ECSE 2500

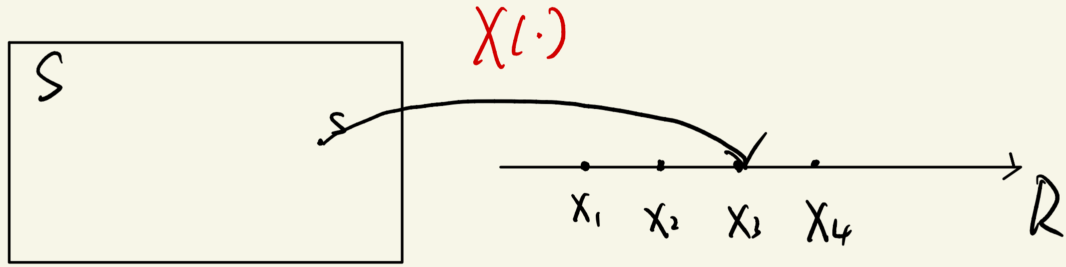
Lec 11

Feb 27

Topic: Cumulative Distribution Function (CDF)

PMF (Discrete RV) $\longrightarrow$ PDF (Continuous)

- [ ] Recall PMF for Discrete RV



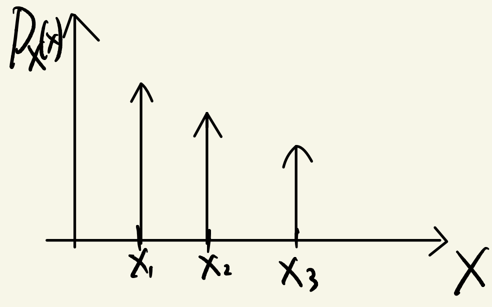
All possible outcomes of X are collected into

$$S_X := \{x_1, x_2, x_3, x_4, \dots\}$$

which can be ∞ many, e.g., Geometric RV

We can write the probability mass function (PMF) as

$$P_X(k) = \begin{cases} \underline{\qquad}, & k = x_1 \\ \underline{\qquad}, & k = x_2 \\ \vdots & \end{cases}$$



$$\sum_{k \in S_X} P_X(k) = 1$$

☐ Today we define Cumulative Distribution Function (CDF) of a RV X:

$$F_X(x) = P(X \le x), \quad -\infty < x < \infty$$

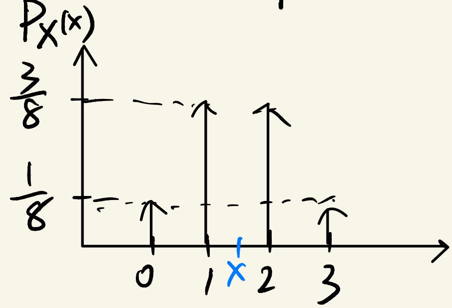
Big case: RV itself

Small case: Value/Outcome of a random variable

That is, what is Probability X is less than a given value? Define CDF for any RV X (whatever discrete or continuous).

Example: Flip coin 3 times, count # of heads.

$$S_X := \{0, 1, 2, 3\}$$



Q: What will be the CDF of X like?

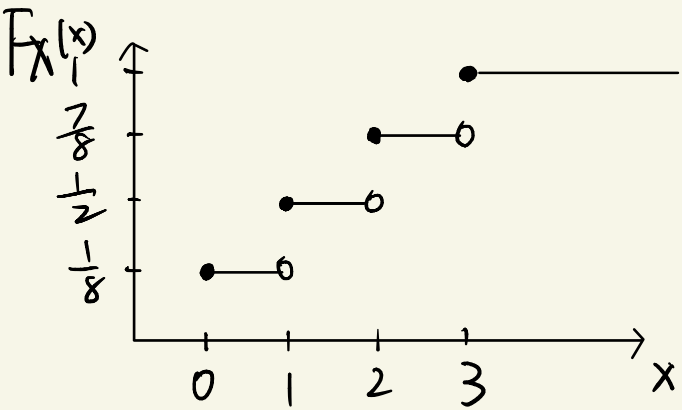
Then we can compute CDF as

$$F_X(x) = 0 \quad , \text{ if } x < 0$$

$$F_X(x) = 1 \quad , \text{ if } x > 3$$

$$F_X(1.5) = \frac{1}{8} + \frac{3}{8} = \frac{1}{2}$$

We can see in fact CDF $F_X(x)$ is a step function



We can also write

$$F_X(x) = \int_{-\infty}^{x_-} P_X(y)\, dy,$$

we sum up the impulse function up to the x (But not including x).

