

Lecture 01: Mathematical Basics (Probability)

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 $f: \Omega \rightarrow \Omega'$ 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

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

- 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 Ω_1 , and the second coordinate lies in Ω_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.

Joint Distribution and Marginal Distributions II

- Let $P_1: \Omega_1 \times \Omega_2 \rightarrow \Omega_1$ 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

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

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

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.*

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

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

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

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 $\mathbb{X}_1$ 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

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

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

Important: Why use Bayes' Rule II

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

$$\mathbb{P} [X_1 = 0] = 1/2$$

Next, we know that

$$\mathbb{P} [X_2 = 0|X_1 = 0] = 2/3$$

Therefore, we have $\mathbb{P} [X_2 = 1|X_1 = 0] = 1/3$. So, we get

$$\begin{aligned}\mathbb{P} [X_1 = 0, X_2 = 1] &= \mathbb{P} [X_1 = 0] \mathbb{P} [X_2 = 1|X_1 = 0] \\ &= (1/2) \cdot (1/3) = 1/6\end{aligned}$$

Independence of Random Variables

- Consider a joint distribution $(\mathbb{X}_1, \mathbb{X}_2)$ over the sample space $\Omega_1 \times \Omega_2$
- The marginal distributions $\mathbb{X}_1$ and $\mathbb{X}_2$ are independent of each other, if for all $x_1 \in \Omega_1$ and $x_2 \in \Omega_2$ we have: If $\mathbb{P}[\mathbb{X}_1 = x_1] > 0$ then

$$\mathbb{P}[\mathbb{X}_2 = x_2] = \mathbb{P}[\mathbb{X}_2 = x_2 | \mathbb{X}_1 = x_1] .$$

- Equivalently, the following condition is satisfied

$$\mathbb{P}[\mathbb{X}_1 = x_1] \cdot \mathbb{P}[\mathbb{X}_2 = x_2] = \mathbb{P}[\mathbb{X}_1 = x_1, \mathbb{X}_2 = x_2] .$$

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	$\mathbb{P}[S = s, P = p]$
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.

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

Probability: First Example IV

The probability that I study.

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

Probability: First Example V

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

$$\begin{aligned}\mathbb{P}[\mathbb{P} = Y \mid S = Y] &= \frac{\mathbb{P}[\mathbb{P} = Y, S = Y]}{\mathbb{P}[S = Y]} \\ &= \frac{1/2}{3/4} = \frac{2}{3}\end{aligned}$$

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	$\mathbb{P}[T = t, \mathbb{B} = b]$
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 $\mathbb{P}[\mathbb{B} = \mathbb{T} \mid \mathbb{T} \leq 7]$?