Lecture 2: Probability Distribution

Statistical Methods for Data Science

Yinan Yu

Department of Computer Science and Engineering

November 4, 2021

Today

1. Probability distribution
   - Why probability distributions?
   - Terminology
   - Some probability distributions that you should know by heart
2. Demo
3. Summary

Learning outcome

- Be able to explain the following concepts: experiment, event, random variable, probability distribution
- Given the PDF/PMF, be able to describe the probability distribution of a continuous/discrete random variable
- Be able to compute conditional probability
- Understand Gaussian distribution and Bernoulli distribution: 1) PDF/PMF; 2) what are the parameters? 3) what happens to the shape of the distribution if we change the parameters?
- Be able to apply the learning routine to study a new probability distribution yourself

Today

1. Probability distribution
   - Why probability distributions?
   - Terminology
   - Some probability distributions that you should know by heart
2. Demo
3. Summary

Why probability distributions?

Histogram vs probability distribution

You need to **estimate** the weight **distribution** of your **1000** ducks without weighing all of them

Histogram vs probability distribution

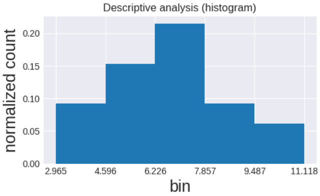
You need to **estimate** the weight **distribution** of your **1000** ducks without weighing all of them, because, well, data collection is expensive.

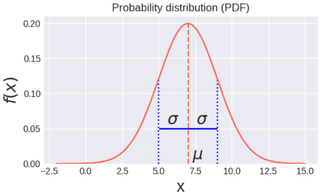
Histogram vs probability distribution

You need to **estimate** the weight **distribution** of your **1000** ducks without weighing all of them, because, well, data collection is expensive. You weighed **20** ducks and you plotted the histogram of the weights.

Histogram vs probability distribution

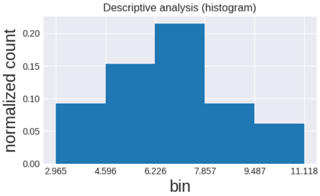
You need to **estimate** the weight **distribution** of your **1000** ducks without weighing all of them, because, well, data collection is expensive. You weighed **20** ducks and you plotted the histogram of the weights. Your best friend Jack looked at the histogram and suggested that you should use a **Gaussian distribution** to make a better **estimation** of the **distribution**.

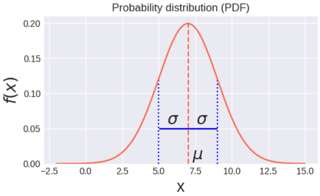




Histogram vs probability distribution

You need to **estimate** the weight **distribution** of your **1000** ducks without weighing all of them, because, well, data collection is expensive. You weighed **20** ducks and you plotted the histogram of the weights. Your best friend Jack looked at the histogram and suggested that you should use a **Gaussian distribution** to make a better **estimation** of the **distribution**.





Histogram vs probability distribution

- Question 1: Why can't I just use **descriptive analysis**, like the **histogram**, to describe the data distribution? Why should I use **probability distributions**?

Histogram vs probability distribution

- Question 1: Why can't I just use **descriptive analysis**, like the **histogram**, to describe the data distribution? Why should I use **probability distributions**?

  To address this question, let's describe the data distribution using a **histogram** and a **Gaussian distribution** to see the difference.

Histogram vs probability distribution

Here are the weights of the 20 ducks in kg

duck id	1	2	3	4	$\cdots$	19	20
weight	6.98	5.43	2.97	7.07	$\cdots$	4.63	7.27

Let's try to describe these ducks using a histogram with 5 bins.

Histogram vs probability distribution

Here are the weights of the 20 ducks in kg

duck id	1	2	3	4	$\cdots$	19	20
weight	6.98	5.43	2.97	7.07	$\cdots$	4.63	7.27

Let's try to describe these ducks using a histogram with 5 bins.

- Divide the range of the data into 5 bins

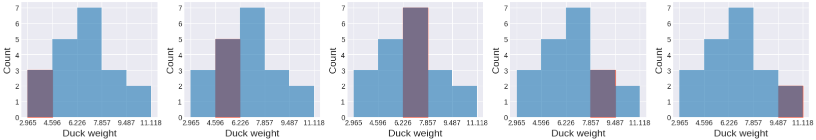
Histogram vs probability distribution

Here are the weights of the 20 ducks in kg

duck id	1	2	3	4	$\cdots$	19	20
weight	6.98	5.43	2.97	7.07	$\cdots$	4.63	7.27

Let's try to describe these ducks using a histogram with 5 bins.

- Divide the range of the data into 5 bins
- There are 3 ducks within the first bin $[2.965, 4.596]$; there are 5 ducks within second bin $[4.596, 6.226]$, etc.



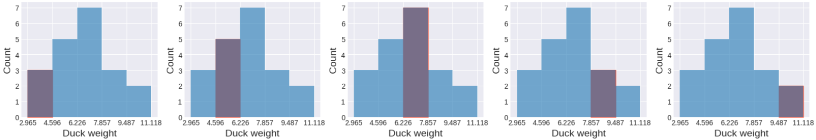
Histogram vs probability distribution

Here are the weights of the 20 ducks in kg

duck id	1	2	3	4	⋯	19	20
weight	6.98	5.43	2.97	7.07	⋯	4.63	7.27

Let's try to describe these ducks using a histogram with 5 bins.

- Divide the range of the data into 5 bins
- There are 3 ducks within the first bin $[2.965, 4.596]$; there are 5 ducks within second bin $[4.596, 6.226]$, etc.



- Now you want to use the histogram to describe the distribution of 1000 ducks

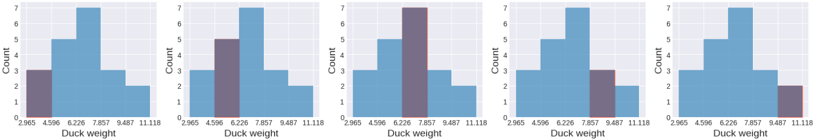
Histogram vs probability distribution

Here are the weights of the 20 ducks in kg

duck id	1	2	3	4	$\cdots$	19	20
weight	6.98	5.43	2.97	7.07	$\cdots$	4.63	7.27

Let's try to describe these ducks using a histogram with 5 bins.

- Divide the range of the data into 5 bins
- There are 3 ducks within the first bin $[2.965, 4.596]$; there are 5 ducks within second bin $[4.596, 6.226]$, etc.



- Now you want to use the histogram to describe the distribution of 1000 ducks
- What is the chance of a duck weighing between 2.965 kg and 4.596 kg?

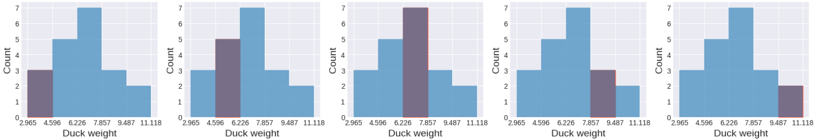
Histogram vs probability distribution

Here are the weights of the 20 ducks in kg

duck id	1	2	3	4	$\cdots$	19	20
weight	6.98	5.43	2.97	7.07	$\cdots$	4.63	7.27

Let's try to describe these ducks using a histogram with 5 bins.

- Divide the range of the data into 5 bins
- There are 3 ducks within the first bin $[2.965, 4.596]$; there are 5 ducks within second bin $[4.596, 6.226]$, etc.



- Now you want to use the histogram to describe the distribution of 1000 ducks
- What is the chance of a duck weighing between 2.965 kg and 4.596 kg? $\frac{3}{20}$

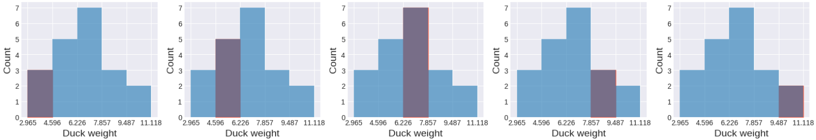
Histogram vs probability distribution

Here are the weights of the 20 ducks in kg

duck id	1	2	3	4	···	19	20
weight	6.98	5.43	2.97	7.07	···	4.63	7.27

Let's try to describe these ducks using a histogram with 5 bins.

- Divide the range of the data into 5 bins
- There are 3 ducks within the first bin $[2.965, 4.596]$; there are 5 ducks within second bin $[4.596, 6.226]$, etc.



- Now you want to use the histogram to describe the distribution of 1000 ducks
- What is the chance of a duck weighing between 2.965 kg and 4.596 kg? $\frac{3}{20}$
  How about between 3.1 kg and 3.4 kg?

