2. Random Variables

We begin the discussion with a very simple example.
Example: Let the random experiment be the toss of a coin and let the sample space associated with the experiment be C = {H, T}, where H and T represent heads and tails, respectively.

Let X be a function such that X(T) = 0 and X(H) = 1.

Thus X is a real-valued function defined on the sample space C which takes us from the sample space C to a space of real numbers D = {0, 1}.

We now formulate the definition of a random variable and its space.

Definition 2.1. Consider a random experiment with a sample space C. A function X, which assigns to each element c ∈ C one and only one number X(c) = x, is called a **random variable**. The **space** or **range** of X is the set of real numbers D = {x : x = X(c), c ∈ C}.

Definition 2.2. We call random variables of the first type **discrete** random variables, while we call those of the second type **continuous** random variables.

Definition 2.2a Consider X is a discrete random variable with a finite space, D = {d1, . . . , dm}. Define the induced probability distribution of X , the function $p_X(di)$ on D by

$$p_X(di) = P[\{c : X(c) = di\}], \text{ for } i = 1, \ldots \qquad (2.1)$$

Definition 2.2b. Suppose, X is a continuous random variable, then D is an interval of real numbers. We determine a nonnegative function $f_X(x)$ such that for any interval of real numbers (a, b) ∈ D, the induced probability distribution of X, $p_X(\cdot)$, is defined as

$$p_X[(a, b)] = P[\{c \in C : a < X(c) < b\}] = \int_a^b fX(x)\,dx \qquad ....(2.2)$$

Definition 2.3 (Cumulative Distribution Function)**.** Let X be a random variable. Then its **cumulative distribution function** (cdf) is defined by $F_X(x)$, where

$$F_X(x) = P_X((-\infty, x]) = P(\{c \in C : X(c) \leq x\}). \qquad \ldots\ldots\ldots\ldots (2.3)$$

We can write, $P(\{c \in C : X(c) \leq x\})$ to $P(X \leq x)$.

Example 2.1. Suppose we roll a fair die with the numbers 1 through 6 on it.
Let X be the upface of the roll.
Then the space of X is {1, 2, . . . , 6} and its pmf is $p_X(i) = 1/6$, for i = 1, 2, . . . , 6.

If $x < 1$, then $F_X(x) = 0$.
If $1 \leq x < 2$, then $F_X(x) = 1/6$.

Example 2.2. Let X be the number chosen at random from the interval (0, 1). In this case the space of X is D = (0, 1). Because the number is chosen at random, it is reasonable to assign

$P_X[(a, b)] = b - a$, for $0 < a < b < 1$. (2.4)

It follows that the pdf of X is

$f_X(x) =$ 1 $0< x < 1$, and = 0 elsewhere.

Then, we have

$$P\left[\left\{X < \frac{1}{8}\right\} \cup \left\{X > \frac{7}{8}\right\}\right] = \int_0^{\frac{1}{8}} dx + \int_{\frac{7}{8}}^1 dx = \frac{1}{4}.$$

We now obtain the cdf of X. First, if

If $x \leq 0$ then $P(X \leq x) = \int_{-\infty}^{x} f_X(t)\,dt = \int_{-\infty}^{x} 0\,dt = 0$, if $0 < x < 1$ then $P(X \leq x) = \int_{-\infty}^{x} f_X(t)\,dt = \int_{-\infty}^{0} f_X(t)\,dt + \int_0^x f_X(t)\,dt = \int_{-\infty}^{0} 0\,dt + \int_0^x 1\,dt = x$, and

if $x \geq 1$ then $P(X \leq x) = \int_{-\infty}^{x} f_X(t)\,dt = \int_{-\infty}^{0} f_X(t)\,dt + \int_0^1 f_X(t)\,dt \int_1^x f_X(t)\,dt = \int_{-\infty}^{0} 0\,dt + \int_0^1 1\,dt + \int_1^x 0\,dt = 0 + 1 + 0 = 1$. So the cumulative distribution function is

$$F_X(x) = P_X((-\infty, x]) = \int_{-\infty}^{x} f_X(t)\,dt = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } 0 \leq x < 1 \\ 1 & \text{if } x \geq 1. \end{cases}$$

Note that we have

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt, \quad \text{for all } x \in R,$$

and $\frac{d}{dx}F_X(x) = f_X(x)$, for all $x \in R$, except for $x = 0$ and $x = 1$.

