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

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

- 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 $\mathbb{X}_2$

- 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 $\mathbb{P}[\mathbb{X}_2 = x_2] > 0$

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 <u>biased towards</u> 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

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

$$\begin{aligned}\mathbb{P}\left[\mathbb{P} = \mathsf{Y}\right] &= \mathbb{P}\left[\mathbb{S} = \mathsf{Y}, \mathbb{P} = \mathsf{Y}\right] + \mathbb{P}\left[\mathbb{S} = \mathsf{N}, \mathbb{P} = \mathsf{Y}\right] \\ &= 1/2 + 0 = 1/2\end{aligned}$$

Probability: First Example IV

The probability that I study.

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

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 $\Omega_2 = \{\mathsf{T}, \mathsf{F}\}$
- 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 $\mathbb{P}\left[\mathbb{B} = \mathsf{T} \mid \mathbb{T} \leqslant 7\right]$?

Birthday Bound I

- 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 $\mathbb{D}_i$ represent the event that the i-th ball falls into a bin that contains no other previous balls.
- Note that the event

$$\mathbb{D}_i \text{ and } \mathbb{D}_{i-1} \text{ and } \cdots \text{ and } \mathbb{D}_1$$

represents the event that the first i balls fall in distinct bins.

- We are interested in computing the following quantity

$$\mathbb{P}\left[\mathbb{D}_m, \mathbb{D}_{m-1}, \ldots, \mathbb{D}_1\right]$$

- Let us observe that the following estimate is correct

$$\mathbb{P}\left[\mathbb{D}_i | \mathbb{D}_{i-1}, \mathbb{D}_{i-2}, \ldots, \mathbb{D}_1\right] = \left(1 - \frac{i-1}{n}\right)$$

  The reasoning is as follows. The conditioning $\mathbb{D}_{i-1}, \mathbb{D}_{i-2}, \ldots, \mathbb{D}_1$ 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 $\frac{n-(i-1)}{n}$.

Birthday Bound III

- By chain rule, we have

$$\mathbb{P}\left[\mathbb{D}_m, \ldots, \mathbb{D}_1\right] = \prod_{i=1}^{m} \mathbb{P}\left[\mathbb{D}_i | \mathbb{D}_{i-1}, \ldots, \mathbb{D}_1\right]$$
$$= \prod_{i=1}^{m} \left(1 - \frac{i-1}{n}\right)$$
$$= \left(1 - \frac{0}{n}\right)\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\cdots\left(1 - \frac{m-1}{n}\right)$$

- 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

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

We shall use the estimate

Claim

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

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

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

- Let us obtain an upper-bound

$$\begin{aligned} P &= \exp \sum_{i=2}^{m} \ln\left(1 - \frac{i-1}{n}\right) \\ &\leqslant \exp\left(\sum_{i=2}^{m} -\frac{i-1}{n} - \frac{(i-1)^2}{2n^2}\right) \\ &= \exp\left(-\frac{(m-1)m}{2n} - \frac{(m-1)(m-1/2)m}{6n^2}\right) \end{aligned}$$

Birthday Bound VI

- Similarly, we can obtain the lower-bound

$$
\begin{aligned}
P &= \exp \sum_{i=2}^{m} \ln \left(1 - \frac{i-1}{n}\right) \\
&\geqslant \exp \left(\sum_{i=2}^{m} -\frac{i-1}{n} - \frac{(i-1)^2}{n^2} \right) \\
&= \exp \left(-\frac{(m-1)m}{2n} - \frac{(m-1)(m-1/2)m}{3n^2} \right)
\end{aligned}
$$

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