ECE 295: Lecture 02 Probability

Spring 2018

Prof Stanley Chan

School of Electrical and Computer Engineering
Purdue University

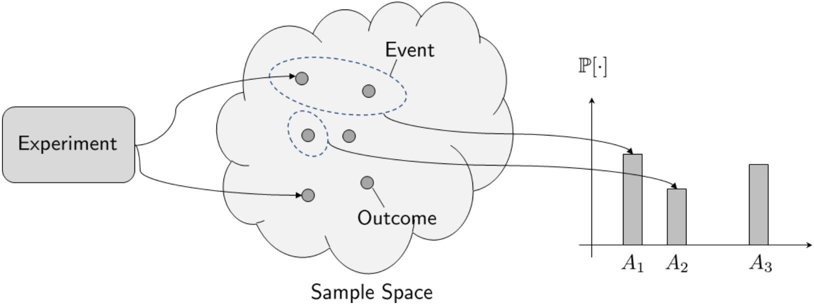
What is Probability?

- It is a
- Always between
- Always the probability of

Example. The probability of getting a Head when tossing a coin:

$$\mathbb{P}(\text{“H”}) =$$

Three Elements of a Probability Model



1. Sample Space
2. Event
3. Probability Law

Probability Distribution

> **"Definition" of Probability Distribution**:
>
> A probability distribution is a histogram when the number of data points go to **infinity**.

When this happens,

$$\text{height of histogram} = \frac{\text{number of times } x_j \text{ happens}}{\text{number of trials}} = \mathbb{P}[X = x_j],$$

So $\mathbb{P}[X = x_j]$ is the probability that we have a state x_j.

(Caution: This "definition" is not general enough to include continuous distributions.)

Random Variable

What is Random Variable?

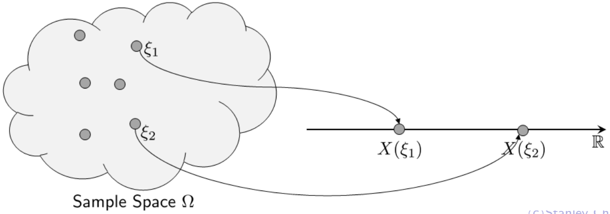
Definition

A **random variable** X is a function $X : \Omega \to \mathbb{R}$ that maps an outcome $\xi \in \Omega$ to a number $X(\xi)$ on the real line.

Why need Random Variable?

- Coin flip: Head or Tail
- Vote: Republican or Democrat
- Alphabet: a, b, c, ..., z

We want to map these outcomes to numbers.

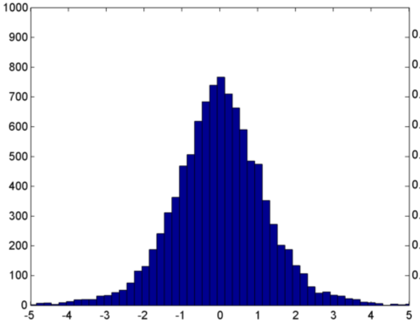


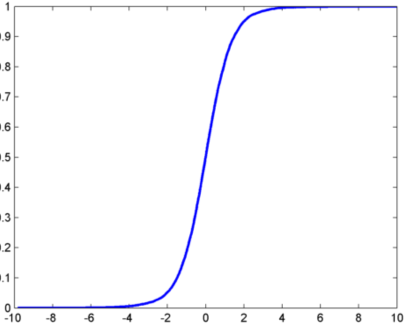
Cumulative Distribution Function

Definition

The **cumulative distribution function** (CDF) of a random variable X is

$$F_X(x) \stackrel{\text{def}}{=} \mathbb{P}[X \leq x]$$





Probability Density Function

Theorem

The **probability density function** *(PDF) is the derivative of the cumulative distribution function (CDF):*

$$p_X(x) = \frac{dF_X(x)}{dx}, \tag{1}$$

if F_X is differentiable at x.

If F_X is not differentiable at x, then $p_X(x)$ is defined as

$$p_X(x) = F_X(x) - \lim_{h \to 0} F_X(x - h). \tag{2}$$

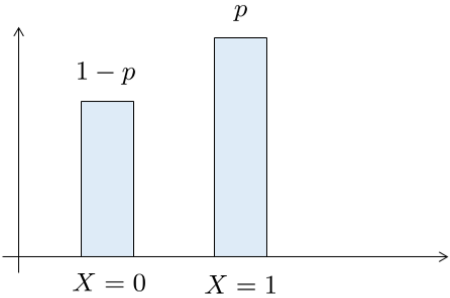
The resulting $p_X(x)$ is called the **probability mass function** (PMF).

