Defining Random Variables

R. Mok BSc(AdvMath), MTeach, GradCertDS

AISNSW Focus Day, November 2021

Table of Contents

Table of Contents

Purpose of Presentation

The Problem

Purpose of Presentation

The Problem

- The Syllabus attempts to define a random variable in terms of what it tries to achieve/model but doesn't actually define what it is:
  - Define and categorise random variables
    - know that a random variable describes some aspect in a population from which samples can be drawn
    - know the difference between a discrete random variable and a continuous random variable

(NSW Education Standards Authority, 2017, p. 47)

Purpose of Presentation

The Problem

- The Syllabus attempts to define a random variable in terms of what it tries to achieve/model but doesn't actually define what it is:
  - Define and categorise random variables
    - know that a random variable describes some aspect in a population from which samples can be drawn
    - know the difference between a discrete random variable and a continuous random variable

  (NSW Education Standards Authority, 2017, p. 47)
- The Syllabus glossary isn't a great definition at all:
  - A random variable is a variable whose possible values are outcomes of a statistical experiment or random phenomenon. (NSW Education Standards Authority, 2017, p. 73)

Purpose of Presentation

The Problem

- The Syllabus attempts to define a random variable in terms of what it tries to achieve/model but doesn't actually define what it is:
  - Define and categorise random variables
    - know that a random variable describes some aspect in a population from which samples can be drawn
    - know the difference between a discrete random variable and a continuous random variable

  (NSW Education Standards Authority, 2017, p. 47)
- The Syllabus glossary isn't a great definition at all:
  - A random variable is a variable whose possible values are outcomes of a statistical experiment or random phenomenon. (NSW Education Standards Authority, 2017, p. 73)

Purpose of Presentation

The Problem

- The Syllabus attempts to define a random variable in terms of what it tries to achieve/model but doesn't actually define what it is:
  - Define and categorise random variables
    - know that a random variable describes some aspect in a population from which samples can be drawn
    - know the difference between a discrete random variable and a continuous random variable

  (NSW Education Standards Authority, 2017, p. 47)
- The Syllabus glossary isn't a great definition at all:
  - A random variable is a variable whose possible values are outcomes of a statistical experiment or random phenomenon. (NSW Education Standards Authority, 2017, p. 73)
- A well known textbook used in schools gives an *example* of a random variable that models tossing 5 coins rather than defines what it is.

Purpose of Presentation

Purpose

Purpose of Presentation

Purpose

- Define probability concepts properly such as Random Variables.

Purpose of Presentation

Purpose

- Define probability concepts properly such as Random Variables.
- Analogous to being able to *compute* derivatives vs. *understand* derivatives, a deeper understanding of Random Variables enhances one's appreciation for the computations and applications.

Purpose of Presentation

(Tao, 2014, pp. 2-3)

There is a certain philosophical satisfaction in knowing *why* things work... you can certainly use things like the chain rule, L'Hôpital's rule, or integration by parts without knowing why these rules work, or whether there are exceptions to these rules. However, one can get into trouble if one applies rules without knowing where they came from and what the limits of their applicability are.

Table of Contents

Preliminaries

Sets

A set is a collection of distinct objects. e.g. $A = \{1, 2, a, \text{Dave}\}$, or $B = \{c, 2, \pi, \mathbb{H}, \mathfrak{g}\}$.

Recall the set operations:

- Union: $A \cup B = \{1, 2, a, c, \pi, \mathbb{H}, \mathfrak{g}, \text{Dave}\}$
- Intersection: $A \cap B = \{2\}$

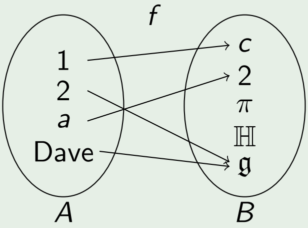
Recall that a set C is a subset of A if all elements of C are also in A, and we write: $C \subseteq A$. Example: $\{1, \text{Dave}\} \subseteq A$.

Preliminaries

Function

A function f from a set A into B, denoted $f : A \to B$ assigns to each $a \in A$ an element $f(a) = b \in B$.
The set A is called the *domain* and the set B is called the *co-domain*.
The *range* is the set $\{f(a) | a \in A\}$.

