Binomial Distribution

1. **Factorial**

$n! = 1 \cdot 2 \cdot 3 \cdot 4 \cdots (n-1) \cdot n$
Special Case: $0! = 1$

Ex.) Find each value

(i) $5! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 120$

(ii) $25! = 1 \cdot 2 \cdot 3 \cdot 4 \cdots 24 \cdot 25 = 1.55 \times 10^{25}$

(iii) $\dfrac{6!}{4!2!} = \dfrac{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6}{4!2!} = \dfrac{4! \cdot 5 \cdot 6}{4!2!} = \dfrac{5 \cdot 6}{2} = 15$

2. **An Example: A Binomial Process in Biology**

Let us assume a population contains a dominant allele A and recessive allele a. Each reproductive cell contains exactly one of the two alleles, either a or A.

Let p represent the probability that a cell contains A and let $q = 1 - p$ be the probability of a.
Thus,
$$P(A) = p, \quad P(a) = q, \text{ and } \quad p + q = 1.$$

If we assume that the reproductive cells meet each other at random, then the process is *independent* of the genetic content of each cell. Therefore,

$$P(AA) = P(A) \cdot P(A) = p \cdot p = p^2$$
$$P(Aa) = P(A) \cdot P(a) = p \cdot q = pq$$
$$P(aa) = P(a) \cdot P(a) = q \cdot q = q^2$$

Note, that biologically speaking Aa is equivalent to aA. So, P(aA) $= pq$. The possible outcomes of the mating of two heterozygous individuals is

	A	a
A	AA	Aa
A	Aa	aa

For the four cases, the sum of the probabilities should equal one.

$$P(AA) + P(Aa) + P(aA) + P(aa) = 1$$
$$p^2 + pq + pq + q^2 = 1$$
$$p^2 + 2pq + q^2 = 1$$
$$(p + q)^2 = 1$$

Recall, $p + q = 1$. Thus, $(p + q)^2 = (1)^2 = 1$.

Now, let's extend this idea in a more general sense

3. **Bernoulli Trials**

By definition, a ***Bernoulli trial*** is a random experiment with exactly two possible outcomes, often labeled a "success" or "failure", in which the probability of success is the same every time the experiment is conducted.

Let p represent the probability of a success, "S," and let q represent the probability of a failure, "F."

Creating a tree diagram illustrates the probability of certain results given three successive trials:

How many ways can there be three consecutive successes?

1

The probability of three successes is

$$P(3\ S) = p^3$$

How many ways can there be two successes in three trials?

3

The probability of two successes is

$$P(2\ S) = 3p^2q$$

How many ways can there be only one success in three trials?

3

The probability of one success is

$$P(1\ S) = 3pq^2$$

How many ways can there be no success in three trials?

1

The probability of no successes is
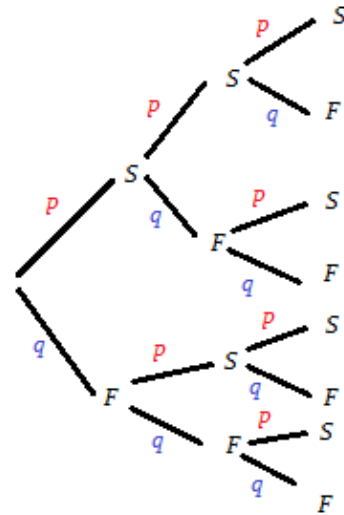
$$P(0\ S) = q^3$$

Taking the sum of the probabilities:

$$p^3 + 3p^2q + 3pq^2 + q^3 = 1$$

Factoring

$$p^3 + 3p^2q + 3pq^2 + q^3 = (p + q)^3 = (1)^3 = 1$$

Summarizing the cases,

Number of Trials	Number of Successes	Number of Ways	Probability	
1	1	1	p	
	0	1	q	$p + q = 1$
2	2	1	p^2	
	1	2	$2pq$	
	0	1	q^2	$(p + q)^2 = 1$
3	3	1	p^3	
	2	3	$3p^2q$	
	1	3	$3pq^2$	
	0	1	q^3	$(p + q)^3 = 1$

4. **Combinations**

The four probabilities listed above for the results of three Bernoulli trials are the terms in the expansion of $(p + q)^3$. The coefficients are the **_binomial coefficients_** $\binom{3}{0}, \binom{3}{1}, \binom{3}{2}, \binom{3}{3}$.

In general,

$$\binom{n}{k} = C(n, k) = \frac{n!}{(n-k)!\,k!}$$

$\binom{n}{k}$ is read as "n choose k" or the number of ways to select k objects out of n distinct objects.

