Chapter 3

Random Variables and Probability Distributions

Section 3.1

Concept of a Random Variable

Definition 3.1

A **random variable** is a function that associates a real number with each element in the sample space.

Example 3.1: Two balls are drawn in succession without replacement from an urn containing 4 red balls and 3 black balls. The possible outcomes and the values y of the random variable Y, where Y is the number of red balls, are

Sample Space	y
RR	2
RB	1
BR	1
BB	0

Examples of RVs

- Consider coin flip
  - $S=\{H,T\}$
- Define random variable $X:S \Rightarrow R$
  - $X(s)=1$ if $s=H$, $X(s)=0$ if $s=T$
- Define random variable $Y:S \Rightarrow R$
  - $Y(s)=0$ if $s=H$, $Y(s)=1$ if $s=T$

Examples of RVs

- Consider roll of a die
  - $S=\{1,2,3,4,5,6\}$
- Define random variable X:S=>R
  - $X(s)=s$, for $s=1,2,3,4,5,6$
- Define another RV Y:S=>R
  - $Y(s)=0$ if $s=1,3,5$, $Y(s)=1$ if $s=2,4,6$
- Define another RV Z:S=>R
  - $Z(s)=2s+1$, for $s=1,\dots,6$

Definition 3.2

If a sample space contains a finite number of possibilities or an unending sequence with as many elements as there are whole numbers, it is called a **discrete sample space**.

- Example
  - Outcomes of a roll of a die
  - S={1,2,3,4,5,6}

Definition 3.3

If a sample space contains an infinite number of possibilities equal to the number of points on a line segment, it is called a **continuous sample space**.

- Example
  - Temperature of Seoul during year of 2018
  - $S=\{x \mid -15 < x < 38\}$

Definition

- Random variable X is said to be discrete RV if the range of X is discrete
  - That is X() takes discrete value
  - Example: S={1,…,6}, X(s)=2s for s in S

Definition

- Random variable X is said to be continuous RV if the range of X is continuous
  - Example: $S = \{s \mid -20 \leq s \leq 40\}$
    - S is temperature in centrigrade
    - Define random variable F such that

$$F = \frac{9}{5}s + 32,\ \ s \in S$$

Definition

- Random variable X is said to be continuous RV if the range of X is continuous
  - Example: $S = \{s | -20 \leq s \leq 40\}$
    - S is temperature in centigrade
    - Define random variable G such that

$$G(s) = \begin{cases} 1, & \text{if } s \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

    - G is continuous? Discrete?

Why use random variable?

- Using sample spaces is not mathematically convenient
  - S={H,T}
- By mapping the sample space (and events) to a real number, using RV, it is much more convenient to mathematically define and use the probability

Probability, Event and RV

- Probability: maps an event to a number in [0,1]
- Consider S={1,2,3,4,5,6}
- What is the probability of a die roll is odd number?
- P({1,3,5})

Probability, Event and RV

- Consider random variable X() such that
- X(s)=0 if s=1,3,5, X(s)=1 if s=2,4,6
- $P(\{s \in S | X(s) = 0\})$
- By properly defining a random variable, we can express the probability associated with events

Probability, Event and RV

- Consider outcome to 2 tosses of a coin
  S={HH,HT,TH,TT}
- What is the probability of at least one H?
- P({HH,HT,TH})
- Let X denote the RV counting the number of H in s,
- X(HH)=2, X(HT)=X(TH)=1, X(TT)=0
- $P(\{s \in S | X(s) \geq 1\})$
- Too complicated, lets write instead

$$P(X \geq 1)$$

Probability, Event and RV

- Let RV X such that
  - S={H,T}, X(s)=1 if s=H, X(s)=0 if s=T
- Let RV Y such that
  - S={1,2,3,4,5,6}, Y(s)=1 if s odd, Y(s)=0 if s even
- X and Y have same distribution
  - same kind of random variable
- So sometimes only the distribution of RV is discussed
  - originating sample space/experiment can be omitted
  - Same for discussing probability of an event

Section 3.2

Discrete Probability Distribution

Definition 3.4

The set of ordered pairs $(x, f(x))$ is a **probability function**, **probability mass function**, or **probability distribution** of the discrete random variable X if, for each possible outcome x,

1. $f(x) \geq 0$,
2. $\sum\limits_{x} f(x) = 1$,
3. $P(X = x) = f(x)$.

Definition 3.4

The set of ordered pairs $(x, f(x))$ is a **probability function**, **probability mass function**, or **probability distribution** of the discrete random variable X if, for each possible outcome x,

1. $f(x) \geq 0$,
2. $\sum_x f(x) = 1$,
3. $P(X = x) = f(x)$.

