Lecture 1. Math Review (1)

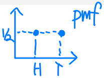
Sim, Min Kyu, Ph.D.
mksim@seoultech.ac.kr

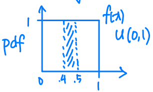
Math Review 1 - Random Variables - pmf/pdf

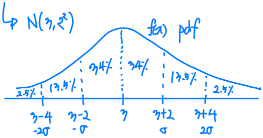
확률변수

- Class of a random variable (r.v.)
  - 이산 $\boxed{\text{Discrete}}$ r.v.: Toss a coin, Throw a dice → table & chart
  - $\boxed{\text{Continuous}}$ r.v.: Uniform r.v., Normal r.v. 연속

	prob
H	1/2
T	1/2



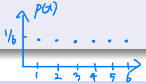




Definition - pmf (probability mass function)

- A pmf for a discrete r.v. X, $p(x)$, is a **function that gives the** $\boxed{\textbf{probability}}$ $=\mathbb{P}$ **that a discrete random variable is exactly equal to some value**, i.e. $\mathbb{P}(X=x)=p(x)$.
- For example, suppose you throw a dice and X be a r.v. of the outcome. What is the pmf of X?

$$p(x)=\mathbb{P}(X=x)=\begin{cases}1/6 & \text{for } x=1,2,3,4,5,6\\ 0 & \text{otherwise}\end{cases}$$

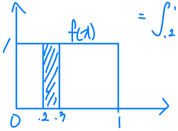


Definition - pdf (probability density function)

- A pdf for a continuous r.v. X, $f(x)$, is a **function that gives the relative likelihood for this continuous random variable to take on a given value,** i.e. $\mathbb{P}(a \le X \le b)=\int_a^b f(x)dx$

$U(0,1)$ $\quad \mathbb{P}(0.2\le X\le 0.3)$

$$=\int_{.2}^{.3} f(x)dx = 0.1$$



Math Review 1 - Random Variables - pmf/pdf

- Properties of pmf/pdf[1]
  - The functions (pmf and pdf) are nonnegative everywhere.
    - i.e. $p(x) \geq 0$, $f(x) \geq 0$
  - Its summation/integral over the entire area is equal to one.
    - i.e. $\sum p(x) = 1$, $\int f(x)dx = 1$.
  - One can derive cdf (cumulative distribution function) from pmf and pdf.
    - (See the definition of cdf.)

1: A mathematical property is a something that always happen.

Math Review 2 - Random Variables - expectation (expected value)

- 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) $\mathbb{E}X^2 = \sum x^2 p(x)$
- 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) $\mathbb{E}X^2 = \int_{-\infty}^{\infty} x^2 f(x)dx$

Math Review 3 - Random Variables - cdf

Definition - cdf (cumulative distribution function)

- 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.**
  - $F(x) := \mathbb{P}(X \le x)$
  - If discrete, $F(x) = \mathbb{P}(X \le x) = \mathbb{P}(-\infty \le X \le x) = \sum_{y=-\infty}^{y=x} \mathbb{P}(X = y) = \sum_{y=-\infty}^{y=x} p(y)$
  - If continuous, $F(x) = \mathbb{P}(X \le x) = \mathbb{P}(-\infty \le X \le x) = \int_{-\infty}^{x} f(y)dy$

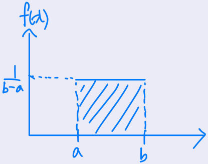
blank

Math Review 4 - Uniform distribution

Definition - Uniform distribution

A random variable (r.v.) follows uniform distribution with parameters a and b

- A continuous random variable X is said to follow **Uniform distribution** with parameter a and b, and write $\boxed{X \sim U(a,b)}$ if



