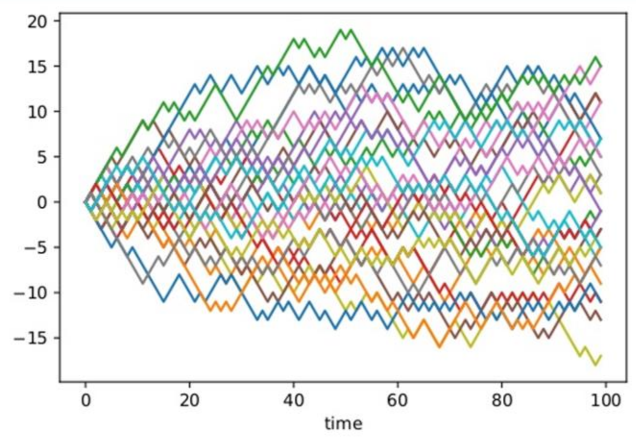
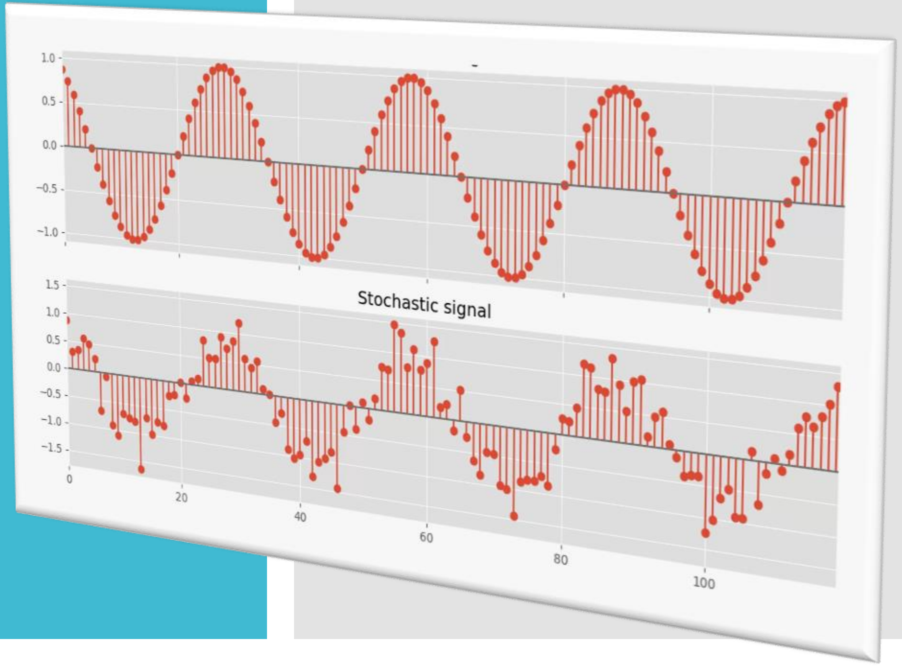



# Stochastic Processes




IS5311

Discrete Mathematics

# CONTENT

- Introduction to Stochastic Processes
- Random Point Processes
- Markov Processes
- Discrete Markov chains
- Continuous Markov Chain
- Random Walks
- Introduction to
  - Branching Processes
  - Queuing Theory
  - Finite state machines

# Introduction to Probability Theory

# Introduction to Probability Theory

- Experiments, Sample Spaces, and Events
- Axioms of Probability
- Assigning Probabilities
- Joint and Conditional Probabilities
- Bayes' Theorem
- Independence
- Discrete Random Variables (Bernoulli, Binomial, Poisson)

# Introduction to Probability Theory

## Experiments, Sample Spaces, and Events

- **Definition: Experiment**
- An experiment is a procedure we perform (quite often hypothetical) that produces some result. Often the letter E is used to designate an experiment
- (e.g., the experiment E5 might consist of tossing a coin five times).

- **Definition: Outcome**
- An outcome is a possible result of an experiment.
- The Greek letter xi ($\xi$ ) is often used to represent outcomes (e.g., the outcome $\xi 1$ of experiment

