EBU5601
Data Design

Probability Fundamentals

Dr Chao Shu, Dr Xiaolan Liu
School of Electronic Engineering and Computer Science
Queen Mary University of London
Sep. 2024

Queen Mary
University of London

Learning Outcomes

- The main outcomes are:
  - [**LO4.1**] Describe and calculate the probability of one event or the probability of a composite event
  - [**LO4.2**] Understand discrete distribution and probability mass functions (PMF)
  - [**LO4.3**] Apply probability mass functions (PMF) to calculate probabilities of events for discrete distributions using SciPy
  - [**LO4.4**] Understand continuous distribution and probability density functions (PDF)
  - [**LO4.5**] Apply cumulative distribution function (CDF) to calculate probabilities of discrete and continuous distributions using SciPy

Probability

- **Probability** is the likelihood of an event occurring.

$$P(events) = \frac{Number\ of\ ways\ event\ can\ happen}{Total\ number\ of\ possible\ outcoms}$$

Example

- When flip a (fair) coin, the probability of getting heads.

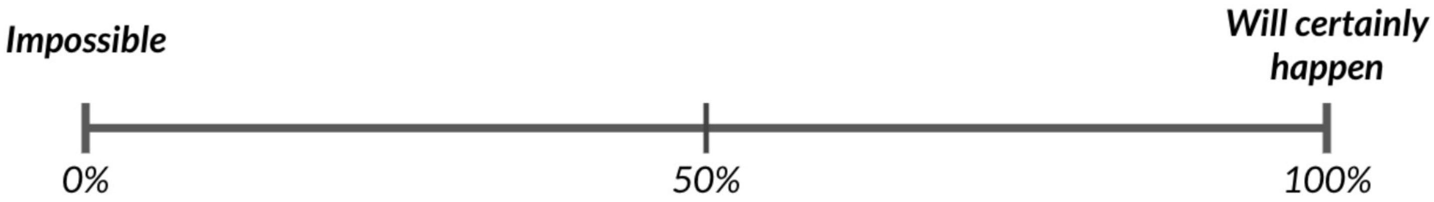
$$P(heads) = \frac{1\ way\ to\ get\ heads}{2\ possible\ outcomes} = 0.5\ or\ 50\%$$

- When roll a (fair) die, the probability of getting a “4”.

$$P(\text{"4"}) = \frac{1\ way\ to\ get\ a\ \text{“4”}}{6\ possible\ outcomes} = \frac{1}{6}$$

Rules

- Probability is always between 0 and 1.



- Complementary Outcomes

$$P(\bar{A}) = 1 - P(A)$$

Example

- The probability of Heads $P(H)=0.6$, what is the probability of getting tails.

$$P(T) = 1 - P(H) = 1 - 0.6 = 0.4$$

Independent & Dependent Events

- **Independent Events**
  - Two events are **independent** if the probability of the second event **isn't** affected by the outcome of the first event

Example

- If we flip a coin three times and it comes up “Heads” each time, what is the chance that the next flip will also be a "Head"?

The chance is still 50%, just like any other flips of the coin. What it did in the past will not affect the current flip.

Independent & Dependent Events

- **Dependent Events**
  - Two events are **dependent** if the probability of the second event **is** affected by the outcome of the first event

Example

- If we draw 2 cards from a deck, what are the chance of getting a King in the first and second draw?

For the 1st card the chance of drawing a King is 4 out of 52

But for the 2nd card:

- If the 1st card was a King, then the chance that the 2nd card to be a King is 3 of the 51.
- If the 1st card was not a King, then the chance that the 2nd card to be a King is 4 of the 51.

Rules

- If our events are independent, then the probability of the string of possible events is the product of those events.

Example

- We have a biased coin with $P(H)=0.6$, if we flip two times, what is the chance of getting two heads in a row?

$$P(H,H) = P(H) * P(H) = 0.6 * 0.6 = 0.36$$

- If we roll a fair die twice, what is the probability that we would get doubles or the same number on both rolls?

$$\mathrm{P} = \frac{1}{6} * \frac{1}{6} * 6 = \frac{1}{6}$$

Probability

Example:

- If we Flip a fair coin twice. What is the probability that we will get exactly one Head?

Answer:

We can simply use a truth table to list all possibilities

Flip 1	Flip 2	Probability of Each Outcome
H	H	0.25
H	T	0.25
T	H	0.25
T	T	0.25

$$P(\#\ Heads\ =\ 1) =\ 0.25 + 0.25 = 0.5\ =\ 2 * 0.5^1 * (1 - 0.5)^{2-1}$$

Probability

Example:

- If we Flip a fair coin three times. What is the probability that we will get exactly one Head?

