Class 12 in-class problems, 18.05, Spring 2022

Concept questions

Concept question 1. Three coins

- *Type $C_{0.5}$ coins are fair, with probability 0.5 of heads*

- *Type $C_{0.6}$ coins have probability 0.6 of heads*

- *Type $C_{0.9}$ coins have probability 0.9 of heads*

A drawer has one of each.

Pick one; toss 5 times; Suppose you get 1 head out of 5 tosses.

What's your best guess for the probability of heads on the next toss?

<div align="center">

(a) *0.1* **(b)** *0.2* **(c)** *0.3* **(d)** *0.4*

(e) *0.5* **(f)** *0.6* **(g)** *0.7* **(h)** *0.8*

(i) *0.9* **(j)** *1.0*

</div>

Solution: Question is answered in the board questions.

Concept question 2. *Suppose instead of saying 1 head in 4 tails, we told you the tosses, in order, were THTTT. Does this affect posterior distribution of the coin type?*

<div align="center">

(a) *Yes* **(b)** *No*

</div>

Solution: No. When the order is not known, the likelihoods will all have a factor of $\binom{5}{1}$, which is absent when the order is known. This will not affect the posterior probabilities.

Board questions

Problem 1. Three coins

- *We have 3 coins with probabilities 0.5, 0.6, and 0.9 of heads.*

- *Pick one at random; toss 5 times.*

- *Suppose you get 1 head out of 5 tosses.*

Compute the posterior probabilities for the type of coin and the posterior predictive probabilities for the results of the next toss.

(a) Specify clearly the set of hypotheses and the prior probabilities.

(b) Compute the prior and posterior predictive distributions, i.e. give the probabilities of all possible outcomes.

(a) Solution: The hypotheses are $C_{0.5}$, the chosen coin has probability 0.5 of landing heads. Likewise for $C_{0.6}$ and $C_{0.9}$.

(b) Solution: Let H and T stand for heads and tails. The prior predictive probability of heads is

$$P(H) = P(H|C_{0.5})P(C_{0.5}) + P(H|C_{0.6})P(C_{0.6}) + P(H|C_{0.9})P(C_{0.9}) = 0.5 \cdot \frac{1}{3} + 0.6 \cdot \frac{1}{3} + 0.9 \cdot \frac{1}{3} = 2/3.$$

So, the prior predictive probability of tails is $P(T) = 1/3$.

The data $\mathcal{D}$ is '1 head and 4 tails'. We first make a Bayesian update table to find the posterior probabilities (of hypotheses).

hypothesis	prior	likelihood	Bayes numerator	posterior
$\mathcal{H}$	$P(\mathcal{H})$	$P(\mathcal{D}|\mathcal{H})$	$P(\mathcal{D}|\mathcal{H})P(\mathcal{H})$	$P(\mathcal{H}|\mathcal{D})$
$C_{0.5}$	1/3	$\binom{5}{1}(0.5)^5$	0.0521	0.669
$C_{0.6}$	1/3	$\binom{5}{1}(0.6) \cdot (0.4)^4$	0.0256	0.329
$C_{0.9}$	1/3	$\binom{5}{1}(0.9) \cdot (0.1)^4$	0.00015	0.002
total	1		0.0778	1

Using the posterior probabilities (of hypotheses) we compute the posterior predictive probabilities of heads and tails.

$$P(H|\mathcal{D}) = 0.669 \cdot 0.5 + 0.329 \cdot 0.6 + 0.002 \cdot 0.9 = 0.53366$$
$$P(T|\mathcal{D}) = 1 - P(H|\mathcal{D}) = 0.46634.$$

So, the posterior predictice probabilities (for all outcomes) are

$$P(\text{heads}|\mathcal{D}) = 0.669 \cdot 0.5 + 0.329 \cdot 0.6 + 0.002 \cdot 0.9 = 0.53366$$
$$P(\text{tails}|\mathcal{D}) = 1 - P(\text{heads}|\mathcal{D}) = 0.46634.$$

Problem 2. Screening test odds
A disease is present in 0.005 of the population.

A screening test has a 0.05 false positive rate and a 0.02 false negative rate.

(a) Give the prior odds a patient has the disease

Assume the patient tests positive

(b) What is the Bayes factor for this data?

(c) What are the posterior odds they have the disease?

(d) Based on your answers to (a) and (b) would you say a positive test (the data) provides strong or weak evidence for the presence of the disease.

