Definition 1.125 (Probability Mass Function (p.m.f.))**.** Let X be a discrete RV with DF F_X and support S. Consider the function $f_X : \mathbb{R} \to \mathbb{R}$ defined by

$$f_X(x) := \begin{cases} F_X(x) - F_X(x-) = \mathbb{P}(X = x), \text{ if } x \in S, \\ 0, \text{ if } x \in S^c. \end{cases}$$

This function f_X is called the probability mass function (p.m.f.) of X.

Example 1.126. Continuing with the Example 1.124, the p.m.f. f_X is given by

$$f_X(x) = \begin{cases} \frac{1}{4}, \text{ if } x = 0, \\ \frac{1}{2}, \text{ if } x = 1, \\ \frac{1}{4}, \text{ if } x = 2., \\ 0, \text{ otherwise.} \end{cases}$$

Remark 1.127*.* Let X be a discrete RV with DF F_X, p.m.f. f_X and support S. Then we have the following observations.

(a) Continuing the discussion from Remark 1.122, we have for all $A \subseteq \mathbb{R}$,

$$\mathbb{P}_X(A) = \mathbb{P}(X \in A) = \sum_{x \in A \cap S} f_X(x).$$

(b) As a special case of the previous observation, note that for $A = (-\infty, x], x \in \mathbb{R}$, we obtain

$$F_X(x) = \mathbb{P}(X \le x) = \mathbb{P}(X \in (-\infty, x]) = \sum_{t \in (-\infty, x] \cap S} f_X(t).$$

Therefore, the p.m.f. f_X is uniquely determined by the DF F_X and vice versa.

(c) To study a discrete RV X, we may study any one of the following three quantities, viz. the law/distribution $\mathbb{P}_X$, the DF F_X or the p.m.f. f_X. Given any one of these quantities, the other two can be obtained using the relations described above.

(d) By Definition 1.121 and Definition 1.125, we have that the p.m.f. $f_X : \mathbb{R} \to \mathbb{R}$ is a function such that

$$f_X(x) = 0, \forall x \in S^c, \quad f_X(x) > 0, \forall x \in S, \quad \sum_{x \in S} f_X(x) = 1.$$

Remark 1.128. Let $\emptyset \neq S \subset \mathbb{R}$ be a finite or countably infinite set and let $f : \mathbb{R} \to \mathbb{R}$ be such that

$$f(x) = 0, \forall x \in S^c, \quad f(x) > 0, \forall x \in S, \quad \sum_{x \in S} f(x) = 1.$$

Then by an argument similar to Proposition 1.45, we conclude that $\mathbb{P}$ as defined below is a probability function/measure on $\mathbb{B}$, where $\mathbb{B}$ denotes the power set of $\mathbb{R}$. For all $A \subseteq \mathbb{R}$, consider

$$\mathbb{P}(A) := \sum_{x \in A \cap S} f(x).$$

By an argument similar to Theorem 1.115, we can then show that the function $F : \mathbb{R} \to \mathbb{R}$ defined by $F(x) := \mathbb{P}((-\infty, x]), \forall x \in \mathbb{R}$ is non-decreasing, right continuous with $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$. By Theorem 1.116, this F is the DF of some RV Y, i.e. $F_Y = F$ and by construction, Y must be discrete with support S and p.m.f. $f_Y = f$.

