Cumulative Distribution Function (CDF)

- A generalization of PMF and PDF.

**Def** The cumulative distribution function (CDF) of a random variable (discrete or continuous) is

\[ F_X(x) = P(X \leq x) \]

CDF for continuous r.v.:

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

Therefore, \( F_X(x) \) is the integration of the PDF from \(-\infty\) to \( x \).

On the other hand, since

\[ F_X(x) = \int_{-\infty}^{x} f_X(t) \, dt \]

by fundamental theorem of calculus, we have

\[ f_X(x) = \frac{d}{dx} \int_{-\infty}^{x} f_X(t) \, dt = \frac{d}{dx} F_X(x) \]
Fundamental theorem of Calculus:
If \( f \) is a continuous function, then
\[
\frac{d}{dx} F(x) = f(x)
\]
if \( F(x) = \int_{a}^{x} f(t) \, dt \).

The previous argument is in fact the original way of defining a PDF:

The probability density function (PDF) of a R.V. \( X \) is defined as
\[
f_X(x) = \frac{d}{dx} F_X(x)
\]
if the derivative exists.

In other words, if the derivative does not exist, then PDF is undefined.

CDF for discrete r.v.:
Recall that PMF is a stream of impulses.

If we follow the same integration/differentiation argument as in the continuous case, we can show that the impulses will be integrated into unit steps
To determine the gap between successive levels in the CDF, we note that

\[ P_1 + P_2 + \ldots + P_X \]

In fact, since PMF is discrete, we have

\[ P(X \leq x) = \sum_{i=1}^{x} P_i \]

E.g. \[ P(X \leq 2) = P(X=1) + P(X=2) \]

and \[ P(X \leq 4) = P(X=1) + P(X=2) + P(X=3) + P(X=4) \]

\[ \text{for } x \geq 0. \]

\[ \text{a r.v.} \]

By adding one additional term we increase the level of CDF by that probability mass.

Note also that the “” is solid on the left hand side. This is because CDF is defined as

\[ \underline{F}_{X}(x) = \mathbb{P}(X \leq x) \]

The equality in “\leq” makes the “” solid on the left.

The \( x \)-axis spacing of the CDF is determined by the elements in the sample space.

E.g. \( S_X = \{1, 3, 4, 7\} \quad S_X = \{2, 4, 6, 8\} \)

\( P_X = \{\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\} \quad P_X = \{\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\} \)
Properties of CDF

Basic ones:
1. \(0 \leq F_X(x) \leq 1\)
2. \(\lim_{x \to +\infty} F_X(x) = 1\)
3. \(\lim_{x \to -\infty} F_X(x) = 0\)
4. \(F_X(x)\) is non-decreasing \(\leq\) Non-decreasing means that \(F_X(x)\) can have
   that regions.
5. \(P(a \leq X \leq b) = F_X(b) - F_X(a)\)
6. \(P(X > x) = 1 - F_X(x)\).

More advanced ones:
7. \(F_X(x)\) is right continuous, i.e., solid "— |—
   \[\lim_{h \to 0} F_X(b + h) = F_X(b)\]

8. \(P(X = b) = F_X(b) - \underline{F_X(b)}\)
   \[= \lim_{h \to 0} F_X(b - h)\]

Remark: (7) and (8) are crucial properties for
   discrete R.V.

Remark: if \(F_X(x)\) is continuous at \(b\), then
   \(F_X(b) = F_X(b^-)\). In this case \(P(X = b) = 0\).
Example

\[ f_X(x) = \begin{cases} 
  c(1-x^2) & , \quad -1 \leq x \leq 1 \\
  0 & , \quad \text{otherwise}
\end{cases} \]

Find \(c\), find \(F_X(x)\).

\[ \int_{-\infty}^{\infty} f_X(x) \, dx = 1 \implies c \int_{-1}^{1} (1-x^2) \, dx = 1 \]
\[ \implies c \left[ x - \frac{1}{3} x^3 \right]_{-1}^{1} = 1 \]
\[ \implies c = \frac{3}{4}. \]

\[ F_X(x) = \int_{-\infty}^{x} f_X(t) \, dt \]
\[ = \begin{cases} 
  0 & , \quad x < -1 \\
  c \int_{-1}^{x} (1-t^2) \, dt & , \quad -1 \leq x \leq 1 \\
  1 & , \quad x > 1
\end{cases} \]
\[ = \begin{cases} 
  0 & \\
  c \left[ t - \frac{1}{3} t^3 \right]_{-1}^{x} & , \quad -1 \leq x \leq 1 \\
  1 & , \quad x > 1
\end{cases} \]
\[ = \begin{cases} 
  0 & , \quad x < -1 \\
  \frac{3}{4} \left( x - \frac{1}{3} x^3 + \frac{2}{3} \right) & , \quad -1 \leq x \leq 1 \\
  1 & , \quad x > 1
\end{cases} \]
Example

\[ F_X(x) = \begin{cases} 
0 & , 
\ x < 0 \\
1 - \frac{1}{4} e^{-2x} & , 
\ x \geq 0 
\end{cases} \]

Find \( f_X(x) \).

First, we can show that \( F_X(0) = \frac{3}{4} \).

There is a discontinuity at \( x = 0 \). Therefore, we can consider three cases:

\( x < 0, \ x = 0, \ x > 0 \).

This leads to

\[ f_X(x) = \begin{cases} 
\frac{d}{dx} F_X(x) & , 
\ x < 0 \\
\mathbb{P}(X = 0) & , 
\ x = 0 \\
\frac{d}{dx} F_X(x) & , 
\ x > 0 
\end{cases} \]

When \( x < 0 \), \( F_X(x) = 0 \). So \( \frac{d}{dx} F_X(x) = 0 \).

When \( x > 0 \), \( F_X(x) = 1 - \frac{1}{4} e^{-2x} \). So \( \frac{d}{dx} F_X(x) = \frac{1}{2} e^{-2x} \).

When \( x = 0 \),

\[ \mathbb{P}(X = 0) = F_X(0) - F_X(0^-) = \frac{3}{4} - 0 = \frac{3}{4}. \]

Therefore,

\[ f_X(x) = \begin{cases} 
0 & , 
\ x < 0 \\
\frac{3}{4} & , 
\ x = 0 \\
\frac{1}{2} e^{-2x} & , 
\ x > 0 
\end{cases} \]