Since $\int_{-\infty}^{x} \delta(y)\,dy = u(x) = \begin{cases} 1, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0 \end{cases}$

we can write the equivalent form of PMF

$$P_X(x) = \sum_{x_k \in S_X} P_X(x_k)\,\delta(x - x_k).$$

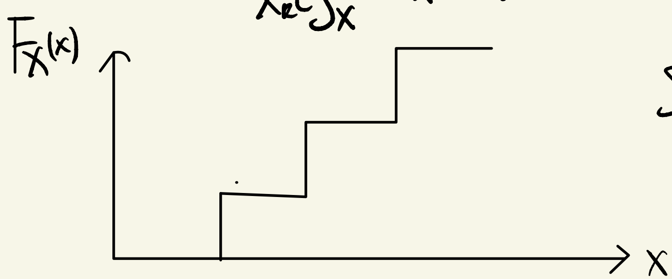
so we can write the equivalent form of CDF

$$F_X(x) = \int_{-\infty}^{x} P_X(y)\,dy$$

$$= \int_{-\infty}^{x} \sum_{x_k \in S_X} P_X(x_k)\,\delta(y - x_k)\,dy$$

$$= \sum_{x_k \in S_X} P_X(x_k) \int_{-\infty}^{x} \delta(y - x_k)\,dy$$

$$= \sum_{x_k \in S_X} P_X(x_k)\,u(x - x_k)$$



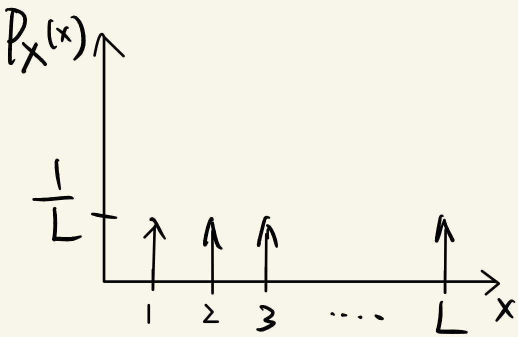
Sum of step functions

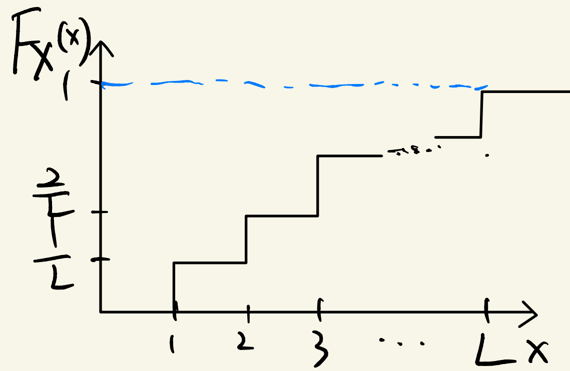
For the coin flipping example we have mentioned, we can write the CDF as

$$F_X(x) = \frac{1}{8}u(x) + \frac{3}{8}u(x-1) + \frac{3}{8}u(x-2) + \frac{1}{8}u(x-3)$$

We have more examples on CDF for other RVs.

$\boxed{\text{Example}}$ Uniform random variable on $\{1, 2, \cdots, L\}$





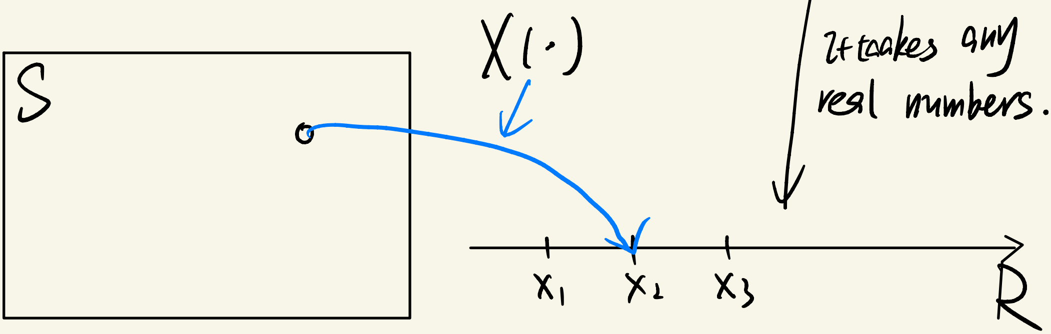
$\boxed{\text{Example}}$ Geometric Random Variable

$$P_X(k) = (1-p)^{k-1} p, \quad k = 1, 2, 3, \cdots$$

$$F_X(k) = \begin{cases} 0, & \text{if } k < 1 \\ p, & \text{if } 1 \le k < 2 \\ p + (1-p)p, & \text{if } 2 \le k < 3 \end{cases}$$

$$\begin{cases} \sum_{z=0}^{n-1} (1-p)^z p \quad \text{if } n \le k < n+1 \\ = p \cdot \dfrac{1-(1-p)^n}{1-(1-p)} = 1-(1-p)^n \end{cases}$$

☐ Now we are ready to talk about continuous RVs.

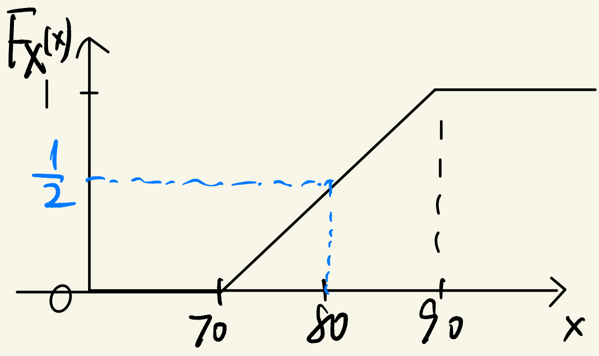


Example Length of time a class lasts.

