1. (2 points) Consider alphabet $X = \{a, b, c, !\}$ with probabilities $p(a) = 0.05$, $p(b) = 0.5$, $p(c) = 0.35$ and $p(!) = 0.1$.

Recall that arithmetic coding can be seen as the application of Shannon-Fano-Elias coding to blocks of $b$ source symbols (where $b$ need not even be given in advance). Hence, the interval $I(x) \subset [0, 1]$ corresponding to message $x$ (a source string) is given by $[F(x), F(x + 1))$, where $x + 1$ denotes the length $b$ message following the actual message in alphabetical order. So for example, $cab + 1 = caca$.

Note that in the above, the cdf $F(x)$ is defined as the sum of the probabilities of length $b$ messages that precede $x$ in the alphabetical order, so $F(cab!) = p(aaaa) + p(aaab) + \ldots + p(aaa!) + p(baaa) + \ldots + p(cabc) + p(cab!)$.

(a) Find out the interval $I(cab!) \subset [0, 1]$ that is used for encoding the message $cab!$ in arithmetic coding.

(b) (1 point) Now consider picking a number from $I(cab!)$ as a codeword for $cab!$. For ease of calculations, we consider codewords that are decimal numbers, not binary (i.e., the encoding alphabet is $\{0, \ldots, 9\}$ instead of $\{0, 1\}$).

i. What is the shortest codeword (decimal number with the least number of decimals) you can find within the interval?

ii. What is the shortest codeword $C = 0.d_1 \ldots d_k$ such that also all its continuations (numbers of the form $0.d_1 \ldots d_k d_{k+1} \ldots d_m$ where $m > k$ and $d_i$ can be arbitrary for $i = k + 1, \ldots, m$) are also within the interval? (This is the property we need for a prefix code.)

(c) (1 point) Same as above, but use binary encoding. To make the calculations less cluttered, use the probabilities $p(a) = 2/32$, $p(b) = 16/32$, $p(c) = 11/32$ and $p(!) = 3/32$.

2. Let the model class be given by Bernoulli distributions where each bit in a sequence, $D = x_1, \ldots, x_n$, is independent with probability $Pr(x_i = 1) = \theta$ for all $1 \leq i \leq n$. The probability of data $D$ then becomes

$$p_\theta(D) = \theta^k (1 - \theta)^{n-k},$$

where $k$ is the number of 1s in $D$.

Evaluate the two-part code-length of sequence 0011 when the parameter values are quantized so that $\theta \in \Theta = \{0.25, 0.5, 0.75\}$, and the parameter is encoded using a code with code-lengths $\ell(0.25) = 2$, $\ell(0.75) = 2$ and $\ell(0.5) = 1$. 

Continues on the next page!
3. Again consider Bernoulli distributions. Evaluate the mixture code-length of sequence 0011 when all parameter values are allowed, so $\theta \in \Theta = [0, 1]$ and the parameter prior (pdf) is uniform, $w(\theta) = 1$. Note that for continuous parameter values, the mixture code-length becomes an integral

$$p^w(D) = \int_{\Theta} p_\theta(D) w(\theta) d\theta.$$ 

You will probably find the “beta-binomial” distribution useful; see, e.g., Wikipedia. Note that Beta(1, 1) is the uniform distribution. (Just ignore the combinatorial $\binom{n}{k}$ term which is the only difference between the binomial and Bernoulli distributions for sequences of length $n$.)

4. Once again Bernoulli. This time it’s NML.

(a) Compute the NML normalizing term $C$ under the full Bernoulli model class, i.e., when $\theta \in \Theta = [0, 1]$. Note that the sum can be computer either by enumerating all the $2^n$ sequences or by combining like terms which leaves $n + 1 = 5$ distinct terms. For large $n$ this makes a big difference.

(b) Evaluate the NML code-length for the sequence $D = 0011$. 
