COGS2020

TUTORIAL 5: PROBABILITY AND RANDOM VARIABLES

Probability Key Terms

- **Sample space** – set of all possible outcomes
- **Outcome** – single *result* of an experiment
- **Event** – a set of outcomes (results) from an experiment
- **Probability of an event** – likelihood of an event,

  0 = no likelihood, 1 = 100% likelihood

Probability Basics

Assumptions/Axioms:

1. Non-negativity: $P(A) \geq 0$
2. Normalisation: $P(S) = 1$
3. Additivity: $P(A \cup B) = P(A) + P(B)$ if A and B are mutually exclusive

Therefore:

$P(\emptyset)=0$, probability of getting nothing, or an outcome outside of S, is 0

P(not A) = 1 - P(A), probability of getting anything but A, is 1 – A

P(A and B) = P(A)*P(B)

P(A or B) = P(A) + P(B) − P(A and B), accounts for the overlap of outcomes between A and B (if any)

Axiom 1: Non-Negativity

The probability of any event A is always **non-negative**:

- $P(A) \geq 0$

This means probabilities cannot be negative and every event has a probability that is at least zero.

Axiom 2: Normalization

The probability of the **entire sample space** (S) is always **1**:

- $P(S) = 1$

This ensures that **something must happen** — the total probability of all possible outcomes is 1.

Axiom 3: Additivity

If two events A and B are **mutually exclusive** (i.e., they cannot occur together):

- $P(A \cup B) = P(A) + P(B)$

This states that the probability of **either** A **or** B happening is simply the sum of their individual probabilities.

∪ = union, or

Probability of the Empty Set

The empty set $\emptyset$ contains **no outcomes**, so:

- $P(\emptyset) = 0$

Example: The probability of flipping a coin getting neither “Heads” nor “Tails” is 0.

Complement Rule

- The probability of **not** A (denoted A^c) is:
  - $P(A^c) = 1 - P(A)$

- The chance that something **does *not*** happen is equal to 1 minus the chance that it *does* happen.

Let's say you flip a coin.

- The chance of getting **Heads** = P(Heads) = 0.5
- So the chance of **not getting Heads** (i.e., getting Tails) =

P(Not Heads) = 1 − 0.5 = 0.5

Complement Rule

In Plain English:

- **P(A)** = The probability of event A happening.
- **P(A^c)** = The probability of event A **not** happening.

- Since something **either happens or it doesn't**, the two probabilities **must add up to 1**.

There's a 30% chance it will rain today.

- So the chance it **won't** rain is:
- 1 - 0.3 = 0.7 (or 70%)

Inclusion-Exclusion (Overlapping Events)

- If A and B are **not** mutually exclusive:

  - $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

P(A∪B) = Probability that A or B happens (or both).

P(A) = Probability that A happens.

P(B) = Probability that B happens.

P(A∩B) = Probability that both A and B happen at the same time.

Why Subtract P(A∩B)?

- When we add P(A) + P(B), we double-count the part where A and B both happen.
  So, we subtract P(A∩B) to fix that.

Inclusion-Exclusion (Overlapping Events)

P(A or B) = P(A) + P(B) − P(A and B)

Example (Real-Life Style):

Imagine we ask a group of people:

- 40% like pizza → P(A) = 0.3 + 0.1 = 0.4
- 30% like burgers → P(B) = 0.2 + 0.1 = 0.3
- 10% like both → $P(A \cap B) = 0.1$

Sample Space	Likes pizza (A)	Likes burgers (B)	Probability
Likes pizza only	✅ Yes	❌ No	0.3
Like burgers only	❌ No	✅ Yes	0.2
Likes both	✅ Yes	✅ Yes	0.1
Likes none	❌ No	❌ No	0.4

So, what's the chance someone likes pizza OR burgers (or both)?

- $P(A \cup B) = 0.4 + 0.3 - 0.1 = 0.6$
- So, there's a 60% chance a random person likes either pizza, burgers, or both.

Probability of Independent Events

- If two events A and B are independent (one does not affect the other):
  - $P(A \cap B) = P(A) \text{ x } P(B)$

P(A∩B) = The chance that both A and B happen.

P(A) = The chance that A happens.

P(B) = The chance that B happens.

- Independent events = Knowing whether A happens tells you nothing about whether B happens (and vice versa).

∩ = intersection, and

Probability of Independent Events

Imagine:

- You **flip a coin**: the chance of **Heads** is 0.5
- You **roll a die**: the chance of rolling a **4** is 1/6

These are **independent** — the coin flip doesn't affect the die roll.

So what's the chance of getting **Heads AND a 4**?

- P(Heads and 4)= P(Heads) × P(4) = 0.5 × 1/6 = 1/12
- So there's a **1 in 12** chance both happen at the same time.

