In-class quiz

Problem Q-1 (20 points)

(a) *Give an example of a Huffman code in which l_i for some code word i is one but $\log(1/p_i)$ is arbitrarily large.*

[6 points] The simplest example is for $M = 2$. Then both code words have length 1, but the probability of symbol i can be arbitrarily close to 0, making $\log(1/p_i)$ arbitrarily large.

(b) *Find an example of a Huffman code with 7 code words in which one code word has length 6 but $\lceil \log(1/p_i) \rceil = 4$.*

[8 points] Since Huffman codes are full, the only possibility is to have two code words of length 6, and one code word of each shorter length down to 1 (see Figure 1).

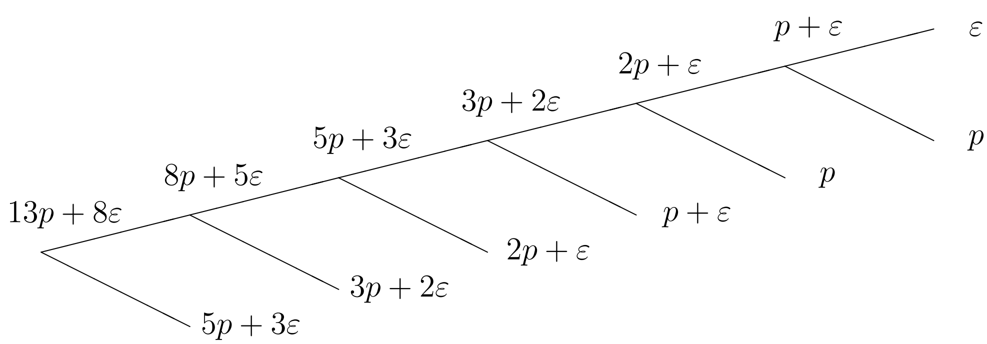

Figure 1: Code tree to maximize p.

Our objective here is to maximize the probability of one of the code words of length 6 while maintaining the constraints of a Huffman code. Let p be the probability of one code word of length 6, and let ε (arbitrarily small) be the probability of the other. We choose the probability of each of the other nodes to be as small as possible, while still allowing the Huffman algorithm to choose the above code.

The figure shows the minimum probability that can be assigned to each leaf node, and results in $1 = 13p + 8\varepsilon$. Since ε can be chosen arbitrarily small, we see that we can choose p to have any desired value greater than 1/13. Thus we can choose $p > 2^{-4}$, which leads to $\lceil \log(1/p_i) \rceil = 4$.

An alternative approach is to set $p = 1/16$ at the outset and then either choose ε as above, or choose any ε small enough and then choose p_1 to make $\sum_i p_i = 1$.

(c) *Explain how this can be generalized to Huffman codes in which $l_i - \lceil \log(1/p_i) \rceil$ is arbitrarily large for an least one code word.*

[6 points] The coefficients of p, moving from right to left on the upper nodes, are $1, 1, 2, 3, 5, 8, 13$. These are the terms of the Fibonnaci series, which increases geometrically as $(1 + \sqrt{5})/2$. More precisely, the nth term of the series is

$$\frac{1}{\sqrt{5}} \left\{ \left[\frac{1+\sqrt{5}}{2} \right]^n - \left[\frac{1-\sqrt{5}}{2} \right]^n \right\}$$

If we extend the argument in part (b) to a tree of length $n-1$ with n code words, we get $-\log p \approx n \log \frac{(1+\sqrt{5})}{2} - \log \sqrt{5}$. This is increasing linearly with n but at a smaller slope than 1. Since the length of the code word is increasing with n with slope 1, the difference between the length and the log pmf is growing without bound[1].

Problem Q-2 (35 points)

(a) *Show that a Huffman code can be rearranged, with no loss in expected length, into a code for which the "binary decimal" numbers associated with the code words are increasing with decreasing code word probabilities.*

[5 points] Any set of code word length that satisfies Kraft can be arranged in binary decimal order. This was shown in lecture 2 when we proved that any set of lengths that satisfied the Kraft inequality could be turned into a prefix-free code with those lengths.

It is not in general possible to assign 0's and 1's in the Huffman construction so as to achieve this ordering property. As an example look at the Huffman code for the symbol probabilities $\{.15, .15, .2, .2, .3\}$.

(b) *Give an intuitive explanation for why the most probable code words,* i.e., *those with* $\left\{ p_{\mathbf{X}^n}(\mathbf{x}^n) \geq 2^{-n(H(X)-\varepsilon)} \right\}$ *are not viewed as typical.*

[6 points] There are very few of the large probability words, so even though they have large probability individually, their aggregate probability is very small.

