

Lab 1

36-350, Statistical Computing

Friday, 31 August 2012

Agenda: Manipulating data objects; using the built-in functions, doing numerical calculations, and basic plots; reinforcing core probabilistic ideas.

General instructions for labs: For questions which ask you to do something, give the command you use to do it. For questions which ask for a plot, have the TA see the plot and check off the question. For questions which ask you to explain, write out an explanation in coherent, complete sentences. (You will be graded on your written explanation, not what you might say to the TA.) Clearly indicate which answer goes with which question.

Point of this lab:

The exponential distribution is defined by its cumulative distribution function

$$F(x) = 1 - e^{-\lambda x}$$

The R function `rexp` generates random variables with an exponential distribution.

`rexp(n=10,rate=5)`

produces 10 exponentially-distributed numbers with rate (λ) of 5. If the second argument is omitted, the default rate is 1; this is the “standard exponential distribution”.

1. (5 points) Generate 100 random values from the standard exponential distribution and store them in a vector `x1`. Find the mean and standard deviation of `x1`.
2. (5 points) Repeat, but changing the rate to 0.1, 0.5, 5 and 10, storing the results in different vectors.
3. (15 points) Make scatterplots of
 - (a) the five means versus the five rates;
 - (b) the standard deviations versus the rates;
 - (c) and the means versus the standard deviations.

Explain, in words, what's going on in these plots.

4. (5 points) Generate one million numbers from the standard exponential distribution and store them in a vector called `y`. Find the mean and standard deviation.
5. (15 points) Plot a histogram of `y`. Does it match the function $1 - e^{-x}$? Should it?
6. (5 points) Create a matrix, `y.mat`, with the values in `y`, 1000 rows, and 1000 columns.
7. (10 points) Explain what happens when you create a histogram of `y.mat`.
8. (2 points) Find the mean of the 371st column of `y.mat`.
9. (8 points) Find the means of all 1000 columns of `y.mat`. Plot the histogram of column means. Explain why its shape does not match the histogram in problem 5.
10. (10 points) Find the mean of all of the entries in `y` which are strictly greater than 1.
11. (10 points) Explain what these two commands do:

```
sum(y > 1)
mean(y[y.mat>1])
```

12. (10 points) Take the square of each number in `y`, and find their mean. Explain this in terms of the mean and standard deviation of `y`. *Hint:* think carefully about the formula R uses to calculate the standard deviation.