**SC531 – Lecture #05**

Examples from Ref. #2

1. A and B alternately throw a pair of dice. A wins if he throws 6 before B throws 7, and B wins if he throws 7 before A throws 6. If the game starts with A, find the probability that A wins.

We know that P(6 is thrown) = 5/36, and P(7 is thrown) = 6/36.

Define event A as “A throws 6”, and B as “B throws 7”.

A wins the game if and only if he throws 6 in the first, third, fifth or any subsequent odd numbered trial. Let AWIN be this event.

So P(AWIN)

= P(A) + P(~A~BA) + P(~A~B~A~BA) + ... (addition rule)

= 5/36 + (31/36) x 5/6 x 5/36 + (31/36)2 x 5/62 x 5/36 + ...

= 5/36 x [1 – 155/216]-1  (infinite geometric series)

= 30/61 🡪 slightly less than 0.5!

Modified version:

A and B alternately throw a pair of dice. Whoever throws 6 first wins. If the game starts with A, find the probability that A wins.

Verify: P(AWIN) = 36/67 🡪 note “first mover advantage".

[BTW: Compiled statistics of good chess players seem to suggest that white wins between 52% and 56% of the time.]

2. Over a binary communication channel, the probability that a transmitted 0 is received as 0 is 0.95, and the probability that a transmitted 1 is received as 1 is 0.90. The probability that a 0 is transmitted is 0.4. Find the probability that: (a) a 1 is received, and (b) a 1 was transmitted, given that 1 is received.

Let A be the event of transmitting 1, and B the event of receiving 1.

(a) Total probability of receiving 1, P(B), is given by:

= P(B|A)P(A) + P(B|~A)P(~A)

= 0.9x0.6 + 0.05x0.4 = 0.56 [assuming no lost bits!]

(b) P(A|B)

= P(A&B) / P(B)

= P(A) x P(B|A) / P(B) [Bayes' rule]

= 27/28

3. A bag contains 5 balls, some white and some black. Two balls are drawn at random from the bag and found to be white. What is the probability that all the balls in the bag are white?

Since two white balls are drawn, the bag must originally contain 2, 3, 4 or 5 white balls. Let Bi represent the corresponding event, with i going from 2 to 5. Let A represent drawing two white balls. Then:

P(A|B2) = 2C2 / 5C2 = 1/10

P(A|B3) = 3C2 / 5C2 = 3/10

P(A|B4) = 4C2 / 5C2 = 6/10

P(A|B5) = 5C2 / 5C2 = 10/10 [denominators are maintained at 10

for clarity]

Therefore by Bayes' theorem:

P(B5|A) = P(B5) x P(A|B5) / P(A) eqn. 1

But we still do NOT have the values of P(B2), P(B3), P(B4) and P(B5), which are needed to calculate the RHS! What should we do?

**WE MUST ASSUME *A PRIORI* THAT THESE FOUR EVENTS ARE EQU-PROBABLE – SINCE WE HAVE *NO EVIDENCE* EITHER WAY!**

Therefore P(B2) = P(B3) = P(B4) = P(B5) = ¼.

Plug into the RHS of eqn. 1, giving the answer ½.

[Exercise: Verify that S P(Bi|A), for i = 2 .. 5, evaluates to 1.]

**Bernoulli trials**

Consider N independent trials of an experiment, say E. Let A be an event defined on the outcomes of E, such that P(A) remains the same across the N trials. Then these trials are called **Bernoulli trials**.

Let the probability of occurrence of A in a single trial be *p*. Then the probability of *r* such “successes” in N trials is NCr *p*r(1- *p*)N-*r*.

Note that the last expression is the typical term of the *binomial* *expansion*.

1. Simple example of a continuous probability distribution – which is correctly known as a ***probability density function***.

p(x)

x = 3

x = 0

a) What is the value of p(3)? Total area under the curve must be 1. Therefore p(3) = 2/3.

b) What is the probability that this RV has value between a and b?

We must integrate p(x)dx between the limits x=a and x=b.

**A trivial example of a joint distribution:**

Throwing a pair of dice, but with the number thrown by each die being considered as a separate RV. Green cells indicate total number = 7.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **1** | **2** | **3** | **4** | **5** | **6** |
| **1** | 1/36 | 1/36 | 1/36 | 1/36 | 1/36 | 1/36 |
| **2** | 1/36 | 1/36 | 1/36 | 1/36 | 1/36 | 1/36 |
| **3** | 1/36 | 1/36 | 1/36 | 1/36 | 1/36 | 1/36 |
| **4** | 1/36 | 1/36 | 1/36 | 1/36 | 1/36 | 1/36 |
| **5** | 1/36 | 1/36 | 1/36 | 1/36 | 1/36 | 1/36 |
| **6** | 1/36 | 1/36 | 1/36 | 1/36 | 1/36 | 1/36 |

Example of **probabilistic inference**, from Ref.#1

Imagine a (highly simplified!) model of a dentist at work.

Three Boolean random variables are relevant: *Toothache*, *Cavity* and *Catch*. Their full 2 x 2 x 2 *joint distribution* is given below, which serves as a probabilistic knowledge base for this example.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Toothache | | ~Toothache | |
|  | Catch | ~Catch | Catch | ~Catch |
| Cavity | 0.108 | 0.012 | 0.072 | 0.008 |
| ~Cavity | 0.016 | 0.064 | 0.144 | 0.576 |

We can see from this joint distribution, for example, that:

(a) The sum of all these probabilities is 1. Of course this must be true.

(b) P(Toothache & Cavity & Catch) = 0.108 [top left entry]

(c) P(~Toothache & ~Cavity & ~Catch) = 0.576 [bottom right entry]

(d) P(Toothache or Cavity ) = 0.280

We get this probability by excluding from the total sum the TWO probabilities corresponding to ~Toothache & ~Cavity.

[You can think of this as an application of de Morgan's rule.]

(e) Applying the definition of conditional probability:

P(Cavity|Toothache)

= P(Cavity & Toothache) / P(Toothache)

= 0.120/0.200

= 0.6

Now imagine **IF** we had a reliable, learned knowledge base – similar to the one here – containing hundreds of random variables!

The prospect is mouth-watering – but we need to remember that reliable and widely applicable data is difficult to come by!

[🡪 *Relationship of learning to statistics*.]

[BTW: Insurance business – for centuries – has relied on huge amounts of carefully compiled data.]

**Marginalisation**

From the above table, we make the following calculation:

P(Cavity) = 0.108 + 0.012 + 0.072 + 0.008 = 0.2

Note that we summed over all possible values of Toothache and Catch. In everyday language, if we ignore the values of Toothache and Catch, then about 20% of the relevant population has cavity.

Such an addition process is called **marginalization** or **summing out**. In a slightly different form, the rule can be expressed as:

P(A, B ..) = S P(A, B ..| X, Y ..) summation over all values of X, Y ..

In this form, the calculation is known as **conditioning**.