6.2: The Binomial Probability Distribution

Bernoulli trials:

A *Bernoulli trial* is an experiment with exactly two possible outcomes. We refer to one of the outcomes as a *success* (*S*) and to the other as a *failure* (*F*). We'll call the probability of success *p* and the probability of failure *q*. In other words,

$P(S) = p$ and $P(F) = 1 - p = q$. (Note that *S* and *F* are complements of one another.)

Example 1: Roll a single die. Consider rolling a 5 to be a success.

S: I roll a 5 Failure: I don't roll a 5

$p = P(S) = \frac{1}{6}$ $q = P(\text{Failure}) = \frac{5}{6}$

Example 2: Roll a pair of dice. Consider success to be rolling a sum of 7, 11, or 12.

Skip

Consider what happens when a Bernoulli trial is repeated several times. Now we can discuss the probability of a particular number of successes.

p = Prob of success
q = prob of failure } on 1 trial

Suppose we repeat a Bernoulli trial 6 times and each time the trial is independent of the others. From the multiplication principle, what is the probability of getting SFSSSF?

same!

$P(SFSSSF) = pqpppq = p^4 q^2$

What is the probability of getting SSSFFS?

$pppqqp = p^4 q^2$

What is the probability of having 4 successes in the 6 trials?

Combinations (this is in 5.5): $C_{n,x}$ is the number of ways we can choose x items from a collection of n items.

$C_{n,x} = \frac{n!}{x!\,(n-x)!}$

next page

There are $C_{6,4}$ ways to choose 4 of the 6 trials to be successful.

Prob of getting 4 success in 6 trials

is $\boxed{C_{6,4}\, p^4 q^2}$.

Example 3: If a single die is rolled 6 times, what is the probability of getting four 5's?

$p = \frac{1}{6}$, $q = \frac{5}{6}$

X = # of success (5's) in 6 trials.

$P(X=4) = C_{6,4}\left(\frac{1}{6}\right)^4\left(\frac{5}{6}\right)^2 = 15\left(\frac{1}{6}\right)^4\left(\frac{5}{6}\right)^2 \approx \boxed{0.0080376}$

$C_{6,4} = \frac{6!}{4!\,2!} = \frac{6\cdot5\cdot4\cdot3\cdot2\cdot1}{(4\cdot3\cdot2\cdot1)(2\cdot1)}$

Binomial experiments:

Now let's write all this down in a more precise way:

A sequence of experiments is called a sequence of Bernoulli trials, or a *binomial experiment*, if

1. The same experiment is repeated a fixed number of times. (Each repetition is called a trial.)
2. Only two outcomes are possible on each trial.
3. The probability of success p for each trial is a constant.
4. All trials are independent.

Probabilities in Bernoulli Trials:

If the probability of success is p and the probability of failure is q, then the probability of exactly x successes in n Bernoulli trials is

$$P(x \text{ successes}) = C_{n,x}\, p^x q^{n-x}.$$

The binomial formula:

For any natural number n,

$$(a+b)^n = C_{n,0}a^n b^0 + C_{n,1}a^{n-1}b^1 + C_{n,2}a^{n-2}b^2 + \ldots + C_{n,n}a^0 b^n$$
$$= \sum_{i=0}^{n} C_{n,i}a^{n-i}b^i$$

no need to know this!

Notice the similarity between this formula and the formula for probabilities in a sequence of Bernoulli trials. This is why such a sequence is called a *binomial experiment*.

Example 4 (c)

at least 1 success (in 10 trials)

$$P(x \geq 1) = P(x=1) + P(x=2) + P(x=3) + \dots + P(x=10)$$

Yuk! Use the complement:

$$P(x=0) = C_{10,0} \left(\frac{1}{6}\right)^0 \left(\frac{5}{6}\right)^{10}$$

$$= 1(1)\left(\frac{5}{6}\right)^{10} \approx 0.161506$$

$$P(x \geq 1) = 1 - P(x=0) = 1 - 0.161506 = \boxed{0.838494}$$

Example 4: If a single fair die is rolled 10 times, what is the probability of

a. Exactly three 4's?

b. At most three 4's?

c. At least one 4?