Theorem 2.1. *Let X be a random variable with cumulative distribution function $F(x)$. Then*

(a) For all a and b, if $a < b$, then $F(a) \leq F(b)$ (F is nondecreasing).
(b) $\lim_{x \to -\infty} F(x) = 0$ (the lower limit of F is 0).
(c) $\lim_{x \to \infty} F(x) = 1$ (the upper limit of F is 1).
(d) $\lim_{x \downarrow x_0} F(x) = F(x_0)$ (F is right continuous).

Proof: We prove parts (a) and (d) and leave parts (b) and (c) for Exercise 1.5.10.
Part (a): Because $a < b$, we have $\{X \leq a\} \subset \{X \leq b\}$. The result then follows from the monotonicity of P; see Theorem 1.3.3.
Part (d): Let $\{x_n\}$ be any sequence of real numbers such that $x_n \downarrow x_0$. Let $C_n = \{X \leq x_n\}$. Then the sequence of sets $\{C_n\}$ is decreasing and $\cap_{n=1}^{\infty} C_n = \{X \leq x_0\}$. Hence, by Theorem 1.3.6,

$$\lim_{n \to \infty} F(x_n) = P\left(\bigcap_{n=1}^{\infty} C_n\right) = F(x_0),$$

which is the desired result. ■

Theorem 2.2. *Let X be a random variable with the cdf F_X. Then for $a < b$,*

$$P[a < X \leq b] = F_X(b) - F_X(a).$$

Proof: Note that

$$\{-\infty < X \leq b\} = \{-\infty < X \leq a\} \cup \{a < X \leq b\}.$$

The proof of the result follows immediately because the union on the right side of this equation is a disjoint union. ■

Theorem 2.3.

For any random variable,

$$P[X = x] = F_X(x) - F_X(x-),$$

for all $x \in R$, *where* $F_X(x-) = \lim_{z \uparrow x} F_X(z)$.

Proof: For any $x \in R$, we have

$$\{x\} = \bigcap_{n=1}^{\infty} \left(x - \frac{1}{n}, x \right] ;$$

that is, $\{x\}$ is the limit of a decreasing sequence of sets. Hence,

$$\begin{aligned} P[X = x] &= P\left[\bigcap_{n=1}^{\infty} \left\{ x - \frac{1}{n} < X \leq x \right\} \right] \\ &= \lim_{n \to \infty} P\left[x - \frac{1}{n} < X \leq x \right] \\ &= \lim_{n \to \infty} [F_X(x) - F_X(x - (1/n))] \\ &= F_X(x) - F_X(x-), \end{aligned}$$

Example 2.3.

Let X have the discontinuous cdf

$$F_X(x) = \begin{cases} 0 & x < 0 \\ x/2 & 0 \leq x < 1 \\ 1 & 1 \leq x. \end{cases}$$

Then

$$P(-1 < X \leq 1/2) = F_X(1/2) - F_X(-1) = \frac{1}{4} - 0 = \frac{1}{4}$$

and

$$P(X = 1) = F_X(1) - F_X(1-) = 1 - \frac{1}{2} = \frac{1}{2}.$$

The value $1/2$ equals the value of the step of F_X at $x = 1$. ■

Example 2.4.

Suppose X has the pmf

$$p_X(x) = \begin{cases} cx & x = 1, 2, \ldots, 10 \\ 0 & \text{elsewhere,} \end{cases}$$

for an appropriate constant c. Then

$$1 = \sum_{x=1}^{10} p_X(x) = \sum_{x=1}^{10} cx = c(1 + 2 + \cdots + 10) = 55c,$$

and, hence, $c = 1/55$. ■

Example 2.5.

