COMP 7180: Quantitative Methods for Data Analytics and Artificial Intelligence

Lecture 6: Introduction to Probability

Lecturer: Dr. Yang Liu and Dr. Jun Qi

Why Probability

- Probability theory is a mathematical framework for representing uncertain statements.

- The laws of probability tell us how AI systems should reason, so we design our algorithms to compute or approximate various expressions derived using probability theory.

- We can use probability and statistics to theoretically analyze the behavior of proposed AI systems.

Why Probability: An Example

- Suppose you are trying to determine if a patient has inhalational anthrax (吸入性炭疽病). You observe the following symptoms:

A. The patient has a cough；

B. The patient has a fever；

C. The patient has difficulty in breathing.

Why Probability: An Example

- You would like to determine how likely the patient is infected with inhalational anthrax given that the patient has a cough, a fever, and difficulty in breathing;

- We are not 100% certain that the patient has anthrax because of these symptoms. We are dealing with uncertainty!

Why Probability: An Example

- Now suppose you order an x-ray and observe that the patient has a wide mediastinum ((胸腔)纵隔);

- Your belief the probability that the patient is infected with inhalational anthrax is now much higher.

Why Probability: An Example

- In the previous slides, what you observed affected your belief that the patient is infected with anthrax;
- This is called reasoning with uncertainty;
- Wouldn't it be nice if we had some methodology for reasoning with uncertainty? In fact, we do ! ☺

What is Probability

- A probability can be regard as a function to estimate the value of every event.

As a function, we should have a domain (定义域). What is the domain?

Given a sample space S: set of all possible outcomes of an experiment.

The domain consists of some subsets of S.

An element E in the domain is called event.

What is Probability

Example: Toss a coin (1 time). Then, the outcome is H or T, where H is the head of a coin and T is the tail of a coin.

Then S= {H, T};

The domain is { {H,T}, {H}, {T}, ∅}.

{H,T}, {H}, {T}, ∅ are called events.

What is Probability

The domain should satisfy some special properties:

- S and $\varnothing$ should be event;
- If E is an event, then E^C is an event (E^C = S-E);
- If E and F are both events, then E∩F is an event, that is event E and event F occur at the same time;
- If E and F are both events, then E∪F is an event, that is event E occur or event F occur .

What is Probability

Example: Toss a coin (1 time). Then, the outcome is H or T, where H is the head of a coin and T is the tail of a coin.

Then S= { H, T}; The domain is { {H,T}, {H}, {T}, ∅}.

{H,T}, {H}, {T}, ∅ are called events.

- S and ∅ should be event;
- $S^C = \varnothing$; $\{H\}^C = \{T\}$; $\{T\}^C = \{H\}$; $\varnothing^C = S$;
- $S \cap \varnothing = \varnothing$; $S \cap \{H\} = \{H\}$; $S \cap \{T\} = \{T\}$; $\{H\} \cap \{T\} = \varnothing$;
- $\{H\} \cup \{T\} = S$; $\{H\} \cup \varnothing = \{H\}$; $\{T\} \cup \varnothing = \{T\}$.

What is Probability

As a function, we should have a range (值域). What is the range?

Given an event E, a probability maps E into [0,1], that is $0 \leq P(E) \leq 1$.

If P(E)=0, then this event E will not occur.

If P(E)=1, then this event E occurs without uncertainty.

What is Probability

Example. Toss a coin (1 time). There are outcomes: H and T, where H is the head of a coin and T is the tail of a coin.

$S = \{H, T\}$; The domain is $\{\{H,T\}, \{H\}, \{T\}, \varnothing\}$.

$\{H,T\}$, $\{H\}$, $\{T\}$, $\varnothing$ are called events.

$P(\{H,T\}) = 1$; $P(\{H\}) = 0.5$; $P(\{T\}) = 0.5$; $P(\varnothing) = 0$.

What is Probability

Probability is a special function, which should satisfy some properities:

- P(S)=1; P($\emptyset$)=0; 0$\leq$P(E) ≤ 1;

- If event E belongs to event F, then P(E)$\leq$P(F);

- Given an event E, then P(E^C) = 1-P(E);

- Given events E and F, then P(E$\cup$F) = P(E)+P(F)-P(E$\cap$F).

