STAT115: Introduction to Biostatistics

University of Otago

Ōtākou Whakaihu Waka

Lecture 5: Probability

Lecture Outline

- We've started interacting with data
- Data summaries: sample mean and standard deviation
- Summaries are limited
  - To go further we need statistical models
    - Use probability to describe the variation in the data
- Over the next few lectures we will look at probability
  - Start today with foundational knowledge
  - Much of this knowledge remains important even in complex applications

Probability: mathematical language of uncertain events

- What is the probability that:
  - A heart attack victim has an enlarged heart?
  - The All Black kicker is successful with their next kick?
  - A rat will choose one reward (out of many) when moving through a maze?
  - A person has a certain genotype?
  - A female skink is a breeder?
  - An earthquake of magnitude 5 or larger occurs this year?
  - A cancer patient will die within 12 months?
  - The sliced ham you got at the supermarket is safe to consume?

Probability

- Setup
  - Random process with a number of possible outcomes
    - Roll a die. Possible outcomes: 1, 2, 3, 4, 5, or 6
    - Flip a coin. Possible outcomes: head or tail
    - Blood group of randomly selected patient. Possible outcomes: A, B, AB, O
    - The set of all possible outcomes is called the *sample space*
- A probability has to satisfy a number of mathematical principles, including:
  - Between 0 and 1
    - We can't have a probability of -0.4 or 1.2
  - Probabilities sum to 1
    - If we observe the random process, we must see one of the possible outcomes
    - If we flip a coin, we must see either a head, or a tail.

Probability

- From here, things get a little murky
  - There are several definitions (or interpretations) of probability
- We can define probability in terms of relative frequency:
  - The probability of an outcome is the proportion of times the outcome occurs if we were to observe the random process a large (infinite) number of times.
    - Imagine a (bored!) person repeatedly tossing a coin[1]

[1] John Edmund Kerrich.

Mutually exclusive outcomes

- Two outcomes are mutually exclusive (or disjoint) if they cannot both happen
  - e.g. a coin flip cannot land on heads and tails
  - e.g. A patient cannot be both A and AB blood group
- The probability of mutually exclusive outcomes can be found with addition
  - For two outcomes A and B that are mutually exclusive:

$$\Pr(A \text{ or } B) = \Pr(A) + \Pr(B)$$

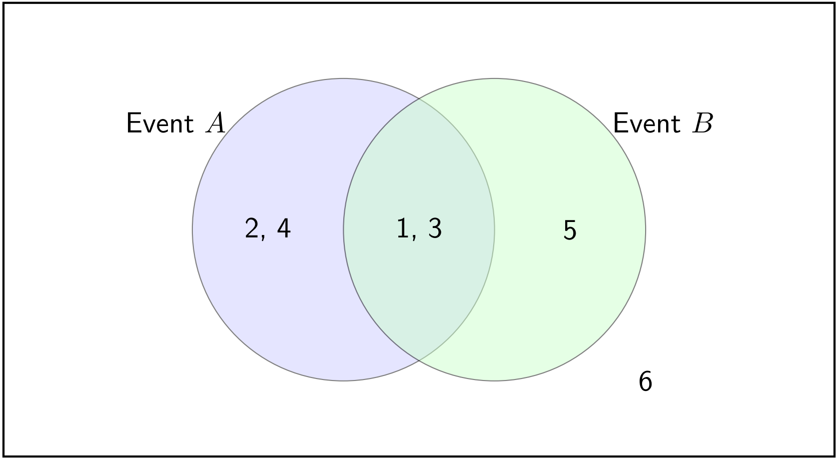
- Die roll: A: roll a 1, B: roll a 6
  - $\Pr(A \text{ or } B) = \Pr(A) + \Pr(B) = 1/6 + 1/6 = 1/3$
- Penguins: A: observe blood group A, B: observe blood group B
  - $\Pr(A \text{ or } B) = \Pr(A) + \Pr(B)$

Events

- We often work with collections of outcomes
  - These are called events
- Examples:
  - Die roll: event A: roll 1, 2, or 4,   event B: roll 5 or 6.
  - Blood group: event C: blood group A or AB,   event D: B or AB
- Events can be mutually exclusive if they have no outcomes in common
  - Events A and B are mutually exclusive
    - $\Pr(A \text{ or } B) = \Pr(A) + \Pr(B) = 3/6 + 2/6 = 5/6$
  - Events C and D are not mutually exclusive
    - What is $\Pr(C \text{ or } D)$?
- An event can comprise a single outcome
  - e.g. the event E: roll a 3

Venn diagram

- Venn diagrams can be used to visualize small sample spaces
- Die roll: event A: roll 4 or less, event B: roll odd number



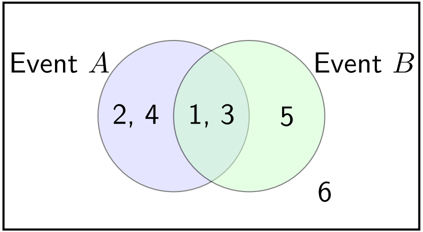
Venn diagram

[1] https://xkcd.com/2090/

Venn diagram and sets

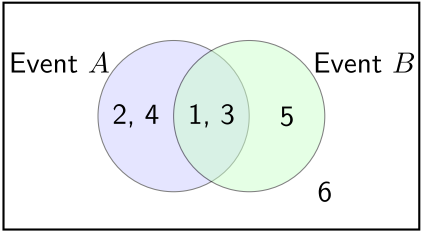
- Venn diagrams are useful when looking at when:
  - Event A or B occurs
    - This is inclusive, i.e. A or B means that event A, B or both A and B occur.
  - Event A and event B occurs

