TA Session 6 08/09/2018

Agenda:

- σ-algebara generated by a random variable
- Simulating CDFs.
- Joint CDFs - Properties and example
- Independence of random variables.
- The Riemann Theory of integration
  - → Why does this fail?
  - → The remedy: Lebesgue theory of integration.
- Summary of independence.

$\boxed{\text{I}}$ σ-algebra generated by a random variable

We shall begin with a simple example. Let

$$\Omega = \{H, T\}$$
$$\mathcal{F} = 2^{\Omega} = \{\phi, \Omega, \{H\}, \{T\}\}$$

Define a rv $X : \Omega \to \mathbb{R}$ as

$$X(H) = 1 \quad , \quad X(T) = 0.$$

Since X is a rv (wrt $\mathcal{F}$, of course), we know that $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}$ for all $B \in \mathcal{B}(\mathbb{R})$. In particular, we have

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} = \begin{cases} \phi, & \text{if } 0 \notin B,\ 1 \notin B \\ \{T\}, & \text{if } 0 \in B,\ 1 \notin B \\ \{H\}, & \text{if } 1 \in B,\ 0 \notin B \\ \Omega, & \text{otherwise.} \end{cases}$$

Let us define another random variable $Y : \Omega \to \mathbb{R}$ as

$$Y(H) = 0, \quad Y(T) = 0.$$

Then, we see that

$$Y^{-1}(B) = \{\omega \in \Omega : Y(\omega) \leq x\} = \begin{cases} \phi, & \text{if } 0 \notin B \\ \Omega, & \text{if } 0 \in B. \end{cases}$$

Thus, we notice that while X as a random variable seems to exploit the full structure of $\mathcal{F}$, Y as a random variable does not. Y seems to exploit only the $\{\phi, \Omega\}$ part of $\mathcal{F}$.

Let us consider one more example to highlight this point. Let

$$\Omega = \{1, 2, 3, 4\}$$

$$\mathcal{F} = \{\phi, \Omega, \{1\}, \{4\}, \{2,3\}, \{1,4\}, \{1,2,3\}, \{2,3,4\}\}.$$

Let us define $X : \Omega \to \mathbb{R}$ as

$$X(\omega) = \begin{cases} \pi/2, & \text{if } \omega = 1 \\ \pi, & \text{if } \omega = 4 \\ 3\pi/2, & \text{if } \omega = 2,3. \end{cases}$$

Then, we see that

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} = \begin{cases} \phi, & \text{if } \pi/2 \notin B,\ \pi \notin B,\ 3\pi/2 \notin B \\ \{1\}, & \text{if } \pi/2 \in B,\ \pi \notin B,\ 3\pi/2 \notin B \\ \{4\}, & \text{if } \pi/2 \notin B,\ \pi \in B,\ 3\pi/2 \notin B \\ \{1,4\}, & \text{if } \pi/2 \in B,\ \pi \in B,\ 3\pi/2 \notin B \\ \{2,3\}, & \text{if } \pi/2 \notin B,\ \pi \notin B,\ 3\pi/2 \in B \\ \{1,2,3\}, & \text{if } \pi/2 \in B,\ \pi \notin B,\ 3\pi/2 \in B \\ \{2,3,4\}, & \text{if } \pi/2 \notin B,\ \pi \in B,\ 3\pi/2 \in B \\ \Omega & \text{otherwise}, \end{cases}$$

Whereas if we define another random variable Y as

$$Y(\omega) = \begin{cases} 0, & \text{if } \omega = 1,2 \\ 1, & \text{if } \omega = 3,4, \end{cases}$$

then we see that

$$Y^{-1}(B) = \{\omega \in \Omega : Y(\omega) \in B\} = \begin{cases} \phi, & \text{if } 0 \notin B,\ 1 \notin B \\ \{1,2\}, & \text{if } 0 \in B,\ 1 \notin B \\ \{3,4\}, & \text{if } 0 \notin B,\ 1 \in B \\ \Omega & \text{otherwise.} \end{cases}$$

Thus, while X exploits the full structure of $\mathcal{F}$, Y exploits only a part of it.

Thus, given $(\Omega, \mathcal{F})$ and an $\mathcal{F}$-measurable random variable X, we saw that X need not exploit the full structure of $\mathcal{F}$ in terms of its pre-image map $X^{-1}(\cdot)$. That part of $\mathcal{F}$ which X fully exploits is called the "σ-algebra generated by X", denoted by $\sigma(X)$. Mathematically,

$$\sigma(X) = \{A \subseteq \Omega : A = X^{-1}(B) \text{ for some } B \in \mathcal{B}(\mathbb{R})\}.$$

It is easy to check that $\sigma(X)$ is indeed a σ-algebra of subsets of Ω (Homework 1, q8). It is a sub-σ-algebra of $\mathcal{F}$. It is the smallest σ-algebra with respect to which X will be a random variable.