Histogram vs probability distribution

- Resolution: the number of bins per kilogram

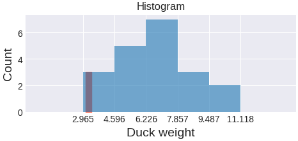
$$\frac{\text{number of bins}}{\textit{range}} = \frac{\text{number of bins}}{\max(\textit{weights}) - \min(\textit{weights})} = \frac{5}{11.118 - 2.965} = 0.61$$

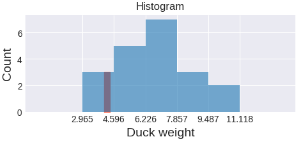
Histogram vs probability distribution

- Resolution: the number of bins per kilogram

$$\frac{\text{number of bins}}{\textit{range}} = \frac{\text{number of bins}}{\max(\textit{weights}) - \min(\textit{weights})} = \frac{5}{11.118 - 2.965} = 0.61$$

  - The chance of weighing between 3.1 and 3.4: $(3.4 - 3.1) \times \textit{resolution} \times \frac{3}{20} = 0.028$
  - The chance of weighing between 4.1 and 4.4: $(4.4 - 4.1) \times \textit{resolution} \times \frac{3}{20} = 0.028$



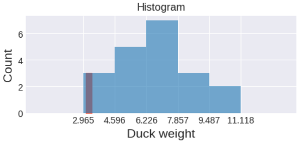


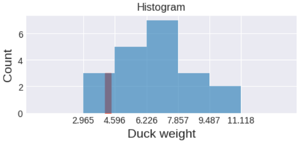
Histogram vs probability distribution

- Resolution: the number of bins per kilogram

$$\frac{\text{number of bins}}{range} = \frac{\text{number of bins}}{\max(weights) - \min(weights)} = \frac{5}{11.118 - 2.965} = 0.61$$

- The chance of weighing between 3.1 and 3.4: $(3.4 - 3.1) \times resolution \times \frac{3}{20} = 0.028$
- The chance of weighing between 4.1 and 4.4: $(4.4 - 4.1) \times resolution \times \frac{3}{20} = 0.028$





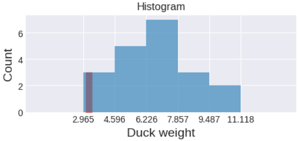
- We lose the resolution within each bin due to quantization

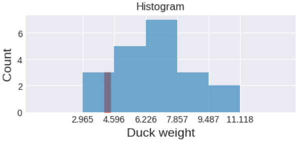
Histogram vs probability distribution

- Resolution: the number of bins per kilogram

$$\frac{\text{number of bins}}{\textit{range}} = \frac{\text{number of bins}}{\max(\textit{weights}) - \min(\textit{weights})} = \frac{5}{11.118 - 2.965} = 0.61$$

  - The chance of weighing between 3.1 and 3.4: $(3.4 - 3.1) \times \textit{resolution} \times \frac{3}{20} = 0.028$
  - The chance of weighing between 4.1 and 4.4: $(4.4 - 4.1) \times \textit{resolution} \times \frac{3}{20} = 0.028$





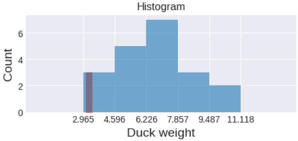
  - We lose the resolution within each bin due to quantization - you don't know what's going on within each bin

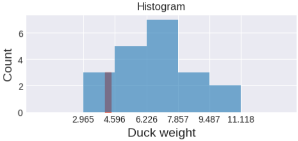
Histogram vs probability distribution

- Resolution: the number of bins per kilogram

$$\frac{\text{number of bins}}{\textit{range}} = \frac{\text{number of bins}}{\max(\textit{weights}) - \min(\textit{weights})} = \frac{5}{11.118 - 2.965} = 0.61$$

- The chance of weighing between 3.1 and 3.4: $(3.4 - 3.1) \times \textit{resolution} \times \frac{3}{20} = 0.028$
- The chance of weighing between 4.1 and 4.4: $(4.4 - 4.1) \times \textit{resolution} \times \frac{3}{20} = 0.028$





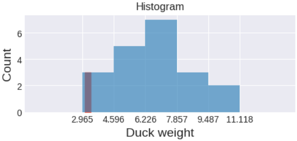
- We lose the resolution within each bin due to quantization - you don't know what's going on within each bin
- And if we increase the number of bins? 5 -> 200

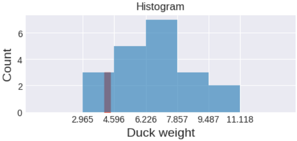
Histogram vs probability distribution

- Resolution: the number of bins per kilogram

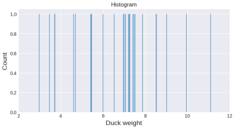
$$\frac{\text{number of bins}}{\textit{range}} = \frac{\text{number of bins}}{\max(\textit{weights}) - \min(\textit{weights})} = \frac{5}{11.118 - 2.965} = 0.61$$

- The chance of weighing between 3.1 and 3.4: $(3.4 - 3.1) \times \textit{resolution} \times \frac{3}{20} = 0.028$
- The chance of weighing between 4.1 and 4.4: $(4.4 - 4.1) \times \textit{resolution} \times \frac{3}{20} = 0.028$





- We lose the resolution within each bin due to quantization - you don't know what's going on within each bin
- And if we increase the number of bins? 5 -> 200



Histogram vs probability distribution

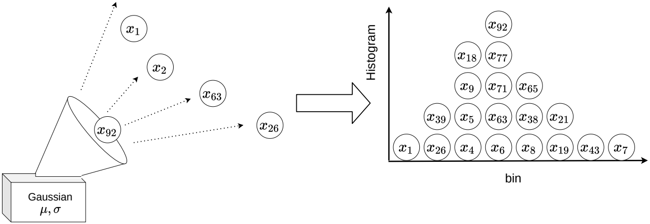
- Descriptive analysis (e.g. histogram) does not **generalize** well to unseen data

Histogram vs probability distribution

- Descriptive analysis (e.g. histogram) does not **generalize** well to unseen data
- Now let's try to use a Gaussian distribution to describe the data

Histogram vs probability distribution

- Descriptive analysis (e.g. histogram) does not **generalize** well to unseen data
- Now let's try to use a Gaussian distribution to describe the data
- First, we **assume** that data is **generated** from a Gaussian distribution



Histogram vs probability distribution

- A Gaussian distribution is described by a **function** that **looks similar** to this histogram

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

Histogram vs probability distribution

- A Gaussian distribution is described by a **function** that **looks similar** to this histogram

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

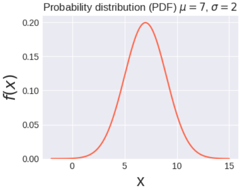
- This **function** is sufficiently defined by two **parameters** μ and σ.

Histogram vs probability distribution

- A Gaussian distribution is described by a **function** that **looks similar** to this histogram

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

- This **function** is sufficiently defined by two **parameters** μ and σ.
- The **shape** of the **function** (e.g. given μ=7 and σ=2):

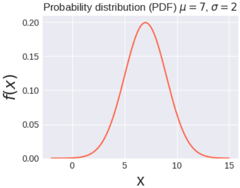


Histogram vs probability distribution

- A Gaussian distribution is described by a **function** that **looks similar** to this histogram

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

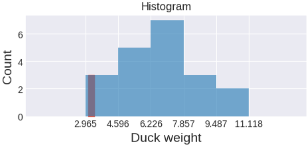
- This **function** is sufficiently defined by two **parameters** μ and σ.
- The **shape** of the **function** (e.g. given μ=7 and σ=2):

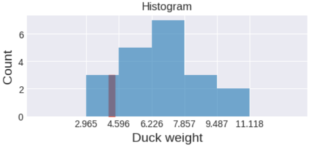


- We will try to use this **function** instead of the histogram to describe the data.

