

## Lecture A2. Probability Review

Sim, Min Kyu, Ph.D., [mksim@seoultech.ac.kr](mailto:mksim@seoultech.ac.kr)



서울과학기술대학교 데이터사이언스학과

# 1 I. Probability

## 2 II. Random Variables

## 3 III. Uniform

## 4 IV. Exponential

## 5 V. Poisson

## 6 VI. Some Exercises

# I. Probability

# Probability

## Definition 1

Probability is a function from “Event” to real number between 0 and 1.

$$\mathbb{P} : S \rightarrow [0, 1],$$

where  $S$  is the set of all possible events.

## Remark 1

- For any event  $E$ ,  $0 \leq \mathbb{P}(E) \leq 1$ .
- For  $E = \emptyset$ ,  $\mathbb{P}(E) = 0$ .
- $\mathbb{P}(S) = 1$ , where  $S$  is whole space, or a space for all possible events.

## Remark 2

- $\mathbb{P}(E_1 \cup E_2) = \mathbb{P}(E_1) + \mathbb{P}(E_2) - \mathbb{P}(E_1 \cap E_2)$ .
- If  $E_1 \cap E_2 = \emptyset$ , then  $\mathbb{P}(E_1 \cup E_2) = \mathbb{P}(E_1) + \mathbb{P}(E_2)$ .

# Conditional Probabilities

## Definition 2

$\mathbb{P}(E|F)$ , “probability of  $E$  given  $F$ ”, is the probability that event  $E$  occurs given that  $F$  has occurred. In a math notation,

$$\mathbb{P}(E|F) = \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(F)}$$

- Ex) (from Ross) Suppose cards numbered one through ten are placed in a hat, mixed up, and then one of the cards is drawn. If we are told that the number on the drawn card is at least five, then what is the conditional probability that it is ten?

# Independent Probability

## Definition 3

If  $\mathbb{P}(E) = \mathbb{P}(E|F)$ , then we say the events  $E$  and  $F$  are independent events.

- That is, the occurrence of  $E$  happening has nothing to do with the occurrence of  $F$ .
- Just as Galilei once said "*And yet it rotates (even though people do not believe so)*".  
People believing whether or not the earth rotates is independent event of earth's rotation.

# Properties

- ❶  $\mathbb{P}(E_1 \cup E_2) = \mathbb{P}(E_1) + \mathbb{P}(E_2)$ , if  $E_1 \cap E_2 = \emptyset$ .
- ❷  $\mathbb{P}(E) = \mathbb{P}(E \cap F) + \mathbb{P}(E \cap F^c)$  ( $\because (E \cap F) \cap (E \cap F^c) = \emptyset$ )
- ❸ If  $F_i \cap F_j = \emptyset$  for all  $i \neq j$  and  $\mathbb{P}(F_1 \cup \dots \cup F_n) = 1$  where  $1 \leq i, j \leq n$ , then  $\mathbb{P}(E) = \sum_{i=1}^n \mathbb{P}(E \cap F_i)$



# Bayes' rule

## Theorem 1

Suppose  $F_1, \dots, F_n$  are **mutually exclusive** events (or,  $F_i \cap F_j = \emptyset$ , or,  $\mathbb{P}(F_i \cap F_j) = 0$ , for any  $i \neq j$ ) and  $\mathbb{P}(F_1 \cup \dots \cup F_n) = 1$  (In other words, exactly one and only one events among  $F_1, \dots, F_n$  will occur.), then the following holds.

$$\begin{aligned} \mathbb{P}(E) &= \mathbb{P}(E \cap F_1) + \dots + \mathbb{P}(E \cap F_n) \\ &= \sum_{i=1}^n \mathbb{P}(E \cap F_i) \\ &= \sum_{i=1}^n \mathbb{P}(E|F_i)\mathbb{P}(F_i) \end{aligned}$$

## Exercise 1

Show that  $\mathbb{P}(A|B \cap C)\mathbb{P}(B|C) = \mathbb{P}(A \cap B|C)$ .

