ISQS 5347 Midterm 1. Closed books, notes, and no electronic devices. Points (out of 100) are in parentheses.

1. Suppose $Y$ has the following distribution:

| $y$ | $p(y)$ |
|---|---|
| 1 | .1 |
| 2 | .4 |
| 3 | .3 |
| 4 | .1 |
| 5 | .1 |

A. (10) Explain briefly how the ACTUAL NUMBERS that are produced by this pdf will look, and give a "random sample" of four numbers from this pdf.

Solution: the actual data will be numbers 1,2,3,4,5, with relatively more 2’s and 3’s. A “random sample” of four numbers is 2,2,1,3.

B. (10) This is supposed to model the waiting time (in minutes) when a person calls for technical support. Critique the model.

Solution: The model presumes that time is either 1, 2, 3, 4 or 5 minutes, with no other options. This seems odd. First, time presumably would be measured in minutes and seconds, not just minutes. But if the reported data are rounded to the nearest minute, there should be some 0’s (when the time was less than 30 seconds) and occasional numbers more than 5 minutes (when the time was more than 5 minutes and thirty seconds). A continuous model with range from 0 to infinity would be more realistic.

C. (10) If the model is valid, then in the next 100 calls, approximately how many are answered in less than 3 minutes? Briefly explain your logic.

Solution: The probability that a call is answered in less than 3 minutes is 0.1+0.4=0.5. So, we expect 50% are answered in less than 3 minutes, or approximately 50 out of the next 100.

2. The data set Work.One looks like this:

VIEWTABLE: Work.One

| | x |
|---|---|
| 1 | 3 |
| 2 | 7 |
| 3 | 5 |
| 4 | 6 |
| 5 | 1 |

The following code is run:

```
data Two;
   set One;
   XGT5 = (x>5);
   XEQ1 = (x=1);
run;
```

A. (10) What does the data set Work.Two look like?

Solution: Like this:

VIEWTABLE: Work.Two

| | x | XGT5 | XEQ1 |
|---|---|---|---|
| 1 | 3 | 0 | 0 |
| 2 | 7 | 1 | 0 |
| 3 | 5 | 0 | 0 |
| 4 | 6 | 1 | 0 |
| 5 | 1 | 0 | 1 |

B. (10) Why is the average of the “XGT5” data values an estimate of a probability?

Solution: The average is the sum (0 + 1 + 0 + 1 + 0) = 2 divided by the number sampled, which is 5. So the average counts that number of occurrences of “X>5” and divides by the total number of trials. The average is 2/5 = 0.4 = 40%, indicating that X is greater than 5, 40% of the time. Thus, the average is an estimate of the probability that X is greater than 5.

3. (10) A pdf is

$$p(x) = (x-x^2)/c, \text{ for } 0<x<1, \quad \text{and}$$
$$p(x) = 0 \quad , \text{ otherwise.}$$

For $p(x)$ to be a valid pdf, the area under the curve must be equal to one. Find the constant $c$.

Solution: We need a $c$ for which $\int_0^1 \frac{x-x^2}{c}dx = 1$. Using integral operations,

$\int_0^1 \frac{x-x^2}{c}dx = \frac{1}{c}\int_0^1 (x-x^2)dx = \frac{1}{c}\int_0^1 xdx - \frac{1}{c}\int_0^1 x^2dx$. Now, $\int_0^1 xdx = \left.\frac{x^2}{2}\right|_0^1 = \frac{1^2}{2} - \frac{0^2}{2} = \frac{1}{2}$, and

$\int_0^1 x^2dx = \left.\frac{x^3}{3}\right|_0^1 = \frac{1^3}{3} - \frac{0^3}{3} = \frac{1}{3}$. Hence $\int_0^1 \frac{x-x^2}{c}dx = \frac{1}{c}\left(\frac{1}{2} - \frac{1}{3}\right) = \frac{1}{c}\left(\frac{1}{6}\right)$. Setting this to 1.0 yields $c$=1/6.

4. (15) Consider the following equation:

$$\hat{x}_{(i-.5)/n} = x_{(i)}$$

Explain the equation.

Solution: The left hand side is the estimated $(i-.5)/n$ quantile of a distribution. The right hand side is the $i^{\text{th}}$ order statistic from a sample of $n$ observations. (The "$n$" is implicit from the left hand side.) So, in words, "the $i^{\text{th}}$ ordered value from a sample of $n$ observations is an estimate of the $(i-.5)/n$ quantile of the distribution from which that $n$ data values were sampled. For instance, in a sample of $n$=4 observations, the second-smallest observation is an estimate of the (2−.5)/4 = 0.375 quantile of the distribution.

5. Suppose

$$F(x) = 1 - 1/x^2 \quad, \text{ for } x>1, \text{ and}$$
$$F(x) = 0 \quad , \text{ for } x \le 1.$$

A. (5) Is this a valid cumulative distribution function (cdf)? Explain.

Solution: The function increases monotonically in $x$, goes to zero as $x$ decreases, and goes to 1.0 as increases. So, yes, it is valid.

B. (5) Assuming it is valid, find the 0.75 quantile.

Solution: Set $0.75 = 1 - 1/x^2$ and solve for $x$: here $x = 2$.

C. (5) Find the pdf. (The pdf is the derivative of the cdf).

Solution: The pdf is given by $p(x) = \mathrm{d}/dx\ (1 - 1/x^2) = d/dx\ (1 - x^{-2}) = 0 - (-2)x^{-3} = 2/x^3$, for $x>1$, and $p(x) = 0$ otherwise.

6. (10) How does simulation illustrate the “model produces data” concept? Give a specific example discussed in class or homework.

Solution: In the car sales example, the real process is the car dealership, its customers who visit, and their decisions to purchase or not purchase a luxury car. This process results in daily car sales data like 1, 0 ,0, 0 ,2, 0, 0, etc. We can mimic these data using a model (the Poisson model) to produce data that are qualitatively similar to the numbers produced in the real study. Using simulation, we can produce many data sets using $\lambda$ to see whether the data produced by the model reasonably match the data produced in the real study, at least for some $\lambda$ values.