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

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

- Check: pdf $\rightarrow$ cdf

- For $X \sim U(a,b)$, show pdf $\longrightarrow$ cdf

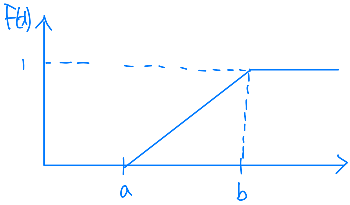
$$f(x) = \begin{cases} 0 & \text{if } x \leq a \\ \frac{1}{b-a} & \text{if } a < x \leq b \\ 0 & \text{if } x > b \end{cases}$$

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

i) $x \leq a$, $F(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f(y)\,dy = \int_{-\infty}^{x} 0\,dy = 0.$ $(\because y \leq x \leq a \Rightarrow y \leq a,\ f(y) = 0)$

ii) $a < x \leq b$, $F(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f(y)\,dy = \int_{-\infty}^{a} f(y)\,dy + \int_{a}^{x} f(y)\,dy = 0 + \int_{a}^{x} \frac{1}{b-a}\,dy = \frac{x-a}{b-a}$

iii) $x > b$, $F(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f(y)\,dy = \int_{-\infty}^{a} f(y)\,dy + \int_{a}^{b} f(y)\,dy + \int_{b}^{x} f(y)\,dy = 0 + 1 + 0 = 1$



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

Math Review 5 - Exponential distribution

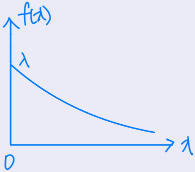
$\mathbb{R}$: set of real number (실수)
$\mathbb{Z}$: set of integer (정수)
$\mathbb{N}$: set of natural number. (자연수)
$\mathbb{R}^+$: set of positive real number

Definition - Exponential distribution

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

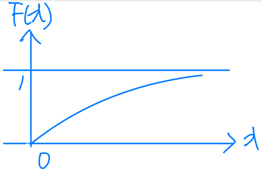
time

$\lambda \in \underbrace{\mathbb{R}^+ \cup \{0\}}_{\text{nonnegative}}$



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

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



- TFAE (The followings are all equivalent)
  - $X \sim exp(\lambda)$
  - pdf $f(x) = \lambda e^{-\lambda x}$, if $x \geq 0$; $f(x) = 0$, otherwise
  - cdf $F(x) = 1 - e^{-\lambda x}$, if $x \geq 0$; $F(x) = 0$, otherwise
- pf) pdf $\rightarrow$ cdf

i) $x<0$, $F(x) = \mathbb{P}(X \leqslant x) = \int_{-\infty}^{x} f(y)dy = \int_{-\infty}^{x} 0\,dy = 0$. ($\because$ $y \leqslant x < 0 \Rightarrow y<0$, $f(y)=0$)

ii) $x \geqslant 0$, $F(x) = \mathbb{P}(X \leqslant x) = \int_{-\infty}^{x} f(y)dy = \int_{-\infty}^{0} f(y)dy + \int_{0}^{x} f(y)dy = \int_{0}^{x} \lambda e^{-\lambda y} dy = \left[-e^{-\lambda y}\right]_{0}^{x} = -e^{-\lambda x} - (-1) = 1 - e^{-\lambda x}$

$$\therefore F(x) = \begin{cases} 0 & \text{if } x<0 \\ 1-e^{-\lambda x} & \text{if } x \geqslant 0 \end{cases}$$

Math Review 6 - Random Variables - summary statistics

- Expectation: $\mathbb{E}X = \int_{-\infty}^{\infty} xf(x)dx$ or $\sum xp(x)$.
- Variance: $Var(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \mathbb{E}(X - \mathbb{E}X)^2$
- Standard Deviation: $sd(X) = \sqrt{Var(X)}$

measure for absolute dispersion

변동계수

- cv (coefficient of variation) $c_X = \frac{sd(X)}{\mathbb{E}X}$
- scv (squared-cv) $c_X^2 = \frac{(sd(X))^2}{(\mathbb{E}X)^2} = \frac{Var(X)}{(EX)^2}$

measure for relative dispersion.