- Let X denote an RV that models outcome of a fair coin toss, where X=0 if H, and X=1 if T
- $P(X=0)=P(X=1)=0.5$
- PMF f is such that $f(x)=0.5$ for $x=0,1$

Definition 3.4

The set of ordered pairs $(x, f(x))$ is a **probability function**, **probability mass function**, or **probability distribution** of the discrete random variable X if, for each possible outcome x,

1. $f(x) \geq 0$,
2. $\sum_{x} f(x) = 1$,
3. $P(X = x) = f(x)$.

Example 3.8: A shipment of 20 similar laptop computers to a retail outlet contains 3 that are defective. If a school makes a random purchase of 2 of these computers, find the probability distribution for the number of defectives.

Definition 3.5

The **cumulative distribution function** $F(x)$ of a discrete random variable X with probability distribution $f(x)$ is

$$F(x) = P(X \le x) = \sum_{t \le x} f(t), \quad \text{for } -\infty < x < \infty.$$

- Note that F(x) is always defined over the entire real number axis

Figure 3.1 Probability mass function plot

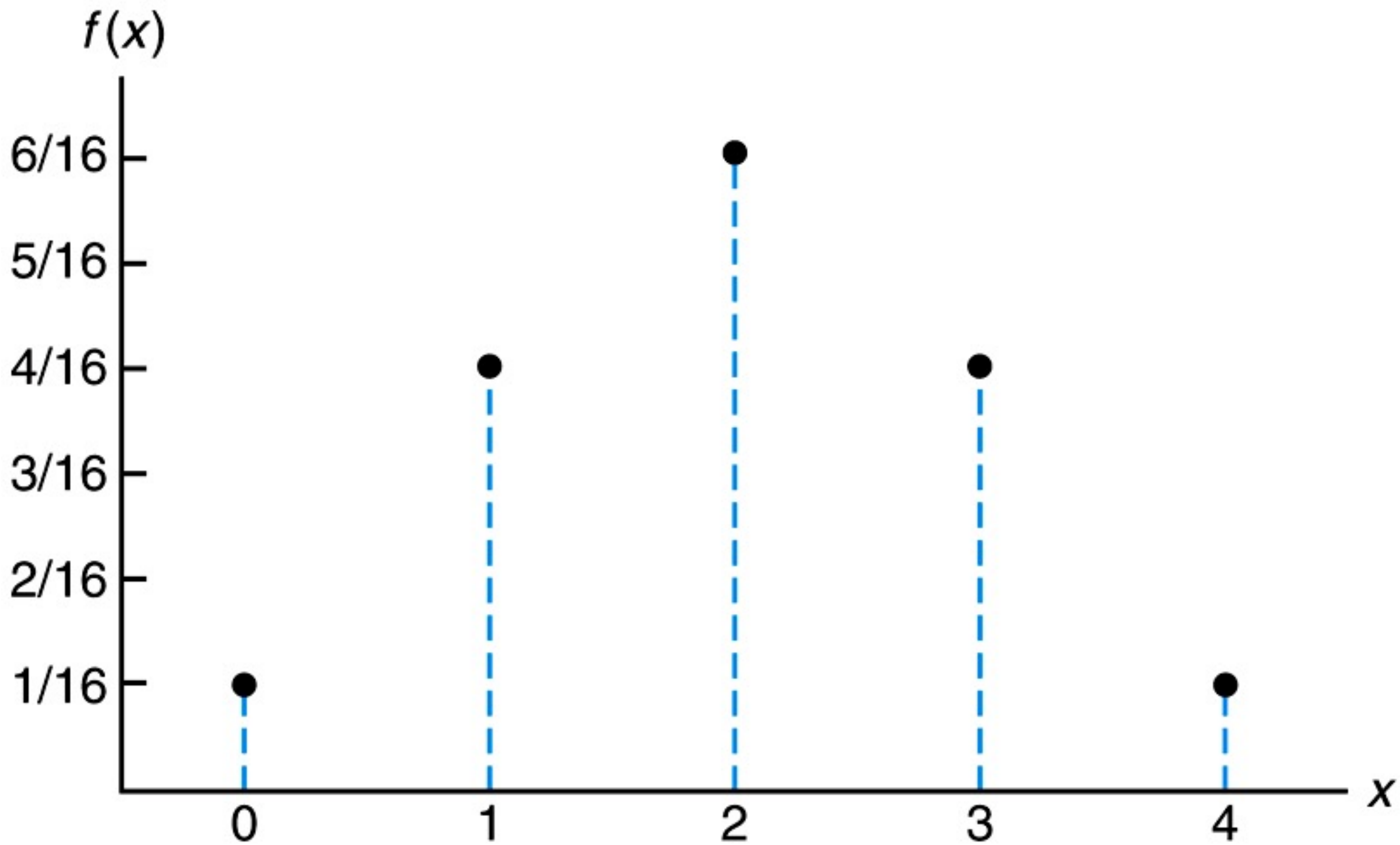


Figure 3.2 Probability histogram

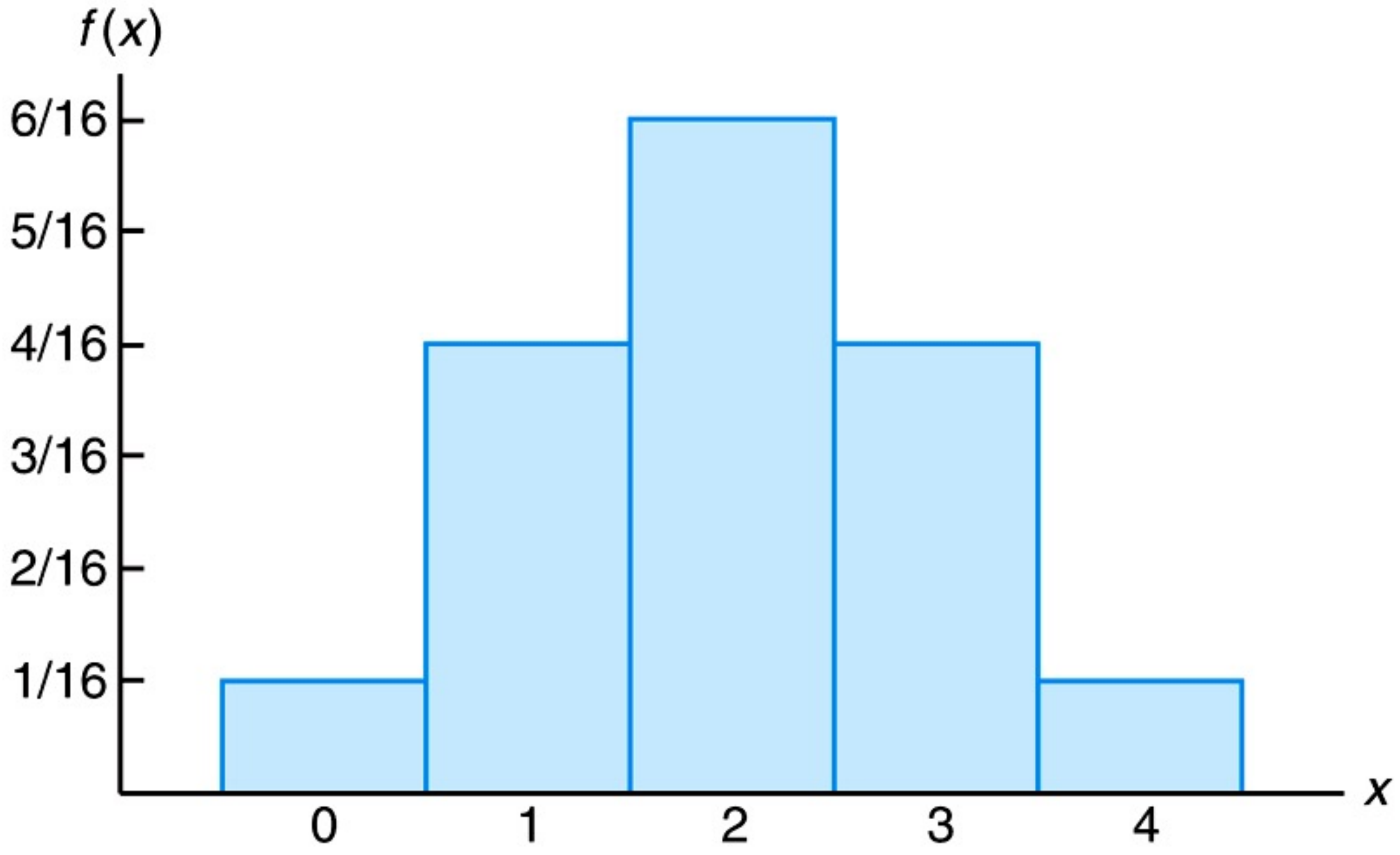
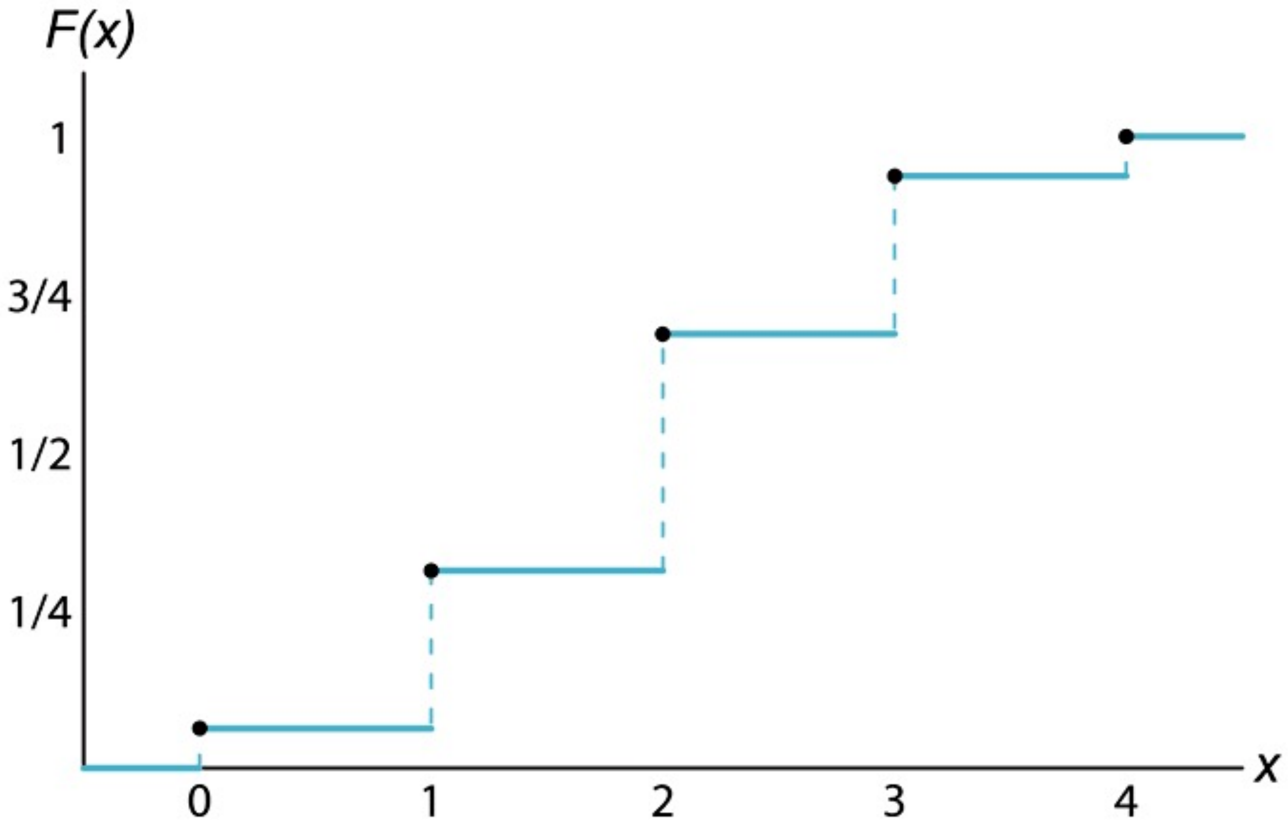