Random Variables (and how they’re related to probability)

- A random variable is a “process” that generates random outcomes
- Process is tied to a population and defined by a probability distribution
  - Outcomes are defined by a sample space (all potential outcomes) and the probability of getting each outcome (visualised in a probability distribution)
- Therefore, random variables and their “behaviour” can be characterised by a probability distribution

Probability Distribution Functions

- Probability distribution *functions* describe the probability of obtaining different values (of the sample space) from a random variable
- In other words, they are **graphs that describe the behaviour of probability distributions**

- Note: You can get all the same info from the original probability distribution graph, but it is easier to understand if we capture it in a different way
- All the functions on the next few slides can be made for each type of probability distribution

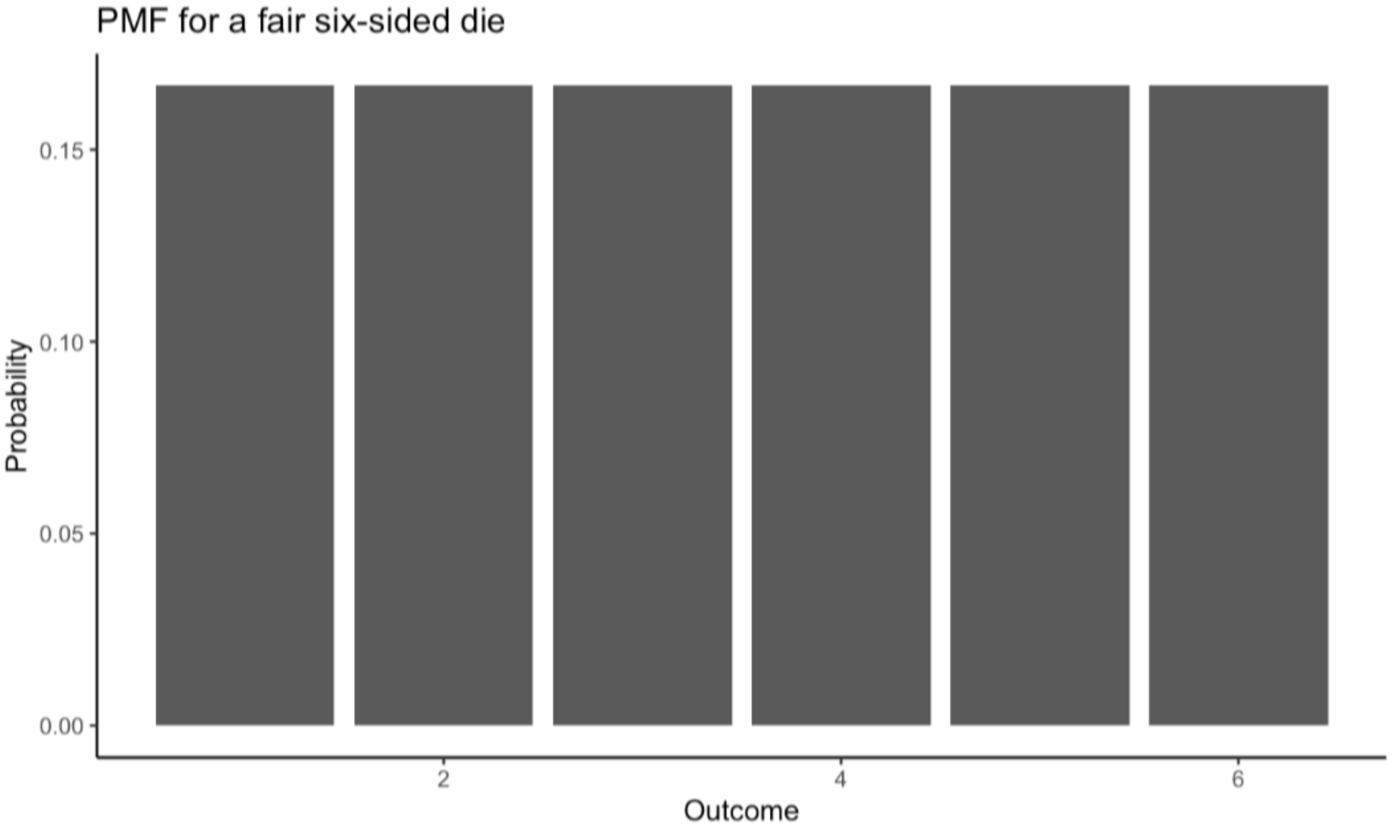
Types of Distribution Functions

Probability Mass Function (PMF)

Probability of each of the outcomes

Sometimes notated as P(X=x)

Big X = random variable
Little x = actual specific outcome



Sample space – all possible outcomes
Outcome sometimes notated as x (little x)

