Information Theory
A couple of slides on probability

Linda Doyle
CTVR

probability rules review

it seems that people are having some problems with manipulating probabilities when we need to do various calculations

we are going to break off here and revise some issues

random variables

- random variables are variables that take on values determined by probability distributions
- two different views about what probability means

view 1

1. **relative frequency**: sample the random variable a great number of times and tally up the fraction of times that each of its different values occur, to arrive at the probability of each

frequentist view and one that predominates in statistics and information theory

two views – view 2

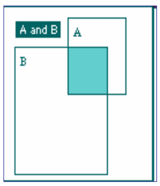
2. **degree-of-belief**: probability is the plausibility of a proposition or likelihood that a particular state (or value of a random variable) might occur, even if its outcome can only be decided once (e.g. the outcome of a particular horse race)

but this kind of view is also important and needed – this one leads to Bayes rule

very basics

- All Probability is Between 0 and 1 (or 100%)
- The *complement* of event A is the event (A does not occur). All simple events in the sample space must either be part of event A or part of the complement of event A.
- P(A) + P(not A) = 1

product rule



P(A and B) = P(B given A) P(A)

joint probability

an example

- To assess the need for day-care services, a large corporation wishes to determine the probability that a female employee has children under the age of 5.
- They take a random sample of employees and determine the proportion who are female. Call this proportion **P(F)**.
- Next, for the female employees, they determine the proportion who have children under 5, Call this proportion **P(C given F)**.
- The proportion who are both female and have children under 5 is:
  P(F and C) = P(F) P(C given F)

Product Rule

p(A,B)

= joint probability of both A and B

= **p(A|B)p(B)**

= **p(B|A)p(A)**

Of course if A and B are independent events they do not depend on each other and

p(A,B) = **p(A)p(B)**

Averaging Rule

- If A is conditionalised on a number of other events B, the total probability of A is the sum of its joint probabilities with all B:
- $p(A) = \sum_B p(A,B) = \sum_B p(A|B)p(B)$

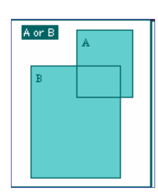
a simple example

A \ B	0	1
0	1/8	1/8
1	1/4	1/2

$p(A{=}0) = \sum_B p(A,B) = p(0, B{=}0) + p(0, B{=}1) = [1/8+1/8] = 2/8 = 1/4$

we used this kind of calculation before in some of the work we did

sometimes you might be interested in A or B



P(A or B) = P(A) + P(B) - P(A and B)

following examples from
http://cnx.org/content/m16847/latest/

- Klaus is trying to choose where to go on vacation. His two choices are: AA = New Zealand and BB = Alaska
- Klaus can only afford one vacation. The probability that he chooses A is P(A) = 0.6 and the probability that he chooses B is P(B)= 0.35.
- P(A and B)=0 because Klaus can only afford to take one vacation
- Therefore, the probability that he chooses either New Zealand or Alaska is P(A OR B)= P(A)+P(B) = 0.6 + 0.35= 0.95.
- Note that the probability that he does not choose to go anywhere on vacation must be 0.050.05.

sample problem to do now

- Carlos plays soccer. He makes a goal 65% of the time he shoots. Carlos is going to attempt two goals in a row in the next game.
- A = the event Carlos is successful on his first attempt => P(A) = 0.65
- B = the event Carlos is successful on his second attempt => P(B) = 0.65.
- Carlos tends to shoot in streaks. The probability that he makes the second goal GIVEN that he made the first goal is 0.90.

sample problem to do now

1. What is the probability that he makes both goals?
2. What is the probability that Carlos makes either the first goal or the second goal?

the probability he makes both goals

- The problem is asking you to find P(A AND B)
- We know that P(B|A) = 0.90
- P(B AND A) = **P(B|A) P(A)**
- = 0.90 * 0.65 = 0.585
- Carlos makes the first and second goals with probability 0.585.

makes either the first goal or the second goal

- The problem is asking you to find P(A OR B)
- P(A OR B) = P(A) + P(B)- P(A AND B)
- = 0.65 + 0.65- 0.585 = 0.715
- Carlos makes either the first goal or the second goal with probability 0.715.

Bayes Rule

- From the product Rule and the symmetry that $p(A,B) = p(B,A)$ is is clear that $p(A|B)p(B)= p(B|A)p(A)$.
- Bayes theorem follows and is

$$p(B|A) = p(A|B)p(B)/p(A)$$

Bayes allows us to reverse the conditionalizing of events

two views – view 2

2. **degree-of-belief**: probability is the plausibility of a proposition or likelihood that a particular state (or value of a random variable) might occur, even if its outcome can only be decided once (e.g. the outcome of a particular horse race)

Bayes rule encapsulates this view

Worked Example

- Suppose that a disease affects $1/1000^{th}$ of all people.
- If you have the disease then a test is positive 95% of the time and negative 5% of the time.
- If you don't have the disease the test is positive 5% of the time.
- How do we interpret these results?

cont.,

- rather than use A and B we use D and H
- D is the data, i.e. the result of a test
- H is the hypothesis you have the disease or $\bar{H}$ is the hypothesis you do not.

a priori information

Before acquiring any data we only know that you have .001 probability of having the disease. $p(H) = 0.001$ and is known as the prior.

doing a test should somehow change the p(H) value – it should update it.

the probability of testing positive

- so there is a chance you can test positive whether you have it or not?
- from the sum rule we can calculate the a priori probability p(D) of testing positive, whatever the truth may be, is:
- $p(D) = p(D|H)p(H)+p(D|\bar{H})p(\bar{H})$
- $p(D) = (.95)(.001)+(.05)(.999)=0.51$

We can now use Bayes Rule

$$p(H|D) = p(D|H)p(H)/p(D)$$

$$p(H|D) = (0.95)(0.001)/(0.051)=0.019$$

This quantity is called the posterior probability because it is computed after the observation of data; it tells us how likely the hypothesis is, given what we observed.

Bayes provides a mechanism for updating our assessment as more data continues to arrive.

summary

this brief overview will should help remind you of some of the main concepts in probability theory that are used in the course