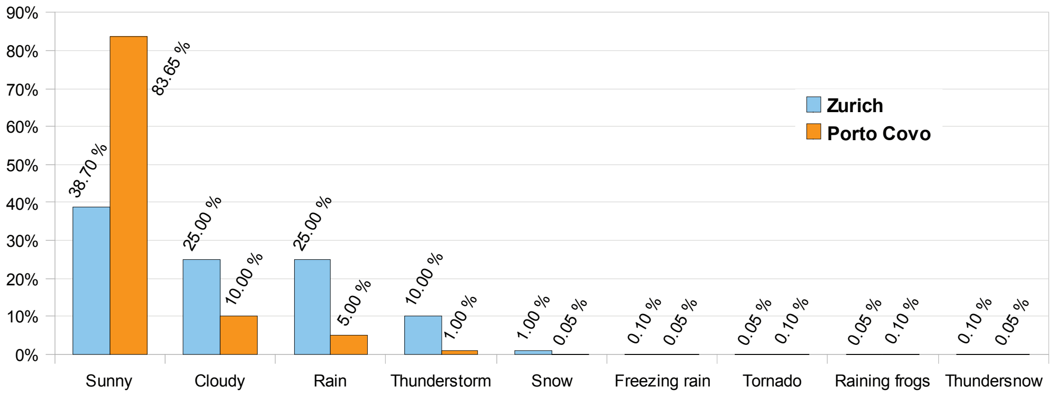
ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Quantum Information Theory
Solutions 2.

HS 2015
Prof. R. Renner

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*

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

Solution.

(a) 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 ϵ we can ignore all the events with probabilites that sum up to ϵ (so that with probability ϵ 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) 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 make a correct guess.

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,

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

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?

Solution.

(a) 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 are right—the exercise did not specify what we meant by "optimal strategy". What are we trying to optimise: the expectation value of our money after 30 days or the probability of having more money than the other player?

Let us first compute your grandfather's expected gain after n days, if he always bids a fraction b of his money on rain. Let $P_R(k)$ be the probability that there are exactly k rainy days. We have

$$\begin{aligned}
\langle £_G \rangle &= \sum_{k=0}^{n} P_R(k)\ (1+b)^k (1-b)^{n-k} \\
&= \sum_{k=0}^{n} \binom{n}{k} 0.8^k\ 0.2^{n-k}\ (1+b)^k (1-b)^{n-k} \\
&= (1 + 0.6\ b)^n,
\end{aligned} \tag{S.1}$$

which is maximised if $b = 1$, i.e., if your grandfather bids all his money on rain every evening. In this case, his expected gain is 1.6^n Of course, his probability of winning anything at all is quite small for large n: only 0.8^n, which is the likelihood of having n rainy days in a row (in which case he winds up with $£2^n$). With probability $1 - 0.8^n$ he will lose all his money on the first sunny day.

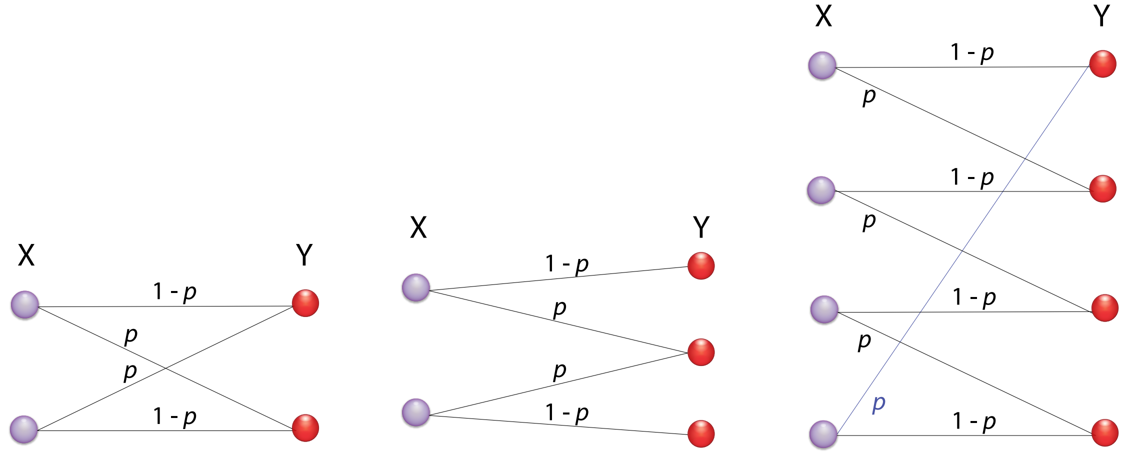
As for you, how much you bid when the radio predicts a sunny day has no impact in your expected gain (because you will guess correctly 50% of the time: $P_{R|\hat{S}} = P_{S|\hat{S}} = 50\%$). So if you choose to keep your money when the forecast is sun, and go all in if it is rain, your expected gain is

$$\begin{aligned} \langle £_Y \rangle &= \sum_{k=0}^{n} P_{\hat{R}}(k)\ 2^k \\ &= \sum_{k=0}^{n} \binom{n}{k} 0.6^k\ 0.4^{n-k}\ 2^k \\ &= 1.6^n, \end{aligned} \tag{S.2}$$

where $P_{\hat{R}}(k)$ is the probability that the radio predicts rain on exactly k of the n days.

