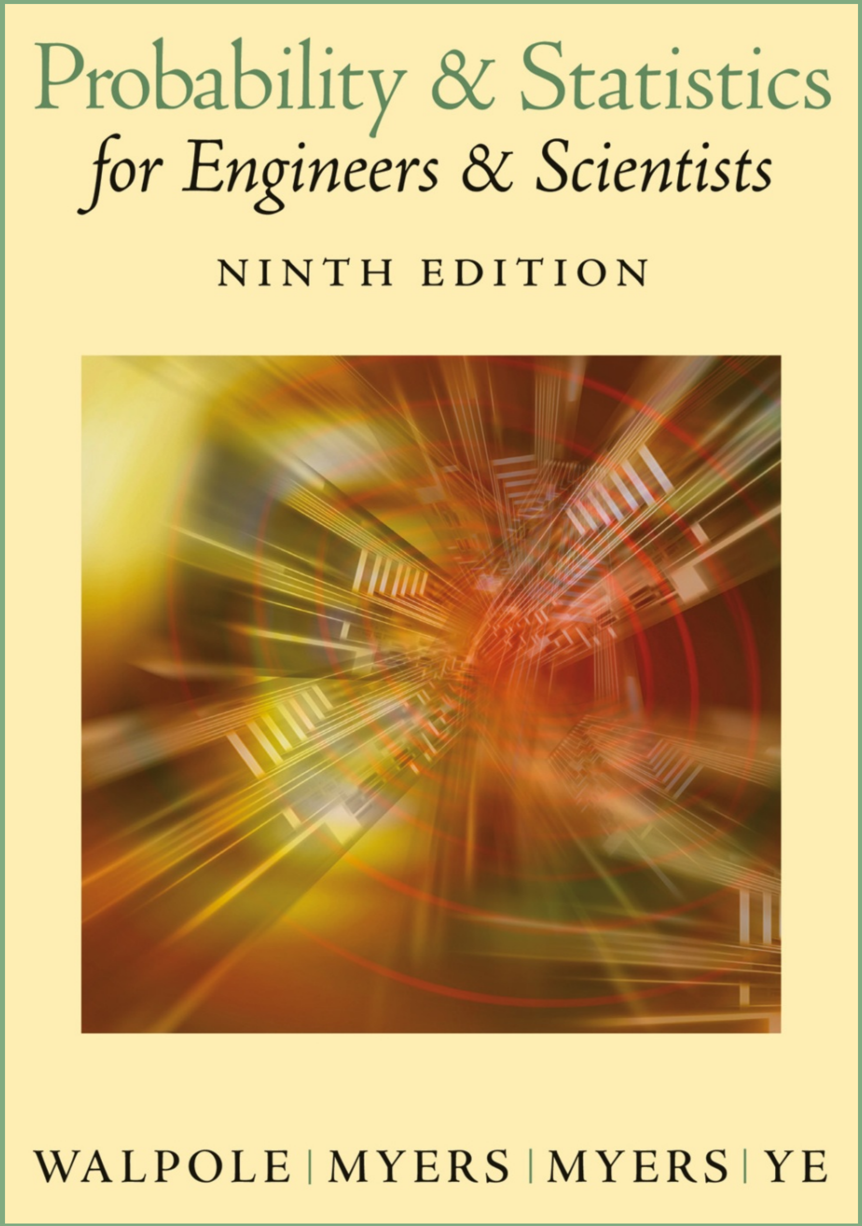
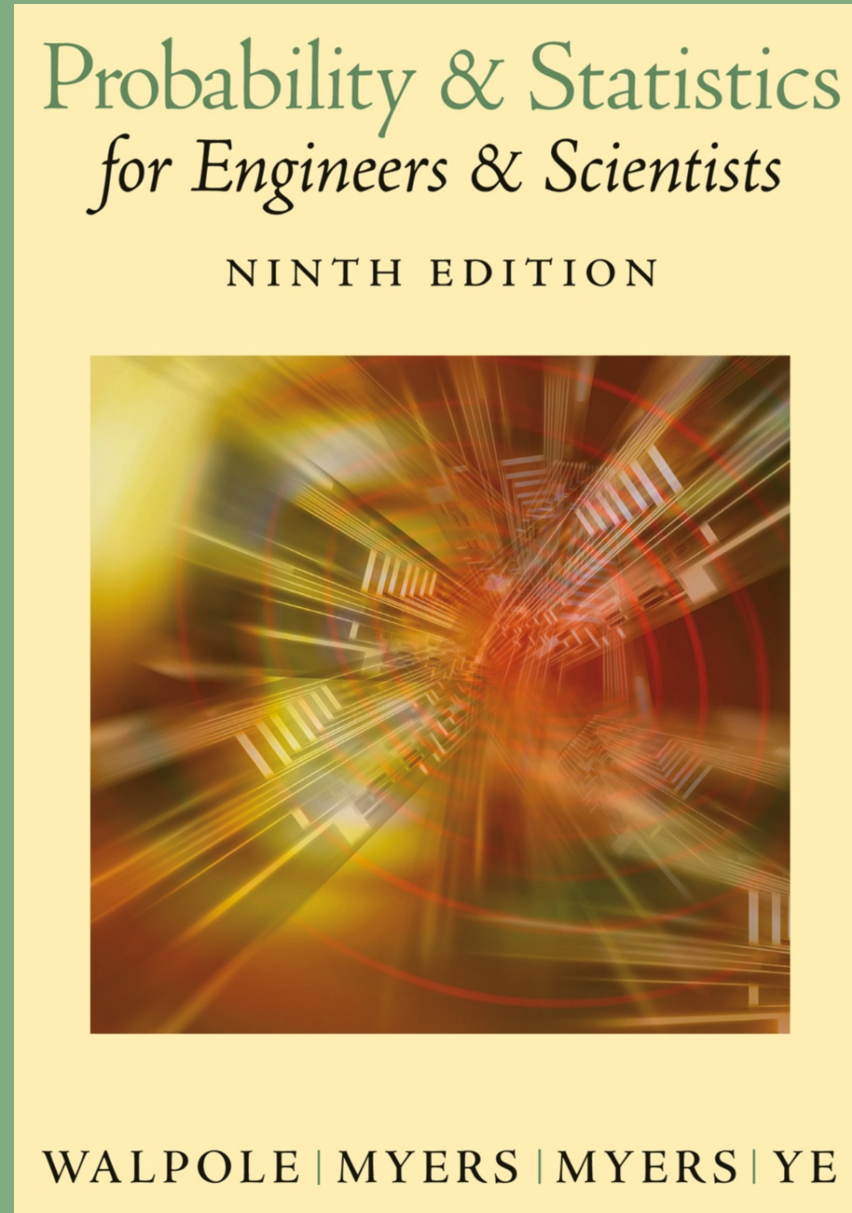
Chapter 3

