To be solved in Class

1. **Geometric distribution.** For a $0 \leq p \leq 1$, let us consider a series of independent events that each have success probability $p$. Let $X$ be the number of trials until the first success.

   (a) Show that $P(X(n) = (1 - p)^n p$.

   (b) Give closed formulas for $\sum_{n=0}^{\infty} np^n$ and $\sum_{n=0}^{\infty} n^2 p^n$.

   **Hint:** Recall the formula for a geometric series $\sum_{n=0}^{\infty} p^n = \frac{1}{1-p}$. Differentiate with respect to $p$ on both sides.

   (c) Show that the entropy $H(X)$ is $\frac{h(p)}{p}$.

   (d) Compute $E[X]$.

   (e) Compute $\text{Var}[X]$.

2. Prove Lemma 1 below stating that the capacity per transmission is not increased if we use a discrete memoryless channel many times. For inspiration, look again at the proof of the converse of Shannon’s noisy-channel coding theorem.

**Lemma 1 (Lemma 7.9.2 in [CT])** Let $Y^n$ be the result of passing $X^n$ through a discrete memoryless channel of capacity $C$. Then, $I(X^n; Y^n) \leq nC$ for all $P_{X^n}$.

**Homework**

1. **Additive noise channel.** Find the channel capacity of the following discrete memoryless channel. On input $X$ from $X = \{0, 1\}$, the output $Y$ is obtained by adding (over the reals) another real random variable $Z$, i.e., $Y = X + Z$ with distribution $P_Z(0) = P_Z(a) = \frac{1}{2}$ independent of $X$. Compute the channel capacity for all possible values of $a \in \mathbb{R}$.

2. **Tall, fat people.** Suppose that the average height of people in a room is 1.5m. Suppose that the average weight is 50kg.

   (a) [1 point] Argue that no more than one third of the population is 4.5m tall.

   (b) [2 points] Find an upper bound on the fraction of people who are simultaneously tall (say, at least 3m) and fat (say, at least 150kg).

3. **Another Kind of Entropy.** In this exercise we consider a different entropy notion. Let $X$ and $Y$ be random variables with joint probability distribution $P_{XY}$. The guessing probability and the min-entropy of $X$ are respectively defined as

   $$\text{Guess}(X) := \max_x P_X(x) \quad \text{and} \quad H_{\min}(X) := -\log \text{Guess}(X).$$
The conditional guessing probability and the conditional min-entropy of $X$ are respectively defined as:

$$\text{Guess}(X|Y) := \sum_y P_Y(y) \text{Guess}(X|Y = y)$$

and

$$H_{\text{min}}(X|Y) := -\log \text{Guess}(X|Y).$$

(a) [1 point] If $X$ has no uncertainty (i.e. $H(X) = 0$), what is $H_{\text{min}}(X)$?

(b) [1 point] If $X$ is uniformly distributed over $\mathcal{X}$, what is $H_{\text{min}}(X)$?

(c) [2 points] Prove that $H_{\text{min}}(X|Y) \geq H_{\text{min}}(X)$.

(d) [3 points] Prove that $H_{\text{min}}(X) \geq H_{\text{min}}(X|Y)$.

(e) [3 points] Prove that $H_{\text{min}}(X|Y) \geq H_{\text{min}}(XY) - \log |\mathcal{Y}|$.

4. **Erasures and errors in a binary channel** Consider a channel with binary inputs that has both erasures and errors. Let the probability of error be $\varepsilon$ and the probability of erasure be $\alpha$, so the channel is as described in Figure 1.

(a) [3 points] Find the channel capacity of this channel.

(b) [1 point] Specialize to the case of the binary symmetric channel ($\alpha = 0$).

(c) [1 point] Specialize to the case of the binary erasure channel ($\varepsilon = 0$).

![Figure 1: Erasures and errors in a binary channel.](image)

5. **Encoder and decoder as part of the channel.** Consider a binary symmetric channel (BSC) with crossover probability $\varepsilon = 0.1$. A possible coding scheme for this channel with two codewords of length 3 is to encode message $a_1$ as 000 and $a_2$ as 111. With this coding scheme, we can consider the combination of encoder, channel, and decoder as forming a new BSC, with two inputs $a_1$ and $a_2$ and two outputs $a_1$ and $a_2$.

(a) [3 points] Calculate the crossover probability of this channel.

(b) [2 points] What is the capacity of this channel in bits per transmission of the original channel?

(c) [1 point] What is the capacity of the original BSC with crossover probability $\varepsilon = 0.1$. Compare the two capacities.

(d) [4 points] Prove the following general result: For any channel, considering the encoder, channel, and decoder together as a new channel from messages to estimated messages will not increase the capacity in bits per transmission of the original channel.

**Hint:** Use Lemma 1 above.