Sets of interest



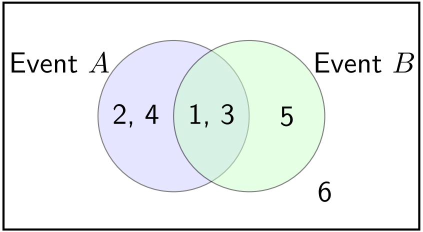
- A or B: 1, 2, 3, 4, 5
- A and B: 1, 3

Probabilities



- What is $\Pr$(A or B)?
  - Same question asked a few slides ago (the events were called C and D then)
  - Events A and B are not mutually exclusive
  - $\Pr(A) + \Pr(B) = 4/6 + 3/6 = 7/6$
    - Clearly incorrect

Probabilities



- What is Pr(A or B)?
  - Probability of observing a 1, 2, 3, 4, or 5: probability of 5/6
- Problem with $\mathrm{Pr}(A) + \mathrm{Pr}(B)$ is that it double counts outcomes 1 and 3
  - Double counting $\mathrm{Pr}(A \text{ and } B)$

General addition rule

- If A and B are any two events, then the probability that at least one of them occurs is

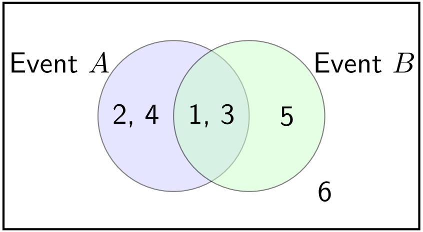
$$\Pr(A \text{ or } B) = \Pr(A) + \Pr(B) - \Pr(A \text{ and } B)$$

- If events A and B are mutually exclusive, then $\Pr(A \text{ and } B) = 0$.
- Example (from previous slide)
  - $\Pr A + \Pr(B) - \Pr(A \text{ and } B) =$ 4/6 + 3/6 - 2/6 = 5/6

Complement

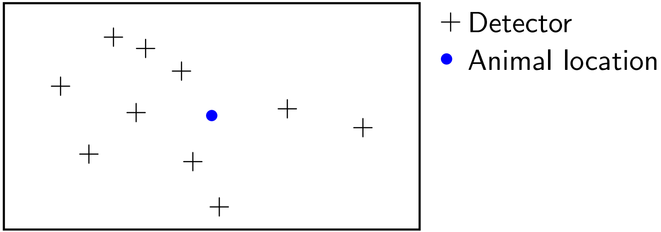
- Compliment: STAT115 students are amazing!
- Complement of event A: the outcomes in the sample space that are not in A
- Roll a die: sample space is $\{1, 2, 3, 4, 5, 6\}$
  - The event E is rolling even: $\{2, 4, 6\}$
  - Its complement $E^{\complement}$ is $\{1, 3, 5\}$
- $\Pr(E) + \Pr(E^{\complement}) = 1$, or $\Pr(E) = 1 - \Pr(E^{\complement})$
  - For the example above: $\Pr(E) = 0.5$, $\Pr(E^{\complement}) = 0.5$
- Complements seem obvious and simple
  - I frequently remind 400-level students how useful they can be

Complements



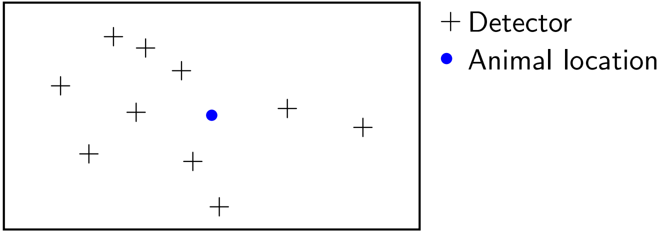
- Complements 'play nice' with Venn diagrams
- What is:
  - $\Pr(A^{\complement})$?
  - $\Pr(B^{\complement})$?
  - $\Pr((A \text{ or } B)^{\complement})$?
  - $\Pr((A \text{ and } B)^{\complement})$?

Complement: real example



- The picture represents an array of 'detectors' (e.g. motion activated camera)
  - Assume we know the probability the animal is detected (in some period of time) for each of the 10 detectors, based on its location
    - $p_1, p_2, \ldots, p_{10}$
- What is the probability it is seen by at least one detector?

Complement: real example



- There are over 1000 possible ways an animal could be detected:
  - Seen at one detector: detector 1, detector 2, detector 3, ...
  - Seen at two detectors: detector 1 & 2, detector 1 & 3
  - etc
- There is only one way an animal cannot be seen
  - Complement of being seen by at least one detector

Summary

- We are working toward statistical model for data
  - Use probability to describe the variation in the data
- Foundational knowledge in probability
  - Outcomes and events
  - Sample space, sets, and complements
  - General addition rule
- Relate probability back to examples
- Tomorrow: everyone bring a coin
  - Explore some (interactive) probability results