



STATISTICS FOR DATA SCIENCE

Bernoulli Distribution

Prof. Uma D

Department of Computer Science and Engineering

STATISTICS FOR DATA SCIENCE

Bernoulli Distribution

Prof. Uma D

STATISTICS FOR DATA SCIENCE

Discrete Distribution



A **Discrete Probability Distribution** relates to **discrete data**.

It is often used to **model uncertain events** where the possible values for the variable are either **attribute** or **countable**.

The common discrete probability distributions are **Bernoulli**, **Binomial** and **Poisson**.

STATISTICS FOR DATA SCIENCE

Bernoulli Distribution

Many real-life events can only have two possible outcomes:

- A **tossed coin** can either have a **head or a tail**.
- A student can either **pass or fail** in an **exam**.
- A **product** can either **pass or fail** in an **inspection test**.



How do you model the given scenario ?

STATISTICS FOR DATA SCIENCE

Bernoulli Distribution

Bernoulli Distribution / Binary Distribution:

A **Bernoulli trial** is a process that result in **one of two possible outcomes**.

Example:

Success or Failure

Tossing a Coin

Any yes or no question

Was the newborn child a girl?

Is the top card of a shuffled deck an ace?



It is used to **model generic probabilistic situations** with just **two outcomes**.

Applications:

A person who can be either healthy or sick with a certain disease.

The state of a telephone at a given time that can be either free or busy.

The preference of a person who can be either for or against a certain political candidate.

Conditions of Bernoulli Distribution:

- It is a **single trial**. The trial can result in one of the two possible outcomes, labelled success and failure.
- **$P(\text{success})=p$**
- Success could mean anything you want to consider as a positive or negative outcome.
- **$P(\text{failure}) = 1-p$**

Given any probability space , for any event(set of outcomes) one can define a Bernoulli trial, corresponding to whether the event occurred or not.

$$X \sim \text{Bernoulli}(p)$$

STATISTICS FOR DATA SCIENCE

Bernoulli Distribution



For any Bernoulli Trial:

A random variable X is defined as:

$X=1$ if **success** occurs, where probability of success is denoted by p

$X=0$ if **failure** occurs, where probability of failure is $(1-p)$

Then X is said to have a **Bernoulli distribution** with probability p .

Note: A Bernoulli random variable can only take values 0 and 1.

Probability distribution of X : $X \sim \text{Bernoulli}(p)$

X	p(X)
---	------

0	1-p
---	-----

1	p
---	---

$$\begin{aligned}\text{Mean} &= \sum xp(x) \\ &= 0(1-p) + 1 * p \\ &= p\end{aligned}$$

$$\begin{aligned}\text{Variance} &= \sum (x-\text{mean})^2 * p(x) \\ &= (0-p)^2 (1-p) + (1-p)^2 * p \\ &= p*(1-p)\end{aligned}$$

STATISTICS FOR DATA SCIENCE

Bernoulli Distribution: Example



Suppose that a student takes a multiple choice test. The test has 10 questions, each of which has 4 possible answers (only one is correct).

If the student blindly guesses the answer to each question, do the questions form a sequence of Bernoulli trials? If so, identify the trial outcomes and the parameter p .

Solution:

For each question, either the answer chosen is correct or incorrect.

$P(\text{Answer is correct})=1/4$; $P(\text{Answer is incorrect})=3/4$

There are only 2 possible outcomes for each question. Hence each question is a **Bernoulli trial**.

Since there are in total 10 questions, we have a **sequence of Bernoulli trials**.

STATISTICS FOR DATA SCIENCE

Bernoulli Distribution: Example



Approximately 1 in 500 Indian adults are teachers. One Indian adult is randomly selected.

What is the distribution of the number of teachers?

Solution:

X represents the **Indian adult** is a **teacher**.

$X \sim \text{Bernoulli}(1/500)$

Therefore, the **probability distribution of X** is

X	p(x) or P(X=x)
---	----------------

0	499/500
---	---------

1	1/500
---	-------

Do It Yourself !!!

Candidate A is running for office in a certain district.

Thirty persons are selected at random from the population of registered voters and asked if they prefer candidate A.

Do the responses form a sequence of Bernoulli trials?
If so identify the trial outcomes and the meaning of the parameter p .



THANK YOU

Prof. Uma D

Department of Computer Science and Engineering