UC Berkeley
Department of Electrical Engineering and Computer Sciences

EE126: Probability and Random Processes

Problem Set 2
Fall 2018

Issued: Wednesday, August 29, 2018   **Due:** Wednesday, September 5, 2018

Problem 1. **Packet Routing**
Consider a system with n inputs and n outputs. At each input, a packet appears independently with probability p. If a packet appears, it is destined for one of the n outputs uniformly randomly, independently of the other packets.

1. Let X denote the number of packets destined for the first output. What is the distribution of X?

2. What is the probability of a collision, that is, more than one packet heading to the same output?

Problem 2. **Compact Arrays**
Consider an array of n entries, where n is a positive integer. Each entry is chosen uniformly randomly from $\{0, \ldots, 9\}$. We want to make the array more compact, by putting all of the non-zero entries together at the front of the array. As an example, suppose we have the array

$$[6, 4, 0, 0, 5, 3, 0, 5, 1, 3].$$

After making the array compact, it now looks like

$$[6, 4, 5, 3, 5, 1, 3, 0, 0, 0].$$

Let i be a fixed positive integer in $\{1, \ldots, n\}$. Suppose that the ith entry of the array is non-zero (assume that the array is indexed starting from 1). Let X be a random variable which is equal to the index that the ith entry has been moved after making the array compact. Calculate $\mathbb{E}[X]$ and $\mathrm{Var}(X)$.

Problem 3. **Message Segmentation**
The number of bytes N in a message has a geometric distribution with parameter p. Suppose that the message is segmented into packets, with each packet containing m bytes if possible, and any remaining bytes being put in the last packet. Let Q denote the number of full packets in the message, and let R denote the number of bytes left over.

1. Find the joint PMF of Q and R. Pay attention on the support of the joint PMF.
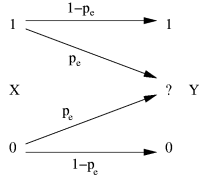
2. Find the marginal PMFs of Q and R.

Figure 1: The channel model for the BEC showing a mapping from channel input X to channel output Y. The probability of erasure is p_e.

3. Repeat part (b), given that we know that $N > m$.

Note: you can use the formulas

$$\sum_{k=0}^{n} a^k = \frac{1 - a^{n+1}}{1 - a}, \text{ for } a \neq 1$$

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1 - x}, \text{ for } |x| < 1$$

in order to simplify your answer.

Problem 4. **Introduction to Information Theory**

Define the *entropy* of a discrete random variable X to be

$$H(X) \overset{\Delta}{=} - \sum_x p(x) \log p(x) = -\mathbb{E}[\log p(X)],$$

where $p(\cdot)$ is the PMF of X. Here, the logarithm is taken with base 2, and entropy is measured in bits.

1. Prove that $H(X) \geq 0$.

2. Entropy is often described as the average information content of a random variable. If $H(X) = 0$, then no new information is given by observing X. On the other hand, if $H(X) = m$, then observing the value of X gives you m bits of information on average.

   Let X be a Bernoulli random variable with $\Pr(X = 1) = p$. Would you expect $H(X)$ to be greater when $p = 1/2$ or when $p = 1/3$? Calculate $H(X)$ in both of these cases and verify your answer.

3. We now consider a **binary erasure channel** (BEC).

   The input X is a Bernoulli random variable with $\Pr(X = 0) = \Pr(X = 1) = 1/2$. Each time that we use the channel the input X will either get get erased with probability p_e, or it will get transmitted correctly with probability $1 - p_e$. Using the character "?" to denote erasures, the output Y of the channel can be written as

   $$Y = \begin{cases} X, & \text{with probability } 1 - p_e \\ ?, & \text{with probability } p_e. \end{cases}$$

   Compute $H(Y)$.