II Joint CDFs - Examples and properties

Let $(\Omega, \mathcal{F}, P)$ be a probability space, and let $X : \Omega \to \mathbb{R}$ be an $\mathcal{F}$-measurable random variable. The CDF of X is defined as a function $F : \mathbb{R} \to [0,1]$, given by

$$F(x) = P\left(\{\omega \in \Omega : X(\omega) \le x\}\right).$$

We now recall some properties that F satisfies.

for proofs see Grimmett

① F is non-decreasing, i.e., if $x \in \mathbb{R}$, $y \in \mathbb{R}$, $x < y$, then $F_X(x) \le F_X(y)$.

② F is right continuous, i.e; $\forall\, x \in \mathbb{R}$, we have

$$\lim_{n \to \infty} F\left(x + \frac{1}{n}\right) = F(x).$$

(A proof of this uses continuity of probability)

③ $\lim_{x \to \infty} F(x) = 1$, $\quad \lim_{x \to -\infty} F(x) = 0$

④ For any $-\infty < x_1 < x_2 < \infty$, we have

$$P(x_1 < X \le x_2) = F(x_2) - F(x_1).$$

On similar lines, we can define the joint CDF of two random variables X and Y. As before, let $(\Omega, \mathcal{F}, P)$ be a probability space, and let X and Y be two $\mathcal{F}$-measurable random variables. Then, the joint CDF of X and Y is defined as a function $F : \mathbb{R}^2 \longrightarrow [0,1]$ given by

$$F(x,y) = P\left(\{\omega \in \Omega : X(\omega) \le x\} \cap \{\omega \in \Omega : Y(\omega) \le y\}\right)$$

for any $x, y \in \mathbb{R}$.

Remark: Since X is $\mathcal{F}$-measurable,

$$\{\omega \in \Omega : X(\omega) \le x\} \in \mathcal{F} \qquad \forall\, x \in \mathbb{R}$$

Since Y is $\mathcal{F}$-measurable,

$$\{\omega \in \Omega : Y(\omega) \le y\} \in \mathcal{F} \qquad \forall\, y \in \mathbb{R}.$$

Thus, for any $x, y \in \mathbb{R}$,

$$\{\omega \in \Omega : X(\omega) \leq x\} \cap \{\omega \in \Omega : Y(\omega) \leq y\} \in \mathcal{F},$$

and therefore $P\left(\{\omega \in \Omega : X(\omega) \leq x\} \cap \{\omega \in \Omega : Y(\omega) \leq y\}\right)$ makes sense.

Sometimes, the following shorthand notations are used to denote the joint CDF of X and Y:

- $P\left(\{\omega \in \Omega : X(\omega) \leq x, Y(\omega) \leq y\}\right)$

In this notation, the comma should be read as "and", which in the language of set theory means intersection.

- $P\left(X \leq x, Y \leq y\right)$.

We shall see some examples of joint CDFs soon. But before we do so, given a function let's say $F(x,y)$, is there any way to check that it is actually a joint CDF? Are there any properties that we can quickly check to assert whether a given function $F(x,y)$ is indeed a joint CDF? Turns out YES!

Properties of joint CDFs

Let $(\Omega, \mathcal{F}, P)$ be a probability space, and let X and Y be two $\mathcal{F}$-measurable random variables. Further, let $F : \mathbb{R}^2 \to [0,1]$ denote the joint CDF of X and Y. Then, F satisfies the following properties:

a) $\lim_{y\to\infty} F(x,y) = F_X(x)$, $\lim_{x\to\infty} F(x,y) = F_Y(y)$

Here, $F_X(\cdot)$ denotes the CDF of X and F_Y denotes the CDF of Y.

Proof: We shall prove only the first one. The other follows on similar lines. Fix any $x \in \mathbb{R}$, and for each $n \geq 1$, define a set A_n as

$$A_n = \{\omega \in \Omega : X(\omega) \leq x\} \cap \{\omega \in \Omega : Y(\omega) > n\}.$$

Notice that

$$A_1 \supseteq A_2 \supseteq A_3 \supseteq \cdots, \text{ and moreover,}$$

$$\lim_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} A_n = \bigcap_{n=1}^{\infty} \{\omega \in \Omega : X(\omega) \leq x\} \cap \{\omega \in \Omega : Y(\omega) > n\}$$

$$= \{\omega \in \Omega : X(\omega) \leq x\} \cap \underbrace{\bigcap_{n=1}^{\infty} \{\omega \in \Omega : Y(\omega) > n\}}_{= \phi \text{ (why?)}}.$$

$$= \phi.$$

By using Continuity of probability, we get

$$0 = P(\phi) = P\left(\lim_{n\to\infty} A_n\right) = \lim_{n\to\infty} P(A_n)$$

$$\Rightarrow \lim_{n\to\infty} P(\{\omega \in \Omega : X(\omega) \leq x\} \cap \{\omega \in \Omega : Y(\omega) > n\}) = 0.$$

$$\Rightarrow \lim_{y\to\infty} P\left(\{\omega \in \Omega : X(\omega) \leq x\} \cap \{\omega \in \Omega : Y(\omega) > y\}\right) = 0 \quad \text{(1)}$$