Example

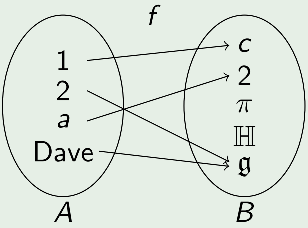


Preliminaries

Function

A function f from a set A into B, denoted $f : A \to B$ assigns to each $a \in A$ an element $f(a) = b \in B$.
The set A is called the *domain* and the set B is called the *co-domain*.
The *range* is the set $\{f(a) | a \in A\}$.

Example



Domain: $\{1, 2, a, \text{Dave}\}$

Co-domain: $\{c, 2, \pi, \mathbb{H}, \mathfrak{g}\}$

Range: $\{c, 2, \mathfrak{g}\}$

Countable and Uncountable Sets

Two types of size (or cardinality) of sets that one encounters in mathematics are:

- Countable: elements in a countable set can be listed (enumerated), e.g the set of integers $\mathbb{Z}$, the set of rational numbers $\mathbb{Q}$, any finite set.
- Uncountable: elements in an uncountable set cannot be listed (enumerated). e.g. the set of real numbers $\mathbb{R}$, the real interval $[0, 1]$.

Table of Contents

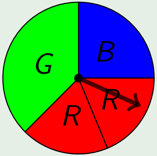
Sample Space

Sample Space (Tao, 2011, p. 192)

The Sample Space is the set of all possible states that a random system could be in, denoted Ω.

Example

Consider the coloured spinner:



What is the sample space?

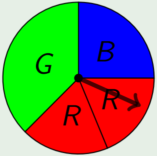
Sample Space

Sample Space (Tao, 2011, p. 192)

The Sample Space is the set of all possible states that a random system could be in, denoted Ω.

Example

Consider the coloured spinner:



What is the sample space? $\Omega = \{B, G, R\}$

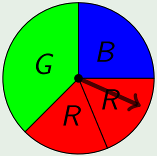
Sample Space

Sample Space (Tao, 2011, p. 192)

The Sample Space is the set of all possible states that a random system could be in, denoted Ω.

Example

Consider the coloured spinner:



What is the sample space? $\Omega = \{B, G, R\}$
It could also be: $\Omega = [0, 2\pi)$

Events and Event Spaces

Event (Tao, 2011, p. 192)

An Event, E, is a subset of the Sample Space Ω. i.e. $E \subseteq \Omega$.

Example

The set $E = \{2, 4, 6\}$ would represent the Event of rolling even numbers on a 6-sided die with sample space $\Omega = \{1, 2, 3, 4, 5, 6\}$.

Event Space (Tao, 2011, p. 192)

An Event Space (also known as a sigma-field), $\mathscr{F}$, is the set of all possible Events that one can measure the probability of.

Example

Continuing with the running example of the spinner on the previous slide, the Event Space would be:

Event Space (Tao, 2011, p. 192)

An Event Space (also known as a sigma-field), $\mathscr{F}$, is the set of all possible Events that one can measure the probability of.

Example

Continuing with the running example of the spinner on the previous slide, the Event Space would be:

- $\mathscr{F} = \{\{\}, \{B\}, \{G\}, \{R\}, \{B, G\}, \{B, R\}, \{G, R\}, \{B, G, R\}\}$

Event Space (Tao, 2011, p. 192)

An Event Space (also known as a sigma-field), $\mathscr{F}$, is the set of all possible Events that one can measure the probability of.

Example

Continuing with the running example of the spinner on the previous slide, the Event Space would be:

- $\mathscr{F} = \{\{\}, \{B\}, \{G\}, \{R\}, \{B, G\}, \{B, R\}, \{G, R\}, \{B, G, R\}\}$
- What would $\mathscr{F}$ be if the Sample Space was $[0, 2\pi)$? (Not in the scope of today's presentation - open for questions at the end of the presentation)

Probability Measure

Probability Measure (Daners, 2012, p. 10)

The probability measure of an Event E, denoted $P(E)$, is a function $P : \mathscr{F} \to [0, 1]$, and:

- For all $n \in \mathbb{N}$, if the events $E_1, E_2, \ldots, E_n \in \mathscr{F}$ are disjoint (i.e. they all have empty intersection), then $P(E_1 \cup E_2 \cup \ldots \cup E_n) = P(E_1) + P(E_2) + \ldots + P(E_n)$.
- $P(\{\}) = 0$
- $P(\Omega) = 1$

