COMP7180: Quantitative Methods for DAAI

(Credits from Prof. Andrew Ng)

(Credits from HKBU)

Course Instructors: Dr. Yang Liu and Dr. Bo Han

Teaching Assistant: Mr. Minghao Li

About Me

- Dr. Bo Han
  - Assistant Professor in HKBU CSD
  - Visiting Scientist in RIKEN AIP
  - Postdoctoral Researcher in RIKEN AIP
  - Ph.D., University of Technology Sydney
- Contact information
  - Email: bhanml@comp.hkbu.edu.hk
  - Homepage: https://bhanml.github.io/
  - Office hours: 2:00pm-3:00pm (Zoom, Sunday)
  - https://hkbu.zoom.us/j/6603117755

Timetable

- Time of our classes
  - 6 weeks from Oct 25 to Nov 29
  - Regular Time: 18:30-21:30 PM (Thu)

- Classroom of our class
  - Lectures: OEE1017 (Thu)

Course Contents

- Continuous and Discrete Random Variables (Week 7) Our Focus
- Conditional Probability and Independence (Week 8)
- Maximum Likelihood Estimation (Week 9)
- Mathematical Optimization (Week 10)
- Convex and Non-Convex Optimization (Week 11)
- Course Review (Week 12)

Course Mode

- Instructor: 3-hour book knowledge
- Office-hour: 0.5~1-hour office hour
- 1 Assignment + 1 Quiz
- Final exam (50% my part)

Learning Outcomes

- COMP7180: To learn the various **quantitative methods** (i.e., **mathematical experience**) necessary for data analytics and artificial intelligence (DAAI).
- Knowledge:
  - Explain the essential concepts in probability and statistics for DAAI
  - Understand the essential concepts in optimization for DAAI
- Professional Skill:
  - Determine suitable quantitative methods for effective data analytics
  - Apply suitable quantitative methods for real-world problem solving

- Compared to COMP7250: To introduce the fundamentals, models and techniques commonly found in machine learning. To gain some **hands-on experience** on developing machine learning solutions.

Assessment Methods

- Continuous Assessment (40%)
  - Assignments and Quizzes
- Examination (60%)
- Important Notices
  - Plagiarism: **Students who plagiarized and who were plagiarized will be given 0 mark**.
  - Final Exam: In order to pass this course, students should attain at least **30% of the final examination mark**.
  - A Cumulative **GPA at least 2.50** for graduation.

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

- Machine learning must always deal with uncertain quantities and sometimes stochastic (nondeterministic) quantities. Uncertainty and stochasticity can arise from many sources.

- (1) Inherent stochasticity in the system being modeled. (2) Incomplete observability. (3) Incomplete modelling.

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.

Credits from Dr. Liu Yang

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!

Credits from Dr. Liu Yang

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}, Ø}.

{H,T}, {H}, {T}, Ø are called events.

What is Probability

The domain should satisfy some speial properties:

- S and Ø 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}, Ø}. {H,T}, {H}, {T}, Ø are called events.

- S and Ø should be event;
- S^C = Ø; $\{H\}^C$ = {T}; $\{T\}^C$={H}; $Ø^C$= S;
- S∩ Ø=Ø; S∩{H}={H}; S∩{T}={T}; {H}∩{T}=Ø;
- {H}∪{T}=S; {H}∪ Ø={H}; {T}∪ Ø={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}, Ø}.

{H,T}, {H}, {T}, Ø are called events.

