Fundamentals of probability

Tutorial/Lab work: Week 1

Model solutions.

Problem 1

1. The data $\mathcal{D}$ (i.e. the observation of 53 heads and 47 tails) tell us that
$$P(\mathcal{D}|p) \propto p^{53}(1-p)^{47}$$
We will obtain the maximum likelihood estimate by taking the derivative of $P(\mathcal{D}|p)$ w.r.t. p and equate it to zero:
$$\begin{aligned}
P(\mathcal{D}|p) &= p^{53}(1-p)^{47} \\
\frac{\mathrm{d}P(\mathcal{D}|p)}{\mathrm{d}p} &= 53p^{52}(1-p)^{47} - 47p^{53}(1-p)^{46} \\
&= \left(\frac{53}{p} - \frac{47}{1-p}\right)\left(p^{53}(1-p)^{47}\right) \\
&= 0 \quad \text{iff } p = 0.53
\end{aligned}$$

2. If the coin is tossed only once, instead of simply stating that $p = 1$ we wish to account for the uncertainty of the observation, which should be very high in this case.

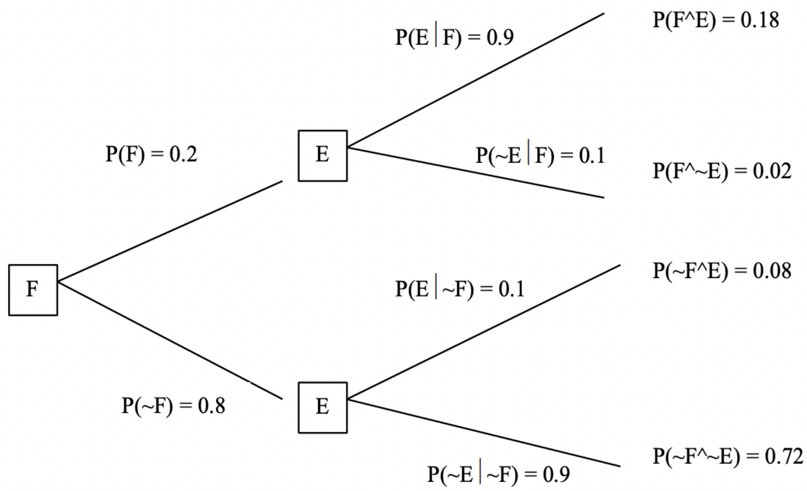
3. If we don't have much data, we are unsure about p. Therefore, we start with a prior distribution over p. In this case, we use a uniform distribution as we do not know anything about coins. Then, we multiply the prior probability of each parameter value by the probability of observing heads given that value. Finally, we scale up[1] all of the probabilities so that their integral[2] tends to 1. This gives us the posterior probability distribution.

   As more data become available, our uncertainty should reduce. After 53 heads and 47 tails we get a very sensible posterior distribution that has its peak[3] at 0.53.

[1] or normalise
[2] area under the curve
[3] even assuming a uniform prior

Problem 2

1. Let's assume E denotes the event that a driver test positive and F denotes the event that a driver has smoked cannabis in the last 72 hrs. Let's visualise this using a chart below:



Using Bayes' theorem:

$$\begin{aligned} P(F|E) &= \frac{P(E|F)\times P(F)}{(P(E|F)\times P(F)) + (P(E|\neg F)\times P(\neg F))} \\ &= \frac{0.9\times 0.2}{(0.9\times 0.2) + (0.1\times 0.8)} \\ &= \frac{0.18}{0.18 + 0.08} \\ &= 0.69 \end{aligned}$$

2. We know the following from the question:

$$\begin{aligned} P(E|F) &= 0.999 \\ P(\neg E|F) &= 0.001 \\ P(E|\neg F) &= 0.2 \\ P(\neg E|\neg F) &= 0.8 \\ P(F) &= 0.2 \\ P(\neg F) &= 0.8 \end{aligned}$$

We are asked to obtain the probability that someone smoked cannabis in the last 72 hours if they have not tested positive, i.e., we have to obtain: $P(F|\neg E)$.

Using Bayes' theorem we have:

$$P(F|\neg E) = \frac{P(\neg E|F)P(F)}{P(\neg E)}$$

We already have the values for the numerator based on the values above. We just have to obtain $P(\neg E)$.