2

4. We defined the entropy of a single random variable as a measure of the uncertainty inherent in the distribution of the random variable. We now extend this definition for a pair of random variables (X, Y), but there is nothing really new in this definition because the pair (X, Y) can be considered to be a single vector-valued random variable. Define the *joint entropy* of a pair of discrete random variables (X, Y) to be

$$H(X, Y) \triangleq -\mathbb{E}[\log p(X, Y)],$$

where $p(\cdot, \cdot)$ is the joint PMF and the expectation is also taken over the joint distribution of X and Y.

Compute $H(X, Y)$, for the BEC. [Later, we will discuss how information-theoretic quantities such as $H(X, Y)$ are fundamentally related to the maximum *rate* at which it is possible to send information across the channel.]

Problem 5. **Soliton Distribution**
This question pertains to the **fountain codes** introduced in the lab.
Say that you wish to send n chunks of a message, $X_1, \ldots, X_n$, across a channel, but alas the channel is a **packet erasure channel**: each of the packets you send is erased with probability $p_e > 0$. Instead of sending the n chunks directly through the channel, we will instead send n packets through the channel, call them $Y_1, \ldots, Y_n$. How do we choose the packets $Y_1, \ldots, Y_n$? Let $p(\cdot)$ be a probability distribution on $\{1, \ldots, n\}$; this represents the **degree distribution** of the packets.

(i) For $i = 1, \ldots, n$, pick D_i (a random variable) according to the distribution $p(\cdot)$. Then, choose D_i random chunks among $X_1, \ldots, X_n$, and "assign" Y_i to the D_i chosen chunks.

(ii) For $i = 1, \ldots, n$, let Y_i be the XOR of all of the chunks assigned for Y_i (the number of chunks assigned for Y_i is called the **degree** of Y_i).

(iii) Send each Y_i across the channel, along with metadata which describes which chunks were assigned to Y_i.

The receiver on the other side of the channel receives the packets $Y_1, \ldots, Y_n$ (for simplicity, assume that no packets are erased by the channel; in this problem, we are just trying to understand what we should do in the ideal situation of *no* channel noise), and decoding proceeds as follows:

(i) If there is a received packet Y with only one assigned chunk X_j, then set $X_j = Y$. Then, "peel off" X_j: for all packets Y_i that X_j is assigned to, replace Y_i with Y_i XOR X_j. Remove Y and X_j (notice that this may create new degree-one packets, which allows decoding to continue).

(ii) Repeat the above step until all chunks have been decoded, or there are no remaining degree-one packets (in which case we declare failure).

In the lab, you will play around with the algorithm and watch it in action. Here, our goal is to work out what a good degree distribution $p(\cdot)$ is.
Intuitively, a good degree distribution needs to occasionally have prolific packets that have high degree; this is to ensure that all packets are connected to at least

one chunk. However, we need "most" of the packets to have low degree to make decoding easier. Ideally, we would like to choose $p(\cdot)$ such that at each step of the algorithm, there is exactly one degree-one packet.

1. Suppose that when k chunks have been recovered ($k = 0, 1, \ldots, N - 1$), then the expected number of packets of degree d (for $d > 1$) is $f_k(d)$. What is the expected number of packets of degree d whose degrees are reduced by one after the $(k + 1)$st chunk is peeled off?

2. Assuming that at each step of the algorithm, there is exactly one degree–one packet, show that $f_{k+1}(1) = f_k(2).2/(N - k)$ for $k = 0, 1, \ldots, N - 1$.

3. We want $f_k(1) = 1$ for all $k = 0, 1, \ldots, n - 1$. Show that in order for this to hold, then for all $d = 2, \ldots, n$ we have $f_k(d) = (n - k)/[d(d - 1)]$. From this, deduce what $p(d)$ must be, for $d = 1, \ldots, n - k$. (This is called the **ideal soliton distribution**.)

   [*Hint*: Your answer to (1) should give you a recurrence between $f_k(d)$, $f_k(d+1)$, and $f_{k+1}(d)$.]

4. Find the expectation of the distribution $p(\cdot)$.

In practice, the ideal soliton distribution does not perform very well because it is not enough to design the distribution to work well in expectation.