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.

Definition 3.2

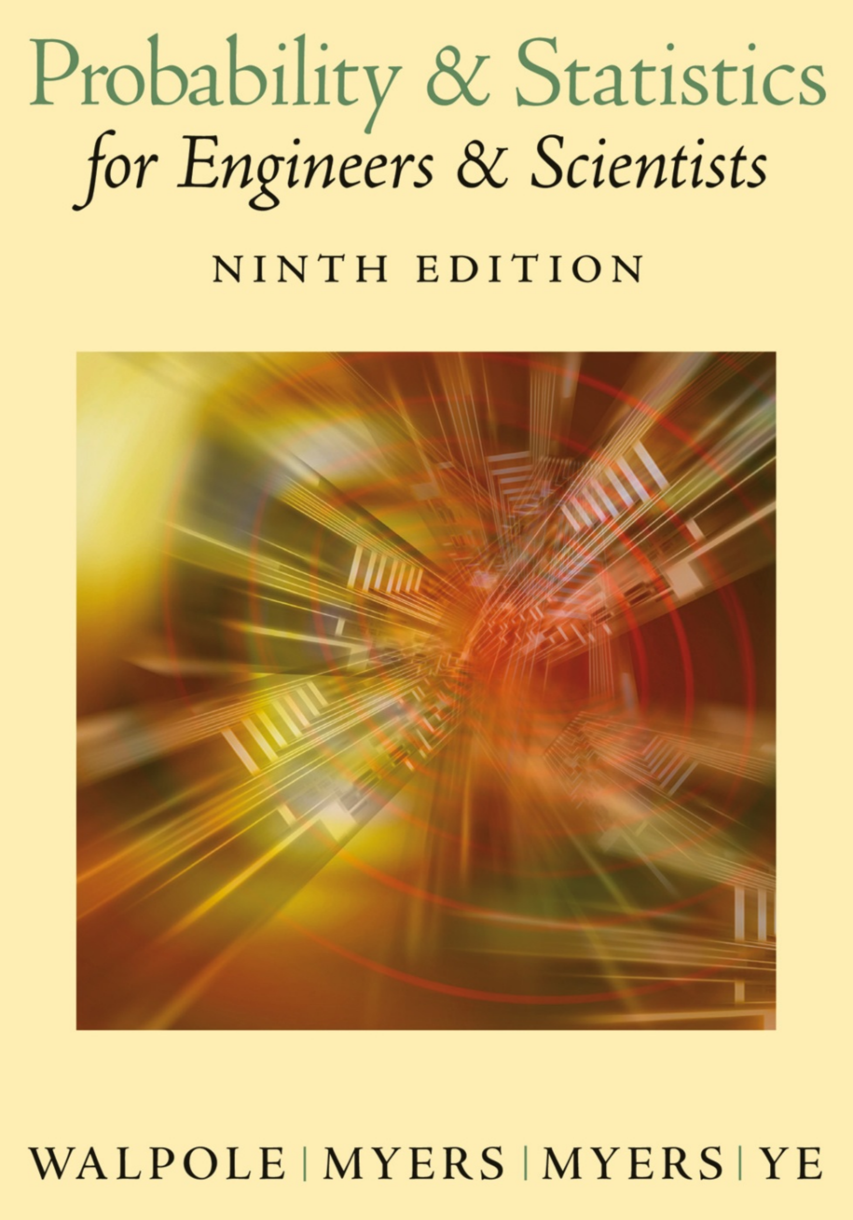
If a sample space contains a finite number of possibilities or an unending sequence with as many elements as there are whole numbers, it is called a **discrete sample space**.

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**.

Section 3.2

Discrete
Probability
Distribution



PEARSON
Addison
Wesley

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. $P(X = x) = f(x)$.