(c) *Assume that there are both intermediate nodes and leaf nodes at some given length l. Prove that each code word of length l has a probability $p \geq q_l/2$ where q_l is the maximum of the probabilities of the intermediate nodes of length l.*

[6 points] For each intermediate node (and in particular the most probable one), both of the immediate descendants of that node have probabilities, say q' and q'' satisfying $q' \leq p$ and $q'' \leq p$, since if $q' > p$ (or $q'' > p$), the node of probability p could be interchanged with the subtree stemming from q' (or q'') with a reduction in the average code word

[1] You were not expected to go through this entire analytic argument.

length. Thus $q_l = q' + q'' \leq 2p$. Thus, $p \geq q_l/2$.

(d) *Let m be the shortest length for which leaf nodes exist (you may assume that all such leaf nodes correspond to atypical n-tuples). Let M_m be the number of leaf nodes of length m. Let $\delta_m \leq \delta$ be the sum of the probabilities of these atypical leaf nodes. Find a lower bound to δ_m in terms of q_m (the maximum of the probabilities of the intermediate nodes of length m) and M_m. Hint: Use part (c).*

[6 points] Let $p_i;\ 1 \leq i \leq M_m$ be the probabilities of the leaf nodes of length m. Then $\sum_{i=1}^{M_m} p_i = \delta_m$. From part (c), $p_i \geq q_m/2$ for each $i;\ 1 \leq i \leq M_m$. Thus

$$\delta_m = \sum_{i=1}^{M_m} p_i \geq \sum_{i=1}^{M_m} \frac{q_m}{2} = \frac{M_m q_m}{2}$$

(e) *Find a lower bound to q_m in terms of δ_m and M_m. Hint: The sum of the probabilities of the intermediate nodes plus leaf nodes at length m must be one.*

[6 points] Let $q_m(i)$ be the probability of the ith intermediate node at length m. Thus $q_m = \max_i\{q_m(i)\}$. The number of intermediate nodes of length m is $2^m - M_m$, so

$$q_m(2^m - M_m) \geq \sum_{i=1}^{2^m - M_m} q_M(i) = 1 - \delta_m$$

$$q_m \geq \frac{1 - \delta_m}{(2^m - M_m)}$$

(f) *Let $\beta_m = M_m/2^m$ be the fraction of nodes at length m that are leaf nodes. Show that*

$$\frac{\beta_m}{1 - \beta_m} \leq \frac{2\delta_m}{1 - \delta_m}$$

[6 points] Combining parts (d) and (e), we have

$$\delta_m \geq \frac{M_m q_m}{2} \geq \frac{M_m(1 - \delta_m)}{2(2^m - M_m)}$$

With the substitution $\beta_m = M_m/2^m$, this becomes

$$\delta_m \geq \frac{\beta_m q_m}{2} \geq \frac{\beta_m(1 - \delta_m)}{2(1 - \beta_m)}$$

which, on rearrangement, is what is to be proven.

Problem Q-3 (30 points)

(a) *Let U be a source output. Find the probability that the distance from U to V_1 exceeds some given number r. Ignore edge effects throughout,* i.e., *assume that the sample value of U is more than r away from the boundary of the region A.*

[5 points] For any given sample value u of U, the distance from u to the sample value v_1 of V_1 is less than or equal to r if v_1 lies within a circle of radius r (*i.e.*, of area πr^2) around u. Since V_1 is uniformly selected over the area A, the probability that V_1 lies within this area is $\pi r^2/A$ (we are using the symbol A both for the area of the region and the region itself). Thus, for each u, $\Pr(\|V_1 - u\| > r) = 1 - \pi r^2/A$.

Many of you tried to approach this problem componentwise (i.e. along each dimension). Since A is unspecified and we are dealing with circular regions around u, that approach doesn't quite work.

Note that it is important to distinguish between,

$$\Pr(\|V_1 - u\| > r) \quad \text{and} \quad \Pr(\|V_1 - U\| > r)$$

Most of you missed this distinction and derived the former and then equated it to the latter.

