

Research on Mathematical Aspects of Binomial and Bernoulli Distribution

Author: Shin Lee

Western Reserve Academy

shin102766@gmail.com

DOI: 10.26821/IJSHRE.9.9.2021.9909

ABSTRACT

The Binomial and Bernoulli Distribution has special mathematical characteristics, many of which are already known. As distributions, they have characteristics and rules that enable them to be used in various circumstances and situations. This study aims to explore and discuss its mathematical features from the ground, analyzing their nature and proving characteristics.

1. Definition

The Bernoulli distribution is a discrete distribution having two possible outcomes labeled by $x = 1$ (success) with probability p and $x = 0$ (failure) with probability $q = 1 - p$, where $0 < p < 1$. If the probability of each outcome remains the same throughout the trials, then such trials are called Bernoulli trials. An experiment having n Bernoulli trials is called a Binomial experiment.

2. Expression

X denotes the number of successes in n trials of a Bernoulli experiment, and it is called a Binomial random variable with its probability distribution called the Binomial probability distribution. The random variable X can take any of the $n + 1$ integer values from 0 through n .

When the Binomial random variable X assumes a value x , then the Binomial probability distribution is given by:

$$F(x) = P(X = x) = \binom{n}{x} p^x q^{n-x}$$

where $x = 0, 1, 2, \dots, n$ and $q = 1 - p$.

When $n = 1$, the Binomial probability distribution becomes Bernoulli probability distribution. Therefore, the Bernoulli probability distribution is given by:

$$F(x) = P(X = x) = \binom{1}{x} p^x q^{1-x}$$

where $x = 0, 1$ and $q = 1 - p$.

3. Derivation of Binomial/Bernoulli Probability Distribution

The experiment has n trials, each of which results in S (success) or F (failure). The sample space has 2^n possible outcomes, each can be represented by a sequence a_1, a_2, \dots, a_n where a_i is either S or F . We aim to find the probability of these outcomes according to the number of successes.

To begin with, we are going to use the generic probability, $P(X = x)$. In this case, each possible outcome consists of exactly x successes with probability p and exactly $n - x$ failures with probability q , and thus $p^x q^{n-x}$.

There is no constraint on which of n trials are successes or failures, so that there exists $\binom{n}{x}$ different sequences of xS s and $n - xF$ s.

Therefore,

$$P(X = x) = \binom{n}{x} p^x q^{n-x} \text{ (Binomial distribution)}$$

$$P(X = x) = \binom{1}{x} p^x q^{1-x} \text{ (Bernoulli distribution).}$$

We have obtained the formula for the Binomial/Bernoulli probability distribution. The binomial probability distribution derives its name from the fact that the probabilities $\binom{n}{x} p^x q^{n-x}$ are the successive terms of the binomial expansion of $(p + q)^n$.

We are going to prove the connection between the binomial probability distribution and the binomial expansion in the following chapters.

4. Binomial/Bernoulli Theorem

We start from our conjecture: For all real numbers p and q , and positive integer n ,

$$(p + q)^n = \sum_{x=0}^n \binom{n}{x} p^x q^{n-x}$$

We can prove this by induction.

First of all, let us start with the base case when $n = 1$.

$$(p + q)^1 = p + q$$

$$\begin{aligned} \sum_{x=0}^1 \binom{1}{x} p^x q^{1-x} &= \binom{1}{0} p^0 q^{1-0} + \binom{1}{1} p^1 q^{1-1} \\ &= q + p = p + q \end{aligned}$$

Then, let us say that hypothetically

$$(p + q)^k = \sum_{x=0}^k \binom{k}{x} p^x q^{k-x}$$

is true for all values. Let us call this statement $P(k)$.