Suppose X has the pdf

$$f_X(x) = \begin{cases} cx^3 & 0 < x < 2 \\ 0 & \text{elsewhere,} \end{cases}$$

for a constant c. Then

$$1 = \int_0^2 cx^3\,dx = c\left[\frac{x^4}{4}\right]_0^2 = 4c,$$

and, hence, $c = 1/4$. For illustration of the computation of a probability involving X, we have

$$P\left(\frac{1}{4} < X < 1\right) = \int_{1/4}^{1} \frac{x^3}{4}\,dx = \frac{255}{4096} = 0.06226. \quad \blacksquare$$

Definition 2.4.

A random variable is a *discrete random variable* if the space (its range) is either finite or countable.

Example 2.6. A lot, consisting of 100 fuses, is inspected by the following procedure. Five of these fuses are chosen at random and tested; if all five "blow" at the correct amperage, the lot is accepted. If, in fact, there are 20 defective fuses in the lot, the probability of accepting the lot is, under appropriate assumptions,

$$\frac{\binom{80}{5}}{\binom{100}{5}} = 0.31931.$$

More generally, let the random variable X be the number of defective fuses among the five that are inspected. The pmf of X is given by

$$p_X(x) = \begin{cases} \frac{\binom{20}{x}\binom{80}{5-x}}{\binom{100}{5}} & \text{for } x = 0, 1, 2, 3, 4, 5 \\ 0 & \text{elsewhere.} \end{cases}$$

Clearly, the space of X is $\mathcal{D} = \{0, 1, 2, 3, 4, 5\}$, which is also its support. This is an example of a random variable of the discrete type whose distribution is an illustration of a **hypergeometric distribution**

