Practice Exercises for Information-Theory 2

Professor Petros Elia, elia@eurecom.fr

February 9th, 2024

1) Let $\mathbb{F}_2^n$ be the set of all binary vectors, and let $\mathcal{A}_\epsilon^n$ be the set of typical sequences based on some non-trivial distribution. Which is true below?
   - a) $Prob(\mathcal{A}_\epsilon^n) \approx 1$ or b) $Prob(\mathcal{A}_\epsilon^n) \ll 1$
   - c) $|\mathcal{A}_\epsilon^n)| \ll 2^n$ or d) $|\mathcal{A}_\epsilon^n)| \approx 1$?

   In the above, choose a) or b) and then choose c) or d). Justify.

2) In a communication setting where X defines the input and Y defines the output, what is the connection between the mutual information $I(X;Y)$ and the channel capacity? Offer some intuition as well as review the proof.

3) Consider a multiple-input multiple-output (MIMO) channel with 2 transmit antennas, 3 receive antennas, and random fading where you can only encode over 1 coherence period. Assume Rayleigh fading where the fading coefficients are iid and distributed as complex-normal random variables with zero mean and unit variance. What is approximately the probability of error if the signal-to-noise ratio is approximately $SNR = 10000$?

4) Can you provide real-life settings (emphasis on the channel behavior) where we must use the metric of ergodic capacity, of ϵ-outage capacity, and of AWGN capacity?

5) Explain the difference between the orthogonal and the non-orthogonal multiple-access settings. What are the practical consequences on the receiver?

6) Consider a multi-antenna communication setting. Let us be able to send 10 different 64-QAM symbols in 4 channel uses. What is the rate of communication in bits per channel use (bpcu)? What would the constellation size be if, under the same conditions, we wished to double the rate.

7) Consider gathering statistics on the frequency with which different 6-tuples of letters appear in the English language. Argue what is (the order of, i.e., approximately) the amount of data that you'd need to gather such statistics that are reliable.

8) Consider the 3-user broadcast channel, with links $|h_1|^2 = 0.7, |h_2|^2 = 0.4, |h_3|^2 = 0.2$, having also a unit power noise at the receivers. Let P_k be the power of the signal meant for user k, and let $P_1 = 10, P_2 = 5$ and $P_3 = 1$. Describe the capacity region, and how this is achieved by describing the method of decoding.

9) What is approximately the doubling rate in a horse race, if your starting bet is 10 euros, you always bet your accumulated funds, and your income in 10 bets is equal to 20000 euros?

10) Derive the Kullback-Leibler distance between two distributions: the first is the fair dice distribution, and the other is the distribution where the equiprobable outcomes $1, 2, 3$ have a double probability than equiprobable outcomes $4, 5, 6$.

11) Consider the coded caching scenario, where a transmitter serves $K = 3$ users via a broadcast shared-link (bottleneck-link), and where the transmitter has access to a library of $N = 6$ files (movies) A, B, C, D, E, F, each of size 1GigaByte (GB). Let each user have a cache of size $M = 2$ GBs. Describe:
    - The placement phase (what data goes into each user's cache)
    - The delivery phase (describe the sequence of XORs sent by the transmitter)
    - What is the total size of all the transmitted XORs together?

12) Consider the setting of horse-races with $m = 4$ horses, where the odds are even. Let the probability that the first horse wins be $p_1 = 1/2$, and the rest equal to $1/6$. Assume you bet 100 euros (starting bet).
    - To be optimal, how much should you bet on the fourth horse (horse number 4)?
    - What is the optimal doubling rate?

13) What is the entropy of a zero-mean Real Gaussian Random Variable with variance 1. How does the entropy change if the variance doubles, and how if the mean shifts to 3?