If we can prove that $P(k + 1)$ is also true, then the statement must be true for all values of k .

$$\begin{aligned} (p + q)^{k+1} &= (p + q)^k (p + q) \\ &= \sum_{x=0}^k \binom{k}{x} p^x q^{k-x} \cdot (p + q) \quad [P(k) \text{ used}] \end{aligned}$$

$$= \sum_{x=0}^k \binom{k}{x} p^{x+1} q^{k-x} + \sum_{x=0}^k \binom{k}{x} p^x q^{k+1-x}$$

$$= \sum_{x=1}^{k+1} \binom{k}{x-1} p^x q^{k+1-x} + \sum_{x=0}^k \binom{k}{x} p^x q^{k+1-x}$$

$$= \sum_{x=1}^k \binom{k}{x-1} p^x q^{k+1-x} + \binom{k}{k} p^{k+1} q^0$$

$$+ \sum_{x=1}^k \binom{k}{x} p^x q^{k+1-x}$$

$$+ \binom{k}{0} p^0 q^{k+1}$$

$$= p^{k+1} + \sum_{x=1}^k \left(\binom{k}{x-1} + \binom{k}{x} \right) p^x q^{k+1-x} + q^{k+1}$$

$$= p^{k+1} + \sum_{x=1}^k \binom{k+1}{x} p^x q^{k+1-x} + q^{k+1}$$

[Combinatorial Identity used]

4.1 Combinatorial Proof of Sub-identity

$$\binom{k+1}{x} = \binom{k}{x-1} + \binom{k}{x}$$

In the identity, $\binom{k+1}{x}$ represents the number of ways of choosing x people out of $k+1$ people. Let one of the $k+1$ people be A. When we look at different combinations of x people, some contain the person A, but the others do not. In order to count the number of combinations of x people including the person A, we only need $x-1$ more people to choose out of k people because the person A must be the one of these x people. This can be represented by $\binom{k}{x-1}$. In order to count the number of combinations of x people where person A is not one of them, we have to choose entire x people out of k people. This can be represented by $\binom{k}{x}$. Therefore, $\binom{k+1}{x}$ can be expressed as the sum of two mutually exclusive cases of $\binom{k}{x-1}$ and $\binom{k}{x}$.

$$= \sum_{x=0}^{k+1} \binom{k+1}{x} p^x q^{k+1-x}$$

Thus, the conjecture made in chapter 4 is true for all positive integer n .

5. Recursive Formula for the Bernoulli Distribution

$$P(X = x) = \frac{n-x+1}{x} \cdot \frac{p}{q} \cdot P(X = x-1)$$

5.1 Proof of Formula

$$P(X = x) = \binom{n}{x} p^x q^{n-x} = \frac{n!}{x!(n-x)!} p^x q^{n-x}$$

$$P(X = x-1) = \binom{n}{x-1} p^{x-1} q^{n-x+1}$$

$$= \frac{n!}{(x-1)!(n-x+1)!} p^{x-1} q^{n-x+1}$$

$$\frac{P(X = x)}{P(X = x-1)} =$$

$$\frac{\frac{n!}{x!(n-x)!} p^x q^{n-x}}{\frac{n!}{(x-1)!(n-x+1)!} p^{x-1} q^{n-x+1}} = \frac{n-x+1}{x} \cdot \frac{p}{q}$$

6. Mean of the Binomial/Bernoulli Probability Distribution

Let X be the random variable with the Binomial/Bernoulli distribution. I will first calculate the mean of the distribution which is denoted by μ . It is defined to be the expected value of the random variable. Therefore, $\mu = E[X]$.

The definition of expected value $E[X]$ of a random variable X is as following:

$$E[X] = \sum_i x_i p_i$$

Before obtaining the generalization of the mean of the distribution, let's compute the expected number of tails after five tosses of a biased coin given that the probability of getting H is 0.6 and the probability of getting T is 0.4 as an example.

$$\begin{aligned} E[X] = \sum_i x_i p_i &= 0 \cdot \binom{5}{0} \left(\frac{2}{5}\right)^0 \left(\frac{3}{5}\right)^5 + 1 \\ &\cdot \binom{5}{1} \left(\frac{2}{5}\right)^1 \left(\frac{3}{5}\right)^4 + 2 \\ &\cdot \binom{5}{2} \left(\frac{2}{5}\right)^2 \left(\frac{3}{5}\right)^3 \\ &+ 3 \cdot \binom{5}{3} \left(\frac{2}{5}\right)^3 \left(\frac{3}{5}\right)^2 + 4 \cdot \binom{5}{4} \left(\frac{2}{5}\right)^4 \left(\frac{3}{5}\right)^1 \\ &+ 5 \cdot \binom{5}{5} \left(\frac{2}{5}\right)^5 \left(\frac{3}{5}\right)^0 = 2 \end{aligned}$$

