More probability

CS311
David Kauchak
Spring 2013

Some material borrowed from:
Sara Owsley Sood and others

Admin

- Assign 3 Tuesday at the beginning of class (in class)
- Should have looked at written 2 by now
- Written 3 out soon
- Mancala tournament: good news and bad news

Another example

Start with the joint probability distribution:

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576

P(*toothache*) = ?

Another example

Start with the joint probability distribution:

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576

P(*toothache*) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2

Another example

Start with the joint probability distribution:

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576

P(¬*cavity* | *toothache*) = ?

Another example

Start with the joint probability distribution:

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576

$$P(\neg cavity \mid toothache) = \frac{P(\neg cavity, toothache)}{P(toothache)} = \frac{0.016+0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4$$

Normalization

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576

Denominator can be viewed as a normalization constant α

P(*CAVITY* | *toothache*) = α **P**(*CAVITY*,*toothache*)
= α [**P**(*CAVITY*,*toothache*,*catch*) + **P**(*CAVITY*,*toothache*,¬ *catch*)]
= α [<0.108,0.016> + <0.012,0.064>]
= α <0.12,0.08> = <0.6,0.4>

unnormalized p(cavity|toothache)   unnormalized p(¬cavity|toothache)

General idea: compute distribution on query variable by fixing evidence variables and summing over hidden/unknown variables

More Probability

In the United States, 55% of children get an allowance and 41% of children get an allowance and do household chores. What is the probability that a child does household chores given that the child gets an allowance?

$$p(chores \mid allow) = p(chores, allow) / p(allow)$$

$$= 0.41 / 0.55 = 0.745$$

Still more probability

- A math teacher gave her class two tests. 25% of the class passed both tests and 42% of the class passed the first test. What is the probability that a student who passed the first test also passed the second test?

Another Example

A patient takes a lab test and the result comes back positive. The test has a false negative rate of 2% and false positive rate of 2%. Furthermore, 0.5% of the entire population have this cancer.

What is the probability of cancer if we know the test result is positive?

Another Example

A patient takes a lab test and the result comes back positive. The test has a false negative rate of 2% and false positive rate of 2%. Furthermore, 0.5% of the entire population have this cancer.

What is the probability of cancer if we know the test result is positive?

p(cancer) = 0.005
p(false_neg) = 0.02
p(false_pos)=0.02

p(cancer | pos) = ?

false negative: negative result even though we have cancer

false positive: positive result even though we don't have cancer

Another Example

p(cancer) = 0.005
p(false_neg) = 0.02
p(false_pos)=0.02

p(cancer | pos) = ?

false negative: negative result even though we have cancer

false positive: positive result even though we don't have cancer

$$p(cancer \mid pos) = \frac{p(cancer, pos)}{p(pos)}$$

Another Example

p(cancer) = 0.005
p(false_neg) = 0.02
p(false_pos)=0.02

p(cancer | pos) = ?

false negative: negative result even though we have cancer

false positive: positive result even though we don't have cancer

1-p(false_neg) gives us the probability of the test correctly identifying us with cancer

$$\frac{p(cancer, pos)}{p(pos)} = \frac{p(cancer)(1 - p(false\_neg))}{p(cancer)(1 - p(false\_neg)) + p(\neg cancer)p(false\_pos)}$$

two ways to get a positive result: cancer with a correct positive and not cancer with a false positive

Another Example

p(cancer) = 0.005
p(false_neg) = 0.02
p(false_pos)=0.02

p(cancer | pos) = ?

false negative: negative result even though we have cancer

false positive: positive result even though we don't have cancer

$$p(cancer \mid pos) = 0.1975$$

Contrast this with p(pos | cancer) = 0.98

Obtaining probabilities

We've talked a lot about probabilities, but not where they come from

- intuition/guess
  - this can be very hard
  - people are not good at this for anything but the simplest problems
- estimate from data!

Estimating probabilities

Total Flips: 10
Number Heads: 5
Number Tails: 5

Probability of Heads:
Number Heads / Total Flips = 0.5

Probability of Tails:
Number Tails / Total Flips = 0.5 = 1.0 – Probability of Heads

The experiments, the sample space and the events must be defined clearly for probability to be meaningful

Theoretical Probability

Maximum entropy principle
- When one has only partial information about the possible outcomes one should choose the probabilities so as to maximize the uncertainty about the missing information
- Alternatives are always to be judged equally probable if we have no reason to expect or prefer one over the other

Maximum likelihood estimation
- set the probabilities so that we maximize how likely our data is

Turns out these approaches do the same thing!

Maximum Likelihood Estimation

$$\frac{\text{Number of times an event occurs in the data}}{\text{Total number of times experiment was run} \atop \text{(total number of data collected)}}$$

Maximum Likelihood Estimation

$$\frac{\text{Number of times an event occurs in the data}}{\text{Total number of times experiment was run} \atop \text{(total number of data collected)}}$$

Rock/Paper/Scissors

http://www.nytimes.com/interactive/science/rock-paper-scissors.html

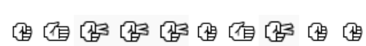
How is it done?

