Lecture 01: Mathematical Basics (Summations & Probability)
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 \( S = n \)
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 $S = \frac{n^2}{2} + o(n^2)$
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 \( S = \frac{n^3}{3} + o(n^3) \)
Do we see a pattern here?

Conjecture: For $k \geq 1$, we have $\sum_{i=1}^{n} i^{k-1} = \frac{n^k}{k} + o(n^k)$.

How do we prove this statement?
Let $f$ be an increasing function

For example, $f(x) = x^{k-1}$ is an increasing function for $k > 1$ and $x \geq 0$
Estimating Summations by Integration II

Basics
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:

\[ \int_{x-1}^{x} f(t)dt \leq f(x) \]

Summing both side from \( x = 1 \) to \( x = n \), we get

\[ \sum_{x=1}^{n} \int_{x-1}^{x} f(t)dt \leq \sum_{x=1}^{n} f(x) \]
The left-hand side of the inequality is

\[ \int_0^1 f(t) \, dt + \int_1^2 f(t) \, dt + \cdots + \int_{n-1}^n f(t) \, dt = \int_0^n f(t) \, dt \]

So, for an increasing \( f \), we have the following lower bound.

\[ \int_0^n f(t) \, dt \leq \sum_{x=1}^n f(x) \]  
(1)
Now, we will upper bound the summation expression. Consider the figure below.
Observation: “Blue area under the curve” is greater than the “Shaded area of the rectangle”

So, we have the inequality:

\[
\int_{x-1}^{x} f(t) \, dt \geq f(x - 1)
\]

Now we sum the above inequality from \(x = 2\) to \(x = n + 1\)

We get

\[
\int_{1}^{2} f(t) \, dt + \int_{2}^{3} f(t) \, dt + \cdots + \int_{n}^{n+1} f(t) \, dt \geq f(1) + f(2) + \cdots + f(n)
\]

So, for an increasing \(f\), we get the following upper bound

\[
\int_{1}^{n+1} f(t) \, dt \geq \sum_{x=1}^{n} f(x) \quad (2)
\]
Theorem

For an increasing function $f$, we have

$$\int_0^n f(t) \, dt \leq \sum_{x=1}^n f(x) \leq \int_1^{n+1} f(t) \, dt$$

Exercise:

- Use this theorem to prove that $\sum_{i=1}^n i^{k-1} = \frac{n^k}{k} + o(n^k)$, for $k \geq 1$

- 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: \( f'(x) \) represents the slope of the curve \( y = f(x) \) at \( x \)

Integration: \( \int_a^b f(t) \, dt \) 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”
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
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.
So, we get

\[
\frac{f(x - 1) + f(x)}{2} \geq \int_{x-1}^{x} f(t) \, dt
\]

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: $\Omega$ is a set of outcomes (it can either be finite or infinite)
- Random Variable: $X$ is a random variable that assigns probabilities to outcomes

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

- The probability that $X$ assigns to the outcome $x$ is represented by

$$P[X = x]$$

Example: In the ongoing example $P[X = \text{Heads}] = 1/3$. 
Let \( f : \Omega \rightarrow \Omega' \) be a function.

Let \( X \) be a random variable over the sample space \( \Omega' \).

We define a new random variable \( f(X) \) is over \( \Omega' \) as follows:

\[
P \left[ f(X) = y \right] = \sum_{x \in \Omega : f(x) = y} P \left[ X = x \right]
\]
Suppose \((X_1, X_2)\) is a random variable over \(\Omega_1 \times \Omega_2\).

Intuitively, the random variable \((X_1, X_2)\) takes values of the form \((x_1, x_2)\), where the first coordinate lies in \(\Omega_1\), and the second coordinate lies in \(\Omega_2\).

For example, let \((X_1, X_2)\) 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.
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 $P_1(X_1, X_2)$ is a probability distribution over the sample space $\Omega_1$.

This is represented simply as $X_1$, the marginal distribution of the first coordinate.

Similarly, we can define $X_2$. 
Let \((X_1, X_2)\) be a joint distribution over the sample space \(\Omega_1 \times \Omega_2\).

