Pattern Recognition

Chapter 02 – Probability Distributions

Content

Content
Introduction https://statisticsbyjim.com/basics/probability-distributions/ , https://www.kdnuggets.com/2020/02/probability-distributions-data-science.html
Bernoulli Distribution
Binomial Distribution
Beta Distribution
Multinomial Distribution
Dirichlet Distribution

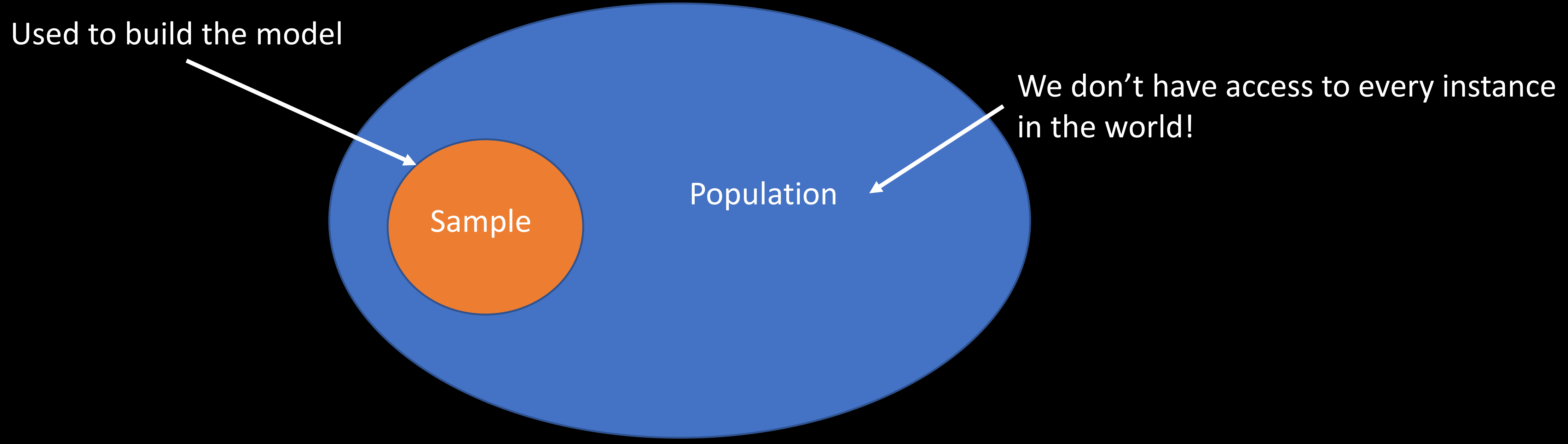
Introduction

- Probability is used for predicting the likelihood of future events.
  - Given a random variable x, what is the probability that $x = 5$?
- Statistics involves the analysis of the frequency of past events.
  - What is the mean of patients with diabetes? How many males with age 35?

- Probability and Statistics allows us to build complex models:
  - Measure the variability of the data
  - Measure the variability of the noise within the data
  - Measure the uncertainty in our model

Introduction

- Our dataset is a **sample** from a **population**.
  - The data samples are used to build models that can be deployed to predict new, unknown instances from the **population**.