Definition 3.5

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

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

Figure 3.1 Probability mass function plot

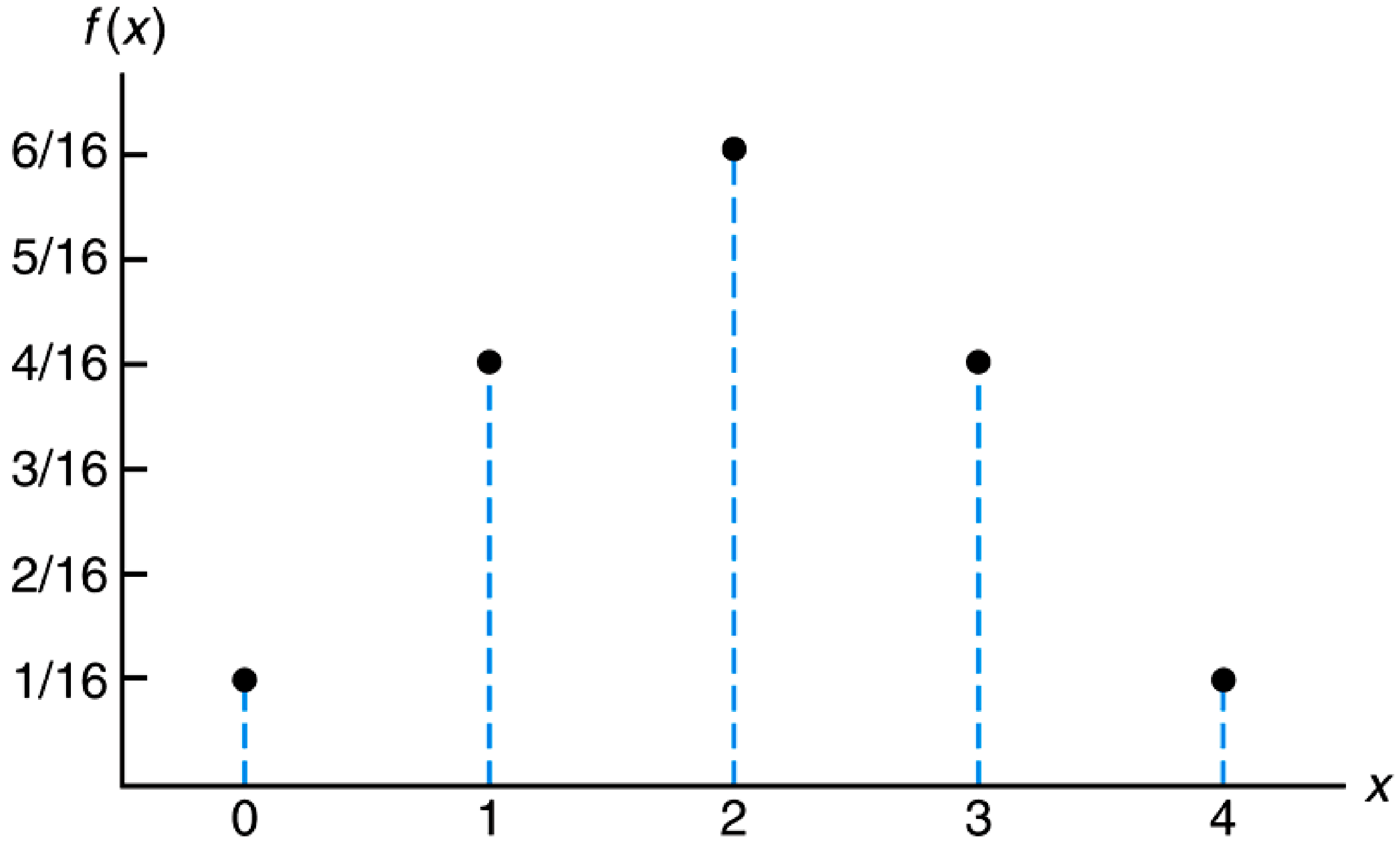


Figure 3.2 Probability histogram

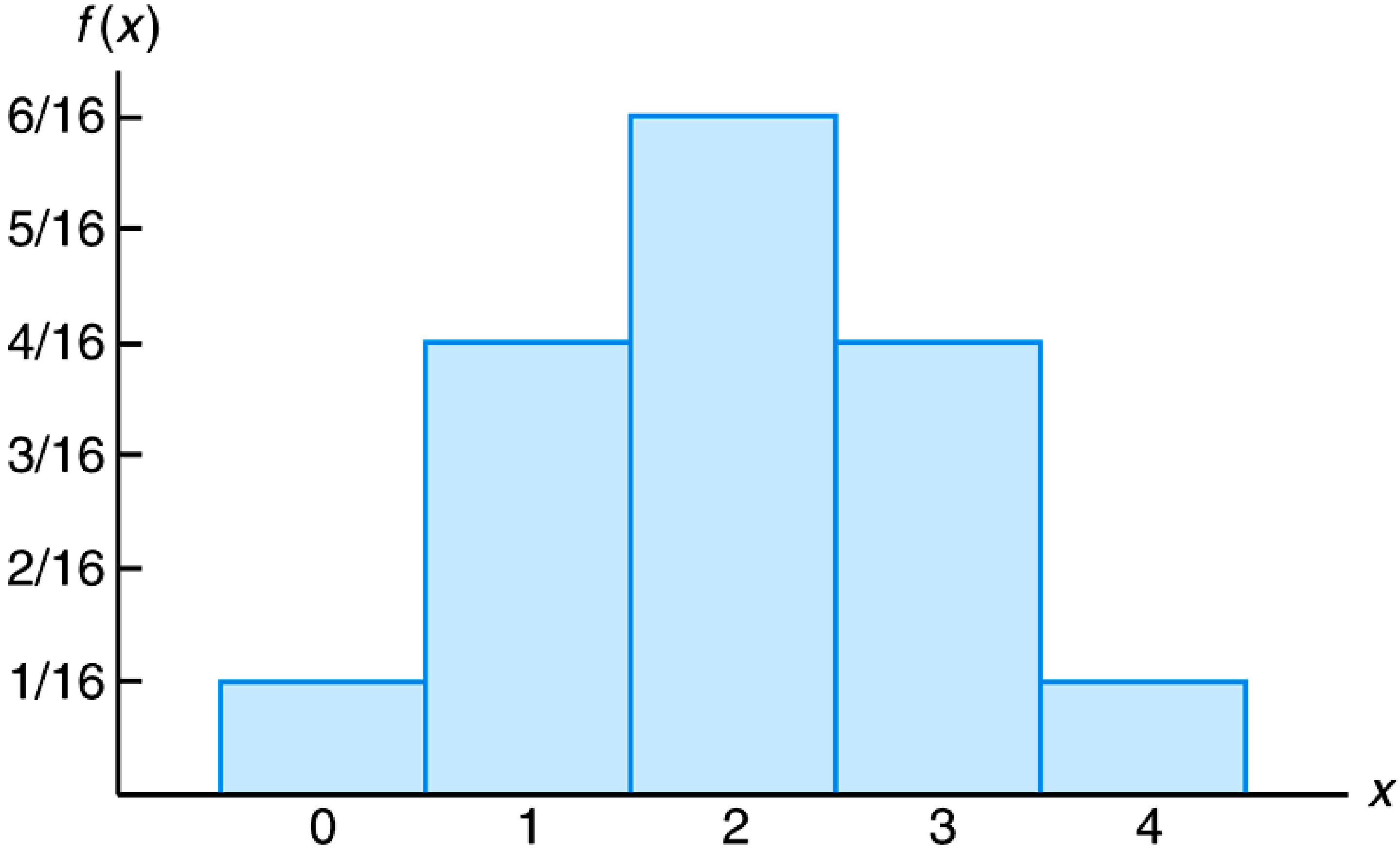
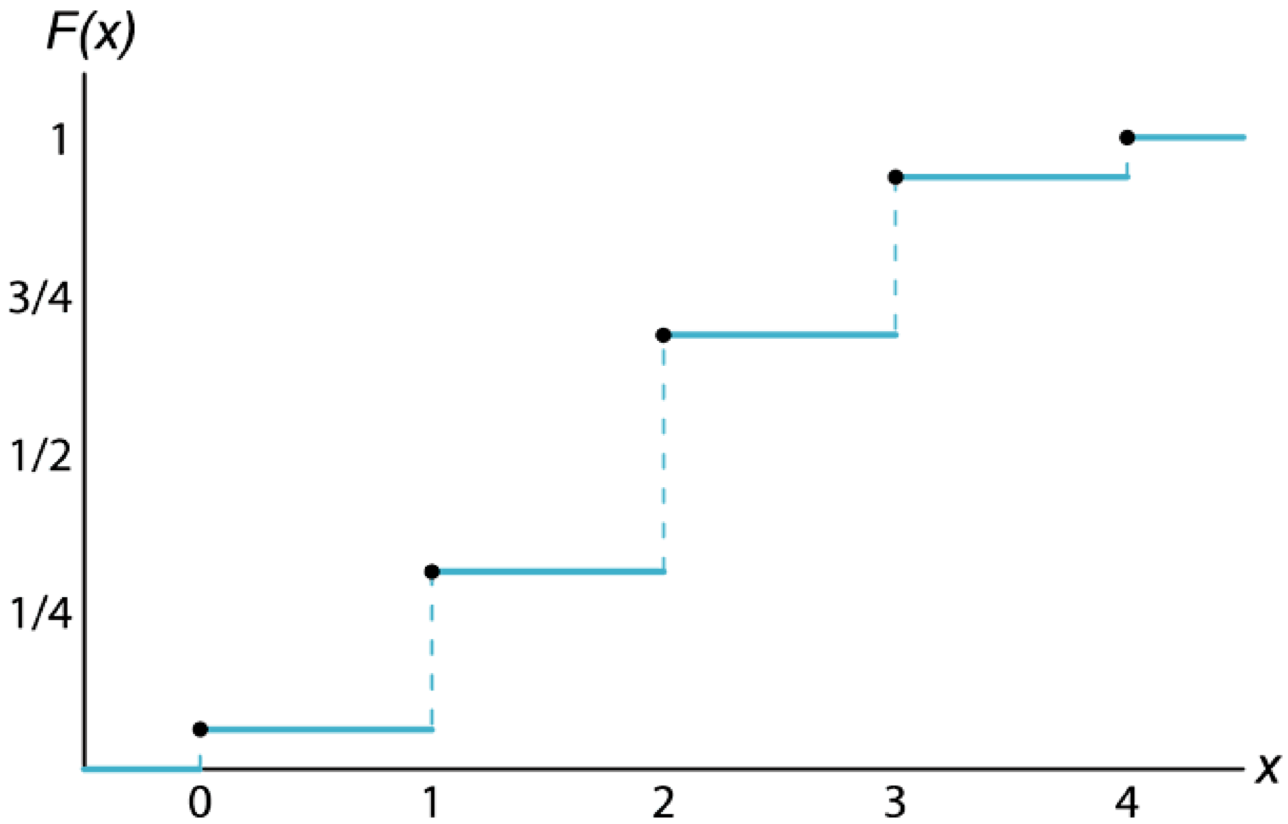