Example

Probability Measure

Probability Measure (Daners, 2012, p. 10)

The probability measure of an Event E, denoted $P(E)$, is a function $P : \mathscr{F} \to [0, 1]$, and:

- For all $n \in \mathbb{N}$, if the events $E_1, E_2, \ldots, E_n \in \mathscr{F}$ are disjoint (i.e. they all have empty intersection), then $P(E_1 \cup E_2 \cup \ldots \cup E_n) = P(E_1) + P(E_2) + \ldots + P(E_n)$.
- $P(\{\}) = 0$
- $P(\Omega) = 1$

Example

Continuing with the running example of the spinner with sample space $\Omega = \{B, G, R\}$, we can measure the probability of the Events with the following function:
$P(\{\}) = 0$, $P(\{B\}) = \frac{1}{4}$, $P(\{G\}) = \frac{3}{8}$, $P(\{R\}) = \frac{3}{8}$. The probability of the other events can be found using the first property listed above.

Summary So Far

Summary So Far

- Sets

Summary So Far

Summary So Far

- Sets
- Functions $f : A \to B$, domain, co-domain, range

Summary So Far

Summary So Far

- Sets
- Functions $f : A \to B$, domain, co-domain, range
- Countable and Uncountable sets

Summary So Far

Summary So Far

- Sets
- Functions $f : A \to B$, domain, co-domain, range
- Countable and Uncountable sets
- Sample Space Ω

Summary So Far

Summary So Far

- Sets
- Functions $f : A \to B$, domain, co-domain, range
- Countable and Uncountable sets
- Sample Space Ω
- Event Space $\mathscr{F}$

Summary So Far

Summary So Far

- Sets
- Functions $f : A \to B$, domain, co-domain, range
- Countable and Uncountable sets
- Sample Space Ω
- Event Space $\mathscr{F}$
- Probability Measure $P : \mathscr{F} \to [0, 1]$
- Note: Many references call the triple $(\Omega, \mathscr{F}, P)$ a Probability Space.

Table of Contents

Random Variable (Daners, 2012, p. 143)

A random variable X is a *function* $X : \Omega \to \mathbb{R}$.

Random Variables

Random Variable (Daners, 2012, p. 143)

A random variable X is a *function* $X : \Omega \to \mathbb{R}$.

- X is a *discrete random variable* if and only if the *range* of X is countable.

Random Variable (Daners, 2012, p. 143)

A random variable X is a *function* $X : \Omega \to \mathbb{R}$.

- X is a *discrete random variable* if and only if the *range* of X is countable.
- X is a *continuous random variable* if and only if the *range* of X is uncountable.

Random Variables

Random Variable (Daners, 2012, p. 143)

A random variable X is a *function* $X : \Omega \to \mathbb{R}$.

- X is a *discrete random variable* if and only if the *range* of X is countable.
- X is a *continuous random variable* if and only if the *range* of X is uncountable.

Useful Notation for Random Variables

Some conventional notation:

Random Variables

Random Variable (Daners, 2012, p. 143)

A random variable X is a *function* $X : \Omega \to \mathbb{R}$.

- X is a *discrete random variable* if and only if the *range* of X is countable.
- X is a *continuous random variable* if and only if the *range* of X is uncountable.

Useful Notation for Random Variables

Some conventional notation:

- $[X = a] = \{\omega \in \Omega | X(\omega) = a\}$

Random Variables

Random Variable (Daners, 2012, p. 143)

A random variable X is a *function* $X : \Omega \to \mathbb{R}$.

- X is a *discrete random variable* if and only if the *range* of X is countable.
- X is a *continuous random variable* if and only if the *range* of X is uncountable.

Useful Notation for Random Variables

Some conventional notation:

- $[X = a] = \{\omega \in \Omega | X(\omega) = a\}$
- $[X < a] = \{\omega \in \Omega | X(\omega) < a\}$

Random Variables

Random Variable (Daners, 2012, p. 143)

A random variable X is a *function* $X : \Omega \to \mathbb{R}$.

- X is a *discrete random variable* if and only if the *range* of X is countable.
- X is a *continuous random variable* if and only if the *range* of X is uncountable.