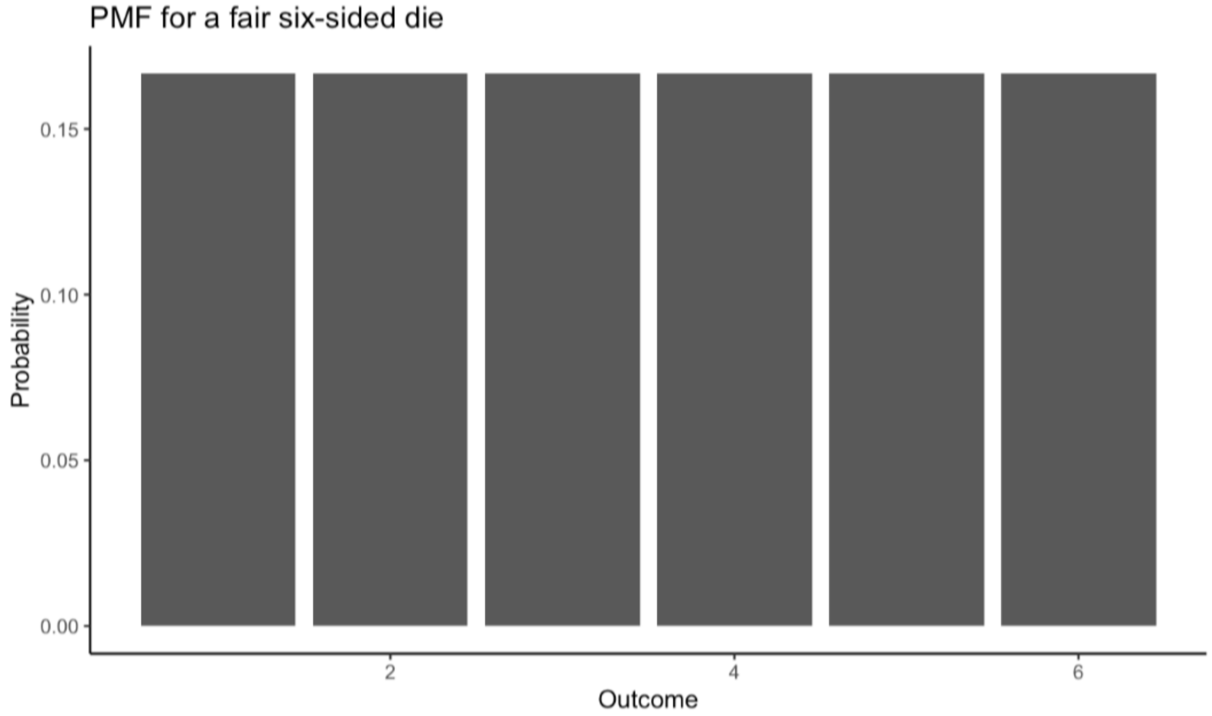
- **Probability Mass Function** – for ***discrete*** random variables, function that tells us the probability/likelihood of each (countable) outcome

Calculating probability in PMF

Probability Mass Function (PMF)

- Can count and add the probability of each singular outcome

Example:



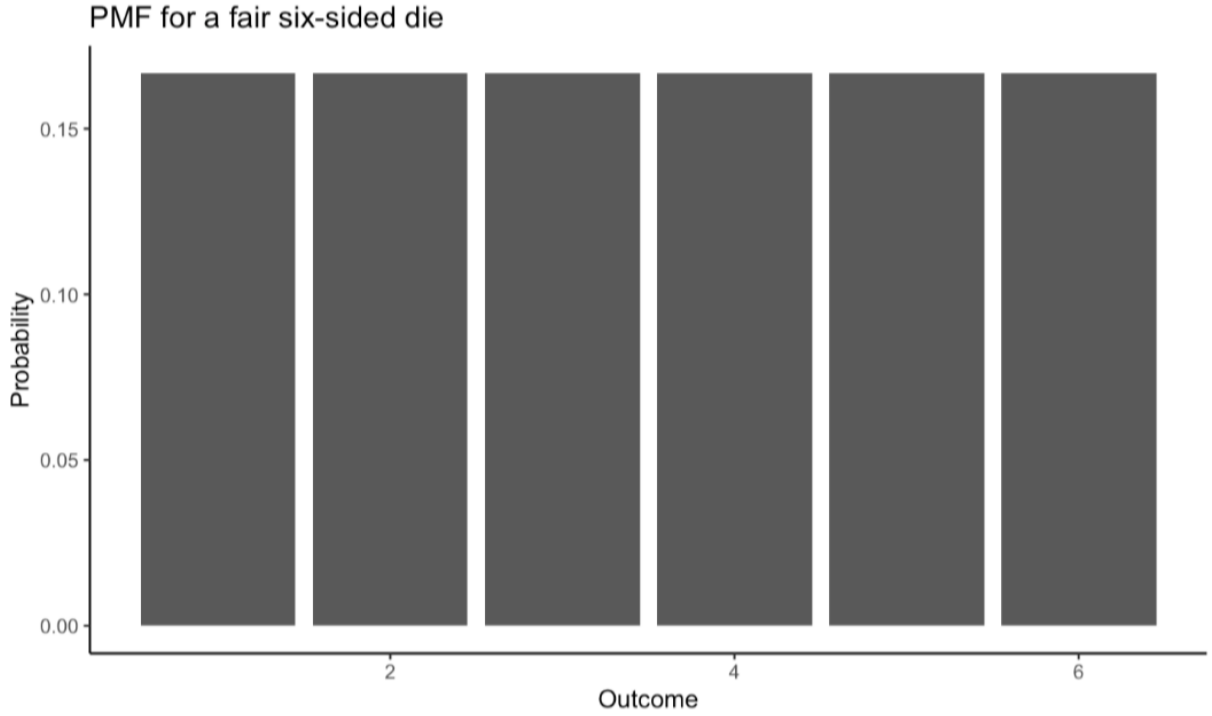
What is the probability of rolling a number greater than 2. Or in different terms, $P(X>2)$?

