Chapter 5 Random Variables. Probability Distributions

- 5.1 Random Variables
- 5.2 Discrete Random Variables and Distributions
- 5.3 Continuous Random Variables and Distributions
- 5.4 Summary

26 February 2018

5.1 Random Variables

A function X whose domain is sample space S and whose range $\mathrm{R} \subseteq \mathbb{R}$ is called a random variable.

That is the variable X takes numerical values for every event in S.

If the sample space $S = \{\omega_1, \omega_2, \cdots, \omega_k, \cdots \omega_n\}$ as domain, then $X(\omega_k)$ is the real number that the function X assigns to the element ω_k.

5.1 Random Variables

Example 1

Let $S = \{1, 2, 3, 4, 5, 6\}$ and define X as follows:

$X(1 \cup 2 \cup 3) = 1, \qquad X(4 \cup 5 \cup 6) = -1.$

X is a random variable defined as a function of the elements of S.

The domain of X is S and its range is the set $\{1, -1\}$.

This is useful to model real variables.
For example, the gain of a player
winning $\$1$ (+1) if the outcome is $1, 2$ or 3
losing $\$1$ (-1) if the outcome is $4, 5$ or 6. ■

5.1 Random Variables

Example 2

Two dice are rolled with the sample space

$S = \{(1,1), (1,2), \cdots, (5,6), (6,6)\}$ containing 36 elements.

Let X denote the random variable whose value for any $\omega \in S$ is the sum of numbers on the two dice.

Find the set $\{\omega : \omega \in S \text{ and } X(\omega) = 5\}$.

5.1 Random Variables

Solution

The range of X is $\{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$.

Each $\omega \in S$ has associated with it exactly one element of the range.

The range of X is made up of the sets

$$\{\omega : \omega \in S \text{ such that } X(\omega) = k, \qquad k = 2, \ldots, 12\}$$

The set $\{\mathrm{X} = 5\} = \{(1, 4), (2, 3), (3, 2), (4, 1)\}$. ■

That is we have 36 possible pairs of results but X can only have 11 different values

5.2 Discrete Random Variables and Distributions

We assign a probability measure on the elements of the sample space S, $P(\omega \in S) \to [0, 1]$.
that is a function from the sample space to the real numbers between 0 and 1.

Note: probability measures must be defined carefully, but this is beyond our needs.

Then since X is a random variable defined as a function of S, we can define the probability $\mathrm{P}(X = k) = P(\omega \in S{:}\, X(\omega) = k)$

5.2 Discrete Random Variables and Distributions

Example 3

A fair coin is tossed two times. Define the variable

$$X = number\ of\ heads\ obtained$$

Find the sample space (domain) and range of X. Then find

$P(X = 1)$.

5.2 Discrete Random Variables and Distributions

Solution

The sample space is $S = \{HH, HT, TH, TT\}$.

The probability measure on S is $P(\omega \in S) = \frac{1}{4}$.

The range of X, *number of Heads*, is $\{0, 1, 2\}$ with

$\mathrm{X}\{\omega = TT\} = 0, \mathrm{X}\{\omega = HT \cup \omega = TH\} = 1$ and $X\{\omega = HH\} = 2$

The required probability is $P(X = 1) = P(\{HT \cup TH\}) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$ ■

5.2 Discrete Random Variables and Distributions

By definition, a random variable X is *discrete* if X assumes only finitely many or countably many values $x_1, x_2, x_3, \cdots$ called the *possible values* of X, with positive probabilities

$$p_1 = P(X = x_1), \quad p_2 = P(X = x_2), \qquad p_3 = P(X = x_3), \cdots$$

whereas the probability $P(X \in I)$ is zero for any interval I containing no possible value.

5.2 Discrete Random Variables and Distributions

The ***probability mass function*** (pmf) of X is, for $j = 1,2,\cdots,$

$$f(x) = \begin{cases} p_j; & x = x_j \\ 0 & \text{otherwise} \end{cases}$$

the cumulative *distribution function* (cdf)

$$F(x) = P(X \leq x) = \sum_{x_j \leq x} p_j$$

where for any given x we sum all the probabilities p_j for which x_j is smaller than or equal to that of x.

