ECMM443
Introduction to Data Science

Dr P Lewis / (from Dr Xiaoyang Wang)

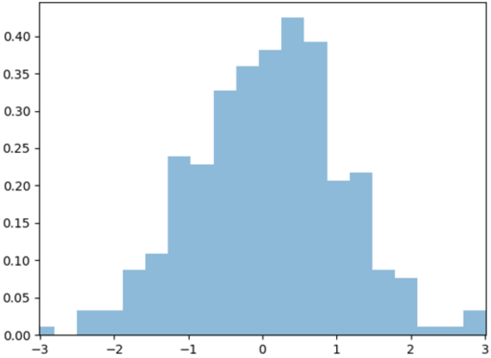
Department of Computer Science

p.lewis2@exeter.ac.uk

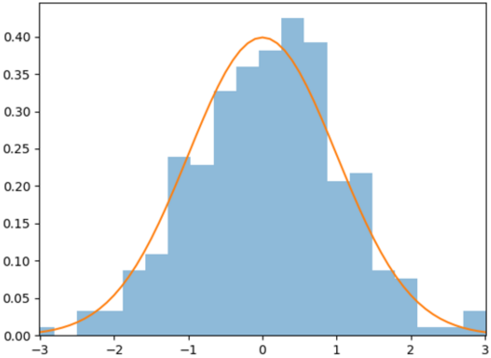
Probability

Working out probabilities can be tricky!
The Monty Hall Problem

Frequentist Probability



Frequentist Probability



Data: Random sample

Curve: Underlining probability distribution

Probability

- Probability is about the **outcomes** of **experiments**.
- Outcomes are random.

Probability

- Probability is about the **outcomes** of **experiments**.
- Outcomes are random.
- Often there are a **finite** number of possible outcomes.
- For example, when flipping a coin the outcomes are
  - Heads (H) or Tails (T)

Probability

- Probability is about the **outcomes** of **experiments**.
- Outcomes are random.
- Often there are a **finite** number of possible outcomes.
- For example, when flipping a coin the outcomes are
  - Heads (H) or Tails (T)
- The set of possible outcomes is called the **sample space**.
  - For a coin tossing experiment S = {H,T}
  - The sample space for a standard die is S = {1,2,3,4,5,6}.

Probability

- An **event** is a particular outcome, or set of outcomes.
  - For example: the event of getting tails {T}, the event of rolling an even number {2,4,6}.

Probability

- An **event** is a particular outcome, or set of outcomes.
  - For example: the event of getting tails {T}, the event of rolling an even number {2,4,6}.
- **Probability** is a way of associating a number with every possible event.

Probability

- An **event** is a particular outcome, or set of outcomes.
  - For example: the event of getting tails {T}, the event of rolling an even number {2,4,6}.
- **Probability** is a way of associating a number with every possible event.

- A **probability function**, P, has to satisfy three properties
  1. $P(S) = 1$
  2. For any event E in S, $P(E) \geqq 0$
  3. If A and B are disjoint events then $P(A \cup B) = P(A) + P(B)$

Probability

- An **event** is a particular outcome, or set of outcomes.
  - For example: the event of getting tails {T}, the event of rolling an even number {2,4,6}.
- **Probability** is a way of associating a number with every possible event.

- A **probability function**, P, has to satisfy three properties
  1. $\boxed{P(S) = 1}$
  2. For any event E in S, $P(E) \geqq 0$
  3. If A and B are disjoint events then $P(A \cup B) = P(A) + P(B)$

Probability

- An **event** is a particular outcome, or set of outcomes.
  - For example: the event of getting tails {T}, the event of rolling an even number {2,4,6}.
- **Probability** is a way of associating a number with every possible event.