Histogram vs probability distribution

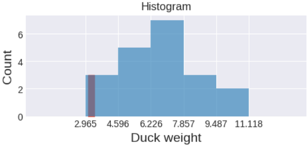
- Describe the distribution:
  - Histogram (using 0.61 bins to describe 1 kg):
    - The chance of $weight \in [3.1, 3.4]$: $(3.4 - 3.1) \times resolution \times \frac{3}{20} = 0.028$
    - The chance of $weight \in [4.1, 4.4]$: $(4.4 - 4.1) \times resolution \times \frac{3}{20} = 0.028$

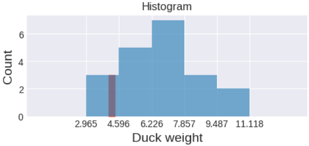




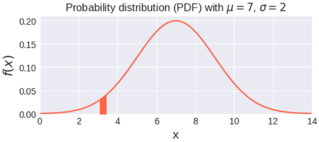
Histogram vs probability distribution

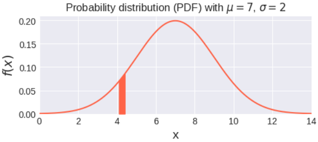
- Describe the distribution:
  - Histogram (using 0.61 bins to describe 1 kg):
    - The chance of *weight* $\in [3.1, 3.4]$: $(3.4 - 3.1) \times resolution \times \frac{3}{20} = 0.028$
    - The chance of *weight* $\in [4.1, 4.4]$: $(4.4 - 4.1) \times resolution \times \frac{3}{20} = 0.028$





  - Gaussian distribution (using **infinite** bins to describe 1 kg):
    - The chance of *weight* $\in [3.1, 3.4]$: $\int_{3.1}^{3.4} f(t)dt = 0.010$
    - The chance of *weight* $\in [4.1, 4.4]$: $\int_{4.1}^{4.4} f(t)dt = 0.023$





Histogram vs probability distribution

- Descriptive analysis: a *histogram* with M bins (e.g. $M = 5$)

Histogram vs probability distribution

- Descriptive analysis: a *histogram* with M bins (e.g. $M = 5$)
- Gaussian distribution: a mathematical function $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

Histogram vs probability distribution

- Descriptive analysis: a *histogram* with M bins (e.g. $M = 5$)
- Gaussian distribution: a mathematical function $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$
- Comparison

	Histogram		Gaussian distribution	
Representation	M values		mathematical function f	
Number of parameters	M	-	2 (μ and σ)	+
Resolution	$\frac{M}{\max(x)-\min(x)}$	-	infinity	+
Analytical properties	No	-	Yes	+
Assumptions	No	+	Yes	-
Can be directly computed from data	Yes	+	Parameters unknown	-

Histogram vs probability distribution

- Descriptive analysis: a *histogram* with M bins (e.g. $M = 5$)
- Gaussian distribution: a mathematical function $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$
- Comparison

	Histogram		Gaussian distribution	
Representation	M values		mathematical function f	
Number of parameters	M	-	2 (μ and σ)	+
Resolution	$\frac{M}{\max(x)-\min(x)}$	-	infinity	+
Analytical properties	No	-	Yes	+
Assumptions	No	+	Yes	-
Can be directly computed from data	Yes	+	Parameters unknown	-

- For your use case, you want to **estimate** the **distribution** of your 1000 ducks without weighing all of them. It is hard to do that from the histogram.

Histogram vs probability distribution

- Descriptive analysis: a *histogram* with M bins (e.g. $M = 5$)
- Gaussian distribution: a mathematical function $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$
- Comparison

	Histogram		Gaussian distribution	
Representation	M values		mathematical function f	
Number of parameters	M	-	2 (μ and σ)	+
Resolution	$\frac{M}{\max(x)-\min(x)}$	-	infinity	+
Analytical properties	No	-	Yes	+
Assumptions	No	+	Yes	-
Can be directly computed from data	Yes	+	Parameters unknown	-

- For your use case, you want to **estimate** the **distribution** of your 1000 ducks without weighing all of them. It is hard to do that from the histogram. The histogram **describes** the data you have seen

Histogram vs probability distribution

- Descriptive analysis: a *histogram* with M bins (e.g. $M = 5$)
- Gaussian distribution: a mathematical function $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$
- Comparison

	Histogram		Gaussian distribution	
Representation	M values		mathematical function f	
Number of parameters	M	-	2 (μ and σ)	+
Resolution	$\frac{M}{\max(x)-\min(x)}$	-	infinity	+
Analytical properties	No	-	Yes	+
Assumptions	No	+	Yes	-
Can be directly computed from data	Yes	+	Parameters unknown	-

- For your use case, you want to **estimate** the **distribution** of your 1000 ducks without weighing all of them. It is hard to do that from the histogram. The histogram **describes** the data you have seen, but it does not **generalize** well to unseen data.

Histogram vs probability distribution

- Descriptive analysis: a *histogram* with M bins (e.g. $M = 5$)
- Gaussian distribution: a mathematical function $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$
- Comparison

	Histogram		Gaussian distribution	
Representation	M values		mathematical function f	
Number of parameters	M	-	2 (μ and σ)	+
Resolution	$\frac{M}{\max(x)-\min(x)}$	-	infinity	+
Analytical properties	No	-	Yes	+
Assumptions	No	+	Yes	-
Can be directly computed from data	Yes	+	Parameters unknown	-

- For your use case, you want to **estimate** the **distribution** of your 1000 ducks without weighing all of them. It is hard to do that from the histogram. The histogram **describes** the data you have seen, but it does not **generalize** well to unseen data.
- Statistical modelling using probability distributions (e.g. a Gaussian distribution) can help you with that!

Histogram vs probability distribution

- Descriptive analysis: a *histogram* with M bins (e.g. $M = 5$)
- Gaussian distribution: a mathematical function $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$
- Comparison

	Histogram		Gaussian distribution	
Representation	M values		mathematical function f	
Number of parameters	M	-	2 (μ and σ)	+
Resolution	$\frac{M}{\max(x)-\min(x)}$	-	infinity	+
Analytical properties	No	-	Yes	+
Assumptions	No	+	Yes	-
Can be directly computed from data	Yes	+	Parameters unknown	-

- For your use case, you want to **estimate** the **distribution** of your 1000 ducks without weighing all of them. It is hard to do that from the histogram. The histogram **describes** the data you have seen, but it does not **generalize** well to unseen data.
- Statistical modelling using probability distributions (e.g. a Gaussian distribution) can help you with that!
- This concludes the question why we use probability distributions instead of histograms to describe your ducks.

Histogram vs probability distribution

- Descriptive analysis: a *histogram* with M bins (e.g. $M = 5$)
- Gaussian distribution: a mathematical function $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$
- Comparison

	Histogram		Gaussian distribution	
Representation	M values		mathematical function f	
Number of parameters	M	-	2 (μ and σ)	+
Resolution	$\frac{M}{\max(x)-\min(x)}$	-	infinity	+
Analytical properties	No	-	Yes	+
Assumptions	No	+	Yes	-
Can be directly computed from data	Yes	+	Parameters unknown	-

- For your use case, you want to **estimate** the **distribution** of your 1000 ducks without weighing all of them. It is hard to do that from the histogram. The histogram **describes** the data you have seen, but it does not **generalize** well to unseen data.
- Statistical modelling using probability distributions (e.g. a Gaussian distribution) can help you with that!
- This concludes the question why we use probability distributions instead of histograms to describe your ducks.

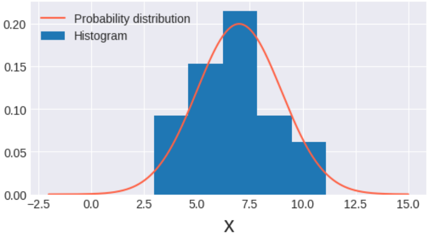
Disclaimer: we used a continuous distribution to illustrate the comparison between a histogram and a probability distribution. A discrete probability distribution differs from a continuous distribution.

Choosing a probability distribution

- Question 2: How do I know which probability distribution I should use to describe the data? How do I know that it should be a Gaussian distribution?

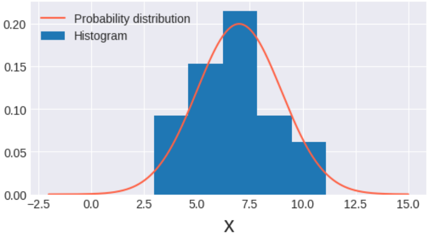
Choosing a probability distribution

- Question 2: How do I know which probability distribution I should use to describe the data? How do I know that it should be a Gaussian distribution?
  - Short answer: if the probability distribution looks like the histogram, then go for it!



Choosing a probability distribution

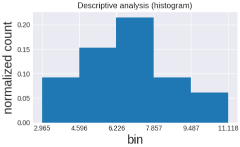
- Question 2: How do I know which probability distribution I should use to describe the data? How do I know that it should be a Gaussian distribution?
  - Short answer: if the probability distribution looks like the histogram, then go for it!

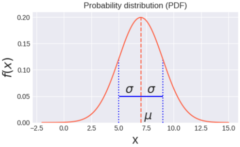


  - Long answer will be given in lecture 3.