Answer:

We can still use a truth table to list all possibilities. (we'll leave it to you to work out the table)

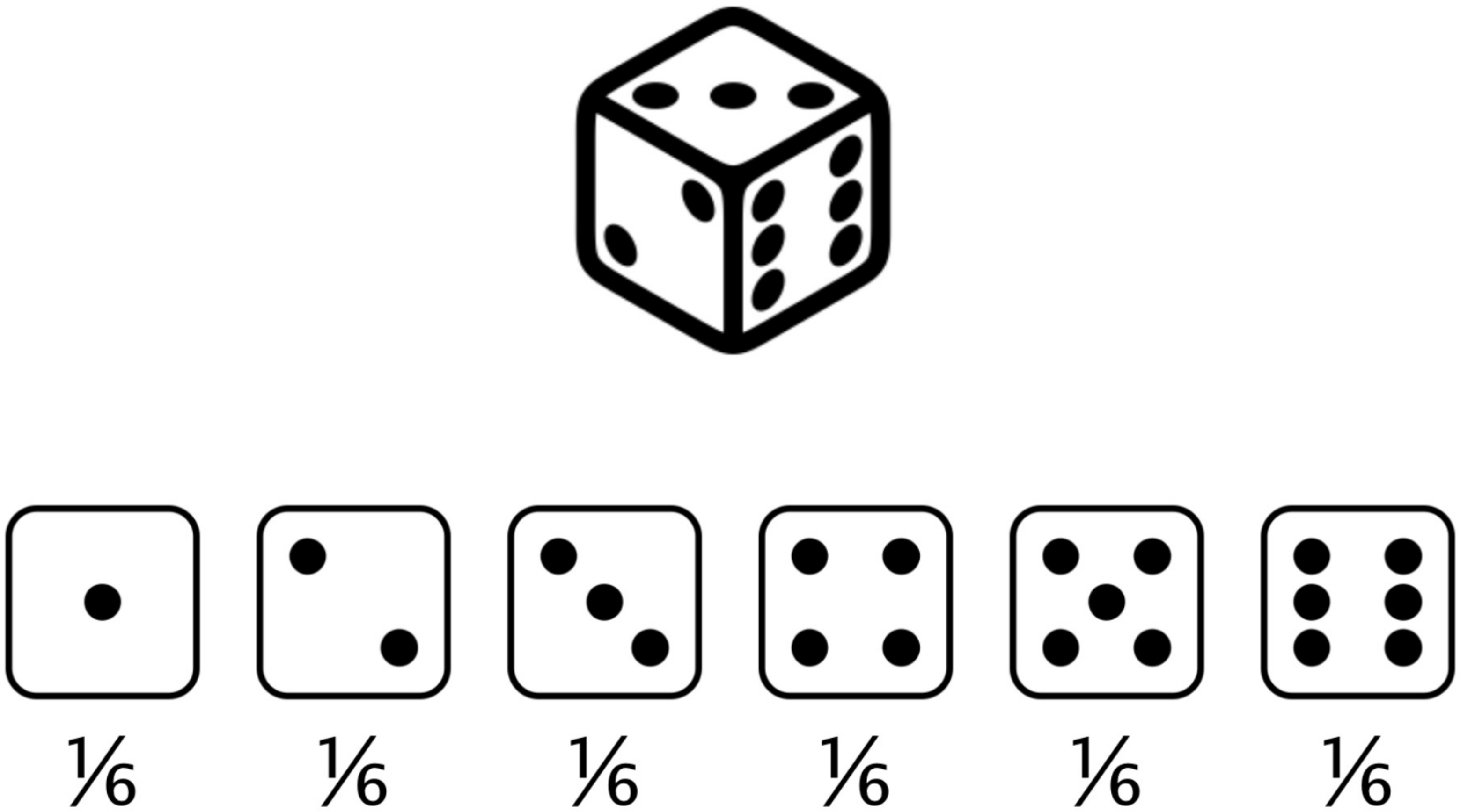
Because there are eight possibilities and they each could occur with the same probability which is ½ * ½ * ½ = 1/8 or 0.125.