Useful Notation for Random Variables

Some conventional notation:

- $[X = a] = \{\omega \in \Omega | X(\omega) = a\}$
- $[X < a] = \{\omega \in \Omega | X(\omega) < a\}$
- $[a < X < b] = \{\omega \in \Omega | a < X(\omega) < b\}$, etc

Random Variables

Random Variable (Daners, 2012, p. 143)

A random variable X is a *function* $X : \Omega \to \mathbb{R}$.

- X is a *discrete random variable* if and only if the *range* of X is countable.
- X is a *continuous random variable* if and only if the *range* of X is uncountable.

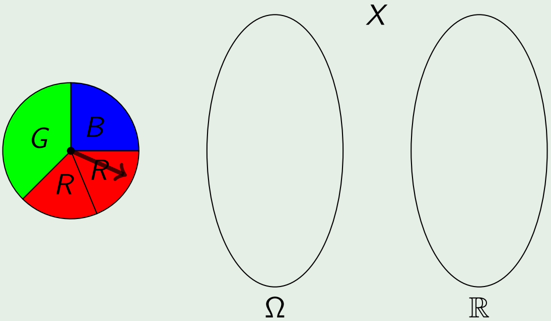
Useful Notation for Random Variables

Some conventional notation:

- $[X = a] = \{\omega \in \Omega | X(\omega) = a\}$
- $[X < a] = \{\omega \in \Omega | X(\omega) < a\}$
- $[a < X < b] = \{\omega \in \Omega | a < X(\omega) < b\}$, etc
- More generally: $[X \in S] = \{\omega \in \Omega | X(\omega) \in S\}$

