# STAT115: Introduction to Biostatistics

University of Otago

Ōtākou Whakaihu Waka

# Lecture 7: Working with Conditional Probability

## Outline

- Look more at (conditional) probability
- Start with another probability exercise

# Birthday challenge

- How many people do you need for a good chance that two share a birthday?
  - Take a guess (but keep it to yourself)
- I'll start asking student's for their birthday until I get a match

# Recap

- In the last lecture we talked about:
  - Joint probability: $\Pr(A \text{ and } B)$
  - Marginal probability: $\Pr(A)$
  - Conditional probability: $\Pr(B \mid A) = \frac{\Pr(A \text{ and } B)}{\Pr(A)}$
- Pick up from where we left off

# Multiplication rule: general

- Last lecture we saw how to find conditional probability

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

- If we rearrange this, we get the general multiplication rule

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

- We can 'switch' $A$ and $B$ so that we also have

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

# Multiplication rule: smallpox example

- Suppose we were told: "96.1% of the residents were not inoculated, and 85.9% of the residents who were not inoculated ended up surviving."
  - What is the probability that a resident was not inoculated and lived?

# Multiplication rule: smallpox example

- Suppose we were told: "96.1% of the residents were not inoculated, and 85.9% of the residents who were not inoculated ended up surviving."
  - What is the probability that a resident was not inoculated and lived?
- $\Pr(I^{\complement}) = 0.961$: 96.1% of the residents were not inoculated
- $\Pr(L \mid I^{\complement}) = 0.859$: 85.9% of the residents who were not inoculated ended up surviving
- $\Pr(I^{\complement} \text{ and } L)$: probability that a resident was not inoculated and lived

$$\begin{aligned} \Pr(I^{\complement} \text{ and } L) &= \Pr(I^{\complement}) \Pr(L \mid I^{\complement}) \\ &= 0.961 \times 0.859 \\ &= 0.825 \end{aligned}$$

# Joint and conditional probability (smallpox example)

- Order doesn't matter for the joint probability
  - $\Pr(A \text{ and } B) = \Pr(B \text{ and } A)$
  - Probability both $A$ and $B$ occur
- Order does matter for the conditional probability
  - $\Pr(A \mid B)$ and $\Pr(B \mid A)$ are two different quantities
- Smallpox: compare $\Pr(L \mid I)$ and $\Pr(I \mid L)$.
  - $\Pr(L \mid I) = \frac{0.038}{0.039} = \frac{238}{244} = 0.975$
    - Probability of a resident living given inoculation
  - $\Pr(I \mid L) = \frac{0.038}{0.863} = \frac{238}{5374} = 0.044$
    - Probability of a resident being inoculated given lived

| | | 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 |

# Marginal probability: law of total probability

- Previously found "intuitively" from contingency table
- To find $\Pr(B)$
  - Sum over possible outcomes that could co-occur with the event $B$
- If there are two outcomes: $A_1$ and $A_2 = A_1^{\complement}$
  - $\Pr(B) = \Pr(A_1 \text{ and } B) + \Pr(A_1^{\complement} \text{ and } B)$

  - Smallpox: find $\Pr(L)$ and $\Pr(I)$

| | | 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 |

# Marginal probability: law of total probability

- Previously found "intuitively" from contingency table
- To find $\Pr(B)$
  - Sum over possible outcomes that could co-occur with the event $B$
- If there are two outcomes: $A_1$ and $A_2 = A_1^{\complement}$
  - $\Pr(B) = \Pr(A_1 \text{ and } B) + \Pr(A_1^{\complement} \text{ and } B)$

| | | 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 |

- Smallpox: find $\Pr(L)$ and $\Pr(I)$
  - $\Pr(L) = \Pr(I \text{ and } L) + \Pr(I^{\complement} \text{ and } L) = 0.038 + 0.825 = 0.863$

# Marginal probability: law of total probability

- Previously found "intuitively" from contingency table
- To find $\Pr(B)$
  - Sum over possible outcomes that could co-occur with the event $B$
- If there are two outcomes: $A_1$ and $A_2 = A_1^{\complement}$
  - $\Pr(B) = \Pr(A_1 \text{ and } B) + \Pr(A_1^{\complement} \text{ and } B)$

