# Lecture 01: Mathematical Basics (Probability and Bound on Summations)

# Probability Basics

- Sample Space ($\Omega$): Set of values (finite, infinite, discrete, continuous)
- Random Variable ($\mathbb{X}$): Assigns probability to entities in the sample space
- Probability: $\mathbb{P}[\mathbb{X} = x]$ represents the probability that the random variable $\mathbb{X}$ outputs $x \in \Omega$

# Joint Distribution and Conditional Distribution

- $(\mathbb{X}, \mathbb{Y})$ is a joint distribution over the sample space $\Omega_{\mathbb{X}} \times \Omega_{\mathbb{Y}}$, i.e. for $(x, y) \in \Omega_{\mathbb{X}} \times \Omega_{\mathbb{Y}}$ the random variable $(\mathbb{X}, \mathbb{Y})$ assigns probability to $(x, y)$
- Joint Distribution $\mathbb{P}[\mathbb{X} = x, \mathbb{Y} = y]$
- Marginal Distribution $\mathbb{P}[\mathbb{X} = x] = \sum_{y \in \Omega_{\mathbb{Y}}} \mathbb{P}[\mathbb{X} = x, \mathbb{Y} = y]$
- Conditional Distribution $(\mathbb{X} | \mathbb{Y} = y)$ is the probability distribution over $\Omega_{\mathbb{X}}$ such that the probability

$$\mathbb{P}\left[\mathbb{X} = x | \mathbb{Y} = y\right] := \frac{\mathbb{P}[\mathbb{X} = x, \mathbb{Y} = y]}{\sum_{x \in \Omega_{\mathbb{X}}} \mathbb{P}[\mathbb{X} = x, \mathbb{Y} = y]} = \frac{\mathbb{P}[\mathbb{X} = x, \mathbb{Y} = y]}{\mathbb{P}[\mathbb{Y} = y]}$$

# Bayes' Theorem

**Theorem (Bayes' Theorem)**

*For a joint distribution* $(\mathbb{X}, \mathbb{Y})$ *over* $\Omega_{\mathbb{X}} \times \Omega_{\mathbb{Y}}$ *and* $\mathbb{P}(\mathbb{Y} = y) > 0$, *we have*

$$\mathbb{P}\left[\mathbb{X} = x | \mathbb{Y} = y\right] = \frac{\mathbb{P}\left[(\mathbb{Y} = y | \mathbb{X} = x\right] \mathbb{P}\left[\mathbb{X} = x\right]}{\mathbb{P}\left[\mathbb{Y} = y\right]}$$

Proof: Cross-multiply and use the definition of conditional probability

# Chain Rule

**Theorem (Chain Rule)**

*For a joint distribution* $(\mathbb{X}_1, \ldots, \mathbb{X}_n)$ *over* $\Omega_{\mathbb{X}_1} \times \cdots \times \Omega_{\mathbb{X}_n}$ *the following holds*

$$\mathbb{P}\left[\mathbb{X}_1 = x_1, \ldots, \mathbb{X}_n = x_n\right] = \prod_{i=1}^{n} \mathbb{P}\left[\mathbb{X}_i = x_i | \mathbb{X}_1 = x_1, \ldots, \mathbb{X}_{i-1} = x_{i-1}\right]$$

Intuition: The probability of $(x_1, \ldots, x_n)$ occurring is equal to the following probabilities

- Probability $\mathbb{X}_1 = x_1$ happens .
- Conditioned on $\mathbb{X}_1 = x_1$ happening, $\mathbb{X}_2 = x_2$ happens
- Conditioned on $(\mathbb{X}_1 = x_1, \ldots, \mathbb{X}_{i-1} = x_{i-1})$ happening, $\mathbb{X}_i = x_i$ happens
- And so on ...

Proof: Use Induction on $n$

# Probability of Sets

- Let $\mathbb{X}$ be a random variable over the sample space $\Omega$
- Let $S \subseteq \Omega$
- The probability a sample drawn according to $\mathbb{X}$ being in $S$ is defined to be

$$\mathbb{X}(S) := \sum_{x \in S} \mathbb{P}\left[\mathbb{X} = x\right]$$

# Functions applied on Random Variables

- Let $\mathbb{X}$ be a random variable over the sample space $\Omega$
- Let $f\colon \Omega \to \Omega'$ be a function
- By $f(\mathbb{X})$ we denote the random variable over $\Omega'$ such that the probability of sampling $y \in \Omega'$ is given by

$$\sum_{x \in \Omega\colon\, f(x)=y} \mathbb{P}\left[\mathbb{X} = x\right]$$

- Intuition: $f(\mathbb{X})$ is a random variable such that $f(\mathbb{X})(y)$, i.e. the probability of outputting $y \in \Omega'$, is the probability that $x \in \Omega$ is sampled by $\mathbb{X}$ such that $f(x) = y$
- Alternately, $f(\mathbb{X})$ is the probability distribution that samples $y \in \Omega'$ with probability

$$\mathbb{X}(f^{-1}(y)),$$

where $f^{-1}(y) \subseteq \Omega$ represents the pre-image of $y$ under $f$

# Expected Outcome

- Let $\Omega \subseteq \mathbb{R}$
- The expected outcome (or, the average) of a random variable $\mathbb{X}$ over $\Omega$ is defined to be

$$\mathbb{E}[\mathbb{X}] := \sum_{x \in \Omega} \mathbb{P}[\mathbb{X} = x] \cdot x$$

Prove the following result

**Theorem (Linearity of Expectation)**

*Let $(\mathbb{X}, \mathbb{Y})$ be a joint distribution. Prove that*

$$\mathbb{E}[\mathbb{X} + \mathbb{Y}] = \mathbb{E}[\mathbb{X}] + \mathbb{E}[\mathbb{Y}]$$

**Theorem**

*Let f be an increasing function over* $[1, n]$*, then the following holds*

$$f(1) + \int_1^n f(x)\,\mathrm{d}x \leqslant \sum_{i=1}^{n} f(i) \leqslant \int_1^n f(x)\,\mathrm{d}x + f(n)$$

Proof: Area under the curve.
Think: Bounds on $H_n := \sum_{i=1}^{n} \frac{1}{i}$
Think: Better bounds for convex upward/downward functions and new approximations