Note: we often refer to the *cdf* simply as *the distribution*

5.2 Discrete Random Variables and Distributions

Example 4 (Continued from Example 1)

S= {1, 2, 3, 4, 5, 6} , $X = 1\ if\ s_i = \{1, 2, 3\}, X = -1\ if\ s_i = \{4, 5, 6, \}$

Obtain the

i. probability mass function $f(x)$, and

ii. cumulative distribution function $F(x)$. Sketch $F(x)$.

Solution

$i.\quad pmf$

$$f(1) = P(X = 1) = P(\{1, 2, 3\}) = \frac{1}{6}+\frac{1}{6}+\frac{1}{6} = \frac{1}{2}$$

$$f(-1) = P(X = -1) = P(\{4, 5, 6\}) = \frac{1}{6}+\frac{1}{6}+\frac{1}{6} = \frac{1}{2}$$

since each outcome has equal probability of occurring.

5.2 Discrete Random Variables and Distributions

Probability mass function: $f(x) = \begin{cases} \frac{1}{2}; & x = -1 \\ \frac{1}{2}; & x = 1 \\ 0 & otherwise \end{cases}$

ii. Cumulative distribution function:

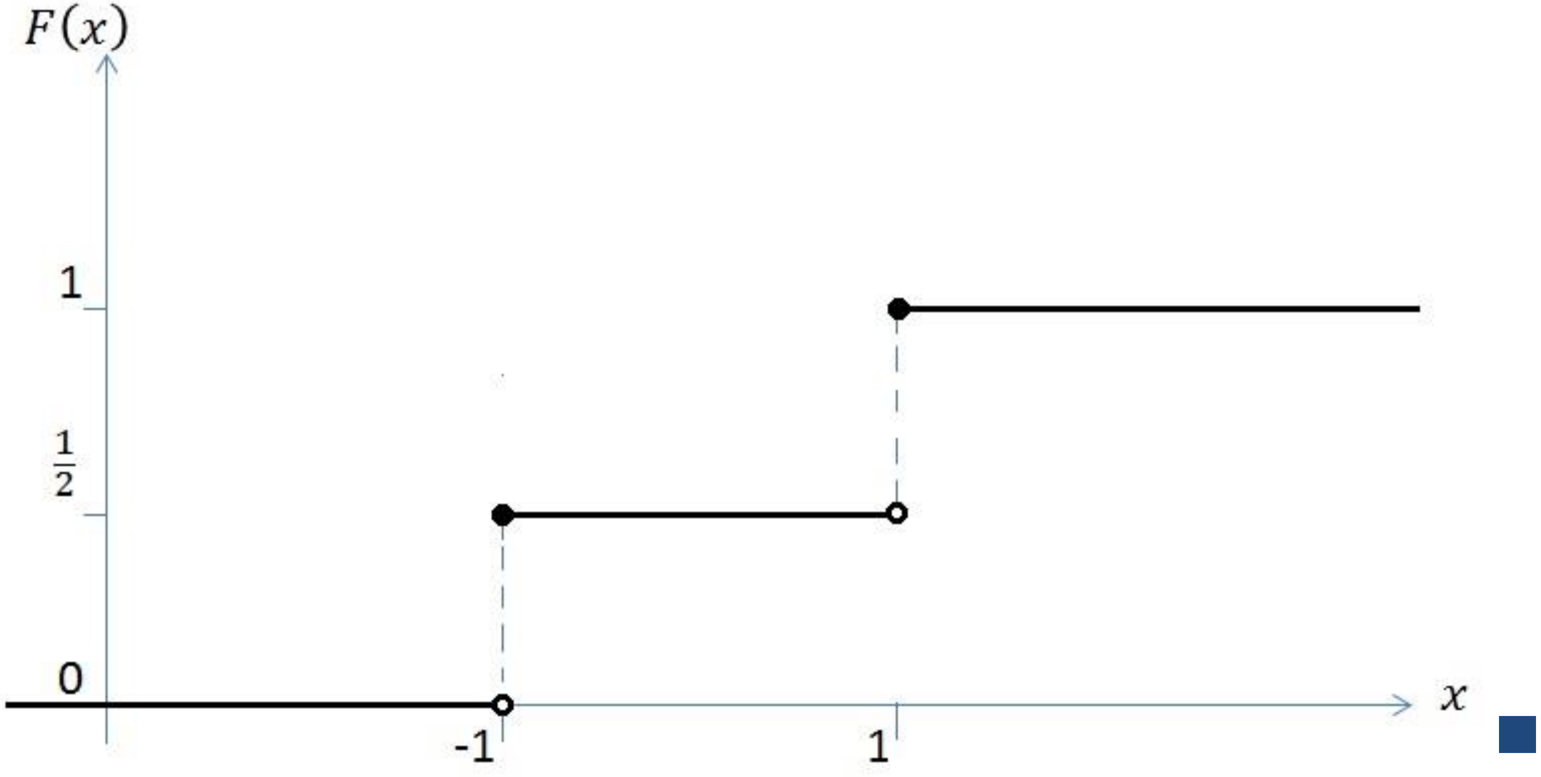
$$F(x) = P(X \leq x) = \begin{cases} 0; & x < -1 \\ \frac{1}{2}; & x = -1 \\ 1; & x = 1 \\ 1; & x > 1 \end{cases}$$