Conclusions: If the two of you follow these strategies (your grandfather always going all in, you only when the forecast is rain, and bidding nothing if the radio says sunny), your expected gain is the same. However, his likelihood of losing everything is huge $(1-0.8^n)$, while you will almost always make money (except with probability 0.4^n, when all forecasts are good and you never risk anything). Only in the unlikely event that it rains every day (and there is at least one sunny forecast) will your grandfather finish with more more money than you.

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, $\{P_X^i\}_i$, 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]

[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.

(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?*

Solution.

(a) The capacity of the binary symmetric channel evaluates to

$$\begin{aligned} C &= \max_{P_X} I(X:Y) = \max_{P_X} H(Y) - H(Y|X) \\ &= \max_{P_X} H(Y) + \sum_{x,y} P_X(x) P_{Y|X=x}(y) \log P_{Y|X=x}(y) \\ &= \max_{P_X} H(Y) - \sum_{x} P_X(x)\ H_{\text{bin}}(p) \\ &= \max_{P_X} H(Y) - \ H_{\text{bin}}(p) \\ &= 1 - H_{\text{bin}}(p), \end{aligned} \tag{S.3}$$

where $H_{\text{bin}}(p)$ is the *binary entropy*, i.e. the entropy of the probability distribution $(p, 1-p)$,

$$H_{\text{bin}}(p) = p \log \frac{1}{p} + (1-p) \log \frac{1}{1-p}.$$

To maximise $H(Y)$ we chose the uniform distribution on the input, $P_X^u = (\frac{1}{2}, \frac{1}{2})$, see part (b).

Similarly, for the symmetric erasure channel, we have

$$\begin{aligned} C &= \max_{P_X} H(Y) - H(Y|X) \\ &= \max_{P_X} H(Y) - H_{\text{bin}}(p) \\ &= 2\frac{1-p}{2} \log \frac{2}{1-p} + p \log \frac{1}{p} - H_{\text{bin}}(p) \\ &= 1 - p. \end{aligned} \tag{S.4}$$

(b) We have

$$\begin{aligned} I(X:Y|B) &= H(Y|B) - H(Y|XB) \\ &= H(Y|B) - H(Y|X) \qquad ^{(*)} \\ &\leq H(Y) - H(Y|X) \qquad ^{(**)} \\ &= I(X:Y), \end{aligned} \tag{S.5}$$

where $^{(*)}$ stands because B is just a label that tells us which probability distribution P_X^i we used, so knowing X is as good as knowing X and B, in the sense that $H(Y|XB) = H(Y|X)$, and $^{(**)}$ comes from the data-processing inequality, $H(Y|B) \leq H(Y)$ (which in lay terms says that extra information cannot hurt).

(c) For symmetric channels, the mutual information between input and output is invariant under permutation of the inputs (that's how they are defined). Look for instance at the symmetric erasure channel. The input distribution $P_X^1 = (0.75, 0.25)$ yields the same mutual information as $P_X^2 = (0.25, 0.75)$.

Not knowing which permutation of P_X was used in the input is equivalent to take a uniform mixture over all possible permutations of P_X. Conveniently, such mixture gives us the uniform distribution:

$$P_X = \begin{pmatrix} P_X(x_1) \\ P_X(x_2) \\ \vdots \\ P_X(x_N) \end{pmatrix}, \qquad \left\{P_X^i\right\}_{i=1,\dots,N!} \text{ permutations of } P_X, \tag{S.6}$$

$$\sum_i^{N!} \frac{1}{N!} P_X^i = \frac{1}{N!} \begin{pmatrix} (N-1)!\ P_X(x_1) + (N-1)!\ P_X(x_2) + \dots + (N-1)!\ P_X(x_N) \\ (N-1)!\ P_X(x_1) + (N-1)!\ P_X(x_2) + \dots + (N-1)!\ P_X(x_N) \\ \vdots \\ (N-1)!\ P_X(x_1) + (N-1)!\ P_X(x_2) + \dots + (N-1)!\ P_X(x_N) \end{pmatrix} \tag{S.7}$$

$$= \frac{1}{N} \begin{pmatrix} \sum_i P_X(x_i) \\ \sum_i P_X(x_i) \\ \vdots \\ \sum_i P_X(x_i) \end{pmatrix} = \begin{pmatrix} 1/N \\ 1/N \\ \vdots \\ 1/N \end{pmatrix}. \tag{S.8}$$

This means that for any input distribution P_X the mutual information always increases if instead you use the uniform distribution. Here, $I(X:Y|B)$ is the mutual information knowing you used P_X (take $B = 1$), and $I(X:Y)$ is the mutual information for the uniform distribution. Conclusion: for symmetric channels the mutual information is maximised if one takes the uniform distribution as input.

(d) Using part (b) and (c), we choose the uniform distribution on X and calculate the capacity:

$$\begin{aligned} C &= \max_{P_X} I(X:Y) = I(X:Y)_{P_X^u} = H(Y) - H(Y|X) \\ &= -\sum_y P_Y(y) \log P_Y(y) + \sum_{x,y} P_{XY}(x,y) \log P_{Y|X}(y) \\ &= -\sum_y \left(\sum_x P_X(x) P_{Y|X=x}(y) \right) \log \left(\sum_x P_X(x) P_{Y|X=x}(y) \right) \\ &\quad + \sum_{x,y} P_X(x) P_{Y|X=x}(y) \log P_{Y|X=x}(y) \\ &= -4 \cdot \frac{1}{4} \log \left(\frac{1}{4} \right) + 4 \cdot \frac{1}{4} H_{\text{bin}}(p) = 2 - H_{\text{bin}}(p). \end{aligned} \tag{S.9}$$