What is Probability

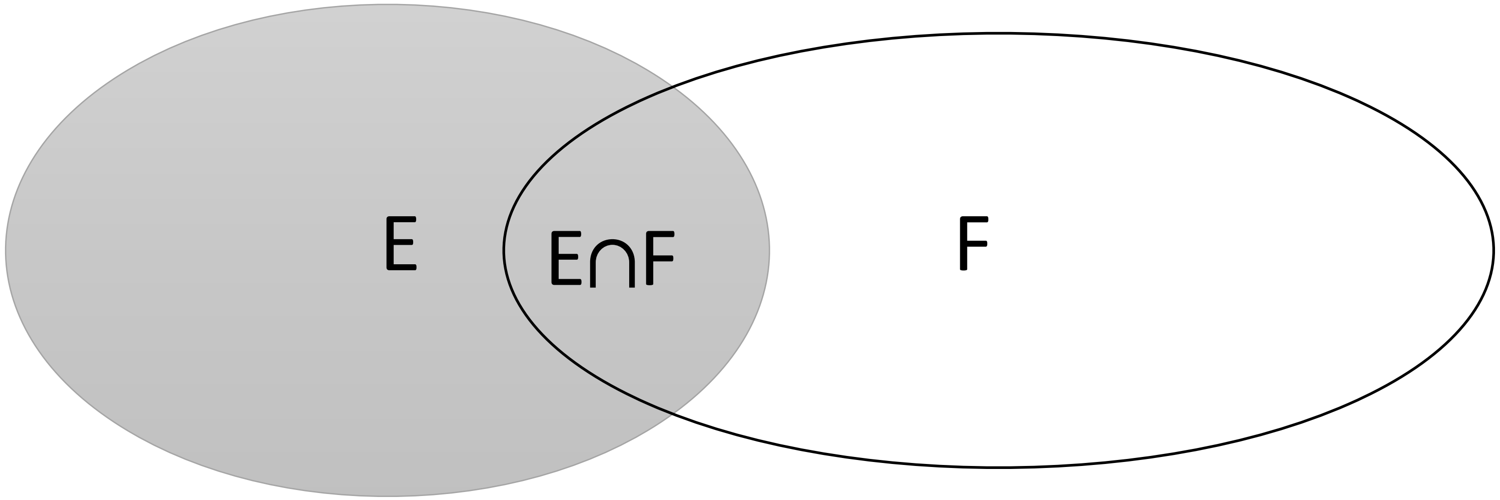
Example. Toss a coin (1 time). There are outcomes: H and T, where H is the head of a coin and T is the tail of a coin.

P({H,T}) =1; P({H}) = 0.5; P({T}) =0.5; P(∅)=0.

- P({H,T})=1; P(∅)=0;
- P({H})= 1-P({T}) and P({H,T}) = 1-P(∅);
- P({H}∪{T}) = P({H})+P({T})-P(∅).

What is Probability

- How to understand $P(E \cup F) = P(E)+P(F)-P(E \cap F)$?



Random Variables

- Generally, it is very complex to represent an event;

- To deal with more complex events, researchers have developed random variables (随机变量).

- Example. Toss a coin (1 time). In the sample space $S=\{H, T\}$, we design a function $X: S \rightarrow \{1,-1\}$ such that $X(H)=1$ and $X(T)=-1$. Then X is

  a random variable.

  Moreover, $P(X=1) = P(\{H\}) = 0.5$ and $P(X=-1) = P(\{T\})=0.5$.

What are Random Variables

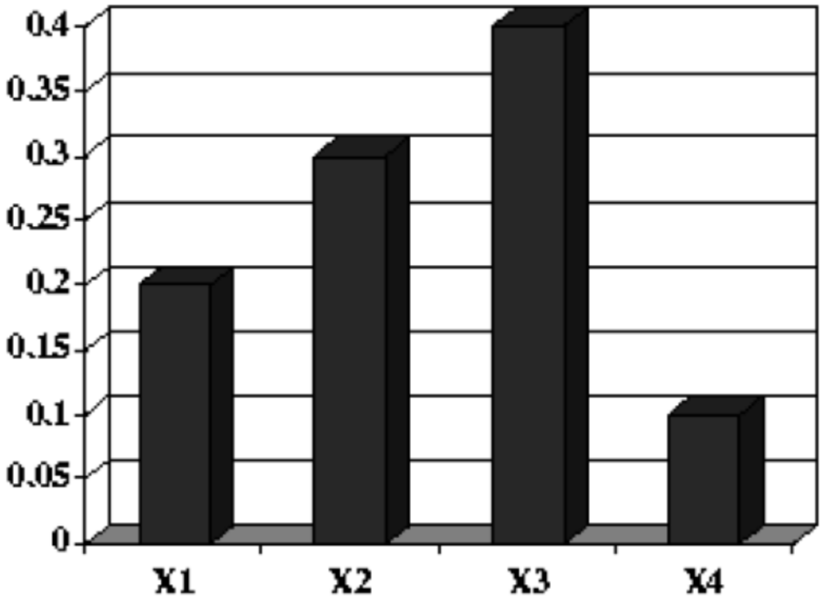
- A random variable is a variable that can take on different values randomly. We typically denote the random variable itself with an uppercase letter in plain typeface, and the values it can take on with lowercase letters.

- For vector-valued variables, we would write the random variable as **X** and one of its values as **x**.

- Random variables may be discrete or continuous. A discrete random variable is one that has a finite or countably infinite number of states. A continuous random variable is associated with a real value.

Probability Distributions

- A probability distribution is a description of how likely a random variable or set of random variables is to take on each of its possible states. The way we describe probability distributions depends on whether the variables are discrete or continuous.



Discrete Variables and PMF

- A probability distribution over discrete variables may be described using a probability mass function (PMF, 概率质量函数)

- The probability mass function maps from a state of a random variable to the probability of that random variable taking on that state.

