§3.3 Discrete Random Variables

Motivation Example. Toss **2** fair 6-sided dice. What is the probability that the **sum** of the numbers equal to 9?

The sample space S has 36 sample points given by

$$S = \left\{ \begin{array}{cccccc} (1,1), & (1,2), & (1,3), & (1,4), & (1,5), & (1,6) \\ (2,1), & (2,2), & (2,3), & (2,4), & (2,5), & (2,6) \\ (3,1), & (3,2), & (3,3), & (3,4), & (3,5), & (3,6) \\ (4,1), & (4,2), & (4,3), & (4,4), & (4,5), & (4,6) \\ (5,1), & (5,2), & (5,3), & (5,4), & (5,5), & (5,6) \\ (6,1), & (6,2), & (6,3), & (6,4), & (6,5), & (6,6) \end{array} \right\}$$

We know that the probability of each sample point is $\frac{1}{36}$.

The event A is given by $A = \{(6,3),(5,4),(4,5),(3,6)\}$

The probability of A is $P(A) = \frac{4}{36}$.

In this example, we only care about the sum of two numbers. So we want to **redefine** the sample space S to a "smaller sample space" in $\mathbb{R}$ using a **random variable** X assigning each outcome a real number,

$$X : S \to \mathbb{R}$$

In our example, we define $X : (a,b) \to a+b$. As a set $X(S) = \{2,3,4,5,6,7,8,9,10,11,12\} \subset \mathbb{R}$, which is called the **range** of X.

If the range of X is a countable subset of $\mathbb{R}$, then X is called a **discrete random variable**.

Definition.

For every discrete random variable X, we define a **probability density function (pdf)** by

$$p_X(k) = P(X=k) := P(\{s \in S | X(s) = k\})$$

$p_X(k)$ is a function from $\mathbb{R}$ to $\mathbb{R}$. (If $k \notin X(S)$, then $p_X(k) = 0$.)

Question: Find probability $p_X(k)$ for all k.

In our example, the probability $p_X(k)$ of each number k in the range is assigned by

k	2	3	4	5	6	7	8	9	10	11	12
$p_X(k)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

The probability that the sum of the numbers equal to 9 is given by $p_X(9) = \frac{4}{36}$.

Question: What is the probability that the sum of the numbers ≤ 4?

It is given by $P(X \leq 4) = \sum_{k=0}^{4} P(X=k) = P(X=0) + P(X=1) \ldots + P(X=4)$

$\approx \frac{6}{36}$

This is another useful function called **cumulative distribution function (cdf)**, defined as

$$F_X(t) = P(X \leq t) := P(\{s \in S | X(s) \leq t\})$$

In our example, find the cumulative distribution function (cdf).

t	2	3	4	5	6	7	8	9	10	11	12
$F_X(t)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{15}{36}$	$\frac{21}{36}$	$\frac{26}{36}$	$\frac{30}{36}$	$\frac{33}{36}$	$\frac{35}{36}$	$\frac{36}{36}$

Example 1. (**Binomial Distribution**) in §3.2

Let the random variable X denote the number of "successes" in n independent trials. Then, the Binomial Distribution can be stated as

$$P(X = k) = \binom{n}{k} \cdot p^k (1-p)^{n-k}$$

for $k = 0, 1, 2, \ldots, n$.
The cumulative distribution function (**cdf**) of X is

$$P(X \leq t) = \sum_{k=0}^{t} \binom{n}{k} \cdot p^k (1-p)^{n-k}$$

Theorem.

