

# Probability using binomial distribution method

Rishabh Verma, Satya Prakash, Sneha Nivedita

**Abstract** This paper proposes the methods of finding probability of an event using the methods of binomial distribution. This paper will consist of the problems involving the binomial distributions. The binomial distribution method will discuss about the type of recursion formula, mean and variance of the binomial distribution.

**Keywords:** binomial distribution, recursion, mean, variance.

## I. INTRODUCTION

In probability theory and statistics, the binomial distribution with parameters  $n$  and  $p$  is the discrete probability distribution of the number of successes in a sequence of  $n$  independent yes/no experiments, each of which yields success with probability  $p$ . A success/failure experiment is also called a Bernoulli experiment or Bernoulli trial; when  $n = 1$ , the binomial distribution is a Bernoulli distribution. The binomial distribution is the basis for the popular binomial test of statistical significance.

The binomial distribution is frequently used to model the number of successes in a sample of size  $n$  drawn with replacement from a population of size  $N$ . If the sampling is carried out without replacement, the draws are not independent and so the resulting distribution is a hyper geometric distribution, not a binomial one. However, for  $N$  much larger than  $n$ , the binomial distribution is a good approximation, and widely used.

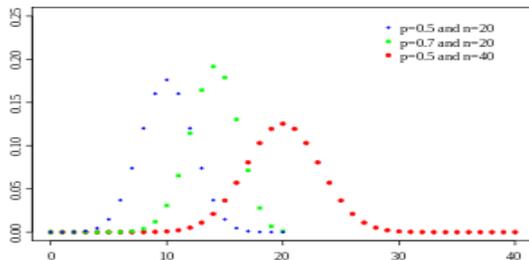


Fig. 1 Probability mass function

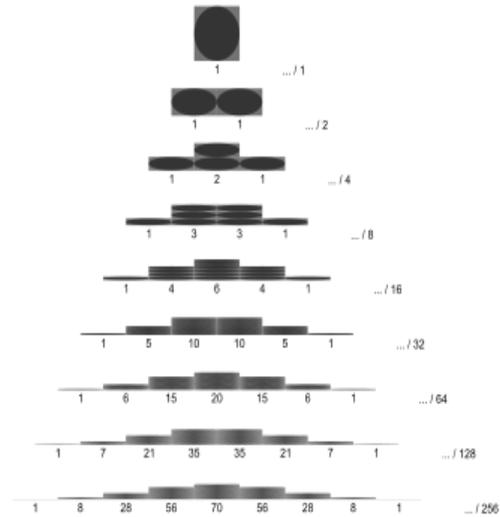


Fig. 2 Pascal's Triangle

The probability distribution function is generally given by the formula

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

Annotations for the formula:

- $n!$ : This starts the count of number of ways event can occur.
- $(n-x)!$ : This ends the count of number of ways event can occur.
- $x!$ : This deletes duplications.
- $p^x$ : This is the probability of success for  $x$  trials.
- $q^{n-x}$ : This is the probability of failure for the  $x$  trials.

## Basic terms

- **Mode:** for a discrete random variable, the value with highest probability (the location at which the probability mass function has its peak); for a continuous random variable, the location at

which the probability density function has its peak.

- Support: the smallest closed set whose complement has probability zero.
- Head: the range of values where the pmf or pdf is relatively high.
- Tail: the complement of the head within the support; the large set of values where the pmf or pdf is relatively low.
- Expected value or mean: the weighted average of the possible values, using their probabilities as their weights; or the continuous analog thereof.
- Median: the value such that the set of values less than the median has a probability of one-half.
- Variance: the second moment of the pmf or pdf about the mean; an important measure of the dispersion of the distribution.
- Standard deviation: the square root of the variance, and hence another measure of dispersion.
- Symmetry: a property of some distributions in which the portion of the distribution to the left of a specific value is a mirror image of the portion to its right.
- Skewness: a measure of the extent to which a pmf or pdf "leans" to one side of its mean.

## II. SPECIFICATIONS

### 1.1. Probability mass function

In general, if the random variable  $X$  follows the binomial distribution with parameters  $n$  and  $p$ , we write  $X \sim B(n, p)$ . The probability of getting exactly  $k$  successes in  $n$  trials is given by the probability mass function:

$$f(k; n, p) = \Pr(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

for  $k = 0, 1, 2, \dots, n$ , where

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

is the binomial coefficient, hence the name of the distribution. The formula can be understood as follows: we want exactly  $k$  successes ( $p^k$ )

and  $n - k$  failures  $(1 - p)^{n-k}$ . However, the  $k$  successes can occur anywhere among