Now let's take a look at the expected value of a random variable associated with the distribution in general using the definition given above.

$$\begin{aligned} \mu &= E[X] = \sum_{x=0}^n x \cdot \binom{n}{x} p^x q^{n-x} \\ &= 0 + 1 \cdot \binom{n}{1} p^1 q^{n-1} + 2 \cdot \binom{n}{2} p^2 q^{n-2} + 3 \\ &\quad \cdot \binom{n}{3} p^3 q^{n-3} + \dots + n \cdot \binom{n}{n} p^n \\ &= npq^{n-1} + 2 \frac{n(n-1)}{2!} p^2 q^{n-2} \\ &\quad + 3 \frac{n(n-1)(n-2)}{3!} p^3 q^{n-3} \\ &\quad + \dots + np^n \\ &= npq^{n-1} + n(n-1)p^2 q^{n-2} \\ &\quad + \frac{n(n-1)(n-2)}{2!} p^3 q^{n-3} + \dots \\ &\quad + np^n \\ &= n \left(pq^{n-1} + (n-1)p^2 q^{n-2} \right. \\ &\quad \left. + \frac{(n-1)(n-2)}{2!} p^3 q^{n-3} + \dots \right. \\ &\quad \left. + p^n \right) \\ &= np \left(q^{n-1} + (n-1)pq^{n-2} + \binom{n-1}{2} p^2 q^{n-3} \right. \\ &\quad \left. + \dots + p^{n-1} \right) \\ &= np(q+p)^{n-1} \\ &= np (\because p+q=1) \end{aligned}$$

The expected value $E[X]$ can be simply abbreviated to np . The expected value of the example with five coin tosses aforementioned can also be computed by the formula $np = 5 \times 0.4 = 2$ more easily.

7. Variance of the Binomial/Bernoulli Probability Distribution

The variance, denoted by σ^2 , is a measure of how far every data point is from the mean of the distribution. It is defined as $E[X^2] - E[X]^2$. It can

be found by calculating the value of the expression npq .

7.1 Proof

$$E[X^2] = \sum_{x=0}^n x^2 \cdot \binom{n}{x} p^x q^{n-x}$$

To simplify:

$$\begin{aligned} E[x^2] - E[x]^2 &= E[x \cdot x] - E[x] \cdot E[x] \\ &= E[x(x-1)] + E[x] - E[x]^2 \end{aligned}$$

$$E[x(x-1)] =$$

$$\sum_{x=0}^n x(x-1) \binom{n}{x} p^x (1-p)^{n-x}$$

$$= \sum_{x=0}^n x(x-1) \frac{n!}{(n-x)! x!} p^x (1-p)^{n-x}$$

$$= \sum_{x=0}^n \frac{n!}{(n-x)! (x-2)!} p^x (1-p)^{n-x}$$

$$= n(n-1)p^2 \sum_{x=0}^n \frac{(n-2)!}{(n-x)! (x-2)!} p^{x-2} (1-p)^{n-x}$$

$$= n(n-1)p^2 \cdot (p+1-p)^{n-2} = n(n-1)p^2$$

$$\begin{aligned} \therefore (\text{Variance}) &= n(n-1)p^2 + np - n^2 p^2 \\ &= np(1-p) = npq \end{aligned}$$

In addition, the standard deviation is the square root of the variance.

8. Mode of Binomial/Bernoulli Probability Distribution

In order to identify the mode of the distribution(s), we will first examine several examples.

First, we'll look at The Binomial Distribution when $n = 5$ and $p = 0.5$.