Maximum Likelihood Estimation

$$\frac{\text{Number of times an event occurs in the data}}{\text{Total number of times experiment was run} \atop \text{(total number of data collected)}}$$

Rock/Paper/Scissors

http://www.nytimes.com/interactive/science/rock-paper-scissors.html

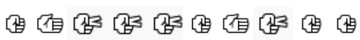
 …

- Analyze the prior choices
- Select probability of next choice based on data

How?

Maximum Likelihood Estimation

$$\frac{\text{Number of times an event occurs in the data}}{\text{Total number of times experiment was run (total number of data collected)}}$$



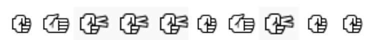
P(rock) =

P(rock | scissors) =

P(rock | scissors, scissors, scissors) =

Maximum Likelihood Estimation

$$\frac{\text{Number of times an event occurs in the data}}{\text{Total number of times experiment was run (total number of data collected)}}$$



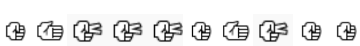
P(rock) = 4/10 = 0.4

P(rock | scissors) = 2/4 = 0.5

P(rock | scissors, scissors, scissors) = 1/1 = 1.0

Maximum Likelihood Estimation

$$\frac{\text{Number of times an event occurs in the data}}{\text{Total number of times experiment was run (total number of data collected)}}$$



P(rock) = 4/10 = 0.4

P(rock | scissors) = 2/4 = 0.5

P(rock | scissors, scissors, scissors) = 1/1 = 1.0

Which of these do you think is most accurate?

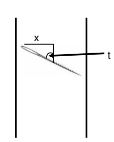
Law of Large Numbers

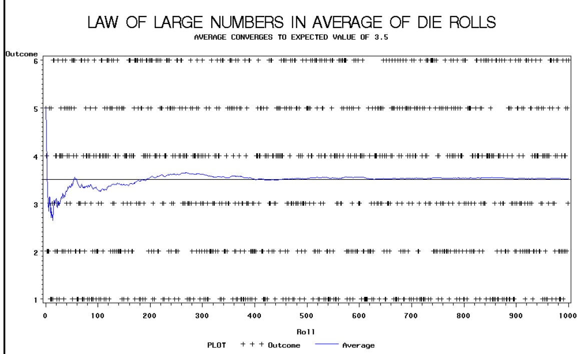
As the number of experiments increases the relative frequency of an event more closely approximates the actual probability of the event.

- if the theoretical assumptions hold

Buffon's Needle for Computing π

- http://mste.illinois.edu/reese/buffon/buffon.html





Large Numbers Reveal Problems in Assumptions

Results of 1,000,000 throws of a die

Number	1	2	3	4	5	6
Fraction	.155	.159	.164	.169	.174	.179

Probabilistic Reasoning

Evidence
- What we know about a situation

Hypothesis
- What we want to conclude

Compute
- P(Hypothesis | Evidence)

Probabilistic Reasoning

Evidence
- What we know about a situation

Hypothesis
- What we want to conclude

Compute
- P(Hypothesis | Evidence)

Credit card application?

Credit Card Application

E is the data about the applicant's age, job, education, income, credit history, etc,

H is the hypothesis that the credit card will provide positive return.

The decision of whether to issue the credit card to the applicant is based on the probability P(H|E).

Probabilistic Reasoning

Evidence
- What we know about a situation

Hypothesis
- What we want to conclude

Compute
- P(Hypothesis | Evidence)

Medical diagnosis?

Medical Diagnosis

E is a set of symptoms, such as, coughing, sneezing, headache, ...

H is a disorder, e.g., common cold, SARS, swine flu.

The diagnosis problem is to find an H (disorder) such that P(H|E) is maximum.

Chain rule (aka product rule)

$$p(X \mid Y) = \frac{P(X,Y)}{P(Y)} \Longrightarrow p(X,Y) = P(X \mid Y)P(Y)$$

We can view calculating the probability of X *AND* Y occurring as two steps:

1. Y occurs with some probability P(Y)
2. Then, X occurs, given that Y has occurred

or you can just trust the math… ☺

Chain rule (aka product rule)

$$p(X \mid Y) = \frac{P(X,Y)}{P(Y)} \Rightarrow p(X,Y) = P(Y \mid X)P(X)$$

We can view calculating the probability of X *AND* Y occurring as two steps:

1. X occurs with some probability P(X)
2. Then, Y occurs, given that X has occurred

or you can just trust the math… ☺

Chain rule

$$p(X,Y,Z) = P(X \mid Y,Z)P(Y,Z)$$
$$p(X,Y,Z) = P(X,Y \mid Z)P(Z)$$
$$p(X,Y,Z) = P(X \mid Y,Z)P(Y \mid Z)P(Z)$$
$$p(X,Y,Z) = P(Y,Z \mid X)P(X)$$

$$p(X_1,X_2,\ldots,X_n) = \ ?$$

Bayes' rule (theorem)

$$p(X \mid Y) = \frac{P(X,Y)}{P(Y)} \Rightarrow p(X,Y) = P(X \mid Y)P(Y)$$

$$p(X \mid Y) = \frac{P(X,Y)}{P(Y)} \Rightarrow p(X,Y) = P(Y \mid X)P(X)$$

$$p(X \mid Y) = \frac{P(Y \mid X)P(X)}{P(Y)}$$

Bayes' rule

Allows us to talk about P(Y|X) rather than P(X|Y)

Sometimes this can be more intuitive

Why?

$$p(X \mid Y) = \frac{P(Y \mid X)P(X)}{P(Y)}$$

Bayes' rule

p(disease | symptoms)

- For everyone who had those symptoms, how many had the disease?
- p(symptoms|disease)
  - For everyone that had the disease, how many had this symptom?

p(good_lendee | credit_features)

- For everyone who had these credit features, how many were good lendees?
- p(credit_features | good_lendee)
  - For all the good lenders, how many had this feature

p(cause | effect) vs. p(effect | cause)

p(H | E) vs. p(E | H)

Bayes' rule

$$p(good\_lendee \mid features) = \frac{P(features \mid good\_lendee)P(good\_lendee)}{P(features)}$$

We often already have data on good lenders, so p(features | good_lendee) is straightforward

p(features) and p(good_lendee) are often easier than p(good_lendee|features)

Allows us to properly handle changes in just the underlying distribution of good_lendees, etc.

Other benefits

Simple lender model:

- score: is credit score > 600
- debt: debt < income

$$p(Good \mid Credit, Debt) = \frac{P(Credit, Dept \mid Good)P(Good)}{P(Credit, Debt)}$$

Other benefits

It's in the 1950s and you train your model "diagnostically" using just p(Good | Credit, Debt).

However, in the 1960s and 70s the population of people that are good lendees drastically increases (baby-boomers learned from their depression era parents and are better with their money)

$$p(Good \mid Credit, Debt)$$

Intuitively what should happen?

Other benefits

It's in the 1950s and you train your model "diagnostically" using just p(Good | Credit, Debt).

However, in the 1960s and 70s the population of people that are good lendees drastically increases (baby-boomers learned from their depression era parents and are better with their money)

$$p(Good \mid Credit, Debt)$$

Probability of "good" should increase, but that's hard to figure out from just this equation

Other benefits

$$p(Good \mid Credit, Debt) = \frac{P(Credit, Dept \mid Good) P(Good)}{P(Credit, Debt)}$$

Modeled using Bayes' rule, it's clear how much the probability should change.

Measure what the new P(Good) is.

When it rains…

Marie is getting married tomorrow at an outdoor ceremony in the desert. In recent years, it has rained only 5 days each year. Unfortunately, the weatherman has predicted rain for tomorrow. When it actually rains, the weatherman correctly forecasts rain 90% of the time. When it doesn't rain, he incorrectly forecasts rain 5% of the time. What is the probability that it will rain on the day of Marie's wedding?

p(rain) = 5/365
p(predicted|rain) = 0.9
p(predicted|¬rain) = 0.05

When it rains…

p(rain) = 5/365
p(predicted|rain) = 0.9
p(predicted|¬rain) = 0.05

$$p(rain \mid predicted) = \frac{p(predicted \mid rain) p(rain)}{p(predicted)}$$

$$= \frac{0.9 * 5/365}{p(predicted)}$$

When it rains…

p(rain) = 5/365
p(predicted|rain) = 0.9
p(predicted|¬rain) = 0.05

$$p(predicted) = p(predicted \mid rain)p(rain) + p(predicted \mid \neg rain)p(\neg rain)$$

$$p(\neg rain \mid predicted) = p(predicted \mid \neg rain)p(\neg rain)$$
$$= 0.05 * 360/365$$

Monty Hall

- 3 doors
  - behind two, something bad
  - behind one, something good
- You pick one door, but are not shown the contents
- Host opens one of the other two doors that has the bad thing behind it (he always opens one with the bad thing)
- You can now switch your door to the other unopened. Should you?

Monty Hall

p(win) initially?
- 3 doors, 1 with a winner, p(win) = 1/3

p(win | shown_other_door)?
- One reasoning:
  - once you're shown one door, there are just two remaining doors
  - one of which has the winning prize
  - 1/2

This is not correct!

Be careful! – Player picks door 1

	winning location		host opens
1/3	Door 1	1/2	Door 2
		1/2	Door 3
1/3	Door 2	1	Door 3
1/3	Door 3	1	Door 2

In these two cases, switching will give you the correct answer. Key: host knows where it is.

Another view

1000 doors

- behind 999, something bad
- behind one, something good

You pick one door, but are not shown the contents

Host opens 998 of the other 999 doors that have the bad thing behind it (he always opens ones with the bad thing)

In essence, you're picking between it being behind your one door or behind any one of the other doors (whether that be 2 or 999)