Parameter estimation and evaluation

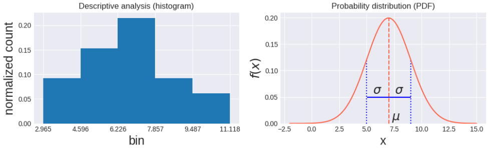
- Question 3: Okay fine, let's say we describe the data with a Gaussian distribution. How do I know what the parameters μ and σ are?





Parameter estimation and evaluation

- Question 3: Okay fine, let's say we describe the data with a Gaussian distribution. How do I know what the parameters μ and σ are?



  - This is done by **parameter estimation**. In lecture 3 & 4, we will talk about the **maximum likelihood estimation (MLE)** and the **maximum a posteriori estimation (MAP)**.

Terminology

Probability distribution

- **Experiment**: an action that leads to one outcome. For example, we weigh a duck and look at its weight. The outcome is weight = 2 kg.

Probability distribution

- **Experiment**: an action that leads to one outcome. For example, we weigh a duck and look at its weight. The outcome is weight = 2 kg.
- **Sample space**: the set of all possible outcomes from the experiment. The sample space of the previous example is any real value between 0 and ∞.

Probability distribution

- **Experiment**: an action that leads to one outcome. For example, we weigh a duck and look at its weight. The outcome is weight $= 2$ kg.
- **Sample space**: the set of all possible outcomes from the experiment. The sample space of the previous example is any real value between 0 and ∞.
- **Event**: a subset of the sample space, for example, a duck weighs between 5kg and 6kg.

Probability distribution

- **Experiment**: an action that leads to one outcome. For example, we weigh a duck and look at its weight. The outcome is weight $= 2$ kg.
- **Sample space**: the set of all possible outcomes from the experiment. The sample space of the previous example is any real value between 0 and ∞.
- **Event**: a subset of the sample space, for example, a duck weighs between 5kg and 6kg.
- **Probability distribution**: the probability of the occurrence of *any* event in the sample space, e.g. P(a duck weighs between a kg and b kg) for any $0 < a < b < \infty$ (not only for $a = 5$ and $b = 6$).

Probability distribution

- **Experiment**: an action that leads to one outcome. For example, we weigh a duck and look at its weight. The outcome is weight = 2 kg.
- **Sample space**: the set of all possible outcomes from the experiment. The sample space of the previous example is any real value between 0 and ∞.
- **Event**: a subset of the sample space, for example, a duck weighs between 5kg and 6kg.
- **Probability distribution**: the probability of the occurrence of *any* event in the sample space, e.g. P(a duck weighs between a kg and b kg) for any $0 < a < b < \infty$ (not only for $a = 5$ and $b = 6$).
- **Random variable X**:
  - Heuristically, X assigns a numerical value to each outcome of the experiment:

$$X : \text{weight} \rightarrow \mathbb{R}$$

  - X follows some underlying probability distribution.
  - **Discrete random variable** and **continuous random variable**: depends on the sample space of the experiment; the underlying distributions are called **discrete distribution** and **continuous distribution**, respectively. For example, weights are continuous so X from this example is a continuous random variable.

Probability distribution

- **Experiment**: an action that leads to one outcome. For example, we weigh a duck and look at its weight. The outcome is weight $= 2$ kg.
- **Sample space**: the set of all possible outcomes from the experiment. The sample space of the previous example is any real value between 0 and ∞.
- **Event**: a subset of the sample space, for example, a duck weighs between 5kg and 6kg.
- **Probability distribution**: the probability of the occurrence of *any* event in the sample space, e.g. P(a duck weighs between a kg and b kg) for any $0 < a < b < \infty$ (not only for $a = 5$ and $b = 6$).
- **Random variable** X:
  - Heuristically, X assigns a numerical value to each outcome of the experiment:
    $$X : \text{weight} \to \mathbb{R}$$
  - X follows some underlying probability distribution.
  - **Discrete random variable** and **continuous random variable**: depends on the sample space of the experiment; the underlying distributions are called **discrete distribution** and **continuous distribution**, respectively. For example, weights are continuous so X from this example is a continuous random variable.
- **Data** x: a value drawn from the **underlying distribution of** X.
  - We use a **capital letter** (e.g. X) to denote a random variable and the corresponding lower case letter (e.g. x) to denote the data generated from the underlying distribution of X.
  - Discrete random variable: categorical data or discrete numerical data
  - Continuous random variable: continuous numerical data

Probability distribution

A probability distribution describes the probabilities of occurrence of all possible events.

Probability distribution

A probability distribution describes the probabilities of occurrence of all possible events.

More precisely, a probability distribution is defined by a function f_X (also denoted as f if neglecting X does not cause confusion), where

Probability distribution

A probability distribution describes the probabilities of occurrence of all possible events.

More precisely, a probability distribution is defined by a function f_X (also denoted as f if neglecting X does not cause confusion), where

- for discrete distribution, the **probability mass function (PMF)** is used, where

$$f_X(x_i) = \boxed{P(X = x_i)}$$

where $0 \leq f_X(x_i) \leq 1$ for all x_i (discrete).

Probability distribution

A probability distribution describes the probabilities of occurrence of all possible events.

More precisely, a probability distribution is defined by a function f_X (also denoted as f if neglecting X does not cause confusion), where

- for discrete distribution, the **probability mass function (PMF)** is used, where

$$f_X(x_i) = \boxed{P(X = x_i)}$$

  where $0 \leq f_X(x_i) \leq 1$ for all x_i (discrete).
- for continuous distribution, the **probability density function (PDF)** is used, where

$$\boxed{P(a \leq X \leq b)} = \int_a^b f_X(x)dx, \ \forall a, b \in \mathbb{R}, a \leq b$$

  where $f_X(x) \geq 0$ for all x (continuous).

Probability distribution

A probability distribution describes the probabilities of occurrence of all possible events.
More precisely, a probability distribution is defined by a function f_X (also denoted as f if neglecting X does not cause confusion), where

- for discrete distribution, the **probability mass function (PMF)** is used, where

$$f_X(x_i) = \boxed{P(X = x_i)}$$

  where $0 \leq f_X(x_i) \leq 1$ for all x_i (discrete).
- for continuous distribution, the **probability density function (PDF)** is used, where

$$\boxed{P(a \leq X \leq b)} = \int_a^b f_X(x)dx, \ \forall a, b \in \mathbb{R}, a \leq b$$

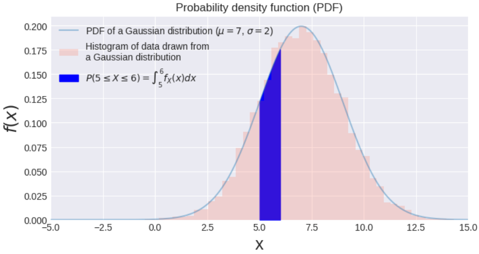
  where $f_X(x) \geq 0$ for all x (continuous).

$\boxed{P(\text{event})}$ is the probability of the **event** occurring.

Example: continuous random variables and PDF

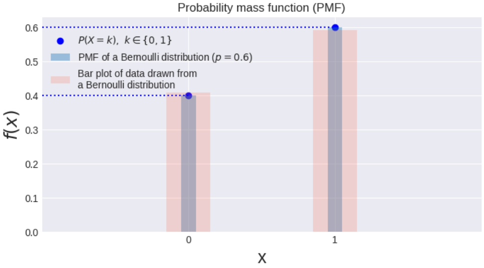
- **Experiment**: you weigh a duck and look at its weight
- **Sample space**: $0 < weight < \infty$
- **Random variable** X : $weight \to \mathbb{R}$
  - $X = x$ if the duck weighs x kg for $0 < x < \infty$
  - X follows a **Gaussian distribution** with parameters μ and σ; denoted as $X \sim \mathcal{N}(\mu, \sigma^2)$
- **PDF**: $f_X(x)$

$$P(a \le X \le b) = \underbrace{\int_a^b f_X(x)dx}_{\text{Integral = area under the PDF curve}} \qquad \forall a, b \in \mathbb{R}, a \le b$$



Example: discrete random variables and PMF

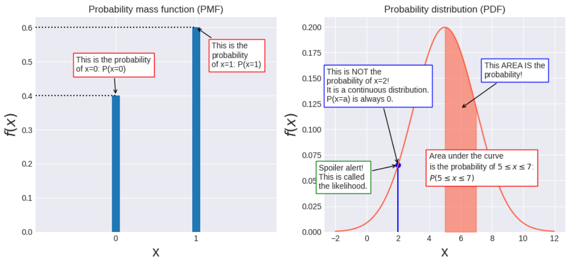
- **Experiment**: you measure the color of the duck.
- **Sample space**: the color can be only red or blue
- **Random variable** X : $color \to \mathbb{R}$
  - $X = \begin{cases} 0, & \text{if a duck is red} \\ 1, & \text{if a duck is blue} \end{cases}$
  - X follows a **Bernoulli distribution** with parameter p; denoted as $X \sim Bernoulli(p)$
- **PMF**: $f_X(x_i) = P(X = x_i)$



Probability distribution

Differences between PMF and PDF

- Discrete distribution $f_X(x_i) = P(X = x_i)$:
  - y-axis represents the probability itself
- Continuous distribution:
  - $P(a \leq X \leq b) = \int_a^b f_X(x)dx$: **y-axis $f(x)$** DOES NOT represent the probability itself.
  - For continuous distributions, **the probability at any given value is always 0**, i.e. $P(X = a) = P(a \leq X \leq a) = \int_a^a f_X(x)dx \equiv 0$. Example: what is the probability of a duck weighing exactly 4.32028374... kg?