For any discrete random variable X, the **pdf** satisfies

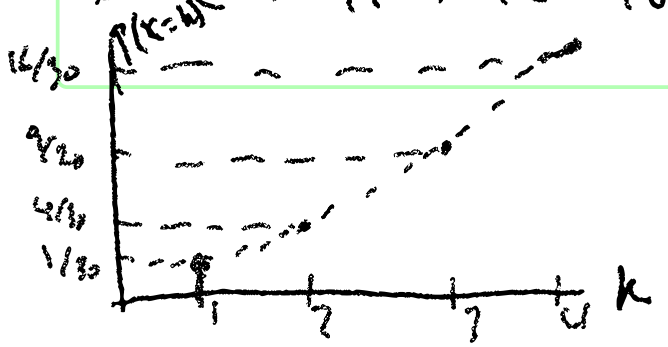
1. $p_X(k) \geq 0$.
2. $\sum_{x \in X(S)} p_X(k) = 1$.

Example 2. If a pdf is given by $p_X(k) = c \cdot k^2$ for $X(S) = \{1, 2, 3, 4\}$. Find c, and graph the pdf function.

$P(X=1) + P(X=2) + P(X=3) + P(X=4)$

$= c + 4c + 9c + 16c = 30c = 1$

$c = 1/30$



$P_X(k) = \frac{1}{30} k^2$

Ex. toss coin 3 times

S	X
HHH	3
HHT	2
HTH	2
HTT	1
THH	2
THT	1
TTH	1
TTT	0

$X = \{0, 1, 2, 3\}$

k	0	1	2	3	tot
$P(X=k)$	1/8	3/8	3/8	1/8	1 ✓

$$P(X=1) = \binom{3}{1}(0.5)(0.5)^2$$

$$= 3(0.5)(0.5)^2$$

$$= \frac{3}{8}$$

pdf $\rightarrow p_X(k) = P(X=k)$

cum $\rightarrow F_X(k) = P(X \le k)$

The definition of the cdf function $F_X(t) = \sum_{k \le t} p_X(k)$ is calculated by the sum of the pdf functions $p_X(k)$.

We can also calculate the pdf $p_X(k)$ using the cdf function $F_X(t)$ by the following proposition.

Theorem.

For any discrete random variable X, the cdf function $F_X(t)$ and the pdf function $p_X(k)$ satisfy

$$p_X(k) = F_X(k) - F_X(k-1).$$

Example 3. Suppose the cumulative distribution function (cdf) is given by

t	1	3	5	6	8	9	10	12	15	16	19
$F_X(t)$	$\frac{1}{39}$	$\frac{1}{36}$	$\frac{2}{37}$	$\frac{3}{35}$	$\frac{1}{11}$	$\frac{2}{13}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{2}{3}$	$\frac{33}{35}$	1

What is the pdf $P(X=9)$?

n

$P(X \le 9) = \frac{2}{13} = P(X \le 8) + P(X=9) = \frac{2}{13} = \frac{1}{11} + n$

$\Rightarrow$ solve for n

Example 4.
Suppose you have \$10 and you go to gamble. Each time, you will either win or lose \$1. Each time, the probability that you will win is $\frac{1}{3}$.

What is the **pdf** of your money situation after 6 times gamble?

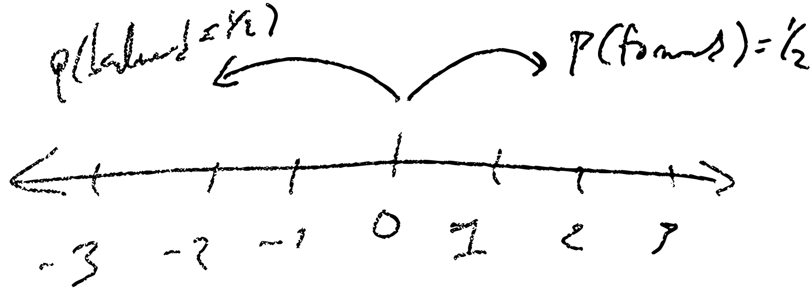
X: the number of times you win. $(X = 0, 1, 2, 3, 4, 5, 6)$ return from winning
The number of losing times is $6 - X$.
In the end, the money in your pocket is $Y = 10 + X - (6-X) = 4 + 2X$
(\$) from losing $6-X$
The probability of win k times in 6 gambles is

$$P(X=k) = \binom{6}{k} (1/3)^k (2/3)^{6-k}$$

k	0	1	2	3	4	5	6
$P(X=k)$	0.0878	0.2634	0.3294	0.2195	0.0823	0.0164	0.0014
$P(Y=4+2k)$	0.0878	0.2634	0.3294	0.2195	0.0823	0.0164	0.0014
$4+2k$	4	6	8	10	12	14	16