Examples of Probability Distributions

Examples

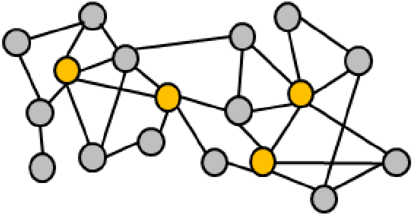
Bernoulli Distribution:

- Two states: $X = 1$ or $X = 0$.
- Flip a coin.
- Probability: $p_X(0)$ or $p_X(1)$.
- We call X a **Bernoulli** random variable.



Example of Bernoulli Distribution

Social Network Analysis



Graph **Matrix**

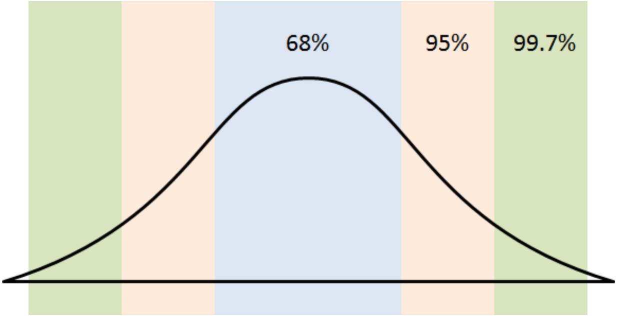
Gaussian Distribution

Also called the **Normal** distribution.

$$p_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} = \mathcal{N}(x \,|\, \mu, \sigma)$$

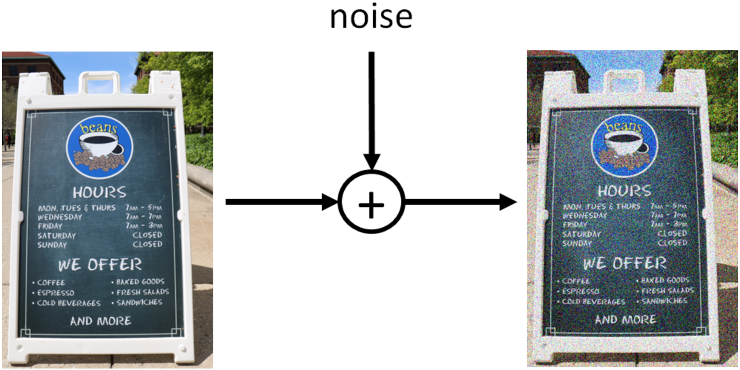
Two parameters:

- μ: mean of the Gaussian
- σ: standard deviation of the Gaussian



Example of Gaussian Distribution

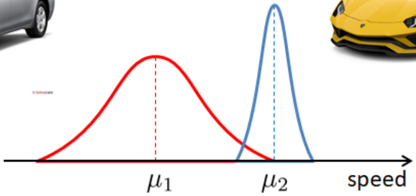
- Noise in cameras can be modeled as Gaussian
- Also Gaussian for communication systems
- A lot of natural phenomena are Gaussian



Example of Gaussian Distribution

Practical questions we can ask:

- You don't see a car, but you measure its speed
- There are only two types of cars: Mini-van and Sports car
- Given the speed, which one would you guess?



Exponential Distribution

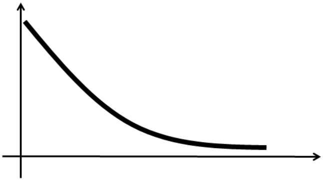
Exponential Distribution:

$$p_X(x) = \lambda \exp\{-\lambda x\}$$

- λ is the rate
- Large λ, decay faster

Usage:

- Use to model inter-arrival time
- Use to model traffic
- Use to model decay processes



Model Selection

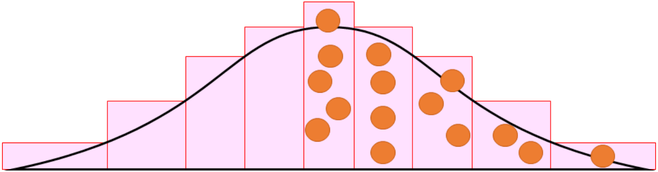
Model Selection

How to Select a Distribution?