Ex.) Find each value

(i) $\binom{5}{2} = \frac{5!}{(5-2)!2!} = \frac{5!}{3!2!} = \frac{5 \cdot 4 \cdot 3!}{3!2!} = \frac{5 \cdot 4}{2} = 10$

(ii) $\binom{8}{4} = \frac{8!}{(8-4)!4!} = \frac{8 \cdot 7 \cdot 6 \cdot 5 \cdot 4!}{4!4!} = \frac{8 \cdot 7 \cdot 6 \cdot 5}{4 \cdot 3 \cdot 2 \cdot 1} = 70$

5. **The Binomial Distribution**

The approach above, creating a table, works well. However, what if there are a large number of Bernoulli trials – say 50? There exists a formulation which provides the probability of any number of successes from any number of trials.

Let there exist n independent Bernoulli trials, where p is the probability of success (which remains constant for all trials), and $q = 1 - p$ is the probability of failure. Then, the probability of having k successes is then

$$f(k; n, p) = \frac{n!}{(n-k)!\,k!}\, p^k q^{n-k}.$$

Ex.) The probability of developing an illness from a recessive allele is 25% if both parents are heterozygotes. What is the probability that two of their three children have the illness?

$n = 3, k = 2, p = 0.25$

$f(k; n, p) = f(2; 3, 0.25)$

$$= \frac{3!}{(3-2)!\,2!}\,(0.25)^2 \cdot (0.75)^{3-2}$$

$$= 3 \cdot 0.0625 \cdot 0.75 = 0.140625 \approx 14.0625\%$$

Ex.) The probability that a person will die within a month after a certain cancer operation is 22%. What are the probabilities that in three such operations zero, one, two, or all three persons will survive?

If the probability of death is 22%, then the probability of survival is $1 - 0.22 = 0.78 = 78\%$. Let k represent the number of survivors.

$f(k; n, p) = f(0; 3, 0.78)$

$$= \frac{3!}{(3-0)!\, 0!}\ (0.78)^0 \cdot (0.22)^{3-0}$$

$$= 1 \cdot 1 \cdot 0.010648 = 0.010648 \approx 1.06\%$$

$f(k; n, p) = f(1; 3, 0.78)$

$$= \frac{3!}{(3-1)!\, 1!}\ (0.78)^1 \cdot (0.22)^{3-1}$$

$$= 3 \cdot 0.78 \cdot 0.0484 = 0.113256 = 11.33\%$$

$f(k; n, p) = f(2; 3, 0.78)$

$$= \frac{3!}{(3-2)!\, 2!}\ (0.78)^2 \cdot (0.22)^{3-2}$$

$$= 3 \cdot 0.6084 \cdot 0.22 = 0.401544 = 40.15\%$$

$f(k; n, p) = f(3; 3, 0.78)$

$$= \frac{3!}{(3-3)!\, 3!}\ (0.78)^3 \cdot (0.22)^{3-3}$$

$$= 1 \cdot 0.474552 \cdot 1 = 0.474552 = 47.46\%$$

The probability nobody survives is approximately 1.06%, exactly one person survives is approximately 11.33%, exactly two people survive is approximately 40.15%, and all three survive is approximately 47.46%.

Two useful technologies to assist in calculating the values include the calculator (here the process of using the TI-83 is detailed) as well as computer software (steps in EXCEL are detailed).

Using the TI-83

- MATH -> PRB
  (enter MATH and move cursor to the far right)
  Options
  4: !   Factorial
  3: nCr   Binomial Coefficient

- DISTR (2nd VARS)
  Options
  0: binompdf   Binomial Distribution

- Examples:
  5!   5 MATH PRB 4 enter ; results in 120
  C(8,5)   8 MATH PRB 3 5 enter ; results in 56
  f(2;3,0.25)   DISTR 0 3,0.25,2) enter ; results in .140625
                              binompdf(3,.25,2)

  On the calculator k is entered last. If k is omitted, all values of k from 0 to n will be displayed.

Using EXCEL 2010

- The command formula **BINOM.DIST** is used to calculate an individual term binomial distribution probability. Note, for EXCEL 2007 or earlier, you should use BINOMDIST (without the period).

- The syntax used in the formula (f_x) area is

  =**BINOM.DIST**(*number_s,trials,probability_s,cumulative*), where

    - *number_s* is the number of successes
    - *trials* is the number of trials
    - *probability_s* is the probability of success, and
    - *cumulative* is a logical value (TRUE or FALSE) that determines the form of the function. If cumulative is TRUE, then BINOM.DIST returns the cumulative distribution function, which is the probability that there are at most number_s successes; if FALSE, it returns the probability mass function, which is the probability that there are number_s successes.

In the previous example, the four probabilities would have been calculated using the following:
=BINOM.DIST(0,3,0.78,FALSE)
=BINOM.DIST(1,3,0.78,FALSE)
=BINOM.DIST(2,3,0.78,FALSE)
=BINOM.DIST(3,3,0.78,FALSE)

whose outputs would are
0.010648, 0.113256, 0.401544, and 0.474552, respectively. These values match those computed by hand in the example.