$F(-1) = P(X \leq -1) = P(X = -1) = \frac{1}{2}$,

$F(1) = P(X \leq 1) = P(X = -1 \cup X = 1) = \frac{1}{2} + \frac{1}{2} = 1$

5.2 Discrete Random Variables and Distributions

The graph of the cumulative distribution function is:



5.2 Discrete Random Variables and Distributions

Properties of the Probability Mass Function (pmf)
1. $f(x) \geq 0$
2. $\sum f(x) = 1$
Properties of the (Discrete) Cumulative Distribution Function
1. $\lim_{x \to -\infty} F(x) = 0$
2. $\lim_{x \to \infty} F(x) = 1$
3. $F(x) = P(X \leq x) = \sum_{x_j \leq x} p_j$
4. F is non-decreasing function
5. $0 \leq F(x) \leq 1$

5.2 Discrete Random Variables and Distributions

Example 5 (Continued from Example 2)

$S = \{(1, 1), (1, 2), \cdots, (5, 6), (6, 6)\}$, X sum of result

Obtain the

i. probability mass function $f(x)$, and

ii. distribution function $F(x)$.

iii. Sketch f(x) and $F(x)$.

5.2 Discrete Random Variables and Distributions

Solution

i. We obtain the following

$f(2) = P(X = 2) = P(\{(1,1)\}) = \frac{1}{36}$

$f(3) = P(X = 3) = P(\{(1,2),(2,1)\}) = \frac{1}{36} + \frac{1}{36} = \frac{2}{36}$

$f(4) = P(X = 4) = P(\{(1,3),(2,2),(3,1)\}) = \frac{1}{36} + \frac{1}{36} + \frac{1}{36} = \frac{3}{36}$

$\vdots \qquad\qquad\qquad \vdots$

$f(12) = P(X = 12) = P(\{(6,6)\}) = \frac{1}{36}$

5.2 Discrete Random Variables and Distributions

i. The probability mass function can be represented with a table:

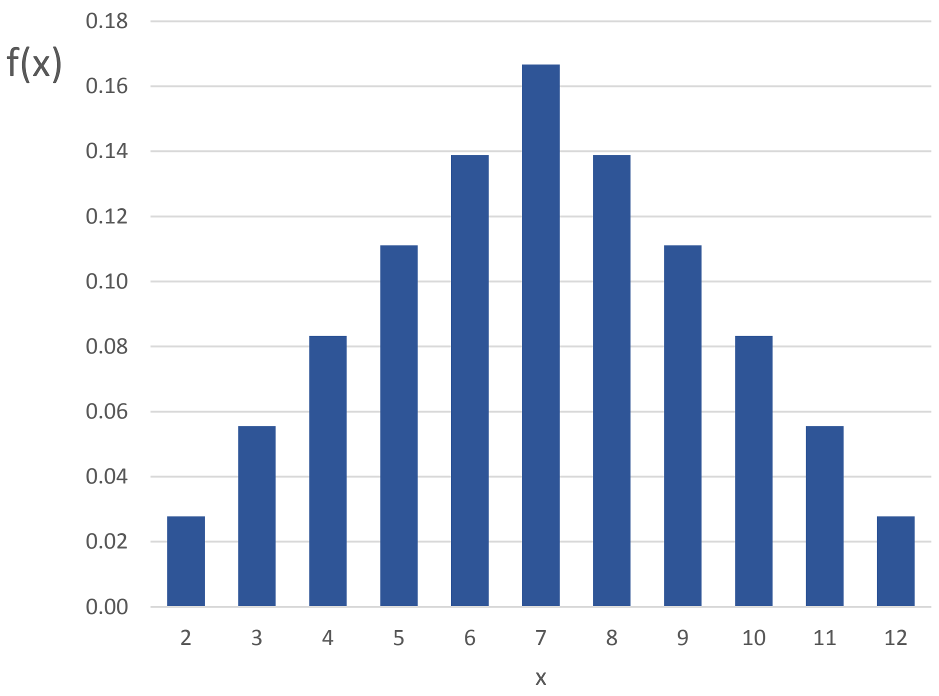
x	2	3	4	5	6	7	8	9	10	11	12
$f(x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

