Lecture 04: 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 $f : \Omega \to \Omega'$ 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

$$P[f(X) = y] = \sum_{x \in \Omega : f(x) = y} P[X = x]$$
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 \((\mathcal{X}_1, \mathcal{X}_2)\) be a joint distribution over the sample space \(\Omega_1 \times \Omega_2\).

We can define the distribution \((\mathcal{X}_1 \mid \mathcal{X}_2 = x_2)\) as follows:

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

\[
P[\mathcal{X}_1 = x_1 \mid \mathcal{X}_2 = x_2] = \frac{P[\mathcal{X}_1 = x_1, \mathcal{X}_2 = x_2]}{\sum_{x \in \Omega_1} P[\mathcal{X}_1 = x, \mathcal{X}_2 = x_2]}\]

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.

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

The random variables \(X_1\) and \(X_2\) are independent of each other if the distribution \((X_1 \mid X_2 = x_2)\) 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

\[
P [X_1 = x_1, \ldots, X_n = x_n] = \prod_{i=1}^{n} P [X_i = x_i \mid X_{i-1} = x_{i-1} \ldots, X_1 = x_1]
\]
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 $X_1$ and then, conditioned on the fact that $X_1 = x_1$, it will state how to sample $X_2$. In such cases, we shall use the Bayes’ rule to calculate

$$
P[X_1 = x_1, X_2 = x_2] = P[X_1 = x_1] P[X_2 = x_2 | X_1 = x_1]
$$

- Let us consider an example.
  - Suppose $X_1$ is a random variable over $\Omega_1 = \{0, 1\}$ such that $P[X_1 = 0] = 1/2$. Next, the random variable $X_2$ is over $\Omega_2 = \{0, 1\}$ such that $P[X_2 = x_1 | X_1 = x_1] = 2/3$. Note that $X_2$ is biased towards the outcome of $X_1$.
  - What is the probability that we get $P[X_1 = 0, X_2 = 1]$?
To compute this probability, we shall use the Bayes’ rule.

\[ P[X_1 = 0] = 1/2 \]

Next, we know that

\[ P[X_2 = 0|X_1 = 0] = 2/3 \]

Therefore, we have \( P[X_2 = 1|X_1 = 0] = 1/3 \). So, we get

\[
P[X_1 = 0, X_2 = 1] = P[X_1 = 0] P[X_2 = 1|X_1 = 0] \\
= (1/2) \cdot (1/3) = 1/6
\]
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.

\[ P[P = Y] = P[S = Y, P = Y] + P[S = N, P = Y] = \frac{1}{2} + 0 = \frac{1}{2} \]
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 $\mathcal{T}$ be the time of the day that I wake up. The random variable $\mathcal{T}$ has sample space $\Omega_1 = \{4, 5, 6, 7, 8, 9, 10\}$

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

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

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

<table>
<thead>
<tr>
<th>$t$</th>
<th>$b$</th>
<th>$\mathbb{P}[T = t, B = b]$</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 $P[B = T | T \leq 7]$?
Consider the following experiment. I sequentially throw $m \ (< n)$ balls into $n$ bins uniformly and independently at random. What is the probability that there exists at least two balls that fall into the same bin?

We shall compute the probability of the complementary event. We shall compute the probability that all $m$ balls fall into distinct bins.

To compute this probability, we define the following event. Let $\mathcal{D}_i$ represent the event that the $i$-th ball falls into a bin that contains no other previous balls.

Note that the event $\mathcal{D}_i \wedge \mathcal{D}_{i-1} \wedge \cdots \wedge \mathcal{D}_1$ represents the event that the first $i$ balls fall in distinct bins.
We are interested in computing the following quantity

\[ P[D_m, D_{m-1}, \ldots, D_1] \]

Let us observe that the following estimate is correct

\[ P[D_i|D_{i-1}, D_{i-2}, \ldots, D_1] = \left(1 - \frac{i-1}{n}\right) \]

The reasoning is as follows. The conditioning \( D_{i-1}, D_{i-2}, \ldots, D_1 \) ensures that the first \((i-1)\) balls fall in distinct bins. We are interested in computing the probability that the \(i\)-th ball falls in a bin that is separate from these \((i-1)\) bins. So, there are \(n-(i-1)\) such bins. The probability that the \(i\)-th ball falls in these bins is \(\frac{n-(i-1)}{n}\).
By chain rule, we have

\[
P[D_m, \ldots, D_1] = \prod_{i=1}^{m} P[D_i | D_{i-1}, \ldots, D_1]
\]

\[
= \prod_{i=1}^{m} \left(1 - \frac{i - 1}{n}\right)
\]

\[
= \left(1 - \frac{0}{n}\right) \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{m - 1}{n}\right)
\]

Next, our objective is to estimate the expression

\[
P = \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{m - 1}{n}\right)
\]
We can re-write this expression as

\[ P = \prod_{i=2}^{m} \left(1 - \frac{i-1}{n}\right) \]

\[ = \prod_{i=2}^{m} \exp \ln \left(1 - \frac{i-1}{n}\right) \]

\[ = \exp \sum_{i=2}^{m} \ln \left(1 - \frac{i-1}{n}\right) \]

We shall use the estimate
Claim

For any $\varepsilon \in [0, 1/2]$ and integer $k \geq 2$, we have

$$\frac{\varepsilon^2}{2} - \frac{\varepsilon^k}{k} - \frac{\varepsilon^k}{k} \leq \ln(1 - \varepsilon) \leq -\varepsilon - \frac{\varepsilon^2}{2} - \frac{\varepsilon^k}{k}$$

Using $k = 2$, we obtain $-\varepsilon - \varepsilon^2 \leq \ln(1 - \varepsilon) \leq -\varepsilon - \varepsilon^2/2$.

Let us obtain an upper-bound

$$P = \exp \sum_{i=2}^{m} \ln \left(1 - \frac{i - 1}{n}\right)$$

$$\leq \exp \left(\sum_{i=2}^{m} -\frac{i - 1}{n} - \frac{(i - 1)^2}{2n^2}\right)$$

$$= \exp \left(-\frac{(m - 1)m}{2n} - \frac{(m - 1)(m - 1/2)m}{6n^2}\right)$$
Similarly, we can obtain the lower-bound

\[
P = \exp \sum_{i=2}^{m} \ln \left( 1 - \frac{i-1}{n} \right)
\]

\[
\geq \exp \left( \sum_{i=2}^{m} - \frac{i-1}{n} - \frac{(i-1)^2}{n^2} \right)
\]

\[
= \exp \left( - \frac{(m-1)m}{2n} - \frac{(m-1)(m-1/2)m}{3n^2} \right)
\]

Note that at \( m = \Theta(\sqrt{n}) \) the probability \( P \) transitions from 0.01 to 0.99