- A **probability function**, P, has to satisfy three properties
  1. $P(S) = 1$
  2. $\boxed{\text{For any event } E \text{ in } S,\ P(E) \geqq 0}$
  3. If A and B are disjoint events then $P(A \cup B) = P(A) + P(B)$

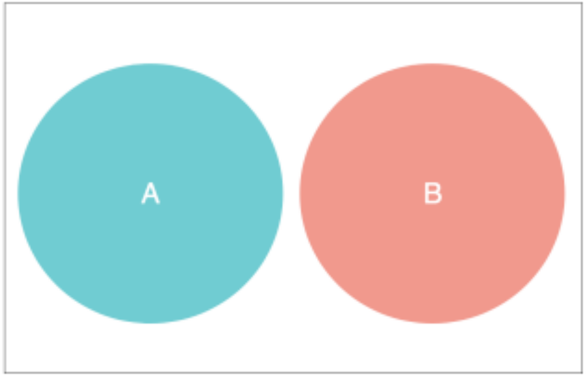
Probability

- An **event** is a particular outcome, or set of outcomes.
    - For example: the event of getting tails {T}, the event of rolling an even number {2,4,6}.
- **Probability** is a way of associating a number with every possible event.

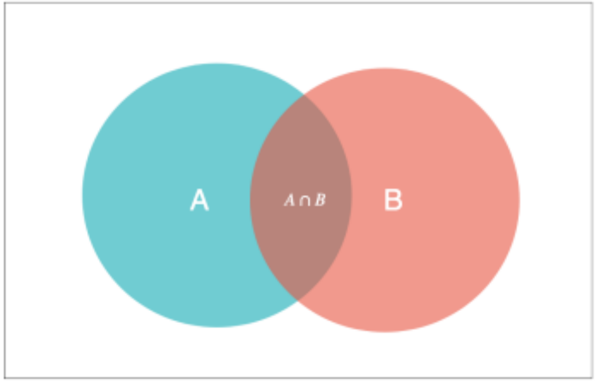
- A **probability function**, P, has to satisfy three properties
    1. $P(S) = 1$
    2. For any event E in S, $P(E) \geqq 0$
    3. $\boxed{\text{If A and B are disjoint events then } P(A \cup B) = P(A) + P(B)}$

Probability

- **Disjoint events** are events, that is sets, which don't have any members in common.
  - For example {2,4} and {1,5} are disjoint.
- P(A ∪ B) is read as the 'probability of A *or* B'.
- The third rule P(A ∪ B) = P(A) + P(B):
  - for events which are disjoint
  - to get the probability of one **or** both happening
  - **add** the probabilities of the separate events.

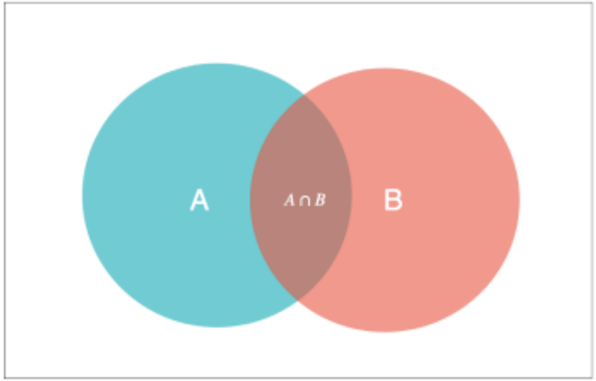


Probability



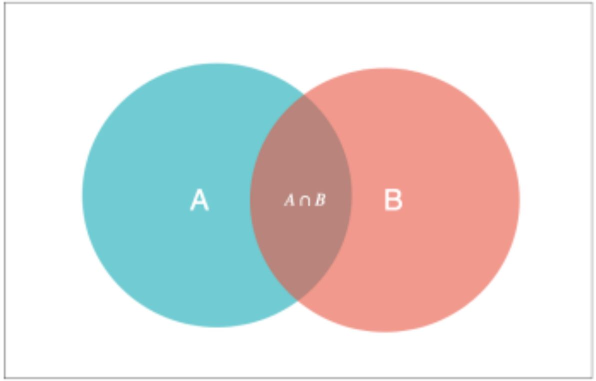
Probability

- The probability of A *and* B happening is written P(A ∩ B).