Further, there are 3 cases have the possibility of having only one Heads: {H, T, T}, {T, H, T}, {T, T, H}

$$P(\#\,Heads\ =\ 1) = 3 * 0.125 = 0.375\ = 3 * 0.5^{1} * (1 - 0.5)^{3-1}$$

Discrete Distributions

- A **discrete distribution** is a probability distribution that depicts the occurrence of discrete (individually countable) outcomes.

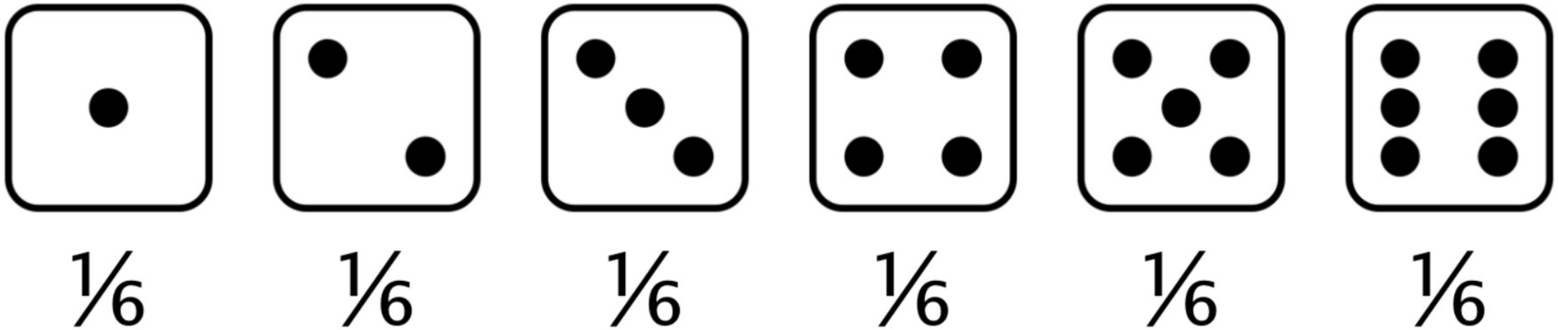


Discrete Distributions

- Die-rolling outcomes can be considered to be a random variable X that follows a **discrete (uniform) distribution**.
- Expect value: mean of a probability distribution

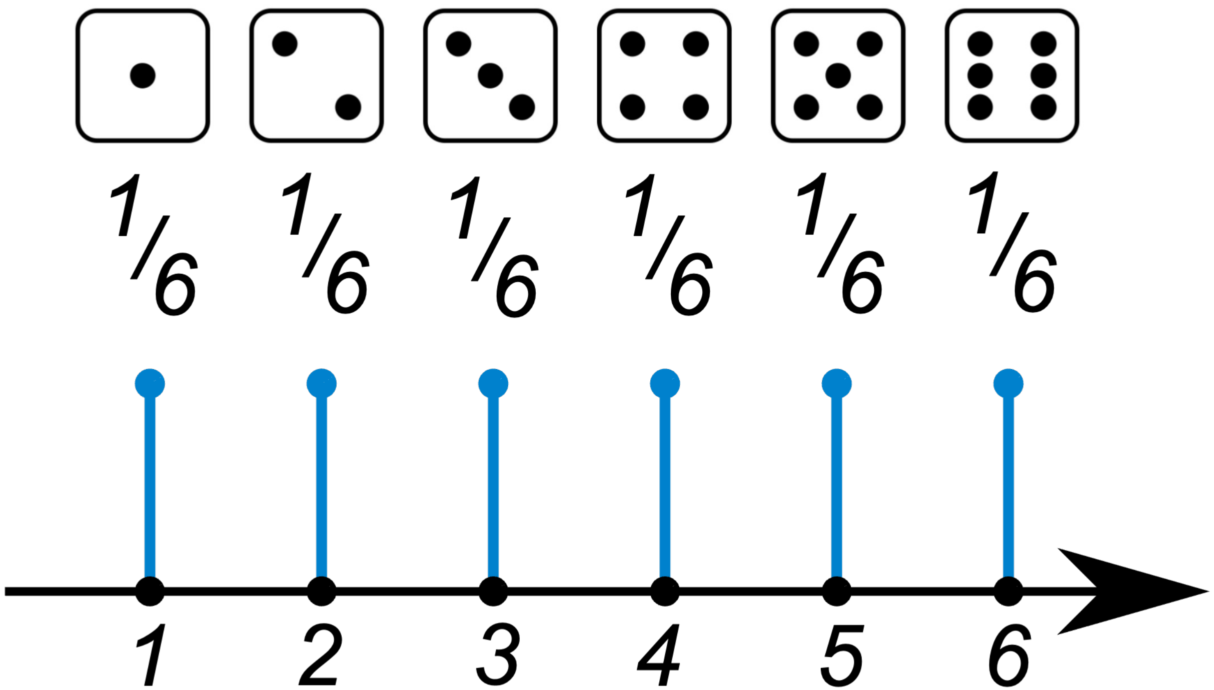
$$E(X) = \sum xP(x)$$

Expected value of a fair die roll =
(1 × 1/6) + (2 × 1/6) + (3 × 1/6) + (4 × 1/6) + (5 × 1/6) + (6 × 1/6) = 3.5