Figure 3.3 Discrete cumulative distribution function



Section 3.3

Continuous Probability Distributions

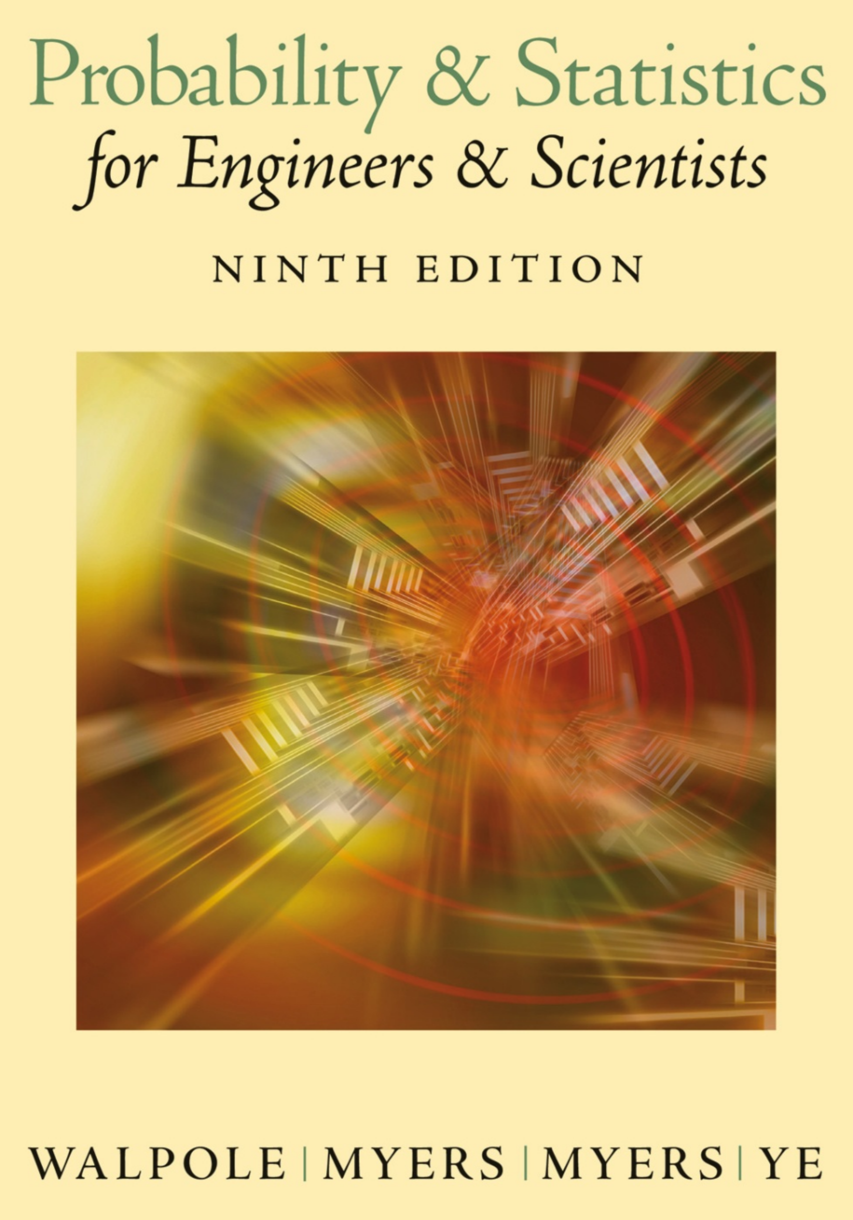


Figure 3.4 Typical density functions

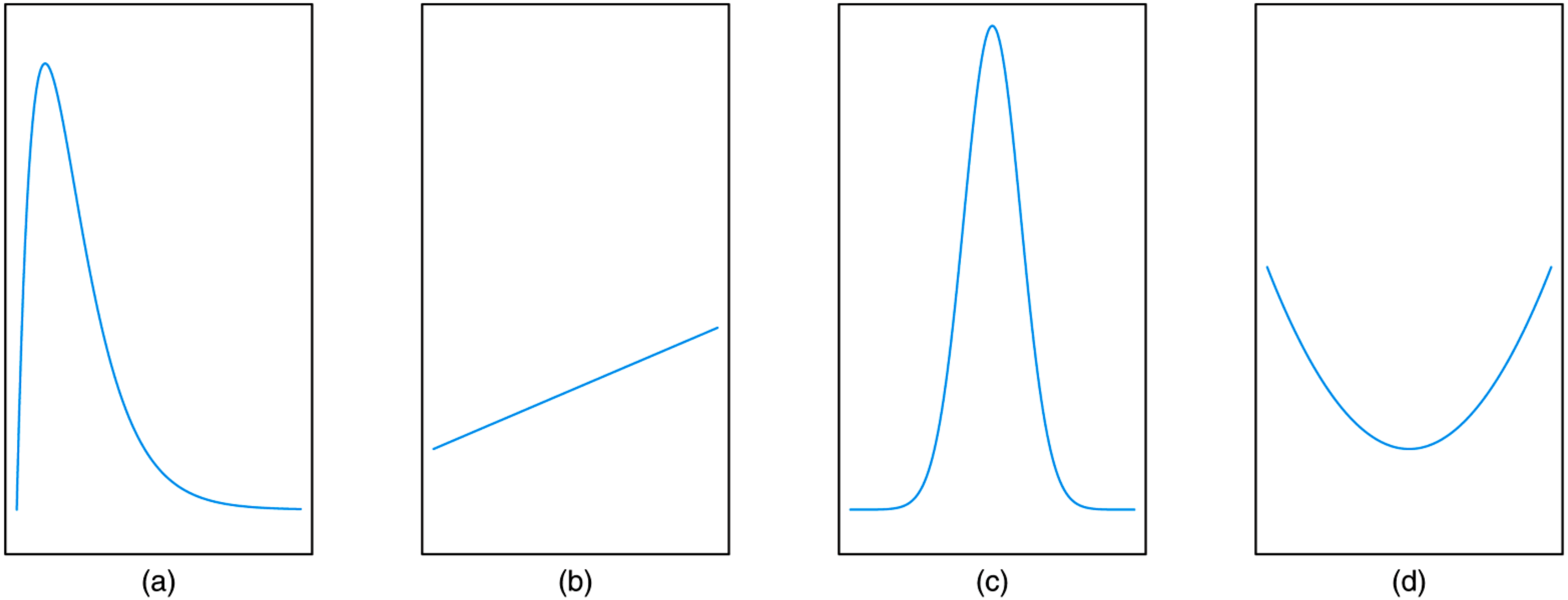
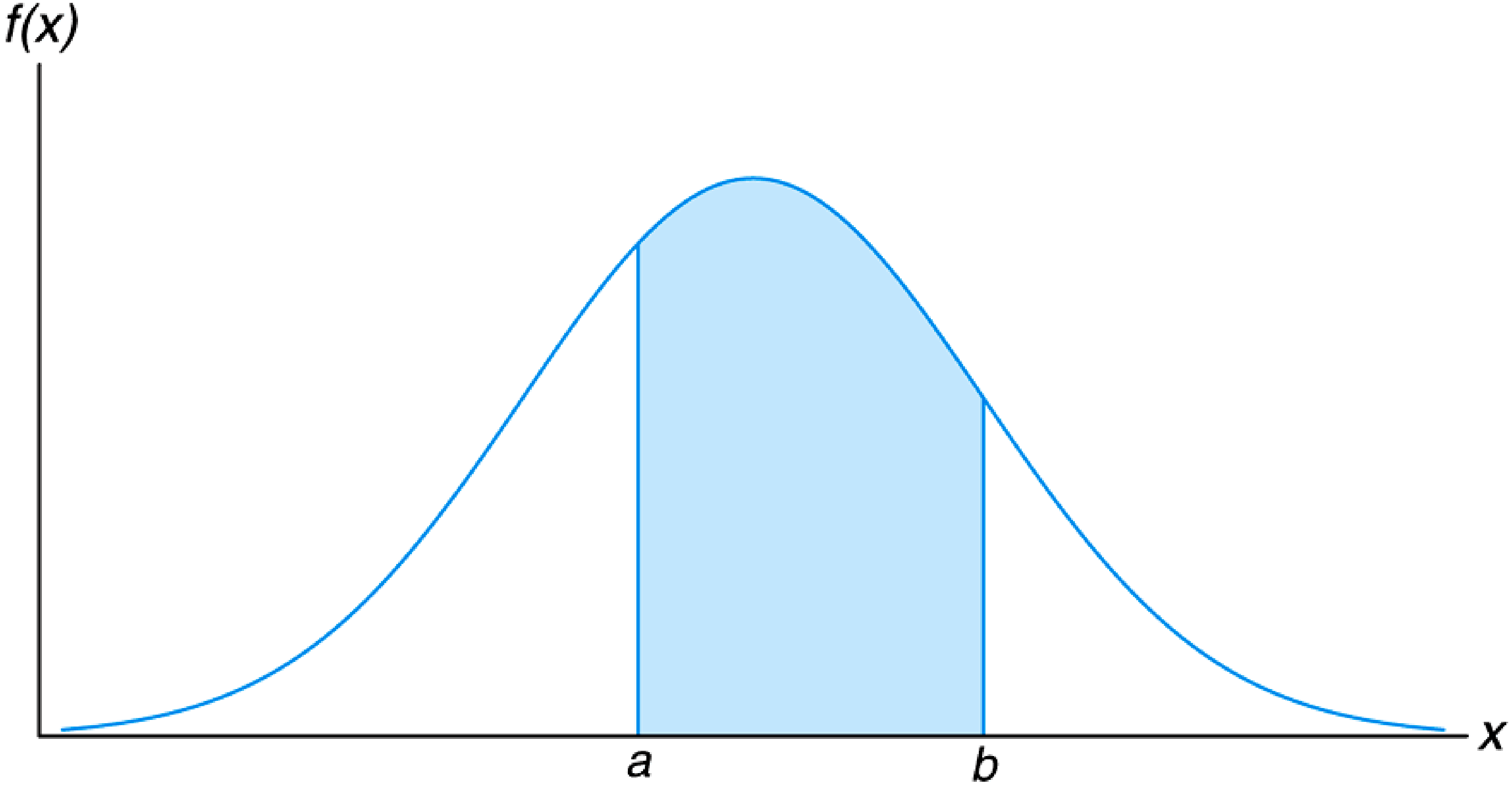


