Stat 280 Sept. 24, 2001

More on Discrete Random Variables and their Probability Mass Functions (Models)

We talked about Bernoulli Trials, and related models:

Bernoulli - Think coin toss or sample of size 1
Binomial - Think sequence of coin tosses, or a sample of size n from a large population, or a sample of size n from a small population with replacement (SWR), 2 types
Geometric - Think of number of trials until first "success"
Negative Binomial - Think of number of trials until rth "success"

Next we had non-Bernoulli trial models:

Hypergeometric - sampling from a small population without replacement, 2 types
Multinomial - sampling from a large population, or sampling with replacement from a small population, r types
Poisson – number of events in a fixed time if average rate does not change

And there were many connections between the model contexts:

Geometric was special case of Negative Binomial
Binomial was approx to Hypergeometric good for large populations (or SWR)
Multinomial reduces to Binomial for r=2
Binomial approximates Poisson as discrete time analogue

For each model we have formulas for the mean and variance, and these also help to reveal the relationships among the models.

And we discussed how to simulate some of these distributions, and how to examine the relationship of a model to its parameters (using pdf and plot in MINITAB). Bit more today.

Final topic in Ch 3: Moment Generating Functions

What is a moment? The kth moment of a rv X is $E(X^k)$.

Suppose we have X as a Bernoulli rv (0-1 representation) with P(1)=.3.

I'll generate 1000 Bernoullis – here are some of them

C1

1	0	1	0	1	0	1	0	1	0	0	0	0	0	0
0	1	0	0	1	0	0	0	1	1	0	0	0	1	1
0	0	1	0	0	1	0	0	0	0	1	0	0	0	1

Now we can compute X^k for each of these easily since they are still 0s and 1s for any k. So the average value of X^k in this case will be the average of the simulated Bernoullis, which for this particular simulation was .304 (close to the long run value of .3).

So all moments of this simple model are .3.

Other models are not so simple:

Try Bin(10,.3):

Here are some of them:

C5

2	3	2	4	2	2	1	4	3	1	5	5	5	5	2
4	2	1	6	3	5	4	4	5	4					

In this case, the sample mean of X^k is 2.997 for k=1, 10.965 for k=2 and 45.79 for k=3.

Is there a formula way to do this, without simulation?

Moment Generating Functions (MGFs).

$M_X(t) = E(e^{Xt})$ is the general definition. For
Bernoulli $M_X(t) = E[e^{1t}.P(X=1) + e^{0t}.P(X=0)] = pe^t+1-p$ where $p=P(X=1)$.
For Binomial $= (pe^t+1-p)^n$ see p 131 for one derivation (or use $M_{X+Y}(t)=M_X(t)M_Y(t)$).

And can derive via derivatives (p 132) that $E(X)=np$ and since $Var(X) = E(X^2)-E^2(X)$
Therefore $Var(X) = np(1-p)$.

MGFs turn the summations implied by the formula for the kth moment into derivatives of the MGF. Of course the MGF itself is needed for this.

This completes Ch 3. More now on programming MINITAB:

1. See e-mail for syntax of commands and subcommands
2. Simulations may be done by the MINITAB command RANDOM if the distribution simulated appears in the list of subcommands of RANDOM. Use the HELP facility in Minitab to find the list. If the desired distribution is not in the list, then you have to program MINITAB as a global macro.

3. Global macros look like this:

```
GMACRO
Name of macro
MINITAB Commands
.
.
ENDMACRO
```

You put these commands in a file that is the same name as the macro, in text format, but with the extension mac.

4. To run the program you issue the MINITAB command (I'll assume the name of macro is "Loop")

%Loop.mac

If the file Loop.mac is in the macro folder within the MINITAB folder, it will run. If the file is on a diskette, or saved somewhere else, you will have to provide the path in the % command, like

%a:\programs\loop.mac

Assignment from Ch 4: Markov Chains

Due at class or tutorial Wed Oct 10: Exercises from Ch 4; 4.1-3, 4.2-8, 4.3-4, 4.4-3(interpretation only), 4.5-4, 4.6-6, 4.7-8

In-class test Oct 1- answers discussed in same class (mark your own).