## II. Random Variables

# discrete vs continuous

## Discrete r.v.

### Definition 4

A random variable  $X$  is called a discrete random variable if (\*complete sentence\*)

- Example
  - flip a coin
  - throw a dice

## Continuous r.v.

### Definition 5

A random variable  $X$  is called a continuous random variable if (\*complete sentence\*)

- Example
  - weights and heights of a person
  - temperature of a room

# pmf and pdf

**pmf(probability mass function)** for a discrete r.v.  $X$

## Definition 6

A pmf  $p(x)$  is a function that gives the **probability** that a discrete random variable is exactly equal to some value.

$$\mathbb{P}(X = x) = p(x)$$

- Suppose you throw a dice and  $X$  be a r.v. of the outcome. What is the pmf of  $X$ ?
  - 1 *answer in a math form:*
  - 2 *answer in a tabular form:*

## pdf(probability density function) for a continuous r.v. $X$

### Definition 7

A pdf  $f(x)$  is a function that gives the **relative likelihood** for this continuous r.v. to take on a given value.

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f(x)dx$$

- Suppose a random variable takes a real values between 0 and 1 with equal likelihood, what is the pdf of  $X$ ?
  - ① *answer in a math form:*
  - ② *draw a graph:*

## Properties of pmf and pdf

- Mathematical property? *Something that always happen.*

### Remark 3

The functions (pmf and pdf) are nonnegative everywhere. That is,

$$p(x) \geq 0 \text{ and } f(x) \geq 0$$

### Remark 4

Its summation or integral over the entire area is equal to one. That is,

$$\sum p(x) = 1 \text{ and } \int f(x)dx = 1$$

- One can derive a cdf (cumulative distribution function) from a pmf or from a pdf.  
(See the upcoming definition of cdf.)

# Expectation

- For a discrete random variable  $X$  with pmf  $p(x)$ 
  - $\mathbb{E}X = \sum xp(x)$
  - $\mathbb{E}[g(X)] = \sum g(x)p(x)$
  - Ex)  $\mathbb{E}X^2 = \sum x^2p(x)$
  - Ex)  $\mathbb{E}(X^2 - 2X) = \sum$
- For a continuous random variable  $X$  with pdf  $f(x)$ 
  - $\mathbb{E}X = \int_{-\infty}^{\infty} xf(x)dx$
  - $\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx$
  - Ex)  $\mathbb{E}X^2 = \int_{-\infty}^{\infty} x^2f(x)dx$



## cdf

## Definition 8

For a random variable  $X$ , the cdf(cumulative distribution function)  $F(x)$  is a function of probability that the random variable  $X$  is found at a value less than or equal to  $x$ .

$$F(x) := \mathbb{P}(X \leq x)$$

- If discrete,

$$F(x) = \mathbb{P}(X \leq x) = \sum_{y=-\infty}^x \mathbb{P}(X = y) = \sum_{y=-\infty}^x p(y)$$

- If continuous,

$$F(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^x f(y)dy$$



## III. Uniform

## Definition 9

A continuous random variable  $X$  is said to follow uniform distribution with parameter  $a$  and  $b$ , and write  $X \sim U(a, b)$  if (reads " $X$  follows a uniform distribution with parameter  $a$  and  $b$ .")

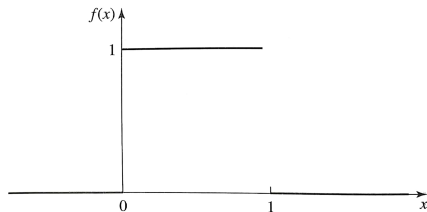
$$\text{pdf } f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

$$\text{cdf } F(x) = \begin{cases} 0 & \text{if } x \leq a \\ \frac{x-a}{b-a} & \text{if } a \leq x \leq b \\ 1 & \text{if } x > b \end{cases}$$

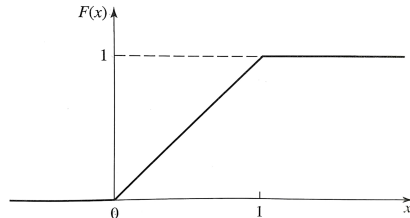
- (The value  $\frac{1}{b-a}$  is chosen as a constant so that it will make integrations over  $-\infty$  to  $\infty$  to be equal to 1.)

