Worksheet 2: outline solutions

Q1 - marginal and conditional distributions

Work out the marginal probability distributions and the $x = a$ conditional probability distribution $P(Y|X = a)$ for

$Y \backslash X$	a	b
1	$\frac{1}{3}$	$\frac{1}{6}$
2	0	$\frac{1}{4}$
3	$\frac{1}{8}$	$\frac{1}{8}$

Solution

Calculating the marginal distribution just requires adding along the rows or columns, if we take X to be the random variable going across, so $\mathcal{X} = \{a, b\}$ and Y the random variable going down, so $\mathcal{Y} = \{1, 2, 3\}$ then the two marginal distributions are, for X

	a	b
	$\frac{11}{24}$	$\frac{13}{24}$

and for Y

	1	2	3
	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$

The conditional distribution, conditioned on $x = a$ is calculated using $p(y|x)p(x) = p(x, y)$ so we divide the $p(x = a, y)$ column by $p(x = a) = 11/24$, hence

	1	2	3
	$\frac{8}{11}$	0	$\frac{3}{11}$

Q2 - working out entropy

A fair coin is flipped until the first head occurs. Let X denote the number of flips required.

1. Find the entropy $H(X)$ in bits. The following expressions may be useful:

$$\begin{aligned}\sum_{n=0}^{\infty} r^n &= \frac{1}{1-r}\\ \sum_{n=0}^{\infty} nr^n &= \frac{r}{(1-r)^2}\end{aligned} \tag{1}$$

2. A random variable X is drawn according to this distribution. Find a sequence of yes-no questions of the form, 'Is X contained in the set S?'. Compare $H(X)$ to the expected number of questions required to determine X. For the most efficient sequence, that is the sequence the shortest expected number of questions, these two numbers will be the same. You will find that the most efficient sequence is very straightforward!

Solution

So, here the set of possible outcomes is $\mathcal{X} = \{1, 2, 3, \ldots\}$ and we need to start by working out $p_X n$ the chance of throwing n flips before getting a head. To get $X = n$ you need to throw $n-1$ tails, this has probability $1/2^{n-1}$ followed by a head, which has probability $1/2$, hence

$$p_X(n) = \frac{1}{2^n} \tag{2}$$

It is easy to check that

$$\sum_{n=1}^{\infty} p_X(n) = \sum_{n=1}^{\infty} \frac{1}{2^n} = \sum_{n=0}^{\infty} \frac{1}{2^n} - 1 = 1 \tag{3}$$

Now, to calculate the entropy, just use the formula:

$$\begin{aligned}H(X) &= -\sum_{n\in\mathcal{X}} p_X(x)\log p_X(x)\\ &= -\sum_{n=1}^{\infty} \frac{1}{2^n}\log 2^{-n} = \sum_{n=1}^{\infty} \frac{n}{2^n}\\ &= 2\end{aligned} \tag{4}$$

It is easy to see that this is the same as the average number of questions asked in order starting with $n = 1$ of the form 'is the answer n?' it would take to find X .

Q3 - A puzzle which lends itself to information type reasoning

Suppose that you have n coins, among which there may or may not be one counterfeit coin. If there is a counterfeit coin it will weigh either less or more than the other coins. The coins are weighed using a balance, any number of coins van be put on each side of the balance, though obviously you will want the same number on each side.

1. Find an upper bound on the number of coins n so that k weighings will find the counterfeit coin, if any, and correctly declare it to be heavier or lighter.

2. What is the coin-weighing strategy for $k = 3$ weighings and 12 coins,

Solution

So given that one of n coins is counterfeit; there are $2n$ possible configurations, numbering the coins one to n, each possibility is either of the form the ith coin is heavier, or the ith coin is lighter. Thus, assuming all possibilities are equally likely, the random variable X giving the identity and type of the bent coin has entropy $H(X) = \log 2n + 1$, the plus one is for the possibility that there is no counterfeit coin. What about weighing, Y, well each weighing involves taking two groups of coins and balancing them and this has three possible outcomes: left heavier, right heavier or balanced. Obivously, depending on what we have already worked out about the coins from previous weighings, these possibilities have different outcomes, for example, at the start, given that one coin is counterfeit, weighing $n/2$ coins against $n/2$ coins can't give balanced and the entropy for this measurement will be one bit. However, we know that $H(Y) < \log 3$; the most uncertain measurement is the one where all possibilities are equally likely.