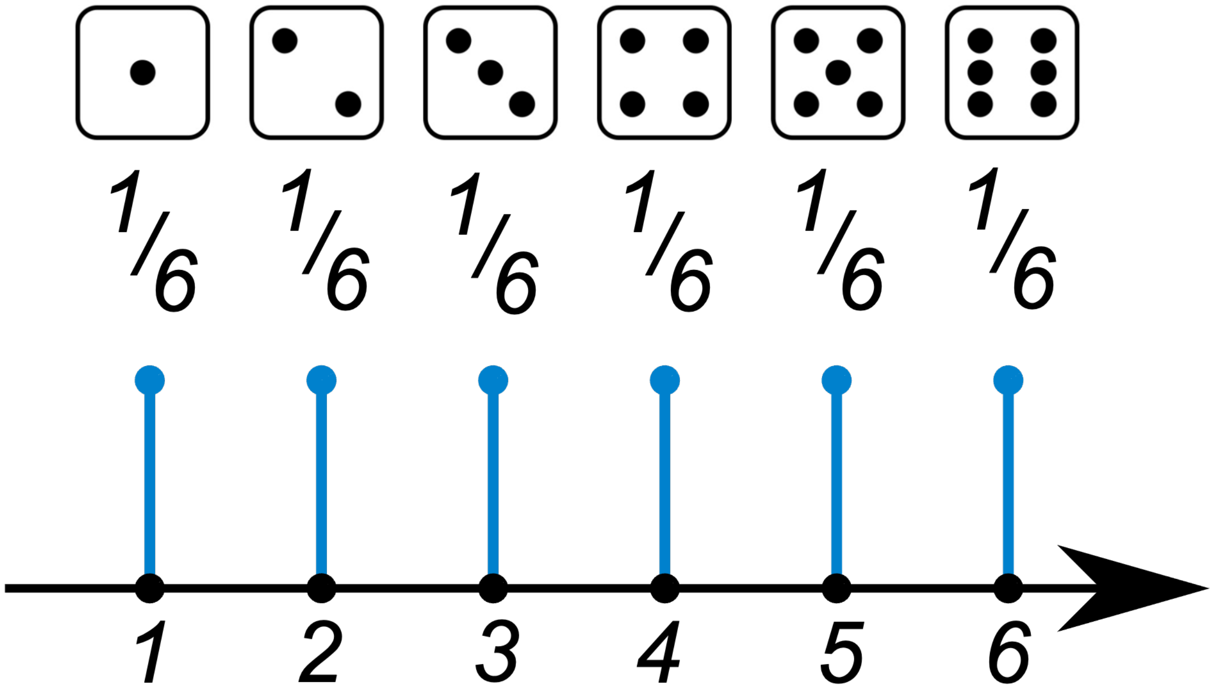
Discrete Distributions

- A **Probability Mass Function (PMF)** is a function that gives the probability that a discrete random variable is exactly equal to some value.



Discrete Distributions

- Probabilities of discrete distributions.
  - Calculate Specific Probabilities: To calculate the probability that X equals a specific value, plug that value into the PMF formula, e.g., $P(X=2)=1/6$
  - Sum or Calculate Cumulative Probabilities: To find cumulative probabilities $(e.g., P(X \leq x))$, sum the probabilities of all values less than or equal to x using the PMF.
  - For example, for a fair die, $P(X \leq 2) = P(X = 1) + P(X = 2) = 1/3$