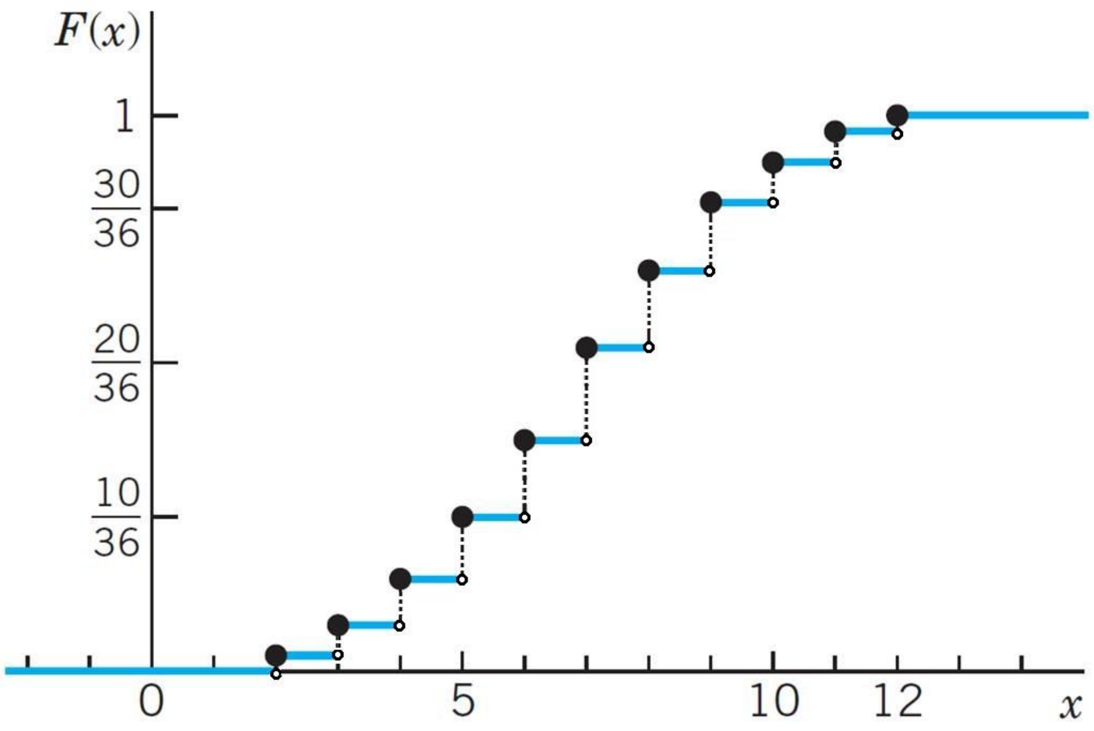
ii. The distribution function can be represented with a table:

x	2	3	4	5	6	7	8	9	10	11	12
$F(x)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{15}{36}$	$\frac{21}{36}$	$\frac{26}{36}$	$\frac{30}{36}$	$\frac{33}{36}$	$\frac{35}{36}$	$\frac{36}{36}$

5.2 Discrete Random Variables and Distributions

iii. The sketch of the probability function and distribution function are:





Notice that the distribution function is *nondecreasing*. ■

5.2 Discrete Random Variables and Distributions

One useful formula for discrete distribution is

$$P(a < X \leq b) = F(b) - F(a) = \sum_{a<x_j\leq b} p_j$$

This is the sum of all probabilities p_j for which x_j satisfies the condition $a < x_j \leq b$.