\$ situation

The mathematical model for this example called one dimensional **Random Walk**. Suppose that a dot sits on an integer number line. The dot starts in the center and start walk. For each step, it either forward or backward, with equal probability. We want to know where is the dot after it has taken k steps.


P(backward = 1/2)
P(forward) = 1/2
-3
-2
-1
0
1
2
3

§3.4 Continuous Random Variables

The range of a **continuous random variable** X is a (piecewise) continuous interval of $\mathbb{R}$.

Motivation Example. Choose a real number randomly from the interval $[0, 2]$ (sample space). If we assume the numbers are equally likely, we have the following the probabilities:

- $P(X \leq 2) = 1$
- $P(X \leq 0.2) = 0.1$
- $P(X \leq 0.02) = 0.01$

  $\vdots$

- $P(X \leq x) = x/2$

We can continue this and $P(X = 0) = 0$.

In fact, $P(X = a) = 0$ for any real number. So, we care about the probability for a interval.

Definition.

The **probability density function (pdf)** of a **continuous** random variable X is a piecewise continuous function $f_X(x)$ satisfying

1. $f_X(x) \geq 0$
2. $\displaystyle\int_{-\infty}^{\infty} f_X(x)dx = 1.$

We also define $f_X(x) = 0$ if x in not in the range of X.

Definition.

The **probability** that X is in an interval $[a, b]$ is

subset of sample space

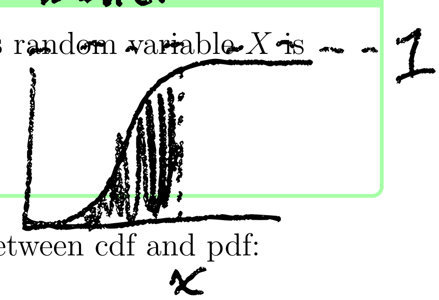
$$P(a \leq X \leq b) = \int_a^b f_X(x)dx$$

inclusive/exclusive doesn't matter

Definition.

The **cumulative distribution function (cdf)** of a **continuous** random variable X is

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)dt$$



From the Fundamental Theorem of Calculus, we have the relation between cdf and pdf:

$$\frac{\partial}{\partial x} F_X(x) = \frac{\partial}{\partial x}\int_{-\infty}^{x} f_X(t)\,dt \qquad F_X'(x) = f_X(x)$$

cdf: $P(X \le x)$

pdf: $P(a \le x \le b)$

Theorem.

$$F'_X(x) = f_X(x)$$

Proof by FTC

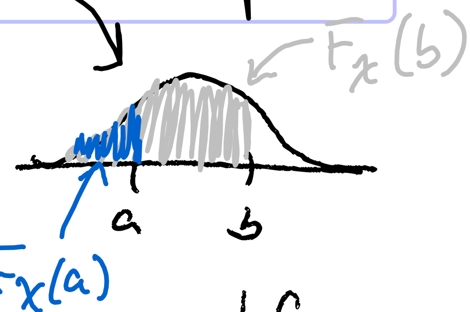
We can use the cdf to find the probability

Theorem.

