

수포자도 도전해 볼 만한

# Mathematics in DeepLearning

---

## Lecture5. Probability and Statistics

Juhee Lee Ph.D.

Research Professor @ ewha womans university

# Probability Theory

Based on Set theory

- Set, universal set, element

# Set Theory & Probability

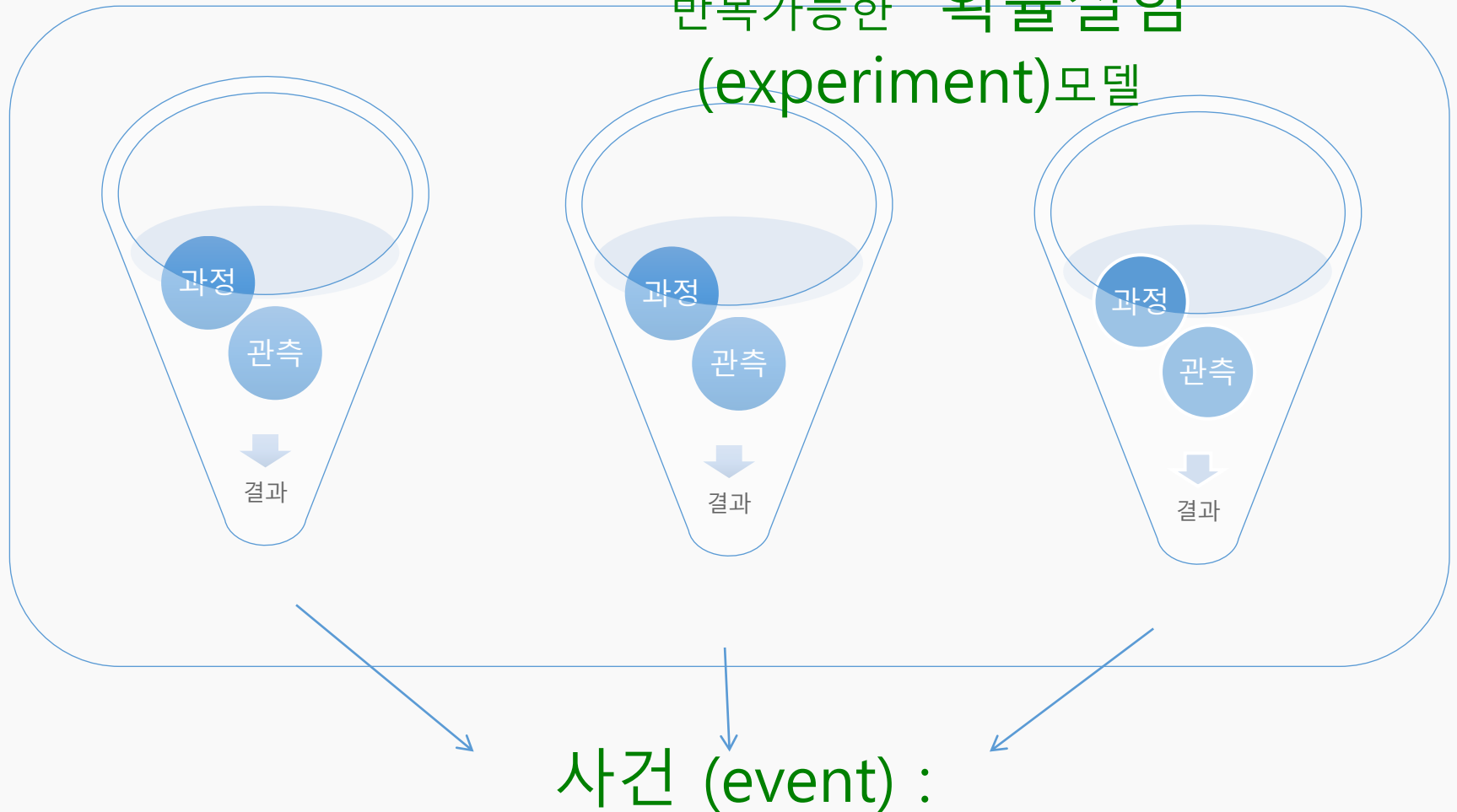
Based on Set theory

집합론	VS	확률론
집합		사건
전체 집합		표본공간
원소		결과

- Set, universal set, element
- event, sample space, outcome
  - Event: set of outcomes
  - Outcome: being an observation
  - Experiment(Procedure, observation)

