Lecture 02: Summations and Probability

Overview

In today's lecture, we shall cover two topics.

1. Technique to approximately sum sequences. We shall see how integration serves as a good approximation of summation of sequences.
2. Basics of Probability. We shall cover Bayes' Rule, chain rule, expectation and linearity of expectation.

Estimating Summation of an Increasing Sequence I

- Suppose f is an increasing function.
- We are interested in finding the following summation

$$S_n = f(1) + f(2) + \cdots + f(n)$$

- For example:
  - For $f(x) = x$, we know that $S_n = n(n+1)/2$
  - For $f(x) = 2x - 1$, we know that $S_n = n^2$.
  - For $f(x) = x^2$, we know that $S_n = n(n+1/2)(n+1)/3$.
  - What if $f(x) = x^3$?
  - What if $f(x) = x \log(x)$?
- Do we have general techniques to perform these summations quickly?

Estimating Summation of an Increasing Sequence II

- We begin with a basic observation

Observation

For an increasing f, we have

$$f(a) \leqslant \int_a^{a+1} f(x)\,\mathrm{d}x \leqslant f(a+1)$$

- For a decreasing f, the inequalities are reversed

Upper Bound.

- Let us apply the basic observation repeatedly

$$f(1) \leqslant \int_1^2 f(x)\,\mathrm{d}x$$
$$f(2) \leqslant \int_2^3 f(x)\,\mathrm{d}x$$
$$\vdots$$
$$f(n) \leqslant \int_n^{n+1} f(x)\,\mathrm{d}x$$

- Summing up both the sides, we get

$$S_n \leqslant \int_1^2 f(x)\,\mathrm{d}x + \cdots + \int_n^{n+1} f(x)\,\mathrm{d}x = \int_1^{n+1} f(x)\,\mathrm{d}x$$

Lower Bound.

- Let us apply the basic observation repeatedly

$$f(1) \geqslant \int_0^1 f(x)\,\mathrm{d}x$$
$$f(2) \geqslant \int_1^2 f(x)\,\mathrm{d}x$$
$$\vdots$$
$$f(n) \geqslant \int_{n-1}^n f(x)\,\mathrm{d}x$$

- Summing up both the sides, we get

$$S_n \geqslant \int_0^1 f(x)\,\mathrm{d}x + \cdots + \int_{n-1}^n f(x)\,\mathrm{d}x = \int_0^n f(x)\,\mathrm{d}x$$

Estimating Summation of an Increasing Sequence V

- We can apply this result directly to several functions f and get the following results
  - Suppose $f(x) = x^c$, for a positive constant c. Then we get

$$\frac{n^{c+1}}{c+1} \leqslant S_n \leqslant \frac{(n+1)^{c+1} - 1}{c+1}$$

  - Try applying it to other functions like $f(x) = x \log(x)$, $f(x) = log(x)$, and $f(x) = \exp(x)$.

Estimating Summation of a Decreasing Sequence

- The basic observation for decreasing function changes to

$$f(a) \geqslant \int_a^{a+1} f(x)\,\mathrm{d}x \geqslant f(a+1)$$

- This implies that

$$\int_0^n f(x)\,\mathrm{d}x \geqslant S_n \geqslant \int_1^{n+1} f(x)\,\mathrm{d}x$$

- Apply this observation to estimate S_n when $f(x) = 1/x$ and $f(x) = x^{-c}$, where c is a positive constant

Food for Thought

- For convex or concave f, we can perform a more precise estimation. Think of using trapeziums to estimate the area of the curve $\int_a^{a+1} f(x)\,\mathrm{d}x$.

Probability Basics

- Sample Space: Ω is a set of outcomes (it can either be finite or infinite)
- Random Variable: $\mathbb{X}$ is a random variable that assigns probabilities to outcomes

Example: Let $\Omega = \{\text{Heads}, \text{Tails}\}$. Let $\mathbb{X}$ be a random variable that outputs Heads with probability $1/3$ and outputs Tails with probability $2/3$

- The probability that $\mathbb{X}$ assigns to the outcome x is represented by

$$\mathbb{P}[\mathbb{X} = x]$$

Example: In the ongoing example $\mathbb{P}[\mathbb{X} = \text{Heads}] = 1/3$.