Probability

- The probability of A *and* B happening is written $P(A \cap B)$.
- When event A and event B, overlap, so they are not disjoint, the probability of A or B, $P(A \cup B)$ is:
  - $P(A \cup B) = P(A) + P(B) - P(A \cap B)$



Conditional Probability

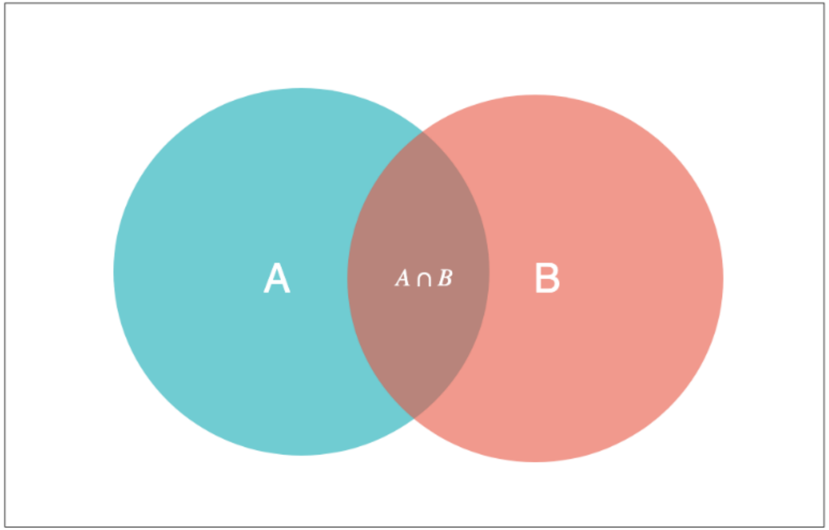
- The probability that A happened given that B already happened.

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

Conditional Probability

- The probability that A happened given that B already happened.

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



Independence

- If

$$P(A) = P(A|B)$$

- Then A and B are independent.

Independence

- If

$$P(A) = P(A|B)$$

- Then A and B are independent.
- The fact that B happened has no effect on the probability of observing A.

Independence

$$P(A) = P(A|B)$$

Independence

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

Independence

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

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

Independence

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

- For independent events, A and B, the probability of A **and** B happening is given by multiplying the individual probabilities.

Random Variables

Random Variables

- A random variable is a number associated with each event in the sample space.

- X = f(event)

Probability Distributions

- Flipping a coin twice gives the following sample space
  - {HH,TH,HT, TT}

Probability Distributions

- Flipping a coin twice gives the following sample space
  - {HH,TH,HT, TT}
- Examples of random variables:
  - the number of heads,
  - the number of tails,
  - the square root of the number of tails...

Probability Distributions

- Flipping a coin twice gives the following sample space
  - {HH,TH,HT, TT}
- Examples of random variables:
  - the number of heads,
  - the number of tails,
  - the square root of the number of tails...
- **Discrete random variables:** only a finite set of possible outcomes
  - flipping a coin
- **Continuous random variables:** where the value is a real number
  - measuring the height of a random person.
  - A random variable is often denoted by capital Roman letters such as X, Y, Z

Bernoulli Random Variable

If we are considering a discrete event where there are only two outcomes (0, 1)

Then we are dealing with a Bernoulli random variable.
This arises in problems such as:

- flipping a coin,
- winning/losing a game,
- success or failure of a treatment etc.

In this case we deal with a single event probability p:

$$P(1) = p \tag{3}$$

Probability Distributions

- The **probability density function**

$$f(x) = P(Y = x)$$

- Is the probability of observing a random variable with the value x.