Conditional probability

Given events A and B,

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

The probability of event A given even B.

Conditional probability example

Example:

- **Experiment**: You ask your ducks to stand in a row again and look at their colors and sizes.
- **Sample space**: The color can be either red or blue; the size can be either slim or chonker.
- **Data**:

duck id	1	2	3	4	5	6
color	red	red	blue	blue	blue	red
size	chonker	slim	slim	chonker	chonker	slim

- **Event**:
  - A: a duck is blue
  - B: a duck is a chonker

Conditional probability example

Example:

- **Experiment**: You ask your ducks to stand in a row again and look at their colors and sizes.
- **Sample space**: The color can be either red or blue; the size can be either slim or chonker.
- **Data**:

duck id	1	2	3	4	5	6
color	red	red	blue	blue	blue	red
size	chonker	slim	slim	chonker	chonker	slim

- **Event**:
  - A: a duck is blue
  - B: a duck is a chonker
- Estimate the conditional probability $P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)}$ from this data

Conditional probability example

Example:

- **Experiment**: You ask your ducks to stand in a row again and look at their colors and sizes.
- **Sample space**: The color can be either red or blue; the size can be either slim or chonker.
- **Data**:

duck id	1	2	3	4	5	6
color	red	red	blue	blue	blue	red
size	chonker	slim	slim	chonker	chonker	slim

- **Event**:
  - A: a duck is blue
  - B: a duck is a chonker
- Estimate the conditional probability $P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)}$ from this data
  - $P(A \text{ and } B)$: the probability that a duck is both blue and a chonker is (10 secs)

Conditional probability example

Example:

- **Experiment**: You ask your ducks to stand in a row again and look at their colors and sizes.
- **Sample space**: The color can be either red or blue; the size can be either slim or chonker.
- **Data**:

duck id	1	2	3	4	5	6
color	red	red	blue	blue	blue	red
size	chonker	slim	slim	chonker	chonker	slim

- **Event**:
  - A: a duck is blue
  - B: a duck is a chonker
- Estimate the conditional probability $P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)}$ from this data
  - $P(A \text{ and } B)$: the probability that a duck is both blue and a chonker is (10 secs)

$$\frac{count(\text{blue and chonker})}{\text{total}} = \frac{2}{6}$$

Conditional probability example

Example:

- **Experiment**: You ask your ducks to stand in a row again and look at their colors and sizes.
- **Sample space**: The color can be either red or blue; the size can be either slim or chonker.
- **Data**:

duck id	1	2	3	4	5	6
color	red	red	blue	blue	blue	red
size	chonker	slim	slim	chonker	chonker	slim

- **Event**:
  - A: a duck is blue
  - B: a duck is a chonker
- Estimate the conditional probability $P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)}$ from this data
  - $P(A \text{ and } B)$: the probability that a duck is both blue and a chonker is (10 secs)

$$\frac{\mathit{count}(\text{blue and chonker})}{\text{total}} = \frac{2}{6}$$

  - $P(B)$: the probability that a duck is a chonker is (10 secs)

Conditional probability example

Example:

- **Experiment**: You ask your ducks to stand in a row again and look at their colors and sizes.
- **Sample space**: The color can be either red or blue; the size can be either slim or chonker.
- **Data**:

duck id	1	2	3	4	5	6
color	red	red	blue	blue	blue	red
size	chonker	slim	slim	chonker	chonker	slim

- **Event**:
  - A: a duck is blue
  - B: a duck is a chonker
- Estimate the conditional probability $P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)}$ from this data
  - $P(A \text{ and } B)$: the probability that a duck is both blue and a chonker is (10 secs)

$$\frac{count(\text{blue and chonker})}{\text{total}} = \frac{2}{6}$$

  - $P(B)$: the probability that a duck is a chonker is (10 secs)

$$\frac{count(\text{chonker})}{\text{total}} = \frac{3}{6}$$

Conditional probability example

Example:

- **Experiment**: You ask your ducks to stand in a row again and look at their colors and sizes.
- **Sample space**: The color can be either red or blue; the size can be either slim or chonker.
- **Data**:

duck id	1	2	3	4	5	6
color	red	red	blue	blue	blue	red
size	chonker	slim	slim	chonker	chonker	slim

- **Event**:
  - A: a duck is blue
  - B: a duck is a chonker
- Estimate the conditional probability $P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)}$ from this data
  - $P(A \text{ and } B)$: the probability that a duck is both blue and a chonker is (10 secs)

$$\frac{count(\text{blue and chonker})}{\text{total}} = \frac{2}{6}$$

  - $P(B)$: the probability that a duck is a chonker is (10 secs)

$$\frac{count(\text{chonker})}{\text{total}} = \frac{3}{6}$$

  - Conditional probability:

Conditional probability example

Example:

- **Experiment**: You ask your ducks to stand in a row again and look at their colors and sizes.
- **Sample space**: The color can be either red or blue; the size can be either slim or chonker.
- **Data**:

duck id	1	2	3	4	5	6
color	red	red	blue	blue	blue	red
size	chonker	slim	slim	chonker	chonker	slim

- **Event**:
  - A: a duck is blue
  - B: a duck is a chonker
- Estimate the conditional probability $P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)}$ from this data
  - $P(A$ and $B)$: the probability that a duck is both blue and a chonker is (10 secs)

$$\frac{count(\text{blue and chonker})}{\text{total}} = \frac{2}{6}$$

  - $P(B)$: the probability that a duck is a chonker is (10 secs)

$$\frac{count(\text{chonker})}{\text{total}} = \frac{3}{6}$$

  - Conditional probability:

$$P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)} = \frac{P(A \cap B)}{P(B)} = \frac{2/6}{3/6} = \frac{2}{3}$$

Conditional probability example (cont.)

An alternative way to estimate $P(A \mid B)$:

- Count the chonkers: 3
- Count blue ducks within the chonker gang: 2
- $P(A \mid B) = \frac{2}{3}$

Conditional probability example (cont.)

An alternative way to estimate $P(A \mid B)$:

- Count the chonkers: 3
- Count blue ducks within the chonker gang: 2
- $P(A \mid B) = \frac{2}{3}$

Side note: from this calculation, can you make a bold statement of the probability distribution of all the ducks in the world? No. It is only an estimation based on the data you got.

Conditional probability example (cont.)

As an exercise, let's define the random variables.

- **Experiment**: You ask your ducks to stand in a row again and look at their colors and sizes.
- **Sample space**: The color can be either red or blue; the size can be either slim or chonker.
- **Data**:

duck id	1	2	3	4	5	6
color	red	red	blue	blue	blue	red
size	chonker	slim	slim	chonker	chonker	slim

- **Event**:
    - A: a duck is blue
    - B: a duck is a chonker

Conditional probability example (cont.)

As an exercise, let's define the random variables.

- **Experiment**: You ask your ducks to stand in a row again and look at their colors and sizes.
- **Sample space**: The color can be either red or blue; the size can be either slim or chonker.
- **Data**:

duck id	1	2	3	4	5	6
color	red	red	blue	blue	blue	red
size	chonker	slim	slim	chonker	chonker	slim

- **Event**:
  - A: a duck is blue
  - B: a duck is a chonker
- **Random variables**: X, Y

Hint: X : color $\rightarrow \mathbb{Z}$, Y : size $\rightarrow \mathbb{Z}$ (10 secs)

Conditional probability example (cont.)

As an exercise, let's define the random variables.

