

1. Review of Probability

Vaughan Sohn

October 5, 2024

Learn from Example: Card game

Random variable and Probability distribution

Chain rule

Bayes rule

Independence

Learn from Example: Card game

Example

- 다음과 같은 3종류의 카드가 존재한다고 가정하자.
 - Red / Red
 - Red / Black
 - Black / Black
- 눈을 감고 하나의 카드를 고른 뒤, 그 카드의 한쪽 면을 확인한다.
- Question: 내가 고른 카드의 한쪽 면의 색상이 "Red" 일 때, 다른 면의 색상도 "Red" 일 확률은 얼마인가?

Prior knowledge

✓이 문제를 풀기 위해서는 "Conditional probability"의 개념에 대해서 이해해야한다.
지금부터 하나하나 확률의 개념들에 대해서 복습하면서 이 예제 문제를 풀어보도록 하자.

Random variable and Probability distribution

Definition 1 (random variable)

A **random variable** is a real-valued *function* of the outcome of the experiment.

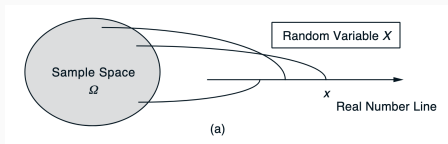
$$X : \Omega \rightarrow \mathbb{R}$$

Definition 2 (sample space, atom)

A **sample space** Ω is a set of all possible outcomes, and the elements of that set are called **atoms**.

✓ meaning: Random variable은 *numerical value*가 아닐 수도 있는 실험 결과(e.g., 동전 앞면이 나왔다.) 를 numerical value로 할당해준다.

- Outcome은 어떠한 **확률**에 의하여 그 값이 결정되는 실험의 결과를 의미한다.
- Random variable과 sample space 그리고 atom간의 관계는 다음과 같다.



Definition 3 (probability law)

Probability distribution assigns to a set A of possible outcomes a *nonnegative* number $P(A)$.

✓ meaning: Random variable이 실험 결과를 numerical value로 할당하는 것처럼, probability distribution은 numerical value를 $(0, 1)$ 사이의 양의 실수값으로 매핑해준다.

$$P : X \rightarrow [0, 1]$$

- 전체 sample space에 대한 확률 분포의 합은 1이 되어야한다.

$$\sum_{x \in \mathcal{X}} P(X = x) = 1$$

- If sample space is finite, we call a probability law as **discrete probability distribution**.

Definition 4 (expectation)

An **expectation** of random variable X is

$$\mathbb{E}[X] = \sum_{x \in \mathcal{X}} x \cdot P_X(x).$$

Definition 5 (variance and standard deviation)

An **variance** of random variable X is

$$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2].$$

And we called the square root of the variance as **standard deviation** of X .

$$\sigma_X = \sqrt{\text{Var}(X)}$$

- X 의 variance는 항상 nonnegative이다.
- Variance는 실제 outcome들이 expectation을 기준으로 얼마나 산재되어있는지를 나타내는 척도이다.

Definition 6 (event)

Event is only a **subset of atoms**.

Theorem 7 (probability of the event)

For any event E , we have

$$P(X \in E) = \sum_{x \in E} P_X(x).$$

✓ meaning: Event는 어떤 특수한 조건을 만족시키는 atom들의 집합일 뿐이다.

- 가능한 모든 event들의 집합을 \mathcal{F} 라고 하면, $|\mathcal{F}|$ 는 $2^{|\Omega|}$ 이다.
- Example: r.v.가 $D \in \{1, 2, 3, 4, 5, 6\}$ 일 때, 주사위를 던졌을 때 그 결과가 3보다 큰 event의 확률은?

$$P(D > 3) =$$

Theorem 8 (union bound)

$$\begin{aligned}P(E_1 \cup E_2) &= P(E_1) + P(E_2) - P(E_1 \cap E_2) \\&= P(E_1 \cap E_2^c) + P(E_1 \cap E_2) + P(E_2 \cap E_1^c) \\&\leq P(E_1) + P(E_2)\end{aligned}$$

- Event를 sample space에 있는 집합으로 생각하면 쉽게 이해할 수 있다.
 \Rightarrow
- application: 실제로 계산하기 어려운 $P(E_1 \cup E_2)$ 대신 $P(E_1) + P(E_2)$ 가 0으로 수렴함을 보여서 좌변도 0으로 수렴함을 증명하는 방식으로 활용한다.

Let X and Y be random variables associated with the *same experiment*, then the sample space of experiment is $\mathcal{X} \times \mathcal{Y}$.

Definition 9 (joint distribution)

The **joint probability distribution** of X and Y is defined by

$$P_{X,Y}(x, y) = P(X = x, Y = y).$$

Definition 10 (marginalization)

The **marginal probability distribution** of X and Y can be obtained from the joint probability distribution,

$$P_X(x) = \sum_{y \in \mathcal{Y}} P_{X,Y}(x, y), \quad P_Y(y) = \sum_{x \in \mathcal{X}} P_{X,Y}(x, y)$$

Conditional distribution

Let X and Y be random variables associated with the *same experiment*, and we consider a probability distribution when given some **condition**.

Definition 11 (conditional distribution)

The **conditional probability distribution** of X **given event** $E; Y = y$ is defined by

$$P_{X|Y}(x|y) = P(X = x|Y = y)$$

Theorem 12

The conditional PMF[] of X given $Y = y$ is related to the joint PMF by*

$$P_{X|Y}(x|y) = \frac{P_{X,Y}(x, y)}{P_Y(y)}$$

✓ meaning: conditional probability는 given event **set** 안에서 각 outcome의 확률을 나타낸다고 생각하면 이해하기 쉽다.