Calculating probability in PMF

Probability Mass Function (PMF)

- Can count and add the probability of each singular outcome

Example:



What is the probability of rolling a number greater than 2. Or in different terms, P(X>2)?

$$P(X > 2) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6}$$
$$= \frac{4}{6} = \frac{2}{3}$$

Probability Mass Function

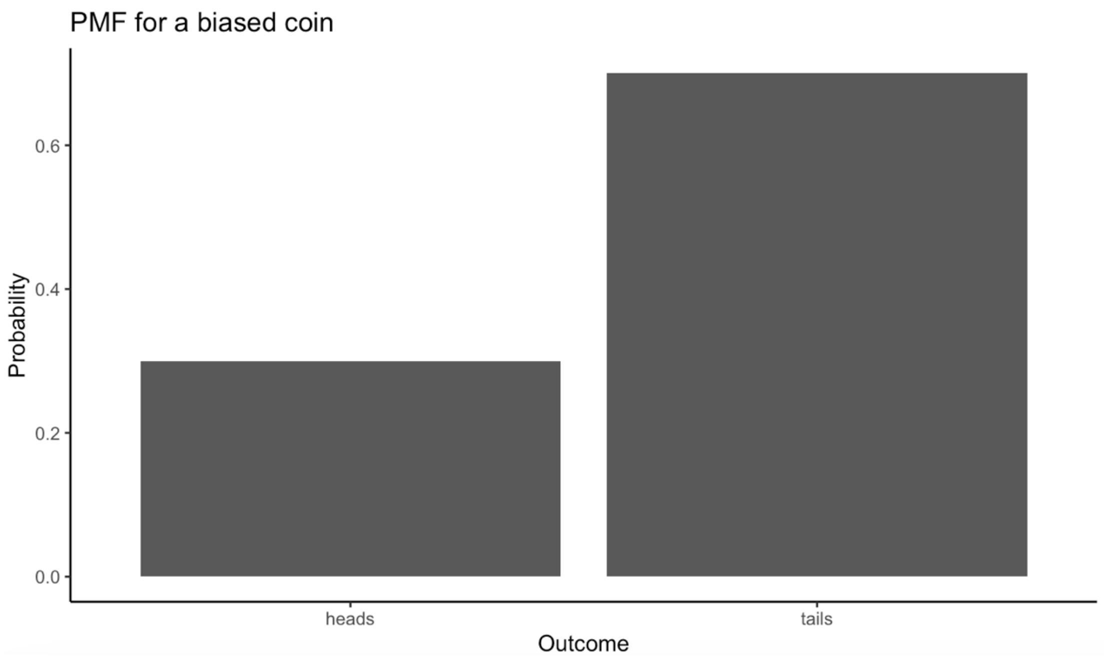
For a discrete random variable, the probability mass function (PMF) gives the probability of each outcome.

$$P(X = x) = \begin{cases} 0.3 & \text{if } x = \text{heads} \\ 0.7 & \text{if } x = \text{tails} \end{cases}$$

P(X=x) = "What's the probability that the random variable X equals some value x?"

A PMF tells you **how likely each outcome is**.

This one says that if you flip this particular **biased coin**, there's a **30% chance of heads** and a **70% chance of tails**.