- **Experiment**: You ask your ducks to stand in a row again and look at their colors and sizes.
- **Sample space**: The color can be either red or blue; the size can be either slim or chonker.
- **Data**:

duck id	1	2	3	4	5	6
color	red	red	blue	blue	blue	red
size	chonker	slim	slim	chonker	chonker	slim

- **Event**:
  - A: a duck is blue
  - B: a duck is a chonker
- **Random variables**: X, Y

Hint: X : color $\rightarrow \mathbb{Z}$, Y : size $\rightarrow \mathbb{Z}$ (10 secs)

$$X = \begin{cases} 0, & \text{duck is red} \\ 1, & \text{duck is blue} \end{cases} \quad \text{and} \quad Y = \begin{cases} 0, & \text{duck is slim} \\ 1, & \text{duck is a chonker} \end{cases}$$

Conditional probability example (cont.)

As an exercise, let's define the random variables.

- **Experiment**: You ask your ducks to stand in a row again and look at their colors and sizes.
- **Sample space**: The color can be either red or blue; the size can be either slim or chonker.
- **Data**:

duck id	1	2	3	4	5	6
color	red	red	blue	blue	blue	red
size	chonker	slim	slim	chonker	chonker	slim

- **Event**:
  - A: a duck is blue
  - B: a duck is a chonker
- **Random variables**: X, Y

Hint: X : color $\rightarrow \mathbb{Z}$, Y : size $\rightarrow \mathbb{Z}$ (10 secs)

$$X = \begin{cases} 0, & \text{duck is red} \\ 1, & \text{duck is blue} \end{cases} \quad \text{and} \quad Y = \begin{cases} 0, & \text{duck is slim} \\ 1, & \text{duck is a chonker} \end{cases}$$

Are they continuous or discrete? (2 secs)

Conditional probability example (cont.)

As an exercise, let's define the random variables.

- **Experiment**: You ask your ducks to stand in a row again and look at their colors and sizes.
- **Sample space**: The color can be either red or blue; the size can be either slim or chonker.
- **Data**:

duck id	1	2	3	4	5	6
color	red	red	blue	blue	blue	red
size	chonker	slim	slim	chonker	chonker	slim

- **Event**:
  - A: a duck is blue
  - B: a duck is a chonker
- **Random variables**: X, Y

Hint: X : color $\rightarrow \mathbb{Z}$, Y : size $\rightarrow \mathbb{Z}$ (10 secs)

$$X = \begin{cases} 0, & \text{duck is red} \\ 1, & \text{duck is blue} \end{cases} \quad \text{and } Y = \begin{cases} 0, & \text{duck is slim} \\ 1, & \text{duck is a chonker} \end{cases}$$

Are they continuous or discrete? (2 secs) Discrete

Conditional probability example (cont.)

As an exercise, let's define the random variables.

- **Experiment**: You ask your ducks to stand in a row again and look at their colors and sizes.
- **Sample space**: The color can be either red or blue; the size can be either slim or chonker.
- **Data**:

duck id	1	2	3	4	5	6
color	red	red	blue	blue	blue	red
size	chonker	slim	slim	chonker	chonker	slim

- **Event**:
  - A: a duck is blue
  - B: a duck is a chonker
- **Random variables**: X, Y

Hint: X : color $\rightarrow \mathbb{Z}$, Y : size $\rightarrow \mathbb{Z}$ (10 secs)

$$X = \begin{cases} 0, & \text{duck is red} \\ 1, & \text{duck is blue} \end{cases} \quad \text{and} \quad Y = \begin{cases} 0, & \text{duck is slim} \\ 1, & \text{duck is a chonker} \end{cases}$$

Are they continuous or discrete? (2 secs) Discrete

Write the conditional probability in terms of the random variables X and Y (10 secs), i.e.

$$P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)} =$$

Conditional probability example (cont.)

As an exercise, let's define the random variables.

- **Experiment**: You ask your ducks to stand in a row again and look at their colors and sizes.
- **Sample space**: The color can be either red or blue; the size can be either slim or chonker.
- **Data**:

duck id	1	2	3	4	5	6
color	red	red	blue	blue	blue	red
size	chonker	slim	slim	chonker	chonker	slim

- **Event**:
  - A: a duck is blue
  - B: a duck is a chonker
- **Random variables**: X, Y

Hint: X : color $\to \mathbb{Z}$, Y : size $\to \mathbb{Z}$ (10 secs)

$$X = \begin{cases} 0, & \text{duck is red} \\ 1, & \text{duck is blue} \end{cases} \quad \text{and } Y = \begin{cases} 0, & \text{duck is slim} \\ 1, & \text{duck is a chonker} \end{cases}$$

Are they continuous or discrete? (2 secs) Discrete

Write the conditional probability in terms of the random variables X and Y (10 secs), i.e.

$$P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)} = P(X = 1 \mid Y = 1) = \frac{P(X = 1 \text{ and } Y = 1)}{P(Y = 1)}$$

Independent events

Two events A and B are independent if and only if

$$P(A \text{ and } B) := P(A \cap B) = P(A)P(B)$$

$$\iff \quad P(A \mid B) = P(A),\ P(B \mid A) = P(B) \text{ (conditional probability)}$$

$$\iff \quad log\left(P(A \text{ and } B)\right) = \log\left(P(A \cap B)\right) = \log(P(A)) + \log(P(B))$$

Bayes' rule

Given events A and B,

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

Just a heads-up!

Summary: Terminology

- Experiment
- Sample space
- Event
- Random variable:
  - Discrete random variable
  - Continuous random variable
- Data
- Probability distribution:
  - Discrete distribution: P(event) is described by the probability mass function (PMF)
  - Continuous distribution: P(event) is described by the probability density function (PDF)

  What are their differences?
- Conditional probability of events
- Independent events
- Bayes' rule

Some probability distributions that you should know by heart

Probability distributions

Probability distribution	Continuous/discrete	Apply to data type
Bernoulli distribution	Discrete	Categorical (nominal)
Categorical distribution	Discrete	Categorical (nominal)
Discrete uniform	Discrete	Numerical (discrete)
Gaussian distribution	Continuous	Numerical (continuous)

Probability distributions

Probability distribution	Continuous/discrete	Apply to data type
Bernoulli distribution	Discrete	Categorical (nominal)
Categorical distribution	Discrete	Categorical (nominal)
Discrete uniform	Discrete	Numerical (discrete)
Gaussian distribution	Continuous	Numerical (continuous)

Recall

- Discrete distribution is described by the probability mass function (PMF)
- Continuous distribution is described by the probability density function (PDF)

Probability distributions

Probability distribution	Continuous/discrete	Apply to data type
Bernoulli distribution	Discrete	Categorical (nominal)
Categorical distribution	Discrete	Categorical (nominal)
Discrete uniform	Discrete	Numerical (discrete)
Gaussian distribution	Continuous	Numerical (continuous)

Recall

- Discrete distribution is described by the probability mass function (PMF)
- Continuous distribution is described by the probability density function (PDF)

For each distribution, you need to know:

- its PMF or PDF: the equation and the shape
- its parameters
- its applications

Probability distributions

Probability distribution	Continuous/discrete	Apply to data type
Bernoulli distribution	Discrete	Categorical (nominal)
Categorical distribution	Discrete	Categorical (nominal)
Discrete uniform	Discrete	Numerical (discrete)
Gaussian distribution	Continuous	Numerical (continuous)

Recall

- Discrete distribution is described by the probability mass function (PMF)
- Continuous distribution is described by the probability density function (PDF)

For each distribution, you need to know:

- its PMF or PDF: the equation and the shape
- its parameters
- its applications
- how to estimate the parameters (next lecture)

Probability distributions

Probability distribution	Continuous/discrete	Apply to data type
Bernoulli distribution	Discrete	Categorical (nominal)
Categorical distribution	Discrete	Categorical (nominal)
Discrete uniform	Discrete	Numerical (discrete)
Gaussian distribution	Continuous	Numerical (continuous)

Bernoulli distribution

In your town everybody has ducks (of course). Ducks in this town ONLY have TWO colors: blue and red. What is the probability distribution we use to describe the duck colors in your town?

Bernoulli distribution

In your town everybody has ducks (of course). Ducks in this town ONLY have TWO colors: blue and red. What is the probability distribution we use to describe the duck colors in your town?

- Let X be a discrete random variable $X = \begin{cases} 0 & \text{a duck is red} \\ 1 & \text{a duck is blue} \end{cases}$

Bernoulli distribution

In your town everybody has ducks (of course). Ducks in this town ONLY have TWO colors: blue and red. What is the probability distribution we use to describe the duck colors in your town?

- Let X be a discrete random variable $X = \begin{cases} 0 & \text{a duck is red} \\ 1 & \text{a duck is blue} \end{cases}$
- Given p the probability of a duck being blue, we can express the probability distribution as follows:

$$P(\text{a duck is red}) = P(X = 0) = 1 - p$$
$$P(\text{a duck is blue}) = P(X = 1) = p$$

Bernoulli distribution

In your town everybody has ducks (of course). Ducks in this town ONLY have TWO colors: blue and red. What is the probability distribution we use to describe the duck colors in your town?