When the variables are set, probabilities of possible cases are as follows:

$$0 \text{ success } 5 \text{ failures: } P(X = 0) = \binom{5}{0} \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^5 = 0.03125$$

$$1 \text{ success and } 4 \text{ failures: } P(X = 1) = \binom{5}{1} \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^4 = 0.15625$$

$$2 \text{ successes and } 3 \text{ failures: } P(X = 2) = \binom{5}{2} \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^3 = 0.3125$$

$$3 \text{ successes and } 2 \text{ failures: } P(X = 3) = \binom{5}{3} \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^2 = 0.3125$$

$$4 \text{ successes and } 1 \text{ failure: } P(X = 4) = \binom{5}{4} \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^1 = 0.15625$$

$$5 \text{ successes and } 0 \text{ failure: } P(X = 5) = \binom{5}{5} \left(\frac{1}{2}\right)^5 \left(\frac{1}{2}\right)^0 = 0.03125$$

The mode of this group of numbers will be $X=2$ and 3 , since they are the ones with the 'biggest probability' and therefore occurring the most. Therefore, even in later cases, the mode will be the one(s) with the biggest probability value.

Let's look at extreme cases.

If $p = 0$, then the mode will be $X=0$, since everything will be 0 except when $X=0$.

If $p = 1$, then the mode will be $X=5$, since everything will be 0 except when $X=5$.

Here's another example:

The Binomial distribution when $n = 6$ and $p = 0.3$.

0 success and 6 failures: $P(X = 0) =$

$$\binom{6}{0} (0.3)^0 (0.7)^6 = 0.11765$$

1 success and 5 failures:

$$P(X = 1) = \binom{6}{1} (0.3)^1 (0.7)^5 = 0.30253$$

2 successes and 4 failures: $P(X = 2) =$

$$\binom{6}{2} (0.3)^2 (0.7)^4 = 0.32413$$

3 successes and 3 failures: $P(X = 3) =$

$$\binom{6}{3} (0.3)^3 (0.7)^3 = 0.18522$$

4 successes and 2 failures: $P(X = 4) =$

$$\binom{6}{4} (0.3)^4 (0.7)^2 = 0.05953$$

5 successes and 1 failure: $P(X = 5) =$

$$\binom{6}{5} (0.3)^5 (0.7)^1 = 0.01021$$

6 successes and 0 failure: $P(X = 6) =$

$$\binom{6}{6} (0.3)^6 (0.7)^0 = 0.00073$$

In this case, the mode of the distribution will be $X=2$.

We can see that $P(X = k)$ is a function that constantly increases until its peak, then decreases constantly until the end. Therefore, there must be a point $X = x$ ($0 \leq x \leq n$) where it is the highest point(s), and therefore is the mode of the distribution. Using this idea, we can generalize the mode by comparing $P(X = k + 1)$ and $P(X = k)$, where k is a point on the function.

To compare, $\frac{P(X=k+1)}{P(X=k)} = \frac{\binom{n}{k+1}p^{k+1}q^{n-k-1}}{\binom{n}{k}p^kq^{n-k}} = \frac{\frac{p}{k+1}}{\frac{q}{n-k}} =$

$$\frac{(n-k)p}{(k+1)q} = \frac{(n-k)p}{(k+1)(1-p)}$$

Let us set three possible cases in which the distribution may be in.

Case 1: $P(X = k + 1) = P(X = k)$

In this case, this means that the distribution is at its highest peak. Therefore, when the values equal, then they are the mode of the distribution.

However, only integer value of $P(X = k + 1)$ or $P(X = k)$ will be the mode, and there are additional requirements:

$$\frac{(n-k)p}{(k+1)(1-p)} = 1$$

$$(n-k)p = (k+1)(1-p)$$

$$np - pk = k - pk + 1 - p$$

$$k = np + p - 1 = p(n+1) - 1$$

$$k + 1 = p(n+1) - 1 + 1 = \lfloor p(n+1) \rfloor$$

Therefore, $p(n+1)$ must be an integer in order for $P(X = k)$ to be the mode.

Case 2: $P(X = k + 1) > P(X = k)$

In this case, the function is still increasing. This indicates that the mode will come later in the distribution, and $X = k$ is not the mode. $X = k + 1 = \lfloor k + 1 \rfloor$ might be the mode, but it is unclear without further information.