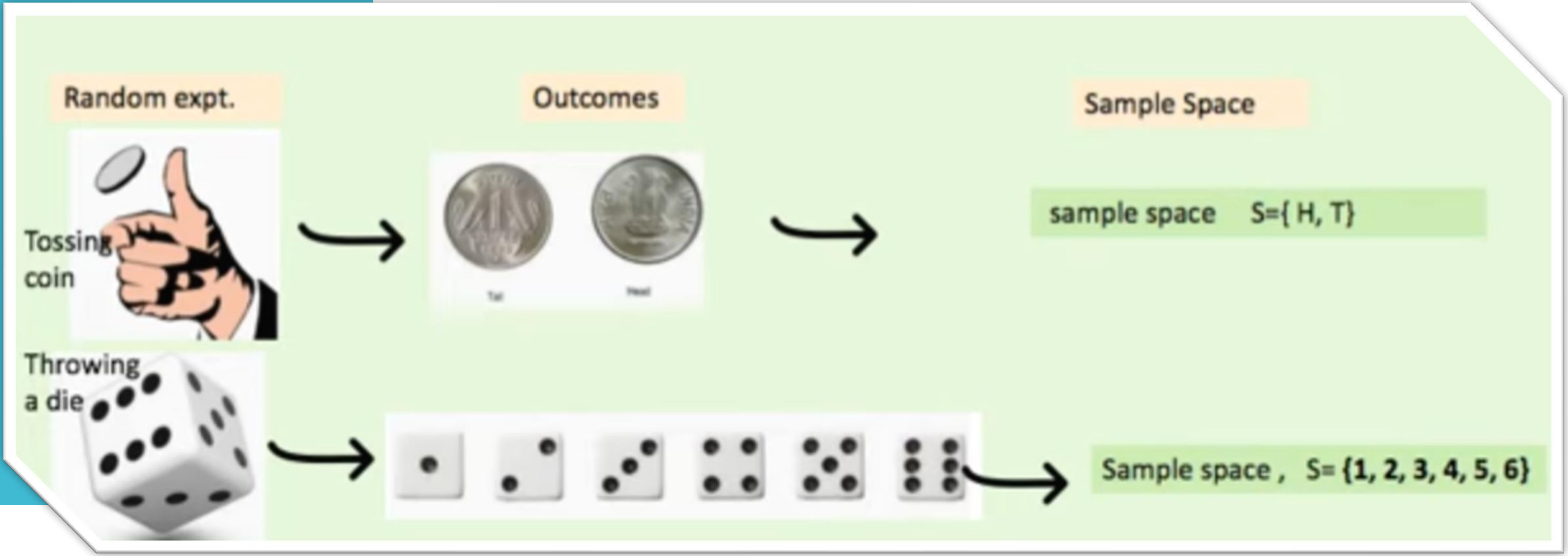
- **Definition: Event**
- An event is a certain set of outcomes of an experiment (e.g., the event C associated with experiment E5 might be C = {all outcomes consisting of an even number of heads}).

# Introduction to Probability Theory

# Experiments, Sample Spaces, and Events


- **Definition: Sample Space**
- The sample space is the collection or set of "all possible" distinct (collectively exhaustive and mutually exclusive) outcomes of an experiment.
- The letter S is used to designate the sample space, which is the universal set of outcomes of an experiment.
- A sample space is called **discrete** if it is a finite or a countably infinite set.
- It is called **continuous** or a continuum otherwise.

# Experiments, Sample Spaces, and Events



# Probability

- probability that it is a measure of the likelihood of various events. So, in general terms, probability is a function of an event that produces a numerical quantity that measures the likelihood of that event.
- **AXIOM 1:** For any event A, Pr (A) ≥ 0 (a negative probability does not make sense).
- **AXIOM 2:** If S is the sample space for a given experiment, Pr (S) = 1 (probabilities are normalized so that the maximum value is unity).
- **AXIOM 3a:** If $A \cap B = \varnothing$, then $\Pr(A \cup B) = \Pr(A) + \Pr(B)$.
- **COROLLARY 1:** Consider M sets $A_1, A_2, \ldots, AM$ that are mutually exclusive, $A_i \cap A_j = \varnothing$ for all $i \neq j$,

$$\Pr\left(\bigcup_{i=1}^{M} A_i\right) = \sum_{i=1}^{M} \Pr(A_i).$$

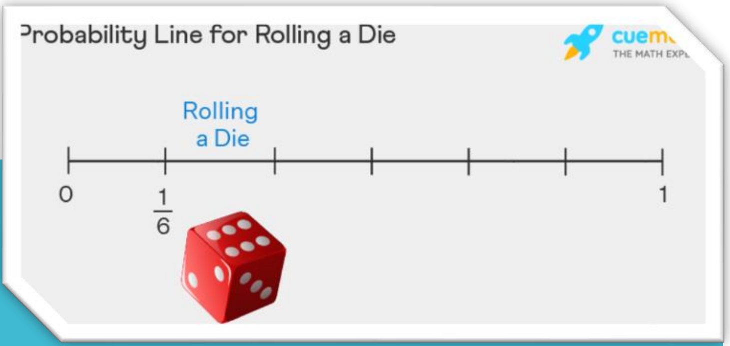
- **AXIOM 3b:** For an infinite number of mutually exclusive sets, $A_i$, i =
- 1, 2, 3, . . . , $A_i \cap A_j = \varnothing$ for all $i \neq j$,

$$\Pr\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \Pr(A_i).$$

- THEOREM 1: For any sets A and B (not necessarily mutually exclusive), $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$.