$n = 10$

Success: a 4 is rolled

Failure: a 4 is not rolled

$p = P(\text{success}) = \frac{1}{6}$

$q = P(\text{Failure}) = \frac{5}{6}$

(a) Exactly 3 successes

$P(X=3) = C_{10,3}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^7$

$= 120\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^7$

$\approx \boxed{0.15505}$

Part (c)

See ~~next~~ previous page

(b) at most 3 successes

$P(X \leq 3) = P(X=0) + P(X=1) + P(X=2) + P(X=3)$

$= C_{10,0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^{10} + C_{10,1}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^9 + C_{10,2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^8 + C_{10,3}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^7$

$= 1(1)\left(\frac{5}{6}\right)^{10} + 10\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^9 + 45\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^8 + 120\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^7$

$= 0.161506 + 0.323011 + 0.290710 + 0.155045$

$= \boxed{0.930272}$

Example 5: In the United States, about 9% of people have blood type B+. If 20 people donate blood, what is the probability that exactly three of them are B+? At least three?

Success: Blood type B+

X = # of people (out of 20) who have B+ blood

$P(\text{success}) = p = 0.09$ (on 1 trial)

$q = 1 - 0.09 = 0.91$

(a) Prob that exactly 3 are B+

$P(X=3) = C_{20,3}(0.09)^3(0.91)^{17}$

$= 1140(0.09)^3(0.91)^{17} \approx \boxed{0.1672}$

The binomial distribution:

We can now generalize Bernoulli trials and determine probability distributions.

In a binomial experiment with *n* trials and probability of success *p*, we can create a binomial distribution table and a histogram, with the variable *x* representing the number of successes.

Example 6: Suppose a fair die is rolled three times and a success is considered to be rolling a number divisible by 3.

On 1 trial: $S = \{1, 2, 3, 4, 5, 6\}$

$n = 3$

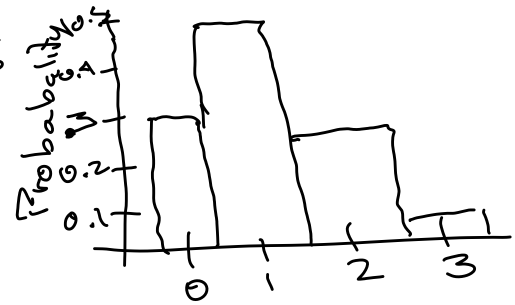
Success: 1 roll a 3 or 6

$E = \{3, 6\}$

$p = P(E) = P(\text{success}) = \frac{1}{3}$

$q = \frac{2}{3}$

x	$P(x)$	
0	$C_{3,0}\left(\frac{1}{3}\right)^0\left(\frac{2}{3}\right)^3 = 1(1)\left(\frac{2}{3}\right)^3 = \frac{8}{27}$	0.3
1	$C_{3,1}\left(\frac{1}{3}\right)^1\left(\frac{2}{3}\right)^2 = 3\left(\frac{1}{3}\right)\left(\frac{4}{9}\right) = \frac{4}{9}$	0.44
2	$C_{3,2}\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^1 = 3\left(\frac{1}{9}\right)\left(\frac{2}{3}\right) = \frac{2}{9}$	0.22
3	$C_{3,3}\left(\frac{1}{3}\right)^3\left(\frac{2}{3}\right)^0 = 1\left(\frac{1}{27}\right)(1) = \frac{1}{27}$	0.04



They should add up to 1: $8/27 + 4/9 + 2/9 + 1/27 = \frac{8}{27} + \frac{12}{27} + \frac{6}{27} + \frac{1}{27} = \frac{27}{27} = 1$

Mean of the binomial distribution:

We can determine the mean number of successes (or the *expected value* of *x*) for any binomial distribution.

> Mean (Expected Value) of the Binomial Distribution:
>
> In a binomial experiment with *n* trials and probability of success *p*, the expected value, or mean, is
>
> $$E(x) = \mu = np.$$
>
> Standard Deviation of the Binomial Distribution:
>
> $$\sigma = \sqrt{np(1-p)} = \sqrt{npq}.$$

Mean for Ex 6:

$\mu = E(x)$
$= 0\left(\frac{8}{27}\right) + 1\left(\frac{4}{9}\right) + 2\left(\frac{2}{9}\right) + 3\left(\frac{1}{27}\right)$
$= 0\left(\frac{8}{27}\right) + 1\left(\frac{12}{27}\right) + 2\left(\frac{6}{27}\right) + 3\left(\frac{1}{27}\right)$
$= 0 + \frac{12}{27} + \frac{12}{27} + \frac{3}{27}$
$= \frac{27}{27} = \boxed{1}$

Example 7: Roll a single fair die 10 times. Consider rolling a 4 to be a success. Find the expected number of successes.

$p = \frac{1}{6}$

$n = 10$

$E(x) = \mu = np = 10\left(\frac{1}{6}\right) = \frac{10}{6} = \frac{5}{3} = 1\frac{2}{3} \approx \boxed{1.67}$

Ex 6: Find mean using formula

$E(x) = \mu = np$
$= 3\left(\frac{1}{3}\right) = \boxed{1}$

1 trial = 1 patient
Success: effective treatment

Example 8: Suppose that a certain cancer treatment has been found to be effective in 63% of patients. If 300 patients undergo the treatment, in how many cases would you expect the treatment to be effective? What is the standard deviation of the variable representing the number of cases in which the treatment is effective?

