Probability and Statistics
Lecture 3: Random Variables and Probability Distributions

to accompany

Probability and Statistics for Engineers and Scientists

Fatih Cavdur

Random Variables

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

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**.

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**.

Random Variables

- Example 3.1: We draw 2 balls from an urn 4 red and 3 black balls without replacement. If we define a random variable X as the number of red balls, we have

Sample Space	$\boldsymbol{X}$
RR	2
RB	1
BR	1
BB	0

Random Variables

- Example 3.3: Consider a situation where components from an assembly line are classified as defective or non-defective.

$$X = \begin{cases} 1, & \text{if component is defective} \\ 0, & \text{otherwise} \end{cases}$$

Random Variables

- Statisticians use sampling plans to either accept or reject batches or lots of material. Suppose that one of those involves sampling independently 10 items from a lot of 100 items in which 12 are defective. We let X be the random variables defined as the number of defective items in the sample where $X = 0,1, \dots, 10$.

Discrete Random Variables

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

Discrete Random Variable

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

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

Discrete Random Variable

- Example 3.8: A shipment of 20 laptop computers contains 3 defectives. If a school makes a random purchase of 2 of these, find the probability distribution for the number of defectives.

Discrete Random Variable

$$f(0) = P\{X = 0\} = \frac{\binom{3}{0}\binom{17}{2}}{\binom{20}{2}} = \frac{68}{95}$$

$$f(1) = P\{X = 1\} = \frac{\binom{3}{1}\binom{17}{1}}{\binom{20}{2}} = \frac{51}{190}$$

$$f(2) = P\{X = 2\} = \frac{\binom{3}{2}\binom{17}{0}}{\binom{20}{2}} = \frac{3}{190}$$

Discrete Random Variables

We thus have the following PDF:

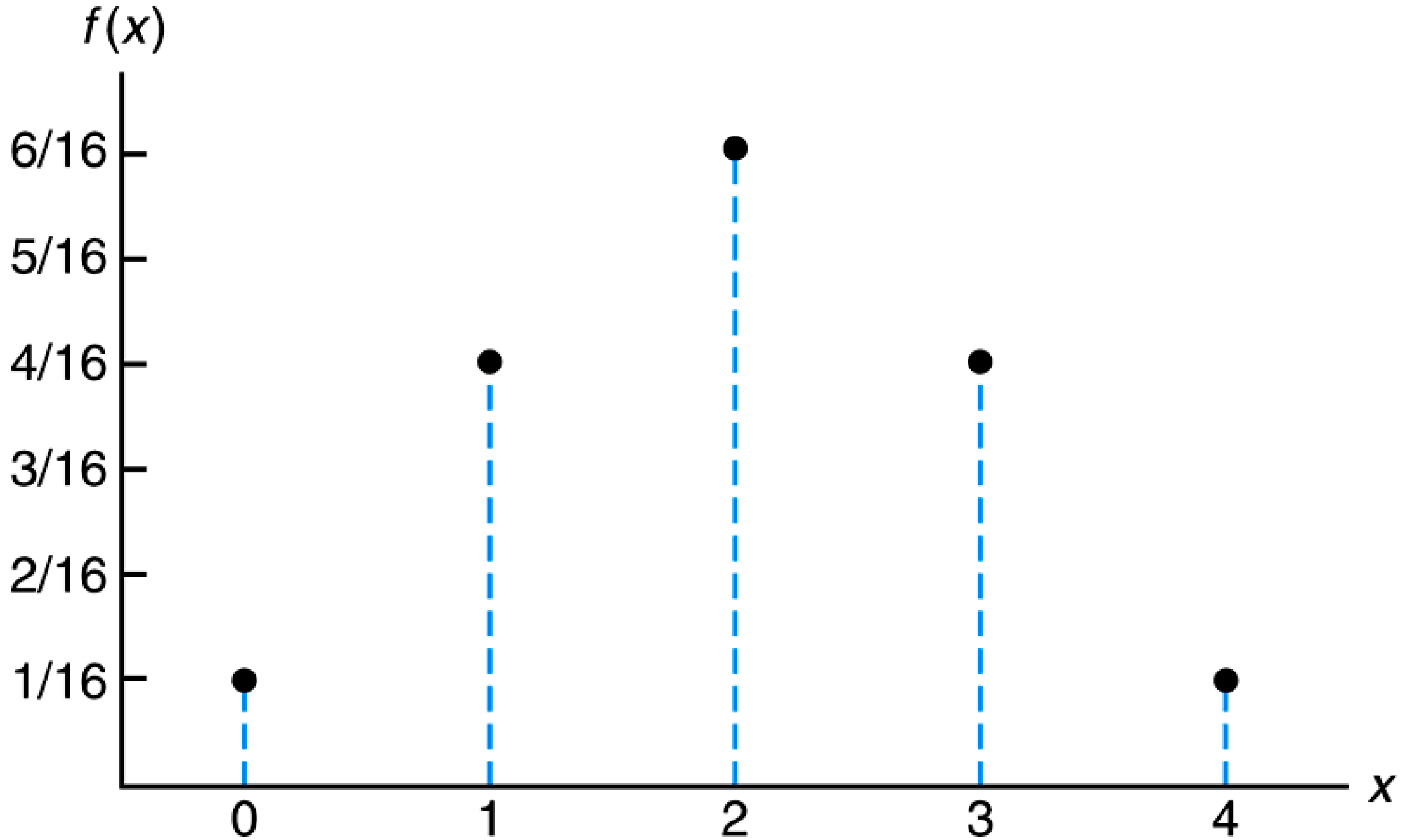
x	0	1	2
$f(x)$	$\frac{68}{95}$	$\frac{51}{190}$	$\frac{3}{190}$