Example 1.129. Take S to be the set of natural numbers $\{1, 2, \cdots\}$ and consider the function $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) := \begin{cases} \frac{1}{2^x}, & \text{if } x \in S, \\ 0, & \text{if } x \in S^c. \end{cases}$$

Then f takes non-negative values with $\sum_{x \in S} f(x) = \sum_{n=1}^{\infty} \frac{1}{2^n} = 1$. Therefore f is the p.m.f. of some RV X with DF F_X given by

$$\begin{aligned} F_X(x) &= \mathbb{P}(X \leq x) = \sum_{t \in (-\infty, x] \cap S} f_X(t) \\ &= \begin{cases} 0, & \text{if } x < 1, \\ \sum_{n=1}^{m} \frac{1}{2^n}, & \text{if } x \in [m, m+1), m \in S. \end{cases} \quad = \begin{cases} 0, & \text{if } x < 1, \\ 1 - \frac{1}{2^m}, & \text{if } x \in [m, m+1), m \in S. \end{cases} \end{aligned}$$

Definition 1.130 (Continuous RV and its Probability Density Function (p.d.f.))**.** An RV X is said to be a continuous RV if there exists an integrable function $f : \mathbb{R} \to [0, \infty)$ such that

$$F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f(t)\, dt, \forall x \in \mathbb{R}.$$

The function f is called the probability density function (p.d.f.) of X.

Remark 1.131*.* Let X be a continuous RV with DF F_X and p.d.f. f_X. Then we have the following observations.

(a) Since f_X is integrable, from the relation $F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt, \forall x \in \mathbb{R}$, we have F_X is continuous on $\mathbb{R}$. In particular, F_X is absolutely continuous. Moreover, for all $a < b$, we have

$$F_X(b) - F_X(a) = \int_{-\infty}^{b} f_X(t)\, dt - \int_{-\infty}^{a} f_X(t)\, dt = \int_{a}^{b} f_X(t)\, dt.$$

(b) Since F_X is continuous, we have

 (i) $F_X(x-) = F_X(x) = F_X(x+), \forall x \in \mathbb{R}$.

 (ii) $\mathbb{P}(X = x) = \mathbb{P}_X(\{x\}) = F_X(x) - F_X(x-) = 0, \forall x \in \mathbb{R}$.

 (iii) $\mathbb{P}(X < x) = F_X(x-) = F_X(x) = \mathbb{P}(X \leq x), \forall x \in \mathbb{R}$.

 (iv) For all $a < b$,

$$\begin{aligned}\mathbb{P}(a < X < b) &= \mathbb{P}(a < X \leq b) = \mathbb{P}(a \leq X < b) = \mathbb{P}(a \leq X \leq b)\\ &= F_X(b) - F_X(a) = \int_{a}^{b} f_X(t)\, dt.\end{aligned}$$

(c) If $A \subset \mathbb{R}$ is finite or countably infinite, then by the finite/countable additivity of $\mathbb{P}_X$, we have

$$\mathbb{P}(X \in A) = \mathbb{P}_X(A) = \sum_{x \in A} \mathbb{P}_X(\{x\}) = 0.$$

(d) By definition, we have $f_X(x) \geq 0, \forall x \in \mathbb{R}$ and

$$1 = \lim_{x \to \infty} F_X(x) = \lim_{x \to \infty} \int_{-\infty}^{x} f_X(t)\, dt = \int_{-\infty}^{\infty} f_X(t)\, dt.$$

Remark 1.132*.* Let $f : \mathbb{R} \to [0, \infty)$ be an integrable function with $\int_{-\infty}^{\infty} f(t)\, dt = 1$. Then the function $F : \mathbb{R} \to [0, 1]$ defined by $F(x) := \int_{-\infty}^{x} f(t)\, dt, \forall x \in \mathbb{R}$ is non-decreasing and continuous with $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$. By Theorem 1.116, this F is the DF of some RV Y, i.e. $F_Y = F$ and by construction, Y must be continuous with p.d.f. $f_Y = f$.

Example 1.133. Let X be an RV with the DF $F_X : \mathbb{R} \to \mathbb{R}$ as discussed in Example 1.118. Here,

$$F_X(x) := \begin{cases} 0, \text{ if } x < 0, \\ x, \text{ if } 0 \leq x < 1, \\ 1, \text{ if } x \geq 1. \end{cases}$$

Then the function $f : \mathbb{R} \to [0, \infty)$ defined by

$$f(x) := \begin{cases} 1, \text{ if } 0 \leq x \leq 1, \\ 0, \text{ otherwise} \end{cases}$$

is an integrable function with $F_X(x) = \int_{-\infty}^{x} f(t)\, dt, \forall x \in \mathbb{R}$. Therefore, X is a continuous RV with p.d.f. f.