- $0 \leq \mathrm{P}(\mathrm{X} = \mathrm{x}) \leq 1$

- $\sum_{\mathrm{x}} \mathrm{P}(\mathrm{X} = \mathrm{x}) = 1$. We refer to this property as being normalized

Discrete Variables and PMF: Examples

- Discrete uniform distribution (均匀分布) is one of the most important discrete distributions
- It is a finite discrete distribution
- Assume that the range is $x_1, x_2 \ldots x_n$, then
- $\mathrm{P}(\mathrm{X} = \mathrm{x_i}) = \frac{1}{n}; \quad \sum_i \mathrm{P}(\mathrm{X} = \mathrm{x_i}) = \sum_i \frac{1}{n} = \frac{n}{n} = 1$

Continuous Variables and PDF

- A continuous variable X is a function;
- Range is not discrete and take values in real number;
- There is a probability density function (PDF) (概率密度函数) $p_X(x)$ such that

1) $p_X(x) \geq 0$;

2) P(a≤X≤b) = $\int_a^b p_X(x)\, dx$;

3) $\int_{-\infty}^{+\infty} p_X(x)\, dx = 1$.

Continuous Variables and PDF

- In principle, variables such as height, weight, and temperature are continuous; in practice, the limitations of our measuring instruments restrict us to a discrete (though sometimes very finely subdivided) world.

- However, continuous models often approximate real-world situations very well, and continuous mathematics (calculus) is frequently easier to work with than mathematics of discrete variables and distributions.

Continuous Variables and PDF: Example

- The weight of a certain animal like a dog.

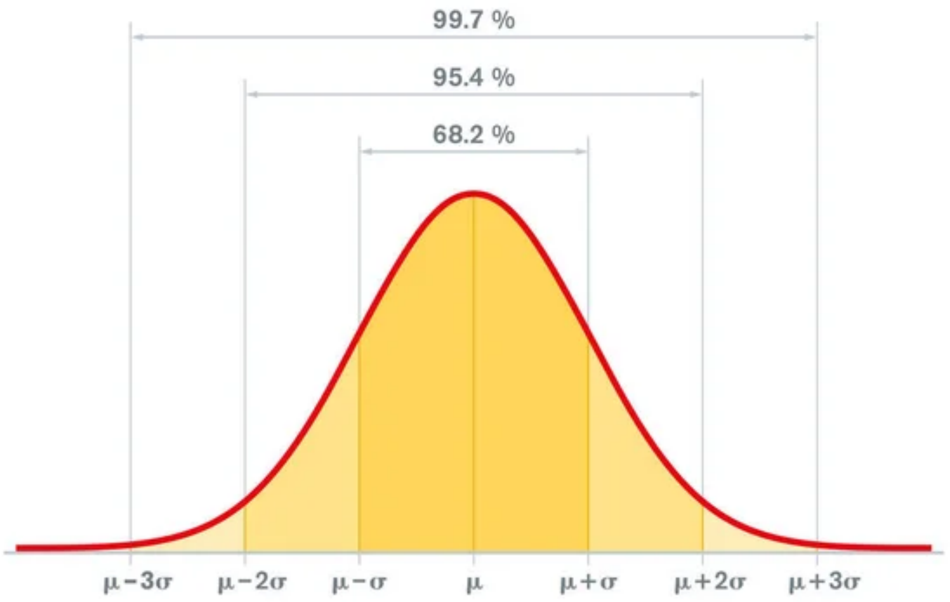
This is a continuous random variable because it can take on an infinite number of values. For example, a dog might weigh 30.333 pounds, 50.340999 pounds, 60.5 pounds, etc.

What is the distribution of dog's weight?

Continuous Variables and PDF

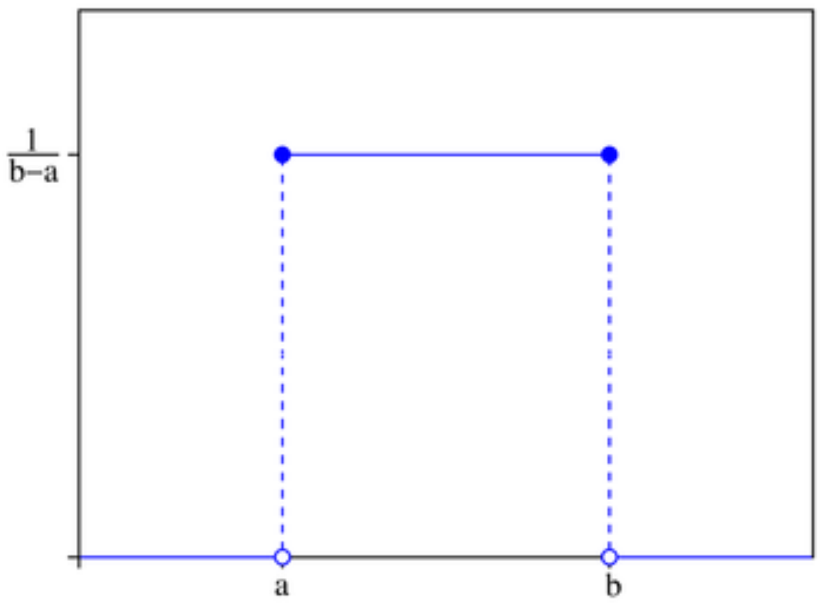
It is similar to a Gaussian distribution. The probability density funtion is

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



Continuous Variables and PDF

- Continuous uniform distribution is one of the most important continuous distributions.
- The probability density function of continuous unfiorm distribution can be written as $p(\mathrm{x}; a, b) = \dfrac{1}{b-a}$



Exercise 1

Classify each random variable as either discrete or continuous

1. The number of applicants for a job.
2. The time between customers entering a checkout lane at a retail store.
3. The temperature of a cup of coffee served at a restaurant.
4. The air pressure of a tire on an automobile.
5. The number of students who actually register for classes at a university next semester.