- Let X be a discrete random variable $X = \begin{cases} 0 & \text{a duck is red} \\ 1 & \text{a duck is blue} \end{cases}$
- Given p the probability of a duck being blue, we can express the probability distribution as follows:

  $P(\text{a duck is red}) = P(X = 0) = 1 - p$
  $P(\text{a duck is blue}) = P(X = 1) = p$
- What is the PMF?

Bernoulli distribution

In your town everybody has ducks (of course). Ducks in this town ONLY have TWO colors: blue and red. What is the probability distribution we use to describe the duck colors in your town?

- Let X be a discrete random variable $X = \begin{cases} 0 & \text{a duck is red} \\ 1 & \text{a duck is blue} \end{cases}$
- Given p the probability of a duck being blue, we can express the probability distribution as follows:
  $P(\text{a duck is red}) = P(X = 0) = 1 - p$
  $P(\text{a duck is blue}) = P(X = 1) = p$
- What is the PMF?
  Merge these two equations:

$$P(X = k) = f_X(k) \equiv f_X(k \mid p) = pk + (1 - p)(1 - k), \; k \in \{0, 1\}, p \in [0, 1]$$

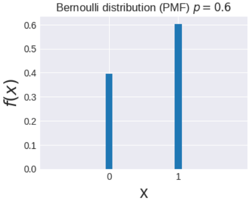
Note: here we use a $|$ to indicate that the parameter p is given.

Bernoulli distribution

- Discrete distribution
- Applies to nominal data with 2 categories
- PMF:
  - Equation

$$f_X(k \mid p) = pk + (1-p)(1-k), k \in \{0, 1\}, p \in [0, 1]$$

  - Shape



- Parameters: p

Probability distributions

Probability distribution	Continuous/discrete	Apply to data type
Bernoulli distribution	Discrete	Categorical (nominal)
Categorical distribution	Discrete	Categorical (nominal)
Discrete uniform	Discrete	Numerical (discrete)
Gaussian distribution	Continuous	Numerical (continuous)

Categorical distribution

In Jack's town, ducks have FOUR colors: blue, red, green and gray. What is the probability distribution of duck colors in Jack's town?

Categorical distribution

In Jack's town, ducks have FOUR colors: blue, red, green and gray. What is the probability distribution of duck colors in Jack's town?

- Given p_1 the probability of a duck being blue, p_2 the probability of a duck being red, p_3 the probability of a duck being green and p_4 the probability of a duck being gray. Note that $p_1 + p_2 + p_3 + p_4 = 1$.

Categorical distribution

In Jack's town, ducks have FOUR colors: blue, red, green and gray. What is the probability distribution of duck colors in Jack's town?

- Given p_1 the probability of a duck being blue, p_2 the probability of a duck being red, p_3 the probability of a duck being green and p_4 the probability of a duck being gray. Note that $p_1 + p_2 + p_3 + p_4 = 1$.
- Let X be a discrete random variable $X = \begin{cases} 1 & \text{a duck is blue} \\ 2 & \text{a duck is red} \\ 3 & \text{a duck is green} \\ 4 & \text{a duck is gray} \end{cases}$

Categorical distribution

In Jack's town, ducks have FOUR colors: blue, red, green and gray. What is the probability distribution of duck colors in Jack's town?

- Given p_1 the probability of a duck being blue, p_2 the probability of a duck being red, p_3 the probability of a duck being green and p_4 the probability of a duck being gray. Note that $p_1 + p_2 + p_3 + p_4 = 1$.
- Let X be a discrete random variable $X = \begin{cases} 1 & \text{a duck is blue} \\ 2 & \text{a duck is red} \\ 3 & \text{a duck is green} \\ 4 & \text{a duck is gray} \end{cases}$
- Now we can express the probability distribution as follows:

$$P(\text{a duck is blue}) = P(X = 1) = p_1$$
$$P(\text{a duck is red}) = P(X = 2) = p_2$$
$$P(\text{a duck is green}) = P(X = 3) = p_3$$
$$P(\text{a duck is gray}) = P(X = 4) = p_4$$

Categorical distribution

In Jack's town, ducks have FOUR colors: blue, red, green and gray. What is the probability distribution of duck colors in Jack's town?

- Given p_1 the probability of a duck being blue, p_2 the probability of a duck being red, p_3 the probability of a duck being green and p_4 the probability of a duck being gray. Note that $p_1 + p_2 + p_3 + p_4 = 1$.
- Let X be a discrete random variable $X = \begin{cases} 1 & \text{a duck is blue} \\ 2 & \text{a duck is red} \\ 3 & \text{a duck is green} \\ 4 & \text{a duck is gray} \end{cases}$
- Now we can express the probability distribution as follows:

$$P(\text{a duck is blue}) = P(X = 1) = p_1$$
$$P(\text{a duck is red}) = P(X = 2) = p_2$$
$$P(\text{a duck is green}) = P(X = 3) = p_3$$
$$P(\text{a duck is gray}) = P(X = 4) = p_4$$

- What is the PMF?

Categorical distribution

In Jack's town, ducks have FOUR colors: blue, red, green and gray. What is the probability distribution of duck colors in Jack's town?

- Given p_1 the probability of a duck being blue, p_2 the probability of a duck being red, p_3 the probability of a duck being green and p_4 the probability of a duck being gray. Note that $p_1 + p_2 + p_3 + p_4 = 1$.
- Let X be a discrete random variable $X = \begin{cases} 1 & \text{a duck is blue} \\ 2 & \text{a duck is red} \\ 3 & \text{a duck is green} \\ 4 & \text{a duck is gray} \end{cases}$
- Now we can express the probability distribution as follows:

  $P(\text{a duck is blue}) = P(X = 1) = p_1$
  $P(\text{a duck is red}) = P(X = 2) = p_2$
  $P(\text{a duck is green}) = P(X = 3) = p_3$
  $P(\text{a duck is gray}) = P(X = 4) = p_4$
- What is the PMF?

$$P(X = k) = f_X(k) \equiv f_X(k \mid p_1, p_2, p_3, p_4) = p_k, \ \sum_{i=1}^{4} p_i = 1, p_i \geq 0, \ k \in \{1, \cdots, 4\}$$

Categorical distribution

In Jack's town, ducks have FOUR colors: blue, red, green and gray. What is the probability distribution of duck colors in Jack's town?

- Given p_1 the probability of a duck being blue, p_2 the probability of a duck being red, p_3 the probability of a duck being green and p_4 the probability of a duck being gray. Note that $p_1 + p_2 + p_3 + p_4 = 1$.
- Let X be a discrete random variable $X = \begin{cases} 1 & \text{a duck is blue} \\ 2 & \text{a duck is red} \\ 3 & \text{a duck is green} \\ 4 & \text{a duck is gray} \end{cases}$
- Now we can express the probability distribution as follows:

  $P(\text{a duck is blue}) = P(X = 1) = p_1$
  $P(\text{a duck is red}) = P(X = 2) = p_2$
  $P(\text{a duck is green}) = P(X = 3) = p_3$
  $P(\text{a duck is gray}) = P(X = 4) = p_4$
- What is the PMF?

$$P(X = k) = f_X(k) \equiv f_X(k \mid p_1, p_2, p_3, p_4) = p_k, \ \sum_{i=1}^{4} p_i = 1, p_i \geq 0, \ k \in \{1, \cdots, 4\}$$

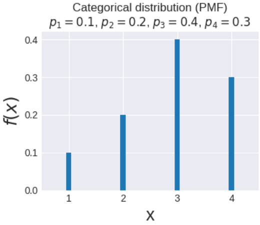
Note: categorical distribution is also called multinoulli distribution. It is a generalization of the Bernoulli distribution.

Categorical distribution

- Discrete distribution
- Applies to nominal data with $n > 0$ categories
- PMF:
  - Equation

$$f_X(k \mid p_1, p_2, \cdots, p_n) = p_k, \ \sum_{i=1}^{n} p_i = 1, p_i \geq 0, \ k \in \{1, \cdots, n\}$$

  - Shape



- Parameters: p_k, $k \in \{1, \cdots, n\}$ for given n; $n-1$ parameters ($\sum_{i=1}^{n} p_i = 1$).