- Examples $C_X = \frac{sd(X)}{\mathbb{E}X}$
  - ex) For an Exponential dist. $X \sim exp(\lambda)$, $c_X =$ $\frac{1}{\lambda} / \frac{1}{\lambda} = 1$
  - ex) For a Normal dist. $X \sim N(\mu, \sigma^2)$, $c_X =$ $\frac{\sigma}{\mu}$
  - ex) For deterministic X, $c_X =$ 0 $(\because sd(X) = 0)$

Math Review 7 - Exponential distribution - properties

- For $X \sim exp(\lambda)$, $\mathbb{E}X = 1/\lambda$, $Var(X) = 1/\lambda^2$, $sd(X) = 1/\lambda$
- Theorems
  - Memoryless property
  - Suppose $X_1 \sim exp(\lambda_1), X_2 \sim exp(\lambda_2)$, and they are independent, then $\mathbb{P}(X_1 < X_2) = \frac{\lambda_1}{\lambda_1+\lambda_2}$
  - If $X_1 \sim exp(\lambda_1), X_2 \sim exp(\lambda_2)$, and they are independent, then $min(X_1, X_2) \sim exp(\lambda_1 + \lambda_2)$

Math Review 8 - Exponential distribution - moments

$\mathbb{E}X$: 1st moment

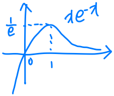
$\mathbb{E}X^2$: 2nd moment

$\vdots$

$\mathbb{E}X^n$: n th moment

부분적분 $\int f g' = f g - \int f' g$

- $\mathbb{E}X = 1/\lambda, Var(X) = 1/\lambda^2$



$X \sim \exp(\lambda)$, $f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & o.w \end{cases}$

$$\mathbb{E}X = \int_{-\infty}^{\infty} x f(x)\,dx = \int_{-\infty}^{0} x f(x)\,dx + \int_{0}^{\infty} x f(x)\,dx = \int_{-\infty}^{0} x \cdot 0\,dx + \int_{0}^{\infty} x \cdot \lambda e^{-\lambda x}\,dx = \lambda \left\{ \left[x \cdot \left(-\frac{1}{\lambda} e^{-\lambda x}\right) \right]_0^{\infty} - \int_0^{\infty} -\frac{1}{\lambda} e^{-\lambda x}\,dx \right\}$$

$$= 0 + \lambda \left[-\frac{1}{\lambda^2} e^{-\lambda x} \right]_0^{\infty} = 0 + \frac{1}{\lambda} = \frac{1}{\lambda}$$

$$\mathbb{E}X^2 = \int_{-\infty}^{\infty} x^2 f(x) = \int_{-\infty}^{0} x^2 \cdot 0\,dx + \int_0^{\infty} x^2 \lambda e^{-\lambda x}\,dx = \lambda \left\{ \left[x^2 \cdot \left(-\frac{1}{\lambda} e^{-\lambda x}\right) \right]_0^{\infty} - \int_0^{\infty} 2x \left(-\frac{1}{\lambda} e^{-\lambda x}\right) \right\}$$

$$= \lambda \left\{ 0 + \frac{2}{\lambda} \int_0^{\infty} x \cdot \lambda e^{-\lambda x}\,dx \right\} = \frac{2}{\lambda}\left(\frac{1}{\lambda}\right) = \frac{2}{\lambda^2}$$

$$Var(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}$$

$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x}\,dx$, $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$, $\Gamma(n) = (n-1)!$ $\quad n \in \mathbb{Z}$

n^{th} moment, $\mathbb{E}X^n = \int_0^{\infty} x^n \lambda e^{-\lambda x}\,dx$, $y = \lambda x$ 치환. $\Rightarrow \mathbb{E}X^n = \int_0^{\infty} \left(\frac{y}{\lambda}\right)^n \lambda e^{-y} \frac{dy}{\lambda} = \frac{1}{\lambda^n} \int_0^{\infty} y^n e^{-y}\,dy$

$= \frac{1}{\lambda^n} \times \Gamma(n+1) = \frac{1}{\lambda^n} \times n! = \frac{n!}{\lambda^n}$

Math Review 9 - Memoryless property

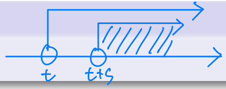
Definition - Memoryless property

조건부 확률.

