Chapter 3

Probability and Information Theory

# **Table of Contents**

1. [Why Probability?](#_Why_Probability?)
2. [Random Variables](#_Random_Variables)
3. [Probability Distributions](#_Probability_Distributions)
   1. [Discrete Variables and Probability Mass Functions](#_Probability_Mass_Function)
   2. [Continuous Variables and Probability Density Functions](#_Probability_Density_Function)
4. Marginal Probability
5. Conditional Probability
6. The Chain Rule of Conditional Probabilities
7. Independence and Conditional Independence
8. Expectation, Variance, and Covariance
9. Common Probability Distributions
   1. Bernoulli Distribution
   2. Multinomial Distribution
   3. Gaussian Distribution
   4. Exponential and Laplace Distributions
   5. The Dirac Distribution and Empirical Distribution
   6. Mixture of Distributions
10. Useful Properties of Common Functions
11. Bayes’ Rule
12. Technical Details of Continuous Variables
13. Information Theory
14. Structured Probabilistic Models

**Important terms:**

* Probability
* Information theory
* Degree of Belief
* Frequentist prob
* Bayesian prob
* Random Variable
* [Probability Mass Function](#PMF)
* [Probability Density Function](#PDF)

## **Introduction**

Probability -> means to represent uncertainty  
Information Theory -> means to quantify amount of uncertainty

## **Why Probability?**

Unlike other branches of Comp. Sc., Machine Learning normally deals with uncertain and stochastic quantities.

Three possibilities of uncertainty:

1. **Model stochasticity,** e.g. dynamics of a sub-atomic particle
2. **Incomplete observability**, e.g. [Monty-Hall Problem](https://www.youtube.com/watch?v=4Lb-6rxZxx0)
3. **Incomplete modelling**, e.g. When a continuous quantity is binned, we lose some information

Why prob? Its more practical to be somewhat uncertain rather than much complex

## **Random Variables**

Random Variable: a variable that can have different possible values. Random means not able to be predicted.

Types of random variables:

1. Continuous random variable
2. Discrete random variable

## **Probability Distributions**

A random variable can take any possible state, but to quantify which state is it more likely to be in; we must use probability distributions. It can be for

* Discrete variables (described as probability mass function *P* )
* Continuous variables (described as probability density function *p* )

### **Probability Mass Function**

*P(x* = x*)* where P is PMF over x

Properties of PMF:

* Domain of {possible values of *x*}

A special kind of PMF is **joint probability distribution,** which models many variables at the same time. For eg. denotes the probability that x = *x* and y = *y* simultaneously.

### **Probability Density Function**