Lecture 01: Mathematical Basics (Summations & Probability)

What I am Assuming

- I am assuming that you know asymptotic notations. For example, the big-O, little-O notations

- Let us try to write a closed form expression for the following summation

$$S = \sum_{i=1}^{n} 1$$

- It is trivial to see that $S = n$

Summation II

- Now, let us try to write a closed form expression for the following summation

$$S = \sum_{i=1}^{n} i$$

- We can prove that $S = \frac{n(n+1)}{2}$
    - How do you prove this statement? (Use Induction? Use the formula for the Sum of an Arithmetic Progression?)
- Using Asymptotic Notation, we can say that $S = \frac{n^2}{2} + o(n^2)$

Summation III

- Now, let us try to write a closed form expression for the following summation

$$S = \sum_{i=1}^{n} i^2$$

- We can prove that $S = \frac{n(n+1)(2n+1)}{6}$
  - Why is the expression on the right an integer? (Prove by induction that 6 divides $n(n+1)(2n+1)$ for all positive integer n)
  - How do you prove this statement? (Use Induction?)
- Using Asymptotic Notation, we can say that $S = \frac{n^3}{3} + o(n^3)$

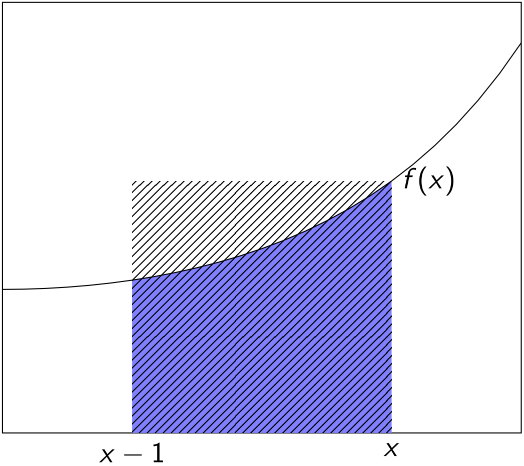
Summation IV

- Do we see a pattern here?
- Conjecture: For $k \geqslant 1$, we have $\sum_{i=1}^{n} i^{k-1} = \frac{n^k}{k} + o(n^k)$.
  - How do we prove this statement?

Estimating Summations by Integration I

- Let f be an increasing function
- For example, $f(x) = x^{k-1}$ is an increasing function for $k > 1$ and $x \geqslant 0$

Estimating Summations by Integration II



Estimating Summations by Integration III

- Observation: "Blue area under the curve" is smaller than the "Shaded area of the rectangle"
  - Blue area under the curve is:

$$\int_{x-1}^{x} f(t)dt$$

  - Shaded area of the rectangle is:

$$f(x)$$

  - So, we have the inequality:

$$\int_{x-1}^{x} f(t)\,\mathrm{d}t \leqslant f(x)$$

  - Summing both side from $x = 1$ to $x = n$, we get

$$\sum_{x=1}^{n} \int_{x-1}^{x} f(t)\,\mathrm{d}t \leqslant \sum_{x=1}^{n} f(x)$$

- The left-hand side of the inequality is

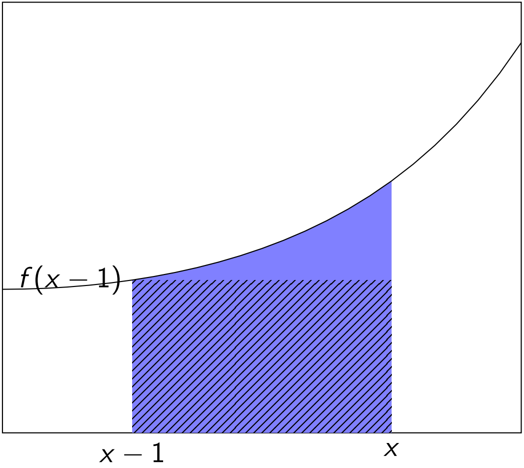
$$\int_0^1 f(t)\,\mathrm{d}t + \int_1^2 f(t)\,\mathrm{d}t + \cdots + \int_{n-1}^{n} f(t)\,\mathrm{d}t = \int_0^n f(t)\,\mathrm{d}t$$

- So, for an increasing f, we have the following lower bound.

$$\int_0^n f(t)\,\mathrm{d}t \leqslant \sum_{x=1}^{n} f(x) \tag{1}$$

Estimating Summations by Integration V

- Now, we will upper bound the summation expression. Consider the figure below



Estimating Summations by Integration VI

- Observation: "Blue area under the curve" is greater than the "Shaded area of the rectangle"
- So, we have the inequality:

$$\int_{x-1}^{x} f(t)\,\mathrm{d}t \geqslant f(x-1)$$

- Now we sum the above inequality from $x = 2$ to $x = n+1$
- We get

$$\int_{1}^{2} f(t)\,\mathrm{d}t + \int_{2}^{3} f(t)\,\mathrm{d}t + \cdots + \int_{n}^{n+1} f(t)\,\mathrm{d}t \geqslant f(1) + f(2) + \cdots + f(n)$$