Figure 3.5 $P(a < X < b)$



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

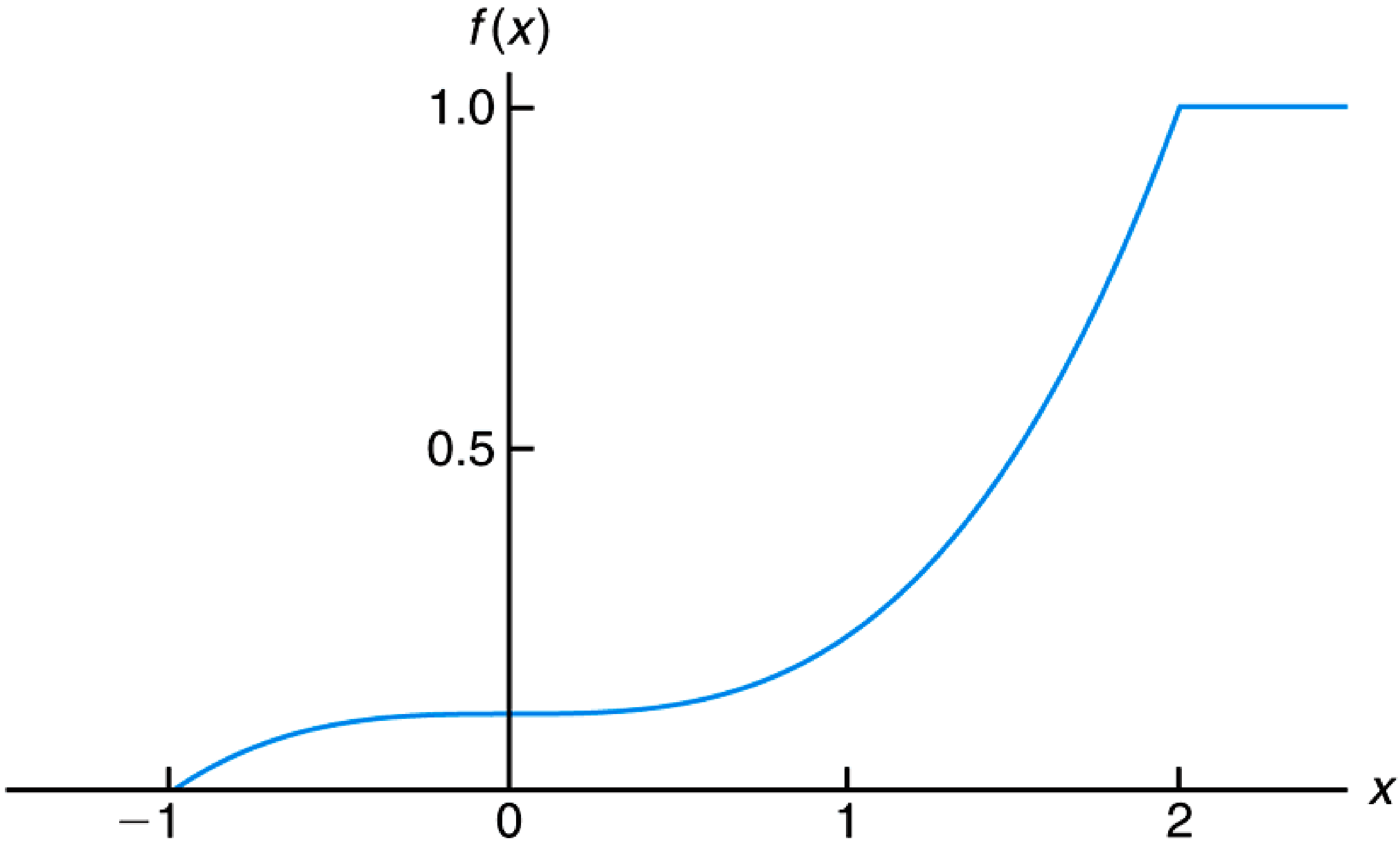
1. $f(x) \geq 0$, for all $x \in R$.
2. $\int_{-\infty}^{\infty} f(x)\,dx = 1$.
3. $P(a < X < b) = \int_a^b f(x)\,dx$.