5.2 Discrete Random Variables and Distributions

Example 6 (Continued from Example 5)

$S = \{(1,1),(1,2),\cdots,(5,6),(6,6)\}$, X sum of result

Using the distribution function only, compute the probability of a sum of at least 4 and at most 8.

Solution

$$P(4 \le X \le 8) = P(3 < X \le 8) = F(8) - F(3)$$

$= \frac{26}{36} - \frac{3}{36} = \frac{23}{36}$ ■

5.2 Discrete Random Variables: problem

Mr Ali hits a target with probability $\mathrm{p_A} = \frac{1}{2}$,

Ms Beatrice hits with probability $\mathrm{p_B} = \frac{1}{3}$.

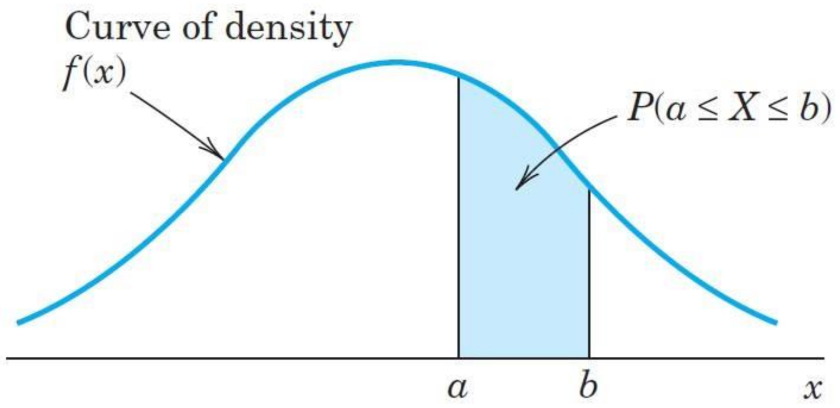
If both shots once, find the range, pmf and cdf for the variable $X = \{number\ of\ hits\}$

5.3 Continuous Random Variables and Distributions

We shall now consider continuous random variables which may take any value on $\mathbb{R}$.

Instead of the pmf, we now have a ***probability density function*** (pdf). This is a continuous function f such that $P(a < X \leq b)$ is equal to the area under the graph of f between $x = a$ and $x = b$.

The probability associated with any particular value $X = a$ is zero. So we need to find the probability of an interval [*X*, *X* + *dX*]



5.3 Continuous Random Variables and Distributions

Note

For a continuous variable a particular value has probability zero: that is $P(X = a) = P(X = b) = 0$. Therefore the following are equivalent

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

However, for consistency with discrete variables we usually write

$$P(x < X \leq y) \blacksquare$$

5.3 Continuous Random Variables and Distributions

The cumulative *distribution function* for a continuous random variable is given by

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(v)\, dv.$$

Furthermore, the probability density function and the distribution function are related by

$f(x) = F'(x).$

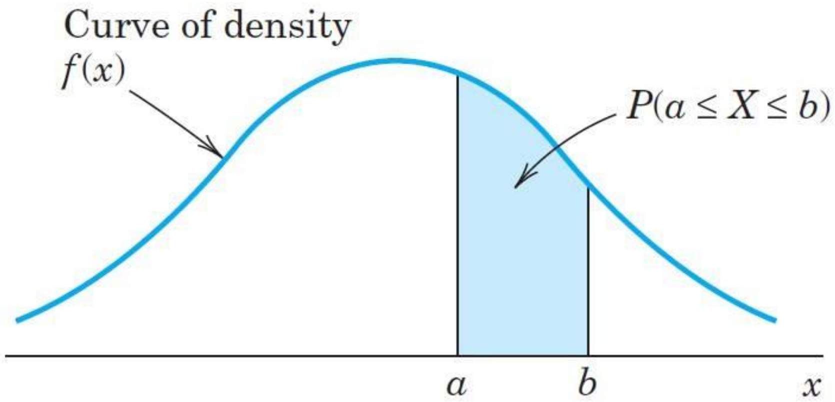
5.3 Continuous Random Variables and Distributions