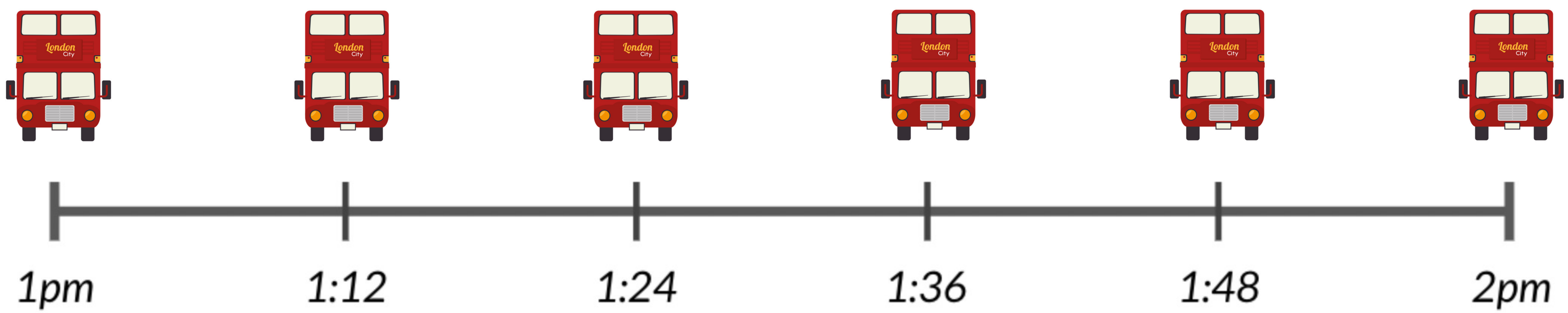


Continuous Distributions

- A **continuous distribution** describes an experiment where there is an arbitrary outcome that lies between certain bounds. The bounds are defined by the parameters, e.g. a and b, which are the minimum and maximum values.

Example: Bus Waiting Time

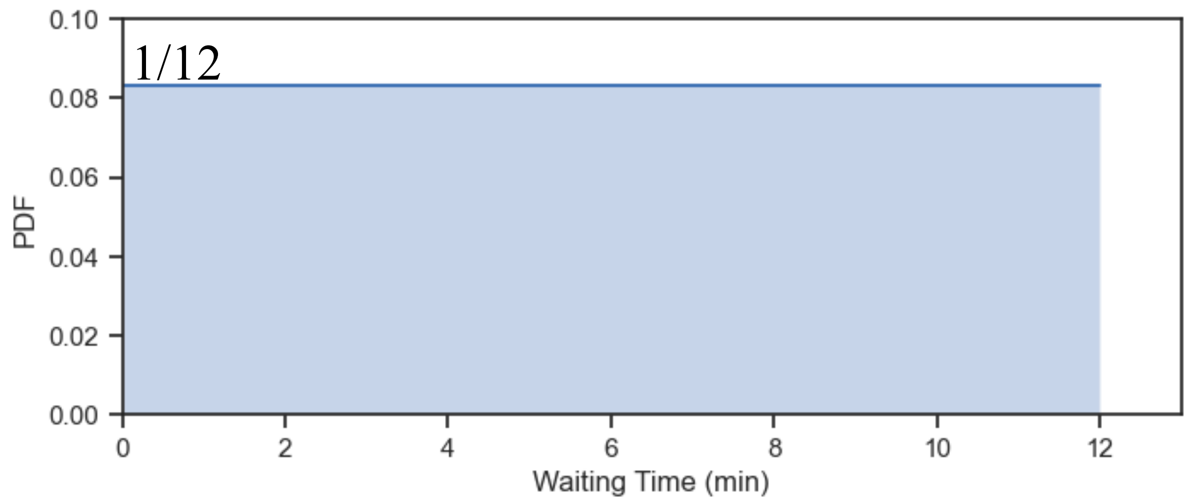
- Suppose a London bus service arrives at one of its stops every 12 minutes. If you show up at the stop at a random time, you could wait for anytime between 0 min (if you just arrive when the bus pulls in) to 12 min if you just arrive when the bus leaves)



Continuous Distributions

- Probabilities of continuous random variables (X) are defined as the area under the curve of its **Probability Density Function** (**PDF**).
- The probability density function is nonnegative everywhere, and the area under the entire curve is equal to 1. The probability that a continuous random variable equals some value is always zero.