- So, for an increasing f, we get the following upper bound

$$\int_{1}^{n+1} f(t)\,\mathrm{d}t \geqslant \sum_{x=1}^{n} f(x) \tag{2}$$

Summary: Estimation of Summation using Integration

Theorem

For an increasing function f, we have

$$\int_0^n f(t)\,\mathrm{d}t \leqslant \sum_{x=1}^{n} f(x) \leqslant \int_1^{n+1} f(t)\,\mathrm{d}t$$

Exercise:

- Use this theorem to prove that $\sum_{i=1}^{n} i^{k-1} = \frac{n^k}{k} + o(n^k)$, for $k \geqslant 1$
- Consider the function $f(x) = 1/x$ to find upper and lower bounds for the sum $H_n = 1 + \frac{1}{2} + \cdots + \frac{1}{n}$ using the approach used to prove Theorem 1

Differentiation and Integration

- Differentiation: $f'(x)$ represents the slope of the curve $y = f(x)$ at x
- Integration: $\int_a^b f(t)\,\mathrm{d}t$ represents the area under the curve $y = f(x)$ between $x = a$ and $x = b$
- Increasing function:
  - Observation: The slope an increasing function is positive
  - So, "f is increasing at x" is equivalent to "$f'(x) > 0$," i.e. f' is positive at x
- Suppose we want to mathematically write "Slope of a function f is increasing"
  - The "slope of a function f" is the function "f'"
  - So, the statement "slope of a function f is increasing" is equivalent to "$(f')' \equiv f''$ is positive"

Concave Upwards Functions

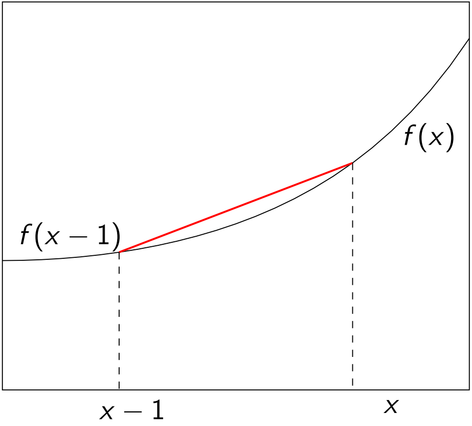
Definition (Concave Upwards Function)

A function f is *concave upwards* in the interval $[a, b]$ if f'' is positive in the interval $[a, b]$.

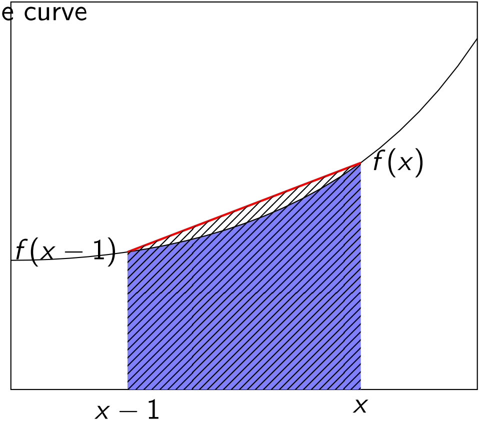
- Example of functions that concave upwards: x^2, $\exp(x)$, $1/x$ (in the interval $(0, \infty)$), $x \log x$ (in the interval $(0, \infty)$)
  - We emphasize that a "concave upwards" function need not be increasing, for example $f(x) = 1/x$ (for positive x) is decreasing

Property of Concave Upwards Function I

- Consider the coordinates $(x-1, f(x-1))$ and $(x, f(x))$
- For a concave upwards function, the secant between the two coordinates is always (on or) above the part of the curve f between the two coordinates



- So, the shaded area of the trapezium is greater than the blue area under the curve



Property of Concave Upwards Function III

- So, we get
$$\frac{f(x-1)+f(x)}{2} \geqslant \int_{x-1}^{x} f(t)\,\mathrm{d}t$$
- Now, use this new observation to obtain a better lower bound for the sum $\sum_{x=1}^{n} f(x)$
- Think: Can you get even tighter bounds?
- Additional Reading: Read on the "trapezoidal rule"

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

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

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

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 $P_1 : \Omega_1 \times \Omega_2 \to \Omega_1$ 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 $(\mathbb{X}_1 \mid \mathbb{X}_2 = x_2)$ as follows
  - This random variable is a distribution over the sample space Ω_1
  - The probability distribution is defined as follows

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

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.*

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

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

Chain Rule

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*

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

Important: Why use Bayes' Rule I

In which context do we foresee to use the Bayes' Rule to compute joint probability?

