STAT115: Introduction to Biostatistics

University of Otago
Ōtākou Whakaihu Waka

Lecture 6: Conditional Probability and Independence

Outline

- Continue to build our knowledge of probability
- Today we look at two (or more) random processes
  - Independence
  - Conditional probability
  - Contingency tables
- Begin with an interactive exercise

Probability is hard

- It can be easy to trick ourselves that probability is easy
  - What is the probability that a die lands on a 4?

Probability is hard

- It can be easy to trick ourselves that probability is easy
  - What is the probability that a die lands on a 4?
  - It goes from easy to difficult very quickly
- In each of the next two lectures:
  - Start with an exercise that *might* be surprising to you
  - Hopefully broaden understanding of probability
- If I told you that I was flipping coins and saw 7 heads in a row
  - Would you think I am telling the truth?
  - How likely is it that I flip a fair coin and get 7 heads in a row?
    - Without getting calculators (or phones!) out
    - Is it closest to: 1 in ten thousand? 1 in 100? 1 in 20?

Exercise

- Everyone stand up
- Flip your coin (when I tell you to)[1]
  - Head: remain standing
  - Tail: sit down
- Those who are still standing, flip again (when I tell you to)
- Repeat until everyone sits down
  - See how many flips the 'best' person can get
- Is that what you were expecting?
- How would this look if we repeated the experiment at Forsyth Barr Stadium?
  - Capacity of about 38,000

[1] If you have forgotten a coin, google 'flip a coin' and you can do it online

Examples

- We might want to know the probability that:
  - A try was scored under the posts given the conversion was successful?
  - A participant undertakes a task left handed given they have a certain genotype?
  - A person has a certain genotype given they have brown eyes?
  - A female skink observed without offspring is a breeder?
  - An aftershock of magnitude 5 or larger occurs within six hours of a earthquake of magnitude 6?
  - Chicken is safe to consume given it has been heated to $60\,^{\circ}\mathrm{C}$ for five minutes?
- Each of these probabilities depends on another variable

Independence

- There are situations where we would expect two random processes to be unrelated
  - Process 1: rolling a die, process 2: flipping a coin
  - Process 1: eye colour of a person, process 2: success of rugby kick (conversion)
- We refer to these as independent
  - Two events A and B are independent if the outcome of one event provides no information about the outcome of the other
    - Knowing that our coin flip landed on heads does not change the probability of rolling a six
- Other processes may not be independent
  - Process 1: stock price of asset A, process 2: stock price of asset B

Independence

- Consider again the two process:
  - Process 1: rolling a die, process 2: flipping a coin
- If A is the event 'roll four or lower', and B is the event 'coin lands head'
  - What is $\Pr(A \text{ and } B)$?

Independence

- Consider again the two process:
  - Process 1: rolling a die, process 2: flipping a coin
- If A is the event 'roll four or lower', and B is the event 'coin lands head'
  - What is $\Pr(A \text{ and } B)$?
- $\Pr(A) = 4/6$ and $\Pr(B) = 1/2$
- Since the two events are independent we can reason that:
  - Event A will occur $4/6$ of the time
  - Event B will subsequently occur $1/2$ of *those* times
  - Event A and B occur together $4/6 \times 1/2 = 2/6$ of the time

Multiplication rule: indpendent processes

- If A and B are independent events, then[2]

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

- Example: suppose that 10% of the population are left handed, and 50% are female. If handedness and sex are independent, then what is the probability that a randomly selected person is right-handed and female?

$$\begin{aligned}\Pr(\text{right-handed and female}) &= \Pr(\text{right-handed})\Pr(\text{female}) \\ &= 0.9 \times 0.5 \\ &= 0.45\end{aligned}$$

[2] This can be extended to more than two events

Diversion: sex and gender

- Often we look at differences to do with sex or gender
- If interest is in biological differences (like the example above)
  - Sex: XX and XY
    - Recognize that intersex individuals exist
    - Usually not accounted for as prevalence is low
- In other applications (e.g. social science) we may be interested in gender
  - Often two genders still used
  - In time, increasingly see wider representation

Conditional probability

- Conditional probability describes the relationship between two events
- The probability of event B given event A has occurred is written $\Pr(B \mid A)$
  - Let T be the event that a try was scored under the posts
  - Let C be the event that the conversion was successful
    - $\Pr(T \mid C)$ is the probability that a try was scored under the posts given the conversion was successful?
  - Let B be the event probability of relapse for some cancer
  - Let O be the event individual was over 60 years old at time of diagnosis
    - $\Pr(B \mid O)$ is the probability that a 60$+$ individual will have relapse

Conditional probability

- We can find $\Pr(B \mid A)$ using

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

- $\Pr(A \text{ and } B)$: joint probability of events A and B occuring
- $\Pr(A)$: marginal probability of event A
- Two events A and B are independent if
  - $\Pr(B \mid A) = \Pr(B)$
  - The event A occurring does not change the probability of B occurring.
- Helpful to look at contingency tables

