Information Theory, Coding and Cryptography
Homework Set 3

Please send your answers – either typeset on a computer, or handwritten (but readable) and scanned, preferably in PDF format – to `bouman@cwi.nl` *before Wednesday March 30th, 23h59. Do not forget to put your name on the first page. Have fun!*

1 Shannon's Noisy-Channel Coding Theorem

Given is a binary symmetric channel (BSC) with crossover probability 0.1. For each (rate, bit-error probability)-pair below, argue (or compute) whether it is achievable using channel coding. (The unit used to specify the rates is bits per channel use.)

a) [1 point] $(0.6, 10^{-3})$

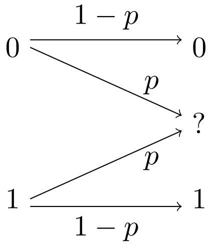
b) [1 point] $(0.4, 0.04)$

c) [1 point] $(0.5, 10^{-17})$

d) [1 point] $(0.7, 0.08)$

2 The Erasure and Z Channel

A simple but important channel model is the *erasure channel*. It is usually depicted as



In this channel, the receiver receives "?" in case a bit is erased, which happens with probability p. Note that there is a form of symmetry; the erasure probability is taken to be the same for zeros and ones.

a) [1 point] For what value of p will it be impossible to send information over the erasure channel? Why?

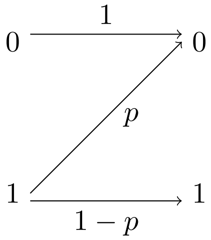
b) [3 points] Derive the capacity of the erasure channel.

The erasure channel is not just a toy model; you'll encounter it for example in the UDP protocol, which is used to send packets over the Internet, where a packet is just a collection of many bits. The UDP protocol computes a checksum of these bits and sends it along with the packet. If, at the receiver side, the bits do not match with the checksum, all bits are declared as "?"-symbols, and the packet is said to be "lost".

Suppose that you use Skype (which uses UDP) and the probability of packet loss is p. The more packet loss, the lower the sound quality. To improve the sound quality, we will compute from every two packets X and Y, where $X = (X_1, X_2, \ldots, X_N)$ for X_i a bit, and Y is defined similarly, a *parity packet*, Z, with $Z_i = X_i \oplus Y_i$ and send it along with the X and Y packets. The receiver can then recover some erased packets using the parity packets.

c) [2 points] Compute the block erasure probability (when at least one packet of the block has been erased, we say the entire block has been erased) for blocks of two packets of the uncoded system, and the block erasure probability of the system with the parity coding.

Yet another basic channel model is the Z-channel, where the "Z" originates from the Z-shaped graph structure of the channel:



The canonical example of an application where this channel occurs naturally is an optical link. A light source at the transmitter side can be switched on ("1") or off ("0"). Along the way, the light signal may be attenuated, such that it is possible that a zero is received while a one was transmitted. It is however not possible that light is received while the light source is switched off. Note that here there is an intrinsic asymmetry between zeros and ones (except in the trivial case where $p = 0$).

d) [6 points] Compute the capacity of the Z-channel for $p = 0.1$. **Hint:** Consider a binary random variable X as input to the Z-channel, with distribution $(q, 1-q)$, and optimize the mutual information (numerically) over q.

3 Lossy Compression

Let X be a uniformly random n-bit string, hence $H(X) = n$. We want to compress X to $Y \in \{0,1\}^\ell$, where $\ell < n$. Hence we cannot compress X losslessly; a positive error probability will be inevitable. In this problem, we will first learn about the best error probability that one can hope for (given some ℓ), then we will analyze the performance (in terms of error probability) of two compression strategies.

a) [5 points] Give a lower bound on $H(X|Y)$ and use the corollary to Fano's inequality in the lecture notes to get a lower bound on the bit-error probability $\bar{p}_e$, where $\bar{p}_e$ is defined as in the lecture notes. **Hint:** Write the inverse of the binary entropy function as $h^{-1}(\cdot)$.

b) [3 points] Suppose that we compress by just discarding the last $n - \ell$ bits of X. Assume that the decompressor will simply guess the missing bits uniformly at random. Compute the average probability of bit error $\bar{p}_e$ obtained using this strategy.