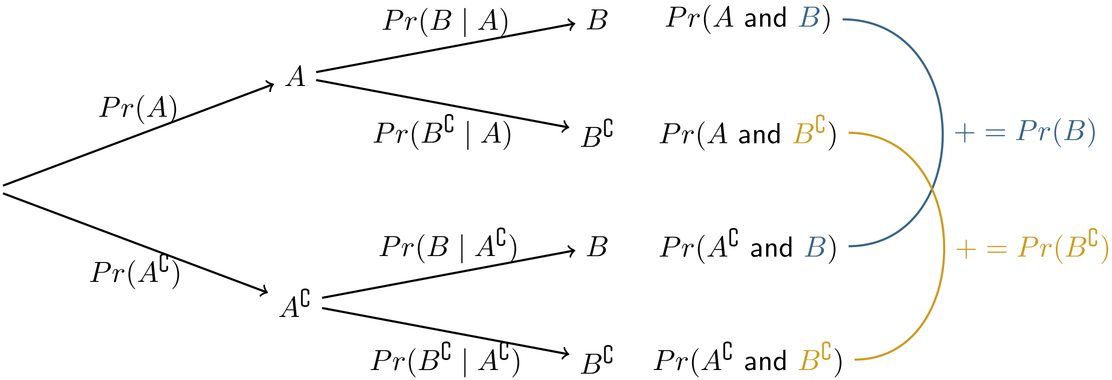
| | | inoculated | | |
|---|---|---|---|---|
| | | yes | no | Total |
| result | lived | 238 | 5136 | 5374 |
| | died | 6 | 844 | 850 |
| | Total | 244 | 5980 | 6224 |

| | | 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 |

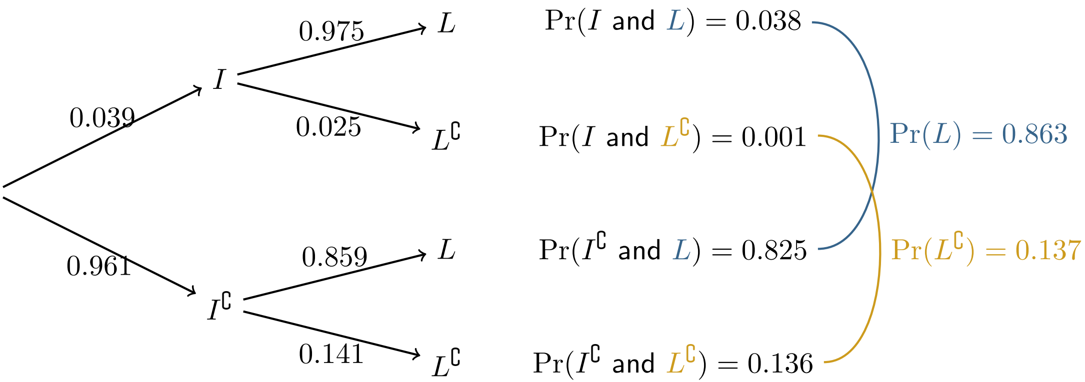
- Smallpox: find $\Pr(L)$ and $\Pr(I)$
  - $\Pr(I) = \Pr(I \text{ and } L) + \Pr(I \text{ and } L^{\complement}) = 0.038 + 0.001 = 0.039$

# Tree diagrams

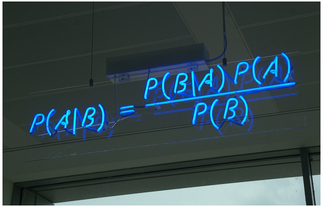
- Tree diagrams are an alternate way to visualize outcomes and probabilities
- General form:



# Tree diagrams: smallpox example



# Bayes' theorem




- An important result in probability theory
  - Underpins a lot of modern statistics/data science/AI
    - Hopefully return to this at the end of the semester

- Idea: find $\Pr(A \mid B)$ from $\Pr(B \mid A)$

$$\Pr(A \mid B) = \frac{\Pr(B \mid A)\Pr(A)}{\Pr(B)} = \frac{\Pr(B \mid A)\Pr(A)}{\Pr(B \mid A)\Pr(A) + \Pr(B \mid A^{\complement})\Pr(A^{\complement})}$$