Exercise 1

Classify each random variable as either discrete or continuous

1. The number of applicants for a job. Discrete
2. The time between customers entering a checkout lane at a retail store. Continuous
3. The temperature of a cup of coffee served at a restaurant. Continuous
4. The air pressure of a tire on an automobile.Continuous
5. The number of students who actually register for classes at a university next semester. Discrete

Exercise 2

Determine whether or not the table is a valid probability distribution of a discrete random variable. Explain fully.

- X = -2, 0, 2, 4. P(X=-2)=0.2, P(X=0) = 0.3, P(X=2) = 0.3, P(X=4)=0.2.

- X = 0, 1, -1. P(X=0) = 0.3, P(X=1) = -0.0001, P(X=-1) = 0.7001.

- X= 1, 2, 3. P(X=1) = 0.2, P(X=2) = 0.2. P(X=3)= 0.2.

- X= 1, 2, 3, 4. P(X=1) = 0.2, P(X=2) = 0.2. P(X=3)= 0.2 P(X=4)= 0.5.

Exercise 2

Determine whether or not the table is a valid probability distribution of a discrete random variable. Explain fully.

- $X = -2, 0, 2, 4$. $P(X=-2)=0.2$, $P(X=0) = 0.3$, $P(X=2) = 0.3$, $P(X=4)=0.2$.

  Yes.

- $X = 0, 1, -1$. $P(X=0) = 0.3$, $P(X=1) = -0.0001$, $P(X=-1) = 0.7001$.

  No. Because $P(X=1)<0$.

- $X= 1, 2, 3$. $P(X=1) = 0.2$, $P(X=2) = 0.2$. $P(X=3)= 0.2$.

  No. Because $P(X=1)+P(X=2)+P(X=3)=0.6<1$.

- $X= 1, 2, 3, 4$. $P(X=1) = 0.2$, $P(X=2) = 0.2$. $P(X=3)= 0.2$ $P(X=4)= 0.5$.

  No. Because $P(X=1)+P(X=2)+P(X=3)+P(X=4)=1.1>1$.

Exercise 3

A discrete random variable X has the following probability distribution:

$X = 1, 3, 4, 70, 80, 90.$

$P(X=1) = 0.1$, $P(X=3)= 0.2$, $P(X=4) = 0.1$, $P(X=70)= 0.3$, $P(X=80) = 0.2$.

What is $P(X=90)$?

What is $P(X<70)$?

What is $P(70 \leq X<90)$?

Exercise 3

A discrete random variable X has the following probability distribution:

$X = 1, 3, 4, 70, 80, 90.$

$P(X=1) = 0.1$, $P(X=3)= 0.2$, $P(X=4) = 0.1$, $P(X=70)= 0.3$, $P(X=80) = 0.2$.

What is $P(X=90)$?

$P(X=90) = 1-P(X=1)-P(X=3)-P(X=4)-P(X=70)-P(X=80)=0.1$

What is $P(X<70)$? $P(X<70)=P(X=1)+P(X=3)+P(X=4)=0.4$

What is $P(90>X\geq 70)$?

$P(90>X\geq 70)=P(X=70)+P(X=80)=0.3+0.2=0.5$

Exercise 4

- A standard dice (骰子) has six sides printed with little dots numbering 1, 2, 3, 4, 5, and 6. Assume that the dice is fair, i.e., each of these outcomes is equally likely. Since there are six possible outcomes, the probability of obtaining any side of the dice is 1/6. Now you are throwing the dice for three times. What is the probability that at least two throws have the same outcome?