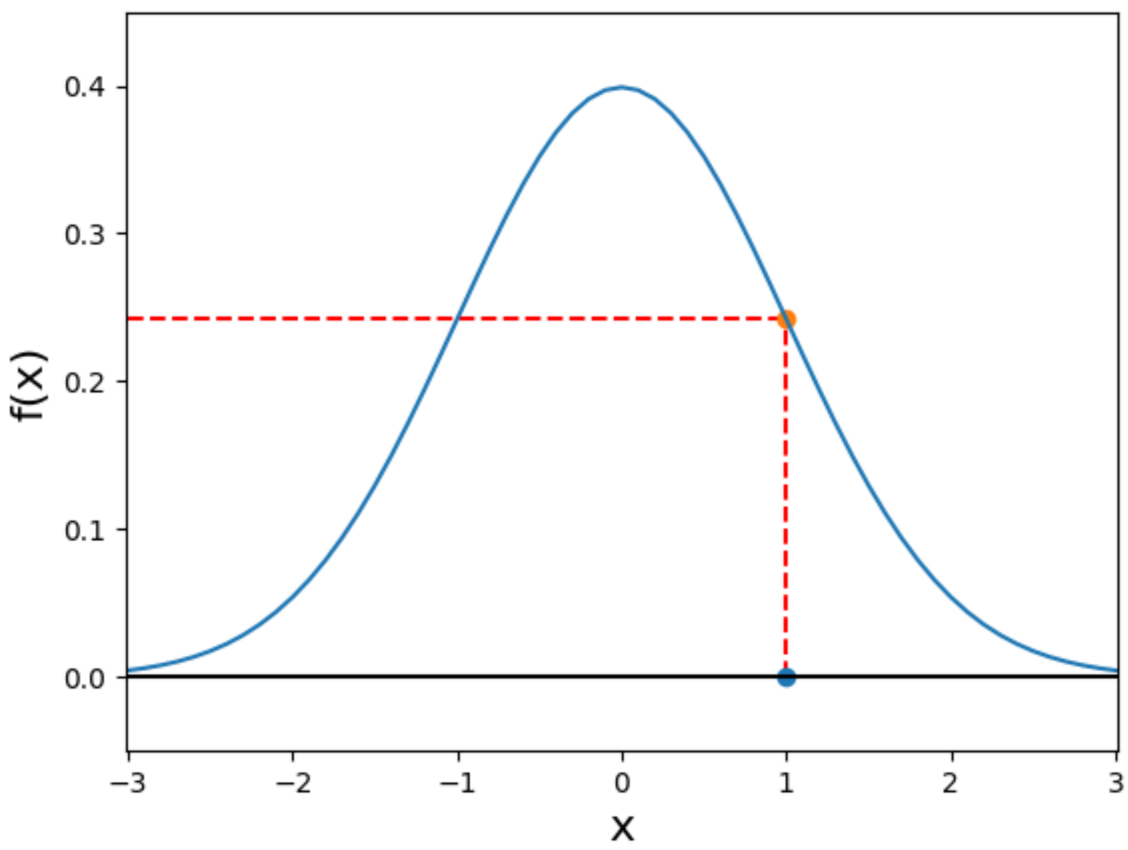
Probability Distributions

- The **probability density function**

$$f(x) = P(Y = x)$$

- Is the probability of observing a random variable with the value x.
- If the random variable is discrete this is also called a **probability mass function**.