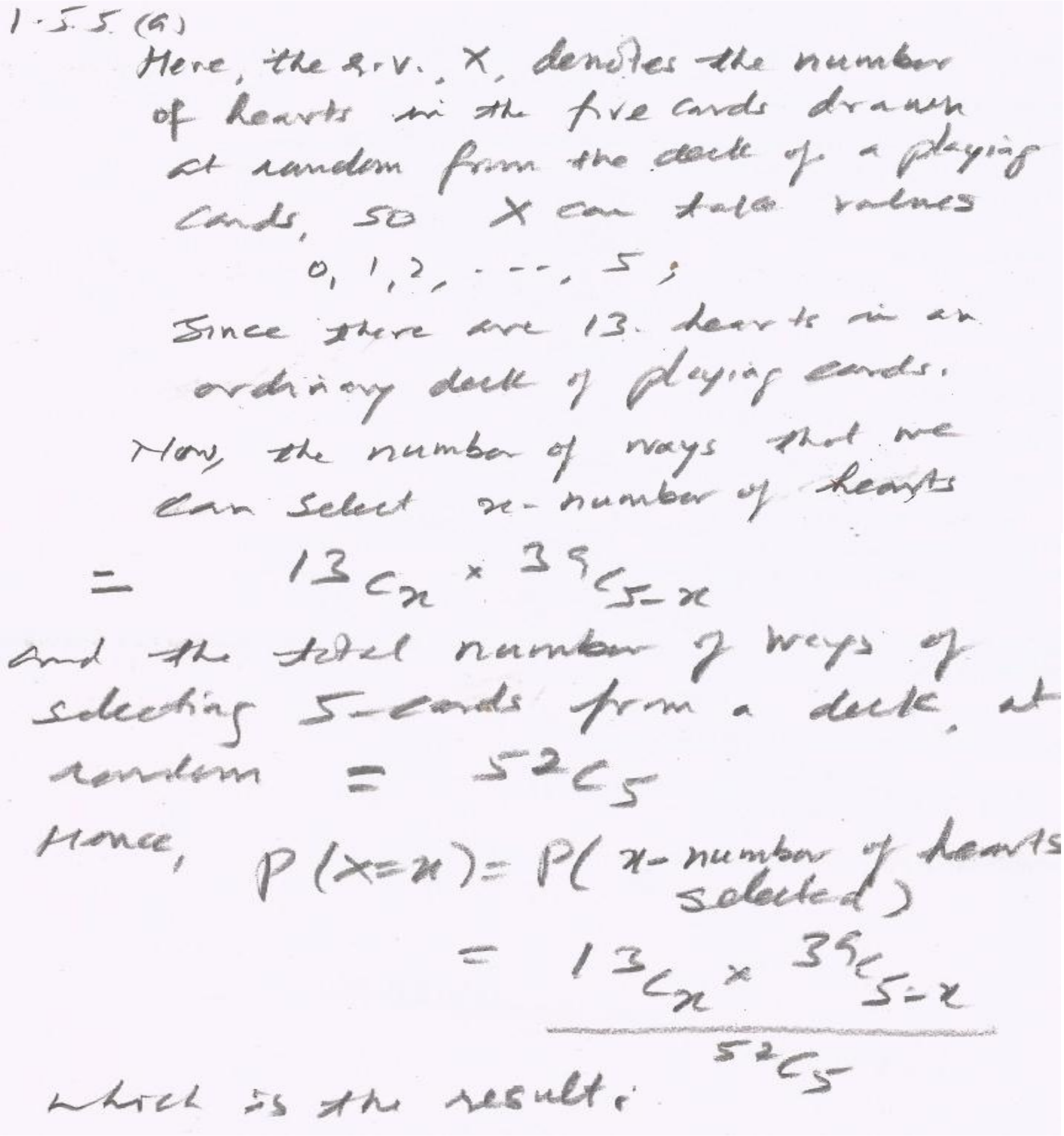
Exercise

1.5.5. Let us select five cards at random and without replacement from an ordinary deck of playing cards.

(a) Find the pmf of X, the number of hearts in the five cards.

(b) Determine $P(X \leq 1)$.

Solution



1.5.5 (a)

Here, the r.v., X, denotes the number of hearts in the five cards drawn at random from the deck of a playing cards, so X can take values

$0, 1, 2, \cdots, 5;$

Since there are 13 hearts in an ordinary deck of playing cards.

Now, the number of ways that we can select x-number of hearts

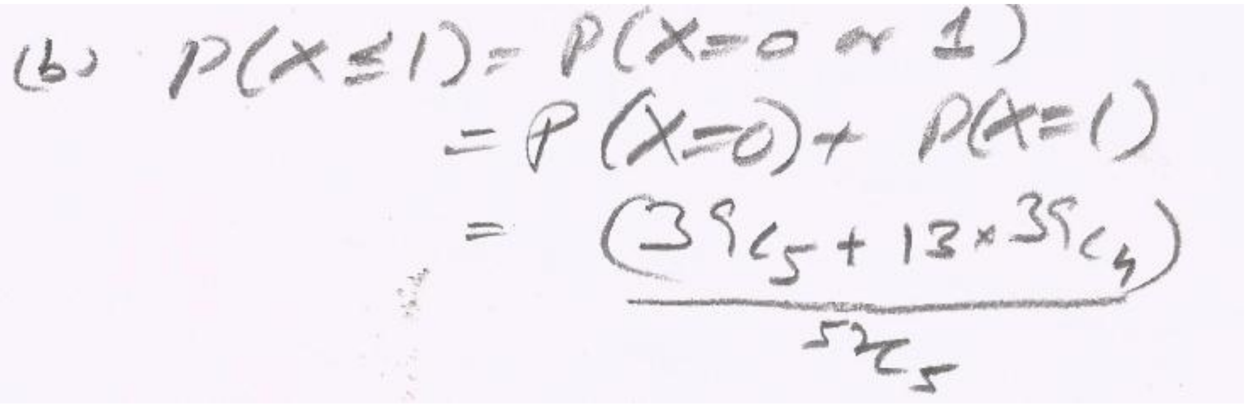
$$= {}^{13}C_x \times {}^{39}C_{5-x}$$

And the total number of ways of selecting 5-cards from a deck at random $= {}^{52}C_5$

Hence, $P(X=x) = P($x-number of hearts selected$)$

$$= \frac{{}^{13}C_x \times {}^{39}C_{5-x}}{{}^{52}C_5}$$

which is the result.