Probability distributions

Probability distribution	Continuous/discrete	Apply to data type
Bernoulli distribution	Discrete	Categorical (nominal)
Categorical distribution	Discrete	Categorical (nominal)
Discrete uniform	Discrete	Numerical (discrete)
Gaussian distribution	Continuous	Numerical (continuous)

Discrete uniform distribution

Meanwhile, back to your town, a team of scientists crunched some numbers and they stated that the number of ducks that each person has follows a uniform distribution between 1 and 1000. What does that mean?

Discrete uniform distribution

Meanwhile, back to your town, a team of scientists crunched some numbers and they stated that the number of ducks that each person has follows a uniform distribution between 1 and 1000. What does that mean?

- Given integers a and b with $a \leq b$ (here we have $a = 1$ and $b = 1000$)

Discrete uniform distribution

Meanwhile, back to your town, a team of scientists crunched some numbers and they stated that the number of ducks that each person has follows a uniform distribution between 1 and 1000. What does that mean?

- Given integers a and b with $a \leq b$ (here we have $a = 1$ and $b = 1000$)
- Let X be a discrete random variable: $X = k$ if a person has k ducks

Discrete uniform distribution

Meanwhile, back to your town, a team of scientists crunched some numbers and they stated that the number of ducks that each person has follows a uniform distribution between 1 and 1000. What does that mean?

- Given integers a and b with $a \leq b$ (here we have $a = 1$ and $b = 1000$)
- Let X be a discrete random variable: $X = k$ if a person has k ducks
- We can express the probability distribution as follows:

$$P(\text{a person has k ducks}) = P(X = k) = \frac{1}{b-a+1}$$

Discrete uniform distribution

Meanwhile, back to your town, a team of scientists crunched some numbers and they stated that the number of ducks that each person has follows a uniform distribution between 1 and 1000. What does that mean?

- Given integers a and b with $a \leq b$ (here we have $a = 1$ and $b = 1000$)
- Let X be a discrete random variable: $X = k$ if a person has k ducks
- We can express the probability distribution as follows:

$$P(\text{a person has k ducks}) = P(X = k) = \frac{1}{b-a+1}$$

- What is the PMF of a discrete uniform distribution?

Discrete uniform distribution

Meanwhile, back to your town, a team of scientists crunched some numbers and they stated that the number of ducks that each person has follows a uniform distribution between 1 and 1000. What does that mean?

- Given integers a and b with $a \leq b$ (here we have $a = 1$ and $b = 1000$)
- Let X be a discrete random variable: $X = k$ if a person has k ducks
- We can express the probability distribution as follows: $$P(\text{a person has k ducks}) = P(X = k) = \frac{1}{b-a+1}$$
- What is the PMF of a discrete uniform distribution?

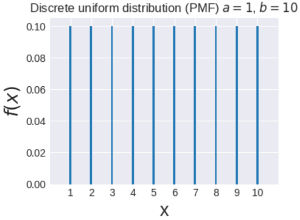
$$P(X = k) = f_X(k) \equiv f_X(k \mid a, b) = \frac{1}{b - a + 1}$$

Discrete uniform distribution

- Discrete distribution
- Applies to discrete numerical data
- PMF:
  - Equation

$$f_X(k \mid a, b) = \frac{1}{b-a+1}, \quad a \leq b, \quad a, b \text{ integers}$$

  - Shape



- Parameters: integers a, b

Probability distributions

Probability distribution	Continuous/discrete	Apply to data type
Bernoulli distribution	Discrete	Categorical (nominal)
Categorical distribution	Discrete	Categorical (nominal)
Discrete uniform	Discrete	Numerical (discrete)
Gaussian distribution	Continuous	Numerical (continuous)

Gaussian (normal) distribution

Still at your town, a scientist told you that the weights of your ducks follow a Gaussian distribution with mean $\mu = 7$ and standard deviation $\sigma = 2$. What does that mean?

Gaussian (normal) distribution

Still at your town, a scientist told you that the weights of your ducks follow a Gaussian distribution with mean $\mu = 7$ and standard deviation $\sigma = 2$. What does that mean?

- Let X be a continuous random variable: $X = x$ if a duck weighs x kg

Gaussian (normal) distribution

Still at your town, a scientist told you that the weights of your ducks follow a Gaussian distribution with mean $\mu = 7$ and standard deviation $\sigma = 2$. What does that mean?

- Let X be a continuous random variable: $X = x$ if a duck weighs x kg
- Let $f_X(x)$ be the PDF, the probability distribution can be expressed as (10 secs):

Gaussian (normal) distribution

Still at your town, a scientist told you that the weights of your ducks follow a Gaussian distribution with mean $\mu = 7$ and standard deviation $\sigma = 2$. What does that mean?

- Let X be a continuous random variable: $X = x$ if a duck weighs x kg
- Let $f_X(x)$ be the PDF, the probability distribution can be expressed as (10 secs):

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

Gaussian (normal) distribution

Still at your town, a scientist told you that the weights of your ducks follow a Gaussian distribution with mean $\mu = 7$ and standard deviation $\sigma = 2$. What does that mean?

- Let X be a continuous random variable: $X = x$ if a duck weighs x kg
- Let $f_X(x)$ be the PDF, the probability distribution can be expressed as (10 secs):

$$P(a \leq X \leq b) = \int_a^b f_X(x)dx \text{ for all } a < b$$

- What is the PDF of a Gaussian distribution with $\mu = 7$ and $\sigma = 2$?

Gaussian (normal) distribution

Still at your town, a scientist told you that the weights of your ducks follow a Gaussian distribution with mean $\mu = 7$ and standard deviation $\sigma = 2$. What does that mean?

- Let X be a continuous random variable: $X = x$ if a duck weighs x kg
- Let $f_X(x)$ be the PDF, the probability distribution can be expressed as (10 secs):

  $P(a \leq X \leq b) = \int_a^b f_X(x)dx$ for all $a < b$
- What is the PDF of a Gaussian distribution with $\mu = 7$ and $\sigma = 2$?

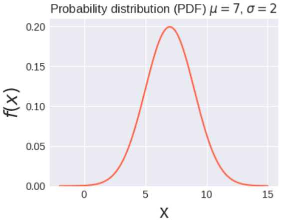
$$f_X(x \mid \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} = \frac{1}{\sqrt{8\pi}} e^{-\frac{1}{2}\left(\frac{x-7}{2}\right)^2}$$

Gaussian (normal) distribution

- Continuous distribution
- Applies to continuous numerical data
- PDF:
  - Equation

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \;\mu \in \mathbb{R}, \;\sigma \in \mathbb{R}^+$$

  - Shape



- Parameters: $\mu, \; \sigma$

Probability distributions

Probability distribution	Continuous/discrete	Apply to data type
Bernoulli distribution	Discrete	Categorical (nominal)
Categorical distribution	Discrete	Categorical (nominal)
Discrete uniform	Discrete	Numerical (discrete)
Gaussian distribution	Continuous	Numerical (continuous)

Hooray!

An important note

An important note

These probability distributions DO NOT ONLY apply to duck related applications!

Check out what data types they apply to!

We are going to talk about more applications in the future

An important note

These probability distributions DO NOT ONLY apply to duck related applications!

Check out what data types they apply to!

We are going to talk about more applications in the future (even though they won't be as important as ducks)

Today

Demo

Code demo

- Image processing
- Natural language processing
- Table with numerical data
- Table with categorical data

Today

So far:

- Data types, data containers, descriptive statistics (e.g. sample mean, sample variance, data quantile), visualization (e.g. histogram)
- Probability distributions, sample space, events, random variables, PMF, PDF, parameters

So far:

- Data types, data containers, descriptive statistics (e.g. sample mean, sample variance, data quantile), visualization (e.g. histogram)
- Probability distributions, sample space, events, random variables, PMF, PDF, parameters

Not yet:

- How to estimate parameters, such as μ and σ in a Gaussian distribution?
- How to choose the probability distribution for a given data set?

So far:

- Data types, data containers, descriptive statistics (e.g. sample mean, sample variance, data quantile), visualization (e.g. histogram)
- Probability distributions, sample space, events, random variables, PMF, PDF, parameters

Not yet:

- How to estimate parameters, such as μ and σ in a Gaussian distribution?
- How to choose the probability distribution for a given data set?

Next:

- Q-Q plot and mathematical modeling

So far:

- Data types, data containers, descriptive statistics (e.g. sample mean, sample variance, data quantile), visualization (e.g. histogram)
- Probability distributions, sample space, events, random variables, PMF, PDF, parameters

Not yet:

- How to estimate parameters, such as μ and σ in a Gaussian distribution?
- How to choose the probability distribution for a given data set?

Next:

- Q-Q plot and mathematical modeling

Before next lecture:

- Quantile
- Probability distributions from today and their parameters
- PMF and PDF

Stay safe!