=x)	1/6	3/8	11/24

$$c) P(Y = 2 | X = 1) = \frac{\frac{1}{6}}{\frac{1}{12}}$$

$$d) p_{X,Y}(2,2) \neq p_X(2) \cdot p_Y(2); \frac{1}{6} \neq \frac{3}{16}$$

Probability Density Function (PDF) of a continuous r.v.



$$P(a \leq X \leq b) = \int_a^b f_X(x)dx$$

$$\int_{-\infty}^{+\infty} f_X(x)dx = 1$$

$$E[X] = \int_{-\infty}^{+\infty} x f_X(x)dx$$

$$Var[X] = \int_{-\infty}^{+\infty} (x-\mu)^2 f_X(x)dx$$

Joint and Marginal PDF

$$P(a \leq X \leq b, c \leq X \leq d) = \int_c^d \int_a^b f_{X,Y}(x,y)dxdy$$

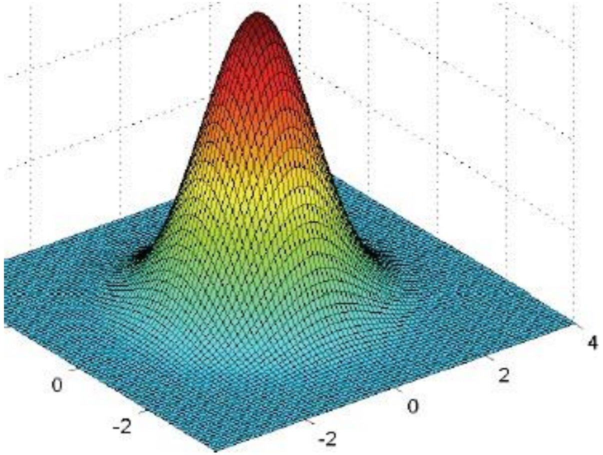
$$\int \int f_{X,Y}(x,y)dxdy = 1$$

$$f_X(x) = \int f_{X,Y}(x,y)dy$$

$$f_Y(y) = \int f_{X,Y}(x,y)dx$$

Conditional PDF

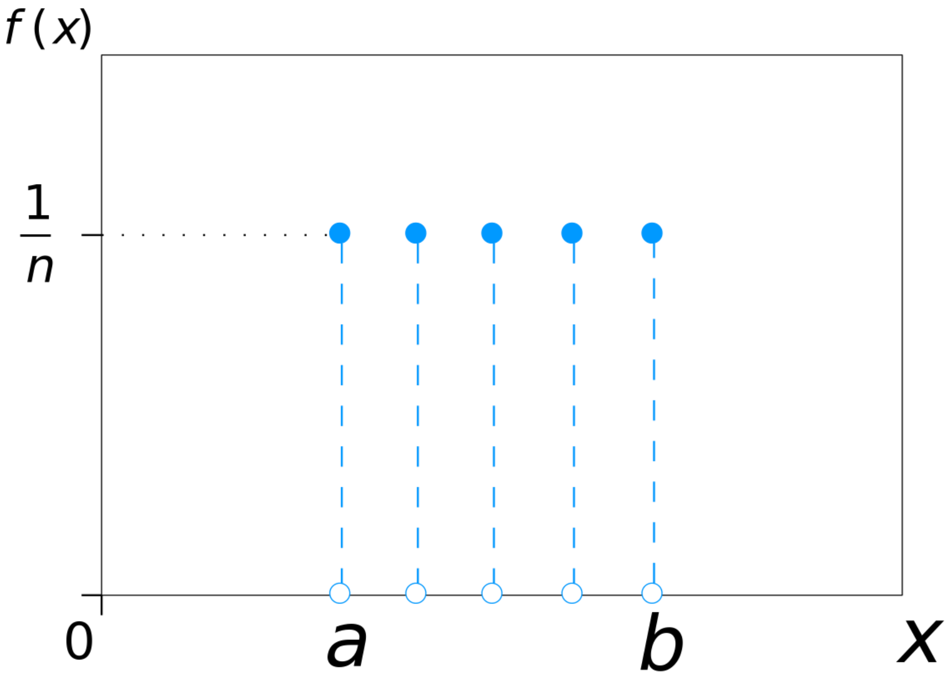
$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$