Contingency table: Titanic

- Contingency tables allow us to compare two (categorical) variables [3]
- Data from the adult passengers on the titanic. Two variables:
  - Sex: male or female
  - Survived: yes or no
- Two tables: the first gives the counts, the second gives proportions

		survived		
		yes	no	Total
Sex	male	338	1329	1667
	female	316	109	425
	Total	654	1438	2092

		survived		
		yes	no	Total
Sex	male	0.162	0.635	0.797
	female	0.151	0.052	0.203
	Total	0.313	0.687	1.000

[3] They can be extended to more than two variables

Contingency table: Titanic

- Two tables: the first gives the counts, the second gives proportions
  - For now, we will treat the proportions as if they are probabilities
  - See better approaches for estimating probabilities from contingency tables later

		survived		
		yes	no	Total
Sex	male	338	1329	1667
	female	316	109	425
	Total	654	1438	2092

		survived		
		yes	no	Total
Sex	male	0.162	0.635	0.797
	female	0.151	0.052	0.203
	Total	0.313	0.687	1.000

- Proportion are found by dividing entries by total, 2092
  - e.g. $316/2092 = 0.151$
  - e.g. $1438/2092 = 0.687$

Contingency table: Titanic

- S: randomly selected passenger survived
- M: randomly selected passenger is male
- Marginal probability
  - $\Pr(M) = 0.797$
  - $\Pr(S) = 0.313$
  - Found in margin of contingency table

		survived		
		yes	no	Total
Sex	male	338	1329	1667
	female	316	109	425
	Total	654	1438	2092

		survived		
		yes	no	Total
Sex	male	0.162	0.635	0.797
	female	0.151	0.052	0.203
	Total	0.313	0.687	1.000

Contingency table: Titanic

- S: randomly selected passenger survived
- M: randomly selected passenger is male
- Joint probabilities
  - $\Pr(M \text{ and } S) = 0.162$
  - $\Pr(M \text{ and } S^\complement) = 0.635$
  - $\Pr(M^\complement \text{ and } S) = 0.151$
  - $\Pr(M^\complement \text{ and } S^\complement) = 0.052$
  - Found in cells of contingency table

		survived		
		yes	no	Total
Sex	male	338	1329	1667
	female	316	109	425
	Total	654	1438	2092

		survived		
		yes	no	Total
Sex	male	0.162	0.635	0.797
	female	0.151	0.052	0.203
	Total	0.313	0.687	1.000

Contingency table: Titanic

- S: randomly selected passenger survived
- M: randomly selected passenger is male
- Conditional probabilities
  - $\Pr(S \mid M) = \frac{\Pr(S \text{ and } M)}{\Pr(M)} = \frac{0.162}{0.797} = \frac{338}{1667} = 0.20$
    - Only consider male row
    - Survival and sex are not independent (why?)
  - $\Pr(S \mid M^{\complement}) = \frac{0.151}{0.203} = \frac{316}{425} = 0.74$
  - Could also find $\Pr(M \mid S)$, ...

		survived		
		yes	no	Total
Sex	male	338	1329	1667
	female	316	109	425
	Total	654	1438	2092

		survived		
		yes	no	Total
Sex	male	0.162	0.635	0.797
	female	0.151	0.052	0.203
	Total	0.313	0.687	1.000

Contingency table: smallpox

- Data are 6224 observations from individuals in Boston in 1721 who were exposed to smallpox[4]. There are two variables:
  - Inoculated: yes or no[5]
  - Result: lived or died

		inoculated		
		yes	no	Total
result	lived	238	5136	5374
	died	6	844	850
	Total	244	5980	6224

		inoculated		
		yes	no	Total
result	lived	0.038	0.825	0.863
	died	0.001	0.136	0.137
	Total	0.039	0.961	1.000

[4] Fenner F. et al. 1988. Smallpox and Its Eradication (History of International Public Health, No. 6). Geneva: World Health Organization. ISBN 92-4-156110-6, p. 257

[5] Exposing a person to the disease in a controlled form

Contingency table: smallpox

- I: individual exposed to smallpox was inoculated
- L: individual exposed to smallpox lived
  - Find marginal probabilities:
    - $\Pr(I)$, $\Pr(L)$, $\Pr(I^{C})$, $\Pr(L^{C})$
  - Find joint probabilities:
    - $\Pr(I \text{ and } L)$, $\Pr(I^{C} \text{ and } L)$, $\Pr(I \text{ and } L^{C})$, $\Pr(I^{C} \text{ and } L^{C})$
  - Find conditional probabilities:
    - $\Pr(L \mid I)$, $\Pr(L \mid I^{C})$
    - Are L and I independent?
  - Find $\Pr(I \text{ or } L)$
    - Is this a meaningful quantity in this example?

		inoculated		
		yes	no	Total
result	lived	238	5136	5374
	died	6	844	850
	Total	244	5980	6224

		inoculated		
		yes	no	Total
result	lived	0.038	0.825	0.863
	died	0.001	0.136	0.137
	Total	0.039	0.961	1.000

Summary

- Looked at two random processes
- Introduced:
  - Independence
  - Multiplication rule (for independent events)
  - Conditional probability
  - Contingency tables
- Still building our knowledge of probability
  - So that we can apply it for statistical modelling
- We will start the next lecture with another probability exercise