Also proof by FTC (See next page)

$$P(a \le X \le b) = F_X(b) - F_X(a)$$

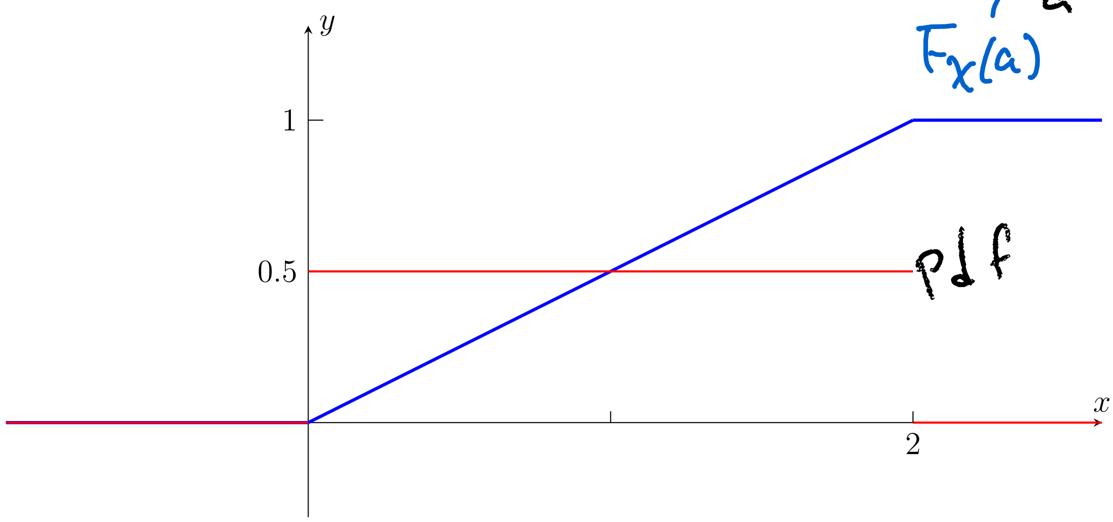
Visual proof



In our motivation example, we have cdf $F_X(x) = P(X \le x) = x/2$.

So, the pdf $f_X(x) = F'_X(x) = 1/2$ for $x \in [0, 2]$.

The graphs for the pdf $f_X(x)$ and the cdf $F_X(x)$:



pdf = derivative of the cdf

The **cdf** function is always a continuous, increasing function. The minimum is 0 and the maximum is 1.

Example 1. Choose a number randomly from the interval $[a, b]$. If we assume the numbers are equally likely. Find pdf $f_X(x)$ and cdf $F_X(x)$.

The cdf function is

The pdf function is

See next page

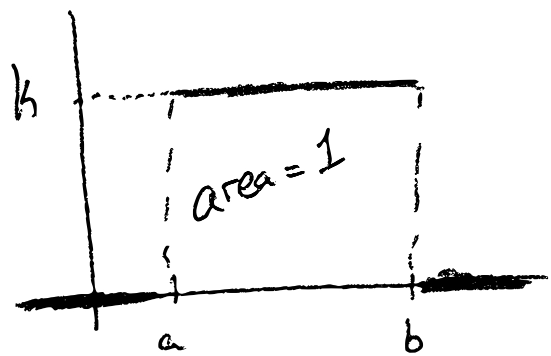
Example 2. Suppose the pdf of a random variable Y is $f_Y(y) = c \cdot y^3$ for $0 \le y \le 2$.

(1) Find c and calculate $P(0 \le Y \le 1)$.

Uniform Distribution

X is a uniform random var $\quad X \sim \text{uniform}(a,b)$

$$f_X(x) = \begin{cases} k & a \le x \le b \ (x \in S) \\ 0 & \text{otherwise} \end{cases}$$



$$\left.\begin{aligned} k(b-a) &= 1 \\ k &= \frac{1}{b-a} \end{aligned}\right\} \Longrightarrow f_X(x) = \begin{cases} \frac{1}{b-a} & x \in S \\ 0 & \text{otherwise} \end{cases}$$