Common Probability Distributions

- Discrete
    - Discrete Uniform
    - Bernoulli
    - Binomial
    - Poisson
- Continuous
    - Continuous Uniform
    - Normal
    - Exponential

Discrete Uniform Distribution

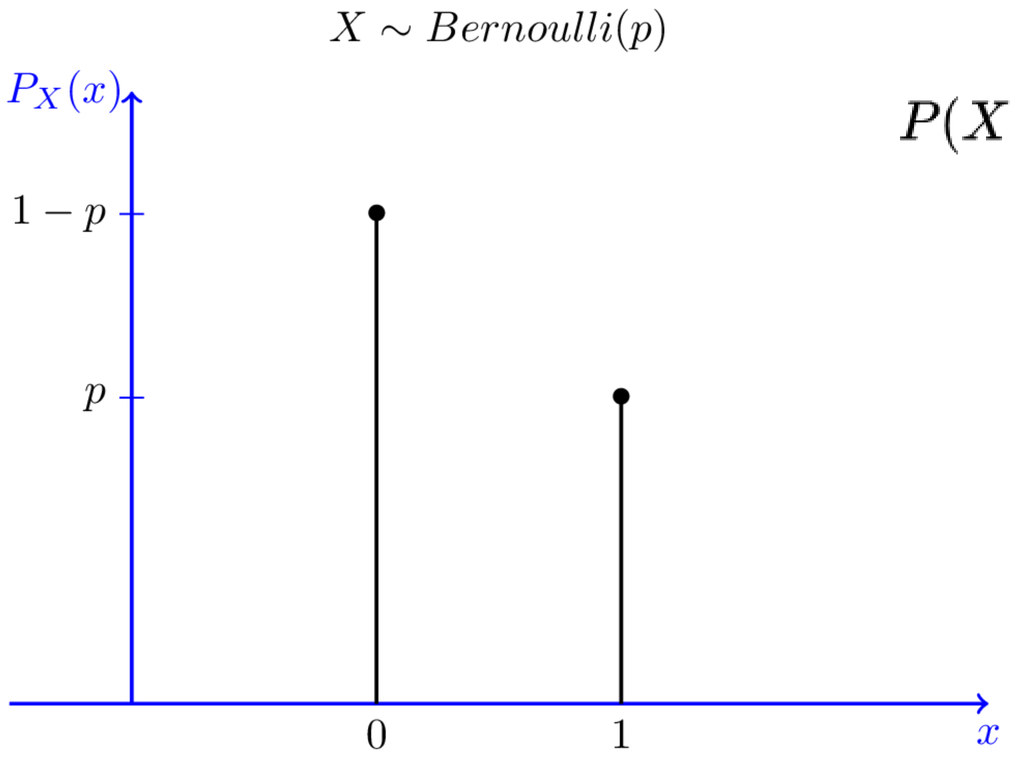


$$P(X = x) = \frac{1}{n}$$

$$E[X] = \frac{a+b}{2}$$

$$Var[X] = \frac{(b-a+1)^2 - 1}{12}$$

Bernoulli Distribution

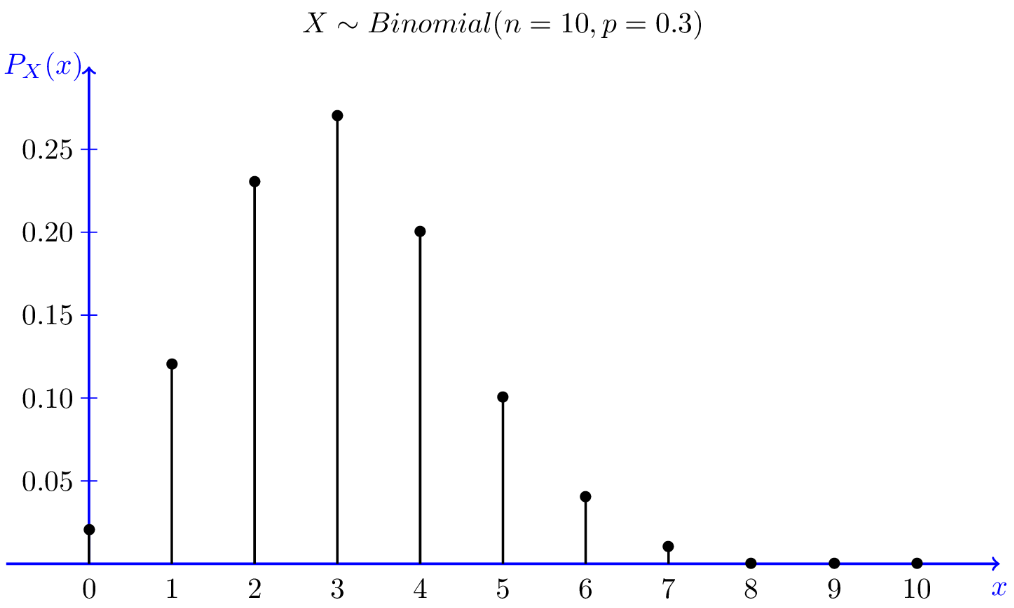


$$P(X=x) = \begin{cases} p & \text{for } x = 1 \\ p-1 & \text{for } x = 0 \end{cases}$$