Example: Bus Waiting Time PDF

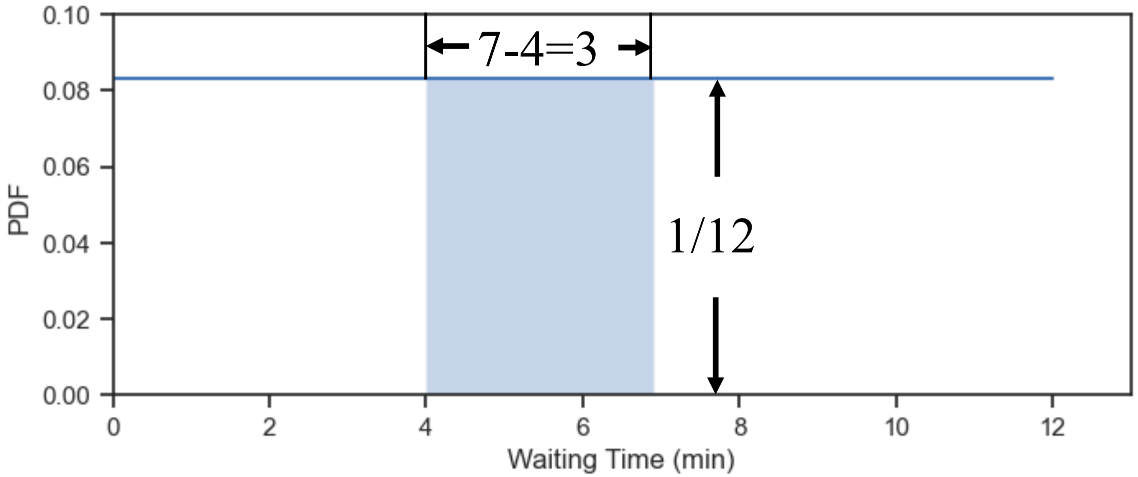


Continuous Distributions

Example: Bus Waiting Time Probability

- Suppose a London bus service arrives at one of its stops every 12 minutes. If you show up at the stop at a random time, you could wait for anytime between 0 min (if you just arrive when the bus pulls in) to 12 min if you just arrive when the bus leaves).
- What is the probability of wait time between 4 min and 7 min?

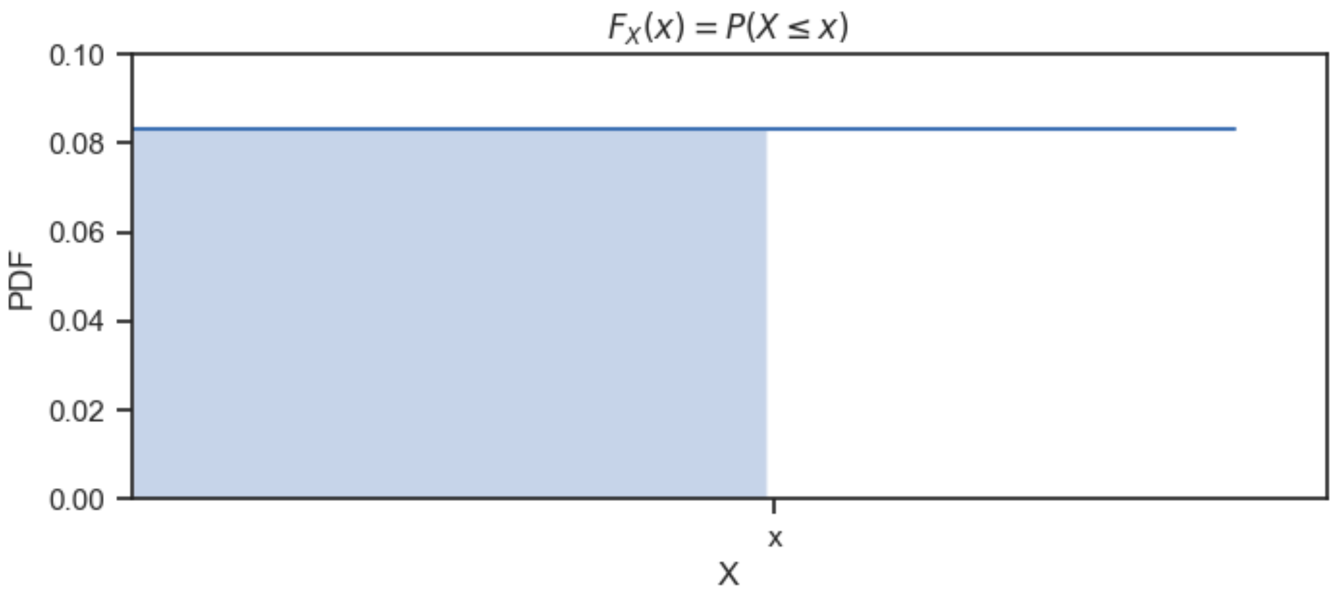
$$P(4 \leq X \leq 7) = (7-4) * \frac{1}{12} = 0.25$$



Continuous Distributions

- The **cumulative distribution function** (**CDF**) of a real-valued random variable X, evaluated at x, is the probability that X will take a value less than or equal to x. $F_X(\mathrm{x}) = \mathrm{P}(\mathrm{X} \leq \mathrm{x})$
- For a scalar continuous distribution, it gives the area under the probability density function from minus infinity to x.