Using basic rules of probability, we have:

$$\begin{aligned}
P(\neg E) &= P(\neg E \cap F) + P(\neg E \cap \neg F) \\
&= P(\neg E|F)P(F) + P(\neg E|\neg F)P(\neg F) \\
&= (0.001x0.2) + (0.8x0.8) \\
&= 0.6402
\end{aligned}$$

Therefore $P(F|\neg E) = \frac{(0.001x0.2)}{0.6402} = 0.00031$

Problem 3

Let us denote:

B = voter from Bury,

C = voter from Croydon,

D = voter from Dover,

W = voted for winner

We want to compute:

$$P(D|W) = \frac{P(D \cap W)}{P(W)}$$

let's get the first thing out of our way

$$\begin{aligned}
P(D \cap W) &= 0.51 \times 0.16 \\
&= 0.0816
\end{aligned}$$

now let's look at the denominator

$$\begin{aligned}
P(W) &= P(B \cap W) + P(C \cap W) + P(D \cap W) \\
&= P(B)P(W|B) + P(C)P(W|C) + P(D)P(W|D) \\
&= 0.46 \times 0.61 + 0.38 \times 0.88 + 0.16 \times 0.51 \\
&= 0.6966
\end{aligned}$$

finally,

$$\begin{aligned}
P(D|W) &= \frac{0.0816}{0.6966} \\
&= 0.1171
\end{aligned}$$

Problem 4

Let us denote:

A = studied AI

B = studied CS

C = did not study AI or CS

D = knows what NN is

The question asks us to obtain the probability $P(D)$ that George knows what a neural network is.

80% of the people who studied AI know what a neural network is, so the probability that someone knows what a NN is given that they studied AI is 0.8, which means $P(D|A) = 0.8$.

40% of the people who studied CS know what a NN is, which means $P(D|B) = 0.4$.

10% of the rest know what a NN is, which means $P(D|C) = 0.1$.

Now, let us suppose that George studied AI with 50% probability, and CS with 20% probability, i.e., $P(A) = 0.5$ and $P(B) = 0.2$.

A, B and C are exhaustive and mutually exclusive[4]. We can now deduce that:

$$P(C) = 1 - P(A) - P(B)$$

[4] George studied only one of these or none

Which is 0.3

Let us now exploit the basic equations of probability theory here:

$$\begin{aligned}
P(D \cap A) &= P(D|A)P(A) \\
&= 0.8 \times 0.4 \\
P(D \cap B) &= P(D|B)P(B) \\
&= 0.4 \times 0.2 \\
P(D \cap C) &= P(D|C)P(C) \\
&= 0.1 \times 0.3
\end{aligned}$$

Since A, B, and C are exhaustive and mutually exclusive:

$$\begin{aligned}
P(D) &= P(D \cap A) + P(D \cap B) + P(D \cap C) \\
&= 0.4 + 0.08 + 0.03 \\
&= 0.51
\end{aligned}$$

Problem 5

The given facts in the question are:

A = picked the biased coin

B = the coin fell as a 'head' 10 times in a row

We want to know the probability that the coin that fell 'head' 10 times in a row is the biased one. That is, the probability that a coin is the biased one given that it has fallen 10 times or $P(A|B)$.

The probability of picking the biased coin is 1 in 1000:

$$P(A) = \frac{1}{1000}$$

so the probability of picking a normal coin is:

$$P(\neg A) = \frac{999}{1000}$$

The probability of a biased coin falling 10 times in a row is 1, which means $P(B|A) = 1$. Similarly, the probability of a normal coin falling heads 10 times in a row is $(\frac{1}{2})^{10}$, which means $P(B|\neg A) = \frac{1}{1024}$.

We are asked to find $P(A|B)$. Using the basic rules of probability, we have:

$$\begin{aligned} P(A|B)P(B) &= P(A \cap B) \\ P(A|B) &= \frac{P(A \cap B)}{P(B)} \end{aligned} \tag{1}$$

Again, with the basic rules we have:
$P(A \cap B) = P(B|A)p(A) = \frac{1}{1000}$.

Also, $P(\neg A \cap B = P(B|\neg A)p(\neg A) = \frac{0.976}{1000} \approx \frac{1}{1000}$.

Finally, from the lecture:

$$\begin{aligned} P(B) &= P(A \cap B) + P(\neg A \cap B) \\ &= \frac{1}{1000} + \frac{1}{1000} \end{aligned}$$

Putting the value of $P(B)$ in Equation 1, we get $P(A|B) = \frac{1}{2}$. Which means, there is a probability of approximately 50% that the coin we picked is biased!

Problem 6

Your possible actions are:

1. Turn witness and if John also does you get 3 years, if he doesn't you get zero.
2. Or you may refuse to turn witness and if John does you get 6 years, and if he doesn't you get 1 year.

According to classical decision theory, you should choose the action that has highest expected utility, in this case the action that minimizes the number of years you expect to spend in jail.

For example, because John is your friend, you might believe:

John turns witness with probability 10%.

John refuses to turn witness with probability 90%.

The best option is for you both to refuse to turn witness against the other. However, you turning witness is your less risky option.

Suppose the probability of John turning witness is P. So, if you turn witness and John also does you get 3 years with probability $P(3 \times P)$; if he doesn't you get 0 years $(0 \times (1 - P))$, for a total of $3P$.

If you don't turn witness but John does you get $6 \times P$, otherwise $1 \times (1 - P)$, for a total of $5P + 1$. Since $3P$ is less than $5P + 1$ for any P, you're always better off turning witness against him[5].

[5] excusing morality and other things

Unfortunately, if he has also studied Decision Theory he will also turn witness against you, and you'll both get 3 years!