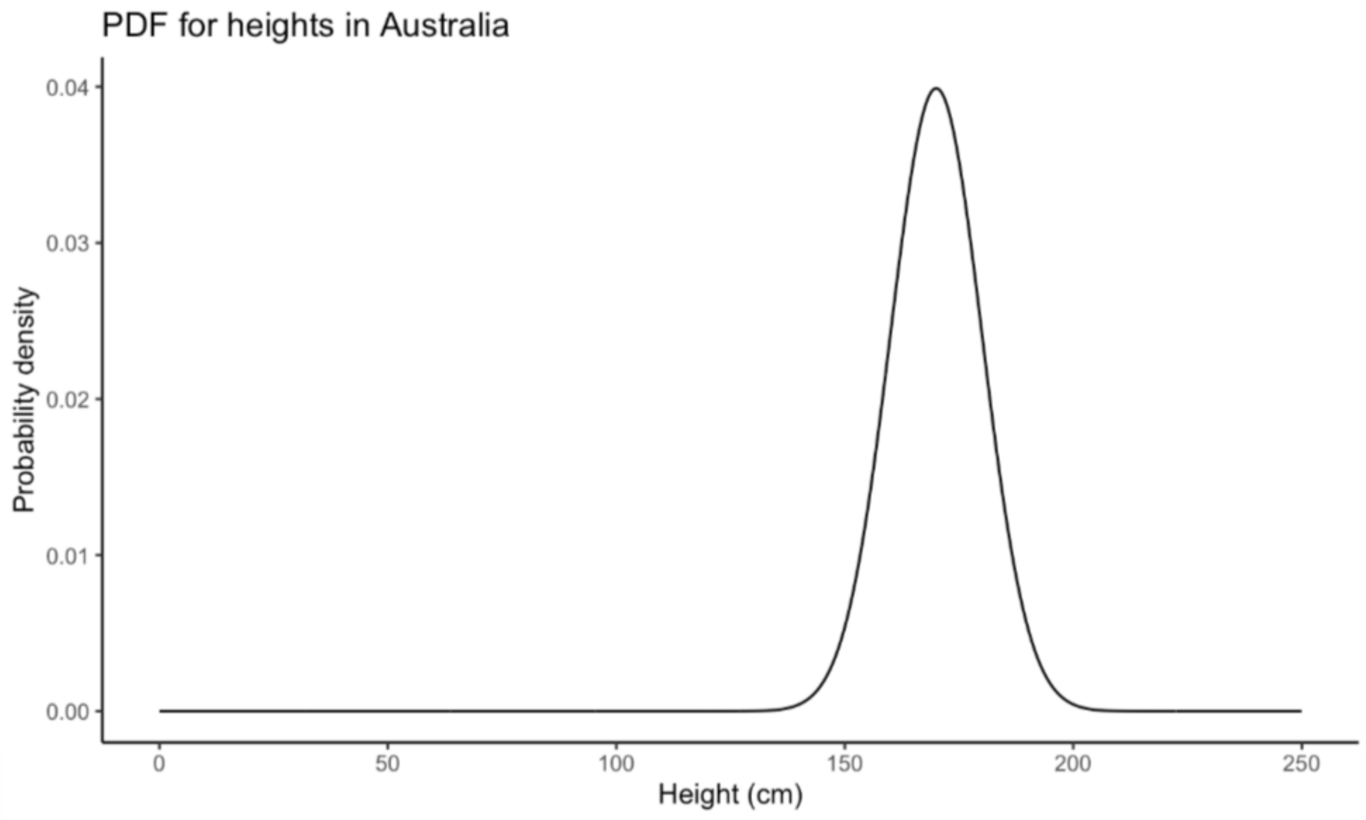
Types of Distribution Functions

Probability Density Function (PDF)

Probability of each of the outcomes

Sometimes notated as f(x)

f(x) representing function of x



Sample space – all possible outcomes
Outcome sometimes notated as x (little x)

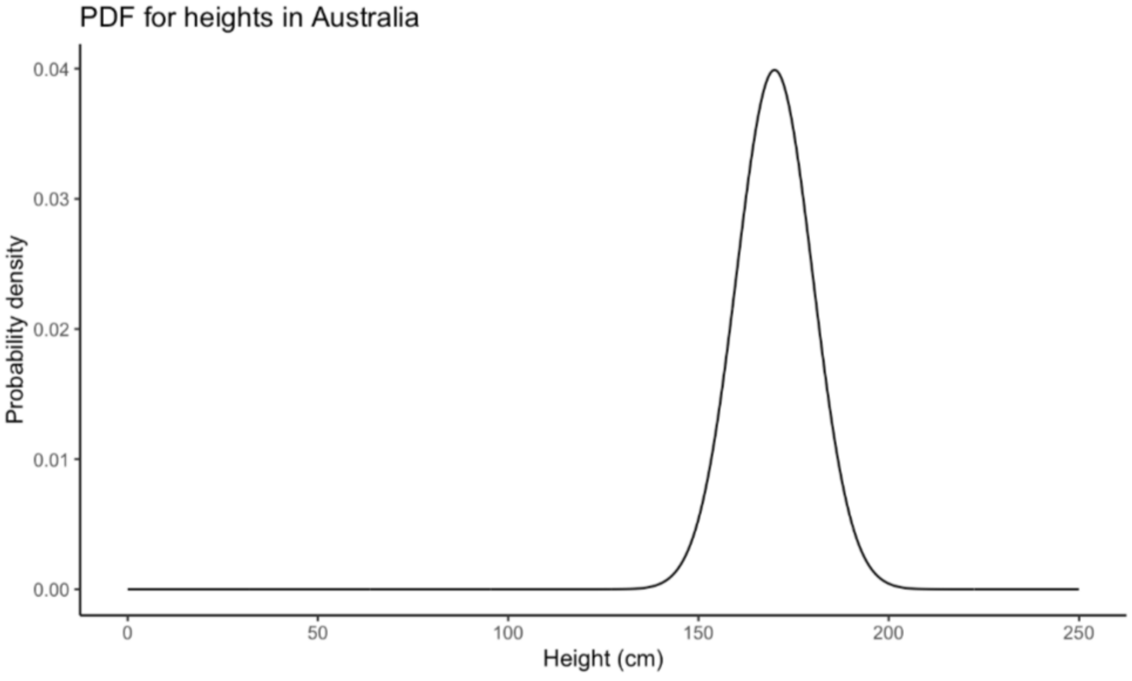
- **Probability Density Function** – for ***continuous*** random variables, function that gives probability *density* of each (uncountable) outcome