Figure 3.3 Discrete cumulative distribution function



Section 3.3

Continuous Probability Distributions

Figure 3.4 Typical density functions

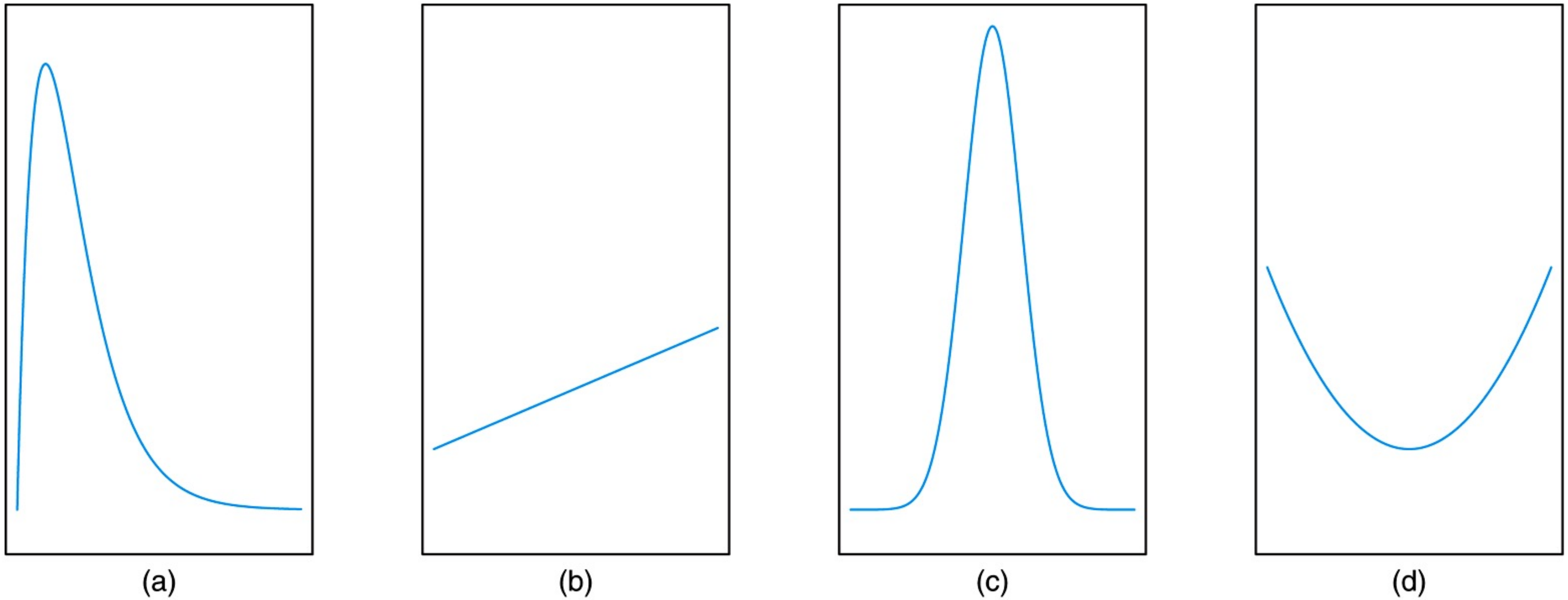
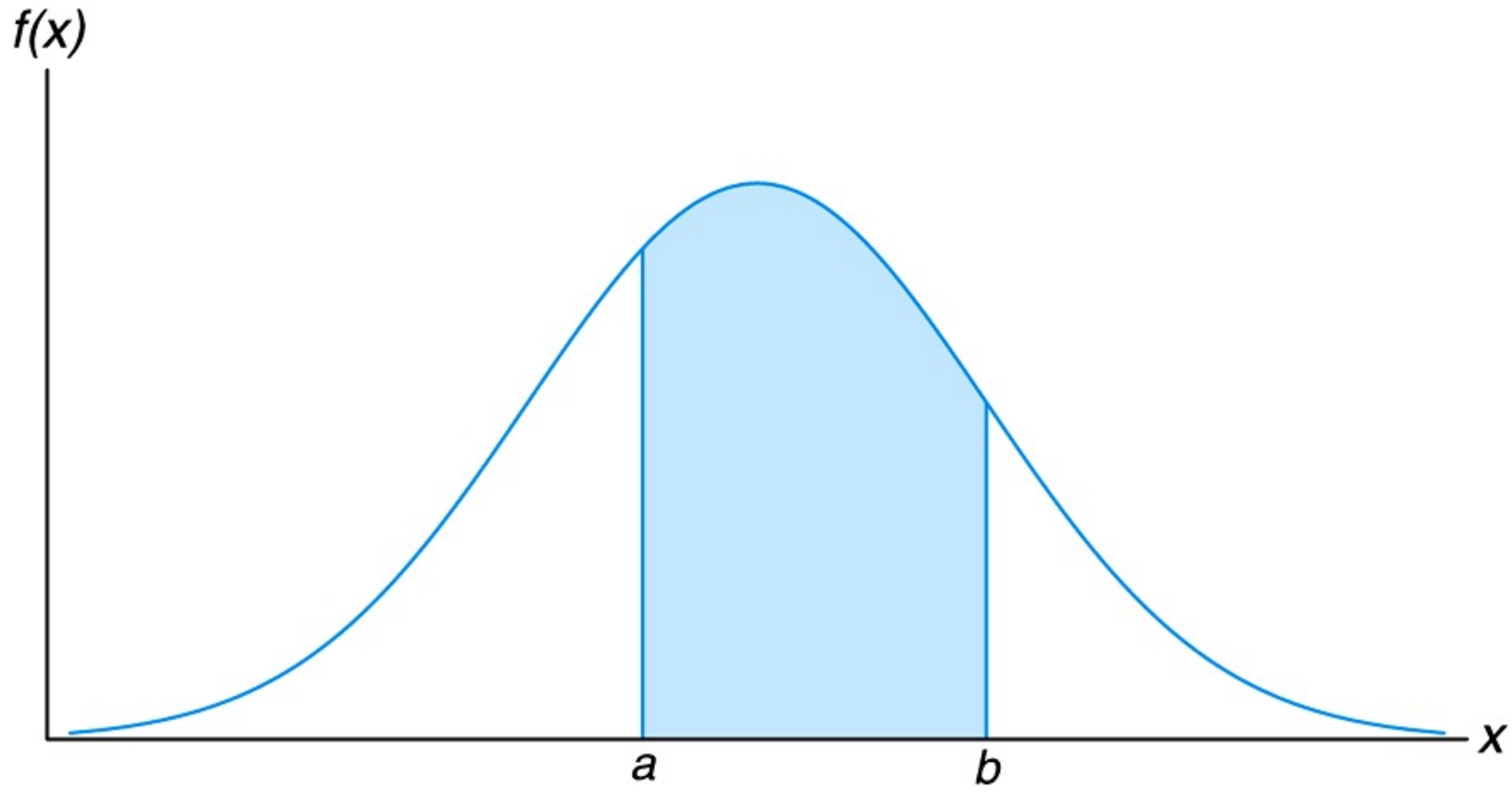


