Chapter 3

Probability and Information Theory

# **Table of Contents**

1. [Why Probability?](#_Why_Probability?)
2. Random Variables
3. Probability Distributions
   1. Discrete Variables and Probability Mass Functions
   2. Continuous Variables and Probability Density Functions
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
* Degree of Belief
* Frequentist prob
* Bayesian prob
* Random Variable

## **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 or 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