Introduction

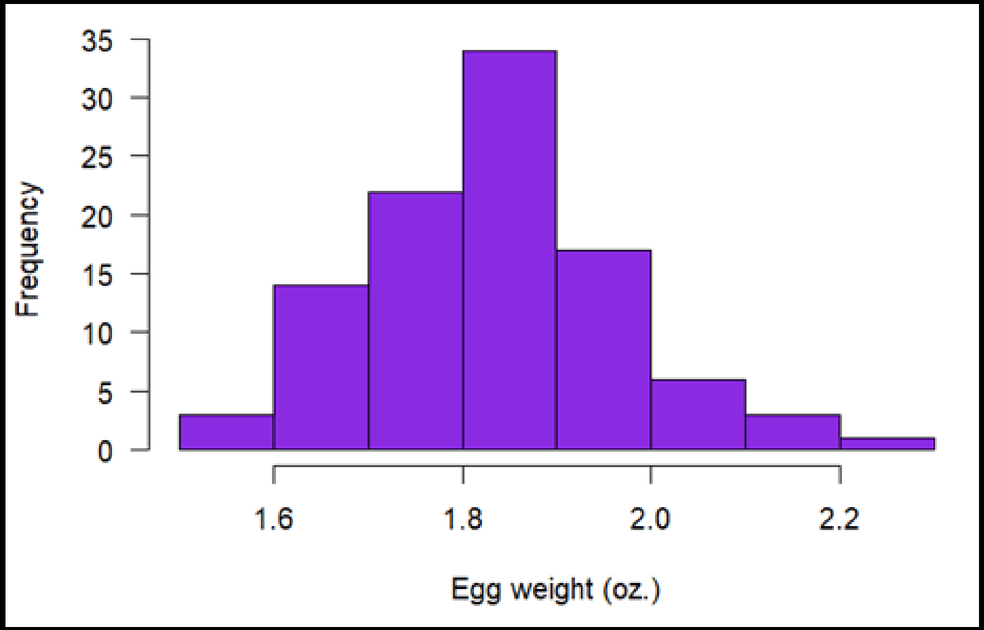
- Datasets can have two types of data:
  - **Numerical**
  - **Categorical** (e.g., {spam, not spam} , {red, green, blue}, names}
- Numerical data can be:
  - **Discrete**: take specific numeric values (number of children, number of courses)
  - **Continuous**: real numbers in any interval (distance, speed, weight, time)

Introduction

- What is a probability distribution?

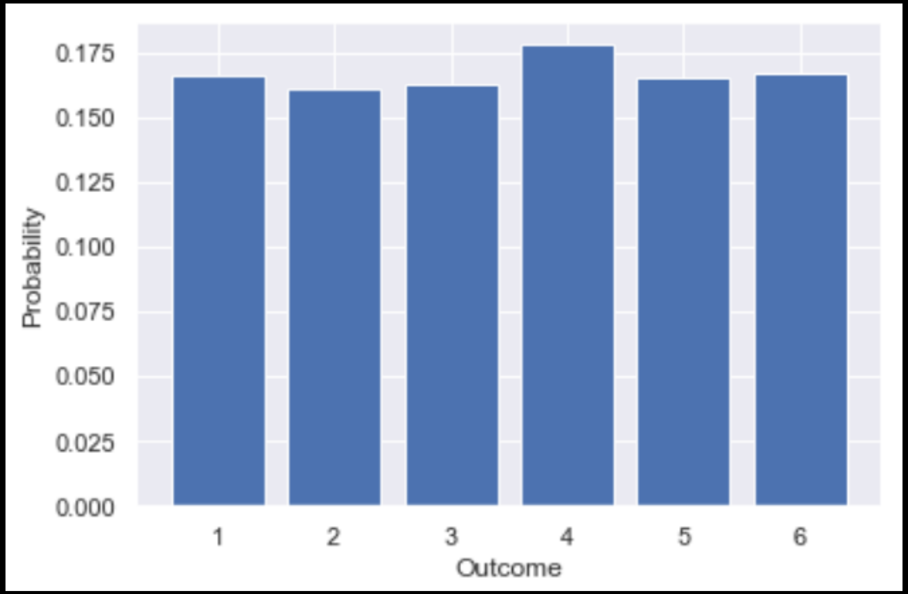
Introduction

- What is a probability distribution?
  A function that describes the likelihood of obtaining all possible values that a random variable can take.
    - For example, measuring the weights of students in the class.
    - As you measure the weights, we create a distribution.
    - If we need to calculate the probability that a random student's weight is between 90KG and 100KG, we have to calculate the likelihood based on the created distribution.



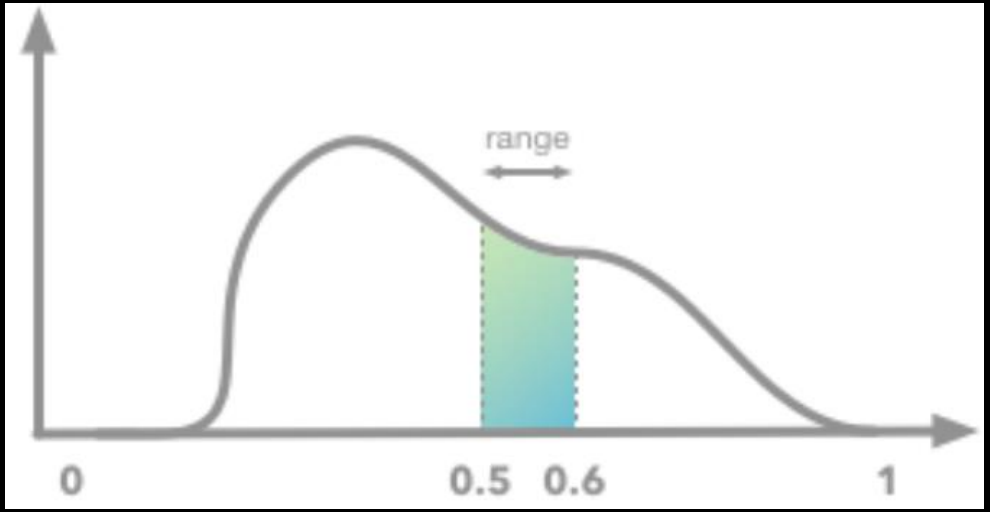
https://www.scribbr.com/statistics/probability-distributions/