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

What is Probability

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

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

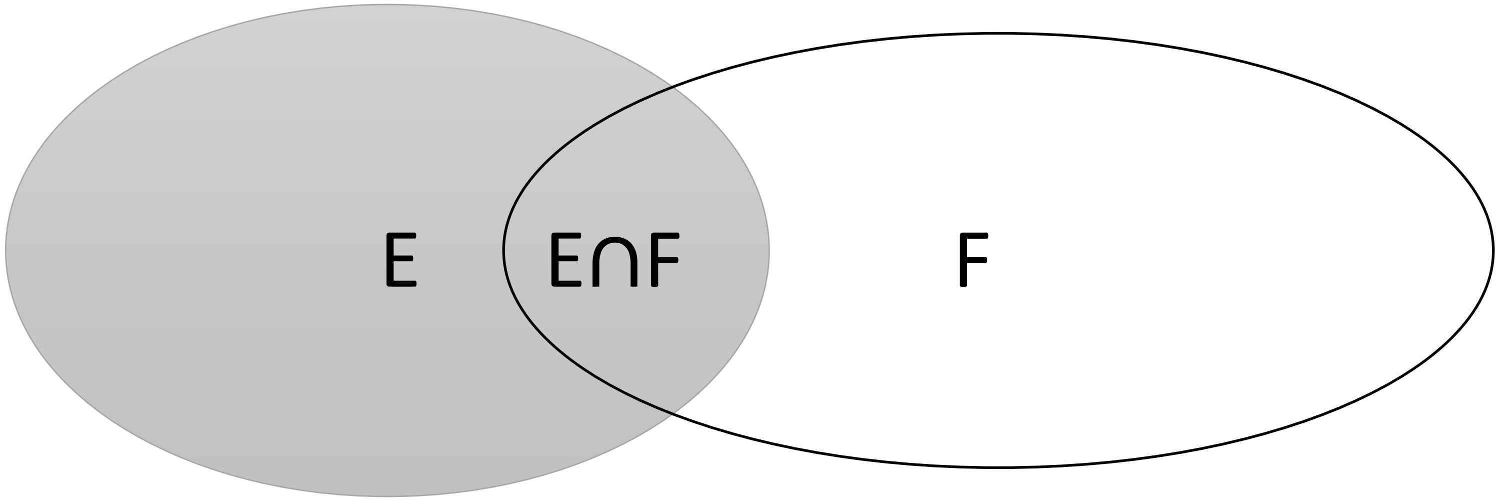
Example. Toss a coin (1 time). There are outcomes: H and T, where H is the head of a coin and T is the tail of a coin.

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

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

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 $X: S\to\{1,-1\}$ 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

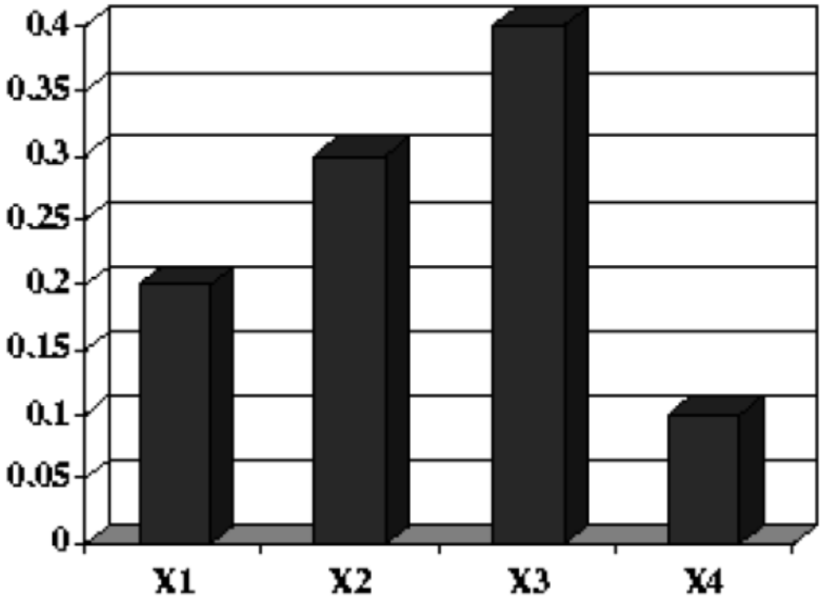
- A random variable is a variable that can take on different values randomly. We typically denote the random variable itself with an uppercase letter in plain typeface, and the values it can take on with lowercase letters.

- 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.



Credits from Dr. Liu Yang

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

A random variable X can be regarded a function:

the domain is the sample space S; but the range is discrete value:

- The range could be finite: $x_1, x_2,\ldots,x_n$;
- The range could be countably infinite: $x_1, x_2, \ldots, x_n,\ldots$

Discrete Variables and PMF: Examples

Discrete Random Variable with finite range:

Toss a coin (1 time).

In the sample space $S=\{H, T\}$, we design a random variable $X: S \to \{1,-1\}$ such that $X(H)=1$ and $X(T)=-1$. Then X is a random variable with finite range.

The probability is $P(X=1)=P(X=-1)= 0.5$.

Discrete Variables and PMF: Examples

Discrete Random Variable with infinite range:

Toss a coin (countably infinity times).

We design a random variable X: X = n means that
the first head appears after throwing n times.

Then X is a random variable with countably infinite range.

The probability is P(X=n) = 0.5^{n}.

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

- $P(X = x_i) = \frac{1}{n}$; $\sum_i P(X = x_i) = \sum_i \frac{1}{n} = \frac{n}{n} = 1$

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 (概率密度函数) $p_{\mathrm{X}}(\mathrm{x})$ such that

1) $p_{\mathrm{X}}(\mathrm{x}) \geq 0$;
2) P(a≤X≤b) = $\int_{a}^{b} p_{\mathrm{X}}(\mathrm{x})\, d\mathrm{x}$;
3) $\int_{-\infty}^{+\infty} p_{\mathrm{X}}(\mathrm{x})\, d\mathrm{x}$=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.