Say this is a continuous random variable uniformly distributed between 70 mins to 90 mins.

$F_X(x) = 0, \quad x < 70$

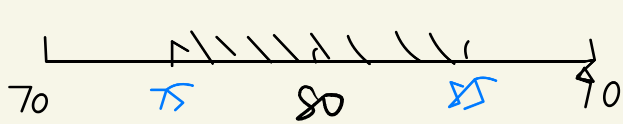
$F_X(x) = 1, \quad x > 90$



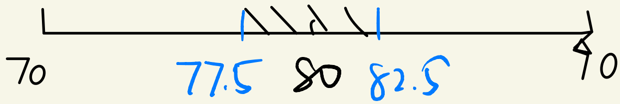
Something counterintuitive!

$$P(X = 80\text{min}) = 0 \quad ?$$

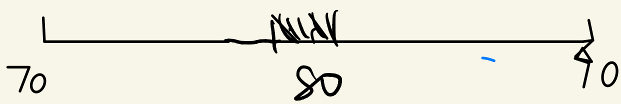
Think of it as follows



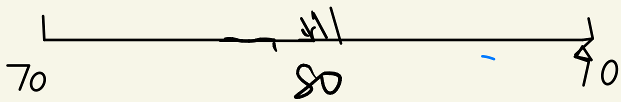
$\frac{1}{2}$



$\frac{1}{4}$



$\frac{1}{8}$



$\frac{1}{16}$

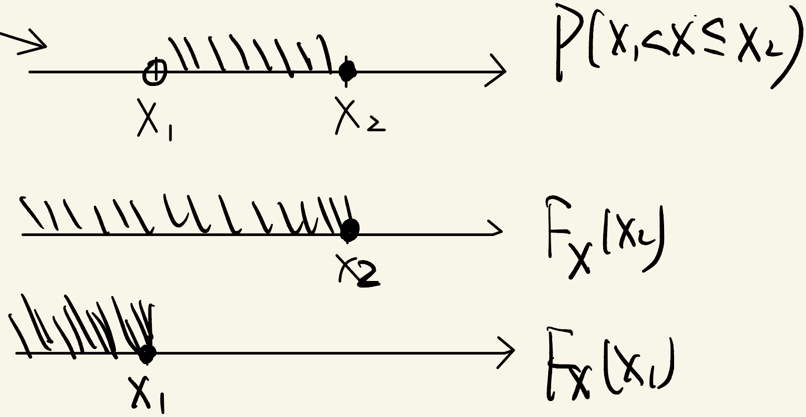
$P(X \in$ within 1 millsec of 80mins$)$ extremely small

$P(X =$ exactly 80 mins$) = 0$

So it make senses that

$$P(x_1 < X \leq x_2) = P(X \leq x_2) - P(X \leq x_1)$$
$$= F_X(x_2) - F_X(x_1)$$

We can think of we can "head off" the probability of any given interval by substituting the CDF at any fixed points.



$$\lim_{\delta \to 0} P(x < X \leq x + \delta) = 0$$

Remark: Therefore, using PMF to describe continuous RV does not provide too much info.

□ Probability Density Function

So in class duration case, we can compute that

$$F_X(x) = \begin{cases} 0, & x < 70 \\ \frac{x-70}{20}, & 70 \le x \le 90 \\ 1, & x > 90 \end{cases}$$

Thus, we have

$$P(75 < X \le 82) = F_X(82) - F_X(75)$$

$$= \frac{12}{20} - \frac{5}{20}$$

$$= \frac{7}{20}$$

Make sense!

Let say we want to $(a, b \in [70, 90])$

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

Continuous RV

$$= \frac{b-70}{20} - \frac{a-70}{20} = \frac{1}{20}(b-a)$$

It is also possible to describe a mixed (both discrete and continuous) RV.

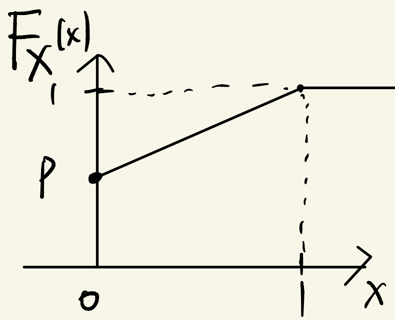
$\boxed{\text{Example}}$ X = # of minutes you wait for a taxi

$P(\text{Taxi is there}) = p$

$P(\text{Taxi is not there}) = 1-p$

If taxi is there, $X=0$

If taxi is not there, assume X is uniformly distributed in $[0,1]$ hour.



What is the CDF of this mixed RV?

$$F_X(0) = P(X=0) = p$$

$$F_X(x) = P(X \le x) = p + (1-p)x \quad , \; 0 < x \le 1$$

$$F_X(x) = P(X \le x) = 1 \quad , \; x > 1$$