Introduction

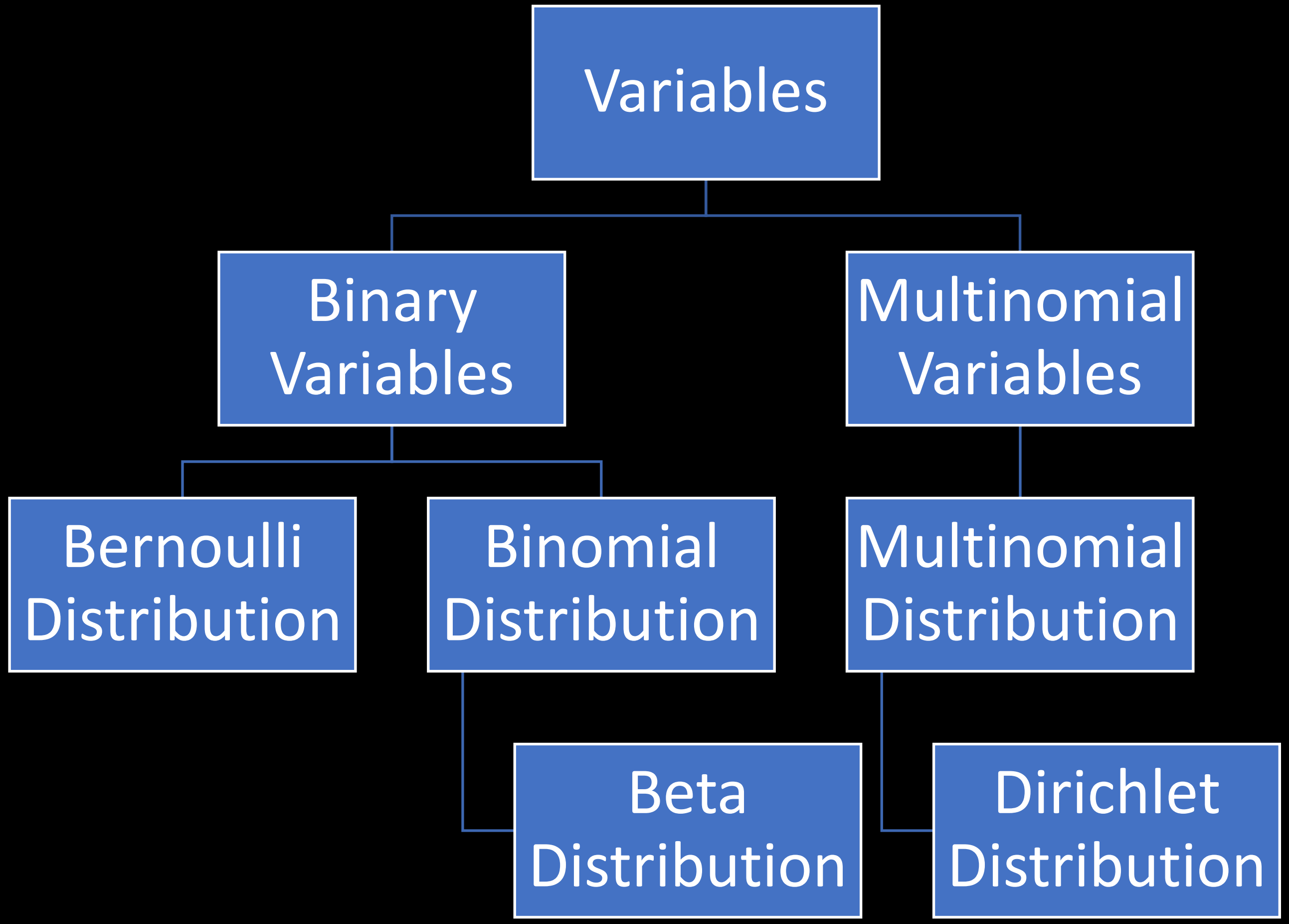


kdnuggets.com/2019/05/probability-mass-density-functions.html

- **Probability Mass Function (PMF)**
  - Describes the probability distribution of a discrete variable (the probability that a discrete random variable can take a specific value).