Figure 3.5 $P(a < X < b)$



Definition 3.6

The function $f(x)$ is a **probability density function** (pdf) for the continuous random variable X, defined over the set of real numbers, if

1. $f(x) \geq 0$, for all $x \in R$.
2. $\int_{-\infty}^{\infty} f(x)\, dx = 1$.
3. $P(a < X < b) = \int_a^b f(x)\, dx$.

Definition 3.6

The function $f(x)$ is a **probability density function** (pdf) for the continuous random variable X, defined over the set of real numbers, if

1. $f(x) \geq 0$, for all $x \in R$.
2. $\int_{-\infty}^{\infty} f(x)\, dx = 1$.
3. $P(a < X < b) = \int_a^b f(x)\, dx$.

Probability that X is in $[x, x + dx)$ is $\approx f(x)dx$

Definition 3.6

The function $f(x)$ is a **probability density function** (pdf) for the continuous random variable X, defined over the set of real numbers, if

1. $f(x) \geq 0$, for all $x \in R$.
2. $\int_{-\infty}^{\infty} f(x)\, dx = 1$.
3. $P(a < X < b) = \int_a^b f(x)\, dx$.

$$\text{Probability that } X \text{ is in } [x, x+dx) \text{ is } \approx f(x)dx$$

$$f(x) \text{ represents the relative weight of how likely } X \text{ is near } x$$

- Probability of taking an exact value x is 0

Example 3.11

Example 3.11: Suppose that the error in the reaction temperature, in °C, for a controlled laboratory experiment is a continuous random variable X having the probability density function

$$f(x) = \begin{cases} \frac{x^2}{3}, & -1 < x < 2, \\ 0, & \text{elsewhere.} \end{cases}$$

.

(a) Verify that $f(x)$ is a density function.

(b) Find $P(0 < X \leq 1)$.

