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,

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

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

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

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

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.

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

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?

If $A_1, A_2, A_3, A_4, \ldots$ are all independent, then $P(A_1 \cap A_2 \cap A_3 \cap A_4 \cap \ldots) = P(A_1)P(A_2)P(A_3)\ldots$

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

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?

"Combination n, x" is the number of ways in which a subset of x items can be selected from a collection of n items. It is denoted $C_{n,x}$ or $C(n,x)$, or ${}_nC_x$.

$C_{n,x} = \frac{n!}{x!(n-x)!}$. ! is a factorial: $6! = 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 720$

$C_{6,4} = \frac{6!}{4!\,2!} = \frac{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{(4 \cdot 3 \cdot 2 \cdot 1)(2 \cdot 1)} = \frac{6 \cdot 5 \cdot \cancel{4!}}{\cancel{4!}\,(2 \cdot 1)} = \frac{30}{2} = 15$

(The number of ways to select 4 of the 6 trials to be successful)

$P(4 \text{ successes}) = C_{6,4}\, p^4 q^2 = \boxed{15 p^4 q^2}$

Note: Here I rounded to 3 significant digits

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

Success: roll a 5 $p = P(\text{Success}) = \frac{1}{6}$, $q = 1 - p = \frac{5}{6}$

Prob of getting 4 successes is

$$P(X=4) = C_{6,4}\, p^4 q^2 = 15\left(\frac{1}{6}\right)^4\left(\frac{5}{6}\right)^2 \approx 0.00803755$$

$$\approx \boxed{0.00804}$$

X = number of successes

Note: If your calculator says 8.03755^{-3} or 8.03755E-3

This means 8.03755×10^{-3}

$= \boxed{0.00803755}$

(calculator is putting it in scientific notation to show more decimal places)

Here I rounded to 6 significant digits

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

$P(X=x) =$

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

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 (This is why we call this distribution the binomial distribution)

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?

Success: ~~A~~ 4 is rolled (on 1 trial)

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

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

$n = 10$

(a) Exactly 3 successes

$P(X=3) = C_{10,3}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^7 = 120\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^7 = \boxed{0.155045}$

(b) ~~At~~ At most 3 successes

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

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

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

$= 0.161506 + 0.323011 + 0.290710 + 0.155045$

$= \boxed{0.930272}$

(c) At least 1 success:

$P(X \ge 1) = P(X=1) + P(X=2) + P(X=3) + \ldots + P(X=10)$ Yuk! Use the complement!

Prob of Complement: $P(X=0) = C_{10,0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^{10} = 1(1)\left(\frac{5}{6}\right)^{10} = 0.161506$ (from (b))

so $P(X \ge 1) = 1 - 0.161506 = \boxed{0.838494}$

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: person has B^+ blood

$p = P(\text{Success}) = 0.09$

$q = 1 - p = 1 - 0.09 = 0.91$

$n = 20$ (each person is a trial)

exactly 3 successes

$P(X=3) = C_{20,3}(0.09)^3(0.91)^{17} = 1140(0.09)^3(0.91)^{17} \approx \boxed{0.1672}$

at least 3 successes

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

Use complement!

Complement: $P(X=0) + P(X=1) + P(X=2)$

See next page

Example 5 cont'd:

$$P(\text{Complement}) = P(X<3) = P(X=0) + P(X=1) + P(X=2)$$

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

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

$$= 0.151645 + 0.299957 + 0.281828$$

$$= 0.73343$$

$$P(X \geq 3) = 1 - P(X<3)$$

$$= 1 - 0.73343 = \boxed{0.26657}$$

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.

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:

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

So variance is

$\sigma^2 = npq$

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

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?

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

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

we would expect about 189 to be effective.

standard deviation: $\sigma = \sqrt{npq} = \sqrt{300(0.63)(0.37)} \approx \boxed{8.36}$

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?

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?

<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.

Success: patient get nephrotoxicity

<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?

$n = 253$

$p = 0.05$

$q = 0.95$

Need to find the standard deviation:

$$\sigma = \sqrt{npq} = \sqrt{253(0.05)(0.95)} = 3.46$$

<u>Note</u>: expected number of nephros. is

$$\mu = np = 253(0.05) = 12.65$$

So we would expect 95% to be within 2 std. devs:

$$\mu + 2\sigma = 12.65 + 2(3.46) = 19.57$$

$$\mu - 2\sigma = 12.65 - 2(3.46) = 5.73$$

A value of 28 is more than 2 std. devs above the mean ($28 > 19.57$), so yes, this result is unusual.