Lec 25 §7.8. Probability

- Probability density,
- Cumulative function
- mean (expectation).
- Some applications

slide lecture.

§ Probability density function:

Def. A function $P(x)$ defined on $[a, b]$ is said to be a probability density (here, a can be $-\infty$, b can be ∞) if

1. $p(x) \geq 0$ for all x
2. $\int_a^b p(x)\,dx = 1.$ ← $[a, b]$ is the possible values of x

* probability density is similar to mass density. except that total probability $= 1$, ALWAYS!.

Def A variable x with probability density $p(x)$ is called a random variable

e.g. $p(x) = \frac{1}{\sqrt{\pi}} e^{-x^2}$ is a probability density function

because
1). $p(x) \geq 0$ for all x.
2). Fact $\int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi}} e^{-x^2} dx = 1.$

e.g. (Probability vs Probability density)

Let, in the Galton board example,

$p(x) = \frac{1}{\sqrt{\pi}} e^{-x^2}$

What is the probability for a small ball to arrive at $x = 1$?

<sol>. ~~At $x=1$, $p(1) = \frac{1}{\sqrt{\pi}} e^{-1}$~~

~~So, the probability is $\frac{1}{\sqrt{\pi}} e^{-1}$.~~

WRONG!

NOTE Probability density $\neq$ probability

(similarly to mass density $\neq$ mass).

At the point $x=1$, the width of the interval $= 0$.

So, the probability for a ball to arrive there is zero. ■

More systematic way
to get probability from probability density:

Integrate the probability density!
(Just like how you get mass from mass density.)

- Probability from Probability density function.

The probability for a random variable takes on values in the interval $\alpha \le x \le \beta$:

$$\mathbb{P}(\alpha \le x \le \beta) = \int_\alpha^\beta p(x)\,dx.$$

just a notation to denote the probability

e.g. $p(x) = \frac{1}{\sqrt{\pi}} e^{-x^2}$

$$\mathbb{P}(x=1) = \int_1^1 p(x)\,dx = \int_1^1 \frac{1}{\sqrt{\pi}} e^{-x^2}\,dx = 0$$

Ex. Let $p(x) = Cx^2$, $0 \le x \le 1$. C = const.
be a probability density.

possible values of x.

① Determine C.

② Find probability for x to be in $\frac{1}{2} \le x \le 1$.

<sol>

① . $1 = \int_0^1 p(x)\,dx = C\int_0^1 x^2\,dx = C\left[\frac{x^3}{3}\right]_0^1 = C\cdot\frac{1}{3}$. $\therefore$ $C=3$

② $\mathbb{P}(\frac{1}{2} \le x \le 1) = \int_{\frac{1}{2}}^1 p(x)\,dx = \int_{\frac{1}{2}}^1 3\cdot x^2\,dx = \left[x^3\right]_{\frac{1}{2}}^1 = 1 - (\frac{1}{2})^3 = \frac{7}{8}$.

C=3 from ①.

• **Cumulative function.**

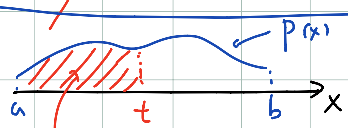
can be $-\infty$ can be ∞.

Let · X be a random variable $a \le X \le b$.

· $p(x)$ be a probability density function on $a \le x \le b$

Define cumulative function

$$\boxed{F(t) = \int_a^t p(s)\,ds \quad \left(= \mathbb{P}(X \le t)\right)}$$



the area $= F(t)$

F.T.C.

NOTE • $F'(t) = p(t)$ if $p(t)$ is continuous.

• $\mathbb{P}(\alpha \le X \le \beta) = \int_\alpha^\beta p(x)\,dx = \underbrace{\int_\alpha^a p(x)\,dx}_{-\int_a^\alpha p(x)dx} + \int_a^\beta p(x)\,dx$

$= F(\beta) - F(\alpha)$.

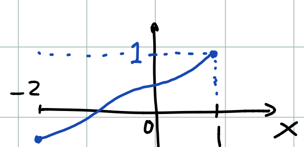
Ex Let · X random variable, $-2 \le X \le 1$.