Definition 3.7

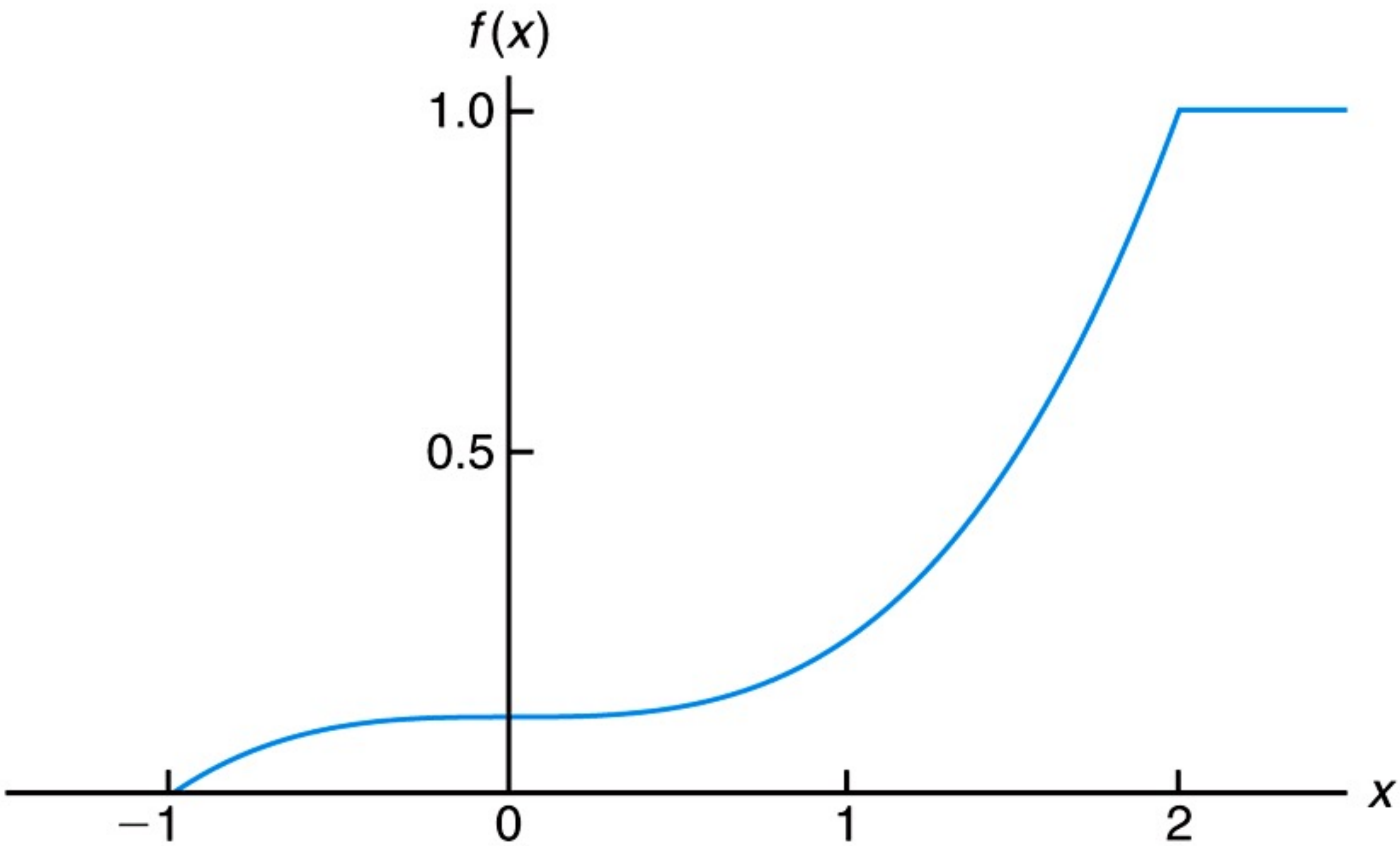
The **cumulative distribution function** $F(x)$ of a continuous random variable X with density function $f(x)$ is

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\, dt, \quad \text{for } -\infty < x < \infty.$$

Example 3.12

Example 3.12: For the density function of Example 3.11, find $F(x)$, and use it to evaluate $P(0 < X \leq 1)$.

Figure 3.6 Continuous cumulative distribution function



Section 3.4

Joint Probability Distributions

Definition 3.8

The function $f(x, y)$ is a **joint probability distribution** or **probability mass function** of the discrete random variables X and Y if

1. $f(x, y) \geq 0$ for all (x, y),
2. $\sum_x \sum_y f(x, y) = 1$,
3. $P(X = x, Y = y) = f(x, y)$.

For any region A in the xy plane, $P[(X, Y) \in A] = \sum \sum_A f(x, y)$.

Example 3.14

Example 3.14: Two ballpoint pens are selected at random from a box that contains 3 blue pens, 2 red pens, and 3 green pens. If X is the number of blue pens selected and Y is the number of red pens selected, find

(a) the joint probability function $f(x, y)$,

(b) $P[(X, Y) \in A]$, where A is the region $\{(x, y) | x + y \leq 1\}$.

Table 3.1 Joint Probability Distribution for Example 3.14

	$f(x,y)$	x: 0	1	2	Row Totals
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$	0	$\frac{3}{7}$
	2	$\frac{1}{28}$	0	0	$\frac{1}{28}$
Column Totals		$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

Definition 3.9

The function $f(x, y)$ is a **joint density function** of the continuous random variables X and Y if

1. $f(x, y) \geq 0$, for all (x, y),
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dx\, dy = 1$,
3. $P[(X, Y) \in A] = \int \int_A f(x, y)\, dx\, dy$, for any region A in the xy plane.

Example 3.15

Example 3.15: A privately owned business operates both a drive-in facility and a walk-in facility. On a randomly selected day, let X and Y, respectively, be the proportions of the time that the drive-in and the walk-in facilities are in use, and suppose that the joint density function of these random variables is

$$f(x, y) = \begin{cases} \frac{2}{5}(2x + 3y), & 0 \leq x \leq 1, 0 \leq y \leq 1, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Verify condition 2 of Definition 3.9.