Exercise 4

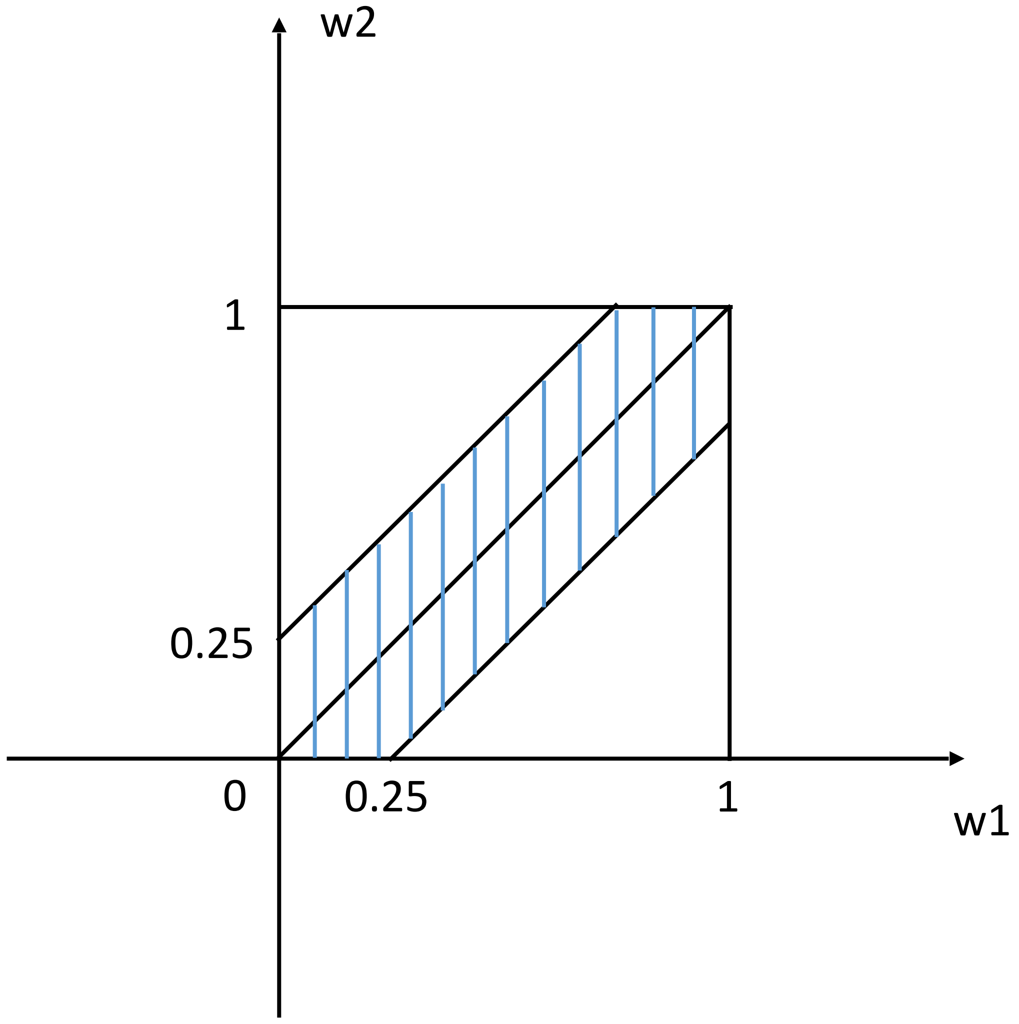
- A standard dice (骰子) has six sides printed with little dots numbering 1, 2, 3, 4, 5, and 6. Assume that the dice is fair, i.e., each of these outcomes is equally likely. Since there are six possible outcomes, the probability of obtaining any side of the dice is 1/6. Now you are throwing the dice for three times. What is the probability that at least two throws have the same outcome?
- Solution: We are required to calculate P(A), where A stands for the event that “at least two throws have the same outcome”. Then event ~A is “all three throws have different outcomes”, and we have P(~A) = (6*5*4)/(6*6*6) = 5/9. So that P(A) = 1 - P(~A) = 4/9.

Exercise 5

- We are in the process of fine-tuning the parameters of a machine learning algorithm, specifically two parameters: w1 and w2. It is established that both w1 and w2 follow a uniform distribution within the range [0,1]. In order for the algorithm to function optimally, it is crucial that the difference between w1 and w2 does not exceed 1/4 (0.25). Now, the question arises: What is the probability of the algorithm successfully meeting this criterion and working effectively?

Exercise 5

- Two parameters: w1 and w2:
  - $0 \leq w1, w2 \leq 1$.
  - $|w1 - w2| \leq 1/4$.
  - $P(|w1 - w2| \leq 1/4) / P(0 \leq w1, w2 \leq 1)$
    $= 1 - (3/4)^2 = 7/16$



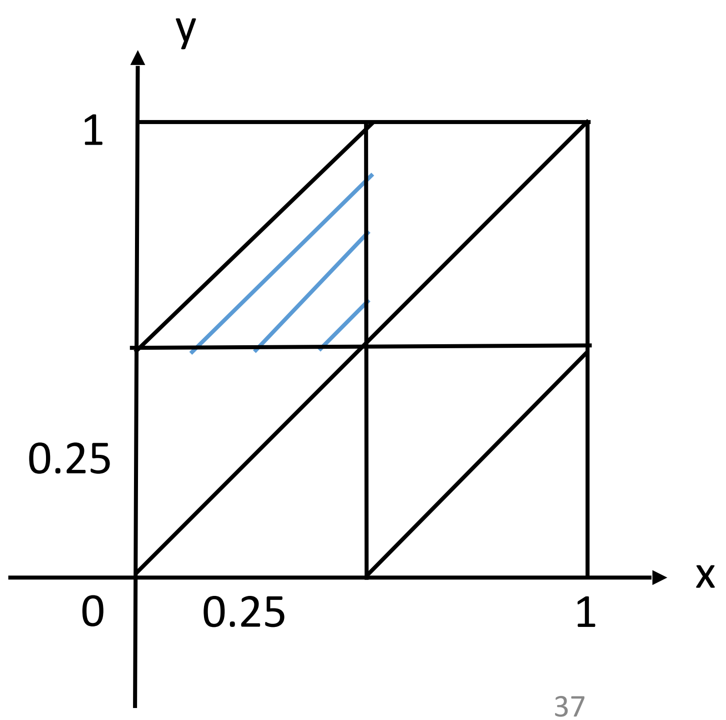
Exercise 6

- Given a line segment AD, we select two points, B and C, on AD. Here, B and C are independently and uniformly distributed along AD. We then break the line segment at points B and C, resulting in three new line segments. The task is to determine the probability of these three new line segments being able to form a triangle.