Probability Distributions



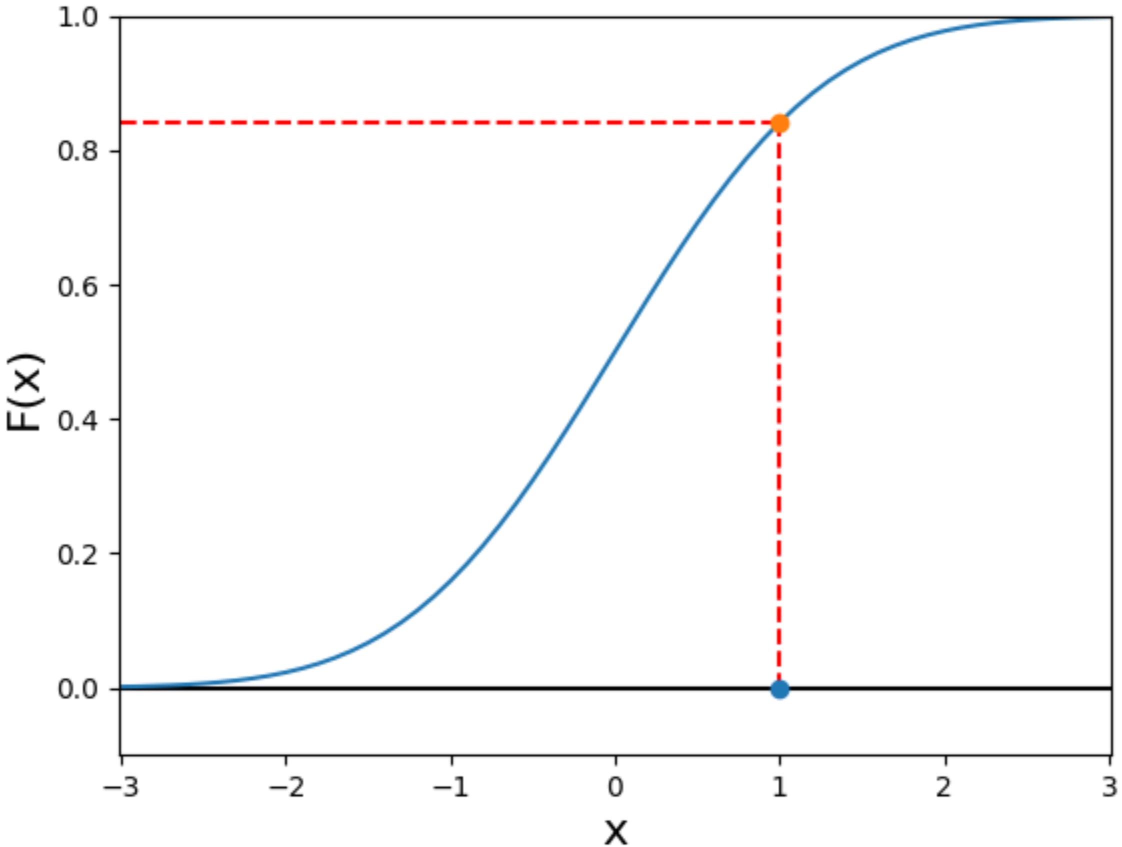
Probability Distributions

- The **cumulative distribution function**

$$F(x) = P(Y \leq x)$$

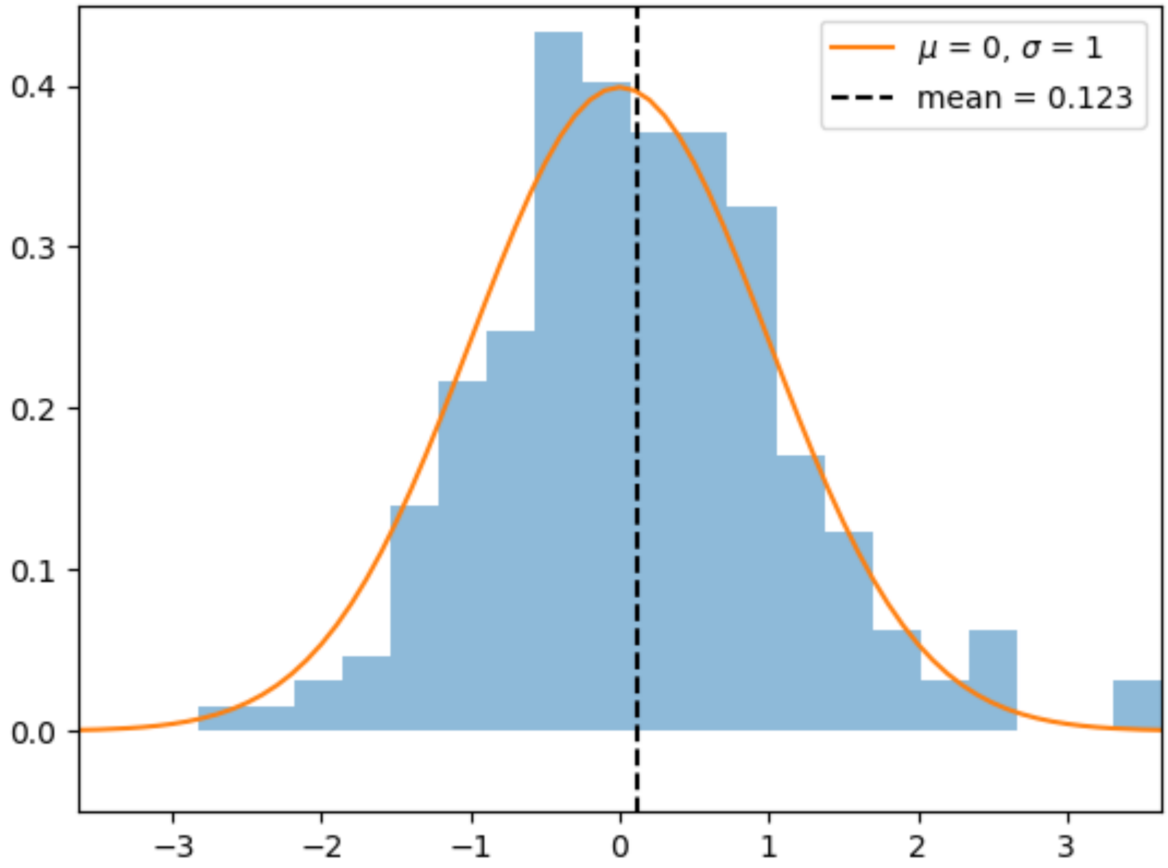
- Is the probability of observing a value of the random variable X that is less than or equal to x.