$p = 0.63$

$q = 1 - 0.63 = 0.37$

$n = 300$

Mean:

$\mu = E(x) = np = 300(0.63) = \boxed{189}$

Std dev: $\sigma = \sqrt{npq} = \sqrt{300(0.63)(0.37)} = \sqrt{69.93} = \boxed{8.36}$

Example 9: An exam has 10 true-false questions. Suppose that you haven't studied all semester, so your only choice is to guess randomly on each question. If you need a 70% to pass, what is your probability of passing?

Each question is a Trial. So $n = 10$.

True/False, so probability of guessing correctly on 1 trial is 0.50.

Success: correct guess.

$p = \frac{1}{2}$, $q = \frac{1}{2}$

70% needed to pass ⇒ must get 7, 8, 9, or 10 questions correct.

$$P(x \geq 7) = P(x=7) + P(x=8) + P(x=9) + P(x=10)$$
$$= C_{10,7}\left(\tfrac{1}{2}\right)^7\left(\tfrac{1}{2}\right)^3 + C_{10,8}\left(\tfrac{1}{2}\right)^8\left(\tfrac{1}{2}\right)^2 + C_{10,9}\left(\tfrac{1}{2}\right)^9\left(\tfrac{1}{2}\right)^1 + C_{10,10}\left(\tfrac{1}{2}\right)^{10}\left(\tfrac{1}{2}\right)^0$$
$$= 120\left(\tfrac{1}{2}\right)^{10} + 45\left(\tfrac{1}{2}\right)^{10} + 10\left(\tfrac{1}{2}\right)^{10} + 1\left(\tfrac{1}{2}\right)^{10}$$
$$= \left(\tfrac{1}{2}\right)^{10}(120 + 45 + 10 + 1) = \frac{1}{2^{10}}(176) = \frac{176}{1024} \approx 0.171875$$
$$\approx \boxed{0.172}$$

(or, $(0.5)^{10}(176) = 0.171875$

Example 10: Suppose the test is multiple choice test with 5 choices on each of 10 questions. If you guess randomly at the answers, what is the probability of passing with at least a 70%? What grade would you expect to receive on the test?

$n = 10$. On each trial, the probability of success is $\frac{1}{5}$.

$p = \frac{1}{5} = 0.2$, $q = \frac{4}{5} = 0.8$

$$P(x \geq 7) = P(x=7) + P(x=8) + P(x=9) + P(x=10)$$
$$= C_{10,7}(0.2)^7(0.8)^3 + C_{10,8}(0.2)^8(0.8)^2 + C_{10,9}(0.2)^9(0.8)^1 + C_{10,10}(0.2)^{10}(0.8)^0$$
$$= 120(0.20)^7(0.8)^3 + 45(0.2)^8(0.8)^2 + 10(0.2)^9(0.8) + 1(0.2)^{10}(1)$$
$$= 7.86432\times10^{-4} + 7.3728\times10^{-5} + 4.096\times10^{-6} + 1.024\times10^{-7}$$
$$= 8.6436\times10^{-4} = \boxed{0.00086436}$$

Need a better study plan!

<u>Rule-of-Thumb Test</u> (for bell shape of the binomial distribution)

For a fixed probability of success *p*, as the number of trials (*n*) in a binomial experiment becomes very large, the probability distribution will approach a bell shape.

If $npq = np(1-p) \geq 10$, we can assume the distribution will be approximately bell-shaped.

Applying the Empirical Rule to decide whether a result is unusual:

From the Empirical Rule, we know that for a bell-shaped distribution, approximately 68% of the data points will lie within 1 standard deviation of the mean, about 95% will lie within 2 standard deviations of the mean, and about 99.7% will lie within 3 standard deviations of the mean.

If an outcome occurs less than 5% of the time, we consider that result to be *unusual*.

Therefore, for a bell-shaped distribution, an outcome is considered *unusual* if it does <u>not</u> lie within 2 standard deviations of the mean.

<u>Example 11:</u> Vancomycin-associated nephrotoxicity has been found to occur in 0-5% of patients treated with vancomycin (http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2813201/). Suppose there is a 0.05 probability that a patient treated with Vancomycin develops nephrotoxicity. Suppose that at a certain hospital, 253 patients are treated with Vancomycin in a year, and 28 of them develop nephrotoxicity. Would this result be considered unusual?