Example 1.134. Consider the DF $F : \mathbb{R} \to [0, 1]$ considered in Example 1.120 given by

$$F(x) = \begin{cases} 0, \text{ if } x < 0, \\ \frac{1}{4} + \frac{x}{2}, \text{ if } 0 \leq x \leq 1, \\ \frac{1}{2} + \frac{x}{4}, \text{ if } 1 < x < 2, \\ 1, \text{ if } x \geq 2. \end{cases}$$

As discussed earlier, F has a discontinuity at the point 0. Therefore, an RV X with DF F is not a continuous RV.

Note 1.135. Given a continuous RV X with p.d.f. f_X, the DF F_X is computed by the formula $F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt, \forall x \in \mathbb{R}$.

Example 1.136. Consider a function $f : \mathbb{R} \to \mathbb{R}$ of the form

$$f(x) = \begin{cases} \alpha x, & \text{if } x \in [-1, 0), \\ \frac{x^2}{8}, & \text{if } x \in [0, 2], \\ 0, & \text{otherwise} \end{cases}$$

for some $\alpha \in \mathbb{R}$. For this f to be a p.d.f. of a continuous RV, two conditions need to be satisfied, viz. $f(x) \geq 0, \forall x \in \mathbb{R}$ and $\int_{-\infty}^{\infty} f(x)\, dx = 1$.
The first condition is satisfied on $(-\infty, -1) \cup [0, \infty)$. For $x \in [-1, 0)$, we must have $\alpha x \geq 0$, which implies $\alpha \leq 0$.
From the second condition, we have $\int_{-1}^{0} \alpha x\, dx + \int_0^2 \frac{x^2}{8}\, dx = 1$. This yields $\alpha = -\frac{4}{3}$, which satisfies $\alpha \leq 0$.
Therefore, for f to be a p.d.f. we must have $\alpha = -\frac{4}{3}$.

In what follows, we consider the question of computing f_X from the DF F_X.

Remark 1.137 (Is the p.d.f. of a continuous RV unique?). Let X be a continuous RV with DF F_X and p.d.f. f_X. Fix any finite or countably infinite set $A \subset \mathbb{R}$ and fix $c \geq 0$. Consider the function $g : \mathbb{R} \to [0, \infty)$ defined by

$$g(x) := \begin{cases} f_X(x), & \text{if } x \in A^c, \\ c, & \text{if } x \in A. \end{cases}$$

Then g is integrable and $F_X(x) = \int_{-\infty}^{x} g(t)\, dt, \forall x \in \mathbb{R}$. Hence, g is also a p.d.f. for X. Therefore, the RV X with DF F_X is a continuous RV with p.d.f. f (or g). For example,

$$g(x) := \begin{cases} 1, & \text{if } 0 < x < 1, \\ 0, & \text{otherwise} \end{cases}$$

is a p.d.f. for X as in Example 1.133. More generally, we may also consider

$$g(x) := \begin{cases} f_X(x), & \text{if } x \in A^c, \\ c_x, & \text{if } x \in A \end{cases}$$

as a p.d.f., where $c_x \geq 0, \forall x \in A$.

Note 1.138. In fact, a p.d.f. f_X for a continuous RV X is determined uniquely on the complement of sets of 'length 0', such as sets which are finite or countably infinite. We do not make a precise statement – this is beyond the scope of this course. However, we consider the deduction of p.d.f.s from the DFs.

The next result is stated without proof.

Theorem 1.139. *Let X be an RV with DF F_X.*

(a) If F_X is differentiable on $\mathbb{R}$ with $\int_{-\infty}^{\infty} F_X'(t)\,dt = 1$, then X is a continuous RV with p.d.f. F_X'.