$$E[X] = p$$

$$Var[X] = p \cdot (1-p)$$

Binomial Distribution

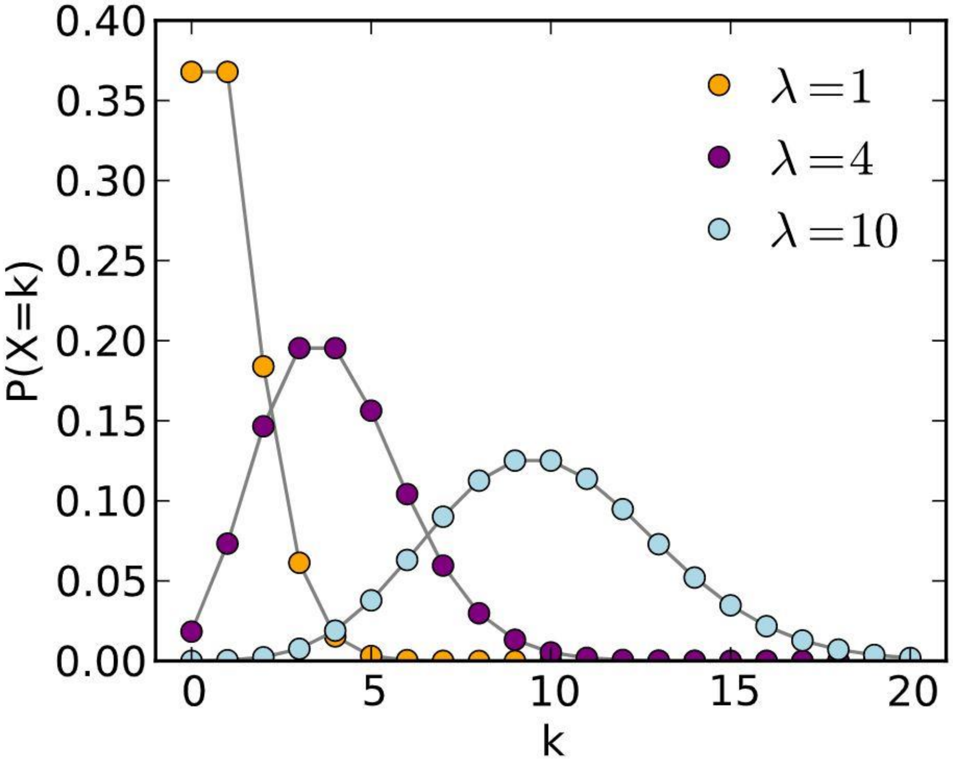


$$P(X = x) = \binom{n}{x} p^x q^{n-x}$$

$$E[X] = np$$

$$Var[X] = npq$$

Poisson Distribution

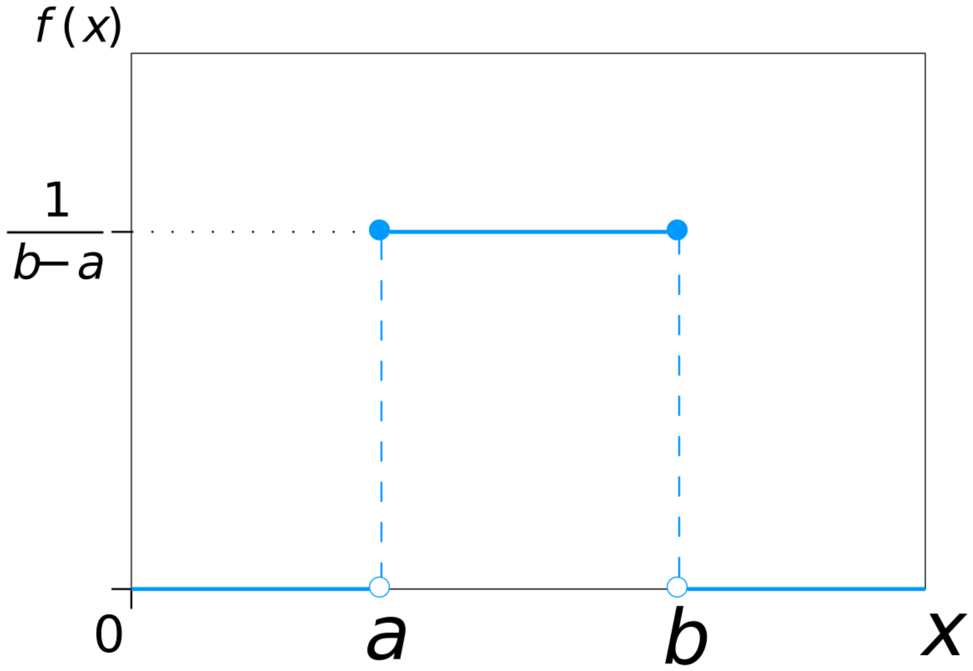