Probability Distributions



Expectation

Expectation



Expectation

- The expectation value of a random variable is denoted

$$\mu = E[x] = \sum_i x_i P(x_i)$$

Expectation

- The expectation value of a random variable is denoted

$$\mu = E[x] = \sum_i x_i P(x_i)$$

- For a Bernoulli random variable

$$E[x] = 0.(1-p) + 1.p = p$$

Expectation

- An important property of expectation is linearity

Expectation

- An important property of expectation is linearity

$$E[X+Y] = \sum_{i,j}(x_i + y_j)P(x_i \cap y_j)$$

Expectation

- An important property of expectation is linearity

$$E[X+Y] = \sum_{i,j} (x_i + y_j) P(x_i \cap y_j)$$

$$= \sum_{i,j} x_i P(x_i \cap y_j) + \sum_{i,j} y_j P(x_i \cap y_j)$$

Expectation

- An important property of expectation is linearity

$$E[X+Y] = \sum_{i,j}(x_i + y_j)P(x_i \cap y_j)$$
$$= \sum_{i,j} x_i P(x_i \cap y_j) + \sum_{i,j} y_j P(x_i \cap y_j)$$
$$= \sum_{i} x_i P(x_i) + \sum_{j} y_j P(y_j)$$

- We used the fact that

$$\sum_{j} P(x_i \cap y_j) = P(x_i)$$

Expectation

- An important property of expectation is linearity

$$
\begin{aligned}
E[X+Y] &= \sum_{i,j} (x_i + y_j) P(x_i \cap y_j) \\
&= \sum_{i,j} x_i P(x_i \cap y_j) + \sum_{i,j} y_j P(x_i \cap y_j) \\
&= \sum_{i} x_i P(x_i) + \sum_{j} y_j P(y_j) \\
&= E[X] + E[Y]
\end{aligned}
$$

Expectation

- Expectation also scales, so that for some constant a

$$
\begin{aligned}
E[aX] &= \sum_i a x_i P(x_i) \\
&= a \sum_i x_i P(x_i) \\
&= aE[X]
\end{aligned}
$$

Expectation

- What is the expected number of heads after three coin tosses?

Expectation

- In general the number of ways to get k successes in n trials is
  n choose k

$$\boxed{\binom{n}{k}} = \frac{n!}{k!(n-k)!}$$

$$n! = n(n-1)(n-2)\ldots 2.1$$

Expectation

- What is the expected number of heads after three coin tosses?

$$E[x] = 0.\binom{3}{0}(1-p)^3 + 1.\binom{3}{1}p(1-p)^2 + 2.\binom{3}{2}p^2(1-p) + 3.\binom{3}{3}p^3$$

Expectation

- What is the expected number of heads after three coin tosses?

$$E[x] = 0.\binom{3}{0}(1-p)^3 + 1.\binom{3}{1}p(1-p)^2 + 2.\binom{3}{2}p^2(1-p) + 3.\binom{3}{3}p^3$$
$$= 0 + \frac{3}{8} + \frac{6}{8} + \frac{3}{8}$$

Expectation

- What is the expected number of heads after three coin tosses?

$$
\begin{aligned}
E[x] &= 0.\binom{3}{0}(1-p)^3 + 1.\binom{3}{1}p(1-p)^2 + 2.\binom{3}{2}p^2(1-p) + 3.\binom{3}{3}p^3 \\
&= 0 + \frac{3}{8} + \frac{6}{8} + \frac{3}{8} \\
&= \frac{3}{2}
\end{aligned}
$$

Variance

- The variance of a random variable is

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

- Expanding

$$Var[x] = \sum_i (x_i - \mu)^2 P(x_i)$$

Variance

- The variance of a random variable is

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

- Expanding

$$Var[x] = \sum_i (x_i - \mu)^2 P(x_i)$$
$$= \sum_i x_i^2 P(x_i) - 2\mu \sum_i x_i P(x_i) + \mu^2 \sum_i P(x_i)$$