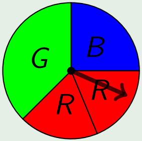
Example of Discrete Random Variable

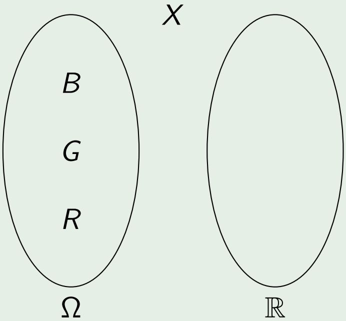
Example



Example of Discrete Random Variable

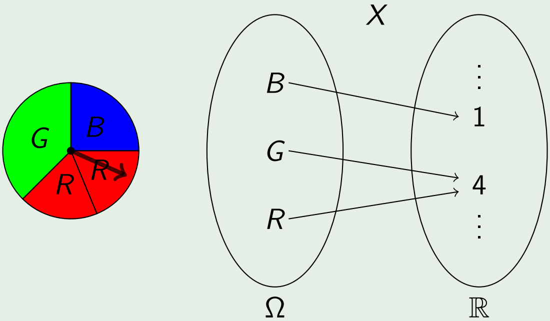
Example





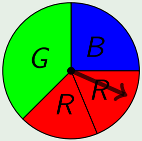
Example of Discrete Random Variable

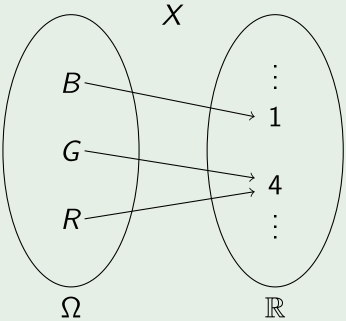
Example



Example of Discrete Random Variable

Example

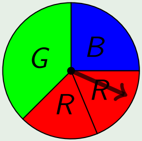


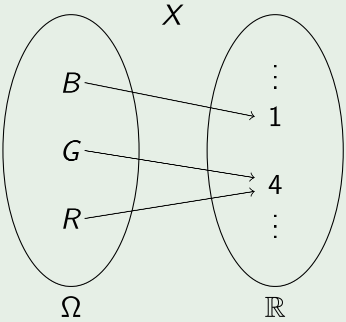


- Range: $\{1, 4\}$

Example of Discrete Random Variable

Example

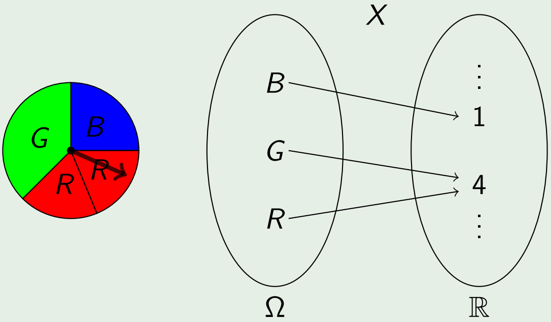




- Range: $\{1, 4\}$
- $[X = 4] = \{G, R\}$

Example of Discrete Random Variable

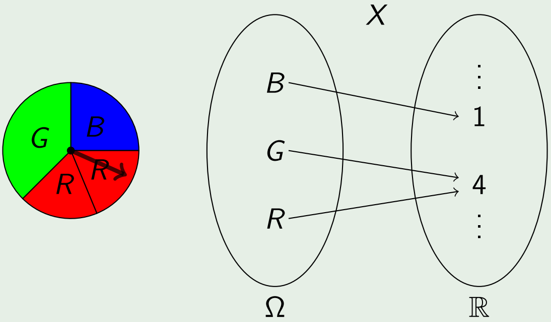
Example



- Range: $\{1, 4\}$
- $[X = 4] = \{G, R\}$
- $[X > -1] = \{B, G, R\}$

Example of Discrete Random Variable

Example



- Range: $\{1, 4\}$
- $[X = 4] = \{G, R\}$
- $[X > -1] = \{B, G, R\}$
- $[X > 100] = \{\}$

Example of Continuous Random Variable

Example

A machine with a slider moving between the values of -2 and 2 inclusive prints a value equal to the square of where the slider lands.

Example of Continuous Random Variable

Example

A machine with a slider moving between the values of -2 and 2 inclusive prints a value equal to the square of where the slider lands.
$X : [-2, 2] \to \mathbb{R}$
$X(\omega) = \omega^2$

Example of Continuous Random Variable

Example

A machine with a slider moving between the values of -2 and 2 inclusive prints a value equal to the square of where the slider lands.
$X : [-2, 2] \to \mathbb{R}$
$X(\omega) = \omega^2$

- Range: $[0, 4]$

Example of Continuous Random Variable

Example

A machine with a slider moving between the values of -2 and 2 inclusive prints a value equal to the square of where the slider lands.
$X : [-2, 2] \to \mathbb{R}$
$X(\omega) = \omega^2$

- Range: $[0, 4]$
- $[X = 4] = \{2, -2\}$

Example of Continuous Random Variable

Example

A machine with a slider moving between the values of -2 and 2 inclusive prints a value equal to the square of where the slider lands.
$X : [-2, 2] \to \mathbb{R}$
$X(\omega) = \omega^2$

- Range: $[0, 4]$
- $[X = 4] = \{2, -2\}$
- $[1 < X \leq 3] = [-\sqrt{3}, -1) \cup (1, \sqrt{3}]$

Why Use Random Variables?

Why Use Random Variables?

- An advantage of using a Random Variable over just working with Sample Spaces (like in Stage 4 Mathematics) is that it allows us to focus on relevant results in the experiment.

Why Use Random Variables?

- An advantage of using a Random Variable over just working with Sample Spaces (like in Stage 4 Mathematics) is that it allows us to focus on relevant results in the experiment.
- In the example of the spinner, let's say spinning a B awards us with \$1 and spinning a R or G awards us with \$4.

Why Use Random Variables?

- An advantage of using a Random Variable over just working with Sample Spaces (like in Stage 4 Mathematics) is that it allows us to focus on relevant results in the experiment.
- In the example of the spinner, let's say spinning a B awards us with \$1 and spinning a R or G awards us with \$4.
- We don't actually care about getting B, G or R - we care about the money!

Table of Contents

Probability of Random Variables

- Recall that a probability measure is a function $P : \mathscr{F} \to [0, 1]$ where $\mathscr{F}$ is the set of all possible Events.

Probability of Random Variables

- Recall that a probability measure is a function $P : \mathscr{F} \to [0, 1]$ where $\mathscr{F}$ is the set of all possible Events.
- The notation we introduced earlier, e.g $[X = 4]$ is what we "sub in" to this function P.

Probability of Random Variables

- Recall that a probability measure is a function $P : \mathscr{F} \to [0,1]$ where $\mathscr{F}$ is the set of all possible Events.
- The notation we introduced earlier, e.g $[X = 4]$ is what we "sub in" to this function P.