Uniform cdf

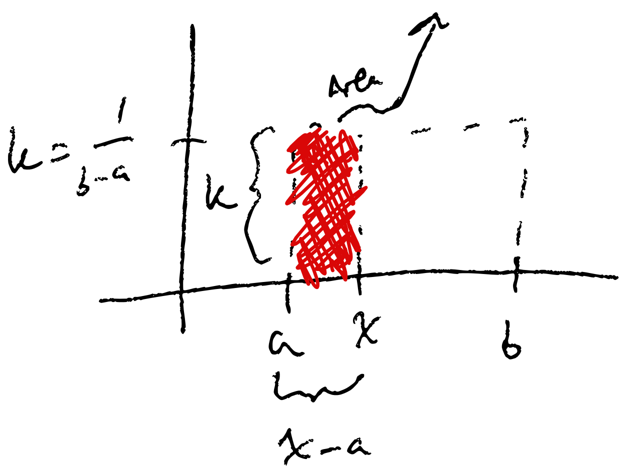
Recall:

Uniform(a, b)

$$f_X(x) = \begin{cases} \frac{1}{b-a} & x \in S \\ 0 & \text{otherwise} \end{cases}$$

Uniform cdf

$$F_X(x) = P(X \leq x) = \frac{x-a}{b-a}$$



$$P(a \le x \le b) = \int_a^b f_x(x)\,dx$$

$$= \text{antiderivative of } f_x(x) \Big|_a^b$$

$$F_x'(x) = f_x(x)$$ (also shown by FTC)

$$\therefore \; F_x(x) = \int f_x(x)\,dx$$

$$= F_x(b) - F_x(a)$$

QED

$f_Y(y) = cy^3$ for $0 \le y \le 2$

find c s.t. $f_Y(y)$ is pdf & find $P(0 \le Y \le 1)$

↳ $\int_0^1 \frac{1}{4} y^3 dy$

$$\int_{-\infty}^{\infty} cy^3 dy \overset{\text{set}}{=} 1 \Rightarrow \int_0^2 cy^3 dy = c\left(\frac{y^4}{4}\Big|_0^2\right) = 4c = 1 \Rightarrow c = \frac{1}{4}$$

$\therefore f_Y(y) = \frac{1}{4} y^3$

(2) Find the cdf $F_Y(y)$.

$$\int F_Y'(y)\,dy = \int f_Y(y)\,dy$$

$$F_Y(y) = \frac{1}{4} \cdot \frac{y^4}{4} = \frac{y^4}{16}, \quad 0 \le y \le 2$$

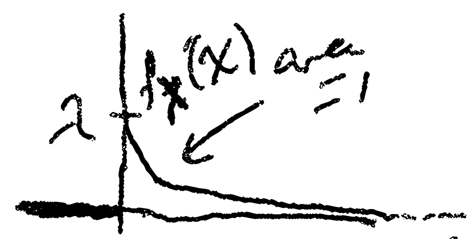
DON'T FORGET SAMPLE SPACE, or piecewise

Example 3. An important continuous distribution is the **exponential distribution** defined as

$$f_X(x) = \lambda e^{-\lambda x} \text{ for } x \ge 0$$

positive value

where λ is a positive parameter.



(1) Check that $f_X(x) \ge 0$ and $\int_{-\infty}^{\infty} f_X(x)dx = 1$.

It is clear that $f_X(x) \ge 0$. For the second equality,

$$P(S) = \int_{-\infty}^{\infty} f_X(x)dx = \int_0^{\infty} \lambda e^{-\lambda x} dx = \left[-e^{-\lambda x}\right]_0^{\infty} = 1$$