$\binom{n}{k}$  different ways of distributing  $k$  successes in a sequence of  $n$  trials. In creating reference tables for binomial distribution probability, usually the table is filled in up to  $n/2$  values. This is because for  $k > n/2$ , the probability can be calculated by its complement as

$$f(k, n, p) = f(n - k, n, 1 - p).$$

Looking at the expression  $f(k, n, p)$  as a function of  $k$ , there is a  $k$  value that maximizes it. This  $k$  value can be found by calculating

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

And comparing it to 1. There is always an integer  $M$  that satisfies

$$(n + 1)p - 1 \leq M < (n + 1)p.$$

$f(k, n, p)$  is monotone increasing for  $k < M$  and monotone decreasing for  $k > M$ , with the exception of the case where  $(n + 1)p$  is an integer. In this case, there are two values for which  $f$  is maximal:  $(n + 1)p$  and  $(n + 1)p - 1$ .  $M$  is the *most probable (most likely)* outcome of the Bernoulli trials and is called the mode. Note that the probability of it occurring can be fairly small.

### 1.2. Cumulative distribution function

The cumulative distribution function can be expressed as:

$$F(k; n, p) = \Pr(X \leq k) = \sum_{i=0}^{\lfloor k \rfloor} \binom{n}{i} p^i (1 - p)^{n-i}$$

Where  $\lfloor k \rfloor$  is the "floor" under  $k$ , i.e. the greatest integer less than or equal to  $k$ .

It can also be represented in terms of the regularized incomplete beta function, as follows.

$$\begin{aligned}
 F(k; n, p) &= \Pr(X \leq k) \\
 &= I_{1-p}(n - k, k + 1) \\
 &= (n - k) \binom{n}{k} \int_0^{1-p} t^{n-k-1} (1 - t)^k dt.
 \end{aligned}$$

1.3. Means and variance

If  $X \sim B(n, p)$ , that is,  $X$  is a binomially distributed random variable,  $n$  being the total number of experiments and  $p$  the probability of each experiment yielding a successful result, then the expected value of  $X$  is

$$E[X] = np,$$

(For example, if  $n=100$ , and  $p=1/4$ , then the average number of successful results will be 25) and the variance

$$\text{Var}[X] = np(1 - p).$$

1.4. Mode and median

Usually the mode of a binomial  $B(n, p)$  distribution is equal to  $\lfloor (n + 1)p \rfloor$ , where  $\lfloor \cdot \rfloor$  is the floor function. However when  $(n + 1)p$  is an integer and  $p$  is neither 0 nor 1, then the distribution has two modes:  $(n + 1)p$  and  $(n + 1)p - 1$ . When  $p$  is equal to 0 or 1, the mode will be 0 and  $n$  correspondingly. These cases can be summarized as follows:

$$\text{mode} = \begin{cases} \lfloor (n+1)p \rfloor & \text{if } (n+1)p \text{ is 0 or a noninteger,} \\ (n+1)p \text{ and } (n+1)p - 1 & \text{if } (n+1)p \in \{1, \dots, n\}, \\ n & \text{if } (n+1)p = n+1. \end{cases}$$

In general, there is no single formula to find the median for a binomial distribution, and it may even be non-unique. However several special results have been established:

- If  $np$  is an integer, then the mean, median, and mode coincide and equal  $np$ .<sup>[3][4]</sup>
- Any median  $m$  must lie within the interval  $\lfloor np \rfloor \leq m \leq \lceil np \rceil$ .<sup>[5]</sup>
- A median  $m$  cannot lie too far away from the mean:  $|m - np| \leq \min\{\ln 2, \max\{p, 1 - p\}\}$ .<sup>[6]</sup>

- The median is unique and equal to  $m = \text{round}(np)$  in cases when either  $p \leq 1 - \ln 2$  or  $p \geq \ln 2$  or  $|m - np| \leq \min\{p, 1 - p\}$  (except for the case when  $p = 1/2$  and  $n$  is odd).<sup>[5][6]</sup>
- When  $p = 1/2$  and  $n$  is odd, any number  $m$  in the interval  $1/2(n - 1) \leq m \leq 1/2(n + 1)$  is a median of the binomial distribution. If  $p = 1/2$  and  $n$  is even, then  $m = n/2$  is the unique median.

III.CONCLUSION

This paper explains about the probability binomial distribution. It also explains how we can use the binomial distribution to find probability of an event.

REFERENCE

- B. S. Everitt: *The Cambridge Dictionary of Statistics*, Cambridge University Press, Cambridge (3rd edition, 2006). ISBN 0-521-69027-7
- Bishop: *Pattern Recognition and Machine Learning*, Springer, ISBN 0-387-31073-8
- den Dekker A. J., Sijbers J., (2014) "Data distributions in magnetic resonance images: a review", *Physica Medica*,