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

Theorem

- Exponential random variable is memoryless.



$$\frac{\mathbb{P}(X>s+t \cap X>t)}{\mathbb{P}(X>t)} = \frac{\mathbb{P}(X>s+t)}{\mathbb{P}(X>t)}$$

이미 s 시간 동안 어떤 사건이 발생하지 않았다고 하더라도 추가로 t 시간이 더 지나서 사건이 발생할 확률은 s 시간과 무관하다.

지수분포: 고장률이 일정(노화, 마모 X)한 상황 모델링
이미 사용한 시간은 무의미. 남은 기대수명도 동일

blank proof

$X \sim \exp(\lambda)$, $\mathbb{P}(X \leq x) = F(x) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & \text{o.w} \end{cases}$ $\quad (s, t \geq 0)$

$$\mathbb{P}(X > s+t \mid X > t) = \frac{\mathbb{P}(X > s+t \cap X > t)}{\mathbb{P}(X > t)} = \frac{\mathbb{P}(X > s+t)}{\mathbb{P}(X > t)} = \frac{1 - \mathbb{P}(X \leq s+t)}{1 - \mathbb{P}(X \leq t)} = \frac{1 - F(s+t)}{1 - F(t)} = \frac{1 - (1 - e^{-\lambda(s+t)})}{1 - (1 - e^{-\lambda t})} = e^{-\lambda(s+t) + \lambda t}$$

$$= e^{-\lambda s} = 1 - (1 - e^{-\lambda s}) = 1 - F(s) = 1 - \mathbb{P}(X \leq s) = \mathbb{P}(X > s) \quad \blacksquare$$

Thus exponential random variable is memoryless.

Math Review 10 - Exponential distribution - property

Theorem

- Suppose $X_1 \sim exp(\lambda_1), X_2 \sim exp(\lambda_2)$, and they are independent, then $\mathbb{P}(X_1 < X_2) = \frac{\lambda_1}{\lambda_1+\lambda_2}$

Example

- 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 $exp(1/5)$.
  1. What is the chance that Smith will be done first?
  2. 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?

Smith $\to$ A, $X_A \sim exp(\lambda_A = 1/3)$

Jones $\to$ B, $X_B \sim exp(\lambda_B = 1/5)$

#1. $\mathbb{P}(X_A < X_B) = \dfrac{\lambda_A}{\lambda_A + \lambda_B} = \dfrac{5}{8}$

#2. No. due to memoryless property

Math Review 11 - Poisson distribution

Definition - Poisson distribution

- A discrete nonnegative random variable X is said to follow a **Poisson distribution** with parameter λ, and write $X \sim Poi(\lambda)$, if pmf

pmf

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

- What is cdf of $Poi(\lambda)$? $F(k) = \mathbb{P}(X \leq k) = \sum_{l=0}^{k} \mathbb{P}(X = l) = \sum_{l=0}^{k} \frac{\lambda^l e^{-\lambda}}{l!}$
- $\mathbb{E}X = Var(X) = \lambda.$

$$\sum_{k=0}^{\infty} \mathbb{P}(X=k) = \sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} \cdot e^{\lambda} = 1$$

$$e = \sum_{k=0}^{\infty} \frac{1}{k!} = \frac{1}{1} + \frac{1}{1} + \frac{1}{1 \cdot 2} + \cdots = \lim_{n \to \infty} \left(1 + \frac{1}{n}\right)^n$$

$$e^{\lambda} = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \quad \text{(Taylor Series)}$$

- We have

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

- $\mathbb{E}X = \lambda$
- $Var(X) = \lambda$ HW.