· $p(x)$ probability density function.

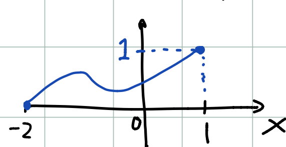
· Assume $p(x)$ is continuous

In the following find all those that cannot be cumulative functions for X.

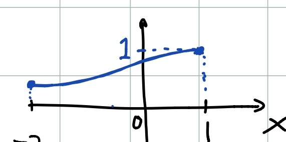
a)



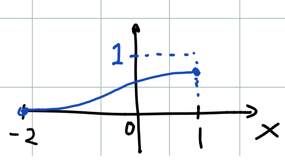
b)



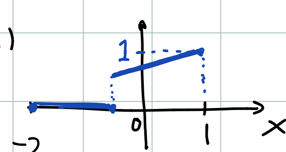
e)



c)



d)



<sol>.

Cumulative function : $F(x) = \int_{-2}^{x} p(s)\,ds$ for $-2 \le x \le 1$

we see

- $F(1) = 1$ ← total probability,
- $F(-2) = 0$ ← $\int_{-2}^{-2} p(s)ds = 0$.
- $F(x) \ge 0$
- $F(x)$ increasing.

} because $p(s) \ge 0$.

- $F'(x) = p(x)$ ← Fundamental theorem of calculus, for continuous $p(x)$.

a) NO because it has negative values

b) NO because it is not increasing.

c) NO because the value at $x=1$, is NOT 1.

d) NO because it is not differentiable. (not even continuous)

e) NO because the value at $x=-2$, is NOT 0.

None of them can be the cumulative function ▨

EX Let . random variable x, $0 \le x < \infty$

$p(x) = e^{-ax}$. $a > 0$. constant.

Find $F(t) = \mathbb{P}(x \le t)$.

the probability for x to be $\le t$.

<sol>.

● Note $F(t) = \int_0^t p(x)\,dx = \int_0^t e^{-ax}\,dx$

$$= -\frac{1}{a}e^{-ax}\Big]_0^t = -\frac{1}{a}\left(e^{-at} - e^{0}\right)$$

$$= -\frac{1}{a}\left(e^{-at} - 1\right).$$

- Need to determine the constant a.

  - Use $1 = \int_0^\infty p(x)dx$.

    This is nothing but

$$1 = \lim_{t\to\infty} F(t) = \lim_{t\to\infty} \Big[\underbrace{-\frac{1}{a}\left(e^{-at}-1\right)}_{\text{found in the above}} \Big]$$

$$= -\frac{1}{a}(0-1) \quad \leftarrow \lim_{t\to\infty} e^{-at} = 0 \text{ for } a>0.$$

$$= \frac{1}{a}.$$

$\therefore \ \underline{a=1}$.

- Tie-up.

$$F(t) = -\frac{1}{1}\left(e^{-\overset{a=1.}{1}\cdot t}-1\right)$$

$$= -\left(e^{-t}-1\right)$$

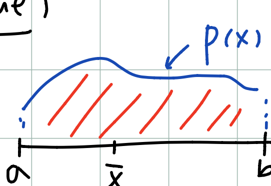
$$= 1-e^{-t}$$

- **Mean** (= Average = Expected Value)

  - random variable x, $a \le x \le b$.
  - probability density $p(x)$.

the mean (or average, or expected value) of the random variable x is

$$\boxed{\bar{x} = \int_a^b x\, p(x)\, dx}$$



※ This is exactly the center of mass of the distribution $p(x)$, (if we view it as mass density)

because

center of mass $= \dfrac{\int_a^b x\, p(x)\, dx}{\int_a^b p(x)\, dx}$ ← this is 1 for a probability density

e.g. (discrete case).

Class midterm average $\bar{k}$ (out of 100 total mark)

$$\bar{k} = \frac{1}{\#\text{ of total students}} \cdot \sum_{k=0}^{100} k \cdot (\#\text{ of students who got score } k)$$

$$= \sum_{k=0}^{100} k \cdot \boxed{\frac{(\#\text{ of students who got score } k)}{(\#\text{ of total students})}}$$

$\therefore$

$$\bar{k} = \sum_{k=0}^{100} k \cdot (\text{Probability of a student to get score } k)$$

● discrete case

average of k $= \sum_k k \cdot (\text{Probability for } k)$

Compare!

