More probability

CS151
David Kauchak
Fall 2010

Some material borrowed from:
Sara Owsley Sood and others

Admin

- Assign 3 due Monday at the beginning of class (in class)

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

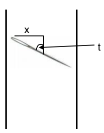
The experiments, the sample space and the events must be defined clearly for probability to be meaningful

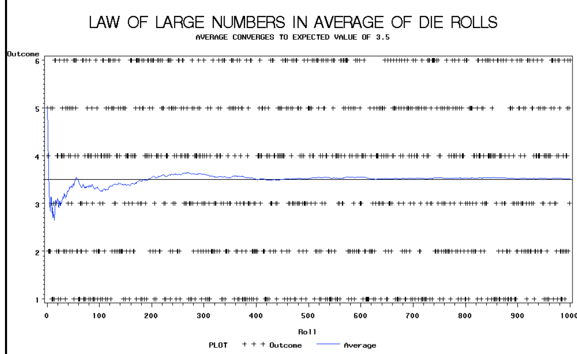
Theoretical Probability

- **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 equi-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!

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)

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).

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)

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

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 occured

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 model 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)

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

Intiuitively, the probability of good should increase, but 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

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

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

Joint distributions

- For an expression with *n* boolean variables e.g. $P(X_1, X_2, \ldots, X_n)$ how many entries will be in the probability table?
  - 2^n
- Does this always have to be the case?

Independence

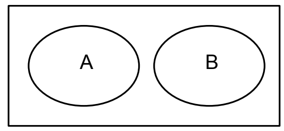
- Two variables are independent if one has nothing whatever to do with the other
- For two independent variables, knowing the value of one does not change the probability distribution of the other variable (or the probability of any individual event)
  - the result of the toss of a coin is independent of a roll of a dice
  - price of tea in England is independent of the result of general election in Canada

Independent or Dependent?

- Catching a cold and having cat-allergy
- Miles per gallon and driving habits
- Height and longevity of life

Independent variables

- How does independence affect our probability equations/properties?



- If A and B are independent (written …)
  - P(A,B) = P(A)P(B)
  - P(A|B) = P(A)
  - P(B|A) = P(B)

Independent variables

- If A and B are independent
  - P(A,B) = P(A)P(B)
  - P(A|B) = P(A)
  - P(B|A) = P(B)

Reduces the storage requirement for the distributions

Conditional Independence

- Dependent events can become independent given certain other events
- Examples,
  - height and length of life
  - "correlation" studies
    - size of your lawn and length of life
- If A, B are conditionally independent of C
  - P(A,B|C) = P(A|C)P(B|C)
  - P(A|B,C) = P(A|C)
  - P(B|A,C) = P(B|C)
  - but P(A,B) ≠ P(A)P(B)