Variance

- The variance of a random variable is

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

- Expanding

$$\begin{aligned} Var[x] &= \sum_i (x_i - \mu)^2 P(x_i) \\ &= \sum_i x_i^2 P(x_i) - 2\mu \sum_i x_i P(x_i) + \mu^2 \sum_i P(x_i) \\ &= E[x^2] - 2\mu^2 + \mu^2 \end{aligned}$$

Variance

- The variance of a random variable is

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

- Expanding

$$\begin{aligned} Var[x] &= \sum_i (x_i - \mu)^2 P(x_i) \\ &= \sum_i x_i^2 P(x_i) - 2\mu \sum_i x_i P(x_i) + \mu^2 \sum_i P(x_i) \\ &= E[x^2] - 2\mu^2 + \mu^2 \\ &= E[x^2] - E[x]^2 \end{aligned}$$

Moments

- The **second moment** of the probability distribution

$$E[x^2]$$

- The **nth moment** of the probability distribution

$$E[x^n] = \sum_i x_i^n P(x_i)$$

- Won’t cover in this course, but moments and related functions are very useful for proving theorems about probability distributions!

Law of Large Numbers

Law of Large Numbers: For a sequence of independent, identically distributed random variables the sample mean gets closer and closer to the true mean $\mu = E[x]$.

Law of Large Numbers

Law of Large Numbers: For a sequence of independent, identically distributed random variables the sample mean gets closer and closer to the true mean μ = E[x].

- The **sample mean** is what we called the average before

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Law of Large Numbers

- We imagine something like a coin tossing experiment, repeated trials where the outcomes are described by the same probability distribution.
  - This is the **independent identically distributed** part

Law of Large Numbers

- We imagine something like a coin tossing experiment, repeated trials where the outcomes are described by the same probability distribution.
    - This is the **independent identically distributed** part
- The sample mean is our best guess at the `typical' result after making n trials.

Law of Large Numbers

- We imagine something like a coin tossing experiment, repeated trials where the outcomes are described by the same probability distribution.
  - This is the **independent identically distributed** part
- The sample mean is our best guess at the `typical' result after making n trials.
- The law of large numbers says: if we do enough experiments the sample mean will be very close to the mean of the distribution.

Bayes Theorem

Bayes' Theorem

- Bayes Theorem allows us to update the probability of an event when we learn additional information

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

Bayes' Theorem

- Bayes Theorem allows us to update the probability of an event when we learn additional information

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

- P(A) and P(B) are called **marginal** probabilities
- P(A) is also sometimes called the **prior** probability
- P(A|B) is posterior

The Monty Hall Problem

Door 1	Door 2	Door 3	Result if stick	Result if switch
Prize	Goat	Goat	Prize	Goat
Goat	Prize	Goat	Goat	Prize
Goat	Goat	Prize	Goat	Prize

In the case of 20 doors, if we don't switch
$P(\text{win the prize}) = \frac{1}{20}$
$P(\text{win a goat}) = 1 - P(\text{win the prize}) = \frac{19}{20}$

The Monty Hall Problem

Door 1	Door 2	Door 3	Result if stick	Result if switch
Prize	Goat	Goat	Prize	Goat
Goat	Prize	Goat	Goat	Prize
Goat	Goat	Prize	Goat	Prize

In the case of 20 doors, if we don't switch
$P(\text{win the prize}) = \frac{1}{20}$
$P(\text{win a goat}) = 1 - P(\text{win the prize}) = \frac{19}{20}$

Having additional information: Monty shows a door with a goat behind it - it changes the decision!

Bayes Theorem

The Monty Hall Problem

Assume we pick door 1, and Monty Hall reveals that door 3 is a goat.

H: behind door 1 is a prize

E: evidence, door 3 is a goat

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)} \tag{1}$$

$P(H)$: prior probability, $\frac{1}{3}$

H^c: complementary event of H, so $P(H \cup H^c) = 1$

$P(E|H)$: given that behind door 1 is a prize, the probability of door 3 is a goat

$P(E)$: the probability that door 3 is a goat

$$P(E) = P(E|H)P(H) + P(E|H^c)P(H^c) \tag{2}$$

Visualisation

https://seeing-theory.brown.edu/index.html