Discrete Random Variables

- Example 3.9: If a car agency sells 50% of its inventory of a car model with side airbags, find a formula for the PDF of the number of cars with side airbags among the next 4 cars sold.

$$f(x) = \frac{1}{16}\binom{4}{x}, \quad x = 0,1,2,3,4$$

Discrete Random Variables (Example)



Discrete Random Variables

- Example 3.9: The CDF of the distribution is

$$F(0) = f(0) = \frac{1}{16}$$

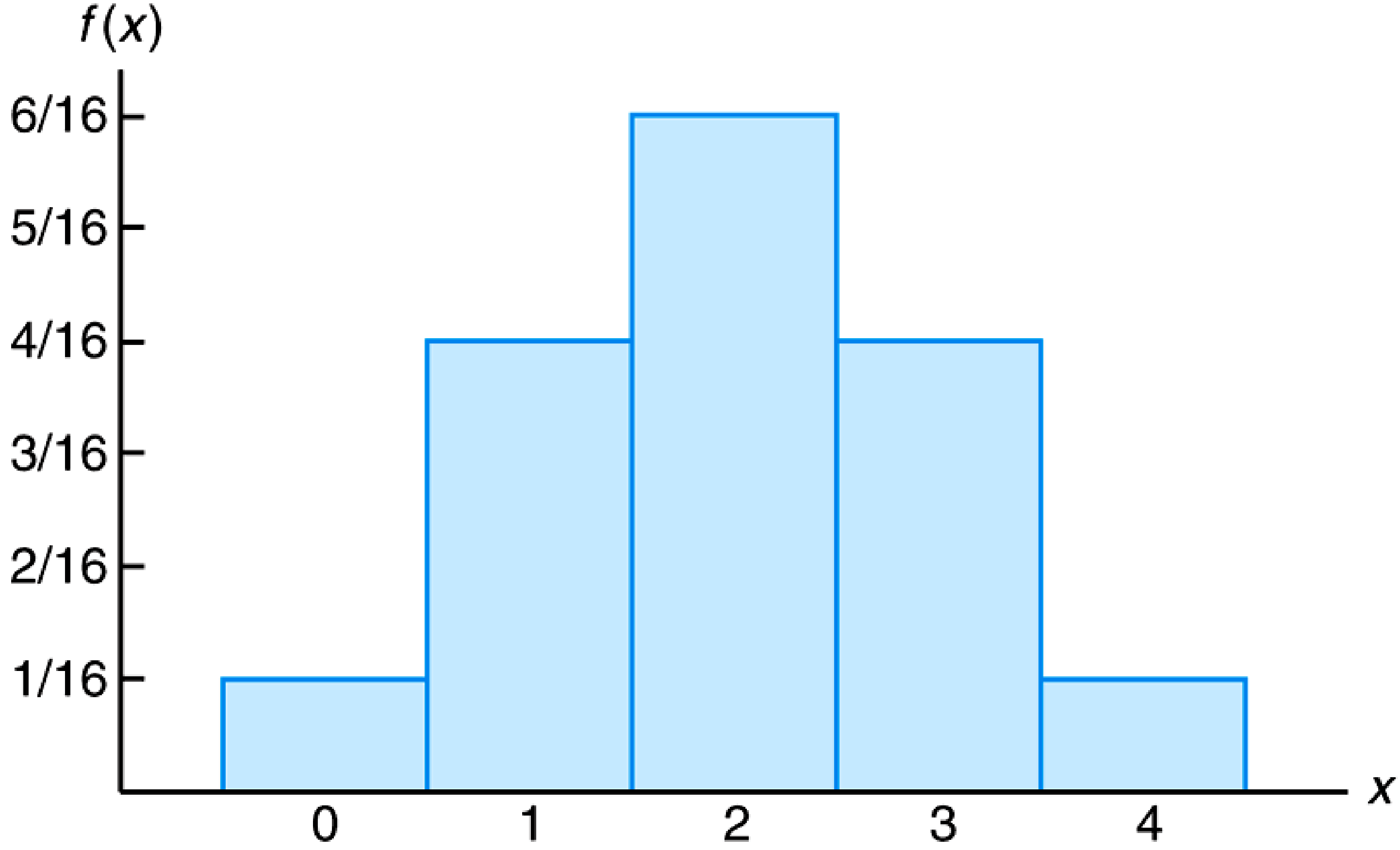
$$F(1) = f(0) + f(1) = \frac{5}{16}$$

$$F(2) = f(0) + f(1) + f(2) = \frac{11}{16}$$

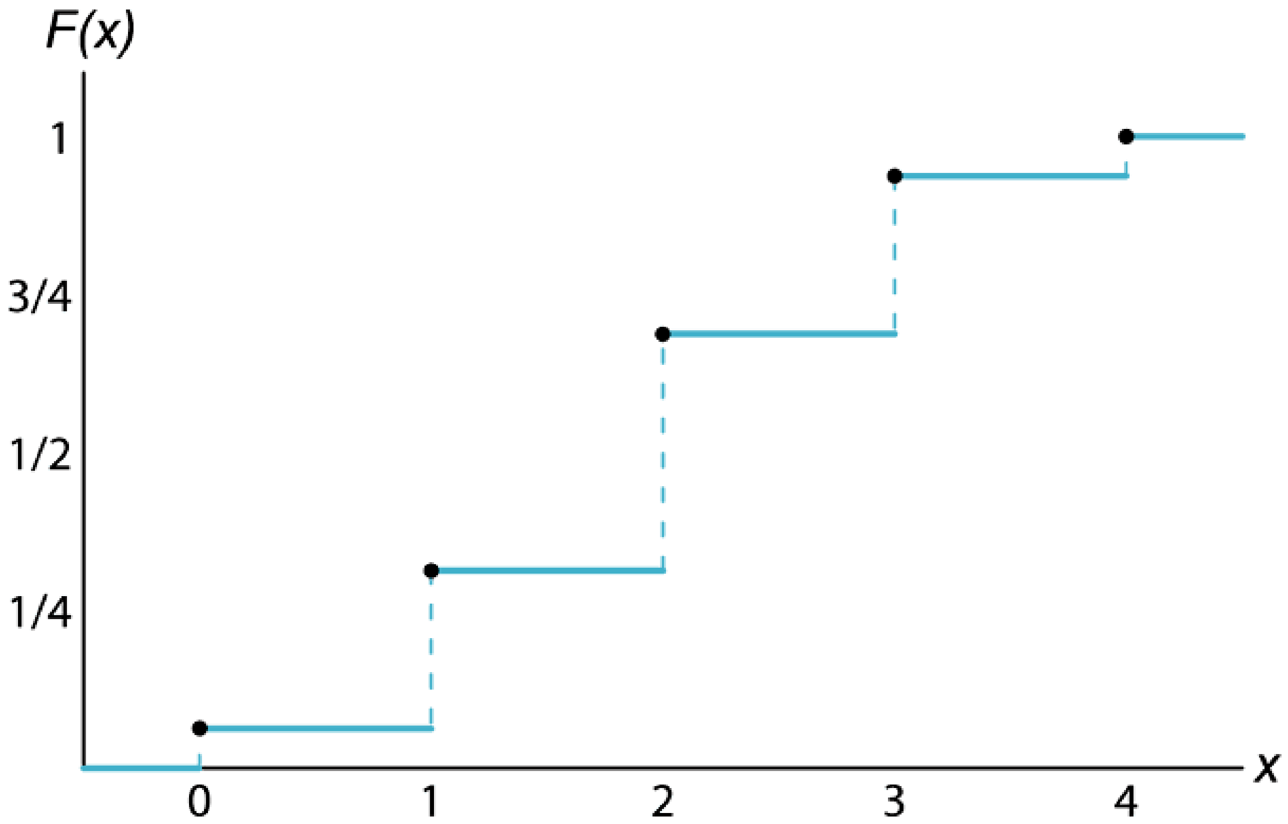
$$F(3) = f(0) + f(1) + f(2) + f(3) = \frac{15}{16}$$