Case 3: $P(X = k + 1) < P(X = k)$

In this case, the function is decreasing. This indicates that the mode is behind the values, and

therefore smaller than k . $X = k + 1$ is not the mode. $X = k$ might be the mode, but it is unclear without further information.

To summarize, the mode of the binomial distribution will be:

mode =

$$\begin{cases} 0 & (p = 0) \\ \lfloor p(n+1) \rfloor & (p(n+1) \notin I) \\ \lfloor \frac{n}{2} \rfloor \text{ and } \lfloor \frac{n+1}{2} \rfloor & (p = \frac{1}{2}) \\ p(n+1), p(n+1) - 1 & (p(n+1) \in I) \\ 1 & (p = 1) \end{cases}$$

9. Median of Binomial and Bernoulli Probability Distribution

Although there is no single way to formalize the median of the distributions, we are going to focus on special cases in which the median can be expressed relatively simple.

When np is an integer, then we can say that the mean, median, and mode of the distribution are equal to np .

An instance of this can be seen when $n = 3$ and $p = \frac{2}{3}$.

In this case, we can line up all the cases:

$$P(X = 0) = \binom{3}{0} \cdot \frac{2^0}{3} \cdot \frac{1^3}{3} = 0.03704$$

$$P(X = 1) = \binom{3}{1} \cdot \frac{2^1}{3} \cdot \frac{1^2}{3} = 0.22223$$

$$P(X = 2) = \binom{3}{2} \cdot \frac{2^2}{3} \cdot \frac{1^1}{3} = 0.44444$$

$$P(X = 3) = \binom{3}{3} \cdot \frac{2^3}{3} \cdot \frac{1^0}{3} = 0.29629$$

The mean of this distribution will always be np .

The mode of this distribution is 2 (0.44444 is the biggest), and the median will be 2 (since $0.03704+0.22223+0.44444 > 0.5$), which is np .

For functions without an integer np value, it is hard to formulize the median of the binomial/Bernoulli distribution.

10. Moment Generating Function (MGF)

We use what we call ‘moments’ of a distribution to describe how a distribution works. Those moments include the first moment, $E[X]$, the second moment, $E[X^2]$, the third, $E[X^3]$, and on until the n th moment $E[X^n]$.

The first two moments were already introduced above, the first moment as the mean of the binomial distribution and the second one as the variance in its equation $E[X^2] - E[X]^2$.

The moments after the first two describes different behaviors of the distribution. For example, the third moment is used to measure the skewness of the distribution, whereas the fourth is used to measure the kurtosis of the distribution. We will derive the formula for both of them later in the chapter.

The MGF of X generates the moments of the distribution in relationship with X .

The Moment Generating Function of X , denoted as M_x , is defined as $M_x(t) = E(e^{tX})$, where the function $E(x)$ notes the expected value as mentioned above.

Extending on the function, $M_x(t) = E(e^{tX}) = \sum e^{tx} P(X = x)$, we can say that the derivative of $M_x(t)$ with respect to t represents the corresponding moments at $t=0$.

We can prove this by considering the function e^x in its taylor series form:

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots$$

Therefore, we can express the term e^{tx} as such:

$$e^{tx} = 1 + tx + \frac{(tx)^2}{2!} + \frac{(tx)^3}{3!} + \dots$$

Since $M_x(t) = E(e^{tX})$,

$$\begin{aligned} M_x(t) &= E\left[1 + tx + \frac{(tx)^2}{2!} + \frac{(tx)^3}{3!} + \dots\right] \\ &= E[1] + tE[x] + \frac{t^2}{2!}E[x^2] + \frac{t^3}{3!}E[x^3] + \dots \\ &\quad \dots \frac{t^n}{n!}E[x^n]. \end{aligned}$$

$$\therefore \frac{d}{dt} M_x(t) =$$

$$\frac{d}{dt} \left(E[1] + tE[x] + \frac{t^2}{2!}E[x^2] + \frac{t^3}{3!}E[x^3] + \dots + \frac{t^n}{n!}E[x^n] \right)$$

$$\begin{aligned} &= E[x] + tE[x^2] + \frac{t^2}{2!}E[x^3] + \dots \\ &\quad + \frac{t^{n-1}}{(n-1)!}E[x^n] \end{aligned}$$

Therefore, if we plug in $t = 0$,

$$\begin{aligned} &\left(\frac{d}{dt} M_x(t) \text{ at } t = 0\right) \\ &= (E[x] + 0 + 0 + 0 + \dots + 0) \\ &= E[x] \end{aligned}$$

Which gives us the first moment.