# 확률은

반복가능한 확률실험  
(experiment)모델



# Probability

Sample space :  $\Omega$

Event

Experiment

# Probability

Definition **Sample space (or State space)  $\Omega$**  is the collection of all possible outcomes under consideration.

Ex. 동전 던지기(1개)

$$\Omega = \{H, T\}$$

Ex. 동전 던지기(2개)

$$\Omega = \{(H, H), (H, T), (T, H), (T, T)\}$$

# Probability

Definition An **Event** is a subset  $A \subset \Omega$  of the sample space

Ex.  $\Omega = \{(H, H), (H, T), (T, H), (T, T)\}$

$$A = \{(H, T), (T, H)\} \subset \Omega$$

# Probability

Definition A set function  $\mathbf{P}$  defined on the set of subsets of  $\Omega$  is called a probability measure if it satisfies these 3 conditions for  $A, B \subset \Omega$

$$(i) \quad P(A) \geq 0$$

$$(ii) \quad P(\emptyset) = 0, P(\Omega) = 1$$

$$(iii) \quad A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$$



# Review of Real analysis

Definition A collection  $\mathcal{F}$  of subsets of a sample space  $\Omega$  is called a  **$\sigma$ -algebra**

If

$$(i) \ \emptyset \in \mathcal{F}$$

$$(ii) \ A^c = \Omega \setminus A \in \mathcal{F} \text{ for any } A \in \mathcal{F}$$

$$(iii) \ A_n \in \mathcal{F}, \ n = 1, 2, \dots \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$$

$$\text{Ex. } \Omega = \{a, b\}, \mathcal{F} = \{\emptyset, \{a\}, \{b\}, \{a, b\}\}$$

Definition A pair  $(\Omega, \mathcal{F})$  is called a **measurable space** if  $\Omega$  is a non-empty set and  $\mathcal{F}$  is a  $\sigma$ -algebra on  $\Omega$

A subset  $A \subset \Omega$  is said to be measurable if  $A \in \mathcal{F}$

$$A_n \in \mathcal{F}, n = 1, 2, \dots \Rightarrow \bigcap_{n=1}^{\infty} A_n \in \mathcal{F}$$

$(\mathcal{F}_i)_{i \in I}$  : a family of  $\sigma$ -algebra on  $\Omega$

$\bigcap_{i \in I} \mathcal{F}_i = \{A \subset \Omega: A \in \mathcal{F}_i, \forall i \in I\}$  is a  $\sigma$ -algebra on  $\Omega$

For any family  $\mathcal{O}$  of subsets of  $\Omega$   
there is a smallest  $\sigma$ -algebra  $\sigma(\mathcal{O})$   
such that  $\mathcal{O} \subset \sigma(\mathcal{O})$

$\mathcal{B}(R^n)$ : Borel set in  $R^n$

- the smallest  $\sigma$ -algebra generated by the open sets in  $R^n$
- 측정가능한 집합

# Probability space

Definition Let  $(\Omega, \mathcal{F})$  be a measurable space

A mapping  $\mu : \mathcal{F} \rightarrow \mathbf{R} \cup \{\infty\}$  is called a **measure**

If

(i)  $\mu(A) \geq 0$

(ii)  $\mu(\emptyset) = 0$

(iii) For any  $A_1, \dots, A_{n_\infty}$  of mutually disjoint sets in  $\mathcal{F}$ ,

$$\mu \left( \bigcup_{n=1}^{\infty} A_n \right) = \sum_{n=1}^{\infty} \mu(A_n)$$

$(\mu, \Omega, \mathcal{F})$ : measure space

# Random Variable

Definition Let  $(\mu, \Omega, \mathcal{F})$ : measure space.