Solution: Let $\mathcal{H}_+$ = 'has disease' and $\mathcal{H}_-$ = 'doesn't have disease'. Let $\mathcal{T}_+$ = positive test

(a) $O(\mathcal{H}_+) = \dfrac{P(\mathcal{H}_+)}{P(\mathcal{H}_-)} = \dfrac{0.005}{0.995} = 0.00503$

(b) The likelihoods are $P(\mathcal{T}_+|\mathcal{H}_+) = 0.98$, $P(\mathcal{T}_+|\mathcal{H}_-) = 0.05$.

So the Bayes factor = likelihood ratio = $\dfrac{P(\mathcal{T}_+|\mathcal{H}_+)}{P(\mathcal{T}_+|\mathcal{H}_-)} = \dfrac{0.98}{0.05} = 19.6$

(c) Posterior odds = Bayes factor $\times$ prior odds = $19.6 \times 0.00504 = 0.0985$

(d) Yes, a Bayes factor of 19.6 indicates a positive test is strong evidence the patient has the disease. The posterior odds are still small because the prior odds are extremely small.

It would have been slower, but we could have computed the posterior odds by first computing the posterior probabilities using a Bayesian update table.

Problem 3. CSI Blood Types
We need to weigh the evidence at a crime scene:

- *Crime scene: the two perpetrators left blood: one of type O and one of type AB*

- *In population 60% are type O and 1% are type AB*

(a) Suspect Oliver is tested and has type O blood.

Compute the Bayes factor and posterior odds that Oliver was one of the perpetrators.

Is the data evidence for or against the hypothesis that Oliver is guilty?

(b) Same question for suspect Alberto who has type AB blood.

Hopefully helpful hints:

For the question about Oliver we have

Hypotheses:
  S = 'Oliver and another unknown person were at the scene'
  S^c = 'two unknown people were at the scene'

Data: D = 'type 'O' and 'AB' blood were found; Oliver is type O'

Prior odds: *These are unknown. We can just say $O(S)$.*

*From 'Information Theory, Inference, and Learning Algorithms' by David J. C. Mackay.

(a) Solution: (Oliver)

The trickiest part of this is seeing that $P(D|S^c) \approx 2 \cdot 0.6 \cdot 0.01$. We will show this below. So,
$$\text{Bayes factor} = \frac{P(D|S)}{P(D|S^c)} = \frac{0.01}{2 \cdot 0.6 \cdot 0.01} = 0.83.$$

Therefore the posterior odds = $0.83 \times$ prior odds. That is,

$$O(S|D) = 0.83 \cdot O(S).$$

Since the odds of his presence decreased this is (weak) evidence of his innocence.

Computing $P(D|S^c)$: Let's use the following notation:
n_O = number of people in the population of type O.
n_{AB} = number of people in the population of type AB.
n = total population.

So $n_O/n = 0.6$ and $n_{AB}/n = 0.01$.

The number of ways to choose two people from the population is $\binom{n}{2} = \dfrac{n(n-1)}{2}$. The number of ways to choose one person of type O who is not Oliver and one of type AB is $(n_O - 1) \cdot n_{AB}$. So

$$P(D|S^c) = \frac{(n_O - 1)n_{AB}}{n(n-1)/2} = 2 \cdot \frac{n_O - 1}{n} \cdot \frac{n_{AB}}{n-1} \approx 2 \cdot 0.6 \cdot 0.01.$$

Here we have assumed that n and n_O are large enough that subtracting 1 has a negligible effect, i.e.

$$\frac{n_O - 1}{n} \approx \frac{n_0}{n} = 0.6 \quad \text{and} \quad \frac{n_{AB}}{n-1} \approx \frac{n_{AB}}{n} = 0.01.$$

(b) For Alberto, the hypotheses are

$A = $ 'Alberto and another unknown person were at the scene'
$A^c = $ 'two unknown people were at the scene'

$$\text{Bayes factor} = \frac{P(D|A)}{P(D|A^c)} = \frac{0.6}{2 \cdot 0.6 \cdot 0.01} = 50.$$

Therefore the posterior odds $= 50 \times$ prior odds. That is,

$$O(A|D) = 50 \cdot O(A).$$

Since the odds of his presence increased this is (strong) evidence of his presence at the scene.

MIT OpenCourseWare

18.05 Introduction to Probability and Statistics
Spring 2022