To get the n th moment, we can simply take more derivatives. For example, to get the second moment

$E[x^2]$, we can take the second derivative and plug in $t = 0$:

$$\begin{aligned} \frac{d^2}{dt^2} M_x(t) &= \\ \frac{d^2}{dt^2} \left(E[1] + tE[x] + \frac{t^2}{2!} E[x^2] + \frac{t^3}{3!} E[x^3] + \dots \right. \\ &\quad \left. + \frac{t^n}{n!} E[x^n] \right) \\ &= 0 + 0 + E[x^2] + tE[x^3] + \dots \\ &\quad + \frac{t^{n-2}}{(n-2)!} E[x^n] \end{aligned}$$

Therefore, by plugging in $t = 0$,

$$\begin{aligned} \left(\frac{d^2}{dt^2} M_x(t) \text{ at } t = 0 \right) \\ = (0 + 0 + E[x^2] + 0 + 0 + 0 + \dots + 0) \\ = E[x^2] \end{aligned}$$

The same would work for the third, fourth and up to the n th moment of the distribution.

11. Moment Generating Function of Binomial and Bernoulli Distribution

The moment generating function for the binomial distribution can be represented as $(pe^t + q)^n$, or in other terms $(pe^t + 1 - p)^n$.

This can be proven through a simple process.

First of all, the probability of a certain case can be represented as follows:

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{(n-x)}$$

And the original formula for the expected value is as such:

$$E[X] = \sum_{x=0}^n x P_x$$

We know that the moment generating function, according to its definition, can be represented as:

$$M_x(t) = E[e^{tx}]$$

Therefore, by substituting with the equation we mentioned earlier, we can come up with this expression:

$$\begin{aligned} M_x(t) &= \sum_{x=0}^n e^{tx} P_x \\ &= \sum_{x=0}^n \binom{n}{x} p^x (1 - p)^{(n-x)} \times e^{tx} \end{aligned}$$

Which can be simplified as:

$$\sum_{x=0}^n \binom{n}{x} (pe^t)^x (1 - p)^{(n-x)}$$

Hence, according to the binomial theorem, we can say this can be simplified to:

$$(pe^t + 1 - p)^n$$

To test this, let us find the first moment of the distribution:

$$\begin{aligned} E[X] &= M'(0) = \frac{d}{dt} (pe^t + 1 - p)^n(0) \\ &= n(pe^t + 1 - p)^{n-1} \cdot pe^t(0) \\ &= n(p + 1 - p)^{n-1} \cdot p \\ &= n(1)p \\ &= np \end{aligned}$$

This is indeed the first moment of the binomial distribution. Let us check the second moment:

$$\begin{aligned} E[X^2] &= M''(0) = \frac{d^2}{dt^2} (pe^t + 1 - p)^n(0) \\ &= \frac{d}{dt} n(pe^t + 1 - p)^{n-1} \cdot pe^t(0) \\ &= n(n-1)(pe^t + 1 - p)^{n-2} \cdot pe^t \cdot pe^t \\ &\quad + n(pe^t + 1 - p)^{n-1} \cdot pe^t(0) \\ &= n(n-1)(p + 1 - p)^{n-2} \cdot p \cdot p + n(p \\ &\quad + 1 - p)^{n-1} \cdot p \\ &= n(n-1)p^2 + np \end{aligned}$$

Since the variance of the distribution is $E[X^2] - E[X]^2 = npq$, we can check:

$$\begin{aligned} E[X^2] - E[X]^2 &= n(n-1)p^2 + np - n^2p^2 \\ &= n^2p^2 - np^2 + np - n^2p^2 = -np^2 + np \\ &= np(-p + 1) \\ &= npq \end{aligned}$$

This is in accordance with what we found out earlier.