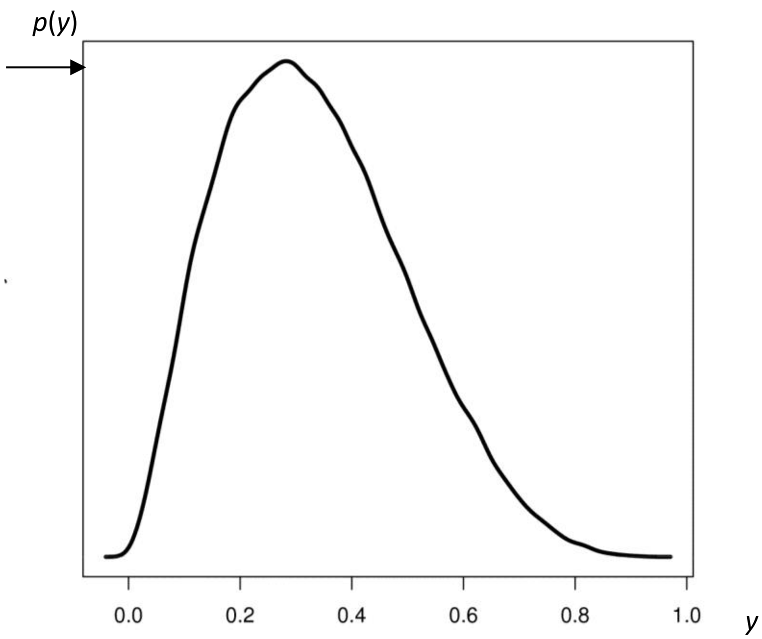
ISQS 5347 Midterm 1   Name:

Closed book, notes, and no electronic devices. **<u>DO NOT WRITE ON THIS EXAM SHEET.</u>**

1. The following graph is a probability distribution function (pdf), $p(y)$.



Each of A. - H. is worth 5 points. Be very brief for all of A. – H.

A. What is the number on the vertical axis shown by the ⟶ ? Give your best guess.

Solution: Around 2.0. The area under the curve looks to be about half of the enclosing rectangle, which has area 1 × 2 = 2.

B. How do you know that this is a continuous pdf?

Solution: Because the possible values of *Y* fall in a continuum; namely, the continuous range from 0 to 1.

C. Guess the value of Pr(Y > .5).

Solution: Maybe around 0.25.