$(x-1=y.)$

$$\mathbb{E}X = \sum x p(x) = \sum_{x=0}^{\infty} x p(x) = \sum_{x=0}^{\infty} x \cdot \frac{\lambda^x e^{-\lambda}}{x!} = 0 \cdot \frac{\lambda^0 e^{-\lambda}}{0!} + \sum_{x=1}^{\infty} \frac{\lambda^x e^{-\lambda}}{(x-1)!} = e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda \cdot \lambda^{x-1}}{(x-1)!} = \lambda \cdot e^{-\lambda} \cdot \sum_{y=0}^{\infty} \frac{\lambda^y}{y!} = \lambda \cdot e^{-\lambda} \cdot e^{\lambda} = \lambda$$

$$\mathbb{E}X^2 = \mathbb{E}[X(X-1)] + \mathbb{E}[X] = \lambda^2 + \lambda$$

$$\mathbb{E}[X(X-1)] = \sum_{x=0}^{\infty} x(x-1) p(x) = \sum_{x=0}^{\infty} x(x-1) \frac{\lambda^x e^{-\lambda}}{x!} = \sum_{x=2}^{\infty} \frac{\lambda^2 \cdot \lambda^{x-2} e^{-\lambda}}{(x-2)!} = \lambda^2 \cdot e^{-\lambda} \cdot e^{\lambda} = \lambda^2$$

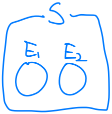
$$Var(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$$

Definition - Probability

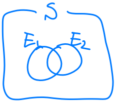
input output

- Probability is a function from **"an event"** to real number between 0 and 1.

- $\mathbb{P} : S \rightarrow [0,1] \cap \mathbb{R}$
- $\mathbb{P}(S) = 1$
- For any event E, $0 \leq \mathbb{P}(E) \leq 1$
- For $E = \emptyset$, $\mathbb{P}(E) = 0$
- If $E_1 \cap E_2 = \emptyset$, then $\mathbb{P}(E_1 \cup E_2) = \mathbb{P}(E_1) + \mathbb{P}(E_2)$.
- $\mathbb{P}(E_1 \cup E_2) = \mathbb{P}(E_1) + \mathbb{P}(E_2) - \mathbb{P}(E_1 \cap E_2)$.



disjoint



Math Review 13 - Conditional Probabilities

Definition - Conditional Probability

- $\mathbb{P}(E|F)$ is the **probability that event E occurs given that F has occurred**.
- $\mathbb{P}(E|F) = \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(F)}$

Example. (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 $\underbrace{\text{least five}}_{F}$, then what is the conditional probability $\underbrace{\text{that it is ten}}_{E}$?

$$P(E|F) = \frac{P(E \cap F)}{P(F)} = \frac{1/10}{6/10} = \frac{1}{6}$$

Math Review 14 - Probabilities - properties

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

Math Review 15 - Bayes' rule

Definition - Bayes' rule

- Definition $F_1, \ldots, F_n$ are a mathematical partition if
  - $F_i \cap F_j = \emptyset$, or, $\mathbb{P}(F_i \cap F_j) = 0$, for any $i \neq j$
  - $\mathbb{P}(F_1 \cup \cdots \cup F_n) = 1$
- In other words, exactly one of the events $F_1, \ldots, F_n$ will occur.
- Then, the following holds.

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

Math Review 16 - Properties of max/min

1. $-max(x, y) = min(-x, -y)$
2. $max(x, y) = -(-max(x, y)) = -min(-x, -y)$
3. $max(x, 0) = -min(-x, 0)$
4. $max(x, y) + z = max(x + z, y + z)$
5. If x^* maximizes(minimizes) $f(x)$, then x^* minimizes(maximizes) $-f(x)$
6. If x^* maximizes(minimizes) $f(x)$, then x^* maximizes(minimizes) $f(x) + \underbrace{g(y) + C}_{\text{constant in terms of } x}.$

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