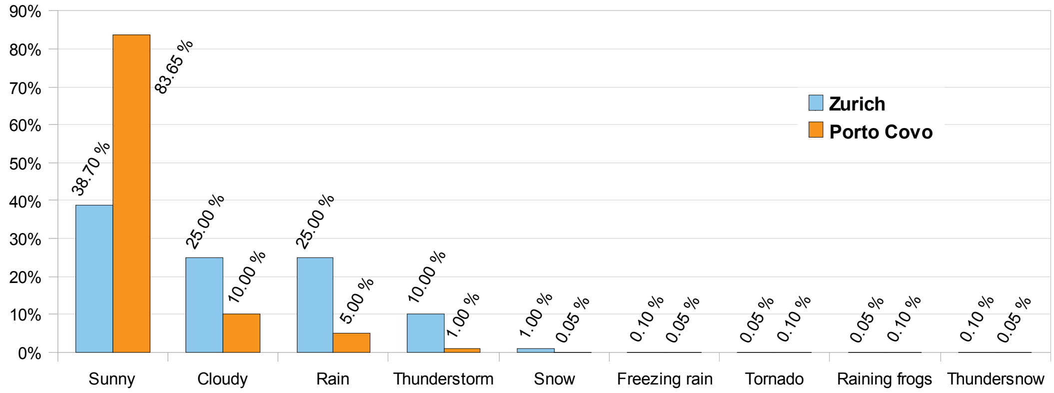
## Exercise 2.1 Entropy as a measure of uncertainty




a) *How long should a list of all weather possibilities be for both cities? What if we dismissed very unlikely events?*

There are 9 possibilities in each case, so the lists would have that size. The max-entropy $H_{\max}(X)_P = \log |P_X|$ give us the logarithm of that value.

If we accept an error tolerance $\epsilon$ we can ignore all the events with probabilites that sum up to $\epsilon$ (so that with probability $\epsilon$ something will happen that is not on our list). For instance, if we choose $\epsilon = 2\%$ we can dismiss the possibility of snow, freezing rain, tornados, raining frogs and thundersnow in Zurich (and our list would have 4 entries) and further ignore thunderstorms in Porto Covo, needing a list with only 3 items. This corresponds to take $\epsilon = 0.02$ and calculate the smooth max-entropy of the two probability distributions defined by the weather possibilities. The size of the lists is $2^{H^{\epsilon}_{\max}}$.

b) *How likely are you to correctly guess the weather in each place? Relate that to the classical min-entropy.*

Our best strategy is, as usual, to bet on the most likely outcome and your probability of winning is precisely the probability that referred outcome occurs – 83.75% in the case of Porto Covo and 38.70% in Zurich. In terms of entropies this is $2^{-H_{\min}}$. We have that $H_{\min}(\text{Porto Covo}) = 0.256$ and $H_{\min}(\text{Zurich}) = 1.37$ : the lower the min-entropy, the more likely we are make a correct guess.

## Exercise 2.2 Mutual Information

a) *Compute the mutual information between your guess and the actual weather, and do the same for your grandfather. Remember that all your grandfather knows is that it rains on 80% of the days. You know that as well and you also listen to the weather forecast and know that it is right 80% of the time and is always correct when it predicts rain.*

The mutual information is given by $I(X:Y) = H(X) - H(X|Y)$ . Let us call your grandfather $G$, you $Y$ and the actual weather $W$. We may also assume you followed the radio forecast (which we saw was an optimal strategy) and we will hold on to the notation $\hat{R}, \hat{S}$ for guesses (both yours and your grandfather's). Then we have, for the grandfather

$$\begin{aligned}
H(W) &= -P(R)\log P(R) - P(S)\log P(S)\\
&= -0.8\log 0.8 - 0.2\log 0.2\\
H(G) &= -P(\hat{R})\log P(\hat{R}) - P(\hat{S})\log P(\hat{S})\\
&= -1.\log 1. - 0. = 0\\
H(GW) &= -P(\hat{R}R)\log P(\hat{R}R) - P(\hat{S}R)\log P(\hat{S}R) - P(\hat{R}S)\log P(\hat{R}S) - P(\hat{S}S)\log P(\hat{S}S)\\
&= -0.8\log 0.8 - 0 - 0 - 0.2\log 0.2\\
H(W|G) &= H(GW) - H(G)\\
&= -0.8\log 0.8 - 0.2\log 0.2\\
I(W:G) &= H(W) - H(H|G)\\
&= 0.
\end{aligned}$$

For your case we will calculate the conditional entropy directly,

$$\begin{aligned}
H(W) &= -0.8\log 0.8 - 0.2\log 0.2\\
H(W|Y) &= -P(\hat{R})\left[P(R|\hat{R})\log P(R|\hat{R}) + P(S|\hat{R})\log P(S|\hat{R})\right]\\
&\quad - P(\hat{S})\left[P(R|\hat{S})\log P(R|\hat{S}) + P(S|\hat{S})\log P(S|\hat{S})\right]\\
&= -0.6[1\log 1 + 0] - 0.4[0.5\log 0.5 + 0.5\log 0.5]\\
&= -0.4\log 0.5\\
I(W:Y) &= H(W) - H(H|Y)\\
&= -0.8\log 0.8 - 0.2\log 0.2 + 0.4\log 0.5\\
&= 0.32.
\end{aligned}$$

b) *You read about a betting game that you think might help you convince your grandfather about the weather forecast (see Cover and Thomas, Chapter 6).*

