Practice Problem Set 4: Random Variables & Distributions

Reading: LLM Chapter 19 (in the 2017 revision), Schaum's Ch 5.1-5.4 and 5.7, 5.9. 6.2, 6.8c. (Note, Schaum's does not cover all the material in this problem set; Schaum's is a good start but LLM is the definitive reference.)

Optional Extra Practice. For the following problems, the parts of the problem that ask for expected value or variance are not relevant to this problem set; we will cover expected value in the next problem set. 5.1-5.2,5.4-5.7,5.11,5.54-5.58,5.61,6.1-6.11,6.44,6.52-6.60, 6.90-6.92

Exercise 1. Suppose that your code crashes on a random input with probability $\frac{1}{8}$. You run your code 12 times.

- If X is a random variable expressing the number of crashes you observe, what kind of random variable is X?
- What's the probability that your code never crashes?
- That you code crashes exactly 3 times?
- That your code crashes at least 80% of the time?
- That you run your code exactly 8 times before it crashes for the first time?

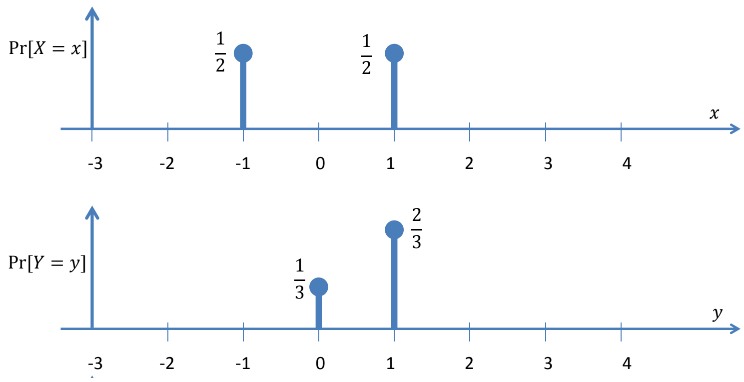
Exercise 2. A bin contains 5 red balls, and 2 blue balls. We select two random balls. For each blue ball we win \$10, for each red ball we lose \$3. Let X be the random variable that measures how much money we win (or owe). What is the range of X? What is the PDF of X? Are you more likely to win or lose money by playing this game?

Exercise 3. Two fair dice are rolled and the absolute value of the difference of the numbers on each die is given by the random variable X. Write down the Range and PDF of X. Is X a uniform random variable?

Exercise 4. A 'sketch' is a set of n counters (*i.e.,* a length n vector of integers). You are building up a sketch of stream of packets as follows; for each packet that you see, you draw a random number from $1, ..., n$ which tells you which counter in the sketch to increment by one.

- Let Y_i be the value in the the i^{th} counter after you see a total of t packets. What is probability distribution of Y_i?
- Let X be the sum of *all* the values in all the counters after you see a total of t packets. Express random variable X in terms of random variables Y_i.
- Let X be the sum of *all* the values in all the counters; what is distribution of X after you see a total of t packets?

Exercise 5. (Problem from a past midterm.) Here are the PDFs of two **independent** random variables, X and Y. Notice that $Range(X) = \{-1, 1\}$ and $Range(Y) = \{0, 1\}$.

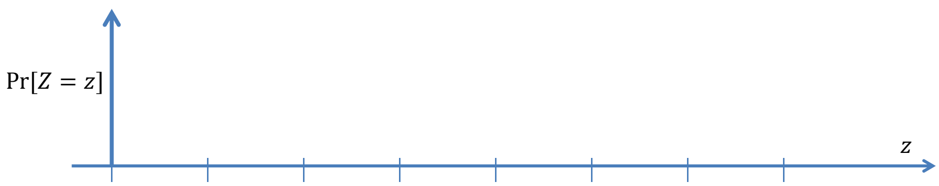


1. Let Z be another random variable such that $Z = 3X + Y$.

   What is the range of Z?

2. Determine $\Pr[Z = 3]$.

3. Plot the PDF of Z on the axes below.



4. This PDF would look different if X and Y were not independent. Given an example of dependent X and Y that result in a different PDF.

Exercise 6. (Another problem from a past midterm.) We have a hash table storing 3500-bit items in 60 bins. To populate this hash table, we have a hash function h that produces uniform random, **independent** outputs such that

$$h : \{0,1\}^{3500} \rightarrow \{1, 2, 3, ..., 60\}$$

To insert item y in the hash table, we simply store it in $h(y)$-th bin.
(Thus, to insert item x with $h(x) = 22$, we store x in the 22^{nd} bin.)
Starting with an empty table, we insert a sequence of items into the table, one after the other.

1. What is the probability that the first item lands in the first bin?

2. Let T be the number of items that we need to insert until **exactly two** items land in the first bin. What is $\Pr[T = 12]$?

3. Suppose each bin can store a maximum of 3 items. Suppose we want to insert $t = 1000$ items into the hash table. What is the probability that **at least one bin** in the hash table will overflow?

Exercise 7. You flip 3 independent, random coins. Let X_i be a random variable that takes on value 1 if coin i comes up heads, and takes on 0 otherwise.

1. Sketch the probability distribution function (PDF) of X_i. (That is, plot $\Pr[X_i = x]$ vs x for every $x \in Range(X_i)$.)

2. Consider the following random variables, for $j = \{1, ..., 7\}$. Let $j = (j_1, j_2, j_3)$ be be the binary representation of the index j. Then

$$Y_j = (X_1 \times j_1) \oplus (X_2 \times j_2) \oplus (X_3 \times j_3)$$

where $\oplus$ is the binary XOR function. Plot the PDF of Y_4, and Y_5. Are the PDFs the same?

3. Provide a counterexample that proves that the set of Y_j for $j = \{1, ..., 7\}$ are NOT 3-wise independent.

4. Are the set of Y_j for $j = \{1, ..., 7\}$ independent? If yes, provide a proof. If no, provide a counterexample.

5. Now let's generalize the example. We have n independent coins, and thus n different $\{0, 1\}$-random variables X_i defined as before. Letting $j = \{1, 2, ..., 2^n - 1\}$ and $j = (j_0, j_1, ..., j_{n-1})$ be the binary representation of index j, let

$$Y_j = \bigoplus_{k=0}^{n-1} (X_k \times j_k)$$

do you expect these Y_j random variables to be independent? 3-wise independent? Explain your answer.