Exercise 6

- Given a line segment AD, we select two points, B and C, on AD. Here, B and C are independently and uniformly distributed along AD. We then break the line segment at points B and C, resulting in three new line segments. The task is to determine the probability of these three new line segments being able to form a triangle.

- Solution: Assume that the coordinates of A, B, C, D are 0, x, y, 1, respectively. Here we have 0 ≤ x, y ≤ 1.
- Let's first assume x < y. Then the three new segments are AB, BC, and CD. Their lengths are x, y-x, 1-y. To form a triangle, the following three inequalities should be satisfied:
- |AB| + |BC| > |CD|, |AB| + |CD| > |BC|, |BC| + |CD| > |AB|,
- x + (y-x) > 1-y,   x + (1-y) > y-x,   (y-x) + (1-y) > x ➔   y > 1/2, y-x < 1/2, x < 1/2.
- P(x < y and form a triangle) = 1/8.
- Similarly, we have P(x > y and form a triangle) = 1/8.
- So P(new line segments being able to form a triangle) = 1/4.



Joint Distribution

- In some practice case, we need to consider multiple randon variables

For example, it is clear that the weight is related to the height. So we are interested in knowing the joint distribution related to dog's weight and height.

- If random variables X, Y are discrete random variables, then the joint distribution of X and Y is

$$P(X=x, Y=y)$$

Joint Distribution

- If random variables X, Y are continuous random variables, then the joint distribution of X and Y is

$$P(a1 \leq X \leq b1,\ a2 \leq Y \leq b2)$$

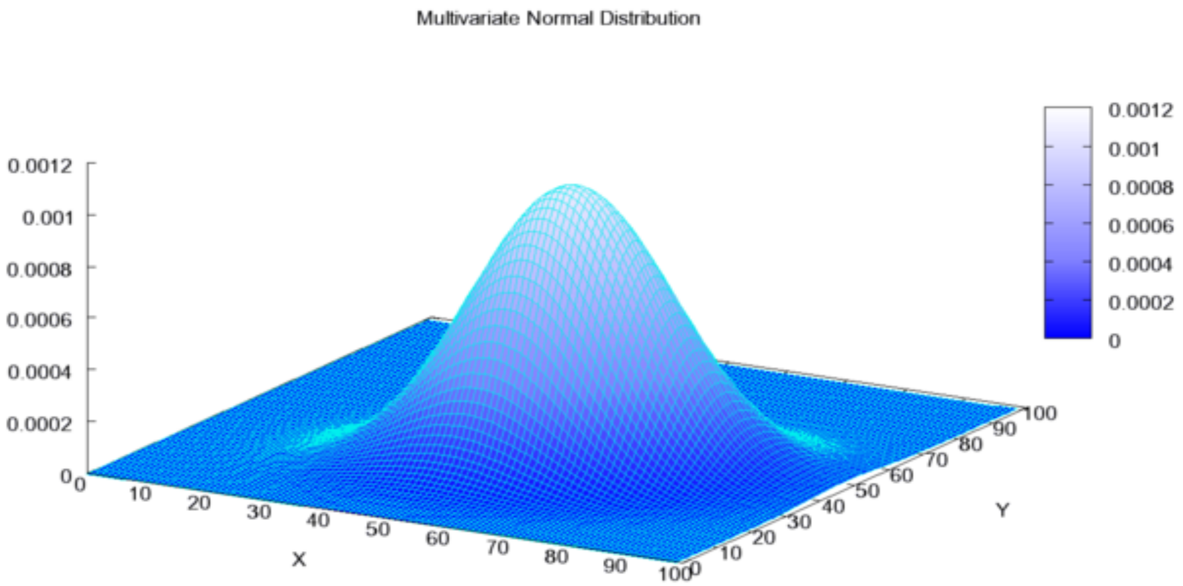
- In fact, when X, Y are continuous random variables, there exists a probability density function for the joint distribution:

$$P(a1 \leq X \leq b1,\ a2 \leq Y \leq b2) = \int_{a1}^{b1} \int_{a2}^{b2} p_{XY}(x,y)\,dx\,dy,$$

where $p_{XY}(x,y)$ is the probability density function.

Joint Distribution

- If X represents the weight of dog and Y represents the height of dog, then the joint distribuion P(X,Y) is similar to a two-dimensional Gaussian distribution.