(i) *Assuming that the race outcomes, your betting strategy and the payout are i.i.d., and considering the asymptotics for $N \to \infty$, show that the* doubling rate *is*

$$W(b,p) = E(\log S(X)) = \sum_{k=1}^{M} p_k \log b_k g_k,$$

*where $S_N = 2^{NW(b,p)}$.*

If the race outcomes $X_1, X_2, ...X_n$ are i.i.d., the functions $\log S(X_1), \log S(X_2), ... \log S(X_n)$ are also i.i.d. Hence, we can apply the weak law of large numbers to see that

$$\frac{1}{n}\log S_n = \frac{1}{n}\sum_{i=1}^{n}\log S(X_i) \to E(\log S(X)) \quad \text{for } n \to \infty.$$

Hence, asymptotically, we find that

$$S_n = 2^{nE(\log S(X))} \tag{1}$$

and so the doubling rate is indeed given by

$$W(b,p) = E(\log S(X)) = \sum_{k=1}^{M} p_k \log b_k g_k,$$

(ii) *Show that the optimal asymptotic doubling rate is*

$$W_{opt}(p) = \sum_i p_i \log(g_i) - H(p).$$

The optimal asymptotic doubling rate is given by the maximum over all betting strategies, which is the maximum over all assignments $\mathbf{b}$:

$$W^*(p) = \max_b W(b,p) = \max_b \sum_{i=1}^{m} p_i \log b_i g_i$$

We use a Lagrange multiplier to maximise subject to the constraint $\sum_i b_i = 1$. The functional can then be written as

$$J(b) = \sum p_i \log b_i g_i + \lambda \sum_i b_i$$

Differentiating with respect to $b_i$ gives

$$\frac{\partial J}{\partial b_i} = \frac{p_i}{b_i} + \lambda.$$

For the maximum, we want that $\frac{\partial J}{\partial b_i} = 0$, and so

$$b_i = \frac{p_i}{\lambda},$$

which, together with the constraint $\sum_i b_i = 1$ yields $\lambda = -1$ and $b_i = p_i$.
It is left to show that this stationary point is indeed a maximum. This can be done in several ways, for example by taking second derivatives (rather involved calculation) or as described in Cover and Thomas, Chapter 6 (much simpler) – we will not repeat their argument here.

(iii) *How does the strategy in (bii) differ from the strategy to optimize your expected wealth after a single bet?*
For a single bet, the best strategy is to bet everything on the horse with highest probability. If that horse is denoted by the index k, the choice is hence $b_k = 1$ and $b_j = 0$ for all $j \neq k$, rather than $\mathbf{b} = \mathbf{p}$ as above.

c) *You would now like to use this newly found knowledge of betting games to show your grandfather that you have more information about the weather. You and your grandfather start with £1. Every night each of you can bet some proportion $b_R$ and $b_S$ of your money on whether it will be rainy of sunny the next day. You must both bet all of your money each night. If your guess was right you double the amount you bet. All winnings will be used in future rounds.*

*Assuming you would repeat this many, many times, in independent rounds, what would your strategy be? And your grandfather's?*

The example is now exactly like the horse race above. To maximize your expected wealth in the asymptotic setting corresponds to maximizing the doubling rate, and so you as well as your grandfather would choose $b = p$. However, note that you and your grandfather assign different probabilities to the events of rain and sunshine as you have different knowledge about the events. Hence, for you the optimal strategy is to bet $b = p_Y$, while for your grandfather it is to bet $b = p_{GF}$, where

$$p_Y = p(\text{weather}|\text{forecast})$$
$$p_{GF} = p(\text{weather}) = (0.8, 0.2)$$

d) *Applying the strategy from c), what is the expected wealth for each of you after the first round?*

For your grandfather, the expected wealth after one round is

$$£1 \cdot \langle S_1 \rangle = £\langle S(X) \rangle = £\sum_i p_i b_i g_i = £2 \cdot (0.8^2 + 0.2^2) = £1.36,$$

For you, the relevant probabilities are conditioned on the forecast, and so we first have to distinguish the two cases:

Forecast predicts rain:

In this case, the relevant probabilities $p_i$ are $(1, 0)$. The expected wealth after one round is hence £$1 \cdot \langle S_R \rangle =$ £2.

Forecast predicts sunny:

In this case, the relevant probabilities $p_i$ are $(0.5, 0.5)$ as established in the last exercise sheet. The expected wealth after one round is hence £$1 \cdot \langle S_S \rangle =$ £$2 \cdot (0.5^2 + 0.5^2) =$ £1.

The overall expected wealth is hence

$$\text{£}1 \cdot \langle S_1 \rangle = \text{£}1 \cdot (p_{\hat{R}} \langle S_R \rangle + p_{\hat{S}} \langle S_S \rangle) = \text{£}(2p_{\hat{R}} + p_{\hat{S}}) = \text{£}1.6,$$

where $p_{\hat{R}} = 0.6$ is the probability of the forecast predicting rain and $p_{\hat{S}} = 0.4$ is the probability of the forecast predicting sunny.

As we can see from this example, even after one round the expected wealth for you is larger than for your grandfather, if both of you are using asymptotically sensible strategies (of course your grandfather could choose to always bet everything on rain. Then his expected wealth after one round would also be £1.6, but of course if the game is played many rounds he will eventually lose all his money). Hence, in this version of the game, listening to the forecast really helps!

e) *Still applying the strategy from c), what is the probability that your grandfather has more money than you after the first round?*

The only case in which your grandfather has more money than you after round 1 is when the forecast predicts sunny, but it nevertheless rains. This happens with probability $p_{\hat{S}} \cdot p(R|\hat{S}) = 0.4 \cdot 0.5 = 0.2$. In this case, your wealth after one round is £0.5 as you will have betted 50/50. Your grandfather will have betted 80/20, and so he has £1.6.