- You have data
- A few candidate distributions
- How to choose?

Main Idea

Trick: Divide the candidate distribution into equal **area** bins

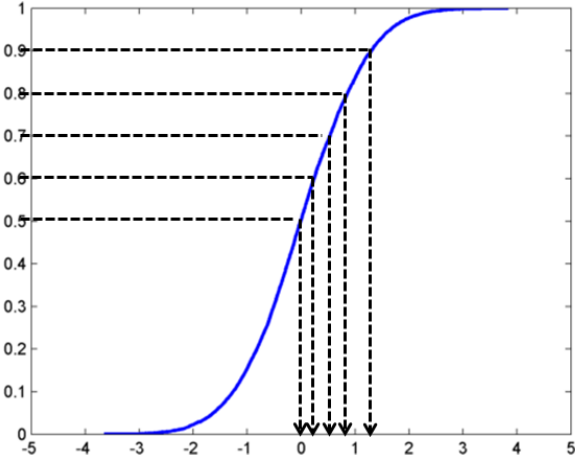


Two sets of numbers:

- Ideal area for each bin
- Actual number of samples fall into each bin

Practice

In practice, you can do this in the CDF domain.



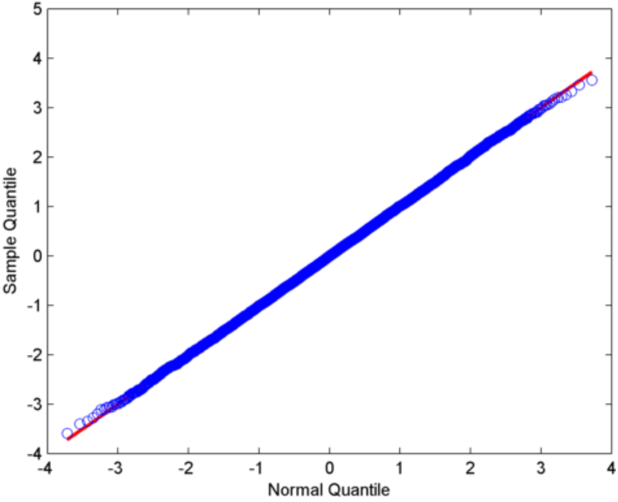
- Equally space cut the probability (y-axis)
- Find correspondingly the value (x-axis)

Algorithm

- Given a set of N data points: $x_1, \ldots, x_N$.
- Sort the numbers as $x_{[1]}, \ldots, x_{[N]}$.
- Create a dummy set $v_1, \ldots, v_N$.
- These v_i's are equally spaced in the range $[0, 1]$.
- Look at your candidate CDF, say $F_Z(z)$.
- Compute $z_i = F_Z^{-1}(v_i)$.
- Plot $x_{[i]}$ VS z_i.

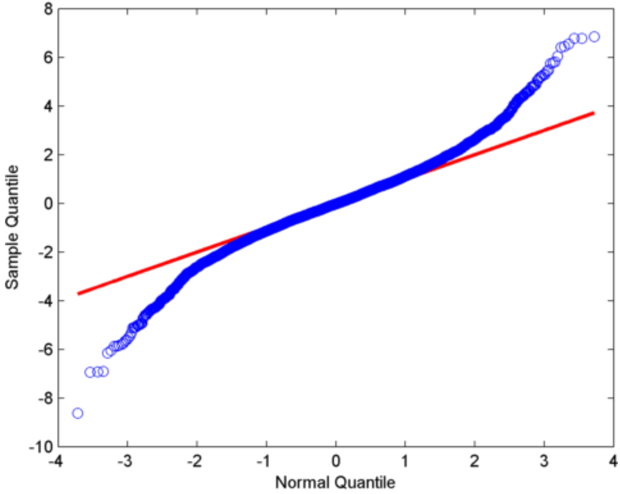
QQ-Plot

- If straight line, then actual fits ideal
- That means your candidate model is good



QQ-Plot

Bad fit:



This type of plot is called the **QQ-plot**.

Summary

- Probability
- Random Variable: A symbolic object. Behind each random variable is a distribution.
- Cumulative Distribution Function
- Probability Density Function
- Three Examples of Probability Distributions
- QQ-Plot