Marginal Distribution

- Using joint distribution, we can construct marginal distribution:
- If random variables X, Y are discrete random variables, then the marginal distributions are

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

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

Marginal Distribution

- If random variables X and Y are continuous random variables, then the marginal distribution with respect to X is

$$\mathrm{P}(\mathrm{a} \leq \mathrm{X} \leq \mathrm{b}) = \int_{a}^{b} \int_{-\infty}^{+\infty} p_{\mathrm{XY}}(\mathrm{x}, \mathrm{y}) d\mathrm{y} d\mathrm{x}$$

and the density function of X is

$$p(\mathrm{x}) = \int_{-\infty}^{+\infty} p_{\mathrm{XY}}(\mathrm{x}, \mathrm{y}) d\mathrm{y}$$

Similarly, we can obtain the marginal distribution with respect to Y and the density function of Y.

Joint Probability and Marginal Probability : An Example

- Joint probabilities can involve any number of variables
- For each combination of variables, we need to say how probable that combination is
- The probabilities of these combinations need to sum to 1
- Once you have the joint probability distribution, you can calculate any probability involving X, Y, and Z

X	Y	Z	P(X,Y,Z)
0	0	0	0.1
0	0	1	0.2
0	1	0	0.05
0	1	1	0.05
1	0	0	0.3
1	0	1	0.1
1	1	0	0.05
1	1	1	0.15

Joint Probability and Marginal Probability: An Example

- P(X=1, Y =1)?
- = P(X=1, Y = 1, Z = 1) + P(X=1, Y = 1, Z = 0) = 0.2

- P(X=1, Y = 0)?
- = P(X=1, Y = 0, Z = 1) + P(X=1, Y = 0, Z= 0) = 0.4

- P(X=1)?
- = P(X=1, Y = 1) + P(X=1, Y = 0)
  =0.6

X	Y	Z	P(X,Y,Z)
0	0	0	0.1
0	0	1	0.2
0	1	0	0.05
0	1	1	0.05
1	0	0	0.3
1	0	1	0.1
1	1	0	0.05
1	1	1	0.15

- Exercise 7: Try to calculate the following by yourself
  - P(Y= 1)

Joint Probability and Marginal Probability: An Example

- Solution:

$P(Y=1) = P(X=1,Y=1,Z=1)+$

$P(X=0,Y=1,Z=1)+$

$P(X=1,Y=1,Z=0)+$

$P(X=0,Y=1,Z=0)$

$= 0.15+0.05+0.05+0.05$

$= 0.3$

X	Y	Z	P(X,Y,Z)
0	0	0	0.1
0	0	1	0.2
0	1	0	0.05
0	1	1	0.05
1	0	0	0.3
1	0	1	0.1
1	1	0	0.05
1	1	1	0.15

Conditional Probability

- Given an event E and an event F,

the condition probability of E given F is

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$

- P(X=1 | Y=1) = P(X = 1, Y = 1) / P(Y = 1)

= 0.2/0.3= 2/3

X	**Y**	**Z**	**P(X,Y,Z)**
0	0	0	0.1
0	0	1	0.2
0	1	0	0.05
0	1	1	0.05
1	0	0	0.3
1	0	1	0.1
1	1	0	0.05
1	1	1	0.15

Conditional Distribution

- Conditional distributions seek to answer the question, what is the probability distribution over Y , when we know that X must take on a certain value x.

If X and Y are discrete random variables, then

$$P(Y = y|X=x)= P(X=x,Y=y)/P(X=x)$$

Conditional Distribution

P(X=1 | Y=1) = P(X = 1, Y = 1) / P(Y = 1) = 0.2/0.3= 2/3

P(Z=0 | X=1) = P(X = 1, Z = 0) / P(X = 1) = 0.35/0.6= 7/12

Exercise 8: Try the calculate the following by yourself:

P(Z=0|X=1, Y=1)?

P(Y=0|X=1, Z=1)?

X	Y	Z	P(X,Y,Z)
0	0	0	0.1
0	0	1	0.2
0	1	0	0.05
0	1	1	0.05
1	0	0	0.3
1	0	1	0.1
1	1	0	0.05
1	1	1	0.15

Conditional Distribution

- Solution

$P(X=1,Y=1) = P(X=1,Y=1,Z=1)+P(X=1,Y=1,Z=0)$
$= 0.15+0.05=0.2$

$P(Z=0|X=1,Y=1) = P(X=1,Y=1,Z=0)/P(X=1,Y=1)$
$= 0.05/0.2=1/4$

$P(X=1,Z=1) = P(X=1,Y=1,Z=1)+P(X=1,Y=0,Z=1)$
$= 0.15+0.1=0.25$

$P(Y=0|X=1,Z=1) = P(X=1,Y=0,Z=1)/P(X=1,Z=1)$
$= 0.1/0.25= 2/5$

X	Y	Z	P(X,Y,Z)
0	0	0	0.1
0	0	1	0.2
0	1	0	0.05
0	1	1	0.05
1	0	0	0.3
1	0	1	0.1
1	1	0	0.05
1	1	1	0.15

Conditional Distribution

- If X and Y are continuous random variables, then

$$P(a2 \le Y \le b2 \mid a1 \le X \le b1) = \int_{a1}^{b1} \int_{a2}^{b2} p_{XY}(x,y)\,dy\,dx \Big/ \int_{a1}^{b1} p_X(x)\,dx,$$

If $p_X(x) > 0$, then P(Y|X=x) is a continuous distribution with probability density function

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

Chain Rule of Conditional Probabilities

- Given n events $E_1, E_2,\ldots,E_n$

$P(E_1 \cap E_2) = P(E_2|E_1)\ P(E_1);$

$P(E_1 \cap E_2 \cap E_3) = P(E_3 \cap E_2|E_1)P(E_1)=P(E_3|E_2 \cap E_1)P(E_2|E_1)P(E_1)$

....

$$P(E_1 \cap E_2 \ldots \cap E_n) = P(E_1)\prod_{i=2}^{n} P(E_i|E_1 \cap E_2 \cap \ldots \cap E_{i-1})$$

Above equation is called *chain rule*.