$$F(4) = f(0) + f(1) + f(2) + f(3) + f(4) = 1$$

Discrete Random Variable (Example)



Discrete Random Variable (Example)



Continuous Random Variable

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

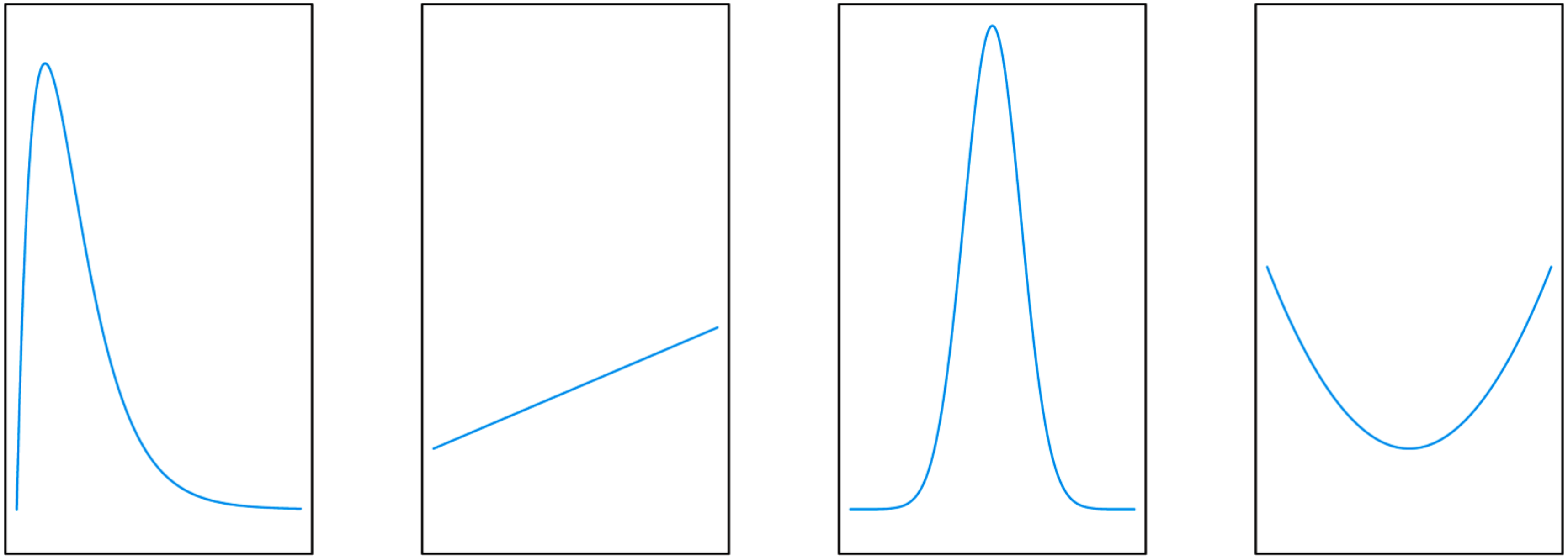
1. $f(x) \geq 0$, for all $x \in R$.
2. $\int_{-\infty}^{\infty} f(x)\, dx = 1$.
3. $P(a < X < b) = \int_a^b f(x)\, dx$.