Example

Continuing with the discrete random variable example of the spinner:

$$P([X = 4]) = P(\{G, R\})$$

Probability of Random Variables

- Recall that a probability measure is a function $P : \mathscr{F} \to [0,1]$ where $\mathscr{F}$ is the set of all possible Events.
- The notation we introduced earlier, e.g $[X = 4]$ is what we "sub in" to this function P.

Example

Continuing with the discrete random variable example of the spinner:

$$P([X = 4]) = P(\{G, R\})$$

Conventionally, writing (followed by [and remembering to close them off is annoying, so we shorten the left side to just $P(X = 4)$.

Probability of Random Variables

- Recall that a probability measure is a function $P : \mathscr{F} \to [0, 1]$ where $\mathscr{F}$ is the set of all possible Events.
- The notation we introduced earlier, e.g $[X = 4]$ is what we "sub in" to this function P.

Example

Continuing with the discrete random variable example of the spinner:

$$P([X = 4]) = P(\{G, R\})$$

Conventionally, writing (followed by [and remembering to close them off is annoying, so we shorten the left side to just $P(X = 4)$. Recall that $P(\{G, R\}) = P(\{G\}) + P(\{R\})$.

Probability of Random Variables

- Recall that a probability measure is a function $P : \mathscr{F} \to [0, 1]$ where $\mathscr{F}$ is the set of all possible Events.
- The notation we introduced earlier, e.g $[X = 4]$ is what we "sub in" to this function P.

Example

Continuing with the discrete random variable example of the spinner:

$$P([X = 4]) = P(\{G, R\})$$

Conventionally, writing (followed by [and remembering to close them off is annoying, so we shorten the left side to just $P(X = 4)$. Recall that $P(\{G, R\}) = P(\{G\}) + P(\{R\})$. Hence,
$P(X = 4) = P(\{G, R\}) = \frac{3}{8} + \frac{3}{8} = \frac{3}{4}$.

Probability Distributions

As stated before, Random Variables should just allow us to focus on results in an experiment, without much care about the Sample Space.

Probability Distributions

As stated before, Random Variables should just allow us to focus on results in an experiment, without much care about the Sample Space. To do this, we can define the Probability Measure to directly interact with the range of the Random Variable (the theory behind how and why this works is beyond the scope of this presentation):

Probability Distributions

As stated before, Random Variables should just allow us to focus on results in an experiment, without much care about the Sample Space. To do this, we can define the Probability Measure to directly interact with the range of the Random Variable (the theory behind how and why this works is beyond the scope of this presentation):

Discrete Random Variable Probabilities

A *Probability Distribution Function* that assigns a probability to each $P(X = k)$ for each k in the range of X is the simplest approach when dealing with Discrete Random Variables.

Probability Distributions

As stated before, Random Variables should just allow us to focus on results in an experiment, without much care about the Sample Space. To do this, we can define the Probability Measure to directly interact with the range of the Random Variable (the theory behind how and why this works is beyond the scope of this presentation):

Discrete Random Variable Probabilities

A *Probability Distribution Function* that assigns a probability to each $P(X = k)$ for each k in the range of X is the simplest approach when dealing with Discrete Random Variables.

Example

As with the spinner, we can measure the probability of the Random Variable with:

k	1	4	Everything else (usually not written)
$P(X = k)$	$\frac{1}{4}$	$\frac{3}{4}$	0

Probability Distributions

Continuous Random Variable Probabilities

A probability measure for a Continuous Random Variable X can be defined by:

$$P(a < X < b) = \int_a^b f(x)\, dx$$

where $f(x)$ is called the *probability density function*.

Note: $P(X = k) = 0$ for any $k \in \mathbb{R}$

Cumulative Distribution Function

Cumulative Distribution Function

The Cumulative Distribution Function $F(x)$ of a continuous random variable X with probability density function $f(x)$ is

$$F(x) = P(X < x) = \int_{-\infty}^{x} f(x)\, dx$$

It accumulates the total of probabilities in the distribution.

Cumulative Distribution Function

Cumulative Distribution Function

The Cumulative Distribution Function $F(x)$ of a continuous random variable X with probability density function $f(x)$ is

$$F(x) = P(X < x) = \int_{-\infty}^{x} f(x)\, dx$$

It accumulates the total of probabilities in the distribution.

