Bayes Theorem, Bayesian Stats, Bayes Nets

Kenneth (Kenny) Joseph

University at Buffalo
Department of Computer Science and Engineering
School of Engineering and Applied Sciences

Reminders

- Corrections due next Tuesday (or you can give them to me today)
  - You don't need to do 5e or 6 (but hey, why not?)
- PA2 grades are out
- PA3 grades out next week I hope
- Quiz 9 out
- PA4 due next Tuesday

Plan

- Review Quiz 8
- Return to end of PCA
- Foundations of Bayesian modeling (IMO)
- Note: The above will likely take 2 lectures
  - Probably, mostly math today, more code on Tuesday

Quiz 8 + PCA ending example

Details for today… a story to keep in mind

1. **Bayes theorem** is a simple probability rule (originally for point probabilities) that is the foundation for…
2. **Bayesian statistics** where the goal is to estimate the posterior distribution of a parameter. One way to do so is through…
3. **MAP Estimation**, although there are others. Many of these alternative approaches can be implemented effectively using…
4. **Probabilistic Programming**

Finally, Bayesian statistical models are often complex, but can be easily represented with

5. **Directed (Probabilistic) Graphical Models,** AKA "Bayesian Networks". However, it is critical to note that even though these are called "Bayesian networks", they **don't have to represent a Bayesian model.** So I'll mostly stick with D-PGM

Bayes Rule / Bayes Theorem

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

A, B = events

$P(A|B)$ = probability of A given B is true

$P(B|A)$ = probability of B given A is true

$P(A), P(B)$ = the independent probabilities of A and B

Bayes Rule / Bayes Theorem

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

A, B	= events
$P(A \mid B)$	= probability of A given B is true
$P(B \mid A)$	= probability of B given A is true
$P(A), P(B)$	= the independent probabilities of A and B

derived
rem with
algebra,
ig deal?

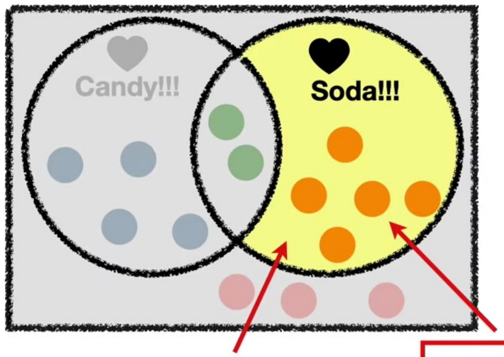
A = does not love candy
B = loves soda

$$p(A \,\&\, B \mid B) = \frac{p(A \,\&\, B \mid A) \times p(A)}{p(B)}$$

$$p(A \,\&\, B \mid A) = \frac{p(A \,\&\, B \mid B) \times p(B)}{p(A)}$$

More From Statquest

$$p(\text{no love } \boldsymbol{c} \text{ \& love } \boldsymbol{s} \mid \text{no love } \boldsymbol{c}) = \frac{p(\text{no love } \boldsymbol{c} \text{ \& love } \boldsymbol{s} \mid \text{love } \boldsymbol{s}) \times p(\text{love } \boldsymbol{s})}{p(\text{no love } \boldsymbol{c})}$$



$$p(\text{love } \boldsymbol{s} \mid \text{no love } \boldsymbol{c}) = \frac{p(\text{no love } \boldsymbol{c} \mid \text{love } \boldsymbol{s}) \times p(\text{love } \boldsymbol{s})}{p(\text{no love } \boldsymbol{c})}$$

Bayes Theorem Example 1: Allergies

		Test +	Test −
Truth	Have allergy	80%	20%
	Not	10%	90%

Test +

$p(\text{Has allergy} \mid \text{test}+) = \frac{p(\text{has} \cap \text{test}+)}{p(\text{test}+)}$

$p(\text{test}+) = p(\text{test}+ \cap \text{has}) + p(\text{test}+ \cap \neg\text{has})$

prior: $p(\text{has}) = 1\%$

$= \frac{p(\text{test}+ \mid \text{has})\, p(\text{has})}{p(\text{test}+ \mid \text{has})\, p(\text{has}) + p(\text{test}+ \mid \neg\text{has})\, p(\neg\text{has})}$

$= \frac{.8 \cdot .01}{.8 \cdot .01 + .99 \cdot .1} = .074$

Example from: https://www.mathsisfun.com/data/probability-false-negatives-positives.html

Bayes Theorem Example 2: Spell Check

Example from: http://www.stat.columbia.edu/~gelman/book/BDA3.pdf, Section 1.4

Moving on

- Bayes Theorem is a way to take two things:
  - What we think we already know about something (our **prior**)
  - What we have learned from data about that thing (our **likelihood**)
- And to use them to update our knowledge of the thing

Bayesian Stats, Definition 1

The goal of Bayesian statistics is to represent prior uncertainty about model parameters with a probability distribution and to update this prior uncertainty with current data to produce a posterior probability distribution for the parameter that contains less uncertainty

Scott M. Lynch: Introduction to Applied Bayesian Statistics and Estimation for Social Scientists (2007) Springer.
Chapter 3

Bayesian Stats, Definition 1

The goal of Bayesian statistics is to represent prior uncertainty about model parameters with a probability distribution and to update this prior uncertainty with current data to produce a posterior probability distribution for the parameter that contains less uncertainty

- Key distinction from above examples: The prior is a **distribution**, so the posterior is too, now.
  - (In our examples thus far, we just have been using point probabilities)

Bayesian Stats, Definition 3

The practice of updating the probability of the value of some parameter θ of model M being the correct value, based on observations (D for data)

- https://www.cs.rice.edu/~ogilvie/comp571/2018/09/13/bayesian-inference.html

I like it b/c it emphasizes the **model**

Bayesian Statistics

Bayes rule

$$p(B|A) = \frac{p(A|B)p(B)}{p(A)}.$$

Bayes stats

likelihood, prior

$$f(\theta|\text{data}) = \frac{f(\text{data}|\theta)f(\theta)}{f(\text{data})},$$

posterior

p(data)

$$f(\text{data}) = \int f(\text{data}|\theta)f(\theta)d\theta;$$

$$\text{Posterior} \propto \text{Likelihood} \times \text{Prior},$$

Bayesian Statistics

update

$$\text{Posterior} \propto \text{Likelihood} \times \text{Prior},$$

what is my model?

1. Set up the full probability model (the **joint**)
2. Condition on observed data (estimate the **posterior**)
3. Evaluate model fit

Probabilistic Programming + An Example

- https://nbviewer.org/github/CamDavidsonPilon/Probabilistic-Programming-and-Bayesian-Methods-for-Hackers/blob/master/Chapter1_Introduction/Ch1_Introduction_PyMC3.ipynb#

Bayesian Statistics

$$\text{Posterior} \propto \text{Likelihood} \times \text{Prior},$$

1. Set up the full probability model (the **joint**)
2. Condition on observed data (estimate the **posterior**)
3. Evaluate model fit