Continuous Random Variable

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

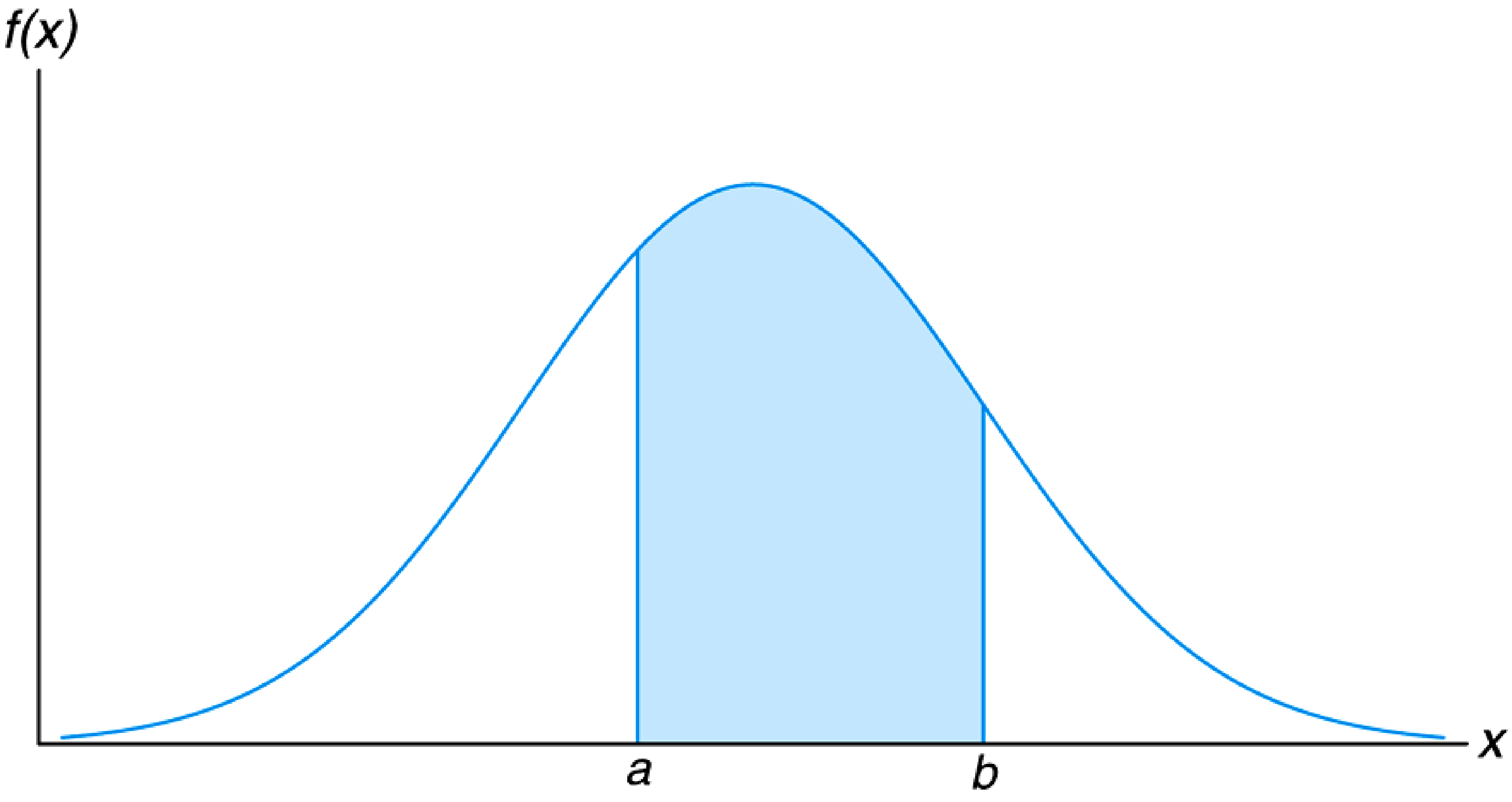
$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\, dt, \quad \text{for } -\infty < x < \infty.$$