# Bayes' theorem

- Example: sleep apnea
- $A$: patient has sleep apnea
- $S$: patient snores
- Question: we know our patient snores. What is the prob they sleep apnea?
  - 90% of patients with sleep apnea snore: $\Pr(S \mid A) = 0.9$
  - 50% of patients without sleep apnea snore: $\Pr(S \mid A^{\complement}) = 0.5$
  - 5% of the population have sleep apnea: $\Pr(A) = 0.05$

# Bayes' theorem

- Example: sleep apnea
- $A$: patient has sleep apnea
- $S$: patient snores
- Question: we know our patient snores. What is the prob they sleep apnea?
  - 90% of patients with sleep apnea snore: $\Pr(S \mid A) = 0.9$
  - 50% of patients without sleep apnea snore: $\Pr(S \mid A^{\complement}) = 0.5$
  - 5% of the population have sleep apnea: $\Pr(A) = 0.05$

$$\begin{aligned}\Pr(A \mid S) &= \frac{\Pr(S \mid A)\Pr(A)}{\Pr(S \mid A)\Pr(A) + \Pr(S \mid A^{\complement})\Pr(A^{\complement})} \\ &= \frac{0.9 \times 0.05}{0.9 \times 0.05 + 0.5 \times 0.95} = 0.087\end{aligned}$$

# Bayes' theorem: example

- We had: $\Pr(S \mid A) = 0.9$
  - Tempting to think that $\Pr(A \mid S)$ will also be high
  - $\Pr(A \mid S) \approx 0.09$ seems surprisingly low
  - Confusing conditional probabilities is a common mistake
- Another perspective:
  - How does the probability of patient having sleep apnea change
- In the general population, we have
  - $\Pr(A) = 0.05$
- After learning that patient snores, the probability increases to
  - $\Pr(A \mid S) \approx 0.09$

# Bayes' theorem: understanding

- It can be difficult to understand why $\Pr(A \mid S)$ is low when $\Pr(S \mid A)$ is high
- Construct a hypothetical ('expected') contingency table
  - Pretend there are 100 000 patients
  - Recall: $\Pr(A) = 0.05$

| | | snores ($S$) | | |
|---|---|---|---|---|
| | | yes | no | Total |
| sleep apnea ($A$) | yes | | | 5000 |
| | no | | | 95 000 |
| | Total | | | 100 000 |

# Bayes' theorem: understanding

- It can be difficult to understand why $\Pr(A \mid S)$ is low when $\Pr(S \mid A)$ is high
- Construct a hypothetical ('expected') contingency table
  - Pretend there are 100 000 patients
  - Recall: $\Pr(S \mid A) = 0.9$

| | | snores ($S$) | | |
|---|---|---|---|---|
| | | yes | no | Total |
| sleep apnea ($A$) | yes | 4500 | 500 | 5000 |
| | no | | | 95 000 |
| | Total | | | 100 000 |

# Bayes' theorem: understanding

- It can be difficult to understand why $\Pr(A \mid S)$ is low when $\Pr(S \mid A)$ is high
- Construct a hypothetical ('expected') contingency table
  - Pretend there are 100 000 patients
  - Recall: $\Pr(S \mid A^{\complement}) = 0.5$

| | | snores ($S$) | | |
|---|---|---|---|---|
| | | yes | no | Total |
| sleep apnea ($A$) | yes | 4500 | 500 | 5000 |
| | no | 47 500 | 47 500 | 95 000 |
| | Total | | | 100 000 |

# Bayes' theorem: understanding

- It can be difficult to understand why $\Pr(A \mid S)$ is low when $\Pr(S \mid A)$ is high
- Construct a hypothetical ('expected') contingency table
  - Pretend there are 100 000 patients

| | | snores ($S$) | | |
|---|---|---|---|---|
| | | yes | no | Total |
| sleep apnea ($A$) | yes | 4500 | 500 | 5000 |
| | no | 47 500 | 47 500 | 95 000 |
| | Total | 52 000 | 48 000 | 100 000 |

- Most of those who snore do not have sleep apnea!
- $\Pr(A \mid S) = \frac{4500}{52\,000} = 0.087$

# Summary

- Looked in more detail at conditional probability
- Generalised the multiplication rule
- Tree diagrams
- Bayes' theorem
  - Using formula
  - Constructing an expected contingency table
- Next: begin exploring how to use probability to model data