We can define the distribution \((X_1 | X_2 = x_2)\) as follows:

- This random variable is a distribution over the sample space \(\Omega_1\).
- The probability distribution is defined as follows:

\[
P[X_1 = x_1 | X_2 = x_2] = \frac{P[X_1 = x_1, X_2 = x_2]}{\sum_{x \in \Omega_1} P[X_1 = x, X_2 = x_2]}
\]

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 \((X_1, X_2)\) be a joint distribution over the sample space \((\Omega_1, \Omega_2)\). Let \(x_1 \in \Omega_1\) and \(x_2 \in \Omega_2\) be such that \(P[X_1 = x_1, X_2 = x_2] > 0\). Then, the following holds.

\[
P[X_1 = x_1 | X_2 = x_2] = \frac{P[X_1 = x_1, X_2 = x_2]}{P[X_2 = x_2]}
\]

The random variables \(X_1\) and \(X_2\) are independent of each other if the distribution \((X_1 | X_2 = x_2)\) is identical to the random variable \(X_1\), for all \(x_2 \in \Omega_2\) such that \(P[X_2 = x_2] > 0\).
We can generalize the Bayes’ Rule as follows.

**Theorem (Chain Rule)**

Let \( (X_1, X_2, \ldots, X_n) \) 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

\[
P[X_1 = x_1, \ldots, X_n = x_n] = \prod_{i=1}^{n} P[X_i = x_i | X_{i-1} = x_{i-1} \ldots, X_1 = x_1]
\]
Let $S$ be the random variable representing whether I studied for my exam. This random variable has sample space $\Omega_1 = \{Y, N\}$.

Let $P$ 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 $(S, P)$ is represented in the next page.
Probability: First Example II

<table>
<thead>
<tr>
<th>$s$</th>
<th>$p$</th>
<th>$P[S = s, P = p]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>Y</td>
<td>1/2</td>
</tr>
<tr>
<td>Y</td>
<td>N</td>
<td>1/4</td>
</tr>
<tr>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>N</td>
<td>N</td>
<td>1/4</td>
</tr>
</tbody>
</table>
Here are some interesting probability computations
The probability that I pass.

\[
= \frac{1}{2} + 0 = \frac{1}{2}
\]
The probability that I study.


\[ = \frac{1}{2} + \frac{1}{4} = \frac{3}{4} \]
The probability that I pass conditioned on the fact that I studied.

\[
P[P = Y \mid S = Y] = \frac{P[P = Y, S = Y]}{P[S = Y]} = \frac{1/2}{3/4} = \frac{2}{3}
\]
Let $T$ be the time of the day that I wake up. The random variable $T$ has sample space $\Omega_1 = \{4, 5, 6, 7, 8, 9, 10\}$

Let $B$ represent whether I have breakfast or not. The random variable $B$ has sample space $\Omega_2 = \{T, F\}$

Our sample space is $\Omega = \Omega_1 \times \Omega_2$

The joint distribution of $(T, B)$ is presented on the next page
### Probability: Second Example II

<table>
<thead>
<tr>
<th>$t$</th>
<th>$b$</th>
<th>$P[T = t, B = b]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>T</td>
<td>0.03</td>
</tr>
<tr>
<td>4</td>
<td>F</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>T</td>
<td>0.02</td>
</tr>
<tr>
<td>5</td>
<td>F</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>T</td>
<td>0.30</td>
</tr>
<tr>
<td>6</td>
<td>F</td>
<td>0.05</td>
</tr>
<tr>
<td>7</td>
<td>T</td>
<td>0.20</td>
</tr>
<tr>
<td>7</td>
<td>F</td>
<td>0.10</td>
</tr>
<tr>
<td>8</td>
<td>T</td>
<td>0.10</td>
</tr>
<tr>
<td>8</td>
<td>F</td>
<td>0.08</td>
</tr>
<tr>
<td>9</td>
<td>T</td>
<td>0.05</td>
</tr>
<tr>
<td>9</td>
<td>F</td>
<td>0.05</td>
</tr>
<tr>
<td>10</td>
<td>T</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>F</td>
<td>0.02</td>
</tr>
</tbody>
</table>
What is the probability that I have breakfast conditioned on the fact that I wake up at or before 7?

Formally, what is $P[B = T | T \leq 7]$?