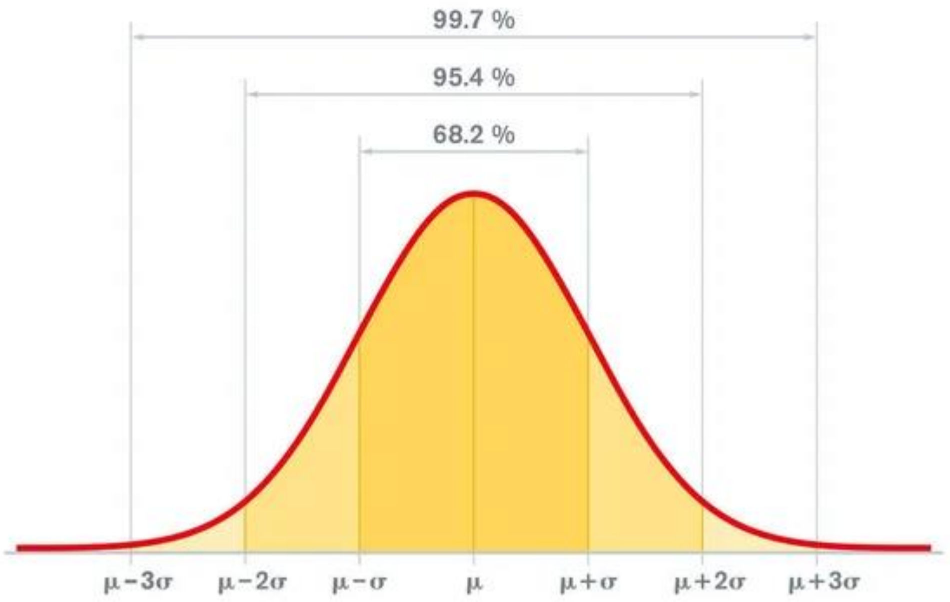
This is a continuous random variable because it can take on an infinite number of values. For example, a dog might weigh 30.333 pounds, 50.340999 pounds, 60.5 pounds, etc.

What is the distribution of dog's weight?

Continuous Variables and PDF

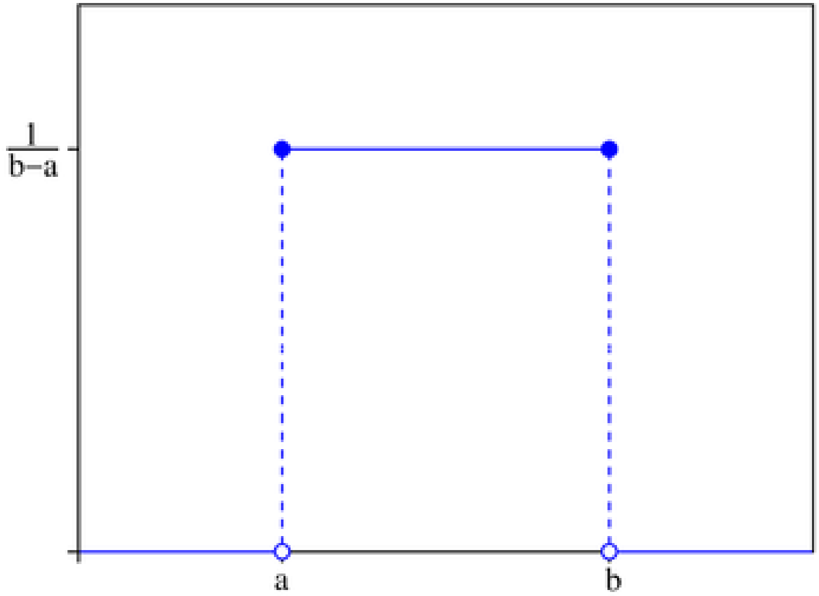
It is similar to a Gaussian distribution. The probability density funtion is

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



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 $p(\mathrm{x}; a, b) = \frac{1}{b-a}$



Exercises

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.