Continuous Random Variable

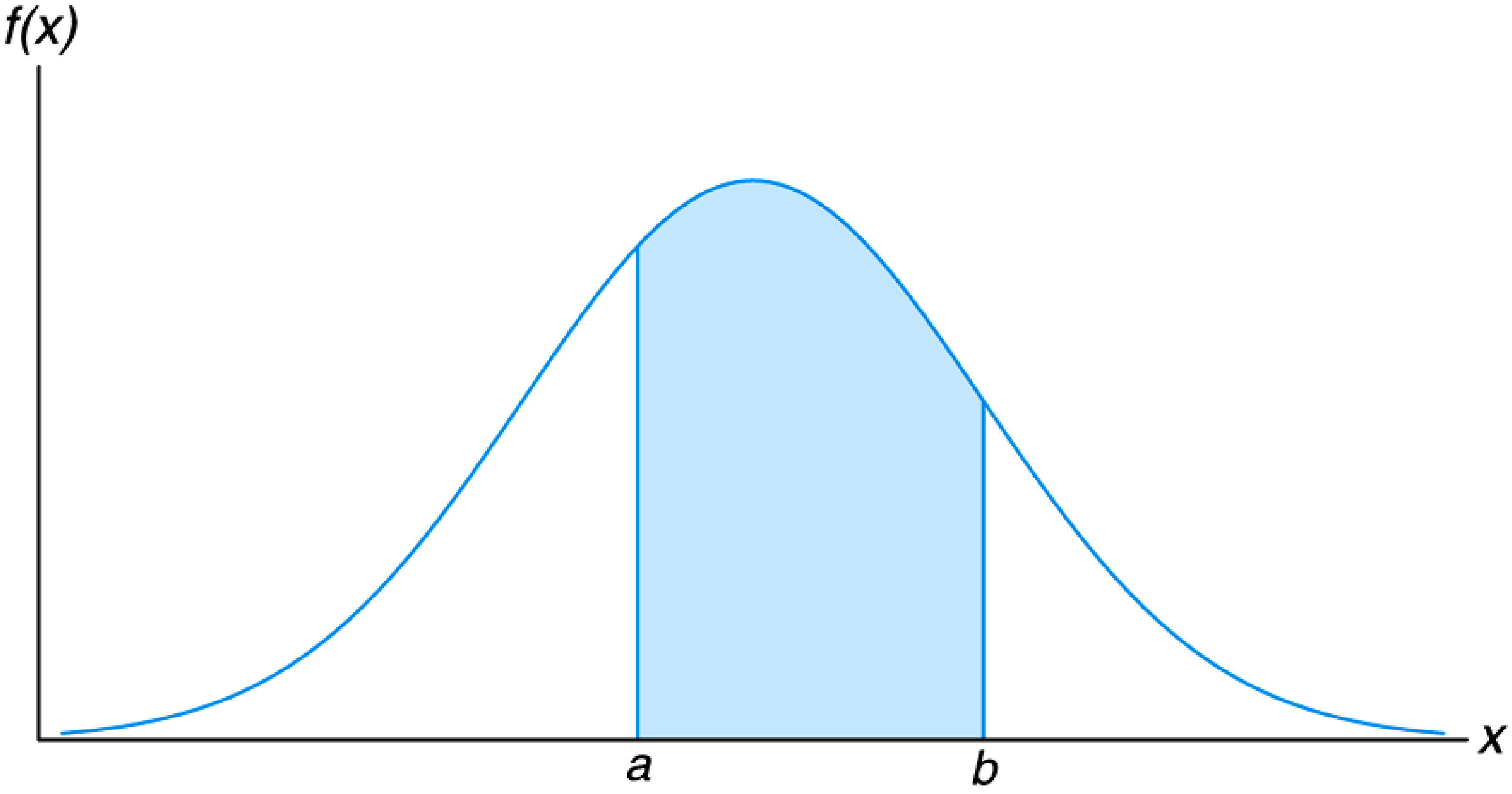


(a) (b) (c) (d)