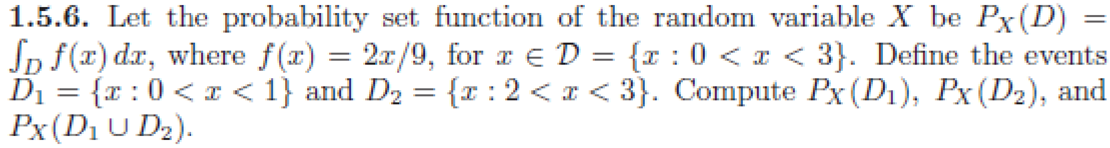


(b) $P(X \le 1) = P(X = 0 \text{ or } 1)$

$= P(X=0) + P(X=1)$

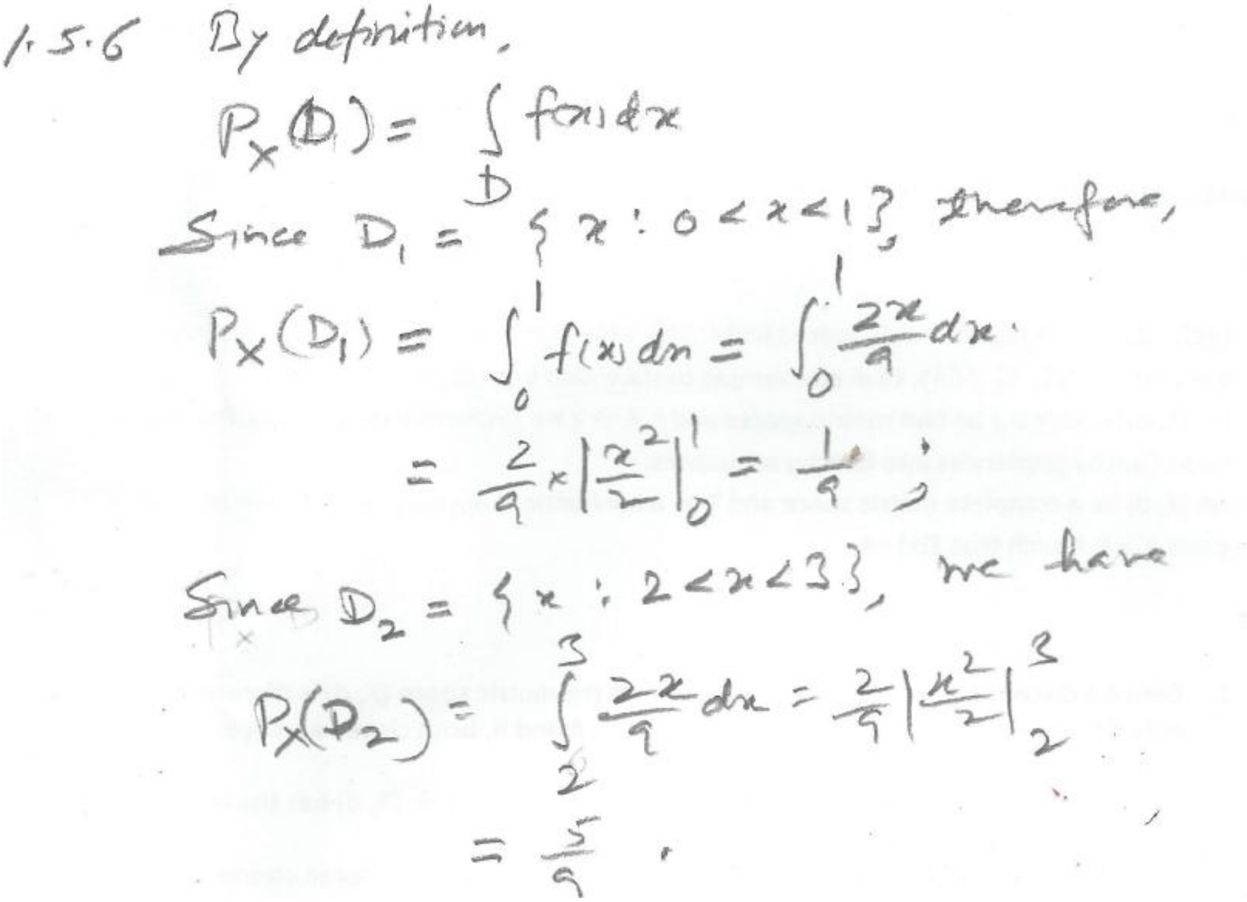
$$= \frac{\left({}^{39}C_5 + 13 \times {}^{39}C_4\right)}{{}^{52}C_5}$$