Function of a Random Variable

- Let $f\colon \Omega \to \Omega'$ be a function
- Let $\mathbb{X}$ be a random variable over the sample space $\mathbb{X}$
- We define a new random variable $f(\mathbb{X})$ is over Ω' as follows

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

Joint Distribution and Marginal Distributions I

- Suppose $(\mathbb{X}_1, \mathbb{X}_2)$ is a random variable over $\Omega_1 \times \Omega_2$.
  - Intuitively, the random variable $(\mathbb{X}_1, \mathbb{X}_2)$ takes values of the form (x_1, x_2), where the first coordinate lies in Ω_1, and the second coordinate likes in Ω_2

For example, let $(\mathbb{X}_1, \mathbb{X}_2)$ represent the temperatures of West Lafayette and Lafayette. Their sample space is $\mathbb{Z} \times \mathbb{Z}$. Note that these two outcomes can be correlated with each other.

Joint Distribution and Marginal Distributions II

- Let $P_1\colon \Omega_1 \times \Omega_2 \to \Omega_1$ be the function $P_1(x_1, x_2) = x_1$ (the projection operator)
- So, the random variable $P_1(\mathbb{X}_1, \mathbb{X}_2)$ is a probability distribution over the sample space Ω_1
- This is represented simply as $\mathbb{X}_1$, the marginal distribution of the first coordinate
- Similarly, we can define $\mathbb{X}_2$

Conditional Distribution

- Let $(\mathbb{X}_1, \mathbb{X}_2)$ be a joint distribution over the sample space $\Omega_1 \times \Omega_2$
- We can define the distribution $(\mathbb{X}_1 | \mathbb{X}_2 = x_2)$ as follows
    - This random variable is a distribution over the sample space Ω_1
    - The probability distribution is defined as follows

$$\mathbb{P}\left[\mathbb{X}_1 = x_1 | \mathbb{X}_2 = x_2\right] = \frac{\mathbb{P}\left[\mathbb{X}_1 = x_1, \mathbb{X}_2 = x_2\right]}{\sum_{x \in \Omega_1} \mathbb{P}\left[\mathbb{X}_1 = x, \mathbb{X}_2 = x_2\right]}$$

For example, conditioned on the temperature at Lafayette being 0, what is the conditional probability distribution of the temperature in West Lafayette?

Bayes' Rule

Theorem (Bayes' Rule)

Let $(\mathbb{X}_1, \mathbb{X}_2)$ be a joint distribution over the sample space (Ω_1, Ω_2). Let $x_1 \in \Omega_1$ and $x_2 \in \Omega_2$ be such that $\mathbb{P}[\mathbb{X}_1 = x_1, \mathbb{X}_2 = x_2] > 0$. Then, the following holds.

$$\mathbb{P}\left[\mathbb{X}_1 = x_1 | \mathbb{X}_2 = x_2\right] = \frac{\mathbb{P}[\mathbb{X}_1 = x_1, \mathbb{X}_2 = x_2]}{\mathbb{P}[\mathbb{X}_2 = x_2]}$$

The random variables $\mathbb{X}_1$ and $\mathbb{X}_2$ are independent of each other if the distribution $(\mathbb{X}_1 | \mathbb{X}_2 = x_2)$ is identical to the random variable $\mathbb{X}_1$, for all $x_2 \in \Omega_2$ such that $\mathbb{P}[\mathbb{X}_2 = x_2] > 0$

We can generalize the Bayes' Rule as follows.

Theorem (Chain Rule)

Let $(\mathbb{X}_1, \mathbb{X}_2, \ldots, \mathbb{X}_n)$ *be a joint distribution over the sample space* $\Omega_1 \times \Omega_2 \times \cdots \times \Omega_n$. *For any* $(x_1, \ldots, x_n) \in \Omega_1 \times \cdots \times \Omega_n$ *we have*

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