Definition 3.7

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

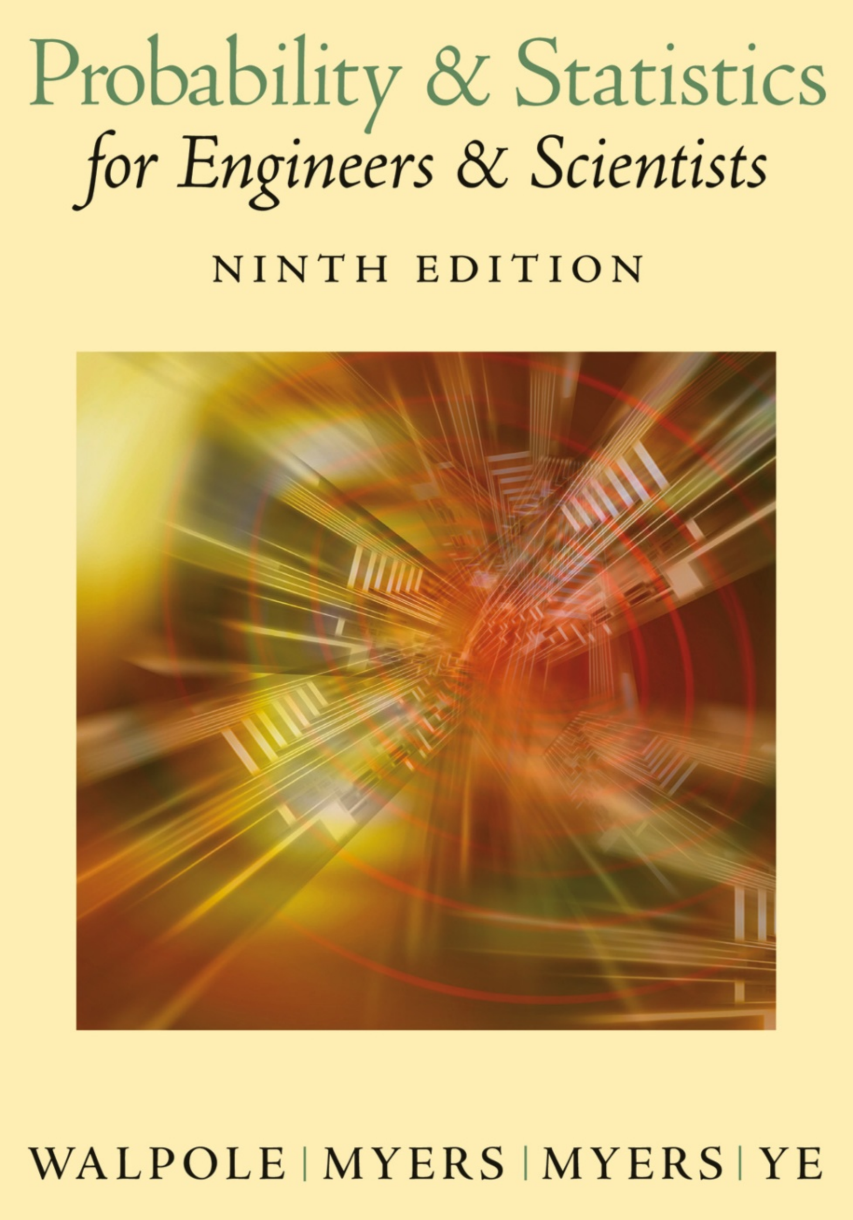
$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\, dt, \quad \text{for } -\infty < x < \infty.$$

Figure 3.6 Continuous cumulative distribution function



Section 3.4

Joint Probability Distributions



Section Abstract

- We generalize the concept of distribution of a single random variable to the case of two r.v.'s where we speak of their joint distribution.

We do so by introducing:

1. The joint p.f. of two discrete r.v.'s
2. The joint p.d.f. of two continuous r.v.'s
3. The joint d.f. for any two r.v.'s

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, $P[(X, Y) \in A] = \sum \sum_{A} f(x, y)$.

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

Discrete Joint Distributions

- Definition by example:

A subcommittee of five members is to be formed randomly from among a committee of 10 democrats, 8 republicans and 2 independent members.