Exercise



1.5.6. Let the probability set function of the random variable X be $P_X(D) = \int_D f(x)\,dx$, where $f(x) = 2x/9$, for $x \in \mathcal{D} = \{x : 0 < x < 3\}$. Define the events $D_1 = \{x : 0 < x < 1\}$ and $D_2 = \{x : 2 < x < 3\}$. Compute $P_X(D_1)$, $P_X(D_2)$, and $P_X(D_1 \cup D_2)$.

Solution:



1.5.6 By definition,

$$P_X(D) = \int_D f(x)\,dx$$

Since $D_1 = \{x : 0 < x < 1\}$, therefore,

$$P_X(D_1) = \int_0^1 f(x)\,dx = \int_0^1 \frac{2x}{9}\,dx$$

$$= \frac{2}{9} \times \left|\frac{x^2}{2}\right|_0^1 = \frac{1}{9},$$

Since $D_2 = \{x : 2 < x < 3\}$, we have

$$P_X(D_2) = \int_2^3 \frac{2x}{9}\,dx = \frac{2}{9}\left|\frac{x^2}{2}\right|_2^3$$

$$= \frac{5}{9}.$$

Now $D_1 \cup D_2 = \{x : 0 < x < 1\} \cup \{x : 2 < x < 3\}$

So, $P_X(D_1 \cup D_2) = P_X(D_1) + P_X(D_2)$

$$= \frac{1}{9} + \frac{5}{9} = \frac{2}{3}.$$

Exercise

1.5.8. Suppose the random variable X has the cdf

$$F(x) = \begin{cases} 0 & x < -1 \\ \frac{x+2}{4} & -1 \leq x < 1 \\ 1 & 1 \leq x. \end{cases}$$

Write an R function to sketch the graph of $F(x)$. Use your graph to obtain the probabilities: (a) $P(-\frac{1}{2} < X \leq \frac{1}{2})$; (b) $P(X = 0)$; (c) $P(X = 1)$; (d) $P(2 < X \leq 3)$.

Solution:

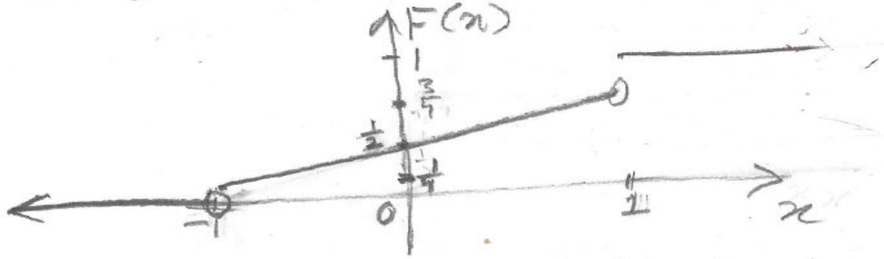
1.5.8. Given $F(x) = \begin{cases} 0, & x < -1 \\ \frac{x+2}{4}, & -1 \leq x < 1 \\ 1 & 1 \leq x. \end{cases}$

