Math 1180:Lab8

Kyle Gaffney

February 26th, 2014

1 Binomial Distribution

Suppose that an event has probability p of happening and probability (1-p) of not happening. We call such a random variable a Bernoulli random variable. If we consider lots of independent Bernoulli events occuring then an interesting statistical question is how many of each type of result should we expect?

Suppose that we flip n many weighted coins where heads comes up with probability p and tails comes up with probability (1-p). Then what is the probability that we will get k many heads?

First we need to count how many ways there are to get k many heads. Of the n coin flips we need to pick k of them to be heads. This sounds like a combinatorial combination (n choose k) and has formula

$$\binom{n}{k} = \frac{n!}{(n-k)!k!}$$

Okay now we need the probability of getting exactly k heads and (n-k) tails. This is simply $(p)^k$ and $(1-p)^{(n-k)}$.

Then Pulling everything together we get that the probability of getting k many heads out of n many coin flips is:

$$\binom{n}{k}(1-p)^{(n-k)}p^k$$

Now as we learned earlier R has an ingrained function to generate numbers from a binomial distribution. The above formula however gives us another way to write our own binomial number generator.

Let us simulate the coin flipping problem where p=0.6, and we are flipping 10 coins and counting how many come up heads, and we do this 1000 times

```
Result=0
for (i in 1:1000){
```

```
for (j in 1:10){
Result[i]=sum(runif(10)<.6)}}
summary(Result)
head(Result)
A=hist(Result,freq=F,breaks=seq(-.5,10.5))
summary(A)
A$counts
A$density
```

Now we can also calculate the expect probability that we would get 6 heads out of ten flips making R do the algebra for us:

```
prob6=factorial(10)/(factorial(10-6)*factorial(6))*(0.6)^6*(0.4)^4
```

How does this compare to the data that we collected?

We also have a new neat way to plot the CDF for our data:

```
barplot(cumsum(A$density),names.arg=seq(0,10))
```

2 Geometric Distribution

The Binomial Distribution gives the probability that a certain number of molecules are still in a cell at time t=T. The Geometric Distribution gives the probability that a molecule is still in the cell at time t in a discrete sense.

Suppose that the probability that a molecule leaves the cell in one timestep is q. Then the probability that the molecule stays in the cell during that timestep is (1-q). So then the probability that the molecule is still in the cell is as follows:

$$p_{t+1} = (1-q) * p_t$$
$$p_t = (1-q)^t$$

What about the probability that the molecule leaves the cell during the t^{th} time step?

$$g_t = q * p_{t-1} = q * (1-q)^{(t-1)} \tag{1}$$

Suppose that we wish to generate a data set of the process of a molecule leaving a cell and want to find the time that the molecule first leaves. Suppose that the probability of the molecule leaving during a time step (1 second) is q=0.2

```
LTime=0
for (i in 1:1000){
LTime[i]=which(runif(100)<.2)[1]-1}
b=hist(LTime,freq=F,breaks=seq(-.0,max(LTime)+.5))
```

How could we plot the probability that the molecule is still in the cell at time T?

3 Assignment for the Week

Suppose that we have a cell that has 20 molecules in it. The probability of a molecule leaving the cell is p=0.45. Create a simulation to collect data on this process and run it 10000 times.

1. What type of distribution are you modelling?
2. Plot a histogram of your results and scale it so that it is an approximation of the PDF for this process
3. Create a plot of the CDF for the data set. From this plot and the data you collected determine what the probability that there are between 4 and 7 molecules still in the cell.