CS115 MIDTERM

Write your name and student number at the top of the page, and on any blue books you use. Each question is worth equal value. HAND IN BOTH THIS QUESTION SHEET AND YOUR BLUE BOOK.

1. Here are five observations of a random variable X: 3,2,4,1,5.

   (a) Draw the empirical probability mass function (aka probabilty density function or PDF) $f_1(x)$ that one can derive directly from the five observations.

   (b) Draw the cumulative distribution function (CDF) $F_1(x)$ for $f_1(x)$.

   (c) Now assume that these 5 data points represent just the beginning of a massive influx of data that would be too big to store permanently. Instead, you will create a fixed set of bins where you're going to build an empirical histogram of the incoming deluge of data. For now assume that we're only going to have 2 bins (not including the ones extending to infinity on either side): $[0, 3)$ and $[3, 6)$. Draw the PDF $f_2(x)$ for the 5 initial samples according to the two bins.

   (d) Draw the CDF $F_2(x)$ corresponding to $f_2(x)$ for the same 5 initial samples.

2. Assume the function `irand()` contains a pseudo-random number generator that returns a non-negative integer in the range [0,`MAXINT`) where `MAXINT` is some very large positive integer. When called repeatedly, `irand()` will return a sequence of integers in that range that satisfy all the "nice" properties of uniform randomness that we discussed in lecture: the sequence covers the range uniformly, touches every integer in the range exactly once before cycling back to the beginning, and "looks" random in the sense there's no obvious way to predict the next number other than it'll be uniformly chosen at random in the range [0,`MAXINT`).

   (a) Assume that you want to choose a single-digit integer (0 through 9 inclusive) uniformly at random. Explain why it would be a bad idea to use the expression `irand()%10` where %10 means "integer mod 10".

   (b) Assume that `irand()` is the only generator at your disposal, and assume there is a function called `srand()` which is documented as a way to "seed" the `irand()` generator. Assume that once you seed the generator `irand()` returns the *next* number in the sequence after the seed. Below is code to create a wrapper around these functions to create `N` independent streams:

```
#define N 10
static int seeds[N];
void StreamSeed(int stream, int value) { seeds[stream] = value; }
int StreamNext(int stream) {
    srand(seeds[stream]); seeds[stream]=irand(); return seeds[stream];
}
void Mystery1(void) {
    int i;
    StreamSeed(0,0);
    for(i=0;i<3;i++) print(StreamNext(0)); // prints the int and a newline
}
void Mystery2(void) {
    int i;
    StreamSeed(0,0); StreamSeed(1,0);
    // print2 prints two space-separated ints on the same line, then a newline
    for(i=0;i<3;i++) print2(StreamNext(0), StreamNext(1));
}
```

Below is the output that happens if I call the `Mystery1()` function:

```
12345
54321
10000
```

In the space below, write the output that you expect to appear if we call the `Mystery2()` function.

3. Starting at time zero minutes, the queue to a single-teller bank is empty. Three customers C_1, C_2, C_3 arrive to the queue at minutes 1, 3, and 6, respectively. Assume each customer requires exactly 3 minutes of service, after which the following customer immediately enters service. (Obviously, customer C_1 enters service immediately upon arrival.) **SHOW YOUR WORK for all questions below.**

   (a) In the space below, draw a long horizontal line representing the time axis from $t = 0$ to $t = 10$. Plot a timeline of this sequence of events: starting at time zero, label each event on the time axis with the time it occurs, which customer is involved, and that event has occurred to that customer.

   (b) Draw another horizontal line below, and above it, draw the function $N(t)$, which is an integer-valued function equal to the number of people in the queue at time t.

   (c) Compute the time-average of queue size between time $t = 0$ and time $t = 10$. You may express your answer either as a fraction or as a real number.

   (d) Compute the population average of the waiting time across the three customers.

   (e) Compute the sample variance of the waiting time across the three customers.

   (f) Say you want to use these three samples to compute a 90% confidence interval on the waiting time using Student's t-distribution, $t_{k,1-\alpha/2}$. What values of k and α would you use?