Chain Rule of Conditional Probabilities

- Chain rule with respect to random variables

Given random variables $X^1, \ldots, X^n$

$$P\left(X^1, \ldots, X^n\right) = P(X^1)\prod_{i=2}^{n} P(X^i | X^1, \ldots, X^{i-1})$$

Calculate P(Z=1|X=1, Y=1):

P(X=1, Y=1, Z=1) =0.15;

P(X=1) = 0.6;

P(Y=1|X=1)= 1/3;

By chain rule, we obtain that

P(Z=1|X=1, Y=1)

= P(X=1, Y=1, Z=1)/(P(Y=1|X=1)P(X=1))

= 0.15/0.2 = 3/4

X	**Y**	**Z**	**P(X,Y,Z)**
0	0	0	0.1
0	0	1	0.2
0	1	0	0.05
0	1	1	0.05
1	0	0	0.3
1	0	1	0.1
1	1	0	0.05
1	1	1	0.15

Chain Rule of Conditional Probabilities

Calculate $P(Z=1|X=0, Y=1)$

$P(X=0, Y=1, Z=1) = 0.05$;

$P(X=0) = 0.4$;

$P(Y=1|X=0) = 0.1/0.4 = 1/4$;

By chain rule, we obtain that

$P(Z=1|X=0, Y=1)$

$= P(X=0, Y=1, Z=1)/(P(Y=1|X=0)P(X=0))$

$= 0.05/0.1 = 1/2$

X	**Y**	**Z**	**P(X,Y,Z)**
0	0	0	0.1
0	0	1	0.2
0	1	0	0.05
0	1	1	0.05
1	0	0	0.3
1	0	1	0.1
1	1	0	0.05
1	1	1	0.15

Chain Rule of Conditional Probabilities

Exercise 9: Please use chain rule to compute

P(Z=1|X=0,Y=0)

X	Y	Z	P(X,Y,Z)
0	0	0	0.1
0	0	1	0.2
0	1	0	0.05
0	1	1	0.05
1	0	0	0.3
1	0	1	0.1
1	1	0	0.05
1	1	1	0.15

Chain Rule of Conditional Probabilities

- Solution

$P(X=0, Y=0, Z=1) = 0.2$;

$P(X=0) = 0.4$;

$P(Y=0|X=0) = (0.1+0.2)/0.4 = 3/4$;

By chain rule, we obtain that

$P(Z=1|X=0, Y=0)$

$= P(X=0, Y=0, Z=1)/(P(Y=0|X=0)P(X=0))$

$= 0.2/0.3 = 2/3$

X	**Y**	**Z**	**P(X,Y,Z)**
0	0	0	0.1
0	0	1	0.2
0	1	0	0.05
0	1	1	0.05
1	0	0	0.3
1	0	1	0.1
1	1	0	0.05
1	1	1	0.15

Exercise 10

- A standard die has six sides printed with little dots numbering 1, 2, 3, 4, 5, and 6. Assume that the die is fair, i.e., each of these outcomes is equally likely. Since there are six possible outcomes, the probability of obtaining any side of the die is 1/6. Now Bob is throwing the dice for three times. Assume that event A is “The outcome of the first throw is 1” and event B is “The summation of outcomes of three throws is no less than 10”. Calculate P(B|A) and show the detailed calculation procedure.

Exercise 10

- A standard die has six sides printed with little dots numbering 1, 2, 3, 4, 5, and 6. Assume that the die is fair, i.e., each of these outcomes is equally likely. Since there are six possible outcomes, the probability of obtaining any side of the die is 1/6. Now Bob is throwing the dice for three times. Assume that event A is “The outcome of the first throw is 1” and event B is “The summation of outcomes of three throws is no less than 10”. Calculate P(B|A) and show the detailed calculation procedure.
- Solution: P(B|A) = P(A,B)/P(A).
- Then event (A and B) is “The outcome of the first throw is 1 and (at the same time) the summation of outcomes of three throws is no less than 10”, which is equivalent to “The outcome of the first throw is 1 and the summation of outcomes of the second and the third throws is no less than 9”.
- There are 10 situations that satisfy this condition: (outcome of the second throw, outcome of the third throw) = (3,6), (4,5), (4,6), (5,4), (5,5), (5,6), (6,3), (6,4), (6,5), (6,6).
- We also know that there are 6*6 = 36 situations of the second and the third throws. Therefore, P(B|A) = P(A,B)/P(A) = 10/36 = 5/18.

Thank You!