We now use the fact that for each $y \in \mathbb{R}$,

$$\{\omega \in \Omega : X(\omega) \le x\} = \left(\{\omega\in\Omega : X(\omega) \le x\} \cap \{\omega \in \Omega : Y(\omega) > y\}\right)$$

$$\uplus$$

$$\left(\{\omega\in\Omega : X(\omega) \le x\} \cap \{\omega\in\Omega : Y(\omega) \le y\}\right).$$

Applying $P(\cdot)$ on both sides and using ①, we get the result.

b) Suppose $x_1, x_2, y_1, y_2 \in \mathbb{R}$ s.t.

$x_1 \le x_2$

$y_1 \le y_2$.

Then,

$$F(x_1, y_1) \le F(x_2, y_2) \quad \text{(monotonicity property)}$$

Proof: If $x_1 \le x_2$, then it follows that

$$\{\omega\in\Omega : X(\omega) \le x_1\} \subseteq \{\omega\in\Omega : X(\omega) \le x_2\}.$$

Similarly, if $y_1 \le y_2$, then

$$\{\omega\in\Omega : Y(\omega) \le y_1\} \subseteq \{\omega\in\Omega : Y(\omega) \le y_2\}.$$

Thus, it follows that

$$\{\omega\in\Omega : X(\omega) \le x_1, Y(\omega) \le y_1\} \subseteq \{\omega\in\Omega : X(\omega) \le x_2, Y(\omega) \le y_2\},$$

and by applying $P(\cdot)$ on both sides and using $P(A) \le P(B)$ if $A \subseteq B$, we get the result.

c) For any $a, b, c, d \in \mathbb{R}$ s.t.

$$-\infty < a < b < \infty$$

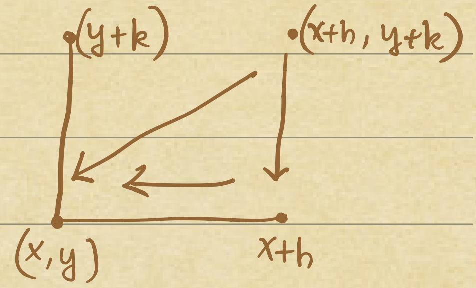
$$-\infty < c < d < \infty,$$ we have

$$P(a < X \leq b,\ c < Y \leq d)$$
$$= F(b,d) - F(a,d) - F(b,c) + F(a,c)$$

Proof: Exercise.

d) F is continuous from above and right, i.e.,

$$F(x+h, y+k) \longrightarrow F(x,y) \text{ as } h, k \downarrow 0.$$



Proof: It suffices to show that

$$F\left(x+\frac{1}{m}, y+\frac{1}{n}\right) \longrightarrow F(x,y) \text{ as}$$
$$m, n \to \infty.$$

Towards this, note that for any fixed $n \in \mathbb{N}$, defining

$$A_{mn} = \left\{\omega \in \Omega : X(\omega) \leq x + \frac{1}{m}\right\} \cap \left\{\omega \in \Omega : Y(\omega) \leq y + 1/n\right\},$$

we have $A_{1n} \supseteq A_{2n} \supseteq \dots$, and moreover,

$$\lim_{m\to\infty} A_{mn} = \{\omega \in \Omega : X(\omega) \leq x\} \cap \left\{\omega \in \Omega : Y(\omega) \leq y + \frac{1}{n}\right\}.$$

Let $B_n = \lim_{m\to\infty} A_{mn}$. Again observe that

$$B_1 \supseteq B_2 \supseteq \dots,$$ and that

$$\lim_{n\to\infty} B_n = \{\omega \in \Omega : X(\omega) \leq x\} \cap \{\omega \in \Omega : Y(\omega) \leq y\}.$$

Apply continuity of probability to get the result.

Examples :

1. $F(x,y) = \begin{cases} 1-e^{-x} - e^{-y} + e^{-x-y}, & x, y \geq 0 \\ 0, & \text{otherwise.} \end{cases}$

Verify that this is indeed a valid joint CDF.

2. $F(x,y) = \begin{cases} 1-e^{-xy}, & 0 \leq x, y < \infty. \\ 0, & \text{otherwise} \end{cases}$

Is this a valid joint CDF ?

