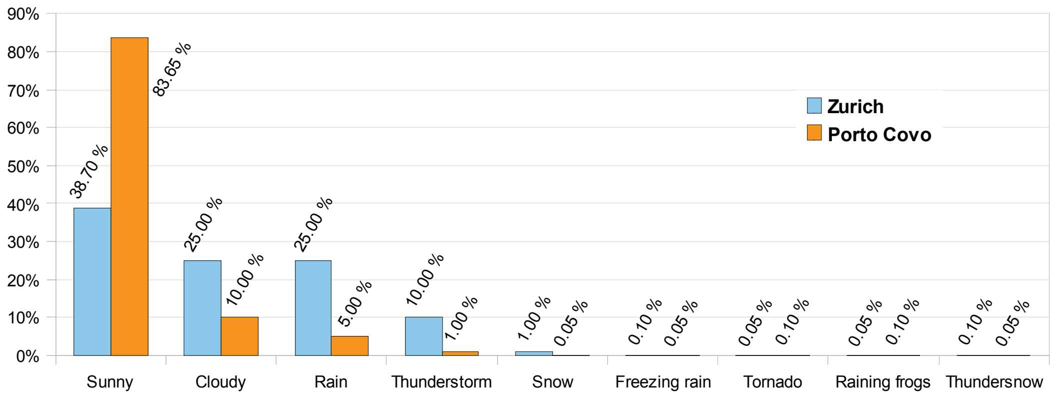
Exercise 1. *Entropy as a measure of uncertainty*

These two graphs represent the probability distributions of the weather conditions for a summer day in Zurich and Porto Covo. We will try to quantify the uncertainty we have about the weather in both cases using some entropy measures. Here $\log \equiv \log_2$.



(a) Suppose you want to make lists of all the weather possibilities in both places (for instance, to decide how many different sets of clothes you need when visiting those places, to be on the safe side). How long would the two lists be?

Realistically, you do not expect snow in Porto Covo or tornados in Zurich on a summer day—you can safely leave those possibilities out of your lists if you allow for a very small error tolerance. How long are the lists if you dismiss very unlikely events? Relate those results to the max-entropy,

$$H_{\max}(X)_P = \log |P_X|, \tag{1}$$

where $|P_X|$ is the size of the support of P_X (i.e. the number of outcomes with non-zero probability), and to its smooth version,

$$H^{\epsilon}_{\max}(X)_P = \min_{Q_X \in \mathcal{B}^{\epsilon}(P_X)} H_{\max}(X)_Q, \tag{2}$$

where the minimum goes over all probability distributions Q_X that are ϵ-close to P_X according to the trace distance.

(b) How likely are you to correctly guess the weather in each place? Relate that to the classical min-entropy of a probability distribution P_X over $\mathcal{X}$ is defined as

$$H_{\min}(X)_P = -\log \max_{x \in \mathcal{X}} P_X(x). \tag{3}$$

Exercise 2. *Mutual Information*

After losing a bet with your Scottish grandfather about whether listening to the radio forecast would help you predict the weather, you have been studying information theory compulsively to try to come up with a clever argument that would make him stop mocking you. You are convinced that even though you did not guess correctly more often than he, you somehow have more *information* about the weather than he does.

(a) The mutual information between two random variables is given by

$$I(X:Y)_P = H(X)_P - H(X|Y)_P, \tag{4}$$

where $H(X)$ is the Shannon entropy of X,

$$H(X)_P = \langle -\log P_X(x)\rangle_x = -\sum_x P_X(x)\log P_X(x) \tag{5}$$

and $H(X|Y)$ is the conditional Shannon entropy of X given Y,

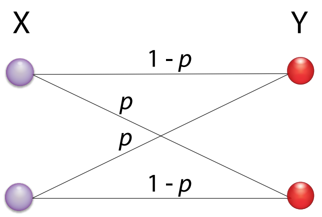
$$\begin{aligned} H(X|Y)_P &= \big\langle -\log P_{X|Y=y}(x)\big\rangle_{x,y} = -\sum_{x,y} P_{XY}(x,y)\ \log P_{X|Y=y}(x) \\ &= H(XY)_P - H(Y)_P. \end{aligned} \tag{6}$$

Compute the mutual information between your guess and the actual weather, and do the same for your grandfather. Remember that your grandfather knows it rains on 80% of the days. You also listen to the forecst, knowing it is right 80% of the time and always correct when it predicts rain.

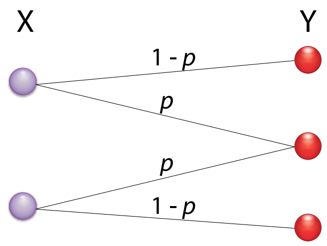
(b) You devise the following betting game to prove that your extra information is useful. You and your grandfather start with £1. Every night each of you can bet part of your money on the next day's weather. If your guess was right you double the amount you bet (e.g., in the first night your grandfather bets £0.2 on rain; if it rains he ends up with £1.2, otherwise with £0.8). Any winnings can be used in future rounds.

What is your optimal strategy for betting, after listening to the weather forecast? What is your grandfather's optimal strategy? After 30 days, what do you expect your total money will be? And your grandfather's?

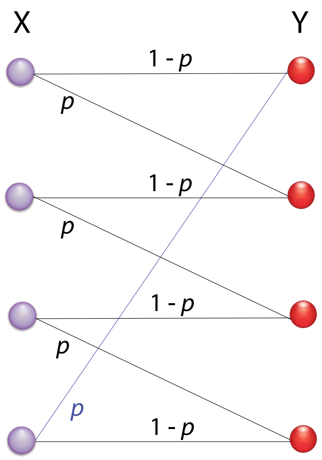
Exercise 3. ***Channel capacity***



(a) Binary Symmetric Channel



(b) Symmetric Erasure Channel



(c) Yet Another Channel

(a) The asymptotic channel capacity is given by

$$C = \max_{P_X} I(X : Y).$$

Calculate the asymptotic capacities of the first two channels depicted above.

(b) We can exploit the symmetries of some channels to simplify the calculation of the capacity. Consider N possible probability distributions as input to a general channel, $\left\{P_X^i\right\}_i$, with the property that $I(X : Y)_{P^i} = I(X : Y)_{P^j}, \forall i, j$. Suppose you choose which distribution to use for the input by checking a random variable, B, with possible values $b = \{1, \dots, N\}$. Show that in this case $I(X : Y|B) \leq I(X : Y)$. [1]

(c) How can you use that to find the probability distribution P_X that maximises the mutual information for symmetric channels?
Hint: Consider $\left\{P_X^i\right\}_i$ permutations of P_X^1.

(d) Using the result from (b), compute the capacity of the last channel. How would you proceed to reliably transmit one bit of information?

[1]Notice that this inequality only holds for the specific case treated here. If X, Y and B are correlated in a different way this inequality does not have to be true.