The probability that a variable X takes values in an interval $[a, b]$ is the area under the *pdf* in that interval. That is,

$$P(a < X \leq b) = \int_a^b f(x)dx = F(b) - F(a)$$

You can look at this as

$$\int_a^b f(x)dx = \int_{-\infty}^b f(x)dx - \int_{-\infty}^a f(x)dx$$



5.3 Continuous Random Variables and Distributions: example

Consider a variable X that takes values in $[0,1]$ with constant density function: $f(x) = k\ if\ 0 < x \leq 1$, $\quad f(x) =, otherwise.$

This is called Uniform variable.

The cdf is $P(X \leq x) = \int_0^x k\ du = x[u]_0^x = kx$ for $0 < x \leq 1$

Since $\mathrm{P(X \leq 1)} = F(1) = k = 1$, we have $k = 1$ and

$$f(x) = \begin{cases} 1;\ 0 < x \leq 1 \\ 0; otherwise \end{cases} \text{ and } F(x) = \begin{cases} 0; x \leq 0 \\ x;\ 0 < x \leq 1 \\ 1; x > 1 \end{cases}$$

5.3 Continuous Random Variables and Distributions: problem

Plot the pdf

$$f(x) = \begin{cases} 1 \; if \; 0 < x \leq 1 \\ 0 \quad otherwise \end{cases}$$

Prove that $P(X \leq x) = x$ by geometric arguments.

5.3 Continuous Random Variables and Distributions

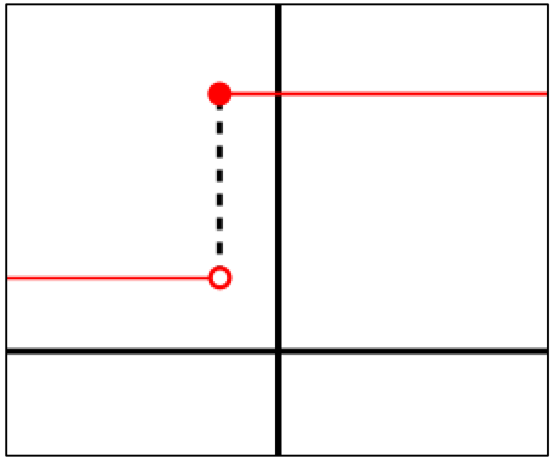
Properties of the Probability Density Function (pdf)
1. $f(x) \geq 0$
2. $\int_{-\infty}^{\infty} f(x)\, dx = 1$
Properties of the (Continuous) Distribution Function
1. $\lim_{x \to -\infty} F(x) = 0$
2. $\lim_{x \to \infty} F(x) = 1$
3. $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(v) dv$
4. F is differentiable [under special conditions], non-decreasing function
5. $0 \leq F(x) \leq 1$
6. $f(x) = F'(x)$

5.3 Continuous Random Variables and Distributions

The cdf is "right continuous". That is, it is allowed to have jumps but only on the left.

That is, for $\delta > 0$ and arbitrary $\epsilon > 0$

$$\lim_{\delta \to 0} |F(X + \delta) - F(x)| < \epsilon$$



this is not required for $\delta < 0$.

If $f(x)$ is continuous then also $F(x)$ is continuous.

In probability we use Labesgue integrals and not Riemann ones.

5.3 Continuous Random Variables and Distributions

Example 7

Let X be a random variable with probability density function $f(x) = 0.75(1 - x^2)$ for $-1 \leq x \leq 1$ and zero otherwise.

i. find the cumulative distribution function.

ii. Find the probabilities $P\left(-\frac{1}{2} < X \leq \frac{1}{2}\right)$ and $P\left(\frac{1}{4} < X \leq 2\right)$.