This allows us to find quantiles easily. For example the median can be found by solving for k in $F(k) = 0.5$.

Cumulative Distribution Function

Cumulative Distribution Function

The Cumulative Distribution Function $F(x)$ of a continuous random variable X with probability density function $f(x)$ is

$$F(x) = P(X < x) = \int_{-\infty}^{x} f(x)\, dx$$

It accumulates the total of probabilities in the distribution.

This allows us to find quantiles easily. For example the median can be found by solving for k in $F(k) = 0.5$.
Note also that by the Fundamental Theorem of Calculus, differentiating the cumulative distribution function $F(x)$ will yield the probability density function $f(x)$.

Expectation and Variance

Often researchers are interested in finding the average (mean) result of a random phenomenon and the spread of the results.

Expectation

For Discrete Random Variables:

$$\mu = E(X) = \sum_{x \in X} xP(X = x)$$

For Continuous Random Variables (Note: NOT IN SYLLABUS):

$$\mu = E(X) = \int_{-\infty}^{\infty} xf(x)\, dx$$

Note that if a and b are constants:

$$E(a + bX) = a + bE(X)$$

Variance

Variance

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

By expanding and simplifying, we can find a simpler formula:

$$\begin{aligned} Var(X) &= E(X^2 - 2X\mu + \mu^2) \\ &= E(X^2) - E(2X\mu) + E(\mu^2) \\ &= E(X^2) - 2\mu E(X) + \mu^2 = E(X^2) - 2\mu^2 + \mu^2 \\ &= E(X^2) - \mu^2 = E(X^2) - E(X)^2 \end{aligned}$$

Standard Deviation

Often to *standardise* the squared units of the Variance to units (not squared), the standard deviation is used:

$$\sigma = \sqrt{Var(X)}$$

Important Considerations

- If a and b are constants: $Var(a + bX) = b^2 Var(X)$

Important Considerations

- If a and b are constants: $Var(a + bX) = b^2 Var(X)$
- Why $E((X - \mu)^2)$ and not $E(X - \mu)$ to measure spread?

Important Considerations

- If a and b are constants: $Var(a + bX) = b^2 Var(X)$
- Why $E((X - \mu)^2)$ and not $E(X - \mu)$ to measure spread?
- Why not $E(|X - \mu|)$?

Important Considerations

- If a and b are constants: $Var(a + bX) = b^2 Var(X)$
- Why $E((X - \mu)^2)$ and not $E(X - \mu)$ to measure spread?
- Why not $E(|X - \mu|)$?
- What is the link between mean and variance in Random Variables with the mean and variance used in Statistics?

Special Distributions Studied in HSC

Some special distributions studied in the HSC are listed below. There is not enough time to cover them in this presentation as they can each take up an entire session in themselves.

- Discrete Probability Distributions

Special Distributions Studied in HSC

Some special distributions studied in the HSC are listed below. There is not enough time to cover them in this presentation as they can each take up an entire session in themselves.

- Discrete Probability Distributions
    - Discrete Uniform Distribution

Special Distributions Studied in HSC

Some special distributions studied in the HSC are listed below. There is not enough time to cover them in this presentation as they can each take up an entire session in themselves.

- Discrete Probability Distributions
    - Discrete Uniform Distribution
    - Bernoulli Distribution $X \sim B(p)$

Special Distributions Studied in HSC

Some special distributions studied in the HSC are listed below. There is not enough time to cover them in this presentation as they can each take up an entire session in themselves.

- Discrete Probability Distributions
    - Discrete Uniform Distribution
    - Bernoulli Distribution $X \sim B(p)$
    - Binomial Distribution $X \sim Bin(n, p)$

Special Distributions Studied in HSC

Some special distributions studied in the HSC are listed below. There is not enough time to cover them in this presentation as they can each take up an entire session in themselves.

- Discrete Probability Distributions
  - Discrete Uniform Distribution
  - Bernoulli Distribution $X \sim B(p)$
  - Binomial Distribution $X \sim Bin(n, p)$
  - Not in syllabus but good to look at: Geometric Distribution $X \sim Geo(p)$

Special Distributions Studied in HSC

Some special distributions studied in the HSC are listed below. There is not enough time to cover them in this presentation as they can each take up an entire session in themselves.