(b) If F_X is differentiable everywhere except on a finite or a countably infinite set $A \subset \mathbb{R}$ with $\int_{-\infty}^{\infty} F_X'(t)\,dt = 1$, then X is a continuous RV with p.d.f. f given by

$$f(x) := \begin{cases} F_X'(x), & \text{if } x \in A^c, \\ 0, & \text{if } x \in A. \end{cases}$$

Note 1.140. Continuing the discussion from Note 1.135, the DF F_X of a continuous RV X may be used to compute the p.d.f. f_X. In Example 1.133, the DF F_X is given by

$$F_X(x) := \begin{cases} 0, & \text{if } x < 0, \\ x, & \text{if } 0 \leq x < 1, \\ 1, & \text{if } x \geq 1. \end{cases}$$

It is differentiable everywhere except at the points 0 and 1. Using Theorem 1.139, we have the p.d.f. given by

$$f(x) := \begin{cases} 1, & \text{if } 0 < x < 1, \\ 0, & \text{otherwise.} \end{cases}$$

Note 1.141. To study a continuous RV X, we may study any one of the following three quantities, viz. the law/distribution $\mathbb{P}_X$, the DF F_X or the p.d.f. f_X. Given any one of these quantities, the other two can be obtained using the relations described above.

Definition 1.142 (Support of a Continuous RV)**.** Let X be a continuous RV with DF F_X. The set

$$S := \{x \in \mathbb{R} : F_X(x+h) - F_X(x-h) > 0, \forall h > 0\}$$

is defined to be the support of X.

Remark 1.143. The support S of a continuous RV X can be expressed in terms of the law/distribution of X as follows.

$$S = \{x \in \mathbb{R} : \mathbb{P}(x-h < X \leq x+h) > 0, \forall h > 0\} = \{x \in \mathbb{R} : \mathbb{P}_X((x-h, x+h]) > 0, \forall h > 0\}.$$

Remark 1.144. The support S of a continuous RV X can be expressed in terms of the p.d.f. f_X as follows.

$$S = \{x \in \mathbb{R} : \int_{x-h}^{x+h} f_X(t)\, dt > 0, \forall h > 0\}.$$

Note 1.145. If $x \notin S$, where S is the support of a continuous RV X, then there exists $h > 0$ such that $F_X(x+h) = F_X(x-h)$. By the non-decreasing property of F_X, we conclude that F_X remains a constant on the interval $[x-h, x+h]$. In particular, $f_X(t) = F_X'(t) = 0, \forall t \in (x-h, x+h)$.

Example 1.146. Consider a continuous RV X with DF $F_X : \mathbb{R} \to [0,1]$ and p.d.f. $f_X : \mathbb{R} \to [0,\infty)$ given by

$$F_X(x) := \begin{cases} 0, & \text{if } x < 0, \\ x, & \text{if } 0 \leq x < 1, \\ 1, & \text{if } x \geq 1. \end{cases}, \qquad f_X(x) := \begin{cases} 1, & \text{if } 0 < x < 1, \\ 0, & \text{otherwise.} \end{cases}$$

To identify the support S, we consider the following cases.

(a) Let $x \in (-\infty, 0)$. Then for all h with $-x > h > 0$, we have $x - h < x + h < 0$ and consequently, $F_X(x+h) - F_X(x-h) = 0 - 0 = 0$. Therefore $x \notin S$.

(b) Let $x \in (1, \infty)$. Then for all $0 < h < x - 1$, we have $1 < x - h < x + h$ and consequently, $F_X(x+h) - F_X(x-h) = 1 - 1 = 0$. Therefore $x \notin S$.