● Continuous case

mean of $x = \bar{x} = \int_a^b x\, p(x)\, dx$

probability = probability density × width.

EX Bacteria colony.

- Each bacterium can live up to maximum 1000 hrs.
- Probability for a bacterium to live less than t hours is $C \cdot e^{2t} + D$.

(C & D are constants.)

Find the expected life time of a bacterium.

<sol>.

- Translate the problem & set-up the goal
  - First, determine the random variable & its range.

    : the random variable t. $0 \le t \le 1000$

    to represent bacterium life time.
  - What is to be found?

    : the expected life time

    = expected value of t.

    $= \bar{t} = \int_0^{1000} t\, p(t)\, dt$

    $p(t)$: the probability density for t.

- set-up strategy

  We will
  - Determine $p(t)$.
  - Then, compute the integral.

- Collect information
- the cumulative function

given

$$F(s) = \int_0^s p(t)\, dt = \mathbb{P}(t \le s) = C \cdot e^{2s} + D$$

- Relate what found to what we want to determine.

relation between $F(t)$ and $p(t)$?

Yes: $F'(t) = p(t)$.

So, $p(t) = F'(t) = (Ce^{2t}+D)' = 2Ce^{2t}$.

- Determine $p(t)$: $p(t) = 2Ce^{2t}$.

How to determine C ?

NOTE $\boxed{0 \le t \le 1000}$

- $F(1000) = \mathbb{P}(t \le 1000) = 1 \Rightarrow 1 = Ce^{2000} + D$. ...... ①

- $F(0) = 0 \Rightarrow 0 = ce^{0} + D$

i.e., $0 = C + D$. ... ②.

From ②, $C = -D$. & from ① $1 = -De^{2000} + D$
$= D(1-e^{2000})$

$$\therefore \quad D = \frac{1}{1-e^{2000}}$$

$$C = -D = \frac{1}{e^{2000}-1}.$$

So, $$p(t) = \frac{2}{e^{2000}-1} e^{2t}.$$

- Perform the calculation as planned.

$$\bar{t} = \int_0^{1000} t\, p(t)\, dt$$

$$= \int_0^{1000} t \cdot \frac{2}{e^{2000}-1} e^{2t}\, dt$$

$$= \frac{1}{e^{2000}-1} \int_0^{1000} 2t e^{2t}\, dt$$ ← Integration by parts.

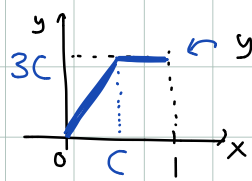
$$= \frac{1}{e^{2000}-1} \left\{ \left[2t \frac{e^{2t}}{2} \right]_0^{1000} - \underbrace{\int_0^{1000} \cancel{2} \cdot \frac{e^{2t}}{\cancel{2}}\, dt}_{\left[\frac{1}{2} e^{2t}\right]_0^{1000}} \right\}$$

$$= \frac{1}{e^{2000}-1} \cdot \left\{ \left(1000\, e^{2000} - 0\right) - \frac{1}{2}\left(e^{2000} - e^{0} \right) \right\}$$

$$= \frac{1}{e^{2000}-1} \left\{ 1000\, e^{2000} - \frac{1}{2} e^{2000} + \frac{1}{2} \right\}$$

$$= \frac{1}{e^{2000}-1} \left\{ \left(1000 - \frac{1}{2}\right) e^{2000} + \frac{1}{2} \right\}$$

$\blacksquare$

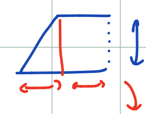
EX 

← $y = p(x)$, $0 \le x \le 1$.

Let $p(x)$ be a probability density function.

Find the cumulative function.

<sol>. • Need to determine C.