- Sometimes, the problem at hand will clearly state how to sample $\mathbb{X}_1$ and then, conditioned on the fact that $\mathbb{X}_1 = x_1$, it will state how to sample $\mathbb{X}_2$. In such cases, we shall use the Bayes' rule to calculate

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

- Let us consider an example.
    - Suppose $\mathbb{X}_1$ is a random variable over $\Omega_1 = \{0, 1\}$ such that $\mathbb{P}\left[X_1 = 0\right] = 1/2$. Next, the random variable $\mathbb{X}_2$ is over $\Omega_2 = \{0, 1\}$ such that $\mathbb{P}\left[X_2 = x_1 | \mathbb{X}_1 = x_1\right] = 2/3$. Note that $\mathbb{X}_2$ is biased towards the outcome of $\mathbb{X}_1$.
    - What is the probability that we get $\mathbb{P}\left[\mathbb{X}_1 = 0, \mathbb{X}_2 = 1\right]$?

Important: Why use Bayes' Rule II

- To compute this probability, we shall use the Bayes' rule.

$$\mathbb{P}\left[\mathbb{X}_1 = 0\right] = 1/2$$

Next, we know that

$$\mathbb{P}\left[\mathbb{X}_2 = 0 | \mathbb{X}_1 = 0\right] = 2/3$$

Therefore, we have $\mathbb{P}\left[\mathbb{X}_2 = 1 | \mathbb{X}_1 = 0\right] = 1/3$. So, we get

$$\begin{aligned}\mathbb{P}\left[\mathbb{X}_1 = 0, \mathbb{X}_2 = 1\right] &= \mathbb{P}\left[\mathbb{X}_1 = 0\right] \mathbb{P}\left[\mathbb{X}_2 = 1 | \mathbb{X}_1 = 0\right] \\ &= (1/2) \cdot (1/3) = 1/6\end{aligned}$$

Probability: First Example I

- Let $\mathbb{S}$ be the random variable representing whether I studied for my exam. This random variable has sample space $\Omega_1 = \{Y, N\}$
- Let $\mathbb{P}$ be the random variable representing whether I passed my exam This random variable has sample space $\Omega_2 = \{Y, N\}$
- Our sample space is $\Omega = \Omega_1 \times \Omega_2$
- The joint distribution $(\mathbb{S}, \mathbb{P})$ is represented in the next page

Probability: First Example II

s	p	$\mathbb{P}[\mathbb{S} = s, \mathbb{P} = p]$
Y	Y	1/2
Y	N	1/4
N	Y	0
N	N	1/4

Probability: First Example III

Here are some interesting probability computations
The probability that I pass.

$$\mathbb{P}[\mathbb{P} = \mathsf{Y}] = \mathbb{P}[\mathbb{S} = \mathsf{Y}, \mathbb{P} = \mathsf{Y}] + \mathbb{P}[\mathbb{S} = \mathsf{N}, \mathbb{P} = \mathsf{Y}]$$
$$= 1/2 + 0 = 1/2$$

Probability: First Example IV

The probability that I study.

$$\begin{aligned}\mathbb{P}\left[\mathbb{S} = \mathsf{Y}\right] &= \mathbb{P}\left[\mathbb{S} = \mathsf{Y}, \mathbb{P} = \mathsf{Y}\right] + \mathbb{P}\left[\mathbb{S} = \mathsf{Y}, \mathbb{P} = \mathsf{N}\right] \\ &= 1/2 + 1/4 = 3/4\end{aligned}$$

Probability: First Example V

The probability that I pass conditioned on the fact that I studied.

$$\mathbb{P}\left[\mathbb{P} = \mathsf{Y} \mid \mathbb{S} = \mathsf{Y}\right] = \frac{\mathbb{P}\left[\mathbb{P} = \mathsf{Y}, \mathbb{S} = \mathsf{Y}\right]}{\mathbb{P}\left[\mathbb{S} = \mathsf{Y}\right]}$$

$$= \frac{1/2}{3/4} = \frac{2}{3}$$

Probability: Second Example I

- Let $\mathbb{T}$ be the time of the day that I wake up. The random variable $\mathbb{T}$ has sample space $\Omega_1 = \{4, 5, 6, 7, 8, 9, 10\}$
- Let $\mathbb{B}$ represent whether I have breakfast or not. The random variable $\mathbb{B}$ has sample space $\Omega_2 = \{T, F\}$
- Our sample space is $\Omega = \Omega_1 \times \Omega_2$
- The joint distribution of $(\mathbb{T}, \mathbb{B})$ is presented on the next page

Probability: Second Example II

t	b	$\mathbb{P}[\mathbb{T} = t, \mathbb{B} = b]$
4	T	0.03
4	F	0
5	T	0.02
5	F	0
6	T	0.30
6	F	0.05
7	T	0.20
7	F	0.10
8	T	0.10
8	F	0.08
9	T	0.05
9	F	0.05
10	T	0
10	F	0.02

Probability: Second Example III

- What is the probability that I have breakfast conditioned on the fact that I wake up at or before 7?

Formally, what is $\mathbb{P}\left[\mathbb{B} = \mathsf{T} \mid \mathbb{T} \leqslant 7\right]$?