$$P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda}$$

$$E[X] = \lambda$$

$$Var[X] = \lambda$$

Continous Uniform Distribution

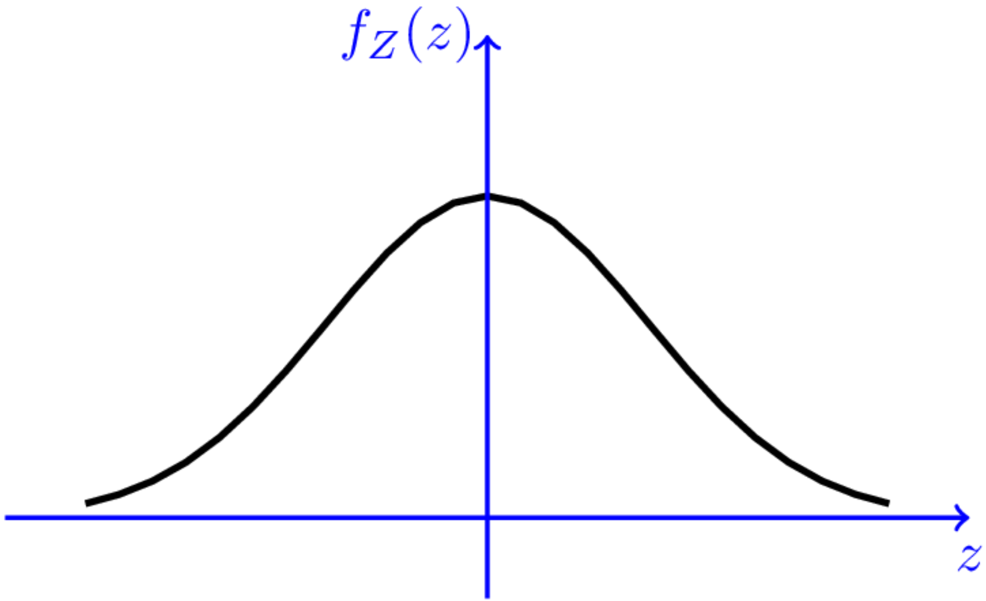


$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } x \in [a, b] \\ 0 & \text{otherwise} \end{cases}$$