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

$\int e^{-\lambda x} = \frac{-1}{\lambda} e^{-\lambda x}$ by u-sub

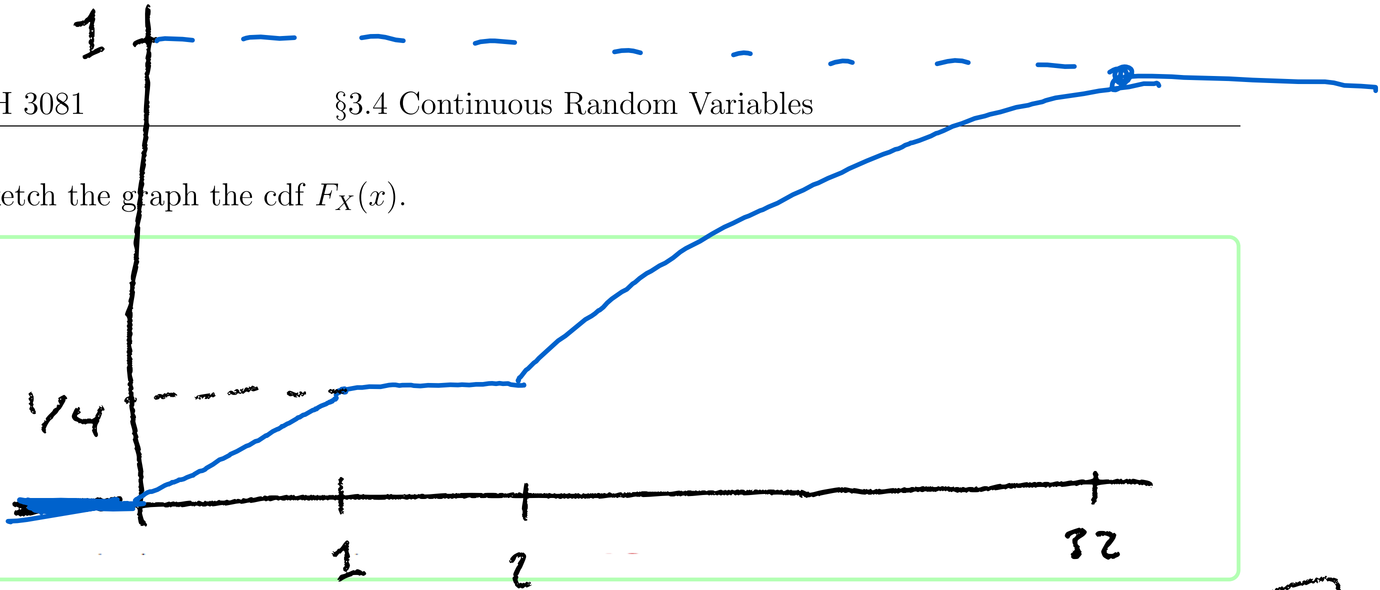
(2) Calculate the cdf $F_X(x)$ of X.

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt = \int_0^x \lambda e^{-\lambda t} dt = \left[-e^{-\lambda t}\right]_0^x = -e^{-\lambda x} - (-1) = 1 - e^{-\lambda x}, \quad x \ge 0$$

Example 4. The cdf of a random variable X is given by

$$F_X(x) = \begin{cases} 0 & x < 0 \\ x/4 & 0 \le x < 1 \\ 1/4 & 1 \le x < 2 \\ \sqrt{2x}/8 & 2 \le x < 32 \\ 1 & x \ge 32 \end{cases}$$

(1) Sketch the graph the cdf $F_X(x)$.



(2) Find $P(1/2 \le X < 4)$.

$$P\left(\tfrac{1}{2} \le X < 4\right) = F_X(4) - F_X\left(\tfrac{1}{2}\right) = \frac{\sqrt{8}}{8} - \frac{1}{8} = \boxed{\frac{\sqrt{8}-1}{8}}$$

(3) Find $P(X > 4)$.

$$P(X>4) = 1 - P(X \le 4) = 1 - F_X(4) = \boxed{1 - \frac{\sqrt{8}}{8}}$$

(4) Find the pdf of X.

The pdf of X is given by

$$f_X(x) = F_X'(x) = \begin{cases} 0 & x < 0 \\ 1/4 & 0 \le x < 1 \\ 0 & 1 \le x < 2 \\ \dfrac{\sqrt{2}}{8} \cdot \dfrac{1}{2} x^{-\frac{1}{2}} & 2 \le x < 32 \\ 0 & x \ge 32 \end{cases}$$