

Probability Distribution

EGI SAFITRI, S.MAT., M.SI

Introduction

Probability is needed, because the truth of a conclusion made from data analysis cannot actually be confirmed to be absolutely correct, because the data is based on samples. The probability distribution function is generally divided into discrete probability distribution and continuous probability distribution.

Discrete Probability Distribution

- Bernoulli/Binomial
- Poisson
- Hipergeometrik

Continuous Probability Distribution

- Normal (Z, t, Chi-Square, dll)

Binomial Probability Distribution

The BI in the word BINOMIAL means two. This refers to each trial or chance, there are only two POSSIBLE outcomes.

The question that usually arises is, when do we use to calculate the odds using the Binomial odds distribution equation? If the event fulfills the properties below then when calculating the probability of the event occurring then the equation used is the odds equation of the binomial distribution.

Binomial Probability Distribution

1. The binomial distribution is a statistical probability distribution that summarizes the likelihood that a value will take one of two independent values under a given set of parameters or assumptions.
2. The basic assumptions of the binomial distribution are that there is only one outcome for each trial, that each trial has the same probability of success, and that each trial is mutually exclusive or independent of each other.
3. The binomial distribution is a common discrete distribution used in statistics, in contrast to continuous distributions, such as the normal distribution.

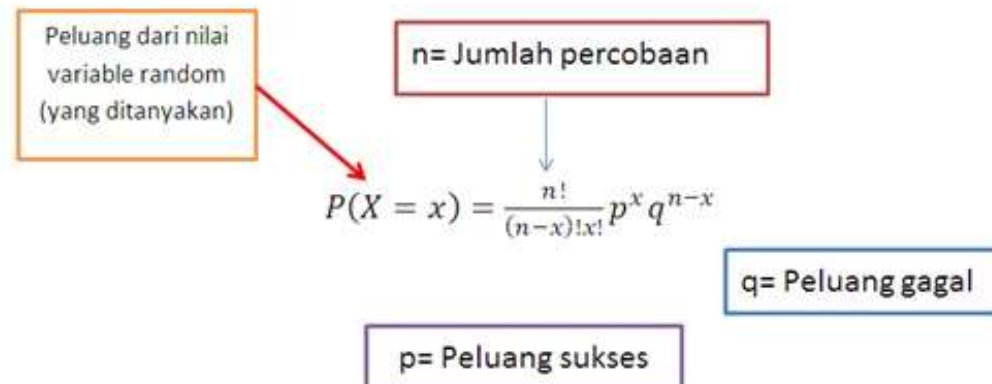
Binomial Probability Distribution

1. The experiment is conducted n times.
2. Each trial has two possible outcomes.
3. The possible outcomes of each trial are the same.
4. The results obtained in the first experiment will not affect the results obtained in other experiments (mutually independent).

Binomial Probability Distribution

The equation for calculating the odds can be seen as follows. Suppose X is a discrete random variable. Then the probability of X is:

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



Binomial Probability Distribution

- If
 - The variable has only **two possible outcomes**
 - The **probability** of each outcome remains **constant**, regardless of previous experiment results

Binomial Distribution

- **Probability of outcomes in a binomial distribution**
 - Probability of success = **p**
 - Probability of failure = **$q = 1 - p$**

Example

- The probability of success (S) in an experiment is $p \rightarrow \text{prob}(S) = p$
- The probability of failure (G) is $q = 1 - p \rightarrow \text{prob}(G) = q$
- **1x experiment:**
 - probability of success: p
 - probability of failure: q
- **2x experiments:**
 - probability of success then success (S, S): pp
 - probability of success then failure (S, G): pq
 - probability of failure then success (G, S): qp
 - probability of failure then failure (G, G): qq

Success-failure 2x Trials

| Number of Successes | Success Combinations | Number of Ways | Probability |
|---------------------|----------------------|----------------|---------------------|
| 2 | SS | 1 | $pp = 1p^2q^0$ |
| 1 | SG or GS | 2 | $pq + qp = 2p^1q^1$ |
| 0 | GG | 1 | $qq = 1p^0q^2$ |

Success-failure 3x Trials

| Number of Successes | Success Combinations | Number of Ways | Probability |
|---------------------|----------------------|----------------|-----------------|
| 3 | SSS | 1 | $ppp = 1p^3q^0$ |
| 2 | SSG, SGS, GSS | 3 | $ppq = 3p^2q^1$ |
| 1 | SGG, GSG, GGS | 3 | $pqq = 3p^1q^2$ |
| 0 | GGG | 1 | $qqq = 1p^0q^3$ |

