EGM101 – Skills Toolbox

Week 8, Part 1: Bayes’ Theorem

Week 8 Outline

1. Bayes' Theorem
2. Hypothesis Testing
3. Parametric Tests
4. Non-parametric Tests
5. ANOVA
6. The Chi-square Distribution

The ~~Cookie~~ Biscuit Problem

- Exercise: if we sample a ~~cookie~~ biscuit at random, what is:
  - $P(\text{choc. chip})$
  - $P(\text{cust. crème})$
  - $P(\text{choc. chip} \mid \text{Bowl 1})$
- Assume we randomly sample from a bowl without looking, and get a custard crème.
  - What is the probability it came from either bowl?
- In other words, what is:
  - $P(\text{Bowl 1} \mid \text{cust. crème})$
  - $P(\text{Bowl 2} \mid \text{cust. crème})$

	Chocolate Chip	Custard Crème	Total
Bowl 1	30	10	40
Bowl 2	20	20	40
Total	50	30	80

$$P(\text{choc. chip}) = \frac{\text{number of choc. chip}}{\text{number of biscuits}} = \frac{50}{80} = 0.625$$

$$P(\text{cust. crème}) = \frac{\text{number of cust. crème}}{\text{number of biscuits}} = \frac{30}{80} = 0.375$$

$$P(\text{choc. chip} \mid \text{Bowl 1}) = \frac{\text{number of choc. chip, Bowl 1}}{\text{number of Bowl 1}} = \frac{30}{40} = 0.75$$

The ~~Cookie~~ Biscuit Problem (cont.)

- Note: can calculate probabilities from a contingency table
- But:
  - What if we only have some of the information?
  - How do the conditional probabilities relate to each other?

	Chocolate Chip	Custard Crème	Total
Bowl 1	30	10	40
Bowl 2	20	20	40
Total	50	30	80

$$P(\text{Bowl 1}|\text{cust. crème}) = \frac{\text{number of cust. crème, Bowl 1}}{\text{number of cust. crème}}$$

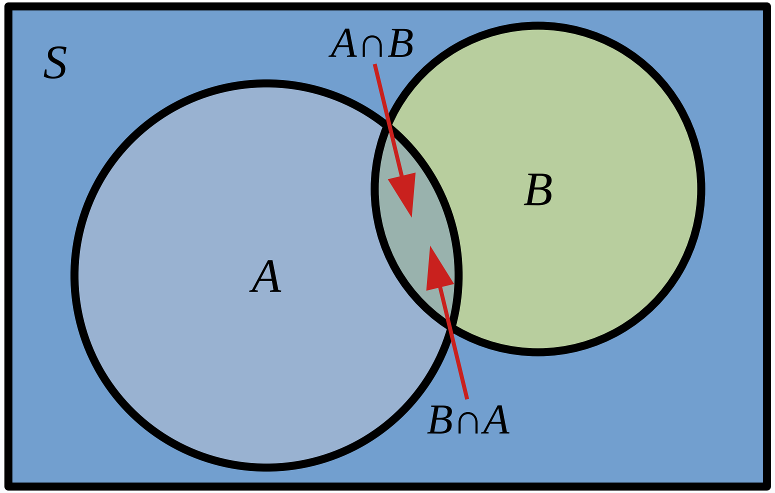
$$= \frac{10}{30} \approx 0.333$$

$$P(\text{Bowl 2}|\text{cust. crème}) = \frac{\text{number of cust. crème, Bowl 2}}{\text{number of cust. crème}}$$

$$= \frac{20}{30} \approx 0.667$$

Conditional Probabilities, Revisited

- Start with events A, B
- Assume that B has happened
  - What is the probability of event A happening, $P(A|B)$?
- We can see:
  - $P(A|B)$ is proportion of $A \cap B$ within B
  - $P(B|A)$ is proportion of $B \cap A$ within A
- Note that $A \cap B = B \cap A$
  - intersection is commutative (order doesn't matter)
  - So, $P(A \cap B) = P(B \cap A)$



$$P(B|A) = \frac{P(B \cap A)}{P(A)} \qquad P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$P(B|A)P(A) = P(A|B)P(B)$$

Bayes' Theorem

- Bayes' Theorem tells us how $P(A|B)$ and $P(B|A)$ are related
- May also see in terms of hypothesis, H and data, D
  - "What is the likelihood of my hypothesis, given the evidence?"
  - Bayesian inference
- Components:
  - Posterior probability: probability of A, given B
  - Likelihood: probability of observing B, given that A occurs
  - Prior probability: the probability of A, regardless of B
  - Marginal likelihood: probability that B occurs, regardless of A

$$\underbrace{\boxed{P(A|B)}}_{\text{posterior}} = \frac{\underbrace{\boxed{P(B|A)}}_{\text{likelihood}}\,\underbrace{\boxed{P(A)}}_{\text{prior}}}{\underbrace{\boxed{P(B)}}_{\text{marginal likelihood}}}$$