- Let X and Y be the numbers of democrats and republicans chosen to serve on the committee, respectively. Their joint p.f. is:

$$P(x,y) := P(X = x, Y = y)$$

$$= C_{10,x} C_{8,y} C_{2,5-x-y} / C_{20,5}$$

$X = 0,1,\ldots,5;\ y = 0,1,\ldots,5;\ x+y \leq 5$

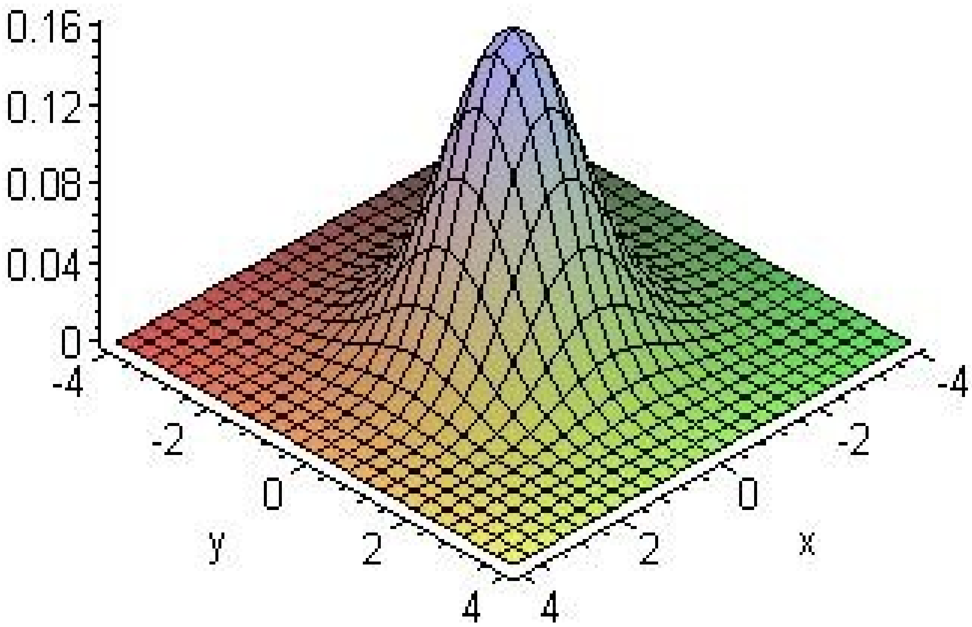
and $x+y \geq 3$

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.

A Typical Shape



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.

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 3.11

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 3.12
(Independence of random Variables)

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

- Roll a fair die once and let X be the number of dots.
- Toss a fair coin X times and let Y be the number of heads.

Find the joint p.m.f. of X and Y.

Example: Table of p(x,y)

x \ y	0	1	2	3	4	5	6	$p_X(x)$
1	1/12	1/12	0	0	0	0	0	1/6
2	1/24	2/24	1/24	0	0	0	0	1/6
3	1/48	3/48	3/48	1/48	0	0	0	1/6
4	1/96	4/96	6/96	4/96	1/96	0	0	1/6
5	1/192	5/192	10/192	10/192	5/192	1/192	0	1/6
6	1/384	6/384	15/384	20/384	15/384	6/384	1/384	1/6
$p_Y(y)$	63/384	120/384	99/384	64/384	29/384	8/384	1/384	1

Marginal distribution of X

x \ y	0	1	2	3	4	5	6	$p_X(x)$
1	1/12	1/12	0	0	0	0	0	1/6
2	1/24	2/24	1/24	0	0	0	0	1/6
3	1/48	3/48	3/48	1/48	0	0	0	1/6
4	1/96	4/96	6/96	4/96	1/96	0	0	1/6
5	1/192	5/192	10/192	10/192	5/192	1/192	0	1/6
6	1/384	6/384	15/384	20/384	15/384	6/384	1/384	1/6
$p_Y(y)$	63/384	120/384	99/384	64/384	29/384	8/384	1/384	1

Marginal distribution of Y

x \ y	0	1	2	3	4	5	6	$p_X(x)$
1	1/12	1/12	0	0	0	0	0	1/6
2	1/24	2/24	1/24	0	0	0	0	1/6
3	1/48	3/48	3/48	1/48	0	0	0	1/6
4	1/96	4/96	6/96	4/96	1/96	0	0	1/6
5	1/192	5/192	10/192	10/192	5/192	1/192	0	1/6
6	1/384	6/384	15/384	20/384	15/384	6/384	1/384	1/6
$p_Y(y)$	63/384	120/384	99/384	64/384	29/384	8/384	1/384	1

Examples

Find the constant c that renders the function $f(x,y)$ into a joint p.d.f.:

- $f(x,y)= cxy^2;\ 0 \le x \le 1;\ 0 \le y \le 1$
  AND zero elsewhere
- $f(x,y)= cxy^2;\ 0 \le x \le y \le 1$
  AND zero elsewhere

Find $P(X>3/4)$ in each case.

Will the Chicken Be Safe?

- A farmer wants to build a triangular pen for his chickens. He sends his son out to cut the lumber and the boy, without taking any thought as to the ultimate purpose, makes two cuts at two points selected at random. What are the chances that the resulting three pieces can be used to form a triangular pen?

Will the Chicken Be Safe?

- Suppose that the length of the lumber is L
- Let X and Y be the distances from the left to the two points where the cuts "occur"
- Two cases: (1) X < Y and (2) Y < X

Case1: pieces have lengths: X, Y-X, L-Y

Case2: pieces have lengths: Y, X-Y, L-X.

By symmetry, both cases have the same probability

Will the Chicken Be Safe?

Case 1: the three pieces form a triangle iff

The length of any piece is less than the sum of the other two. We get the restrictions:

$X < Y$

$X < (Y-X) + (L-Y) \Rightarrow X < L/2$

$(Y-X) < X + (L-Y) \Rightarrow (Y-X) < L/2$

$L-Y < X + (Y-X) \Rightarrow Y > L/2$

Will the Chicken Be Safe?

- So the conditions are

$X < Y; X < L/2; (Y-X) < L/2; Y > L/2.$

- Then integrate the joint pdf of X and Y in the domain determined by those conditions.

Note: (X,Y) is a point taken uniformly from the square $0 < X < L, 0 < Y < L$,

Hence it has denisty $f(x,y) = 1/L^2$ in that square and zero elsewhere