STAT 270 Chapter 3 Sections 4,5,and 6 = Discrete Probability Models

The models we introduce in these sections are very useful: Binomial, Hypergeometric, Negative Binomial, Poisson. We will motivate the models today – next day we will explore the characteristics of the models (like mean and variance).

Binomial models probabilities for the number of "successes" in n Bernoulli trials.
Negative Binomial models probs for the number of Bernoulli trials until the kth "success".
Hypergeometric models the probs for number of a certain kind selected when sampling is without replacement.
Poisson models the probs for the number of events that occur in a certain length of time.

(The Geometric model is not featured separately only because it is a special case of the Negative Binomial, k=1)

To describe these in more detail, lets recall what "Bernoulli Trials" are. Think of a series of coin tosses – what are the features? "trial" just means "experiment" in this context.
i) independent trials
ii) each trial has an outcome that is either event E or not E.
iii) probability of E has a fixed probability p (i.e. not depending on the trial number)

Any sequence of trials satisfying i), ii), iii) is called Bernoulli trials. Note that each trial can be described by a "Bernoulli" random variable.

A "Binomial experiment" (p 120) is a sequence of Bernoulli trials in which the number of trials is fixed in advance (and usually denoted n). So the parameters of a Binomial experiment are the two values n,p. And the number of successes is the Binomial random variable whose probabilities are given by the Binomial Distribution.

So, for example, the Binomial model (which we will show soon) tells us that, with a fair coin, the chance of getting exactly 5 heads in 10 tosses is .246, and if the coin is biased so that P(head)=.6, the probability is then only .201.

```
> dbinom(5,10,.5)
[1] 0.2460938
> dbinom(5,10,.6)
[1] 0.2006581
```

But who cares about the number of heads in 10 coin tosses? However, the same model applies to sampling from a large population (as in an opinion survey), or counting speeders on a freeway or terrorists at the airport!

The formula for the binomial probabilities is (p 123):

$P(X=x)=C_{n,x}\, p^x(1-p)^{(n-x)}$   x=0,1,2,...,n

which looks a little intimidating but actually has a simple structure.

Consider a sequence of 5 Bernoulli trials - for example tosses of a coin for which P(Head)=p.

What is the probability of a sequence of outcomes T H H T H ? Surely it is, by independence, $(1-p)pp(1-p)p = p^3(1-p)^2$

This particular sequence has 3 Hs and 2 Ts. But any sequence with 3Hs and 2Ts must have this same probability. As long as we are focussing on outcomes with 3Hs and 2Ts, we have equally likely outcomes, and to compute the probability of one or other of these outcomes occurring, we just need to count up the number of sequences like THHTH. In this case we can figure out an exhaustive list, but it would be nice to have a formula for the number of ways this can happen. But we do! Every sequence of length 5 with 3 Hs and 2 Ts can be specified by specifying the order numbers of the 3 Hs. There are five possibilities {1,2,3,4,5} and we need to choose 3 of them (without regard to order – just which subset are we choosing). If we had chosen the subset {2,3,5} we would have selected THHTH. But there are $C_{3,5}$ ways to select this subset, and so there are $C_{3,5}$ ways to create the length 5 sequence with 3 Hs and 2 Ts. So P(number of Hs in 5 tosses =3)= $C_{3,5}$ times $p^3(1-p)^2$. The general case is the formula

$P(X=x)=C_{n,x}\, p^x(1-p)^{(n-x)}$   x=0,1,2,...,n

for the binomial probabilities.

So if p = .5 (to make the calculation easy) and n=5 then P(no of Hs = 3) = $C_{5,3}(.5)^3(1-.5)^2$ = 10/32. Clearly the formula simplifies the calculation. However, the formula is not too difficult to derive or explain, and so you should know how to do this.

Some R commands for this:

```
> dbinom(3,5,.5)
```

```
[1] 0.3125
> 10/32
[1] 0.3125
```

Let's see if you understand the binomial experiment, and at the same time I will suggest a use for use of the binomial probability formula.

Suppose we select a random sample of 25 names from the thousands of subscribers to the ABC magazine. We ask them to rate the magazine's recent issue on a 1 to 5 scale with 5 being "very satisfied". The magazines objective is to have 80 percent of its subscribers "very satisfied". The results of the survey are that 15 are very satisfied and others less so. What is the chance that 80 percent of the subscribers are very satisfied and yet only 15 of the 25 in the sample express this view?

A useful calculation that will help to answer this question is

P(X=1,2,....,15) = $\sum_{x=1}^{15} C_{x,25}.8^{x}(1-.8)^{25-x}$ and after a lot of arithmetic, this turns out to be 0.017. In other words, the chance that there would be this few "very satisfied" or less, IF the true proportion of subscribers is 0.8, is less than .02 – not likely. We really would expect more than 15 "very satisfied" if the true proportion in the entire list were 0.80. This suggests that the magazine still has some work to do to reach its goal – the 0.8 is probably not true.

R for above calculation

```
> pbinom(15,25,.8)
[1] 0.01733187
```

The logic above may seem complex – we will get more of that later in the course. The main thing for now is that you know how to recognize a binomial experiment and also how to compute P(X=x).

--------------------------------------------------------------------------------

Next model for us is the Negative Binomial. Here is another unfortunate jargon anomaly. The Negative Binomial is not a Binomial!

Let Y be the number of failures in Bernoulli trials, P(success)=p, until the kth success occurs. For example, how many times do we have to toss a fair coin to achieve 5 successes? Here k=5. Actually we have defined Y as the number of failures, so the number of trials until the kth success will be Y+k.

