STATISTICS FOR DATA SCIENCE

Bernoulli Distribution

Prof. Uma D
Prof. Suganthi S
Prof. Silviya Nancy J

Department of Computer Science and Engineering

STATISTICS FOR DATA SCIENCE

Bernoulli Distribution

Prof. Uma D
Prof. Suganthi S
Prof. Silviya Nancy J

Topics to be covered...

- Discrete Probability Distributions
- Bernoulli Distribution
- Examples
- Mean and Variance of Bernoulli 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**.

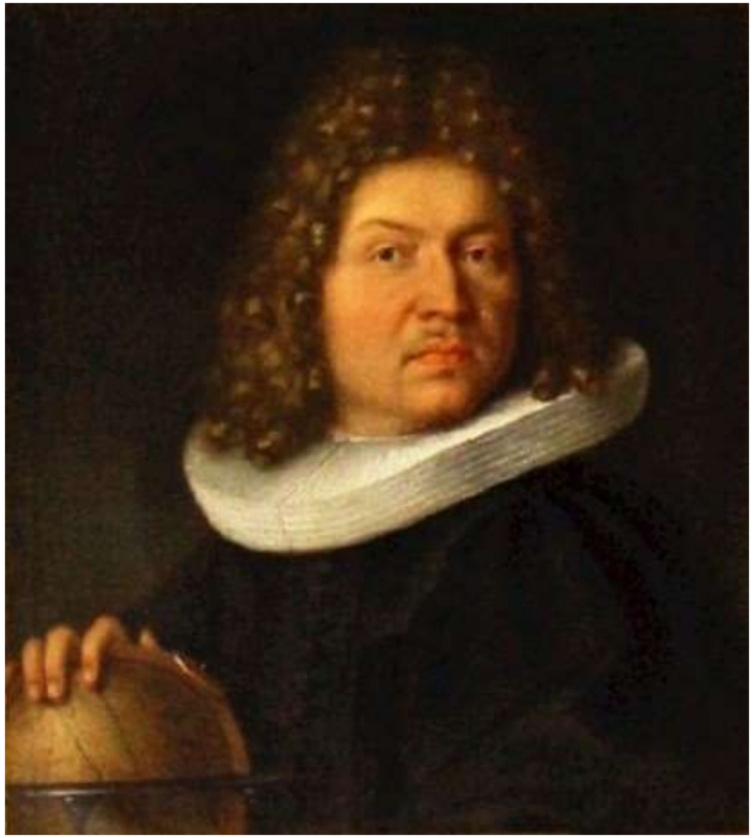
Source: www.citoolkit.com

Discrete Probability Distribution

- **Discrete random variables take on only a finite or countable number of values.**

- A discrete probability distribution is a table (or a formula) listing all possible values that a discrete variable can take on, together with the associated probabilities.

Bernoulli Distribution



Jacob Bernoulli
(Swiss mathematician of the 17th century.) (1654 – 1705)
Discovered constant e

Source: https://en.wikipedia.org/wiki/Jacob_Bernoulli

Bernoulli Trials

Boy? Girl? Heads? Tails? Win? Lose?

Do any of these sound familiar?

When there is the possibility of only two outcomes occuring during any single event, it is called a Bernoulli Trial.

Jacob Bernoulli, a profound mathematician of the late 1600s, from a family of mathematicians, spent 20 years of his life studying probability. During this study, he arrived at an equation that calculates probability in a Bernoulli Trial.

His proofs are published in his 1713 book Ars Conjectandi (Art of Conjecturing).

Source: http://www.math.wichita.edu/history/topics/probability.html#bern-trials

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 ?

Bernoulli Distribution - Conditions

- Its a Single Trial.
- The trial can result in one of the two possible outcomes, labelled success and failure.
- P(success) = p

The term "success" in this sense consists in the result meeting specified conditions

- P(failure) = 1 – p

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

Bernoulli Distribution - Notation

X ~ Bernoulli(p)

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.

Examples

Approximately 1 in 200 Indian adults are Doctors. One Indian adult is randomly selected. What is the distribution of the number of doctors?

Examples

Solution:

X – represents the Indian adult is a doctor.

X ~ Bernoulli (1 /200)

Probability Distribution of X

X	p(x)
0	199/200
1	1/200

Examples

A coin has probability 0.5 of landing heads when tossed.

Let X = 1 if the coin comes up heads, and X = 0 if the coin comes up tails. What is the distribution of X?

Examples

Solution:

- Since $X = 1$ when heads comes up, heads is the success outcome.
- The success probability, $P(X = 1)$, is equal to 0.5.
- Therefore $X \sim \text{Bernoulli}(0.5)$.

Examples

A die has probability 1/6 of coming up 6 when rolled. Let X = 1 if the die comes up 6, and X = 0 otherwise. What is the distribution of X?

Examples

Solution:

- The success probability is $p = P(X = 1) = 1/6$.
- Therefore $X \sim \text{Bernoulli}(1/6)$.

Examples

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.

Examples

Solution:

For each question, Either the answer chosen is correct or incorrect

P(Answer is correct) = ¼

P(Answer is incorrect) = ¾

- Hence 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.

Examples

Ten percent of the components manufactured by a certain process are defective. A component is chosen at random. Let $X = 1$ if the component is defective, and $X = 0$ otherwise. What is the distribution of X?

Examples

Solution:

- The success probability is $p = P(X = 1) = 0.1$.
- Therefore $X \sim \text{Bernoulli}(0.1)$.

Examples – Joining two Bernoulli Random Variables

At a certain fast-food restaurant,25% of drink orders are for a small drink 35% for a medium drink,40% for a large drink

X = 1 if a randomly chosen order is for a small drink and 0 otherwise.

Y = 1 if a randomly chosen order is for a medium drink and 0 otherwise.

Z = 1 if a randomly chosen order is for a small or a medium drink and 0 otherwise.

1.Find Probability distribution of X , Y, Z.

2.Is it possible for both X and Y to be equal to 1?

3.Does $p_Z = p_X + p_Y$?

4.Does Z = X + Y?

Solution:

1. $X \sim \text{Bernoulli}(0.25)$, $Y \sim \text{Bernoulli}(0.35)$, $Z \sim \text{Bernoulli}(0.60)$

2. Is it possible for both X and Y to be equal to 1?

No. If the order is for small drink, it cannot also be for a medium drink. Orders are mutually exclusive.

3. Yes. $p_z = 0.60 = 0.25+0.35 \quad = p_x + p_y$

Mutually exclusive events. For one order if X occurs Y cannot occur and vice versa.

4. No $Z \neq X + Y$.

Mean and Variance of a Bernoulli Random Variables

Mean and Variance of a Bernoulli Random Variable

It is easy to compute the mean and variance of a Bernoulli random variable. If $X \sim \text{Bernoulli}(p)$, then, using Equations (2.29) and (2.30) (in Section 2.4), we compute

$$\begin{aligned}\mu_X &= (0)(1-p) + (1)(p)\\ &= p\end{aligned}$$

$$\begin{aligned}\sigma_X^2 &= (0-p)^2(1-p) + (1-p)^2(p)\\ &= p(1-p)\end{aligned}$$

Summary

If $X \sim \text{Bernoulli}(p)$, then

$$\mu_X = p \tag{4.1}$$

$$\sigma_X^2 = p(1-p) \tag{4.2}$$

Bernoulli Distribution

X	p(X)
0	1-p
1	p

Mean $= \sum x p(x)$

$= 0(1-p) + 1 * p$

$= \mathbf{p}$

Variance $= \sum (x\text{-mean})^2 * p(x)$

$= (0-p)^2 (1-p) + (1-p)^2 * p$

$= \mathbf{p*(1-p)}$

Example - Mean and Variance of a Bernoulli Random Variables

Ten percent of the components manufactured by a certain process are defective. A component is chosen at random. Let $X = 1$ if the component is defective, and $X = 0$ otherwise. What is the distribution of X? find the mean and variance.

Example - Mean and Variance of a Bernoulli Random Variables

Solution

Since $X \sim$ Bernoulli(0.1), the success probability p is equal to 0.1. Using Equations (4.1) and (4.2), $\mu_X = 0.1$ and $\sigma_X^2 = 0.1(1 - 0.1) = 0.09$.

Bernoulli Distribution Applications

For all its simplicity, the Bernoulli random variable is very important. In practice, it is used to model generic probabilistic situations with just two outcomes, such as:

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

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

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

THANK YOU

Prof. Uma D
Prof. Suganthi S
Prof. Silviya Nancy J

Department of Computer Science and Engineering