⇒

- 현재 우리가 관심있는 r.v. X (i.e., | 왼쪽에 위치한 변수)에 대한 conditional PMF의 합은 1이다. 즉, 여전히 확률로서 동작한다.

$$\sum_{x \in \mathcal{X}} P_{X|Y}(x|y) = 1$$

- 그러나 r.v. Y 에 대해서는 그 합이 반드시 1이 되지는 않는다.

$$\sum_{y \in \mathcal{Y}} P_{X|Y}(x|y) \neq 1 \text{ at least, not necessarily.}$$

Example

- 다음과 같은 3종류의 카드가 존재한다고 가정하자.
 - Red / Red
 - Red / Black
 - Black / Black
- 눈을 감고 하나의 카드를 고른 뒤, 그 카드의 한쪽 면을 확인한다.
- Question: 내가 고른 카드의 한쪽 면의 색상이 "Red" 일 때, 다른 면의 색상도 "Red" 일 확률은 얼마인가?

Solution:

- 다음과 같이 2개의 r.v.를 정의하자.
 - 첫번째로 확인한 카드의 색상: $X_1 \in \{R, B\}$
 - 카드의 반대쪽 면의 색상: $X_2 \in \{R, B\}$
- 그렇다면, 눈을 감고 고른 카드의 색상을 확인하는 실험결과가 "Red" 라는 것을 알고있기 때문에 $X_1 = R$ 임을 알 수 있다.
- 즉, 우리는 다음의 확률을 계산하면 된다.

$$P_{X_2|X_1}(R|R) = \frac{P_{X_1, X_2}(R, R)}{P_{X_1}(R)} =$$

Chain rule

Theorem 13 (chain rule)

By definition of conditional PMF, we can write joint PMF as

$$\begin{aligned} P_{X,Y}(x,y) &= P_{X,Y}(x,y) \frac{P_Y(y)}{P_Y(y)} = P_{X|Y}(x|y)P_Y(y) \\ &= P_{X,Y}(x,y) \frac{P_X(x)}{P_X(x)} = P_{Y|X}(y|x)P_X(x) \end{aligned}$$

In general, for any set of N variables

$$P(x_1, \dots, x_N) = \prod_{n=1}^N P(x_n | x_1, \dots, x_{n-1}).$$

Theorem 14 (marginalization)

By chain rule, we can get probability distribution on some r.v. from joint PMF as

$$P_X(x) = \sum_{y \in \mathcal{Y}} \sum_{z \in \mathcal{Z}} P_{X,Y,Z}(x, y, z)$$

* Proof:

Bayes rule

Theorem 15 (Bayes rule)

From the chain rule and marginalization, we get bayes rule

$$P(x|y) = \frac{P(y|x)P(x)}{\sum_x P(y|x)P(x)}$$

✓ meaning: 사전확률 $P(y|x), P(x)$ 로부터 사후확률을 추정할 수 있다.

applications:

- Let Y be a symptom, X be a disease and interpret the Bayes rule.
- Let Y data and X what we want to infer from the data.

Independence

Definition 16 (independence)

Random variables X and Y are **independent** $X \perp Y$ if know about the value of X (i.e., conditioning event ($X = x$)) tell us *nothing* about Y .

$$P_{Y|X}(y|x) = P_Y(y)$$

Definition 17 (conditionally independence)

Random variables X and Y are **conditionally independent** if know about the value of X tell us *nothing* about Y **given** Z .

$$P_{Y|X,Z}(y|x,z) = P_{Y|Z}(y|z)$$

- 만약 두 r.v.가 independence라면, joint PMF를 인수분해 할 수 있다. [*]

$$P_{X,Y}(x,y) = P_X(x)P_Y(y)$$

- 만약 두 r.v.가 conditionally independence라면, joint PMF를 인수분해 할 수 있다.

$$P_{X,Y|Z}(x,y|z) = P_{X|Z}(x|z)P_{Y|Z}(y|z)$$

Example of Independence and Conditionally Independence

Example

- 2개의 코인 C_1, C_2 가 있을 때, 각 코인에 대한 PMF가 다음과 같다.
 - $P_{C_1}(H) = 0.5, \quad P_{C_1}(T) = 0.5$
 - $P_{C_2}(H) = 0.7, \quad P_{C_2}(T) = 0.3$
- 즉, C_1 은 fair coin이고 C_2 는 앞면의 확률이 더 높은 unfair coin이다.
- 두 coin중에서 어떤 coin을 사용할 것인지를 랜덤하게 결정하기 위한 랜덤변수 $Z \in \{1, 2\}$ 를 가정하자. ($Z \sim U$)
- Z 의 value에 의해 결정된 coin을 2번 flip하고 첫번째 결과에 대한 랜덤변수 X , 두번째 결과에 대한 랜덤변수 Y 를 정의하자.
- Question 1: $X \perp Y$?
- Question 2: $X \perp Y$, given Z ?

Solution:

Notations

- sample space: Ω
- random variable: X, Y, Z, \dots
- numerical value: x, y, z, \dots
- probability for arbitrary outcome: $P(X = x) = P_X(x)$
- expectation of random variable X : $\mathbb{E}[X]$

- D. Bertsekas and J. Tsitsikilis, Introduction to Probability
- Lecture notes for EE623: Information Theory (Fall 2024)