Ans: No. To see this, assume that F as defined is indeed a valid joint CDF of X and Y. Then, we can derive the CDF of X by

$$F_X(x) = \lim_{y \to \infty} F(x,y)$$

$$= \begin{cases} 1, & \text{if } x \geq 0 \\ 0, & x < 0. \end{cases}$$

This implies that $P(X=0) = 1$. On similar lines, we get $P(Y=0) = 1$. Thus, clearly, we have $P(X \leq 1, Y \leq 1) = 1$.

However, from the definition, we have

$$P(X \leq 1, Y \leq 1) = F(1,1) = 1-e^{-1} = 1-\frac{1}{e} \neq 1.$$

This is a contradiction. Hence, F is not a valid joint CDF.

Remark : Given a joint CDF, we know that properties a) – d) above must be true. It turns out that the converse is also true, i.e., any function $F : \mathbb{R}^2 \to \mathbb{R}$ that satisfies a) - d) above must be a joint CDF. It is this important converse result which we used to check whether a given F is indeed a joint CDF or not (We do not provide a proof of this converse result)

<u>Example</u> :

Let $(\Omega, \mathcal{F}, P)$ be a probability space, and let $A, B \in \mathcal{F}$ be two events. Define two $\mathcal{F}$-measurable random variables X and Y as

$$X = \mathbb{1}_A \,, \quad Y = \mathbb{1}_B .$$

What is the joint CDF of X and Y?

<u>Ans:</u> Let the joint CDF be denoted by $F(x,y)$, $x, y \in \mathbb{R}$.

Then,

$$F(x,y) = P(X \leq x, Y \leq y) = P\left(\mathbb{1}_A \leq x, \mathbb{1}_B \leq y\right)$$

$$= \begin{cases} 0, & \text{if } x < 0, \; y < 0 \\ 1, & \text{if } x \geq 1, \; y \geq 1 \\ P(A^c \cap B^c), & 0 \leq x < 1, \; 0 \leq y < 1 \\ P(A^c), & 0 \leq x < 1, \; y \geq 1 \\ P(B^c), & x \geq 1, \; 0 \leq y < 1 \end{cases}$$

Independence of random variables

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let X and Y be two $\mathcal{F}$-measurable random variables. Then, X and Y are said to be independent if and only if

$$F(x,y) = F_X(x) \cdot F_Y(y) \qquad \text{for all } x, y \in \mathbb{R},$$

where $F(x,y)$ denotes the joint CDF of X and Y,

$F_X(\cdot)$ denotes the CDF of X, and

$F_Y(\cdot)$ denotes the CDF of Y.

Equivalently, one can define independence of X and Y as:

X and Y are independent if and only if:

$\{\omega \in \Omega : X(\omega) \in A\}$ and $\{\omega \in \Omega : Y(\omega) \in B\}$ are independent for every possible choice of $A, B \in \mathcal{B}(\mathbb{R})$, i.e.,

$$P\big(X \in A,\ Y \in B\big) = P\big(X \in A\big) \cdot P\big(Y \in B\big) \quad \forall\ A, B \in \mathcal{B}(\mathbb{R}).$$

Remark: ① If X and Y are jointly discrete, and their joint pmf is denoted by $p_{X,Y}(x,y)$, $x, y \in \mathbb{R}$, then X and Y are independent iff (if and only if)

$$p_{X,Y}(x,y) = p_X(x) \cdot p_Y(y) \qquad \forall\ x, y \in \mathbb{R},$$

where $p_X(\cdot)$ and $p_Y(\cdot)$ denote the pmfs of X and Y respectively.

(2) If X and Y are jointly continuous, and $f_{X,Y}(x,y)$ denotes the joint pdf of X and Y, then X and Y are independent if and only if

