Intro to Information Theory

Bonus Problems

Math Circle

February 25, 2018

1. Let X and Y be independent, Bernoulli($\frac{1}{4}$) random variables. That is, X and Y both take values on $\{0,1\}$ and $\Pr(X=1)=\Pr(Y=1)=\frac{1}{4}$. Let

$$Z = X + Y \mod 2.$$

   (a) What are $H(X)$, $H(Y)$ and $H(Z)$ in binary bits? Use $\log_2(3) = 1.58$ and $\log_2(5) = 2.32$.

   (b) Instead of X and Y being Bernoulli($\frac{1}{4}$), let X and Y be independent, Bernoulli(p) with $0 < p < \frac{1}{2}$. Again, let $Z = X + Y \mod 2$. Show that $H(Z) > H(X)$.

(c) What are $H(\{X, Y\})$, $H(\{X, Z\})$ and $H(\{Y, Z\})$ in terms of $H(X)$, $H(Y)$ and $H(Z)$? Recall that the entropy of a set of random variables is found by summing over all the possible outcomes that the set of random variables can take. See Problems 3c and 3d from the first packet for other examples with entropy of a set of random variables.

(d) Recall from the previous packet that $H(f(X)) \leq H(X)$ for any deterministic function $f(\cdot)$ and any discrete random variable X. Show that if f is invertible, then $H(f(X)) = H(X)$. Hint: Let g be the inverse of f (that is, $g(f(x)) = x$) and use the above property twice.

(e) How can you use the property from Problem 1d plus Property 3(g)ii from the previous packet to find the entropies in Problem 1c?

2. A **geometric** random variable describes the number of independent Bernoulli trials until a success occurs. The distribution is given by

$$\Pr(X = k) = (1 - p)^{(k-1)} p,\ k \geq 1.$$

The expected value of X is

$$E[X] = \sum_{k \geq 1} k \Pr(X = k) = \frac{1}{p}.$$

(a) What is the entropy of the geometric random variable if $p = \frac{1}{2}$ (the trials are equally likely to succeed or fail)?

(b) Consider the following code for a geometric random variable. Let $C(k)$ have $k - 1$ zeros followed by 1 one. The codewords for 1, 2, 3, 4, 5 ... are 1, 01, 001, 0001, 00001 ....

i. Is this code prefix-free?

ii. What is the average codeword length of this code for a general p?

iii. How does the average codeword length compare to $H(X)$ in general (see 4(g)i from the first packet)?

iv. How does the average codeword length compare to $H(X)$ for the code above and a geometric random variable with $p = \frac{1}{2}$?

(c) Consider a geometric random variable with $p = 1 - \frac{\sqrt{2}}{2}$.

i. What is the probability that X is 1 **or** 2? That is, find $\Pr(X = 1) + \Pr(X = 2)$.

ii. What is the probability that X is $2m-1$ **or** $2m$? That is, find $\Pr(X = 2m-1) + \Pr(X = 2m)$.

iii. Extra credit: Verify that $H(X) \approx 2.98$. To do this by hand, you only need to know that $p \approx 0.29$ and $\log_2(p) \approx -1.77$ (you should be able to find the exact solution for $\log_2(1-p)$).

(d) Consider the following code for a geometric random variable. First, for a given k, write k as $2m-1$ or $2m$. The codeword is then the codeword for m from Problem 2b plus a 1 if k is odd or a 0 if k is even. The codewords for 1, 2, 3, 4, 5 ... are 11, 10, 011, 010, 0011, ...

i. Draw a tree that includes at least the first 8 codewords. Verify that the code is prefix-free.

ii. For a geometric random variable with $p = 1 - \frac{\sqrt{2}}{2}$, what is the average codelength for this code? How does this compare to $H(X)$? Hint: Use the results from Problem 2c.