(c) Let $x \in (0,1)$. For any $0 < h < \min\{x, 1-x\}$, we have $0 < x - h < x + h < 1$ and consequently, $F_X(x+h) - F_X(x-h) = (x+h) - (x-h) = 2h > 0$. For $h \geq \min\{x, 1-x\}$, at least one of $x-h, x+h$ is in $(0,1)^c$ and hence $F_X(x+h) - F_X(x-h) > 0$. Therefore $x \in S$.

(d) Let $x = 0$. Then for any $h > 0$, we have $F_X(0+h) - F_X(0-h) = F_X(0+h) > 0$. Then $0 \in S$. By a similar argument, $1 \in S$.

From the above discussion, we conclude that $S = [0,1]$.

Remark 1.147 (Identifying discrete/continuous RVs from their DFs). Suppose that the distribution of an RV X is specified by a given DF F_X. In order to check if X is a discrete/continuous RV, we use the following steps.

(a) Identify the set $D = \{x \in \mathbb{R} : F_X(x-) < F_X(x+)\} = \{x \in \mathbb{R} : \mathbb{P}(X = x) > 0\}$ of discontinuities of F_X. Recall that D is a finite or a countably infinite set.

(b) If D is empty, then F_X is continuous on $\mathbb{R}$. By verifying the hypothesis of Theorem 1.139 or otherwise, check if there exists a p.d.f.. If a p.d.f. exists, then X is a continuous RV. Otherwise, X is not a continuous RV.

(c) If F_X has at least one discontinuity, then F_X is not continuous on $\mathbb{R}$ and hence X cannot be a continuous RV. For X to be a discrete RV X, we must have

$$\sum_{x \in D} [F_X(x+) - F_X(x-)] = \sum_{x \in D} \mathbb{P}(X = x) = 1.$$

If the above condition is satisfied, X is a discrete RV. Otherwise, X is not a discrete RV.

Note 1.148. Cantor function (also known as the Devil's Staircase) is an example of a continuous distribution function, which is not absolutely continuous. In this case, the DF F is not representable as $\int_{-\infty}^{x} f(t)\,dt$ for any non-negative integrable function. We do not discuss these types of examples in this course.

Note 1.149. Consider the DF $F : \mathbb{R} \to [0,1]$ considered in Example 1.120 given by

$$F(x) = \begin{cases} 0, \text{ if } x < 0, \\ \frac{1}{4} + \frac{x}{2}, \text{ if } 0 \leq x \leq 1, \\ \frac{1}{2} + \frac{x}{4}, \text{ if } 1 < x < 2, \\ 1, \text{ if } x \geq 2. \end{cases}$$

As discussed in Example 1.123 and Example 1.134, an RV with DF F is neither discrete nor continuous.

Definition 1.150 (Quantiles and Median for an RV)**.** Let X be an RV with DF F_X. For any $p \in (0,1)$, a number $x \in \mathbb{R}$ is called a quantile of order p if the following inequalities are satisfied, viz.

$$p \leq F_X(x) \leq p + \mathbb{P}(X = x).$$

A quantile of order $\frac{1}{2}$ is called a median.

Note 1.151. A quantile need not be unique. Refer to problem set 4 for explicit examples.

Notation 1.152. We write $\mathfrak{z}_p(X)$ to denote a quantile of order p.

Notation 1.153. The quantiles of order $\frac{1}{4}$ and $\frac{3}{4}$ for an RV X are referred to as the lower and upper quartiles of X, respectively.

Note 1.154. The inequalities mentioned in Definition 1.150 can be restated as

$$\mathbb{P}(X \leq x) \geq p, \quad \mathbb{P}(X \geq x) \geq 1 - p.$$

Note 1.155. Let X be a continuous RV with DF F_X. Then a quantile of order p is a solution to the equation $F_X(x) = p$, since $\mathbb{P}(X = x) = 0, \forall x \in \mathbb{R}$. Moreover, if F_X is strictly increasing, then $\mathfrak{z}_p(X)$ is unique for all $p \in (0,1)$.