Success-Failure

- **3x experiments:**
 - probability of success on the 3rd experiment: qqp
 - probability of success in any one experiment: $pqq + qpq + qqp$
- **5x experiments:**
 - probability of 2 successes: $ppqqq + pqpqq + \dots + qqppp$

$$\binom{5}{2} p^2 q^3 = 10p^2 q^3$$

Example

Suppose we have a coin that consists of two sides, front and back. Suppose we draw the coin 10 times. In the first draw, the possible outcomes are either obverse or reverse. On the second draw, the result may be either the obverse or the reverse. And so on. Each time we draw, the probability of the result is the same, only two, the front side or the back side. Out of ten tries, what is the chance that the front side will appear twice?

Contoh

Given:

Number of trials = $n = 10$

- Probability of success = probability of a coin landing heads in each trial = $p = 0.5$
- Probability of failure = probability of not landing heads = $q = 1 - p = 0.5$

Question:

In 10 trials, what is the probability that heads appear exactly 2 times?

That is:

$$P(X = 2) = \frac{10!}{(10 - 2)! \cdot 2!} \cdot (0.5)^2 \cdot (1 - 0.5)^8 = 0.0439$$

Exercise

A PDAM (water utility company) found that the risk (probability) of a water supply disruption on any given day is 7%.

Using a **binomial distribution approach**, calculate:

- The risk of having 2 days of disruption in a 30-day month.
- The risk of having **at least 2** days of disruption in 30 days.
- The probability that the water supply is smooth throughout the month (i.e., no disruptions at all in 30 days).

Multinomial Probability Distribution

Multinomial distribution specifically handles events that have **many different outcomes**.

The **binomial distribution** is a special case of the multinomial distribution, where only **two possible outcomes** exist.

Multinomial distributions are **not limited** to events with just different outcomes.

They can also categorize outcomes on a **continuous scale** with **different levels** (e.g., high, medium, low).

Example:

Water level in a storage tank can be classified as "**acceptable**" or "**unacceptable**".

Hence, the multinomial distribution is widely applicable in **process control**.

Multinomial Probability Distribution

Binomial trials, as previously discussed, involve a trial with **two possible outcomes**, namely **success** and **failure**.

It becomes a **multinomial trial** when each trial can result in **more than two possible outcomes**.

A binomial experiment becomes a multinomial experiment if each trial can result in more than two possible outcomes.

For example, factory production results can be categorized as good, defective, or repairable.

If a trial can produce k types of outcomes E_1, E_2, \dots, E_k ,

with corresponding probabilities p_1, p_2, \dots, p_k ,

then the multinomial distribution over random variables X_1, X_2, \dots, X_k

(denoting the number of occurrences of events E_1, E_2, \dots, E_k)

in n independent trials is given by:

$$f(x_1, x_2, \dots, x_k; p_1, p_2, \dots, p_k, n) = \frac{n!}{x_1! x_2! \cdots x_k!} \cdot p_1^{x_1} \cdot p_2^{x_2} \cdots p_k^{x_k}$$

With the conditions:

$$\sum_{i=1}^k x_i = n \quad \text{and} \quad \sum_{i=1}^k p_i = 1$$

Hypergeometric Distribution

The Hypergeometric Distribution is not as difficult as it sounds at first. It is essentially a discrete probability distribution framework. It describes the probability of achieving a certain amount of success in a series of draws from a finite population with no returns.

Difference between binomial distribution and hypergeometric distribution

- The difference is in the way the sample is drawn.
- In the binomial distribution, the mutually independent nature of repetition is required, and the repetition must be done with replacement.
- Meanwhile, the hypergeometric distribution does not require the mutually independent nature of repetition and is done without replacement.

Application to hypergeometric distributions

It is found in a variety of fields, and is most commonly used in acceptance sampling, electronic testing, quality assurance, etc.

In many of these fields, tests are conducted on tested items that eventually become damaged and cannot be returned. Thus, sampling must be done without return

Hypergeometric Distribution

$$P(X = k) = \frac{C(K, k) \cdot C(N - K, n - k)}{C(N, n)}$$

This formula is used to calculate the probability of **k successes** (desired outcomes) in **n draws, without replacement**, from a limited population of size **N**, containing **K** objects with the desired feature.

'C' represents the **binomial coefficient**.