STAT 100 Lecture 10: Bernoulli Trials and the Binomial Distribution

Nate Strawn

http://www.math.umd.edu/∼nstrawn/

Last Time...

1. Expected Value of a Random Variable

$$\mu = E(X) = \sum x_i f(x_i)$$

2. Variance and Standard Deviation of a Random Variable
   - $\sigma^2 = E(X^2) - E(X)^2 = \sum x_i^2 f(x_i) - \left(\sum x_i f(x_i)\right)^2$
   - $\sigma = \sqrt{\sigma^2}$

Today's Agenda

1. Bernoulli Trials (Section 5.5)
2. The Binomial Distribution (Section 5.6)
3. Examples

Successes and Failures

Often, only two outcomes can be attributed to a process. Here are just a few examples:

- An industrial process produces processors that can be either usable or defective.
- Polled individuals can be either for or against the death penalty.
- STD tests look for the presence or absence of antibodies in blood.
- Can you name a few?

Daniel Bernoulli

In 1776, Swiss mathematician Daniel Bernoulli published his analysis of the mortality of smallpox[1].

[1]Blower, S. (2004), D, Bernoulli's "An attempt at a new analysis of the mortality caused by smallpox and of the advantages of inoculation to prevent it", Reviews of Medical Virolology, 14: 275-288

Bernoulli Trials

Definition

A process is called a Bernoulli Trial if

1. *The process has two outcomes: Success (S) or Failure(S).*
2. *For each trial,* $p = P(S)$ *is the probability of Success and* $q = 1 - p = P(F)$ *is the probability of Failure. Both* p *and* q *are fixed for all time.*
3. *Trials are independent of each other.*

Given a rachet set with three bits that are pretty close in size, you'll randomly choose one everytime you start a new job. You only have to whip out this set once every year (holidays, moves, etc) and you're only concerned with the first try. Either the bit fit the bolt or it doesn't. Is this process a Bernoulli Trial? Why or why not?

Suppose someone picks out 5 movies. Of the movies, 3 are good, but 2 are awful. We pick one to watch at random and determine whether the movie is bad or good. Then we pick another one at random and watch it to see if it is good or bad. We continue these trials. Of course, we never view the same movie again. Are these movie screenings Bernoulli Trials? Why or why not?

What if the movies are chosen from Netflix's collection of 70,000 titles and we suppose that 30,000 of them are awful?

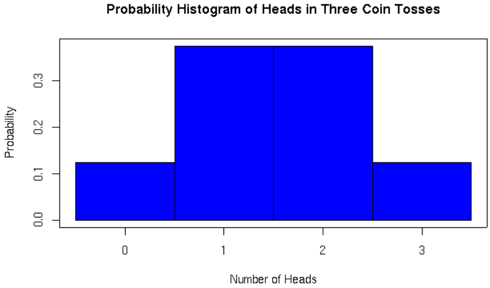
Approximate Bernoulli Trials

Rule of Thumb

If one randomly samples a small proportion of a large, dichotomous population without replacement, then the trials behave very much like Bernoulli Trials. In general, this should work if less than 10 percent of the population is sampled.

Coin Tosses as Bernoulli Trials

Let's label Heads as a Success and Tails as a Failure. Then, X = Number of Heads = Number of Successes is a random variable on the sample space consisting of the outcomes from three coin tosses. This means that we have already computed the probability distribution of the number of successes arising from 3 Bernoulli trials with $p = 1/2$.



The Distribution of n Bernoulli Trials

Now suppose we run n Bernoulli Trials, each with probability of success p. If $X =$ Number of Successes, then the probability distribution of X can be computed by reasoning as follows:

1. Choose x many of the n trials to be successes.
2. Use independence of the trials to compute the probability of this particular choice.
3. Note that there are 2^n possible success/failure outcomes.