(b) Find $P[(X, Y) \in A]$, where $A = \{(x, y) \mid 0 < x < \frac{1}{2}, \frac{1}{4} < y < \frac{1}{2}\}$.

Definition 3.10

The **marginal distributions** of X alone and of Y alone are

$$g(x) = \sum_{y} f(x, y) \quad \text{and} \quad h(y) = \sum_{x} f(x, y)$$

for the discrete case, and

$$g(x) = \int_{-\infty}^{\infty} f(x, y)\, dy \quad \text{and} \quad h(y) = \int_{-\infty}^{\infty} f(x, y)\, dx$$

for the continuous case.

Example 3.17

Example 3.17: Find $g(x)$ and $h(y)$ for the joint density function of Example 3.15.

Definition 3.11

Let X and Y be two random variables, discrete or continuous. The **conditional distribution** of the random variable Y given that $X = x$ is

$$f(y|x) = \frac{f(x,y)}{g(x)}, \text{ provided } g(x) > 0.$$

Similarly, the conditional distribution of X given that $Y = y$ is

$$f(x|y) = \frac{f(x,y)}{h(y)}, \text{ provided } h(y) > 0.$$

Definition 3.11

Let X and Y be two random variables, discrete or continuous. The **conditional distribution** of the random variable Y given that $X = x$ is

$$f(y|x) = \frac{f(x,y)}{g(x)}, \text{ provided } g(x) > 0.$$

Similarly, the conditional distribution of X given that $Y = y$ is

$$f(x|y) = \frac{f(x,y)}{h(y)}, \text{ provided } h(y) > 0.$$

$f(y|x)$ is distribution of Y, so $\sum_y f(y|x) = 1$

$f(y|x)$ is function of x and y, but $f(y|x = c)$ is function of y only

Definition 3.11

Let X and Y be two random variables, discrete or continuous. The **conditional distribution** of the random variable Y given that $X = x$ is

$$f(y|x) = \frac{f(x,y)}{g(x)}, \text{ provided } g(x) > 0.$$

Similarly, the conditional distribution of X given that $Y = y$ is

$$f(x|y) = \frac{f(x,y)}{h(y)}, \text{ provided } h(y) > 0.$$

For continuous RVs, $f(x|Y = y)$ may seem strange, because prob. of $Y = y$ is 0

But $Y = y$ is something already observed (prior information)

Table 3.1 Joint Probability Distribution for Example 3.14

	$f(x, y)$	$x = 0$	$x = 1$	$x = 2$	Row Totals
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$	0	$\frac{3}{7}$
	2	$\frac{1}{28}$	0	0	$\frac{1}{28}$
Column Totals		$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

Find $g(x)$ and $h(y)$

Find $f(x|y = 0)$ and $f(y|x = 1)$

Example 3.19

Example 3.19: The joint density for the random variables (X, Y), where X is the unit temperature change and Y is the proportion of spectrum shift that a certain atomic particle produces, is

$$f(x, y) = \begin{cases} 10xy^2, & 0 < x < y < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Find the marginal densities $g(x)$, $h(y)$, and the conditional density $f(y|x)$.

(b) Find the probability that the spectrum shifts more than half of the total observations, given that the temperature is increased by 0.25 unit.

Definition 3.12

Let X and Y be two random variables, discrete or continuous, with joint probability distribution $f(x, y)$ and marginal distributions $g(x)$ and $h(y)$, respectively. The random variables X and Y are said to be **statistically independent** if and only if

$$f(x, y) = g(x)h(y)$$

for all (x, y) within their range.

Definition 3.13

Let $X_1, X_2, \ldots, X_n$ be n random variables, discrete or continuous, with joint probability distribution $f(x_1, x_2, \ldots, x_n)$ and marginal distribution $f_1(x_1), f_2(x_2), \ldots, f_n(x_n)$, respectively. The random variables $X_1, X_2, \ldots, X_n$ are said to be mutually **statistically independent** if and only if

$$f(x_1, x_2, \ldots, x_n) = f_1(x_1)f_2(x_2)\cdots f_n(x_n)$$

for all $(x_1, x_2, \ldots, x_n)$ within their range.

Example 3.21

Example 3.21: Show that the random variables of Example 3.14 are not statistically independent.

Example 3.22

Example 3.22: Suppose that the shelf life, in years, of a certain perishable food product packaged in cardboard containers is a random variable whose probability density function is given by

$$f(x) = \begin{cases} e^{-x}, & x > 0, \\ 0, & \text{elsewhere.} \end{cases}$$

Let X_1, X_2, and X_3 represent the shelf lives for three of these containers selected independently and find $P(X_1 < 2, 1 < X_2 < 3, X_3 > 2)$.