Ex.) An experimental medical procedure is 86% successful. What is the probability that it fails on three of twelve patients in a hospital?

If the probability of success is 86%, then the probability of failure is $1 - 0.86 = 0.14 = 14\%$. Let k represent the number of fails.

$$f(k; n, p) = f(3; 12,0.14)$$
$$= \frac{12!}{(12 - 3)!\, 3!}\, (0.14)^3 \cdot (0.86)^{12-3}$$
$$\approx 220 \cdot 0.002744 \cdot 0.2573274173 \ldots = 0.15534 \ldots \approx 15.53\%$$

An alternative is to determine the probability of successes if there are 3 failures in the 12 trials. This would imply there are $12 - 3 = 9$ successes. The probability of success on 9 of the 12 patients (here k is the number of successes) is

$$f(k; n, p) = f(9; 12,0.86)$$

$$= \frac{12!}{(12 - 9)!\, 9!}\, (0.86)^9 \cdot (0.14)^{12-9}$$
$$\approx 220 \cdot 0.2573274173 \cdot 0.002744 \ldots = 0.15534 \ldots \approx 15.53\%$$

Ex.) Given that 4% of the population carries the recessive CF allele. What is the probability that in a group of 50 people (a) nobody, (b) one person, (c) two people, and (d) three or more people carry the allele?

Let k represent the number of people that carry the allele.

(a) Nobody carries the allele $\Rightarrow k = 0$
$$f(k; n, p) = f(0; 50,0.04)$$
$$= \frac{50!}{(50 - 0)!\, 0!}\, (0.04)^0 \cdot (0.96)^{50-0}$$
$$\approx 1 \cdot 1 \cdot 0.1298857 \ldots \approx 12.99\%$$

(b) One person carries the allele $\Rightarrow k = 1$
$$f(k; n, p) = f(1; 50,0.04)$$
$$= \frac{50!}{(50 - 1)!\, 1!}\, (0.04)^1 \cdot (0.96)^{50-1}$$
$$\approx 50 \cdot 0.04 \cdot 0.1352977016 \ldots = 0.27059 \ldots \approx 27.06\%$$

(c) Two people carry the allele $\Rightarrow k = 2$
$$f(k; n, p) = f(2; 50,0.04)$$
$$= \frac{50!}{(50 - 2)!\, 2!}\, (0.04)^2 \cdot (0.96)^{50-2}$$
$$\approx 1225 \cdot 0.0016 \cdot 0.1409351058 \ldots = 0.27623 \ldots \approx 27.62\%$$

(d) Three or more people carry the allele $\Rightarrow k \geq 3$
This implies *at least* 3 people have the disease. You could find the probabilities for all values of k from 3 to 50 and sum up the results or use the complement of *at least* 3 which is *no more than* 2 (so, $k = 0$, $k = 1$, and $k = 2$).

From parts (a)-(c) those probabilities were found. Summing their results gives the solution to part (d):
$$P(at\ least\ 3) = 1 - \big(P(0) + P(1) + P(2)\big)$$
$$\approx 1 - (.1299 + .2706 + .2762) = 1 - (0.6767) = 0.3233 = 32.33\%$$

Ex.) It has been determined that the chance of becoming infected by a disease after contact with a particular parasite is 15%. If 50 individuals were in contact with the parasite, then how many individuals tested positive for the disease if the calculated binomial distribution is 3.28%.

In this example, the goal is to determine the number of individuals that have become infected with the disease; or, the number of successes, k, if a success is an infection.

This means you have to solve for k in the following formula,

$$f(k; n, p) = f(k; 50, 0.15)$$
$$= \frac{50!}{(50 - k)! \, k!} \, (0.15)^k \cdot (0.85)^{50-k} = 0.0328$$

and, quite frankly, this can be difficult.

However, using EXCEL can make the process much easier.

In the first column, make a list of the possible number of successes, $k = 0$ to $k = 50$. In the second column, calculate the binomial distribution (using BINOM.DIST) for each corresponding value of k.

Below is a small section of the output.

k	Distribution
0	0.000295765
1	0.002609688
2	0.011283064
3	0.031858062
4	0.066058629
5	0.107248127
6	0.141946051
7	0.157452762
8	0.149348576
9	0.122992945
10	0.088989013
11	0.057105249
12	0.03275154
13	0.016894459
14	0.007879349
15	0.003337136
16	0.001288233
17	0.00045467

Scanning down the distribution column, it appears when $k = 12$, the probability is approximately equal to 3.28%. Thus, you may concluded 12 individuals of the 50 became infected with the disease.