5.3 Continuous Random Variables and Distributions

Solution

Integrating,

$$0.75\int_{-1}^{x}(1-v^2)dv = 0.5 + 0.75x - 0.25x^3,$$

therefore

$$F(x) = \begin{cases} 0 \;\; ; \;\; x \le -1 \\ 0.5 + 0.75x - 0.25x^3 \;\; ; \;\; -1 < x \le 1 \\ 1 \;\; ; \;\; x > 1 \end{cases}$$

We use the distribution function to obtain the probabilities:

$$P\left(-\tfrac{1}{2} \le X \le \tfrac{1}{2}\right) = P\left(-\tfrac{1}{2} < X \le \tfrac{1}{2}\right) = F\left(\tfrac{1}{2}\right) - F\left(-\tfrac{1}{2}\right) = 0.6875$$

$$P\left(\tfrac{1}{4} \le X \le 2\right) = P\left(\tfrac{1}{4} < X \le 2\right) = F(2) - F\left(\tfrac{1}{4}\right) = 0.3164 \quad \blacksquare$$

5.3 Continuous Random Variables and Distributions

Example 8

Given that the probability density function for a random variable X is

$$f(x) = \begin{cases} 4x & ; \quad 0 \le x \le \frac{1}{2} \\ -4x + 4 & ; \quad \frac{1}{2} < x \le 1 \end{cases}$$

Obtain the distribution function $F(x)$. Sketch $F(x)$.

5.3 Continuous Random Variables and Distributions

Solution

From the pdf, we integrate to obtain:

Between 0 and $\frac{1}{2}$, $F(x) = \int_0^x 4u\, du = 2\mathrm{x}^2$;

Between $\frac{1}{2}$ and 1,

$$F(x) = \int_0^x f(u)du = [2x^2]_0^{\frac{1}{2}} + \int_{\frac{1}{2}}^x (4 - 4u)du$$

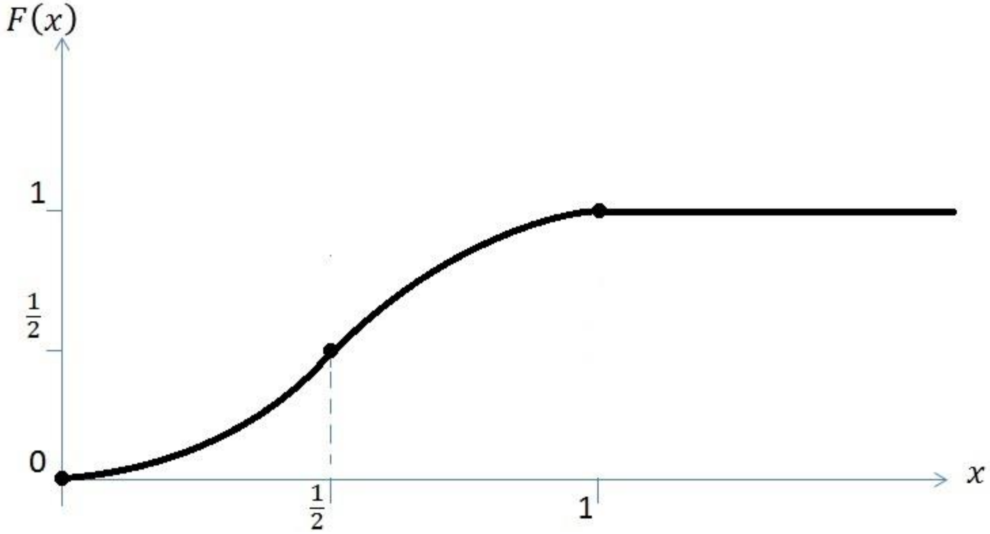
$$= -2x^2 + 4x - 1;$$

5.3 Continuous Random Variables and Distributions

Hence

$$F(x) = \begin{cases} 0; \quad x < 0 \\ 2x^2; \quad 0 \le x \le \frac{1}{2} \\ -2x^2 + 4x - 1; \quad \frac{1}{2} \le x \le 1 \\ 1; \quad x > 1 \end{cases}$$

The graph of the distribution function is:



5.4 Summary

- Discrete: probability mass function (pmf) $P(X = x_j)$
- Continuous: probability density function (pdf) $f(x)$
  - not a probability!
- Cumulative distribution functions (cdf) $F(X) = P(X \leq x)$
  - discrete $F\left(x_j\right) = \sum_{i \leq j} P(X = x_i)$
  - continuous $F(x) = \int_{-\infty}^{x} f(u)du$
- $P(a < X \leq b) = F(b) - F(a)$