$$f_{X,Y}(x,y) = f_X(x)\, f_Y(y) \qquad \forall\, x, y \in \mathbb{R},$$

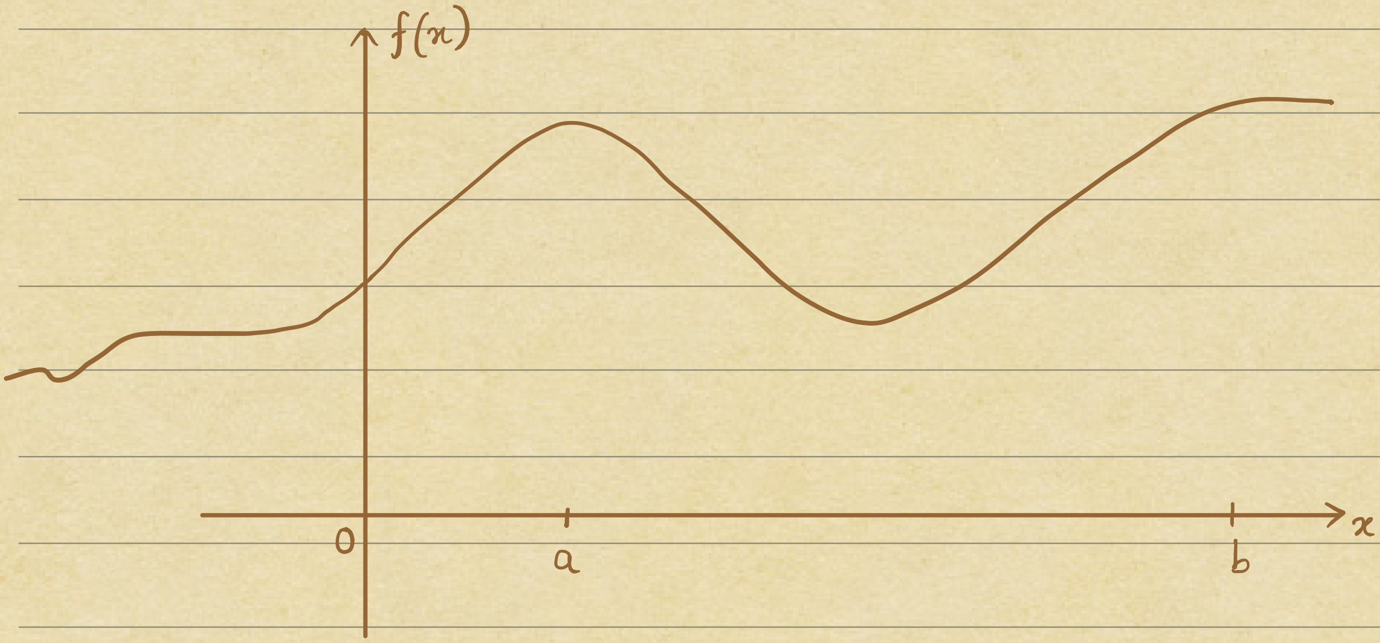
where $f_X(\cdot)$ and $f_Y(\cdot)$ denote the pdfs of X and Y respectively.

Expectations of Random Variables

We now come to a very important topic, that is of considering integration of random variables. Before we embark upon this, we shall begin by reviewing the theory of integration we have learnt in calculus.

Review of integration – The Riemann Integral

Let us consider a continuous function $f : \mathbb{R} \to \mathbb{R}$ whose sketch is as follows:



What do we mean by the integral of $f(x)$ from $x=a$ to $x=b$?

We have learned that this stands for the area under the function curve $f(x)$ from $x=a$ to $x=b$. But there arises two questions in mind:

a) How do we compute the above area?

b) Can we compute such an "area under the curve" for any function $f: \mathbb{R} \to \mathbb{R}$? Are there functions for which we cannot compute such an area at all?

We answer these questions one by one.

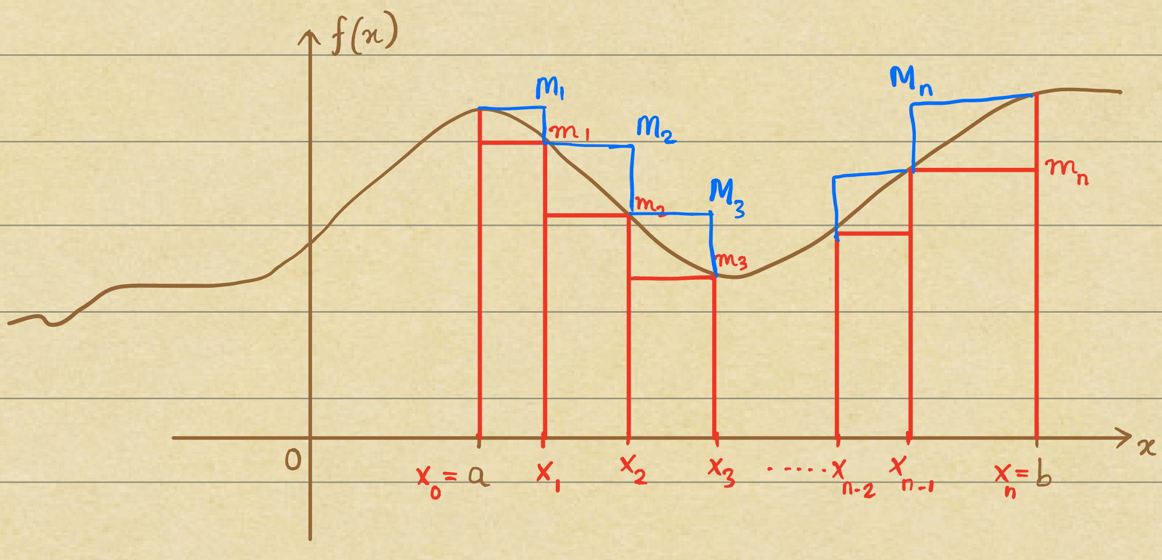
Answer to a):

Mathematicians came up with the following way to compute the area under the curve $f(x)$ from $x=a$ to $x=b$:

First, divide the interval $[a,b]$ into smaller sub-intervals. Label the end points of each sub-interval as

$$x_0 = a < x_1 < x_2 < \cdots < x_{n-1} < x_n = b.$$

as shown below:



The set

$$P := \{x_0 = a, x_1, x_2, \dots, x_{n-1}, x_n = b\}$$

is called a "partition" of the interval $[a, b]$.