# $U(0, 1)$

• pdf



• cdf



## Exercise 2

$X \sim U(10, 20)$ , then what is  $F(10)$ ? and  $F(15)$ ?

## IV. Exponential

## Definition 10

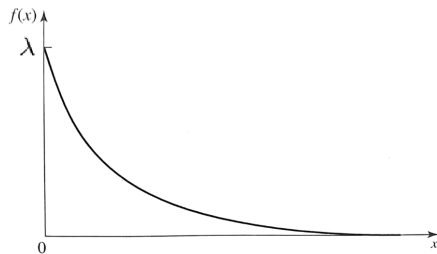
A nonnegative continuous random variable  $X$  is said to follow exponential distribution with parameter  $\lambda$  and write  $X \sim \exp(\lambda)$ , if

$$\text{pdf } f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

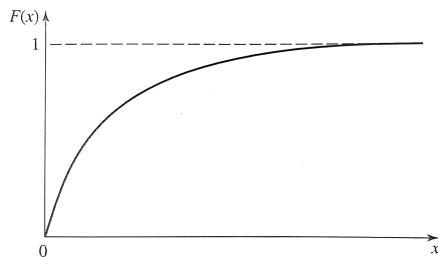
$$\text{cdf } F(x) = \begin{cases} 1 - e^{-\lambda x}, & \text{if } x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$



## ● pdf



## ● cdf



- TFAE (The followings are equivalent - *If any one of them is true, the rest of them are all true. If any one of them is false, then the rest of them are all false.*)

①  $X \sim \exp(\lambda)$

② pdf  $f(x) = \lambda e^{-\lambda x}$ , if  $x \geq 0$ ;  $f(x) = 0$ , otherwise

③ cdf  $F(x) = 1 - e^{-\lambda x}$ , if  $x \geq 0$ ;  $F(x) = 0$ , otherwise

### Exercise 3

*Prove that pdf  $\rightarrow$  cdf*

## More statistics (통계량) for r.v.

- Variance:  $Var(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2$
- Standard Deviation:  $sd(X) = \sqrt{Var(X)}$
- Coefficient of Variation (cv):  $cv(X) = \frac{sd(X)}{\mathbb{E}X}$
- (A cv measures 'relative' variation and often more meaningful.)
- Examples
  - ① cv of  $N(10, \sqrt{10}^2)$  vs  $N(10000, 10^2)$ .
  - ② For an exponential r.v.,  $X \sim exp(\lambda)$ , then  $cv(X) =$  (next slide)
  - ③ For a normal r.v.  $X \sim N(\mu, \sigma^2)$ ,  $cv(X) =$
  - ④ For deterministic variable  $X$ ,  $c_x =$

## Properties of exponential distribution

- Statistics for a r.v.  $X \sim \exp(\lambda)$ ,
  - ①  $EX = 1/\lambda$
  - ②  $Var(X) = 1/\lambda^2$
  - ③  $cv(X) = 1$  (Exponential r.v. has c.v. equal to 1, always.)
- Theorems
  - ① Memoryless property
  - ② Suppose  $X_1 \sim \exp(\lambda_1)$ ,  $X_2 \sim \exp(\lambda_2)$ , and independent, then
 
$$\mathbb{P}(X_1 < X_2) = \frac{\lambda_1}{\lambda_1 + \lambda_2}$$
  - ③ If  $X_1 \sim \exp(\lambda_1)$ ,  $X_2 \sim \exp(\lambda_2)$ , and independent, then
 
$$\min(X_1, X_2) \sim \exp(\lambda_1 + \lambda_2)$$

## Exercise 4

Show that  $EX = 1/\lambda$

## Exercise 5

Show that  $\text{Var}(X) = 1/\lambda^2$ . (Hint: need to do  $EX^2$  first)



## Common usage of exponential distribution

- Since exponential r.v. is continuous and nonnegative, it is frequently used to describe time.
- In such cases, the parameter  $\lambda$  can be understood as “rate” as an inverse value of the time.
- For example,  $X \sim \exp(\lambda)$  and  $\mathbb{E}X = 1/\lambda = 5$  yrs imply  $\lambda = 1/5$  per year.
- For example, consider the series of events that occur in every time interval that follows exponential distribution with mean 3 minutes. (In post office, you expect one customer arrival in 3 minutes on average and the time follows exponential distribution). We say that it follows exponential arrival times with rate  $1/3$  per minute.