A function  $X : \Omega \rightarrow \mathbf{R}^n$  is said to be **measurable** if  $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}$  for any  $B \in \mathcal{B}(\mathbf{R}^n)$

$(\Omega, \mathcal{F}, P)$ : probability space

Definition A real measurable function  $X : (\Omega, \mathcal{F}) \rightarrow (\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n))$  is called a **random variable**

$(\Omega, \mathcal{F}, P)$ : probability space

Definition Two events  $A, B \in \mathcal{F}$  are said to be **independent** if

$$P(A \cap B) = P(A)P(B)$$

Ex.  $\Omega = \{(H, H), (H, T), (T, H), (T, T)\}$ ,  $H, T \rightarrow \frac{1}{2}$

$$A = \{(H, H), (H, T)\}, B = \{(H, T), (T, T)\}$$

$$A \cap B = \{(H, T)\}$$

$$P(A) = \frac{1}{2}, P(B) = \frac{1}{2} \quad P(A \cap B) = \frac{1}{4}$$

$(\Omega, \mathcal{F}, P)$ : probability

Definition  $n$  events  $A_1, \dots, A_n \in \mathcal{F}$  are said to be **not independent** even if

$$P(A \cap A_2 \cap A_n) = P(A)P(A_2)P(A_n)$$

Ex.  $\Omega$  = 두개의 주사위던질때

$A$  = 첫번째 주사위  $\{1,2,3\}$

$B$  = 두번째 주사위  $\{3,4,5\}$

$C$  = 두 결과의 합이 9

$$P(A \cap B \cap C) = P(A)P(B)P(C)$$

$(\Omega, \mathcal{F}, P)$ : probability

Definition **Let**  $X$  and  $Y$  be two random variable on  $\Omega$

We say that  $X$  and  $Y$  are independent if

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

$$\Leftrightarrow$$

$$P(\{X \in A\} \cap \{Y \in B\}), \quad \forall A, B \in \mathcal{B}(R)$$



$$\text{Ex. } \Omega = \{(H, H), (H, T), (T, H), (T, T)\}$$

$$w = \{w_1, w_2\} \in \Omega, w_i \in \{H, T\}, i = 1, 2$$

$$X(w) = 1 \text{ if } w_1 = H, 0 \text{ o.w}$$

$$Y(w) = 1 \text{ if } w_2 = H, 0 \text{ o.w}$$

=> Independent

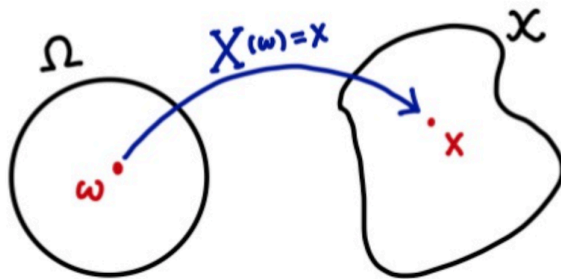
$$P(X = 0, Y = 1) = P(X^{-1}(1) \cap Y^{-1}(0)) = P(A \cap B) = P(A)P(B) = P(X = 0)P(Y = 1)$$

$$A = \{(H, H), (H, T)\}, B = \{(H, T), (T, T)\}$$

Check  $P(X = 0, Y = 0)P(X = 1, Y = 1)P(X = 1, Y = 0)$

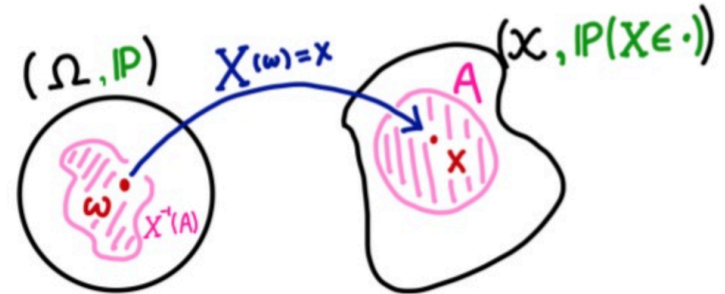
확률변수  $X$  는  $\omega$  를  $\mathcal{X}$  상의 원소  $x$  로 사상(mapping) 하는 함수입니다