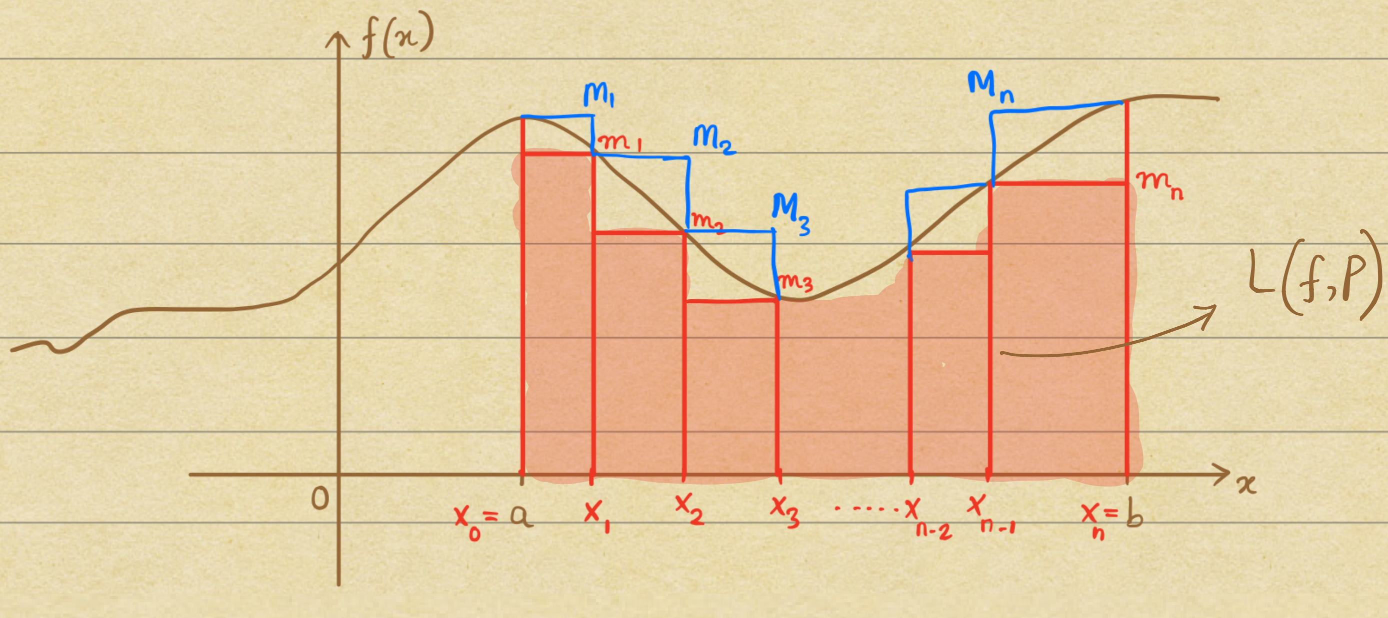
- Next, compute the minimum value of $f(x)$ in each sub-interval created by this partition. Let m_i denote the minimum value of f in the subinterval $[x_{i-1}, x_i]$, $i = 1, \dots, n$.