Now, imagine drawing up a weighing statedgy, you are going to do k weighings Y_1, Y_2 to Y_k. The outcome of a weighing is determined by the value of x, the idenity and type of the bent coin, so $H(Y_1, Y_2, \ldots, Y_k|X) = 0$. We have

$$\begin{aligned} H(X) + H(Y_1, Y_2, \ldots, Y_k|X) &= H(X, Y_2, \ldots, Y_k) \\ &= H(Y_1, Y_2, \ldots, Y_k) + H(X|Y_1, Y_2, \ldots, Y_k) \end{aligned} \quad (5)$$

If we have a statedgy that locates and types the counterfeit, there should be no uncertainty in X given the Y_i so $H(X|Y_1, Y_2, \ldots, Y_k) = 0$. So, if we are able to find and type the counterfeit

$$H(X) = H(Y_1, Y_2, \ldots, Y_k) \tag{6}$$

but, from the independence theorem and the bound above

$$H(X) = H(Y_1, Y_2, \ldots, Y_k) \leq H(Y_i) \leq k \log 3 \tag{7}$$

and hence

$$n \leq \frac{3^k - 1}{2} \tag{8}$$

Hence, if it is possible to identify and type the coin in k weighings, we know we have less than $(3^k - 1)/2$ coins. This bound may not be sharp, for particular values of k it may not be possible to choose a stategy so each Y has $H(Y) = \log 3$ or so that the entropy of the joint distribution is equal to the sum of the entropies of the marginal distribution. However, we do have a bound.

For $k = 3$ we have $n \leq 13$, in fact, there doesn't seem to be a solution for $n = 13$; there is one for $n = 12$. Lets start by numbering the coins from one to 12. The first weighing is $g_1 = \{1, 2, 3, 4\}$ versus $g_2 = \{5, 6, 7, 8\}$. If g_1 is heavier; then weigh $g_3 = \{1, 2, 5\}$ versus $g_4 = \{3, 4, 6\}$. Thus g_3 and g_4 each have two coins which must be heavier if they are counterfeit and the remaining two coins, 7 and 8, must be lighter. If g_3 is heavier than g_4 this can only be because either 1 or 2 is heavier, or 6 is lighter; weighing 1 or 2 settles this, if one is heavier than the other, it is the bent coin, if they balance, 5 is. If g_3 and g_4 balance then the counterfeit is either 7 or 8 and weighing them gives the answer. Finally, if g_1 and g_2 balance the counterfeit coin must be one of $\{9, 10, 11, 12\}$; start by weighting $g_5 = \{9, 10\}$ against $g_6 = \{11, 1\}$: 1 is known not to be counterfiet. If g_5 and g_6 balance then the coin can only be 12 and weighing this against 1 gives the answer, otherwise, say g_5 is heavier than, either one of 9 and 10 is heavy, or 11 is light, weighing 9 against 10 sorts this out.

Q4 - Working out entropy and information

Let $p(x, y)$ be given by $p(0, 0) = p(0, 1) = p(1, 1) = 1/3$ and $p(1, 0) = 0$. Find $H(X)$, $H(Y)$, $H(X|Y)$, $H(Y|X)$, $H(X, Y)$, $H(Y) - H(Y|X)$ and $I(X; Y)$.

Solution

So this is just a lot of calculating. First lets work out the two marginal distributions: $p_X(0) = 2/3$, $p_X(1) = 1/3$, whereas $p_Y(0) = 1/3$ and $p_Y(1) = 2/3$, hence

$$H(X) = H(Y) = \frac{1}{3}\log_2 3 + \frac{2}{3}\log_2 \frac{3}{2} \approx 0.92 \tag{9}$$

Now for the conditional distributions $p(0|Y = 0) = 1$, $p(1|Y = 0) = 0$, but $p(0|Y = 1) = p(1|Y = 1) = 0.5$ and hence

$$\begin{aligned} H(X|Y = 0) &= 0 \\ H(X|Y = 1) &= 1 \end{aligned} \tag{10}$$

and hence, given $p_Y(1) = 2/3$ we have $H(X|Y) = 2/3$. $H(Y|X)$ is the same. Now

$$H(X, Y) = 3\frac{1}{3}\log_2(3) \approx 1.58 \tag{11}$$

Finally we could use $I(X; Y) = H(X)+H(Y)-H(X, Y) \approx 0.25$ or $I(X, Y) = H(X|Y) - H(X) \approx 0.25$

Q5 - A question about information in the brain

Answer just one of these two questions, each is worth equal marks but the second is much harder than the first, so you'd be better off doing the first unless you are particularly interested in this topic. Both papers are available in the paper repository in the github.

1. The original idea of estimating neural information by binning spike trains was spread across several papers, but one of the main references is **?**. One aspect of this paper we didn't discuss is the use of extrapolation to estimate the information as the number of samples becomes large based on the behaviour for smaller numbers of samples. Can you give a short, up to five line, summary of what this involves.

2. In **?** there is a deep commentary on how information in neural data is computed. This is a very difficult paper and the mathematics towards the end is hard. The aim of this question is to read the paper and offer a three or four line overall summary of what the paper is trying to do.

Solution

No fixed answer, you need to read the paper!