# Memoryless property

## Definition 11

A r.v.  $X$  is *memoryless*, if  $\mathbb{P}(X > s + t | X > t) = \mathbb{P}(X > s)$ , for  $s, t \geq 0$ .

## Theorem 2

*Exponential random variable is memoryless.*

- For example, we shall assume that bus arrival time follows  $\exp(1/5)$ . In other words, it follows an exponential distribution and its expected arrival time is 5 minutes. You are waiting for bus, and have been waiting for 3 minutes, what is the probability that bus will not come in 5 minutes from now.

## Exercise 6

*Prove the previous theorem.*

### Theorem 3

Suppose  $X_1 \sim \exp(\lambda_1)$ ,  $X_2 \sim \exp(\lambda_2)$  and they are independent, then

$$\mathbb{P}(X_1 < X_2) = \frac{\lambda_1}{\lambda_1 + \lambda_2}$$

- Smith and Jones came to post office together and they are served by two clerks, A and B, respectively. Server A has service time following  $\exp(1/3)$  and server B has service time following  $\exp(1/5)$ . What is the chance that Smith will be done first?
- Suppose that Smith came to post office earlier than Jones by 2 minutes, but Smith was still being served at the moment that Jones started to being served. Would this assumption change your previous answer? Why or why not?

## V. Poisson

## Definition 12

A discrete random variable  $X$  is said to follow Poisson distribution with parameter  $\lambda$ , and write  $X \sim \text{poi}(\lambda)$ , if its pmf is

$$\mathbb{P}(X = k) = \frac{\lambda^k e^{-\lambda}}{k!} \text{ for } k = 0, 1, 2, \dots$$

- What is cdf of  $\text{poi}(\lambda)$ ?

## Remark 5

$$\mathbb{E}X = \text{Var}(X) = \lambda$$

## Exercise 7

For  $X \sim \text{poi}(\lambda)$ , prove that  $\mathbb{E}X = \lambda$ .

- cf)  $e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots = \sum_{n=0}^{\infty} \frac{x^n}{n!}$
- pf) We have  $\mathbb{P}(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$  for  $k = 0, 1, 2, \dots$ , and

$$\begin{aligned} \mathbb{E}X &= \sum_{x=-\infty}^{\infty} xp(x) \text{ (this is common for all discrete r.v.)} \\ &= \end{aligned}$$



## VI. Some Exercises



# Maximum and minimum

## Definition 13

$$x \wedge y = \min(x, y) = \begin{cases} x & , \text{ if } x \leq y \\ y & , \text{ otherwise} \end{cases}$$

$$x \vee y = \max(x, y) = \begin{cases} x & , \text{ if } x \geq y \\ y & , \text{ otherwise} \end{cases}$$

$$x^+ = \max(x, 0) \text{ (positive part)}$$

- Ex)  $x \wedge 25 = \min(x, 25)$
- Ex)  $(25 - x)^+ = \max(25 - x, 0)$

## Exercise 8

For  $X \sim U(20, 40)$ , evaluate  $\mathbb{E}[X \wedge 25]$  and  $\mathbb{E}[(25 - X)^+]$ .



## Exercise 9

For  $X \sim Poi(8)$ ,

①  $\mathbb{P}(X = 0) =$

②  $\mathbb{P}(2 \leq X \leq 4) =$

③  $\mathbb{P}(X > 2) =$

## Exercise 10

For  $X \sim \exp(7)$ , evaluate  $\mathbb{E}[\max(X, 7)]$ .



## Exercise 11

For  $X \sim \exp(8)$ , find  $x^*$  such that  $F(x^*) = 0.6$ .

## Exercise 12

For  $X \sim U(10, 20)$ , find  $x^*$  such that  $F(x^*) = 0.7$ .





"Man can learn nothing unless he proceeds from the known to the unknown. - Claude Bernard"