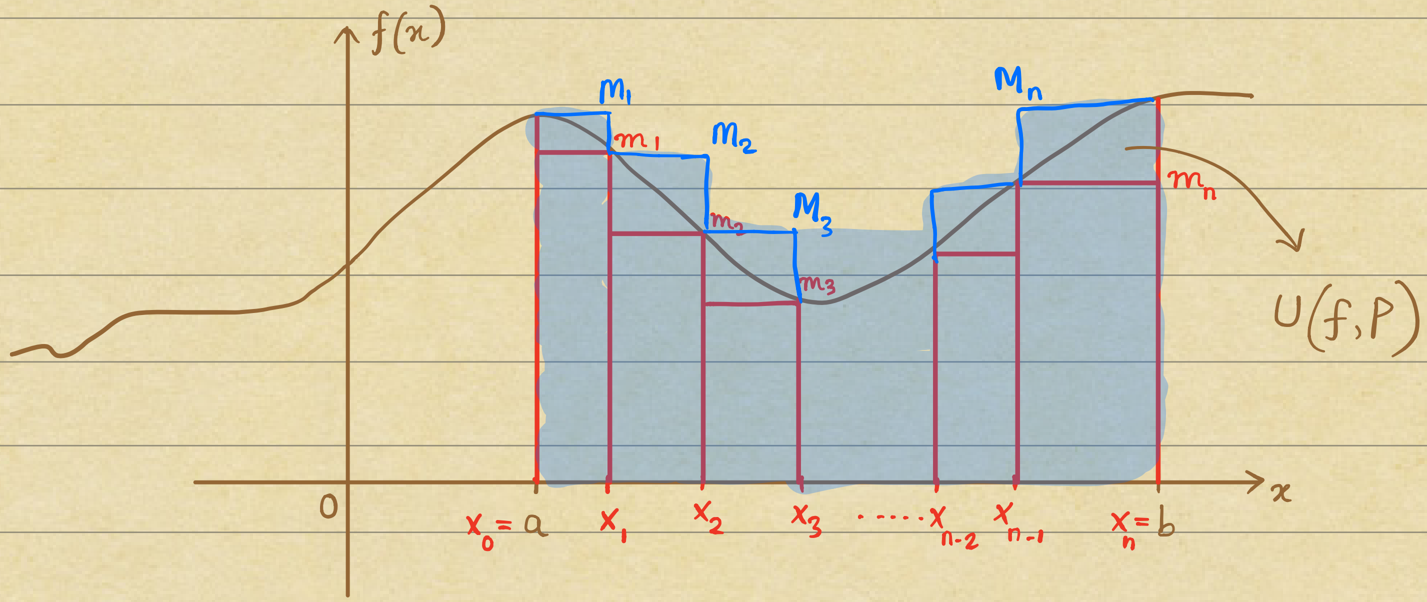
- Also compute the maximum value of $f(x)$ in each sub-interval. Let M_i denote the maximum value of f in the sub-interval $[x_{i-1}, x_i]$, $i = 1, \dots, n$.

- Now, evaluate two quantities $L(f, P)$ and $U(f, P)$ as follows:

$$L(f, P) = \sum_{i=1}^{n} m_i \cdot |x_i - x_{i-1}|$$ (represents total area under all red rectangles)

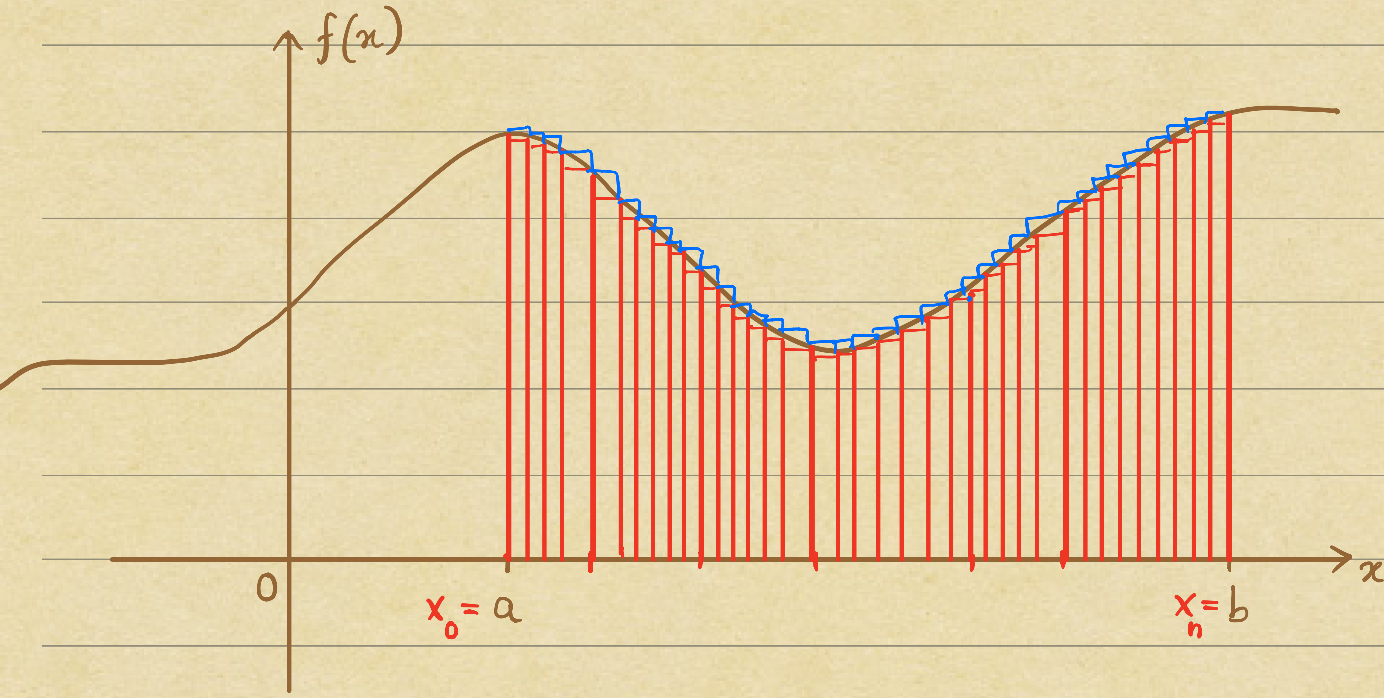
$$U(f, P) = \sum_{i=1}^{n} M_i \cdot |x_i - x_{i-1}|$$ (represents total area under all blue rectangles)