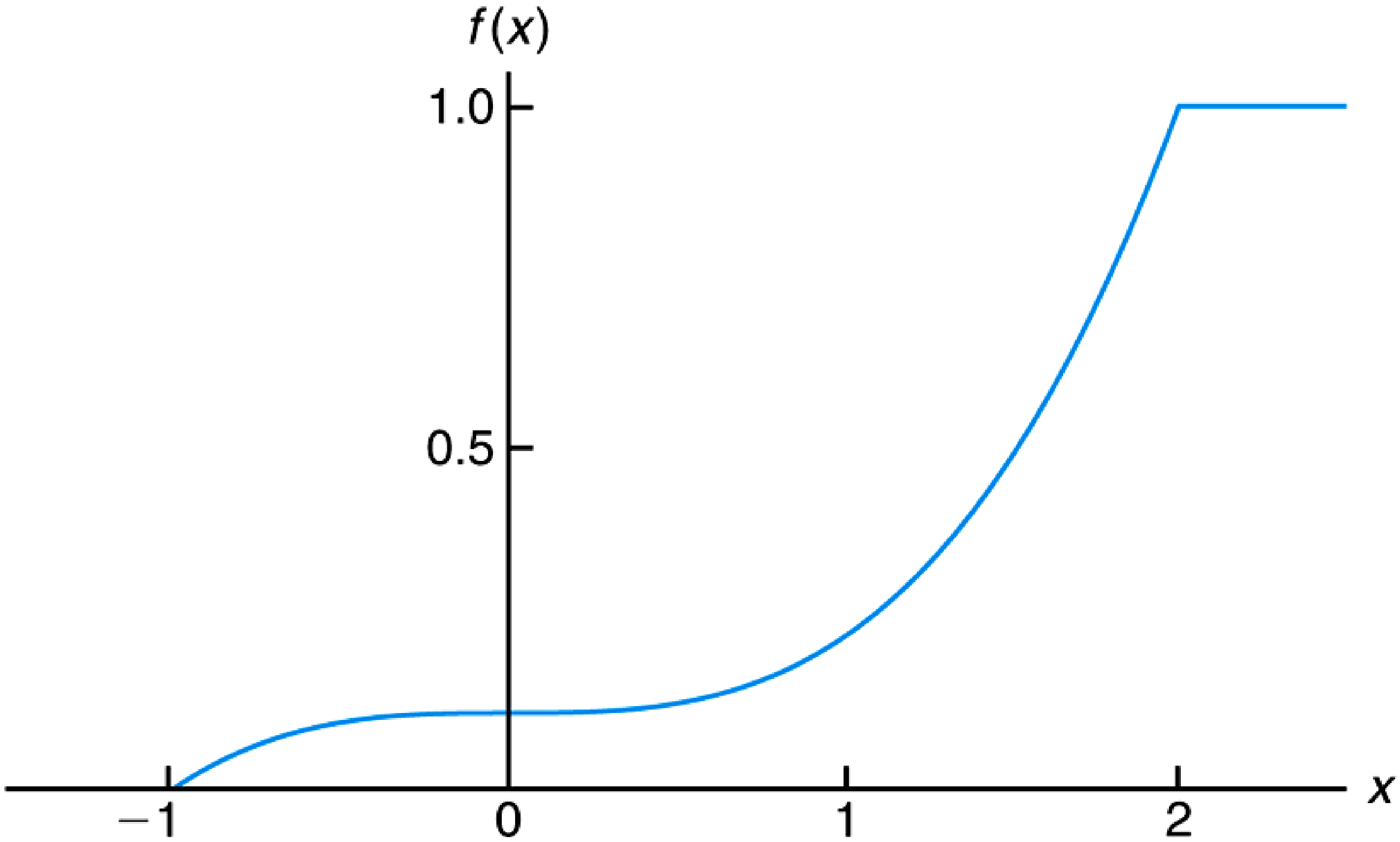
Continuous Random Variable



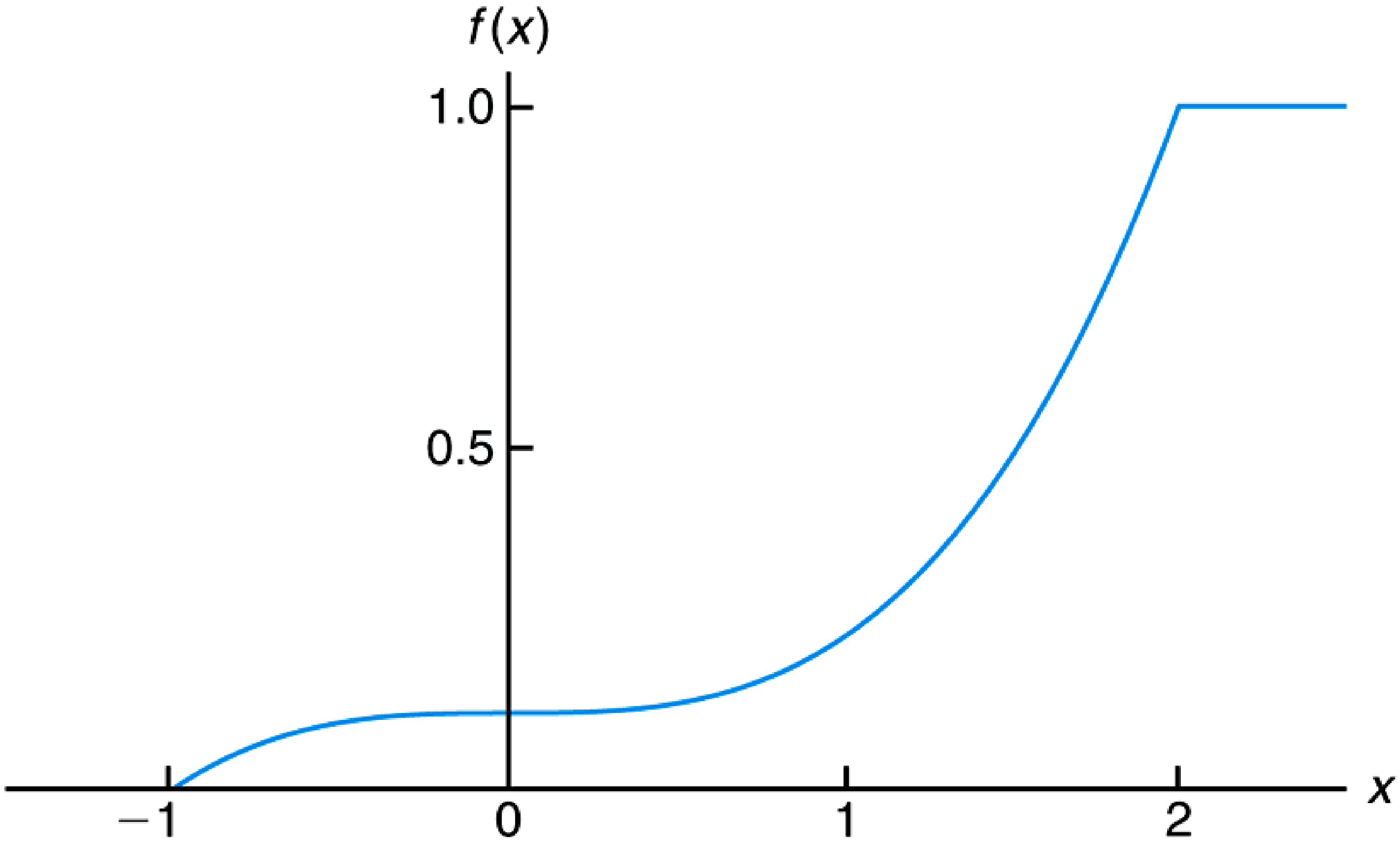
Continuous Random Variable



Continuous Random Variable



Continuous Random Variable



Joint Probability Distributions (discrete)

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

Joint Probability Distribution Example (discrete)

	$f(x,y)$	$x=0$	$x=1$	$x=2$	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

Joint Probability Distributions (continuous)

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.

Marginal Distributions

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.

Conditional Distributions

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

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

Statistical Independence

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.

Statistical Independence

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.

Statistical Independence

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.

End of Lecture

Thank you! Questions?