D. Express Pr(*Y* > .5) as an integral.

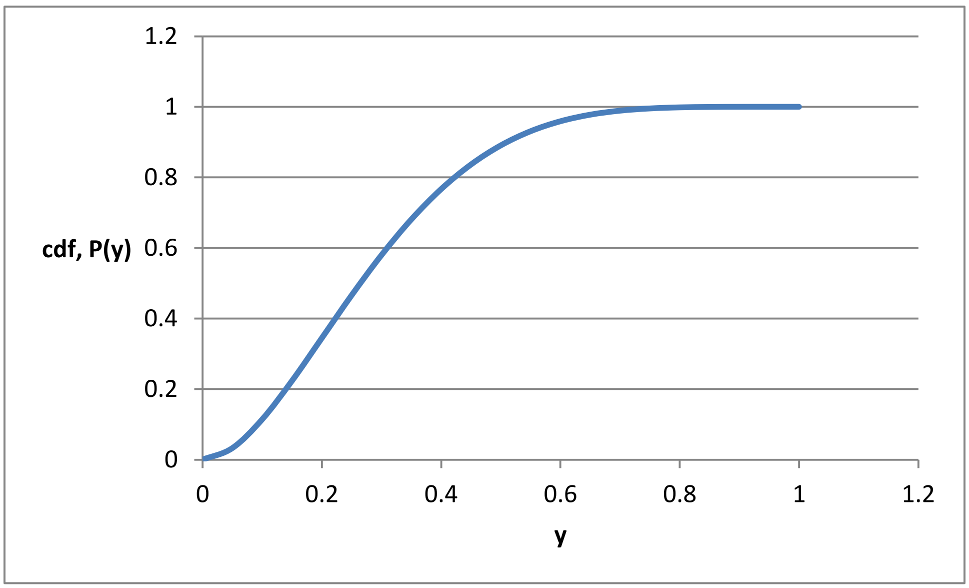
Solution: $\int_{.5}^{1} p(y)dy$

E. Guess the value of the median.

Solution: Around 0.4

F. Draw a graph of the cumulative distribution function corresponding to this pdf. Label the axes and put numbers on them.

Solution: Something like this:



G. Write down a sample of 10 data* values that might be randomly produced from this pdf.

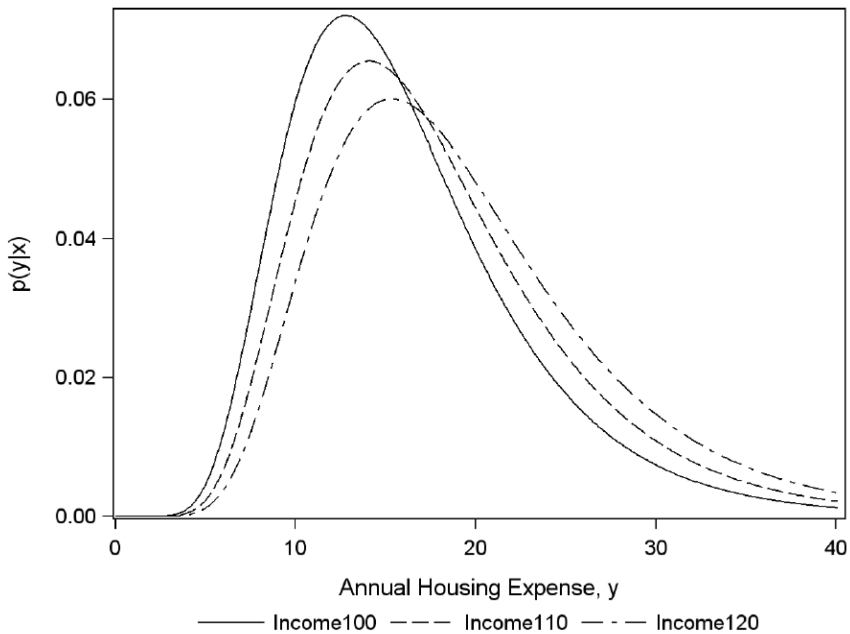
Solution: 0.22, 0.09, 0.31, 0.51, 0.49, 0.60, 0.03, 0.23, 0.47, 0.11

H. The peak of the pdf curve *p*(*y*) shown in the graph locates the *mode* of the distribution. Write an equation involving a derivative that can be solved to find the mode.

Solution: $\partial p(y)/\partial y = 0$.

2. (15) Income (X) and house expense (Y) are dependent. What does this mean? Use the *definition* of dependence in your answer, and make your answer specific to income and house expense. Draw relevant graphs to illustrate your answer.

   Solution: The distribution of house expense changes for different values income. Here is a picture showing how these distributions might look for three different values of income.



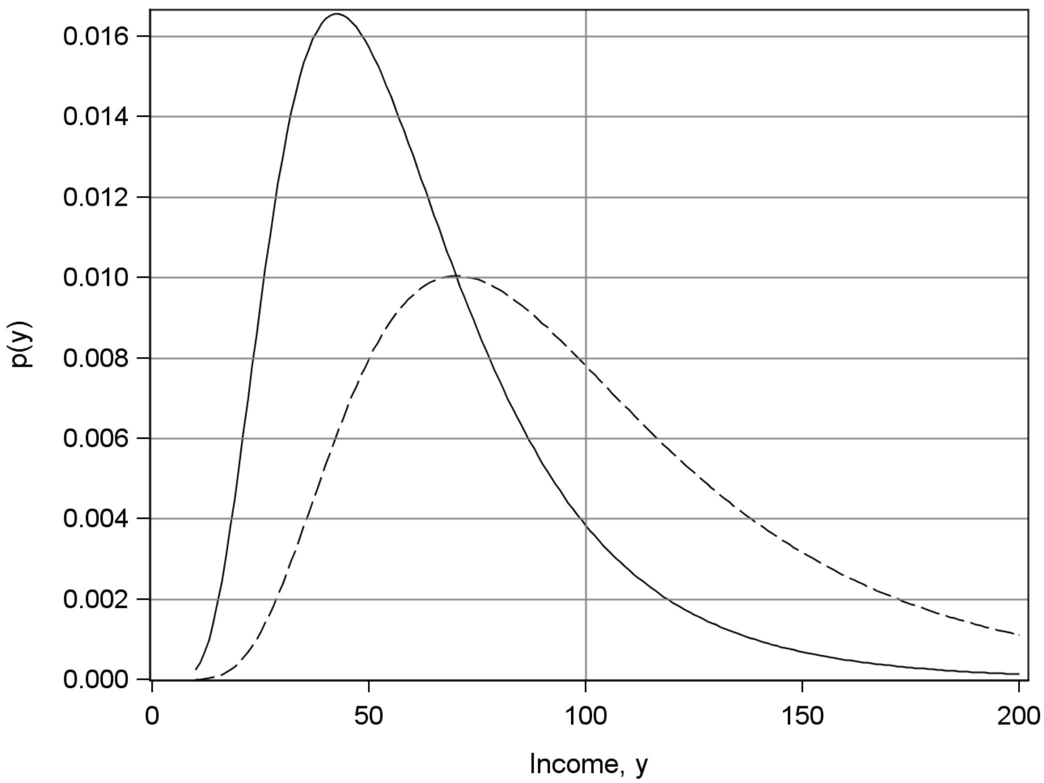
3. (15) Hans says that charitable contributions have the "Lutz distribution," one that he likes. Describe carefully how you would use simulation to estimate the probability that charitable contributions are more than \$5000, assuming you have a computer random number generator than can simulate data* from the Lutz distribution.

   Solution: Use the computer to randomly generate a large number, say 10,000 numbers from the Lutz distribution. Count how many of those 10,000 numbers are more than 5000, and divide the result by 10,000.

4. (15) *Model produces data*. Pick a real example of your own choosing, one where the DATA are variable, but not one involving dice or coins or other games of chance, and not one from the class or book. Explain this concept, *model produces data*, in terms of your chosen example.

   Solution: Every time a soccer player kicks the ball at the goal, the outcome is variable: $Y = 0$ (no goal) or $Y = 1$ (goal). A model to produce data that looks like these data is the Bernoulli model, which produces 0/1 data that look just like the real soccer data. There is very little scoring in soccer, so most of the Y's will be 0's. For the Bernoulli model to be realistic, the value of π must therefore be very small.

5. (15) An Internet marketing company thinks that 50% of the visitors to a web site click on the banner advertisement. Based on customer surveys, they have been able to estimate the income distributions for customers who click on their banner ad, and for customers who do not click. The following graph shows these distributions: The solid line indicates non-clickers, and the dashed line indicates clickers.



Based on this graph and market research, about what percent of people with income = 100 will click? Provide relevant explanatory math along with your answer.

Solution: p(click | income = 100) $\propto$ p(income = 100 | click) * p(click). When click = yes, we get (roughly, from visual inspection), p(click = yes | income = 100) $\propto$ p(income = 100 | click = yes) p(click = yes) = .008*.5 = .004. When click = no, we get p(click = no | income = 100) $\propto$ p(income = 100 | click = no) p(click = no) = .004*.5 = .002. So the probability is about .004/(.004 + .002) = 2/3 = 67%. In other words, according to this model, about 2/3 of the people with income 100 will click.