2 Classical Information

Exercises

2.1 Consider a sequence of 10 letters of the English alphabet, and a sequence a 26 digits from the set $\{0, 1, ..., 9\}$. Which has the greater information content, if all sequences are equiprobable?

2.2 A TV speaker has a vocabulary of 10,000 words. He utters 1000 words at random (repetitions allowed). Compare the information content of his speech with that of a TV picture of 500×600 pixels with 16 brightness levels each one.

2.3 Decomposability of the entropy. Let $H(p_1, ..., p_n) := \sum_i p_i \log_2 p_i$. Prove that

$$\begin{aligned}
&H(p_1, ..., p_n) = H((p_1 + ...p_r), (p_{r+1} + ... + p_n))+\\
&(p_{1,r} := p_1 + ... + p_r)H(p_1/p_{1,r}, ..., p_r/p_{1,r})+\\
&(p_{r+1,n} := p_{r+1} + ... + p_n)H(p_{r+1}/p_{r+1,n}, ..., p_n/p_{r+1,n}).
\end{aligned}$$

2.4 Entropic quantum uncertainty principle. Let A, B be two quantum observables, with eigenvectors $\{|a\rangle\}, \{|b\rangle\}$. Let $F(A, B) := \sup_{a,b} |\langle b|a\rangle|$ the maximum fidelity between any couple $|a\rangle, |b\rangle$. Let ψ be a state vector. If $p_{\psi,A}, p_{\psi,B}$ denote the distribution probabilities $\{|\langle a|\psi\rangle|^2\}, \{|\langle b|\psi\rangle|^2\}$, then it can be shown that

$$H(p_{\psi,A}) + H(p_{\psi,B}) \geq 2 \log_2 \frac{1}{F(A, B)}.$$

Prove the following weaker result:

$$H(p_{\psi,A}) + H(p_{\psi,B}) \geq 2 \log_2 \frac{2}{1 + F(A, B)}.$$

2.5 Let $f(x)$ be a convex function of a real variable. Prove that Jensen's inequality is equivalent to the assertion that the center of gravity of any collection of particles $\{(x_1, y_1), ..., (x_N, y_N)\}$ on the curve $y = f(x)$ lies above the curve.

2.6 N random cubes have average volume $\langle V \rangle = 1000$. The average of the lengths of their sides is $\langle \ell \rangle = 10$. What can be said of the size of the largest of these N cubes?

2.7 Conditioning reduces entropy. Prove that $H(X|Y,Z) \leq H(X|Y)$.

2.8 The entropy distance between the random variables X and Y is defined as

$$d_H(X,Y) := H(X|Y) + H(Y|X).$$

i/ Prove that d_H satisfies the triangle inequality

$$d_H(X,Z) \leq d_H(X,Y) + d_H(Y,Z).$$

However it is not a distance. ii/ Why?

2.9 Let X be a real-valued random variable. Prove that $H(X^2|X) = 0$, but $H(X|X^2)$ is not necessarily zero.

2.10 Suppose X is the value obtained by throwing a fair die. Let Y be 1 if the value of X is odd and 0 otherwise. Compute $H(X), H(Y), H(X|Y), H(Y|X)$.

2.11 Principle of maximum entropy (maxent). Let X be a random variable with real values $\{x_1, ..., x_N\}$. Let $\mu := \langle X \rangle$ its expectation value. Prove that, given μ, the probability distribution which maximizes the entropy $H(X)$ satisfies

$$p_j := P(X = x_j) = A\mathrm{e}^{\alpha x_j},$$

where A, α are determined by $\langle X \rangle = \mu$, $\sum_i p_i = 1$.

2.12 Chaining rule for conditional entropies. Prove that

$$H(X_1, X_2, ..., X_n|Y) = \sum_i H(X_i|Y, X_1, ..., X_{i-1}).$$

2.13 Prove that the entropy distance satisfies

$$d_H(X,Y) := H(X,Y) - H(X:Y).$$

2.14 Mutual information is not necessarily subadditive: $H(X,Y:Z) \nleq H(X:Z) + H(Y:Z)$. Give an example.

2.15 Chain rule for mutual information. Prove that

$$H(X:Y,Z) = H(X:Y) + H(X:Z|Y),$$

where $H(X:Z|Y) := H(X|Y) + H(Z|Y) - H(X,Z|Y)$.

2.16 Data processing can only destroy information. A Markov chain is a sequence of random variables $X_1 \to X_2 \to ... \to X_n \to ...$ such that X_n is independent of $X_1, ..., X_{n-2}$, given X_{n-1}. I.e.

$$P(X_n = x_n|X_{n-1} = x_{n-1}, ..., X_1 = x_1) = \\ = P(X_n = x_n|X_{n-1} = x_{n-1}).$$

In particular, for such chains

$$P(x_1, ..., x_n) = P(x_n|x_{n-1})...P(x_3|x_2)P(x_2|x_1)P(x_1).$$

The data processing theorem states that if $X \to Y \to Z$ is Markov, then

$$H(X:Z) \leq H(X:Y) \leq H(X).$$

In other words, if X is a random variable subject to noise, producing Y, then data processing (further actions on our part) cannot be used to increase the mutual information between the output of the process and the original information X. Prove it.

2.17 Prove that $X \to Y \to Z \implies Z \to Y \to X$, and hence the data pipelining inequality for any Markov chain $X \to Y \to Z$:

$$H(Z:X) \leq H(Z:Y) \leq H(Y).$$

Any information that Z shares with X is also shared by Z and Y, i.e. the information is pipelined from X to Z through Y.

2.18 Prove that strong additivity $\implies$ subadditivity.

2.19 Consider a real random variable X with the Cauchy probability distribution

$$P(x) = \frac{1}{\pi} \frac{1}{x^2 + 1}.$$

Prove that the probability distribution of the average $Z := (X_1 + ... + X_n)/n$ of n independent copies of X is again

$$P(z) = \frac{1}{\pi} \frac{1}{z^2 + 1}.$$

Comment in relation with the weak law of large numbers.

2.20 In a game over a chessboard, player A has to guess where player B has placed a queen. A is allowed six questions which B has to answer truthfully by a yes/no reply. Find a strategy by which A can always win the game, but winning is not assured if only five questions are permitted. Generalize to an $n \times n$ board.

2.21 The weighing problem and why entropy considerations allow designing optimal strategies. You are given a balance and 9 apparently equal balls. You are told that one ball weights different than the rest and asked to find which ball it is and whether it is heavier or lighter. Devise a strategy with 3 weighings. Generalize to n balls and k weighings, and prove that necessarily $3^k \gtrsim 2n$. (As a matter of fact, a strategy exists which allow to spot the odd ball whenever $3^k \geq 2n + 3$.)

2.22 Suppose you have to find out the value of a random variable X by asking questions to the guy who performs the experiment and can only answer yes/no. Prove that the average number of questions in any successful strategy must be $\geq H(X)$.