$$E[X] = \frac{a+b}{2}$$

$$Var[X] = \frac{1}{12}(b-a)^2$$

Gaussian Distribution

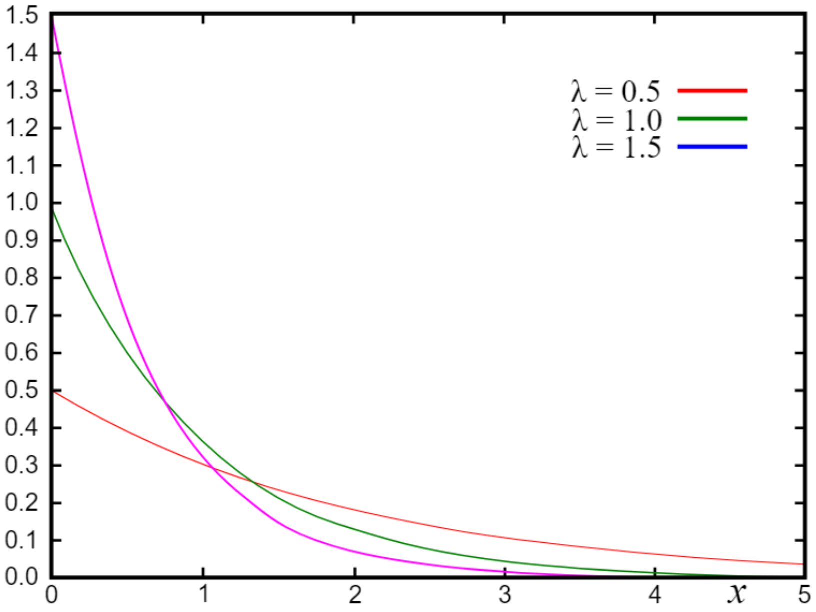


$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$$

$$E[X] = \mu$$

$$Var[X] = \sigma^2$$

Exponential Distribution

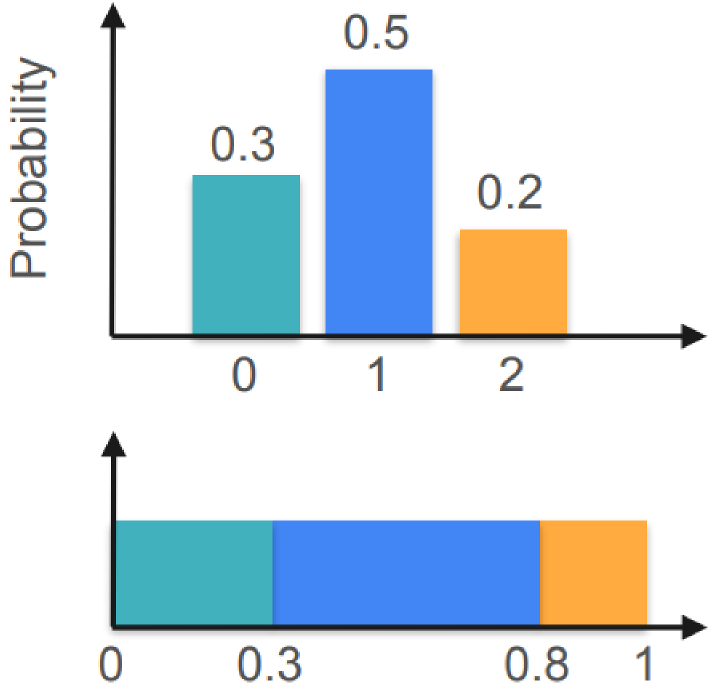


$$f(x) = \begin{cases} \lambda \exp(-\lambda x), x \geq 0 \\ 0, x < 0 \end{cases}$$