The pdf, $f_X(x)$, of X, is given by

$$f_X(x) = \frac{d}{dx}(F_X(x)), \text{ for } x \in R - \{-1, 1\}^*$$

$$= \begin{cases} \frac{1}{4}, & -1 < x < 1 \\ 0, & x > 1 \end{cases}$$

The graph of cdf, $F(x)$, is



(* Clearly, there are two points of discontinuities, at $x = -1$ and $x = 1$, respectively)

(a) We know that

$$P(a < X \leq b) = P(a \leq X \leq b)$$

$$= F_X(b) - F_X(a).$$

So, $P(-\frac{1}{2} < X \leq \frac{1}{2}) = F_X(\frac{1}{2}) - F_X(-\frac{1}{2})$

$$= \frac{5}{8} - \frac{3}{8} = \frac{1}{4}.$$

(b) Since, $P(X=a) = F_X(a) - F_X(a-)$, so,

$$P(X=0) = F_X(0) - \lim_{b \to 0\uparrow} F_X(b)$$

$$= \frac{1}{2} - \lim_{b \to 0\uparrow} \frac{1}{2} = \frac{1}{2} - \frac{1}{2} = 0$$

(c) $P(X=1) = F_X(1) - F_X(1-)$

$$= 1 - \lim_{x \to 1-} \left(\frac{x+2}{4}\right)$$

$$= 1 - \frac{3}{4} = \frac{1}{4}.$$

(d) $P(2 < X \leq 3)$

$= F_X(3) - F_X(2)$

$= 1 - 1$, since $F(x) = 1$, if $x \geq 1$.

$= 0$.

Definition 2.4 (Discrete Random Variable). *We say a random variable is a* **discrete random variable** *if its space is either finite or countable.*

Definition 2.5 (Continuous Random Variables). *We say a random variable is a* **continuous random variable** *if its cumulative distribution function $F_X(x)$ is a continuous function for all $x \in R$.*

Definition 2.6 (Probability Mass Function (pmf)). *Let X be a discrete random variable with space $\mathcal{D}$. The* **probability mass function** *(pmf) of X is given by*

$$p_X(x) = P[X = x], \quad \text{for } x \in \mathcal{D}.$$

Note that pmfs satisfy the following two properties:

$$\text{(i) } 0 \leq p_X(x) \leq 1 \text{ , } x \in \mathcal{D}, \text{ and (ii) } \textstyle\sum_{x \in \mathcal{D}} p_X(x) = 1.$$

Definition 2.7

The **support** of a continuous random variable X consists of all points x such that $f_X(x) > 0$. As in the discrete case, we often denote the support of X by $\mathcal{S}$.

Remarks:

If X is a continuous random variable, then probabilities can be obtained by ntegration; i.e.,

$$P(a < X \leq b) = F_X(b) - F_X(a) = \int_a^b f_X(t)\,dt.$$

Also, for continuous random variables,

$$P(a < X \leq b) = P(a \leq X \leq b) = P(a \leq X < b) = P(a < X < b).$$

And, note that pdfs satisfy the two properties

(i) $f_X(x) \geq 0$ and (ii) $\int_{-\infty}^{\infty} f_X(t)\,dt = 1$.

The second property, of course, follows from $F_X(\infty) = 1$.

Example 2.6 Let the random variable be the time in seconds between incoming telephone calls at a busy switchboard. Suppose that a reasonable probability model for X is given by the pdf

$$f_X(x) = \begin{cases} \frac{1}{4}e^{-x/4} & 0 < x < \infty \\ 0 & \text{elsewhere.} \end{cases}$$

Note that f_X satisfies the two properties of a pdf, namely, (i) $f(x) \geq 0$ and (ii)

$$\int_0^\infty \frac{1}{4}e^{-x/4}\,dx = -e^{-x/4}\Big|_0^\infty = 1.$$

For illustration, the probability that the time between successive phone calls exceeds 4 seconds is given by

$$P(X > 4) = \int_4^\infty \frac{1}{4}e^{-x/4}\,dx = e^{-1} = 0.3679.$$