CS3236: Homework 2

Pierre Senellart ⟨pierre.senellart@nus.edu.sg⟩

Due Monday, 25 August, 2pm

Submission: Submission can be made in person before the beginning of the lecture, or by upload to the IVLE Workbin. You may submit handwritten answers (possibly scanned), but they must be **clearly readable**. Answers that are not readable may receive only partial marks. Late answers are still accepted on Tuesday (on the IVLE Workbin only), but receive only 50% of the possible marks. Any later submissions receive no marks. Assignments are graded out of 20 points.

1. (3 points) Do exercise 2.30 in the book.

2. (3 points) Do exercise 2.37 in the book.

3. (2 points) Consider an n-bit string $x \in \{0,1\}^n$ in which each bit $x_j \in \{0,1\}$ is independently chosen with probability $p(X_j = 0) = p(X_j = 1) = 1/2$. Compute the Shannon entropy of $H(X)$ in terms of n.

4. (3 points) Consider an n-bit string $x \in \{0,1\}^n$ in which the all zero string is chosen with probability $P_X(x = 00\ldots0) = 1/2$ and with probability $1/2$ we choose one of the remaining strings uniformly at random. Compute the Shannon entropy of $H(X)$ in terms of n.

5. For the following two exercises, you should read 2.7 carefully if you haven't done so. You will show your first result about entropy – namely that forgetting information increases entropy, that is, uncertainty!

   a) (2 points) Show that if f is convex, then $-f$ is concave. (Hint: f is concave on an interval (a,b) if for all $x_1, x_2 \in (a,b)$ and $0 \leq \lambda \leq 1$ we have $f(\lambda x_1 + (1-\lambda)x_2) \geq \lambda f(x_1) + (1-\lambda) f(x_2)$.)

   b) (5 points) Show that the Shannon entropy is strictly concave. That is, if we consider the entropy of a mixture of distributions in which the probability of x is given by $P_X(x) = \lambda Q_X(x) + (1-\lambda) Q'_X(x)$ with $1 \geq \lambda \geq 0$ and Q_X and Q'_X being two *distinct* probability distributions we have

   $$H(P_X) = -\sum_x P_X(x) \log P_X(x) \geq \lambda H(Q_X) + (1-\lambda) H(Q'_X)\,.$$

   with equality only for $\lambda = 0$ and $\lambda = 1$.

c) (2 points) Explain why the fact that the Shannon entropy is concave implies that forgetting information can only increase the entropy.

6. (Bonus) If you are curious, the following two exercises may give you some additional insight into the Shannon entropy.

   a) (Easier) Use Jensen's inequality to prove that $H(X) \leq \log |X|$ where
   $$|X| = |\{x \mid P_X(x) > 0\}| \, .$$

   b) (More tricky) Show that the entropy of any joint probability distribution P_{XY} satisfies $H(X, Y) \leq H(X) + H(Y)$.