The Binomial Distribution

Definition

The Binomial Distribution of the parameters n and p (n is the number of trials and p is the success rate) is described by the function (for $0 \leq x \leq n$)

$$f(x) = P[X = x] = \binom{n}{x} p^x (1-p)^{n-x} = \binom{n}{x} p^x q^{n-x}.$$

Formula

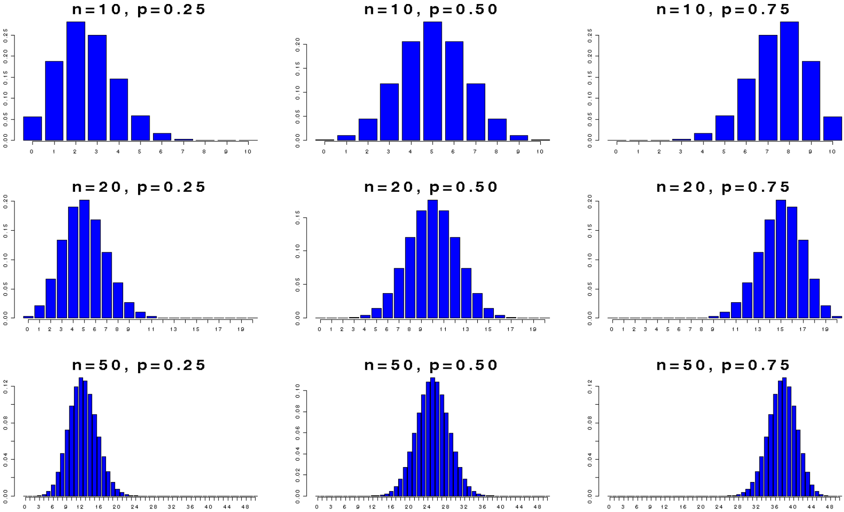
The Binomial Distribution with n trials and success rate p satisfies

$$\mu = np$$

and

$$\sigma^2 = npq.$$

The Binomial Distribution for Different n and p



The offspring of two people with heterozygous resitance alleles will be resistant to malaria with probability 0.75. Suppose these two people have five offspring.

1. What is the probability that all of the offspring are resistant to malaria?
2. What is the expected number of offspring with resistance to malaria?
3. For this distribution, determine the standard deviation.

Using Binomial Tables

First, locate the Binomial Table corresponding to the n parameter of the Binomial Distribution. Then, use the top row to locate the correct column for the p parameter. The c^{th} entry of that column corresponds to the quantity $P[X \leq c]$. Thus, you can compute $P[X = c] = P[X \leq c] - P[X \leq c - 1]$. Below is the Binomal Table for $n = 5$.

	p										
c	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95
0	0.774	0.590	0.328	0.168	0.078	0.031	0.010	0.002	0.000	0.000	0.000
1	0.977	0.919	0.737	0.528	0.337	0.188	0.087	0.031	0.007	0.000	0.000
2	0.999	0.991	0.942	0.837	0.683	0.500	0.317	0.163	0.058	0.009	0.001
3	1.000	1.000	0.993	0.969	0.913	0.813	0.663	0.472	0.263	0.081	0.023
4	1.000	1.000	1.000	0.998	0.990	0.969	0.929	0.832	0.672	0.410	0.226
5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Example

1. *Using the* $n = 5$ *Binomal table, compute the probability of 4 successes in 5 trials with* $p = 0.60$
2. *Compute the probability of more than 3 successes in 5 trials if* $p = 0.8$.
3. *Compute the probability of more than 2 successes, but less than 4 successes in 5 trials if* $p = 0.30$.

	p										
c	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95
0	0.774	0.590	0.328	0.168	0.078	0.031	0.010	0.002	0.000	0.000	0.000
1	0.977	0.919	0.737	0.528	0.337	0.188	0.087	0.031	0.007	0.000	0.000
2	0.999	0.991	0.942	0.837	0.683	0.500	0.317	0.163	0.058	0.009	0.001
3	1.000	1.000	0.993	0.969	0.913	0.813	0.663	0.472	0.263	0.081	0.023
4	1.000	1.000	1.000	0.998	0.990	0.969	0.929	0.832	0.672	0.410	0.226
5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

For Next Time

- Read Section 5.5 and 5.6 from Johnson and Bhattacharyya
- Group Problems:

Group	1	2	3	4	5
Problem	5.45	5.47	5.53	5.55	5.57
Group	6	7	8	9	10
Problem	5.59	5.61	5.67	5.69	5.77