ENGR 113: Lab 3
Probability Distributions in R

Objectives:

1. To explore the binomial, Poisson, uniform, and exponential distributions.

2. To learn learn relevant commands in R for probability distributions.

Generating Random Numbers in R

The basis for most random number generation is the uniform distribution. To get a random value in [0,1] with uniform probability over the interval, enter {`runif(1)`}, which will return one value. To get 100 of them, use `runif(100)`. Better yet, we will save a vector of 100 samples and take a look at its properties. Enter {`x <- runif(100)`}. Now look at its five number summary {`summary(x)`} and histogram {`hist(x)`} (don't forget to open `x11()` first). Are the median and quartiles about where you expect them to be? Does the histogram show a roughly uniform distribution? These are random numbers, so they may not match quite what you expect. As you use a larger sample, they should look more like you expect. Try {`x <- runif(10000)`} and then look at the summary and histogram and see that they should be pretty close to what theory predicts.

What if you want numbers uniformly distributed over a different interval, such as [10,20]? `runif()` can take two more arguments, a minimum and maximum for the interval. Thus {runif(1,10,20)} will produce a single random value between 10 and 20. You can try 10,000 of them and see that they are uniformly distributed between 10 and 20.

To get random values from other distributions, the syntax is similar. All of the commands start with "r" and then there is some abbreviation for the name of the distribution. Try {`x <- rbinom(1000,100,.5)`} to get 1000 observations of a binomial with $n = 100$ and $p = \frac{1}{2}$, e.g., the result of flipping a fair coin 100 times. Then try {`summary(x)`} and {`hist(x)`}. Do the plots look like you expect? With $p = .5$, the histogram should be symmetric (although it may look a little different depending on the particular bins). What is the theoretical mean of this distribution? Is the mean of your sample close? How about the theoretical variance? Compare the sample variance with {`var(x)`}. Now try a different binomial, such as {`x <- rbinom(1000,100,.8)`}. Check its summary and histogram. Notice how it is skewed, and see that its mean is where you expect it.

Poissons and exponentials are similar. To get a single random Poisson with mean 10, you would use `rpois(1,10)`. To get a single random exponential with rate 10, you use `rexp(1,10)`. Note that an exponential with rate 10 has mean $\frac{1}{10}$. To convince yourself that this is true, try {`summary(rexp(1000,10))`} and {`summary(rexp(1000,.1))`}.

Note: if you ever forget what the arguments are for a command in R, you can just type the name of the command, and it will usually remind you. Try {`rbinom`}.

Working with the distributions

For a discrete distribution, such as a binomial or a Poisson, the probability density function gives the probability that the random variable will be equal to a particular value. R will give you these numbers with functions starting with "d". For example, to see the probability that a binomial with $n = 5$ and $p = .25$ is equal to 3, enter {`dbinom(3,5,.25)`} and see that this gives the same result as we saw in class in the example on guessing on a multiple choice test. R will also give you cumulative probabilities, i.e., the probability that a random draw will be less than or equal to a particular value. For example, $P(X \leq 3)$ is {`pbinom(3,5,.25)`}. Notice that this is $1 - P(X \geq 4)$, so you can also compare this result to the example in class where we computed $P(X \geq 4)$. The syntax is comparable for the Poisson distribution, `dpois()` and `ppois()`, although instead of giving x, n and p, you only need to give x and λ.

For a continuous distribution, such as the exponential (or the normal, as we will see in chapters 6, 7, and beyond), the probability density function is a function that gives relative likelihoods, and that when integrated over an interval gives the probability that the variable will be in that interval. `dexp()` gives the density function for the exponential distribution. Usually of more interest is the cumulative distribution function, the probability that the random variable is less than or equal to a particular value, i.e., $P(X \leq x)$. Just like with the discrete variables, we can use a "p" prefix in R. So to get the probability that a random exponential with rate 10 (mean $\frac{1}{10}$) is less then .05, enter {`pexp(.05,10)`}. Recall that the CDF for an exponential is $F(x) = 1 - e^{-\lambda x}$. Does the result from R match what you get from evaluating the CDF yourself? For another example, what is the probability that a random exponential with rate 10 is between .4 and .5? To get the answer, enter {`pexp(.5,10) - pexp(.4,10)`}.