Calculating probability in PDF

Probability Density Function (PDF)

- Can NOT count and add the probability of each singular outcome – need to calculate area under the curve

Example:



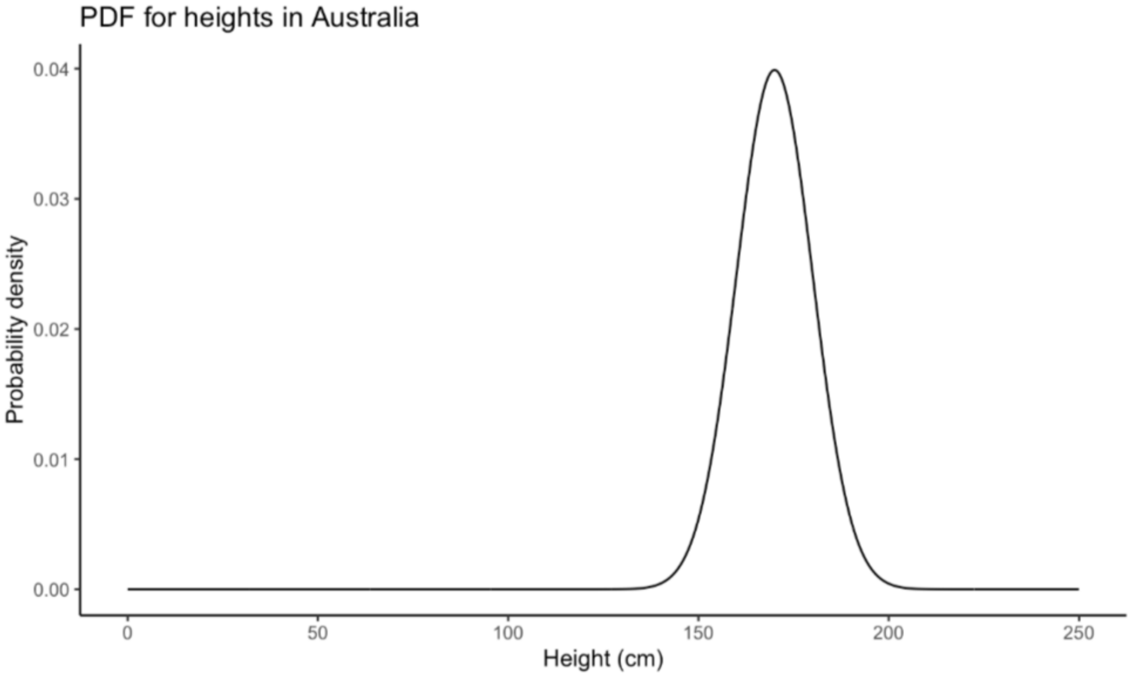
What is the probability that someone's height (from this population) will be greater than 175cm, or in different terms P(X > 175)?

Calculating probability in PDF

Probability Density Function (PDF)

- Can NOT count and add the probability of each singular outcome – need to calculate area under the curve

Example:

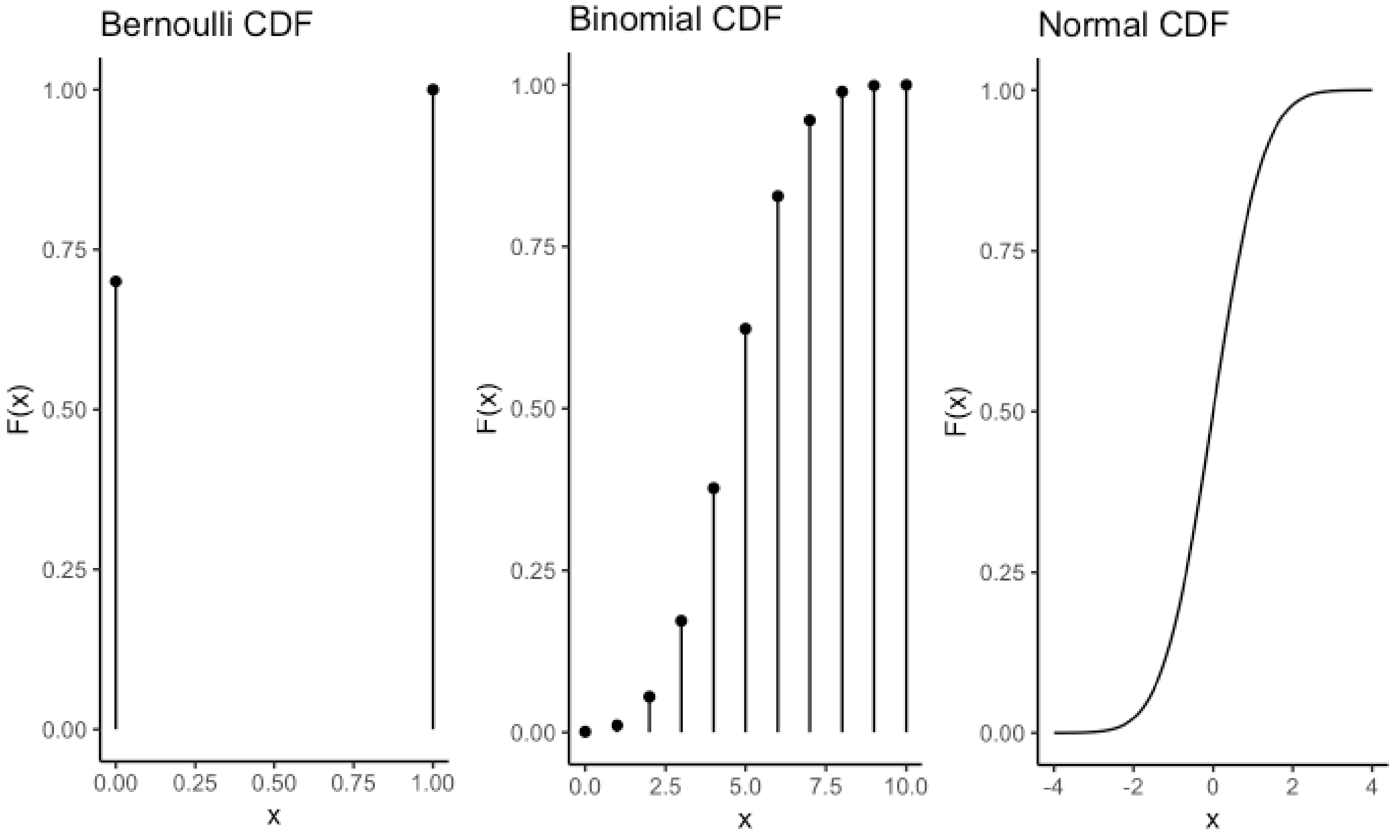


What is the probability that someone's height (from this population) will be greater than 175cm, or in different terms P(X > 175)?

Unable to calculate by hand (without fancy maths), but can calculate using R!

