

INFORMATION THEORY

Master of Logic, Master AI, Master CS, University of Amsterdam, 2018

TEACHER: Christian Schaffner, TAs: Yfke Dulek, Esteban Landerreche, Kyah Smaal

Team homework set 2

Unless otherwise stated, you should provide exact answers rather than rounded numbers (e.g., $\log 3$ instead of 1.585) for non-programming exercises.

Problem 1: Fun with dice (6 pt)

Consider the following random experiment with two fair (regular six-sided) dice. First, the first die is thrown – let the outcome be A . Then, the second die is thrown until the outcome has the same parity (even, odd) as A . Let this final outcome of the second die be B . The random variables X , Y and Z are defined as follows:

$$X = (A + B) \bmod 2, \quad Y = (A \cdot B) \bmod 2, \quad Z = |A - B|.$$

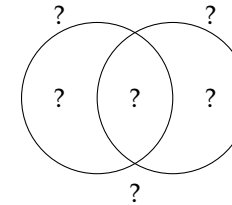
- (a) (1pt) Find the joint distribution P_{AB} .
- (b) (1pt) Determine $H(X)$, $H(Y)$ and $H(Z)$.
- (c) (1pt) Compute $H(Z|A = 1)$.
- (d) (1pt) Compute $H(AB)$, i.e. the joint entropy of A and B .
- (e) (2pt) A random variable M describes whether the sum $A + B$ is strictly larger than seven, between five and seven (both included) or strictly smaller than five. How much entropy is present in M ?

Problem 2: Mutual information (2pt)

Let the binary random variables X and Y be defined by the joint probability distribution $P_{XY}(00) = 3/7$, $P_{XY}(01) = 1/7$, $P_{XY}(10) = 1/7$, $P_{XY}(11) = 2/7$.

Draw the entropy diagram for X and Y , and compute $H(X)$, $H(Y)$, $H(XY)$, $H(X|Y)$, $H(Y|X)$, and $I(X;Y)$. Reading off a value from the

diagram is a valid ‘computation’, as long as you clearly state how you did it.



Problem 3: Entropy of functions of a random variable (3pt)

Let X be a random variable, and let f be a function of X .

- (a) (1pt) Show that $H(f(X)|X) = 0$.
- (b) (2pt) Show that $H(f(X)) \leq H(X)$. *Hint:* use the chain rule.

Problem 4: Relative and cross entropy (8pt)

- (a) (5pt) Prove that for any two distributions P and Q over \mathcal{X} , $D(P||Q) \geq 0$, and that equality holds if and only if $P = Q$.
- (b) (1pt) We have seen that the mutual information can be expressed in terms of the relative entropy, i.e. that $I(X;Y) = D(P_{XY}||P_X \cdot P_Y)$. Use (a) and this fact to show that $H(X|Y) \leq H(X)$.
- (c) (2pt) We have seen a relation between relative entropy and cross entropy in Intro/Team Quiz 02. Use this relation to express the mutual information $I(X;Y)$ in terms of Shannon entropies of X and Y (such as $H(X)$, $H(Y)$, $H(X|Y)$, $H(Y|X)$, $H(YX)$) and of the cross entropy H_c of P_{XY} and $P_X \cdot P_Y$.

Problem 5: Programming (8pt)

Note: You do not have to hand in your code. As before, describe briefly what you did to reach your answer, what choices you made, and how you treated edge cases if those arose.

- (a) (1pt) In the [first programming quiz](#) for this module, you computed the entropy of sampling a single letter from the story of Jack and the beanstalk.

If you instead decided to sample a single *word* from the same story, do you expect the entropy to be higher or lower compared to sampling a single letter?

- (b) (3pt) Verify your answer from (a) by writing a program that computes the entropy explicitly. Use [the same text file](#) as input.

- (c) (1pt) In the [second programming quiz](#) for this module, you computed the entropy of the second letter given the first when sampling a pair of letters from the story of Jack and the beanstalk. Let P_{XY} denote the joint distribution where X is the first and Y is the second letter (again neglecting the beginning and end of the text).

Explain in words why $P_X = P_Y$.

- (d) (2pt) For P_{XY} the bigram distribution from above, write a program to compute the cross entropy $H_C(P_{XY}, P_X \cdot P_Y)$ between the bigram distribution and the independent single-letter distribution.

- (e) (1pt) Express in words what this quantity $H_C(P_{XY}, P_X \cdot P_Y)$ means.