- Eqn for C.

total probability

$$1 = \int_0^1 p(x)\,dx = \text{Area}(\triangle\ 3c,\ c) + \text{Area}(\square\ 3c,\ 1-c)$$

$$= \frac{1}{2} 3c^2 + (1-c)\cdot 3c$$

$$= \frac{1}{2} 3c^2 + 3c - 3c^2$$

$$= 3c - \frac{1}{2}\cdot 3c^2$$

$$\therefore \quad \frac{1}{2} 3c^2 - 3c + 1 = 0$$

$$3c^2 - 6c + 2 = 0$$

$$3c^2 - 6c + 3 - 3 + 2 = 0$$
$$3(c^2 - 2c + 1) - 3 + 2 = 0$$
$$3(c-1)^2 - 1 = 0.$$
$$3(c-1)^2 = 1$$
$$(c-1)^2 = \frac{1}{3}$$
$$c - 1 = \pm\sqrt{\frac{1}{3}}$$
$$c = 1 \pm \sqrt{\frac{1}{3}}.$$

- Apply quadratic formula

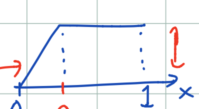
$$c = 1 \pm \sqrt{\frac{1}{3}}.$$

- But $0 \le c \le 1$.

So, choose $\underline{c = 1 - \sqrt{\frac{1}{3}}}$

• Now the cumulative function is

$$F(t) = \int_0^t p(x)\,dx$$

where $p(x)$: 

$3c = 3(1 - \sqrt{\frac{1}{3}})$.

$\triangle\ 3c,\ c$ slope $= 3$

$c = 1 - \sqrt{\frac{1}{3}}$

<u>Two cases</u>:

- If $t < C = 1-\sqrt{\frac{1}{3}}$, then $F(t) = \int_0^t \overset{\text{slope } 3}{3x}\, dx$

$$= \frac{3}{2}x^2\Big]_0^t = \frac{3}{2}t^2.$$

- If $t > C = 1-\sqrt{\frac{1}{3}}$, then

$$F(t) = \int_0^C p(x)\,dx + \int_C^t p(x)\,dx$$

$$= \frac{1}{2}\cdot 3\cdot\left(1-\sqrt{\tfrac{1}{3}}\right)^2 + \int_{1-\sqrt{\frac{1}{3}}}^t 3\cdot\left(1-\sqrt{\tfrac{1}{3}}\right)dx$$

$$= \frac{3}{2}\left(1-\sqrt{\tfrac{1}{3}}\right)^2 + 3\cdot\left(1-\sqrt{\tfrac{1}{3}}\right)\cdot\left(t-1+\sqrt{\tfrac{1}{3}}\right).$$

- Finally,

$$\boxed{F(t) = \begin{cases} \frac{3}{2}t^2 & \text{if } 0 \le t < 1-\sqrt{\frac{1}{3}} \\ \frac{3}{2}\left(1-\sqrt{\frac{1}{3}}\right)^2 + 3\left(1-\sqrt{\frac{1}{3}}\right)\cdot\left(t-1+\sqrt{\frac{1}{3}}\right) & \text{if } 1-\sqrt{\frac{1}{3}} \le t \le 1. \end{cases}}$$

- To make sure we get the correct answer, check $1 = F(1)$:

$$\frac{3}{2}\left(1-\sqrt{\tfrac{1}{3}}\right)^2 + 3\left(1-\sqrt{\tfrac{1}{3}}\right)\left(1-1+\sqrt{\tfrac{1}{3}}\right) = \frac{3}{2}\left(1-\sqrt{\tfrac{1}{3}}\right)^2 + 3\left(1-\sqrt{\tfrac{1}{3}}\right)\cdot\sqrt{\tfrac{1}{3}}$$

$$= \frac{3}{2}\left(1-2\cdot\sqrt{\tfrac{1}{3}}+\tfrac{1}{3}\right) + 3\sqrt{\tfrac{1}{3}} - 3\cdot\tfrac{1}{3} = \frac{3}{2} \cancel{-3\sqrt{\tfrac{1}{3}}} + \frac{1}{2} \cancel{+3\sqrt{\tfrac{1}{3}}} - 1 = 1.$$