- Discrete Probability Distributions
  - Discrete Uniform Distribution
  - Bernoulli Distribution $X \sim B(p)$
  - Binomial Distribution $X \sim Bin(n, p)$
  - Not in syllabus but good to look at: Geometric Distribution $X \sim Geo(p)$
- Continuous Probability Distributions

Special Distributions Studied in HSC

Some special distributions studied in the HSC are listed below. There is not enough time to cover them in this presentation as they can each take up an entire session in themselves.

- Discrete Probability Distributions
  - Discrete Uniform Distribution
  - Bernoulli Distribution $X \sim B(p)$
  - Binomial Distribution $X \sim Bin(n, p)$
  - Not in syllabus but good to look at: Geometric Distribution $X \sim Geo(p)$
- Continuous Probability Distributions
  - Continuous Uniform Distribution

Special Distributions Studied in HSC

Some special distributions studied in the HSC are listed below. There is not enough time to cover them in this presentation as they can each take up an entire session in themselves.

- Discrete Probability Distributions
  - Discrete Uniform Distribution
  - Bernoulli Distribution $X \sim B(p)$
  - Binomial Distribution $X \sim Bin(n, p)$
  - Not in syllabus but good to look at: Geometric Distribution $X \sim Geo(p)$
- Continuous Probability Distributions
  - Continuous Uniform Distribution
  - Normal Distribution $X \sim N(\mu, \sigma^2)$

Special Distributions Studied in HSC

Some special distributions studied in the HSC are listed below. There is not enough time to cover them in this presentation as they can each take up an entire session in themselves.

- Discrete Probability Distributions
  - Discrete Uniform Distribution
  - Bernoulli Distribution $X \sim B(p)$
  - Binomial Distribution $X \sim Bin(n, p)$
  - Not in syllabus but good to look at: Geometric Distribution $X \sim Geo(p)$
- Continuous Probability Distributions
  - Continuous Uniform Distribution
  - Normal Distribution $X \sim N(\mu, \sigma^2)$
- Covered in HSC Science Extension
  - t Distribution
  - χ^2 Distribution
  - F Distribution
- Not covered in any HSC syllabus but really cool for exploration
  - Poisson Distribution

Table of Contents

Double Dice

Two 6-sided dice are thrown and the result on the top face are added.

1. What is the sample space?
2. How can we define a random variable here that is appropriate to the scenario?
3. What subset of the sample space does $[X < \pi]$ represent?
4. Find $P(X < \pi)$.

A Geometric Distribution Example

Let X be a random variable that represents the number of tosses of a four-sided die until a 3 is thrown.

1. What is the sample space?
2. How can we define a random variable that represents the scenario?
3. Find $P(X = 1)$, $P(X = 2)$, $P(X = x)$.
4. Find $E(X)$.

Independent Binomial Probability

John tosses a coin 4 times and independently Paul tosses a coin 4 times.

1. How can we define the random variables for John and Paul's results?
2. What distribution do these random variables have?
3. In random variable notation, how can we represent the probability that they toss the same number of heads?
4. Find the probability that they toss the same number of heads.

Note: for numbers higher than 4, you may need to prove binomial identities such as $\sum_{k=0}^{n} \binom{n}{k}^2 = \binom{2n}{k}$

Distribution Functions

A continuous random variable is distributed according to the probability density function $f(x)$ such that:

$$f(x) = \begin{cases} kx^2 + 1 \text{ if } 0 \leq x \leq 3 \\ 0 \text{ otherwise} \end{cases}$$

where k is some constant.

1. Find the value of k.
2. Find the cumulative distribution function.

Table of Contents

Bibliography

Daners, D. (2012). Measure theory and fourier analysis - lecture notes for MATH3969. School of Mathematics and Statistics, The University of Sydney.

NSW Education Standards Authority. (2017). Mathematics advanced stage 6 syllabus. NSW Education Standards Authority.

Tao, T. (2011). An introduction to measure theory. *Graduate Studies in Mathematics*, *126*, 192.

Tao, T. (2014). *Analysis i* (3rd ed.). Hindustan Book Agency (India).

Source files and PDF for this presentation can be found on: https://github.com/moksifu/aisnsw_focus_day_2021 under the MIT Licence.