

# Question Set 1 — Probability and Bayes Solutions

## Question 1

- (i) Sum rule ( $A$  and  $B$  are mutually exclusive, otherwise a third term would be needed);
- (ii) Product rule (completely general, but entire equation is within context of  $Z$ );
- (iii) Bayes' rule (the hypothesis is that  $a = 3$  AND  $b = 4$ , the data is that  $c = 5$ );
- (iv) It's the marginal likelihood formula, which involves the sum and product rules. The data is that  $x = 3$  and the two hypotheses are that  $y = 3$  and  $y = 4$ .

## Question 2

- (a) In this question it's best to assign the likelihoods just by imagining that you knew which hypothesis was true, and assigning a probability to describe how confidently you would predict  $D$ .

If we knew  $S$  were true, then the footprint would definitely be size 10. So  $P(D|S) = 1$ . If we knew  $M$  were true, there'd be a 50% chance that the footprint would be size 10. So  $P(D|M) = 0.5$ . If we knew  $U$  were true, then we'd predict a size 10 footprint with probability 10%, so  $P(D|U) = 0.1$ .

- (b) By three applications of Bayes' rule, we have

$$P(S|D) = \frac{P(S)P(D|S)}{P(D)} \tag{1}$$

$$= \frac{0.5 \times 1}{P(D)} \tag{2}$$

$$P(M|D) = \frac{P(M)P(D|M)}{P(D)} \tag{3}$$

$$= \frac{0.3 \times 0.5}{P(D)} \tag{4}$$

$$P(U|D) = \frac{P(U)P(D|U)}{P(D)} \tag{5}$$

$$= \frac{0.2 \times 0.1}{P(D)} \tag{6}$$

To get  $P(D)$ , we can use the fact that these three posterior probabilities should add to 1. Equivalently, the marginal likelihood is the sum of the prior×likelihood values:

$$P(D) = 0.5 \times 1 + 0.3 \times 0.5 + 0.2 \times 0.1 \quad (7)$$

$$= 0.67. \quad (8)$$

Therefore, the three posterior probabilities are

$$P(S|D) = \frac{0.5 \times 1}{0.67} = 0.7463 \quad (9)$$

$$P(M|D) = \frac{0.3 \times 0.5}{0.67} = 0.2239 \quad (10)$$

$$P(U|D) = \frac{0.2 \times 0.1}{0.67} = 0.0299. \quad (11)$$

The data made  $S$  more plausible,  $M$  a bit less plausible, and  $U$  a lot less plausible.

### Question 3

- (a) The four prior probabilities are 0.5, 0.1, 0.2, and 0.2. The likelihood for  $N$  is 0.1 by arbitrary assumption. For curbing, there are three possible vowels, but none of them are ‘EH’, so  $P(D_1|C) = 0$ . For overdrive, there are two possible vowels, so the probability of ‘EH’ given overdrive is  $P(D_1|O) = 0.5$ . For edge there are four possible vowels, so the probability of ‘EH’ given edge is  $P(D_1|E) = 0.25$ .

To use Bayes’ theorem, we need the marginal likelihood. This is

$$P(D_1) = 0.5 \times 0.1 + 0.1 \times 0 + 0.2 \times 0.5 + 0.2 \times 0.25 \quad (12)$$

$$= 0.2 \quad (13)$$

Therefore, by four applications of Bayes’ rule, the posterior probabilities of the four hypotheses are:

$$P(N|D_1) = \frac{0.5 \times 0.1}{0.2} = 0.25 \quad (14)$$

$$P(C|D_1) = \frac{0.1 \times 0}{0.2} = 0 \quad (15)$$

$$P(O|D_1) = \frac{0.2 \times 0.5}{0.2} = 0.5 \quad (16)$$

$$P(E|D_1) = \frac{0.2 \times 0.25}{0.2} = 0.25 \quad (17)$$

The data ruled out curbing, made overdrive and edge more plausible, and neutral less plausible.

- (b) We already did this, but here it is again...

$$P(D_1) = P(N)P(D_1|N) + P(C)P(D_1|C) + P(O)P(D_1|O) + P(E)P(D_1|E).$$

- (c) As the question suggests, you can use the posterior probabilities after taking into account one piece of data as the new prior probabilities for when you want to take a second piece of data into account. The four likelihoods for this second piece of data are  $1/3$ ,  $0$ ,  $1$ , and  $1$ . The marginal likelihood is

$$0.25 \times 1/3 + 0 \times 0 + 0.5 \times 1 + 0.25 \times 1 = 5/6. \quad (18)$$

The posterior probabilities are

$$P(N|D_1, D_2) = \frac{0.25 \times 1/3}{5/6} = 0.1 \quad (19)$$

$$P(C|D_1, D_2) = \frac{0 \times 0}{5/6} = 0 \quad (20)$$

$$P(O|D_1, D_2) = \frac{0.5 \times 1}{5/6} = 0.6 \quad (21)$$

$$P(E|D_1, D_2) = \frac{0.25 \times 1}{5/6} = 0.3 \quad (22)$$

The fact that the note was loud supports overdrive and edge.

- (d) You can get the same result from (c) by going back to the original priors, and doing a single update which considers both pieces of data together. If you do things this way, the likelihoods need to be the product of the likelihoods for the two pieces of data. In this case, the data is that the vowel was ‘EH’ AND that the note was loud.

The marginal likelihood is

$$\begin{aligned} P(D_1, D_2) &= P(N)P(D_1, D_2|N) + P(C)P(D_1, D_2|C) \\ &\quad + P(O)P(D_1, D_2|O) + P(E)P(D_1, D_2|E) \\ &= 0.5 \times 0.1 \times 1/3 + 0.1 \times 0 \times 0 \\ &\quad + 0.2 \times 0.5 \times 1 + 0.2 \times 0.25 \times 1 \\ &= 1/6 \end{aligned}$$

and the posterior probabilities are

$$P(N|D_1, D_2) = \frac{0.5 \times 0.1 \times 1/3}{1/6}$$

$$P(C|D_1, D_2) = \frac{0.1 \times 0 \times 0}{1/6}$$

$$P(O|D_1, D_2) = \frac{0.2 \times 0.5 \times 1}{1/6}$$

$$P(E|D_1, D_2) = \frac{0.2 \times 0.25 \times 1}{1/6}$$

which gives the same results as in part (c).

These two ways of calculating the posterior for a hypothesis  $H$  given the two datasets can be expressed as

$$P(H|D_1, D_2) = \frac{P(H|D_1)P(D_2|D_1, H)}{P(D_2|D_1)} = \frac{P(H)P(D_1, D_2|H)}{P(D_1, D_2)}. \quad (23)$$