**SC531 – Lecture #04**

Let B1, B2 ... Bn be a set of mutually exclusive events, and let A be another event in the sample space. We assume that (i) A is caused by any of the events Bi, and that (ii) all the possible causes of A are included in Bi. Remember causality pervades the universe.

Then P(A) = S P(A & Bi) sum rule over i = 1, 2 .... n

= S P(A|Bi)P(Bi)

The terms on the RHS, for i = 1,2 ... n, are values which we assume to be “known” – or learnt by experience.

Then the summation can be understood as a kind of weighted sum, with terms in the weighted sum being of the form P(A|Bi)P(Bi). This relationship is called the rule of total probability.

Examples from Ref. #2

1. Bolts are produced by machines A, B and C. Machine A produces TWICE as many bolts as machines B and C. The respective defect rates of the three machines are 2%, 2% and 4%, respectively.

From a huge stock of these bolts, one is drawn at random. What is the probability that it is defective?

Let the events A, B, C be defined as the chosen bolt being from machine A, B, C respectively, and D as the bolt being defective.

From the given data, P(A) = ½, P(B) = P(C) = ¼.

Also, P(D|A) = P(D|B) = 0.02, P(D|C) = 0.04.

So P(D) = P(D|A)P(A) + P(D|B)P(B) + P(D|C)P(C)

Plugging in the values, P(D) = 0.01 + 0.005 + 0.01 = 0.025.

2. An urn contains 10 white and 3 black balls, while a second urn contains 3 white and 5 black balls. Two balls are drawn at random from the first urn and transferred to the second. Then a single ball is drawn at random from the second urn. Find the probability that it is white.

Let B1, B2, B3 respectively be events representing 2 white balls, two black balls, and 1 white + 1 black balls being drawn from urn 1 and transferred to urn 2.

P(B1) = 10C2 /13C2 = 15/26

P(B2) = 3C2 /13C2 = 1/26

P(B3) = 10C1 x 3C1 /13C2 = 10/26 [Note these numbers add to 1.]

Let W represent the event of a white ball being drawn at the end.

Then P(W|B1) = 5/10, P(W|B2) = 3/10, P(W|B3) = 4/10.

Applying the rule of total probability, P(W) = 59/130.

\*\*\*

The famous **Bayes’ theorem** (or **rule**) follows by re-applying the basic definition of conditional probability to the above *in reverse*:

P(Bi|A) = P(Bi & A) / P(A) = P(A|Bi)P(Bi) / S P(A|Bi)P(Bi) [sum as above]

Note the crucial fact here: Instead of talking about P(A|Bi), we have switched to talking about P(Bi|A).

In one direction, we see P(effect|cause). This is our knowledge.

In the converse direction, we see P(cause|effect). This is our diagnosis.

Converse of the previous two examples:

1. If a defective bolt is drawn from the combined lot, what are the probabilities that it was produced by machines A, B, C?

P(A|D) = P(A & D) / P(D) = 0.01/0.025 = 0.4

P(B|D) = P(B & D) / P(D) = 0.005/0.025 = 0.2

P(C|D) = P(C & D) / P(D) = 0.01/0.025 = 0.4 [Numbers add to 1.]

Note that here, in the converse case, we already have a defective bolt in hand. We are trying to diagnose its probable cause.

2. If a white ball is drawn from urn 2, what is the probability that two white balls were initially moved from the urn 1 to urn 2?

P(B1|W) = P(B1 & W) / P(W) = P(W|B1) x P(B1) / P(W)

= (5/10)x(15/26)/(59/130) = 15/59 × 5/2 = 75/118

Similarly,

P(B2|W) = P(B2 & W) / P(W) = P(W|B2) x P(B2) / P(W)

= (3/10)x(1/26)/(59/130) = 1/59 × 3/2 = 3/118

P(B3|W) = P(B3 & W) / P(W) = P(W|B3) x P(B3) / P(W)

= (4/10)x(10/26)/(59/130) = 10/59 × 2 = 40/118

Given a white ball in hand, we are trying to infer or diagnose what probably happened in the initial transfer from urn 1 to urn 2. Note that the answers add up to 1.

In both the above cases, note that we used the same numerical values as before, but in a different way.

Example from Ref. #1

Earthquake

Burglary

Alarm

Mahesh calls

Ramesh calls

A burglar alarm is installed in my house. Ramesh and Mahesh are my neighbours.

Arrows indicate possible causes. Thus we can assume that we “know” – or have learnt – the values of P(A|E), P(A|B), P(R|A) and P(M|A) etc. We assume that E or B do not directly affect R or M calling.

Let us assume the following probabilities:

A. Characteristics of the neighbourhood, assumed known *a priori*.

P(B) = 0.001

P(E) = 0.002

B. characteristics of the alarm

P(A|B&E] = 0.95

P(A|B&~E] = 0.94

P(A|~B&E] = 0.29

P(A|~B&~E] = 0.001

C. characteristics of the neighbours

P(R|A) = 0.90 [Note: The numbers in A, B, C do

P(R|~A) = 0.05 NOT add to 1!]

P(M|A) = 0.70

P(M|~A) = 0.01

D. Simple “forward" question: What is the probability that Ramesh calls on a given day, if burglary occurs but no earthquake?

Answer: 0.94 x 0.9

E. Typical “converse” question: On a given day, if Ramesh calls but Mahesh does not, what is the probability that a burglary has occurred?

How do we do this “converse" or “diagnostic" calculation? See Ref. #1, under the topic “Bayesian networks”.

**Probability distribution** of a discrete random variable:

Usually, along the X axis, we show values of the random variable; along the Y axis, we show the corresponding probabilities. But any other format will do, provided the meaning is clear and unambiguous.

Simple examples:

Tossing a single unbiased coin: P(head) = P(tail) = 0.5

Rolling a single fair die: P(i) = 1/6, for i = 1, 2 ... 6.

Rolling a pair of fair dice:

P(2) = P(12) = 1/36

P(3) = P(11) = 2/36 [denominator is deliberately left at 36]

P(4) = P(10) = 3/36

P(5) = P(9) = 4/36

P(6) = P(8) = 5/36

P(7) = 6/36

**Continuous random variables**: We make use of *probability density*, in the sense that p(*x*)*dx* gives the probability that the variable has value in the infinitesimal interval from *x* to *x*+*dx*.

**Joint probability distribution**

Each entry gives the probability of two (or more) variable having the values specified in the entry. Can be in tabular form. In the continuous case, the probability density takes the form p(x,y...).