5.3: 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.

Success: a 5 is rolled
Failure: a 5 is not rolled

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

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

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?

$P(S) = p$
$P(F) = q = 1-p$

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

What is the probability of getting SSSFFS?

$P(SSSFFS) = pppqqp = p^4 q^2$

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

How many ways to choose 4 of the 6 trials to be successful? $C_{6,4}$ or ${}_6C_4$

$P(4 \text{ successes in } 6 \text{ trials}) = C_{6,4}\, p^4 q^2$. This is denoted as $P(X=4)$, where X is the number of successes.

on one trial: sample space: $S = \{1, 2, 3, 4, 5, 6\}$

$E = \{5\}$. $p = P(E) = \frac{1}{6}$

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

S: success: a 5 is rolled

F: failure: a 5 is not rolled

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

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

X = # of successes in 6 trials

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

$= 15\left(\frac{1}{6}\right)^4\left(\frac{5}{6}\right)^2$

$\approx \boxed{0.008038}$

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

$$P(X = x) = C_{n,x}\, p^x q^{n-x}, \text{ or equivalently, } P(X = x) = \binom{n}{x} p^x q^{n-x}.$$

The binomial formula:

no need to know this!

For any natural number n,

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

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: 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?

On 1 trial

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

$E = \{4\}$

$P(E) = \frac{1}{6}$

Success: a 4 is rolled on 1 trial

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

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

$n = 10$ (# of trials)

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

$q = \frac{5}{6}$

(a) Exactly three successes

$P(X=3) = C_{10,3}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^7 = 120\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^7 \approx 0.15504536 \approx \boxed{0.155}$

(b) At most 3 successes

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

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

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

$= 0.16151 + 0.32301 + 0.29071 + 0.15505$

$= 0.93028$

$\approx \boxed{0.9303}$

see next page for (c)

Often we need to select a random sample from a larger population. To do this, we would select each item/person in a separate trial, without replacement. (We do not want to select the same person twice). Because the selected individuals are not replaced, the population size decreases by 1 with each trial. For that reason, this process is not a binomial experiment. (The probability of success is not the same for all the trials.) Instead, the probabilities follow a *hypergeometric distribution*.

However, as long as the sample does not exceed 5% of the population, the distribution will be extremely close to the binomial distribution, and it is acceptable to use the binomial distribution as an approximation.

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

Success: person is B+

$p = P(\text{Success}) = 0.09$, $q = 1 - 0.09 = 0.91$

$n = 20$

(a) exactly three successes

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

$= 1140(0.09)^3(0.91)^{17}$

$\approx \boxed{0.1672}$

(b) at least 3 successes

E: $P(X \ge 3) = P(X=3) + P(X=4) + P(X=5) + \ldots P(X=20)$

Use complement:

E^c: $P(X=0) + P(X=1) + P(X=2) = P(X \le 2)$

$= C_{20,0}(0.09)^0(0.91)^{20} + C_{20,1}(0.09)^1(0.91)^{19} + C_{20,2}(0.09)^2(0.91)^{18}$

$= 1 \cdot 1(0.91)^{20} + 20(0.09)(0.91)^{19} + 190(0.09)^2(0.91)^{18}$

$= 0.15164 + 0.29996 + 0.28183 = 0.73343$

Subtract from 1 to get $P(E)$; $P(E) = P(X \ge 3) = 1 - 0.73343$

$= \boxed{0.26657}$

Example 4 (cont'd)

(c) At least one 4 (at least 1 success)

$$P(X \geq 1) = P(X=1) + P(X=2) + P(X=3) + \ldots + P(X=10)$$

Use the complement instead!

Complement: $P(X=0)$

$$P(X=0) = C_{10,0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^{10} = 1 \cdot 1 \left(\frac{5}{6}\right)^{10} = 0.16151$$

$$P(X \geq 1) = 1 - P(X=0) = 1 - 0.16151 = \boxed{0.8385}$$

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.

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

Mean and standard deviation of the binomial distribution:

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

<u>Mean (Expected Value) of the Binomial Distribution</u>:

In a binomial experiment with n trials and probability of success p, the expected value, or mean, is

$$E(x) = \mu = np.$$

<u>Standard Deviation of the Binomial Distribution</u>:

$$\sigma = \sqrt{np(1-p)} = \sqrt{npq}.$$

<u>Example 7:</u> Roll a single fair die 10 times. Consider rolling a 4 to be a success. Find the expected number of successes. (Find the mean, or expected value, of the number of successes)

$n = 10$

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

$q = \frac{5}{6}$

<u>Mean:</u> $\mu = E(x) = np = 10\left(\frac{1}{6}\right) = \frac{10}{6} = \frac{5}{3} = \boxed{1.67}$

<u>Std. deviation of X:</u> (X = # of successes)

$\sigma = \sqrt{npq} = \sqrt{10\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)} \approx \sqrt{1.389} \approx \boxed{1.179}$

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?

$n = 300$

Success: treatment is effective

$p = P(\text{Success}) = 0.63$

$q = 1 - p = 1 - 0.63 = 0.37$

X = number of patients for which treatment is effective

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

We would expect it to be effective in about 189 patients out of 300.

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

Note:

$\boxed{\text{Variance} = \sigma^2 = npq = 69.93}$

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?

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?