Types of Distribution Functions

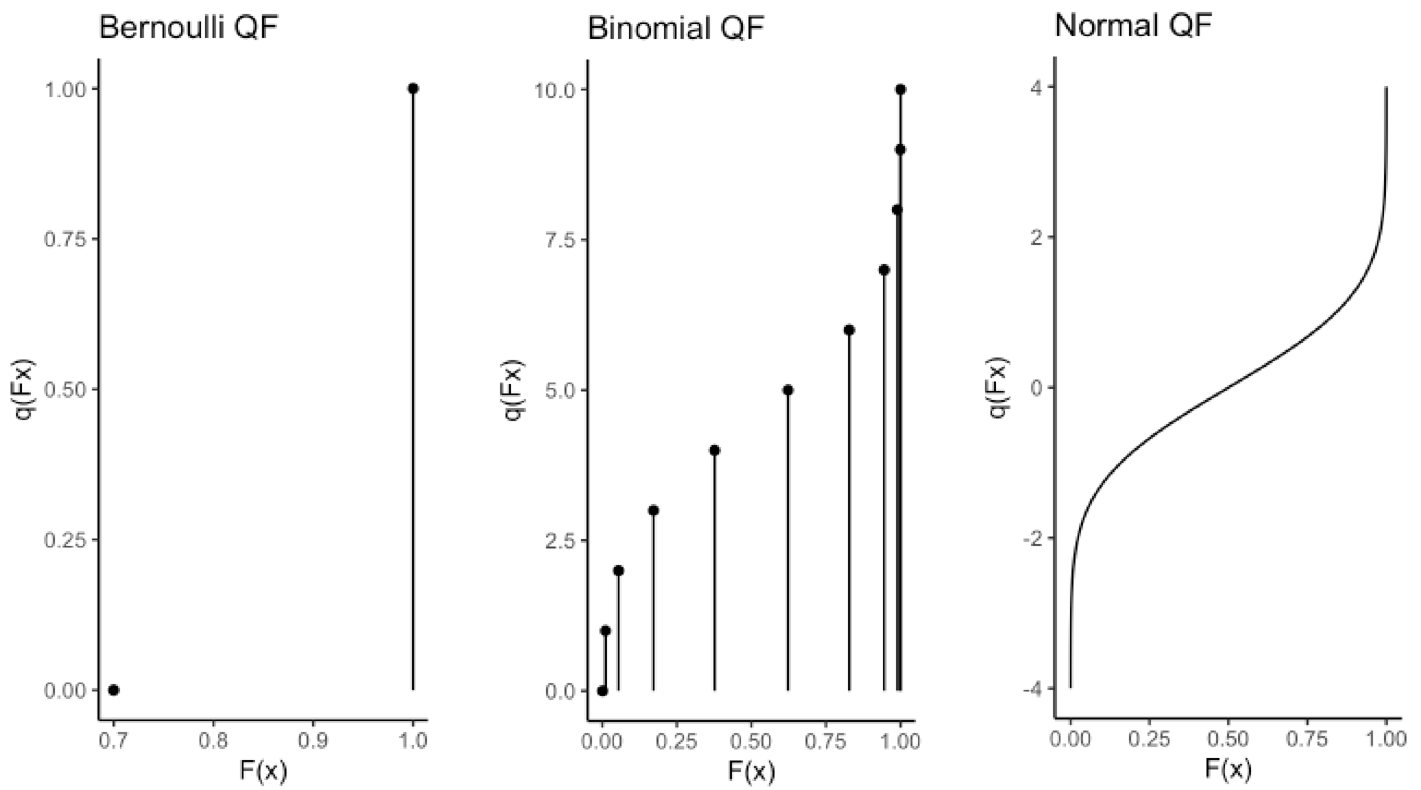
Cumulative Distribution Function (CDF)



- **Cumulative Distribution Function** – gives the cumulative probability of getting values less than or equal to a specific value
- Represents area under the curve in a function form

Types of Distribution Functions

Quantile Function (QF)



- **Quantile Function** – gives the value that corresponds to a specified probability/percentile
- Inverse of a CDF – gives X value instead of probability

R functions used for probability calculations

Calculating probability of specific values:

- dbinom(x, n, p), PMF of binomial distribution
- dnorm(x, mean, sd, lower.tail = T/F), PDF of normal distribution *

Calculating probability of a range of values:

- pbinom(x, n, p, lower.tail = T/F), CDF of binomial dist, P(X ≤ x) | P(X > x)
- pnorm(x, mean, sd, lower.tail = T/F), CDF of normal distribution

Calculating specific values that correspond to a probability/percentile:

- qbinom(q, n, p, lower.tail = T/F)
- qnorm(q, mean, sd, lower.tail = T/F)

** Note that because we cannot count each and every single outcome of a norm dist, trying to isolate a particular point on a PDF is not recommended. The probability of getting any single particular outcome in a continuous random variable is so small, that it is practically 0. P(X=x)=0*

Why do probability distributions matter in statistics/research?

- Probability distributions are **used to model the population** in some way, and **tells us how data is expected to behave**
- Moments are “descriptives” or characteristics of a probability distribution – these moments are **key characteristics that define/determine the form of the distribution**
- Moments (e.g. expected value, variance, etc.) can be used to estimate outcomes, and run statistical/hypothesis tests…

Types of Moments/Descriptives

Note: these are the moments of a random variable that form a normal distribution.

$X \sim N(\mu_X, \sigma_X^2)$

- Expected Value, E(X) or μ_X

Binomial

$$E(X) = \sum_{i=1}^{n} x_i P(x_i) = \mu_X$$

Normal

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

- Variance, Var(X) or σ_X^2

Binomial

$$\mathrm{Var}(X) = \sum_{i=1}^{n} (x_i - \mu_X)^2 P(x_i) = \sigma_X^2$$

Normal

$$\mathrm{Var}(X) = \int_{-\infty}^{\infty} (x - \mu_X)^2 f(x)\, dx = \sigma_X^2$$

Using these moments, we can make a probability distribution and model the population in some way .... More on this next week!

Key Takeaways + What's next?

- Sample statistics (e.g. mean, sd, etc.) and graphs/plots (e.g. histogram, bar graph) made from sample data are all *estimates* of the true population statistics
- How does this relate to probability distributions?
  - Remember – probability distributions are used to model the population, so **moments of a probability distribution (e.g. expected value) are therefore used to model moments/descriptives of the population (e.g. population mean)**

....

- Next week will cover more on *how* probability distributions are used to model the population, and how that is relevant in null hypothesis testing!

Recommendation: check out old tutorial resources (tutorial 5 worksheet) to get hands-on practice with these concepts