MDL exercises #1

1. The Cost of Rounding, and how to Avoid it.

   Let P be a distribution on a set $\mathcal{X}$. We say that a prefix code C *corresponds* to P if for all $x \in \mathcal{X}$, we have $L_C(x) = \lceil -\log P(x) \rceil$.
   Now consider the uniform distribution P on the set $\{a, b, c\}$.

   a) Show that there exists a prefix code C that corresponds to P.

   b) Now consider strings from $\{a, b, c\}^{100}$. An easy way to code such strings is by using C a hundred times in a row. The result of this procedure is a new prefix code which we will call C'. Similarly, P can be extended to a distribution on 100 outcomes by defining $P'(x^n) = \prod_{i=1}^{100} P(x_i)$.
   Show that we no longer have for all strings z of length 100 that $L_{C'}(z) = \lceil -\log P'(z) \rceil$.

   c) Does this mean that there is no prefix code that corresponds to P'? If you think that there is one, then describe this code. And/or does this mean that there is no distribution that corresponds to C'? If you think that there is one, then describe this distribution.

2. No hypercompression.

   Consider a distribution P on a set of messages $\mathcal{X}$. Let $C : \mathcal{X} \to \{0, 1\}^*$ be a prefix code on $\mathcal{X}$ (i.e., $\mathcal{Y} = \{C(x) : x \in \mathcal{X}\}$ is a prefix free set). As in the Kolmogorov Complexity book, $l(\cdot)$ denotes the length of a binary sequence and $d(\cdot)$ denotes the size of a set. For convenience, we define $L_C(x) := l(C(x))$.

   a) Given that P is uniform, show that for *any* code C, the probability that we are able to compress an outcome by more than k bits is less than 2^{-k}. That is, we have that $P(L_C(X) < \log d(\mathcal{X}) - k) \leq 2^{-k}$.

   b) The no hypercompression inequality is a generalisation of the previous result to arbitrary P: it states that $P(L_C(X) \leq -\log P(X) - k) \leq 2^{-k}$. Prove this using Markov's inequality, which states that for nonnegative random variables X we have $P(X \geq a) \leq E[X]/a$. Hint: consider the ratio between the probability of x under P (where $x \in \mathcal{X}$) and under the distribution that corresponds to L_C.

3. Maximum likelihood.

   a) The Bernoulli probability of a sequence with n_0 zeroes and n_1 ones is $\theta^{n_1}(1-\theta)^{n_0}$. Compute the maximum likelihood estimator for the parameter, that is the value of θ that maximizes this probability.

   b) Compute the maximum likelihood estimator $\hat{\theta} = (p_{[1|0]}, p_{[1|1]})$ for a binary first order Markov chain.

   c) The numbers $x_1, \ldots, x_n$ are sampled from an exponential distribution, which has density function $f(x) = \lambda e^{-\lambda x}$. Compute the maximum likelihood value for λ.

   d) Suppose that we model data with a uniform distribution on the real numbers between a and b. Given outcomes $x_1, \ldots, x_n$, compute the maximum likelihood values for a and b.

4. Draw X_1, X_2, X_3 from an order 1 Markov chain. Is X_3 dependent on X_1? What if you know the value of X_2? Base your answer on the definition of independence: X_3 is independent of X_1 iff for all values x_1 and x_3 that the corresponding random variables can take we have that $P(x_3 \mid x_1) = P(x_3)$.

5. The ELISA test for AIDS is used in America to screen blood donations. If a person actually carries HIV, experts estimate that the test gives a positive result 97.7% of the time. If a person does not carry HIV, ELISA gives a negative result 92.6% of the time. Estimates are that 0.5% of the American public carry HIV, 77% of which are male. Evelyn Average has just tested positive on ELISA and is scared out of her wits. What is the probability that she is infected? Hint: do this by relating the quantity of interest, $P(D \mid E)$, to the available knowledge, $P(E \mid D)$ and $P(E \mid D^c)$, where D is the event that she is infected with the disease, E is the event that she tests positive on ELISA, and D^c is the complement of event D.