Example: Bus Waiting Time CDF



Continuous Distributions

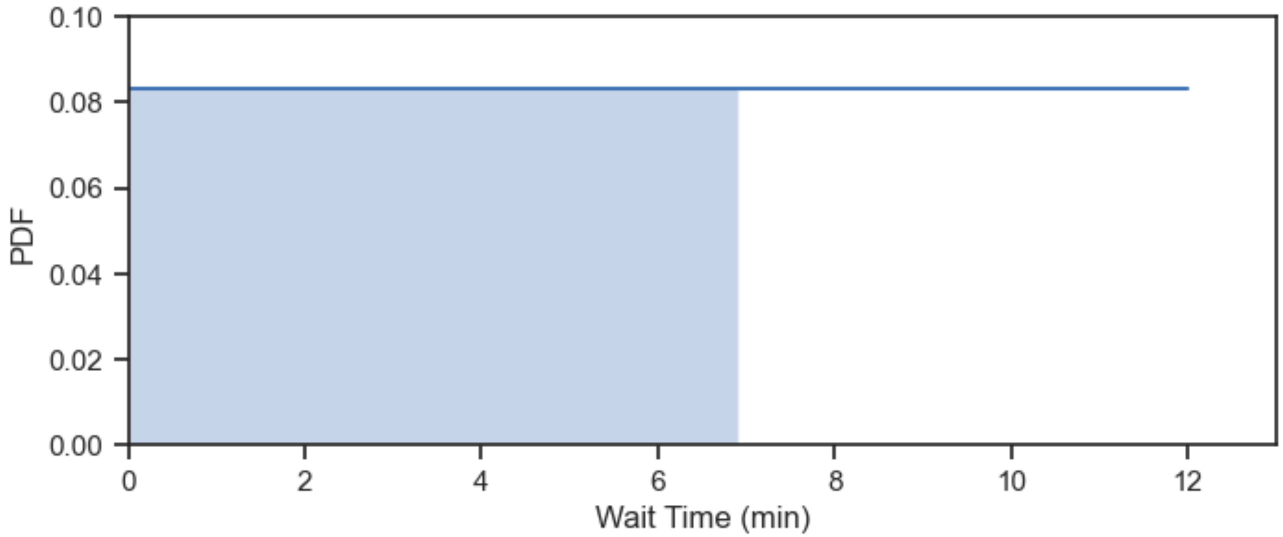
Demo: Bus Waiting Time Probabilities using Python

1. What is the probability of waiting time less than or equal to 7 min?

$$P(X \leq 7) = 7/12 = 0.583$$

```
1  stats.uniform.cdf(7, 0, 12)
✓ 0.0s

0.5833333333333334
```



Continuous Distributions

Demo: Bus Waiting Time Probabilities using Python

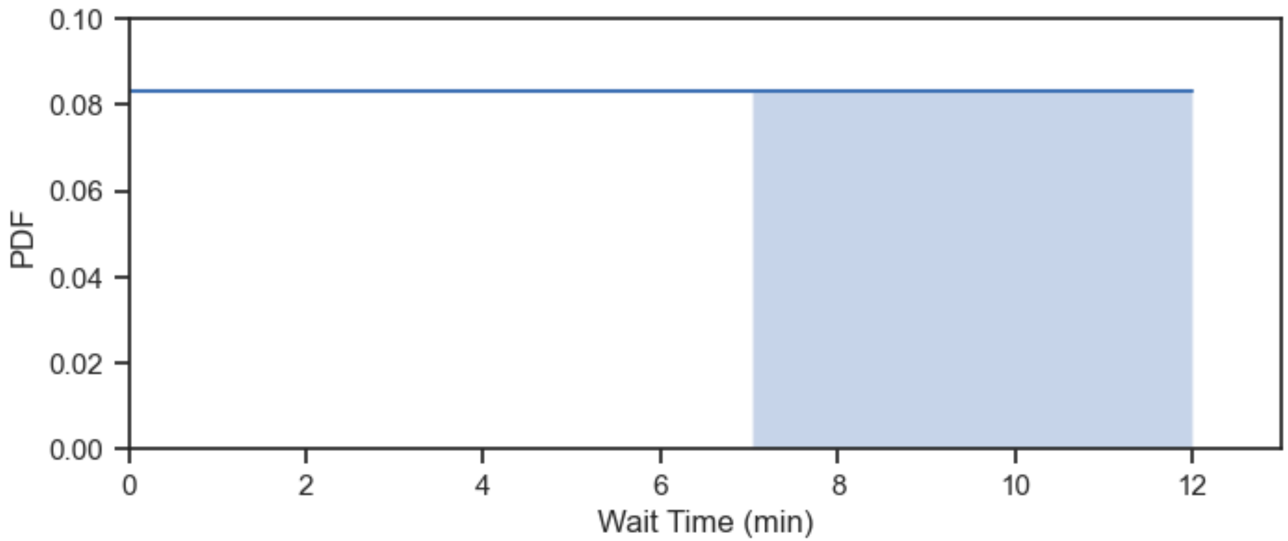
2. What is the probability of waiting time larger than or equal to 7 min?