$$E[X] = \frac{1}{\lambda}$$

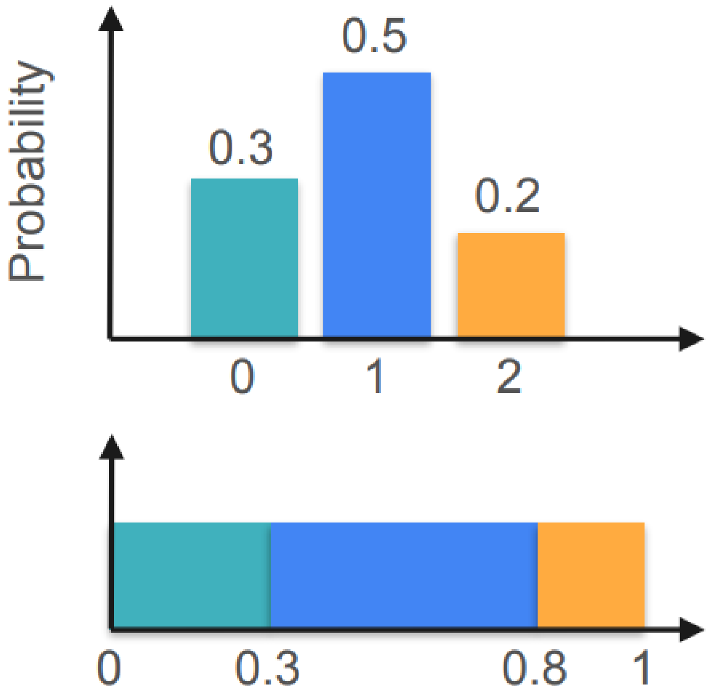
$$Var[X] = \frac{1}{\lambda^2}$$

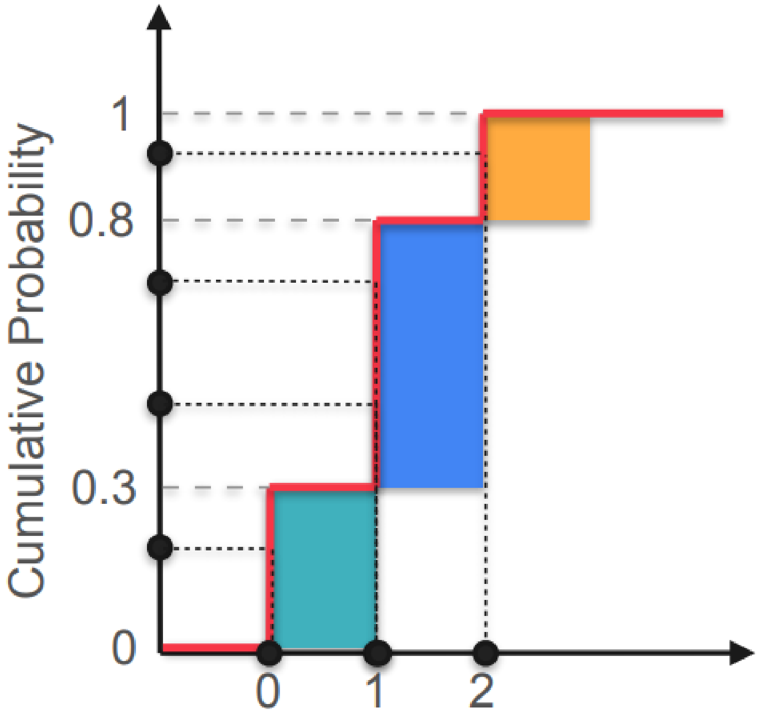
Sampling from a Distribution



- **Step 1**: generate a random number between 0 and 1
- **Step 2**: find out which interval the number belongs to
  - [0, 0.3)
  - [0.3, 0.8)
  - [0.8, 1]
- **Step 3**: Assign an outcome based on the interval

Sampling from a Distribution





Sampling from a Distribution