$$X : \omega \mapsto x \in \mathcal{X}$$



확률분포는  $X$  가 어떻게 mapping 되는지 보여주는  $\mathcal{X}$  상의 측도입니다

$$\mathbb{P}(X^{-1}(A)) = \mathbb{P}(X \in A)$$



그리고 확률변수  $X$  는 **H** 와 **T** 를 각각 1 과 0 으로 보내는 걸로 정의합니다

$$X(\mathbf{H}) = 1, \quad X(\mathbf{T}) = 0$$

이 때 확률분포는 다음과 같이  $\mathcal{X} = \{0, 1\}$  위에 똑같이 정의됩니다

$$\mathbb{P}(X = 1) = p, \quad \mathbb{P}(X = 0) = 1 - p$$

고로 확률측도는 사실 확률분포와 전혀 다르지 않습니다 😊

$$\mathbb{P}(\mathbf{H}) = \mathbb{P}(X = 1), \quad \mathbb{P}(\mathbf{T}) = \mathbb{P}(X = 0)$$

확률분포를 다른 용어로  $X$  의 법(law of  $X$ ) 이라고도 부릅니다

단지 확률값의 측정을 어떤 공간에서 하는지, 그리고 확률변수가 묵시적으로 정의되어 있다는 점에서 차이가 있을 뿐입니다. 앞으로는 두 용어를 혼용해서 쓰겠습니다.

$(\Omega, \mathcal{F}, P)$ : probability space

Definition The **conditional probability**  $P(A|B)$  of  $A$  given  $B$  is defined by

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \text{ for } \forall A \in \mathcal{F}$$

If  $A$  and  $B$  are independent

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = P(A)$$

# Distribution (law)

$(\Omega, \mathcal{F}, P)$ : probability space

Definition The distribution

(law)  $P_X$  of a random variable  $X: (\Omega, \mathcal{F}) \rightarrow (R^n, \mathcal{B}(R^n))$  is given by

$$P_X(B) = P(X \in B) = P(\{\omega \in \Omega: X(\omega) \in B\})$$

For each Borel set  $B$

$P_X$  is also probability measure on  $(\Omega, \mathcal{F})$

The distribution  $F_X$  is defined by

$$F_X(x) = P_X((-\infty, x]) = P(X \leq x), x \in R$$

$$E_X. \Omega = \{(H, H), (H, T), (T, H), (T, T)\}$$

$$Z = X + Y \in \{0, 1, 2\}, F_Z(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{4}, & 0 \leq x < 1 \\ \frac{3}{4}, & 1 \leq x < 2 \\ 1, & x \geq 2 \end{cases}$$

$X_1, X_2, \dots, X_n$ : random variables

Definition The vector of random variables  $X = (X_1, X_2, \dots, X_n)$  is called a random vector.

The distribution function  $F_X$  is defined by  $P(X_1 \leq x_1, \dots, X_n \leq x_n)$  for  $X = (X_1, X_2, \dots, X_n) \in R^n$

Theorem. The random variables  $X_1, X_2, \dots, X_n$  are independent  $\Leftrightarrow F_{X_1}(x_1) \cdots F_{X_n}(x_n)$

# Expectation

$(\Omega, \mathcal{F}, P)$ : probability space

$$\Omega = \{w_1, w_2, \dots, w_n\}, P(w_i) = P_i, i = 1, \dots, n$$

Definition  $X: \Omega \rightarrow R$  a random

vector,  $P(\Omega) = \sum_{i=1} P_i = 1$

The expectation of  $X$  is defined by

$$E(X) = \sum_{i=1}^n P_i X(w_i) = \sum_{i=1}^n P_i x_i .$$

$$X(w_i) = x_i \text{ for } i = 1, \dots, n$$

# Expectation

Ex. Two coin,  $X$ : the number of heads  
 $E(X)=?$

- Expectation is a linear functional

$$E(aX + bY) = aE(X) + bE(Y)$$



# Expectation

We may assume that  $x_1, \dots, x_n$  for random variable  $X$  such that  $x_1 < \dots < x_n$