$$P(X \geq 7) = 1 - P(X \leq 7) = 1 - \frac{7}{12} = \frac{5}{12} = 0.417$$

```
1 - stats.uniform.cdf(7, 0, 12)
```

✓ 0.0s

```
0.41666666666666663
```



Continuous Distributions

Demo: Bus Waiting Time Probabilities using Python

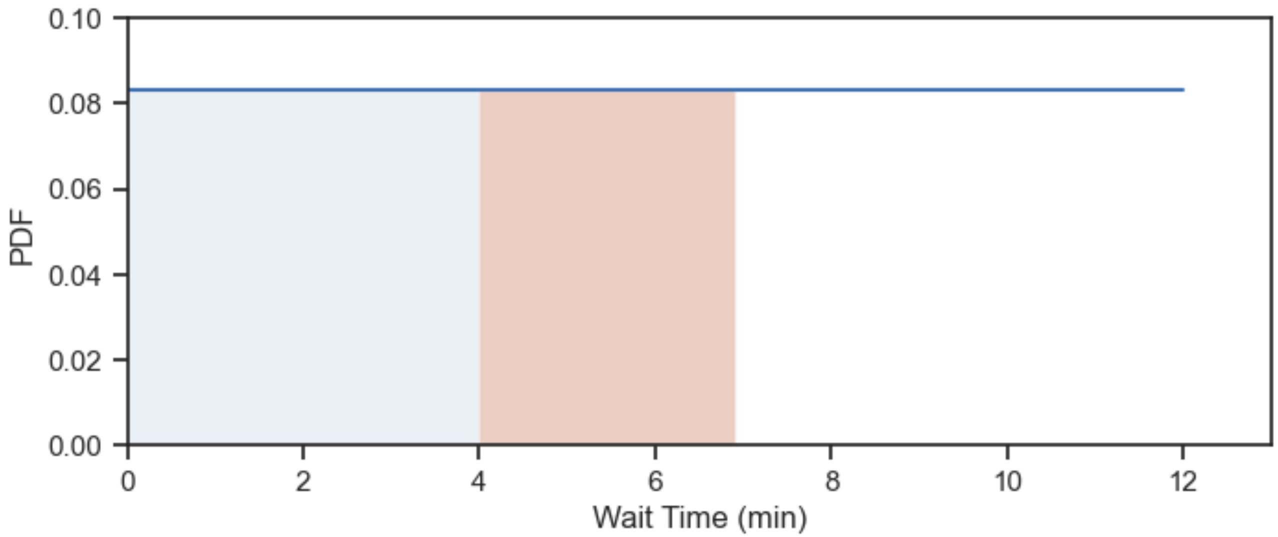
3. What is the probability of wait time between 4 min and 7 min?

$$P(4 \leq X \leq 7) = P(X \leq 7) - P(X \leq 4) = \frac{7}{12} - \frac{4}{12} = 0.25$$

```
stats.uniform.cdf(7, 0, 12) - stats.uniform.cdf(4, 0, 12)
```

✓ 0.0s

```
0.25000000000000006
```



RECAP

Probability Fundamentals

Recap

- **Probability**
  - Basic properties
  - Complementary outcomes
  - Independent and dependent events
- **Discrete Distribution**
  - Discrete normal distribution
  - PMF
- **Continuous Distribution**
  - Continuous normal distribution
  - PDF
  - CDF

Questions

Use student forum on QM+

chao.shu@qmul.ac.uk
xiaolanliu@qmul.ac.uk