Exercises

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

Exercises

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

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

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

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

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

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

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

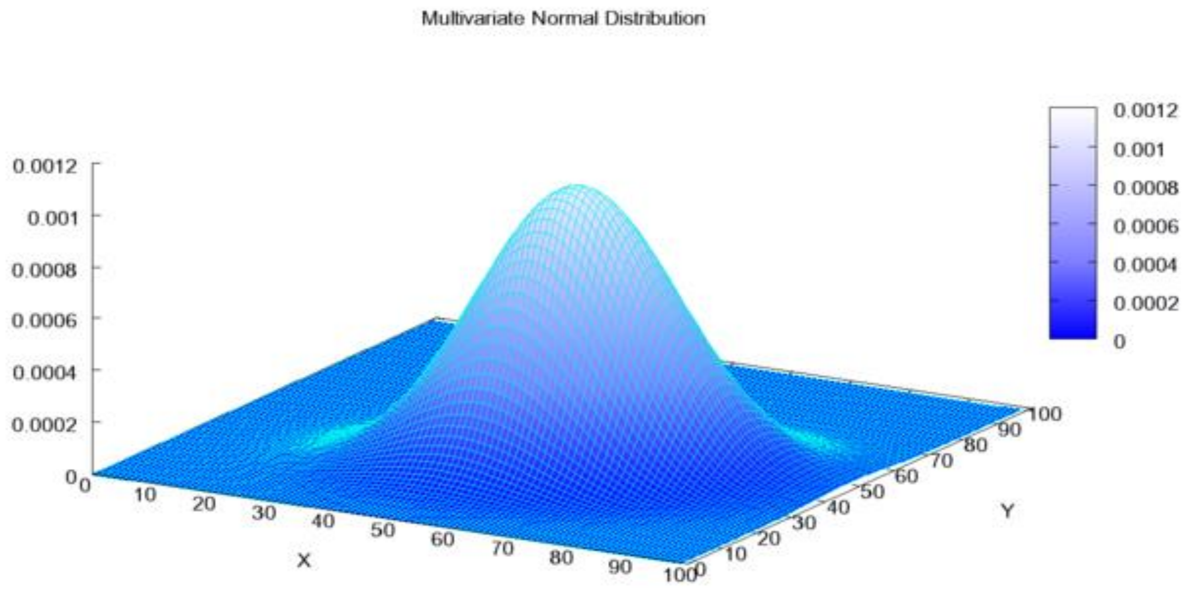
- In fact, when X, Y are continuous random variables, there exists a probability density function for the joint distribution:

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

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

Joint Distribution

- If X represnets 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.



Joint Probability

- If random variable X is continuous random variable, and Y is discrete random variable, then the joint distribution of X and Y is

$$P(a1 \leq X \leq b1,\ Y=y)$$

In fact, when X, is continuous random variable, there exists a probability density function for the joint distribution:

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

where $p_{\mathrm{XY}}(x,y)$ is continuous with respect to x, but discrete with respect to y.

Marginal Probability

- Using joint distrbution, 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 Probability

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

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

and the density function of X is

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

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

Marginal Probability

- If random variable X is continuous random variable, and Y is a discrete random variable, then the marginal distribution with respect to X is

$$P(a \leq X \leq b) = \sum_y \int_a^b p_{XY}(x, y)dx$$

and the density function of X is

$$p(x) = \sum_y p_{XY}(x, y)$$

The marginal distribution with respect to Y is

$$P(Y=y) = \int_{-\infty}^{+\infty} p_{XY}(x, y)dx$$

Joint Probability and Marginal Probability : An Example

- Joint probabilities can be between 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

Credits from Dr. Liu Yang

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

- Try
  - $P(Y=1)$

Credits from Dr. Liu Yang

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

Credits from Dr. Liu Yang

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(X=0 | Y=1) = P(X = 0, Y = 1) / P(Y = 1) = 0.1/0.3= 1/3

P(Y=1 | X=1) = P(X = 1, Y = 1) / P(X = 1) = 0.2/0.6= 1/3

P(Y=0 | X=1) = P(X = 1, Y = 0) / P(X = 1) = 0.4/0.6= 2/3

P(Z=1 | X=1) = P(X = 1, Z = 1) / P(X = 1) = 0.25/0.6= 5/12

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

Try:

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

Credits from Dr. Liu Yang

Condition Distribution

- Solution

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

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

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

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

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

Credits from Dr. Liu Yang

Conditional Distribution

- If X and Y are continuous random variables, then

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

So as a1 and b1 get close to x, we obtain that

If $p_X(x) > 0$, then

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

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

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

Chain Rule of Conditional Probabilities

- Given n events E_1, E_2,…,E_n

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

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

  ………..

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

Above equation is called *chain rule*.

Chain Rule of Conditional Probabilities

- Chain rule with respect to random variables

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

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

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

Credits from Dr. Liu Yang

Chain Rule of Conditional Probabilities

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

$P(X=0) = 0.4;$

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

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

Credits from Dr. Liu Yang

Chain Rule of Conditional Probabilities

Exercise: 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

Credits from Dr. Liu Yang

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

Credits from Dr. Liu Yang

Thank You!