It turns out that (Compare p 132)

P(Y=y) = $C_{(k-1),(y+k-1)}\ p^{k}(1-p)^{y}$ where y=0,1,2,3,....

This looks a bit like the Binomial probability but note that the combinatorial coefficient is selecting k-1 items from the variable number y+k-1, wheras in the Binomial we were selecting the variable number x from the fixed number n.

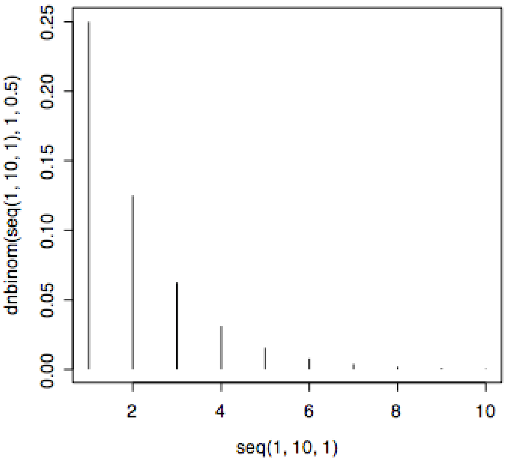
Again, lets see why this formula has the form it has:

First of all, the probability for a particular sequence of outcomes in which the kth success occurs just after y failures is $p^k(1-p)^y$. But there are many such sequences. we have to consider the y+k-1 trials before the kth success, and think of how many ways we can select the order numbers for the k-1 successes. But that number is clearly $C_{(k-1),(y+k-1)}$. QED!

Now lets look at the special case when k=1. In other words, what are the probabilities for the various sequences of failures that we might get before the *first* success. This random variable, Z say, is the geometric distribution, and the above formula specializes to

$P(Y=y) = p(1-p)^y$ where y=0,1,2,3,....

For p=.5, it looks like this:



[R is > plot(seq(1,10,1),dnbinom(seq(1,10,1),1,.5),type="h")

If you toss a fair coin, the chance that you have exactly one failure before the first success is 0.25, as can be read from the graph.

----------------------------------------------------------------------

Next, The Hypergeometric Distribution

The setting is this. A population of N items, m of them a type 1 and n of them are type 2, and N=n+m. Now we take a random sample, *without replacement*, of k items from this population.

The hypergeometric random variable X in this scenario is the number of type 1 items drawn in the sample of size k.

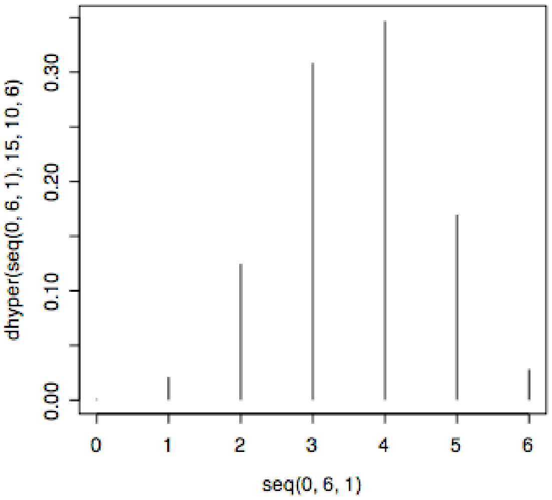
See the formula p 130. It will look a little different that the following but is actually the same with different symbols.

$P(X = x) = \frac{C_{x,n}C_{(k-x),m}}{C_{k,(n+m)}}$ for any x in the possible range.

p 130 has "possible range" defined but to explain – you cant draw more type 1 items than there are in the population, and you cant draw more than the total sample size k. Similarly for the type 2 items. This is what the constrained range of x involves.

We have already seen examples of this distribution – See Example 2.23 on p 72-73.
Here is a graph for that instance (N=25, m=15, n=10, k=6)

```
> plot(seq(0,6,1),dhyper(seq(0,6,1),15,10,6),type="h")
```



Next, the Poisson Probability Model

Suppose you in a position to observe the southward migration of grey whales heading down the west coast. Suppose we know that they pass our view point at an average rate of 1 whale per minute. Of course in a particular minute, there may be 0,1,2,3,... whales pass by – it is only the average rate that is 1 per minute. The Poisson distribution provides probabilities for these possibilities – in this case P(X=0)=.37, P(X=1)=.37, P(X=2)=.18, P(X=3)=.06, ...

Of course there are certain assumptions about the nature of the stream of events that must be satisfied in order for this probability law to apply. The nice feature of this distribution is that these assumptions are easy to understand and to judge applicability in a given situation.

In general, suppose the average rate that events occur is λ, then the probability that x events occur in one time unit is

$P(X = x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$ for x=0,1,2, ...

(For t time units, the mean would be λt in place of λ in the formula).

So P(X=0)=$e^{-1}$ = .37 (since e = 2.72 approx) as claimed earlier.

To get this to work, the assumptions are roughly (details p 137)

1. probability of an event in (t,t+Δt) is proportional to Δt for Δt small
2. only zero or one pulse in each small interval (t,t+Δt)
3. disjoint intervals have independent numbers of events

Might this apply to

1. Traffic on King George Highway past 100 Ave during 8 am – 8 pm ? No.
2. Number of rain storms in Vancouver during January. Maybe.
3. Number of requests of the help line during 10am-11am. Yes.