Worksheet 1 - 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	$\frac{1}{16}$	$\frac{1}{2}$
2	0	$\frac{1}{4}$
3	$\frac{1}{16}$	$\frac{1}{8}$

Solution

The marginal distributions come from summing the rows or columns so

X	a	b
	$\frac{1}{8}$	$\frac{7}{8}$

and

Y	1	2	3
	$\frac{9}{16}$	$\frac{1}{4}$	$\frac{3}{16}$

Finally, $P(X = a) = 1/8$ so

$Y|X = a$	1	2	3
	$\frac{1}{2}$	0	$\frac{1}{2}$

Q2 - working out entropy

For the above distribution work out $H(X)$, $H(Y)$, $H(X|Y)$, $H(Y|X)$, $H(X, Y)$, $H(Y) - H(Y|X)$ and $I(X; Y)$.

Solution

So all that happens here, again and again, is that you need to work out something like

$$H(X) = -\sum_i p_i \log p_i \tag{1}$$

So

$$H(X) = -\frac{1}{8}\log\frac{1}{8} - \frac{7}{8}\log\frac{7}{8} \approx 0.544 \tag{2}$$

and

$$H(Y) = -\frac{9}{16}\log\frac{9}{16} - \frac{1}{4}\log\frac{1}{4} - \frac{3}{16}\log\frac{3}{16} \approx 1.42 \tag{3}$$

and

$$H(X,Y) = -\frac{1}{16}\log\frac{1}{16} - \frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\log\frac{1}{4} - \frac{1}{16}\log\frac{1}{16} - \frac{1}{8}\log\frac{1}{8} = 1.875 \tag{4}$$

The conditional distributions are more complicated to calculate since we need to work out one for each condition:

$$H(Y|X=a) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{2}\log\frac{1}{2} = 1 \tag{5}$$

and

$$H(Y|X=b) = -\frac{4}{7}\log\frac{4}{7} - \frac{2}{7}\log\frac{2}{7} - \frac{1}{7}\log\frac{1}{7} \approx 1.38 \tag{6}$$

so

$$H(Y|X) \approx \frac{1}{8} + \frac{7}{8}1.38 \approx 1.33 \tag{7}$$

and

$$H(X|Y=1) = -\frac{1}{9}\log\frac{1}{9} - \frac{8}{9}\log\frac{8}{9} \approx 0.50 \tag{8}$$

and

$$H(X|Y=2) = 0 \tag{9}$$

and

$$H(X|Y=3) = -\frac{1}{3}\log\frac{1}{3} - \frac{2}{3}\log\frac{2}{3} \approx 0.92 \tag{10}$$

which together tells us

$$H(X|Y) \approx \frac{9}{16}0.5 + \frac{3}{16}0.92 \approx 0.45 \tag{11}$$

Finally

$$H(Y) - H(Y|X) \approx 1.42 - 1.33 = 0.09 \tag{12}$$

and this should be equal $I(X;Y)$; we could work this out another way using

$$I(X;Y) = H(X) + H(Y) - H(X,Y) \approx 1.33 + 0.54 - 1.87 = 0.09 \tag{13}$$

Q3 - working out entropy

The World Series is a competition held each year in North America between two baseball teams. The series consists of between four and seven games, terminating if either team wins four games. Thus, the set of outcomes includes sequences like AAAA, ABAAA and ABABABA. Let X be the random variable representing the outcome and Y the number of games played. Assuming the teams are equally matched and the games are independent, what are $H(X)$, $H(Y)$, $H(X|Y)$ and $H(Y|X)$?

Solution

So if the two teams are equally matched then the four game series each have probability $1/2^4$; the five series $1/2^5$ and so on. There are two four game series, $\{AAAA\}$ and $\{BBBB\}$, for five games there are four choose one series for each of an A and B victory because the last game has to be the fourth by the winning team, for six games there are two times five choose two, and so on.

This gives us a table:

games	probability per series	number of series	total probability
4	$\frac{1}{16}$	2	$\frac{1}{8}$
5	$\frac{1}{32}$	8	$\frac{1}{4}$
6	$\frac{1}{64}$	20	$\frac{5}{16}$
7	$\frac{1}{128}$	40	$\frac{5}{16}$

and hence

$$H(Y) = -\frac{1}{8}\log\frac{1}{8} - \frac{1}{4}\log\frac{1}{4} - \frac{5}{16}\log\frac{5}{16} - \frac{5}{16}\log\frac{5}{16} \approx 1.92 \tag{14}$$

whereas for X you are summing over all the outcomes, so for the length four outcomes you have two possibilites and this adds

$$-\frac{2}{\times}\frac{1}{16}\log_2\frac{1}{16} = \frac{1}{8}\times 4 \tag{15}$$

to the total whereas the 8 length five outcomes add

$$-\frac{8}{\times}\frac{1}{32}\log_2\frac{1}{32} = \frac{1}{4}\times 5 \tag{16}$$

and putting it all together we get

$$H(X) = \frac{1}{8}4 + \frac{1}{4}5 + \frac{5}{16}6 + \frac{5}{16}7 \approx 5.82 \tag{17}$$

Now $H(Y|X) = 0$ since the outcome determines the series length, the other way around is more complicated. For a given length, that is a given value of Y, all the outcomes are equally likely so the conditioned entropy $H(X|Y = y)$ equals the number of elements, hence summing over the conditioned entropies we get

$$H(X|Y) = \frac{1}{8}\log 2 + \frac{1}{4}\log 8 + \frac{5}{16}\log 20 + \frac{5}{16}\log 40 \approx 3.89 \tag{18}$$

Q4 - the average entropy

Work out the average entropy for the distribution with two events $\{x_1, x_2\}$ and $p(x_1) = p$ and $p(x_2) = 1 - p$ under the assumption that each value of p is equally likely.

Solution

So $H(p) = -p\log p - (1-p)\log 1 - p$ so to get get average entropy we need

$$\langle H\rangle_p = \int_0^1 H(p)dp = -\int_0^1 p\log pdp - \int_0^1 (1-p)\log(1-p)dp \tag{19}$$

Given that the choice of which probability to call p and which to call $1 - p$ is arbitrary you might expect

$$\int_0^1 p\log pdp = \int_0^1 (1-p)\log(1-p)dp \tag{20}$$

and this is easy to check by substituting $q = 1 - p$. Hence

$$\langle H\rangle_p = -2\int_0^1 p\log pdp \tag{21}$$

To make the integration easier lets use the entropy measured in nats:

$$\langle H_e\rangle_p = -2\int_0^1 p\ln pdp \tag{22}$$

It is easy to convert back since $\log p = \ln p/\ln 2$. Now use integration by parts with $u = \log p$ and $dv = pdp$ so

$$\int_0^1 p \ln p dp = \frac{p^2}{2} \ln p]_0^1 - \int_0^1 \frac{p^2}{2}\frac{1}{p}dp = -\left.\frac{p^2}{4}\right]_0^1 = -\frac{1}{4} \tag{23}$$

and hence

$$\langle H \rangle_p = \frac{1}{2\ln 2} \approx 0.72 \tag{24}$$