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)}$$

Testing and Type I, Type II Errors

- Test: testing for presence/absence of a condition
  - Positive: condition is present
  - Negative: condition is not present
- Sensitivity: probability of a true positive
  - Type I Error: test incorrectly says positive
- Specificity: probability of a true negative
  - Type II Error: test incorrectly says negative
- Prevalence: probability of someone having the condition

		Reality	
		Condition present	Condition not present
Measurement	Condition present	Correct (True Positive)	Type I (False Positive)
Measurement	Condition not present	Type II (False Negative)	Correct (True Negative)

The Prosecutor's Fallacy

- Assume: my fingerprint is found on a (now-empty) biscuit tin.
- Fingerprint test:
  - Sensitivity, $P(\mathrm{Pos}|\mathrm{G})$: 1 in 1,000,000
  - Specificity, $P(\mathrm{Pos}|\mathrm{NG})$: 1 in 1,000,000
  - Number of entries in database: 10,000,000
- Given a match (positive result), what are the chances that I am truly guilty, $P(\mathrm{G}|\mathrm{Pos})$?
  - Note: $P(\mathrm{Pos})$ includes both true positives **and** false positives!
- Prosecutor's fallacy: assuming that $P(\mathrm{G}|\mathrm{Pos}) \approx P(\mathrm{Pos}|\mathrm{NG})$

	Positive	Negative
Guilty	1	0
Not Guilty	10	9,999,990

$$P(\mathrm{G}|\mathrm{Pos}) = \frac{P(\mathrm{Pos}|\mathrm{G})P(\mathrm{G})}{P(\mathrm{Pos})}$$

$$= \frac{P(\mathrm{Pos}|\mathrm{G})P(\mathrm{G})}{\boxed{P(\mathrm{Pos}|\mathrm{G})P(\mathrm{G})} + \boxed{P(\mathrm{Pos}|\mathrm{NG})P(\mathrm{NG})}}$$

very small!

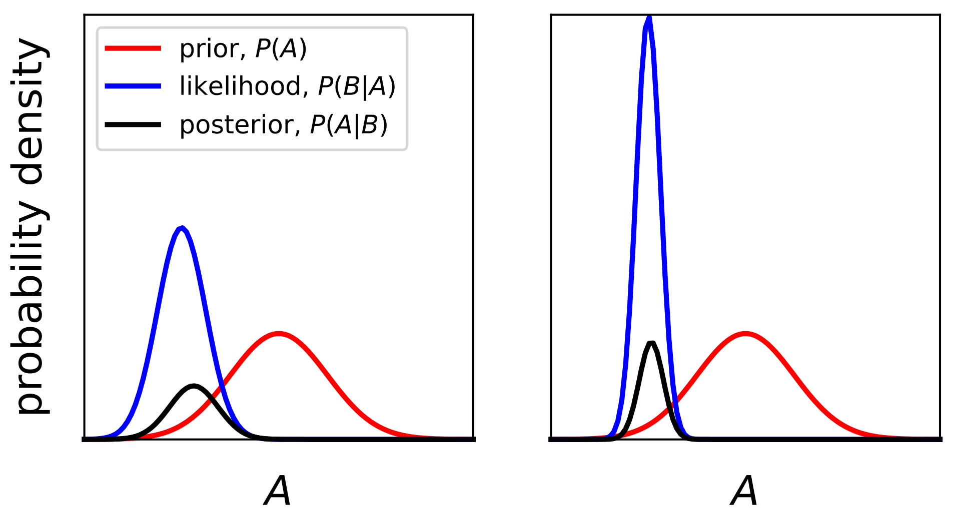
$$= \frac{\boxed{\frac{999999}{1000000} * \frac{1}{10000000}}}{\boxed{\frac{999999}{1000000} * \frac{1}{10000000}} + \boxed{\frac{10}{9999990} * \frac{9999999}{10000000}}}$$

also very small, but ~10x larger

$$= \frac{1}{11}$$

What Just Happened?

- Probability distributions
  - If likelihood or the prior distribution are very wide, so is the posterior
  - In other words, there are a larger range of possible outcomes
- Another way of thinking of Bayes' Theorem:
  - The more alternative explanations there are, the less likely your chosen explanation is.
  - Or: the less sure you are of the evidence, the less likely the hypothesis is.



Summary

- Bayes’ Theorem:
  - Gives us a way to relate conditional probabilities
  - Allows us to calculate the likelihood of an explanation, given the evidence
- If there are many alternative explanations for our evidence, our chosen explanation is less likely
- Above all, beware false positives!

Additional Resources

- Bergstrom and West, “The Susceptibility of Science” (Chapter 9)
- Bayes’ Theorem: What’s the big deal? [Scientific American]
- Bayesian Inference [Seeing Theory]
- Bayes theorem, the geometry of changing beliefs [3Blue1Brown]
- P Values and the Prosecutor’s fallacy [UW iSchool]