Statistics 341 Name: Solutions
Exam #2
October 17, 2008

1. (6 pts) The negative binomial distribution has the following probability distribution function.

$$p(y) = \binom{y-1}{r-1} p^r (1-p)^{y-r} \qquad y = r, r+1, r+2, \ldots$$

Explain how the probability distribution function for the negative binomial distribution was derived.

The random variable Y is the number of trial on which the rth success occurs. If the yth success is on the rth trial, $y-1$ successes had to have occurred on the previous $r-1$ trials. So we have the statement

$$P(Y = y) = P((y-1 \text{ successes on first } r-1 \text{ trials}) \cap (\text{success on trial } y))$$

For the negative binomial distribution, the trials are independent, and so the probability above becomes.

$$P(Y = y) = P(y-1 \text{ successes on first } r-1 \text{ trials}) P(\text{success on trial } y)$$

The probability of a success on the yth trial = p and the probability of $y-1$ successes on the first $r-1$ trials is a binomial probability with $n = r-1$ and p. So we have

$$P(Y = y) = \binom{r-1}{y-1} p^{y-1} (1-p)^{r-1-(y-1)} p = \binom{r-1}{y-1} p^y (1-p)^{r-y}$$

2. (13 pts) Five cards are dealt without replacement from a deck of 52 cards (4 suits, 13 values per suit). Let the random variable Y be the number of hearts in the five card hand.

   (4 pts) The random variable Y has a hypergeometric distribution with $r = 13$, $N = 52$, $n = 5$.

   (a) (3 pts) What is the probability the five card hand will contain 3 or more hearts?
   This is the probability $P(Y \geq 3)$, which can be calculated in R using the formula

   `sum(dhyper(3:5, 13, 39, 5)) = 0.0928.`

   (b) (3 pts) Find the expected value of the number of hearts in the five card hand.
   For a hypergeometric distribution, $E(Y) = n\left(\frac{r}{N}\right) = 5\left(\frac{13}{52}\right) = 1.25$

   (c) (3 pts) Find the variance of the number of hearts in the five card hand.
   For a hypergeometric distribution,

$$V(Y) = n\left(\frac{r}{N}\right)\left(\frac{N-r}{N}\right)\left(\frac{N-n}{N-1}\right) = 5\left(\frac{13}{52}\right)\left(\frac{39}{52}\right)\left(\frac{47}{51}\right) = 0.8640$$

3. (12 pts) Roll two 10 sided dice (sides numbered 0 through 9). Let the random variable Y be the minimum value of the two dice. The probability distribution function of Y is

y	0	1	2	3	4	5	6	7	8	9
$p(y)$	19/100	17/100	15/100	13/100	11/100	9/100	7/100	5/100	3/100	1/100

(a) (5 pts) Find the expected value of the random variable Y.
The expected value is

$$
\begin{aligned}
E(Y) &= \sum_{y=0}^{9} y p(y) \\
&= \frac{0(19) + 1(17) + 2(15) + 3(13) + 4(11) + 5(9) + 6(7) + 7(5) + 8(3) + 9(1)}{100} \\
&= 2.85
\end{aligned}
$$

(b) (7 pts) Find the variance of the random variable Y.
First you will need to find $E(Y^2)$ and then use it for the variance calculation.

$$
\begin{aligned}
E(Y^2) &= \sum_{y=0}^{9} y^2 p(y) \\
&= \frac{0^2(19) + 1^2(17) + 2^2(15) + 3^2(13) + 4^2(11) + 5^2(9) + 6^2(7) + 7^2(5) + 8^2(3) + 9^2(1)}{100} \\
&= 13.65
\end{aligned}
$$

$$
\begin{aligned}
V(Y) &= E(Y^2) - (E(Y))^2 \\
&= 13.65 - (2.85)^2 \\
&= 5.5275
\end{aligned}
$$

4. (10 pts) In 2006, cyclist Floyd Landis won the sports premier event, the Tour de France. After completing the race, the World Anit-Doping Agency released test results which indicated that Landis had high testosterone levels in his urine after one of the race stages. After further litigation, the International Court of Arbitration for Sport upheld the test results, stripped Landis of the Tour de France win, and banned him from professional cycling for two years.

   During the course of the 2006 Tour de France, Landis submitted 8 urine specimens for analysis. The test results on these 8 specimens depend on the accuracy of the testing procedures. For this problem, assume Landis was NOT doping and that each test result is independent of the others. Let the probability of a correct no-doping test result given the assumption that Landis was not doping to be 0.95. Let the random variable Y be the number of correct no-doping test result out of the 8 tests Landis submitted during the 2006 Tour de France.

   (3 pts) The random variable Y has a binomial distribution with $n = 8$ and $p = 0.95$.

   (a) (3 pts) What is the probability that seven or more of the tests would have the correct no-doping test result?
   This is the probability $P(Y \geq 7)$. You can find this probability in R with the formula

   ```
   sum(dbinom(7:8, 8, 0.95)) = 0.9428
   ```

   (b) (4 pts) Describe the shape of the distribution of the random variable Y. In your answer, describe what mechanism you used to investigate the shape of the distribution.

   The distribution is skewed to the left. There are several ways you could have determined the shape of the distribution. One is by looking at the probabilities in the problem above. Almost all of the probability for this random variable is at 7 or 8, which are the two largest possible values for the random variable. Clearly, the probability of the remaining values 0 through 6 will be very small, indicating a skewed to the left distribution. You could have generated many values from this distribution using the rbinom command and then graphed a histogram of these values. This distribution would have been skewed to the left. Finally, you could have plotted the values of y versus the corresponding probabilities in dbinom. For y between 0 and 8, the probabilities would have started very small, and then increased to the largest probabilities at $y = 8$.

5. (9 pts) Flood insurance is determined by the federal government through the development of flood plains. One of the flood plains is called a 100 year flood plain. The means that the chance of experiencing a significant flooding event in a given year is 0.01. Assume that significant flooding events between years are independent and let the random variable Y be the number of years to experience the first significant flood.

   (3 pts) The random variable Y has a negative binomial distribution with $r = 1$ and $p = 0.01$.

   (a) (3 pts) Find the expected number of years to experience the first significant flood.
   $E(Y) = \frac{r}{p} = \frac{1}{0.01} = 100.$

   (b) (3 pts) Find the variance of the number of years to experience the first significant flood.
   $V(Y) = \frac{r(1-p)}{p^2} = \frac{0.99}{0.01^2} = 9900$

6. (10 pts) In the state of Illinois, there is a lottery game called Little Lotto. The Little Lotto game is based on the selection of five balls without replacement from a bin of 39 balls. Each ticket costs \$1, and several types of tickets will win money. Here are the types of tickets and the amount of winnings for each type of ticket.

Type of Ticket	Winnings
Matching all 5 numbers	\$100,000
Matching 4 out of 5 numbers	\$100
Matching 3 out of 5 numbers	\$10
Matching 2 out of 5 numbers	\$1

Determine the probability distribution function for the random variable Y, which is the winnings in playing the Little Lotto.

There are 5 possible winning values in playing the Little Lotto (\$100,000, \$100, \$10, \$1, \$0). The probability of obtaining the values is given by the hypergeometric distribution with $r = 5$, $N = 39$ and $n = 5$. To win \$100,000, you will match 5 numbers which has probability dhyper(5, 5, 34, 5). To win \$100, you will match 4 out of the 5 numbers, which has probability dhyper(4,5,34,5). To win \$10, you will match 3 out of the 5 numbers, which has probability dhyper(3,5,34,5) and to win \$1, you will match 2 out of the 5 numbers, which has probability dhyper(2,5,34,5). If you match 0 or 1 numbers, you will not win anything and the probability of this occurrence is sum(dhyper(0:1, 5, 34, 5)).