- **Probability Density Function (PDF)**
  - Describes the probability of a continuous variable.
  - Probabilities need to be integrated over the given range.


Variables
Binary Variables
Multinomial Variables
Bernoulli Distribution
Binomial Distribution
Multinomial Distribution
Beta Distribution
Dirichlet Distribution

Content

Content
Introduction https://statisticsbyjim.com/basics/probability-distributions/ , https://www.kdnuggets.com/2020/02/probability-distributions-data-science.html
Bernoulli Distribution
Binomial Distribution
Beta Distribution
Multinomial Distribution
Dirichlet Distribution

Bernoulli Distribution

- Bernoulli distribution is a discrete probability distribution where variables can have value 0 or 1.

$$Bern(x|\mu) = \mu^x(1-\mu)^{1-x}$$

  - x is the random variable
  - μ is the probability that $x = 1$

$$p(x = 1|\mu) = \mu$$
$$p(x = 0|\mu) = 1 - \mu$$

- The mean (expected value) of the distribution: $E(x) = \mu$
- The variance of the distribution: $var[x] = \mu(1-\mu)$

Bernoulli Distribution

- **Example**: Use Bernoulli distribution to compute the likelihood that a random binary variable x = 1, given that the expected value that $x = 1$ is 70%

- **Solution**:

We have $\mu = 0.7$

$$Bern(x = 1|\mu = 0.7) = \mu^x(1 - \mu)^{1-x}$$
$$= \ 0.7^1(1 - 0.7)^{1-1} = 0.7 * 0.3^0$$
$$= 0.7 \ * 1 = 0.7 = \mu$$

Bernoulli Distribution

- **Example**: Use Bernoulli distribution to compute the likelihood that a random binary variable x = 0, given that the expected value that $x = 0$ is 70%

- **Solution**:

We have $\mu = 0.7$

$$
\begin{aligned}
Bern(x = 0|\mu = 0.7) &= 1 - Bern(x = 1|\mu = 0.3) \\
&= 1 - \ \mu^x(1-\mu)^{1-x} \\
&= 1 - [0.3^1(1-0.3)^{1-1}] = 1 \ - 0.3 \\
&= 0.7 = 1 - \mu
\end{aligned}
$$

Bernoulli Distribution

- Given a dataset $D = \{x_1, x_2, \dots, x_N\}$, the likelihood function is

$$p(D|\mu) = \prod_{n=1}^{N} \mu^{x_n}(1-\mu)^{1-x_n}$$

- We notice that the function depends on the parameter μ.
  - What is the best value of μ that maximizes the likelihood?

Bernoulli Distribution

- Given a dataset $D = \{x_1, x_2, \dots, x_N\}$, the likelihood function is
$$p(D|\mu) = \prod_{n=1}^{N} \mu^{x_n}(1-\mu)^{1-x_n}$$
- We notice that the function depends on the parameter μ.
  - What is the best value of μ that maximizes the likelihood?
$$\mu_{ML} = \frac{1}{N}\sum_{n=1}^{N} x_n$$
  - μ_{ML} is called a **maximum likelihood estimator**.

Bernoulli Distribution

- **Example**: Given a dataset $D = \{0, 1, 0, 0, 0, 1\}$ use Bernoulli distribution to compute what the likelihood that random variable $x = 1$.

- **Solution**:

1. Compute $\mu_{ML} = \frac{0+1+0+0+0+1}{6} = \frac{1}{3} = 0.33$
2. Compute the likelihood:

$$Bern(x = 1|\mu = 0.33) = 0.33^1(1 - 0.33)^{1-1} = 0.33$$

Bernoulli Distribution

The idea of Maximum Likelihood Estimation
is to select the parameters (e.g., μ) that make the
observed data is most likely to happen.

Bernoulli Distribution

Go to code