Notice that $L(f,P) \leq$ area under f from $x=a$ to $x=b$ $\leq U(f,P)$.

- Next, create a finer partition of $[a,b]$, say Q. For this partition, again compute $L(f,Q)$ and $U(f,Q)$. Repeat this process for finer and finer partitions.



- Mathematicians were able to show that if Q is a finer partition than P, then

$$L(f,P) \leq L(f,Q) \quad \text{and} \quad U(f,P) \geq U(f,Q).$$

In other words:

Total Area under red rectangles keeps monotonically increasing towards area under f as the partitions keep getting finer

and

Total area under blue rectangles keeps monotonically decreasing towards area under f as the partitions keep getting finer

Will the increasing and decreasing sequences eventually meet? This is what mathematicians asked. They came up with the following conclusion:

- For some functions, these sequences meet. The common value at the meeting point was known as the "area under f" or the "integral of f" from $x=a$ to $x=b$, denoted $\int_a^b f(x)\,dx$.

- For some functions, these sequences do not meet. In such a case, the "area under a function" for such functions does not exist / cannot be defined.

This method of integration that involves partitioning the x-axis is known as Riemann integration.

Examples of functions for which Riemann integral is well defined:

$f(x) = \sin x$, $\cos x$, $x \log x - x$, x^2, and all the nice functions for which we have carried out integration till now.

Example of a function for which Riemann integral is not defined

Let $f : [0,1] \to \mathbb{R}$ be defined as

$$f(x) = \begin{cases} 1, & \text{if } x \in \mathbb{Q} \\ 0, & \text{if } x \notin \mathbb{Q}. \end{cases} \quad (\text{Here } \mathbb{Q} : \text{set of rationals})$$

Then, we see that $L(f, P) = 0$ for any partition P of $[0,1]$.

$$U(f, P) = 1$$

Therefore, there is no hope of the sequences meeting.

So, mathematicians were looking for a way to remedy this problem of Riemann integral not being well-defined for some important functions of interest. Lebesgue then came up with the following idea:

Instead of partitioning x-axis as in Riemann integration, partition the y-axis !!

Out of this idea came a new theory of integration.

The Lebesgue Theory of Integration

Lebesgue developed the theory in 3 stages:

1. For "simple" functions
2. For non-negative functions
3. For arbitrary measurable functions.

(You have seen this in class.)

Let $(\Omega, \mathcal{F}, P)$ be a probability space, and let $X : \Omega \to \mathbb{R}$ be an $\mathcal{F}$-measurable random variable. Then, the Lebesgue integral of X (also called expectation of X), denoted by $E[X]$, is defined as

$$E[X] = \int_{\Omega} X(\omega)\, dP(\omega).$$

Going back to the example for which Riemann integration failed, taking $\Omega = [0,1]$, $\mathcal{F} = \mathcal{B}([0,1])$ and P any probability measure, we get

$$E[X] = \int_{\mathbb{Q} \cap [0,1]} 1 \cdot dP(\omega) = P(\mathbb{Q} \cap [0,1]),$$

which is well defined.

Back to independence of rvs

If X and Y are independent, then

$$E[XY] = E[X] \cdot E[Y].$$

Is the converse true? See hw2, q4.

Summary of independence of rvs

- $(\Omega, \mathcal{F}, P)$ probability space
- X, Y $\mathcal{F}$-measurable random variables.

X and Y independent

$$\iff F_{X,Y}(x,y) = F_X(x)\, F(y) \quad \forall x,y \in \mathbb{R}$$

$$\overset{\text{if jointly discrete}}{\iff} p_{X,Y}(x,y) = p_X(x)\, p_Y(y) \quad \forall x,y \in \mathbb{R}$$

$$\underset{\text{Continuous}}{\overset{\text{if jointly}}{\iff}} f_{X,Y}(x,y) = f_X(x)\, f_Y(y) \quad \forall x,y \in \mathbb{R}$$

$$\iff P(X\in A, Y\in B) = P(X\in A)\cdot P(Y\in B) \quad \forall A,B \in \mathcal{B}(\mathbb{R})$$

$$\implies E[XY] = E[X]\cdot E[Y].$$

$$\not\Longleftarrow$$