(b) *Find the probability that the distance from U to each of the $\{V_j\}$ (*i.e., *to the closest of the $\{V_j\}$) exceeds r.*

[5 points] Since $V_1, \ldots, V_M$ are independent, the events $\|V_j - u\| > r$ are independent events for any given u. Thus, for any given u,

$$\Pr(\|V_j - u\| > r \text{ for all } j; 1 \le j \le M) \;=\; (1 - \pi r^2/A)^M$$

Since this is the same for all sample values u of U,

$$\Pr(\|V_j - U\| > r \text{ for all } j; 1 \le j \le M) \;=\; (1 - \pi r^2/A)^M$$

(c) *Assume that r^2/A is extremely small and approximate the probability in (b) as $e^{-Mg(r)}$ for the appropriate function $g(r)$.*

[5 points] For any $\varepsilon > 0$, $(1-\varepsilon)^M = e^{M\ln(1-\varepsilon)}$. For ε very small, $\ln(1-\varepsilon) \approx \varepsilon$, so $(1-\varepsilon)^M \approx e^{-M\varepsilon}$. Using $\pi r^2/A$ for ε, this becomes

$$(1 - \pi r^2/A)^M \approx e^{-M\pi r^2/A}$$

(d) *Let R be the error when the source output is represented as the closest quantization point. Express the distribution function of the random variable R in terms of your answer to c.*

[7 points] The distribution function of R is $F_R(r) = \Pr(R \le r) = 1 - \Pr(R > r)$. However, $R > r$ means that $\|V_j - U\| > r$ for all $j; 1 \le j \le M$, which is the quantity found in part (c). Thus

$$F_R(r) = 1 - e^{-M\pi r^2/A}$$

(e) *Find the mean square error. The mean square error here is averaged over both the source output and the random choice of quantizer points. Compare your result with that of a quantizer using a square of quantization regions.*

[6 points] From part (d), the probability density for R is

$$f_R(r) = \frac{2\pi M r}{A}\, exp\left\{\frac{-\pi M r^2}{A}\right\}$$

The mean square error *per dimension* is then $(1/2)\int r^2 f_R(r)dr$. If we substitute y for r^2, this simplifies to

$$\text{MSE} = \frac{1}{2}\int_{y=0}^{\infty} \frac{\pi M y}{A} \exp\left\{\frac{-\pi M y}{A}\right\} dy = \frac{A}{2\pi M}$$

[2 points] For the 2D quantizer using square regions, each of area A/M, the MSE per dimension is $(1/12)(A/M)$. Thus, the random choice of quantization points in 2D is not as good as the much more straight forward uniform scalar quantizer.

Some students attempted to find the distribution of R^2 and then find its mean. While this is not necessary here, it is useful to know. See the solutions to HW 5.2(e) to see how to do this.

Several students were confused regarding the limits of integration in this part (and the range of r in defining the distribution in part (d)). If we derived a precise distribution for r, then r clearly cannot exceed the diameter of region A (the diameter is simply the largest distance between any two points in A). However, in part (d), we *approximated* the distribution function by $1 - e^{-M\pi r^2/A}$. For this to be a valid distribution, r *must* range from 0 to infinity. This is why the limits of integration are 0 to infinity.

The other way of looking at this is that under the assumption $r^2 << A$, the diameter of A is much larger than r. This, combined with the fact that the distribution is falling exponentially in r is why integrating from 0 to infinity is justified.

Some of you forgot to normalize both the MSEs per dimension leading to an incorrect comparison of the random and square case.

Problem Q-4 (15 points)

(a) *Express the coefficients* $\{u_k\}$ *as inner products involving* $u(t)$, $\{\boldsymbol{\theta}_k\}$, *and* $\{A_k\}$.

[7 points] As done in the notes several times, we have

$$\begin{aligned}\int_{-\infty}^{\infty} u(t)\theta_j^*(t)\,dt &= \int_{-\infty}^{\infty}\sum_k u_k\theta_k(t)\theta_j^*(t)\,dt \\ &= \sum_k u_k \int_{-\infty}^{\infty}\theta_k(t)\theta_j^*(t)\;dt = u_j A_j\end{aligned}$$

Thus,

$$u_k = \frac{1}{A_k}\langle \boldsymbol{u}, \boldsymbol{\theta}_k\rangle$$

(b) *Find the energy* $\|\mathbf{u}\|^2 = \int_{-\infty}^{\infty} |u(t)|^2 dt$ *in the simplest form you can in terms of* $\{u_k\}$, $\{\theta_k\}$ *and* $\{A_k\}$.

[8 points] Again, as before,

$$\begin{aligned}\int_{-\infty}^{\infty} u(t)u^*(t)\,dt &= \int_{-\infty}^{\infty} u(t)\sum_k u_k^*\theta_k^*(t)\,dt \\ &= \sum_k u_k^* \int_{-\infty}^{\infty} u(t)\theta_k^*(t)\ dt = \sum_k |u_k|^2 A_k\end{aligned}$$

Common errors in this part had to do with scaling. People either forgot to scale or assumed that all of the A_js were identical.