The distribution function  $F_X$  is the given by

$$F_X(x_i) = P_1 + \dots + P_{i-1} \quad \text{for } x_{i-1} < x < x_i$$

$$\text{with } F_X(x) = \begin{cases} 0, & x < x_1 \\ 1, & x \geq x_n \end{cases}$$

In particular,  $P_i = F_X(x_i) - F_X(x_{i-1}), i = 2, \dots, n,$   
 $P_1 = F_X(x_1)$

$$E(X) = \sum_{i=1}^n P_i x_i = x_1 F_X(x_1) + x_2 (F_X(x_2) - F_X(x_1)) + \dots$$

# Expectation

$$E(X) = \sum_{i=1}^n P_i x_i = x_1 F_x(x_1) + x_2 (F_x(x_2) - F(x_1)) \\ + \cdots + x_n (F_x(x_n) - F_x(x_{n-1}))$$

Take,  $x_0 < x_1$  and let  $F_X: [x_0, X_n] \rightarrow R$  with

$$f_X(x_i) = \frac{F_x(x_n) - F_x(x_{n-1})}{x_i - x_{i-1}}, \quad F_x(x_0) = 0$$

# Expectation

$$\begin{aligned} E(X) &= x_1 F_X(x_1) + \sum_{i=2}^n x_i \frac{F_x(x_n) - F_x(x_{n-1})}{x_i - x_{i-1}} (x_i - x_{i-1}) \\ &= \int x f_x(x) dx \end{aligned}$$

# Density function

$(\Omega, \mathcal{F}, P)$ : probability space , *random variable*  $X$

Definition If the distribution function  $F_X$  is differentiable , then the expectation of  $X$  can be written by

$$E(X) = \int_{-\infty}^{\infty} x f_x(x) dx \text{ where } f_x(x) = \frac{d}{dx} F_X(x)$$

We call  $f_x$  the **density function** for  $X$

$f_x(x)$ : non-negative,  $\int_{-\infty}^{\infty} f_x(x) dx = 1$

$$F_X(x) = \int_{-\infty}^x f_x(x) dx$$

Consider indep. Rv  $X$  and  $Y$  with finitely many values  $x_1, \dots, x_n$  and  $y_1, \dots, y_m$  Then

$$E(XY) = \sum_{i=1}^n \sum_{j=1}^m x_i y_j P(X = x_i, Y = y_j) = E(X)E(Y)$$

Definition The variance  $var(X)$  of a random variable  $X$  is defined by

$$var(X) = E(X - E(X))^2 = E(X^2) - E(X)^2$$

$X$  : random variable

$$g(t) = (t - E(X))^2$$

$$var(X) = E(X - E(X))^2 = E(g(x)) = \int_{-\infty}^{\infty} g(x) f_x(x) dx$$

# Higher moments

Definition a random variable  $X$  with density function  $f$

$$E(x^n) = \int_{-\infty}^{\infty} x^n f_x(x) dx$$

Is called the  $n$ -th moment of  $X$ , where  $n \in \mathbb{N}$

$$\int_{-\infty}^{\infty} g(x) f_x(x) dx$$

$XE(X)$ : central moment of  $X$ ,  $\sigma = \sqrt{\text{Var}(X)}$

$Y = (X - E(X))/\sigma$  's moment : standard moment's

$$E(Y) = 0 \quad E(Y^2) = 1,$$

Definition *The covariance*

*$Cov(X, Y)$  of two random variable  $X, Y: \Omega \rightarrow R$  is defined by*

$$Cov(X, Y) = E((X - E(X))(Y - E(Y))) = E(XY) - E(X)E(Y)$$

$X, Y$ : independent,  $Cov(X, Y) = 0$



# Statistics

- Probability theory based on Set theory
- A branch of mathematics dealing with the collection, analysis, interpretation, presentation, and organization of data

Q & A