Lab Project 3: The Binomial Distribution

Course : Introduction to Probability and Statistics, Math 113 Section 3234
Instructor: Abhijit Champanerkar
Date: Nov 7th 2012

Theoretical Binomial Probability Distribution

Let the probability of a success be p and of failure be q. The formula (to be covered in class) for the probability of k successes in n trials is

$$P(k) = p^k q^{n-k} {}_nC_k \quad \mu = n\,p \quad \sigma = n\,p\,(1-p)$$

The command to compute ${}_nC_k$ in **R** is `choose(n,r)`. For example:

```
>choose(8,5)
```

In the example below, with 8 question multiple choice test with straight guessing, probability of getting one correct answer is

$$P(1) = (1/2)^1(1/2)^{8-1}{}_8C_1 = (1/2)^8 \cdot 8 = 8/256 = 0.03125$$

Using R to create Binomial Distributions

R can easily produce binomial random numbers. We can then *simulate* various experiments easily on the computer.

For example, lets consider a True/False test with 8 questions. If a student simply guesses at each question, the number of correct answers on the test will be a binomial random number. The number of Bernoulli trials is $n = 8$, the probability of getting a correct answer for this student is $p = 1/2$ and getting it wrong is $q = 1/2$. To simulate 5 such students taking the test, use the **R** command:

```
> rbinom(5,8,.5)
[1] 3 4 4 2 5
```

The results show that the first student got 3 correct answers, the second two got 4 correct answers, the fourth student got only 2 correct answers and the fifth (lucky) student got 5 of the 8 questions correct. **Note that since the command generates random numbers, you may get different numbers from above**. In general we can use the following command `rbinom(s,n,p)`, where we substitute `s`=number of students, `n`=number of questions (trials) and `p`=probability of the correct answer. For example, we can easily simulate 1000 students taking this test (and answering randomly).

```
> testdata = rbinom(1000,8,0.5)
```

Lets look at the relative frequency histogram of `testdata` with the rectangles centered at number of correct answers.

```
> hist(testdata, prob=T, breaks=c(-.5,.5,1.5,2.5,3.5,4.5,5.5,6.5,7.5,8.5))
```

Lets compare the numerical value of $P(1)$ to the theoretical value computed above.

```
sum(testdata==1)/1000
```

Multiple Choice Test : Write the answers to the questions below on the next page in the space allocated.

Consider a multiple choice test with 20 questions, 4 possible choices for each question and only one correct answer.

1. What is n, p and q in this case ?
2. Use **R** to simulate the number of correct answers for 500 students taking this test if every student simply guesses at each question.
3. If you need 10 correct answers to pass this test, for your sample of 500 students, how many passed the test?
4. Plot and print a relative frequency histogram of the simulated data with your name on it.
5. Compare the simulated value of P(10) to the theoretical value of P(10).
6. Find the mean of your data. Compare this to the theoretical mean for the binomial distribution. Are they close?
7. If a Professor gives the test to a class of 100 students and finds that 75% of the students passed, explain, using statistics and probability, why the Professor can conclude that the student's were not simply guessing at the answers.

Lab Project 3

Please write your name, fill in the values, tear off and hand to instructor.

Name: ______________________________

Multiple Choice Test

1	$n =$ $\quad\quad$ $p =$ $\quad\quad$ $q =$	
3	____________ students passed test	
5	Simulated value $P(10) =$	Theoretical value $P(10) =$
6	Simulated mean $=$	Theoretical mean $=$

Hand in the histogram for Question 4.

Answer to 7: