University of Waterloo
Faculty of Mathematics

Centre for Education in
Mathematics and Computing

Senior Math Circles
March 10, 2010
Distribution and Probability III

Today we are going to investigate certain distributions. Warning - the typical distribution definitions are slightly changed to better suit the audience.

Binomial Distribution

Recall:
In how many ways can we arrange the letters of the word AABBA? $\frac{5!}{3!2!}$
Why? Because we don't count the different arrangements of the As and Bs.

If events A and B are independent then $Pr(A)Pr(B) = Pr(AB)$.

Two events A and B are mutually exclusive if $Pr(AB) = 0$.

Example:
Coin tosses are independent. What is the probability we get a head and a tail on two flips of a coin?
$Pr(HT \text{ or } TH) = Pr(HT) + Pr(TH) = Pr(H)Pr(T) + Pr(T)Pr(H) = \frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{2} = \frac{1}{2}$.

Hence if we flip a coin 3 times, what is the probability we get heads twice?
First, there are $\binom{3}{2} = \frac{3!}{2!1!}$ arrangemtns of HHT. The probability of HHT is $\left(\frac{1}{2}\right)^3$.
Putting these together gives $\binom{3}{2}\left(\frac{1}{2}\right)^3 = \frac{3}{8}$.

The binomial distribution is a generalized version of this situation:

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

where x is the number of successes, p is the probability of success, and n is the number of trials.

Example:
What is the probability that you get exactly 2 questions right in a multiple choice test consisting of 4 questions and 5 answer choices (each) if you guess at each question?

$f(x) = \binom{n}{x} p^x (1-p)^{n-x}$; where $n = 4, x = 2, p = \frac{1}{5}$
$f(2) = \binom{4}{2} \left(\frac{1}{5}\right)^2 \left(1 - \frac{1}{5}\right)^2$
$f(2) = 15.36\%$

Binomial experiments are those which involve:
T - two outcomes
I - independent trials
M - multiple trials
S - same probability of success

Expectation

The expectation of X is defined as $E(X) = \sum_x xPr(X = x)$.

In this case, the binomial, $E(X) = \sum_x xPr(X = x) = \sum_x x \binom{n}{x} p^x (1-p)^{n-x} = np$.
Why? This makes more sense by considering a simple example. How many heads would you expect if I flipped a fair coin 10 times?
Answer: 5. How did you get it? You took the probability, $\frac{1}{2}$ and multiplied by the number of trials, 10.

Negative Binomial Distribution

Consider a binomial experiment in which the last trial is a success.

In such a case we have $n - 1$ trials and $x - 1$ successes that can occur in any order. The last one must be a success.
Hence the probability function is:

$$f(x) = \binom{n-1}{x-1} p^{x-1} (1-p)^{n-x} \times p = \binom{n-1}{x-1} p^x (1-p)^{n-x}$$

where $n = k + x$ and k is the number of failures.

Example:
The world series (baseball) ends when one team wins 4 games. If two teams are equally matched and games are independent, what is the probability that the world series ends in 7 games?

$f(x) = \binom{n-1}{x-1} p^x (1-p)^{n-x}$
$f(4) = \binom{7-1}{4-1} \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^{7-4}$
$f(4) = \binom{7-1}{4-1} \left(\frac{1}{2}\right)^7$

$f(4) = 15.6\%$

Hypergeometric Distribution

Suppose we relax the independence assumption of a binomial. One way we might do this is to perform sampling without replacement instead of sampling with replacement.

If we continue to have two outcomes over multiple trials then the hypergeometric distribution comes into play. The probability function is:

$$f(x) = \frac{\binom{n}{x}\binom{m}{y}}{\binom{n+m}{x+y}}$$

In this case we select $x + y$ objects from a set of $n + m$ objects, where we select x objects from type N and y objects from type M.

Example:
What is the probability that we draw 3 aces in a 4 card hand?

$$f(x) = \frac{\binom{n}{x}\binom{m}{y}}{\binom{n+m}{x+y}}$$

$$f(3) = \frac{\binom{4}{3}\binom{48}{1}}{\binom{52}{4}}$$

$f(3) = 0.071\%$

Here we are choosing 3 aces from a total of 4 in a deck, and 1 other card from the remaining 48 cards.

Poisson Distribution

Experiments/processes involving events that occur over time and/or space independently at a constant rate such that no two can occur at the same time are considered to have a poisson distribution.

The probability function in this case is:

$$f(x) = \frac{e^{-\mu}\mu^x}{x!}$$

Where x is the number of events and μ is the average number of events in a particular period of time.

Examples:
Assuming that people arriving at a line up arrive according to a poisson process on average 4 in a minute, what is the probability that in a mintue we see 5 people?

$f(x) = \frac{e^{-\mu}\mu^x}{x!}$
$f(5) = \frac{e^{-4}4^5}{5!}$
$f(5) = 15.6\%$
Note: μ must have the same units as the unit of time of interest.

Assuming that people arriving at a line up arrive according to a poisson process on average 4 in a minute , what is the probability that in <u>2 minutes</u> we see 5 people?
$f(5) = \frac{e^{-8}8^5}{5!}$
$f(5) = 9.2\%$

Binomial vs. Poisson
The main difference between a binomial and poisson is that:

a) a binomial has an n which is known and fixed.

b) in a binomial we can count both failures and successes, but in a poisson we can only count events.

Problem Set

1. The geometric probability function is a special case of the negative binomial which occurs when the first success is on the last trial. Show the probability function of a geometric is given by: $f(n) = p(1-p)^{n-1}$ where n is from 1 to infinity.

2. In a multiple choice test consisting of 4 questions and 5 answer choices (each), what is the probability that I pass (where a pass is 50%)?

3. Earthquakes recorded in Ontario each year follow a poisson process with an average of 6 per year. Find the probability that 7 earthquakes will be recorded in a 2 year period?

4. At a nuclear power station an average of 8 leaks of heavy water are reported per year. Find the probability of 2 or more leaks in 1 month if leaks follow a poisson process.

5. Show that the geometric probability function in (1) is a probability function. (ie. show that it sums to 1 and is always positive)

6. The discrete uniform distribution is defined by the probability function, $f(x) = \dfrac{1}{b-a+1}$ where $x = a, a+1, ..., b-1, b$.

   (a) Show that $f(x)$ satisfies the criterion of a probability function. (see 5)

   (b) What is $E(X)$? (Hint: What is the sum of the digits from 1 to n?)

   (c) What is $E(X^2)$? (Hint: What is the sum of the digits from 1^2 to n^2?)

7. In Lotto 6/49 a player selects a set of 6 numbers (with no repeats) from the set {1,2,...,49}. In the lottery draw 6 numbers are selected at random. Find the probability function for X, which represents how many of your selected numbers that are drawn.

8. In a crate there are 12 cans. Of the twelve, 3 have dents. What is the probability that in randomly selecting 4 cans, 2 of your cans have dents?

9. 200 people are at a party. What is the probability that 2 of them were born on Jan 2?

Answers

1. The probability function of a negative binomial is: $f(x) = \binom{n-1}{x-1} p^x (1-p)^{n-x}$

   We only have 1 success, $x = 1$: $f(n) = \binom{n-1}{1-1} p^1 (1-p)^{n-1}$

   Since $\binom{n-1}{0} = 1$, we get: $f(n) = p(1-p)^{n-1}$

2. $Pr(X \geq 2) = f(2) + f(3) + f(4)$

   $Pr(X \geq 2) = \binom{4}{2}\left(\frac{1}{5}\right)^2\left(1-\frac{1}{5}\right)^2 + \binom{4}{3}\left(\frac{1}{5}\right)^3\left(1-\frac{1}{5}\right)^1 + \binom{4}{4}\left(\frac{1}{5}\right)^4\left(1-\frac{1}{5}\right)^0$

   $Pr(X \geq 2) = 18.08\%$

3. $f(7) = \dfrac{e^{-12}12^7}{7!}$

4. $Pr(X \geq 2) = 1 - Pr(X < 2) = 1 - \dfrac{\left(\frac{8}{12}\right)^0 e^{-\frac{8}{12}}}{0!} - \dfrac{\left(\frac{8}{12}\right)^1 e^{-\frac{8}{12}}}{1!} = 14.43\%$

5. $\displaystyle\sum_n p(1-p)^{n-1} = \frac{p}{1-(1-p)} = 1$

6. (a) $\displaystyle\sum_{x=a}^{b} \frac{1}{b-a+1} = \frac{1}{b-a+1}\sum_{x=a}^{b} 1 = \frac{b-a=1}{b-a+1} = 1$

   (b) $\displaystyle E(X) = \sum_{x=a}^{b} \frac{x}{b-a+1}$

   $\displaystyle E(X) = \frac{1}{b-a+1}\sum_{x=a}^{b} x$

   $\displaystyle E(X) = \frac{1}{b-a+1}\sum_{i=1}^{b-a+1} a+i-1$

   $\displaystyle E(X) = \frac{1}{b-a+1}\left(a(b-a+1) - (b-a+1) + \sum_{i=1}^{b-a+1} i\right)$

   $\displaystyle E(X) = \left(a - 1 + \frac{1}{b-a+1}\sum_{i=1}^{b-a+1} i\right)$

   $\displaystyle E(X) = \left(a - 1 + \frac{1}{b-a+1}\left(\frac{(b-a+1)(b-a+2)}{2}\right)\right)$

   $\displaystyle E(X) = \left(a - 1 + \frac{(b-a+2)}{2}\right)$

   $\displaystyle E(X) = \left(\frac{a+b}{2}\right)$

(c) $E(X) = \sum_{x=a}^{b} \frac{x^2}{b-a+1}$

$$E(X) = \frac{1}{b-a+1} \sum_{x=a}^{b} x^2$$

$$E(X) = \frac{1}{b-a+1} \left(\sum_{x=1}^{b} x^2 - \sum_{x=1}^{a-1} x^2 \right)$$

$$E(X) = \frac{1}{b-a+1} \left(\frac{b(b+1)(2b+1)}{6} - \frac{a(a+1)(2a+1)}{6} \right)$$

7. Think of your number as the S objects and the remainder as F objects. Then X has a hypergeometric distribution with $N = 49$ and $r = 6$ and $n = 6$ so:

$$Pr(X = x) = \frac{\binom{6}{x}\binom{43}{6-x}}{\binom{49}{6}} \text{ where } x = 0, 1, ..., 6$$

For example, you win a prize if you match all 6 numbers. The chance of that is $\frac{\binom{6}{6}}{\binom{49}{6}}$ or about 1 in 13.9 million.

8. $Pr(X = 2) = \frac{\binom{9}{2}\binom{3}{2}}{\binom{12}{4}} = 21\%$

9. This is binomial with $p = \frac{1}{365}$

Hence, $f(2) = \binom{200}{2} \left(\frac{1}{365}\right)^2 \left(1 - \frac{1}{365}\right)^{198} = 8.7\%$