Now consider the following compression strategy. We take a (n, ℓ) channel code $\mathcal{C}$ (i.e. the encoder maps ℓ-bit words to n-bit codewords) for the binary symmetric channel having rate $R = \ell/n$. But now, we will use this channel code in "in reverse", in that we pass X through the *decoder* of this code. By doing so, the output will have length ℓ and we obtain a compressor with rate $1/R$.

c) [3 points] Assume that $\mathcal{C}$ is ε-capacity-achieving for some small positive ε, meaning that $R + \varepsilon$ equals the capacity of some BSC with crossover probability q. Find an expression for q in terms of n, ℓ and ε.

The decoder of $\mathcal{C}$ "expects" that the bits entering the decoder are coming out of a BSC with parameter q, hence it tries to correct roughly a q-fraction of the bits. In our setting, the codeword found by the decoder will differ from X on average in a q-fraction of the bits. Hence, the q that you have (hopefully) found in (b) is exactly the average bit error probability that we are looking for.

d) [3 points] Plot the bit error probabilities found in (a), (b) and (c) (on the y-axis) as a function of the rate R (on the x-axis) with range $R \in [1/3, 1]$, by taking $\varepsilon = 0.01$.

4 Typicality

Consider a random source S that emits independent and identically distributed binary symbols X_i. (We view these output symbols as random variables.) Let $p := \Pr[X_i = 1]$.

a) [1 point] Suppose the source emits N symbols, and let $\boldsymbol{X} := (X_1, \ldots, X_N)$ be the bit string composed of these symbols. What is the expected Hamming weight of $\boldsymbol{X}$?

The set of N-bit strings whose Hamming weight is βN-close to the expected Hamming weight of an N-bit string coming from S is called the *typical set* $\mathcal{T}_\beta^N(S)$:

$$\mathcal{T}_\beta^N(S) := \{\mathbf{b} \in \{0,1\}^N : |\mathrm{wt}(\mathbf{b})/N - p| \le \beta\}$$

b) [3 points] Give a (non-trivial) upper bound for the probability that $\boldsymbol{X}$ is *atypical* (i.e. not in the typical set defined above). **Hint:** use that

$$\Pr[\,|\tfrac{1}{N}\sum X_i - \mu| > \delta] \le 2\Pr[\tfrac{1}{N}\sum X_i - \mu > \delta].$$

c) [2 points] Compute or upper bound the size of the typical set $\mathcal{T}_\beta^N(S)$. **Hint:** have a look at Lemma 4 in the notes.

5 Privacy Amplification

Fix a finite field $\mathbb{F}_q$. Consider the family of functions

$$\mathcal{F} := \{f : \mathbb{F}_q^{d+1} \to \mathbb{F}_q\},$$

with functions f defined by

$$f : x \mapsto x_0 + \sum_{i=1}^{d} x_i r_i,$$

where $x = (x_0, \dots, x_d) \in \mathbb{F}_q^{d+1}$ and where each $r_i \in \mathbb{F}_q$.

a) [1 point] What is the cardinality of $\mathcal{F}$?

Let F be random over $\mathcal{F}$, i.e. obtained by replacing the r_i in the definition of f by the random variables R_i, which are uniformly distributed over $\mathbb{F}_q$.

b) [4 points] Prove that the family $\mathcal{F}$ is universal, i.e. prove that for any $x, x' \in \mathbb{F}_q^{d+1}$ such that $x \neq x'$, it holds that

$$P[F(x) = F(x')] \leq 1/q$$

The conditional guessing probability and the conditional min-entropy are respectively defined as:

$$\mathrm{Guess}(X|Y) := \sum_y P_Y(y) \max_x P_{X|Y=y}(x)$$

and

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

c) [4 points] Prove that $H_{\min}(X|Y) \leq H_2(X|Y) \leq H(X|Y)$.

Let X be a uniformly distributed n-bit string, held by Alice, and she wants to derive a cryptographic key from it. However, Eve holds Y, which is obtained by passing X though a BSC with crossover probability 0.15.

d) [3 points] Compute $H_2(X|Y)$.

Alice applies privacy amplification with the help of the family $\mathcal{F}$, resulting in a shorter, but almost uniform key S. Let F be the function that Alice selected at random from $\mathcal{F}$.

e) [2 points] Find the maximum length of the extracted key, such that it has statistical security (i.e. statistical distance from being uniform) of $< 10^{-6}$.