ISQS 5347 Midterm 1, Spring 2010

Instructions: Closes book and notes, no electronic devices. Put answers on blank paper. Points out of 100 are in parentheses.

1.(30) The paradigm of statistical science is that many real processes are well modeled using probabilistic models. A probabilistic model produces data, just like the real process produces data. Give an example of a real process that produces data, and describe a probabilistic model that produces data that is similar to the data produced by your real process. Make sure that your model is probabilistic; very little credit will be given for models that are not probabilistic.

Solution: One example is the death process discussed in class. On any given day, there will be a number of deaths in the elderly facilities. Some days 0, some 1, some 2, etc. A model that produces data that looks very much like the real data is the Poisson model: It also produces 0's, 1's, 2's, etc., in frequencies that look very close to those observed in the real process.

Other examples discussed in class: (i) Coin flips look like data produced by a Bernoulli distribution; (ii) stock returns look like data produced by a normal distribution (although perhaps better, they look more like data produced by a distribution with heavier tails than the normal distribution); (iii) stoplight times look like data produced by the U(0,2) distribution; (iv) questionnaire responses look like they are produced by a generic discrete distribution; (v) waiting times look like they are produced by an exponential distribution.

2.(10) Using calculus, find x that gives the minimum value of $f(x) = (2–x)^2 + (10–x)^2$.

Solution: The minimum occurs where the slope is zero. Using the additivity property and the chain rule, the slope is $(d/dx)f(x) = 2(2–x)( –1) + 2(10–x)( –1)$; setting this to zero implies

$2(2–x)( –1) + 2(10–x)( –1) =0$; or equivalently

$(2–x) + (10–x) =0$; or equivalently

$12 – 2x = 0$; or equivalently $x = 6$. So $x=6$ minimizes the function. (Recall from class that this least squares function is minimized using the mean value, which is $(2+10)/2=6$).

3.(20) Give examples of a discrete random variable and of a continuous random variable. Write your answers so that it is clear that the variables are indeed "random."

Solution: An example of a discrete random variable is the outcome you will see when you toss a die. (It can be 1,2,3,4,5, or 6). An example of a continuous random variable is the time you will have to wait until customer service answers. (It can be any number greater than 0 minutes.) Other examples: Discrete: coin, deaths, survey response on questionnaire. Continuous: height, weight, stock return or earnings, calories in beer, income, house price.

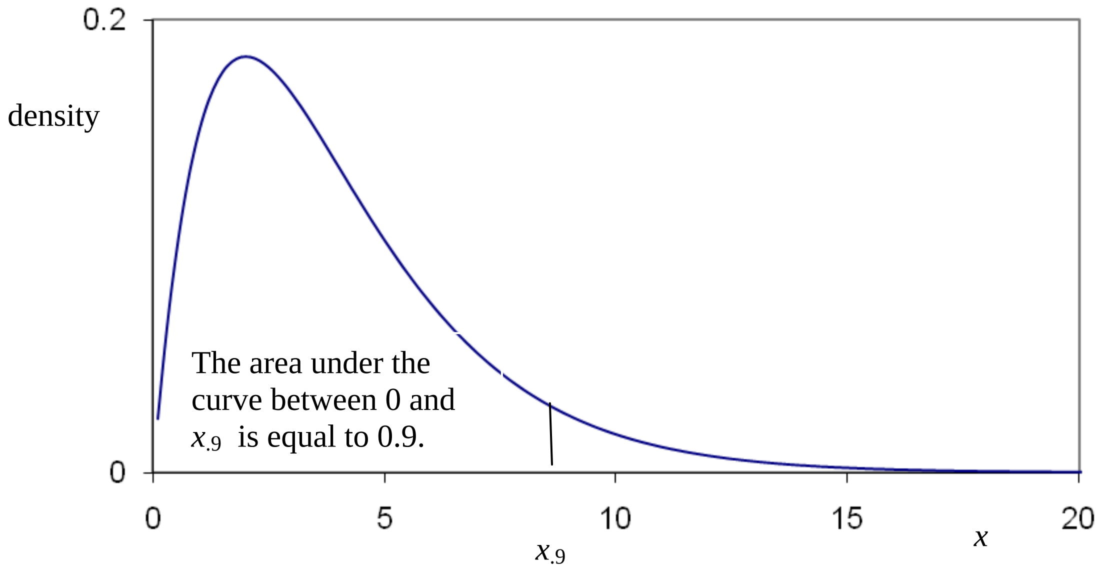
4. What does the following code do? Explain lines 1, 2 and 3.

```
data new;                   /* line 1 (7) */
   set mylib.grunfeld;      /* line 2 (6) */
   if Comp = "GM";          /* line 3 (7) */
run;
```

Solution:
Line 1 tells SAS to create a data set called "new" in the "work" library; this is a temporary library.
Line 2 tells SAS to use uses the existing data set "grunfeld" that is in the library called "mylib" to create the data set "work.new".
Line 3 tells SAS to keep only those records in mylib.grunfeld for which the variable called "Comp" is equal to GM.

5.(20) Draw a graph showing a continuous **non-normal** probability distribution function, as well as the location of the 0.9 quantile of that distribution. Show how "0.9" appears in the graph.

Solution:


0.2
density
The area under the curve between 0 and $x_{.9}$ is equal to 0.9.
0
0
5
10
15
20
$x_{.9}$
$x$