- THEOREM 2: $\Pr(A) = 1 - \Pr(A)$.

- THEOREM 3: If $A \subset B$, then $\Pr(A) \leq \Pr(B)$.




# Assigning Probabilities

# Classical Approach


## Atomic Outcomes

- In many experiments, it is possible to specify all of the outcomes of the experiment in terms of some fundamental outcomes, which we refer to as atomic outcomes.
- The most basic events that cannot be decomposed into simpler events.
- M atomic outcomes of an experiment E as $\xi_1, \xi_2, \cdots, \xi_M$. These atomic events are taken to be mutually exclusive and exhaustive.
- That is, $\xi_i \cap \xi_j = \varnothing$ for all $i \neq j$, and $\xi_1 \cup \xi_2 \cup \cdots \cup \xi_M = S$. Then by Corollary 2.1 and Axiom 2.2,

$$\Pr(\xi_1 \cup \xi_2 \cup \cdots \cup \xi_M) = \Pr(\xi_1) + \Pr(\xi_2) + \cdots + \Pr(\xi_M) = \Pr(S) = 1$$

- If each atomic outcome is to be **equally probable**, then assigning each a probability of **$\Pr(\xi i) = 1/M$** for there to be equality in the preceding equation.
- This approach to assigning probabilities is referred to as the classical approach.

# Issues with Classical Approach when Assigning Probabilities

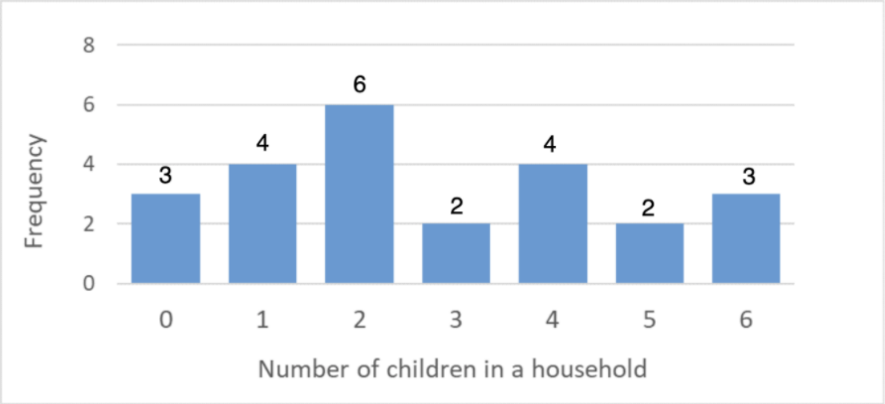
- **Example of rolling two fair dice**
- Defining the set of atomic outcomes **incorrectly** as the different sums that can occur on the two die faces. If we assign an equally likely probability to each of these outcomes, then we arrive at the assignment
- $\Pr(\text{sum} = 2) = \Pr(\text{sum} = 3) = \cdots = \Pr(\text{sum} = 12) = 1/11.$
- The problem here is that the atomic events we have assigned are not the most basic outcomes and can be decomposed into simpler outcomes.

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| 6 | 7 | 8 | 9 | 10 | 11 | 12 |

- **Example of measuring the height of an arbitrarily chosen student**
- Suppose we consider an experiment that consists of measuring the height of an arbitrarily chosen student in your class and rounding that measurement to the nearest inch.
- The atomic outcomes of this experiment would consist of all the heights of the students in your class.
- However, it would not be reasonable to assign an equal probability to each height.
- Those heights corresponding to very tall or very short students would be expected to be less probable than those heights corresponding to a medium height.
- So, how then do we assign probabilities to these events? The problems associated with the classical approach to assigning probabilities can be overcome by using the relative frequency approach.

# Assigning Probabilities

# Relative Frequency Approach




- The relative frequency approach requires that the experiment we are concerned with be repeatable, in which case, the probability of an event, A, can be assigned by repeating the experiment a large number of times and observing how many times the event A actually occurs.

- If we let n be the number of times the experiment is repeated and $n_A$ be the number of times the event A is observed, then the probability of the event A can be assigned according to

$$\boxed{\Pr(A) = \lim_{n \to \infty} \frac{n_A}{n}.}$$

# Example

## Relative Frequency Approach

&

## The Law of Large Numbers

Imagine rolling a die 12 times and writing down how many fives you get. The relative frequency is the ratio of how many fives you got (number of favorable outcomes), to how many times you rolled the die (number of trials). If you got three fives, the relative frequency is

$$\frac{3}{12} = \frac{1}{4} = 0.25$$

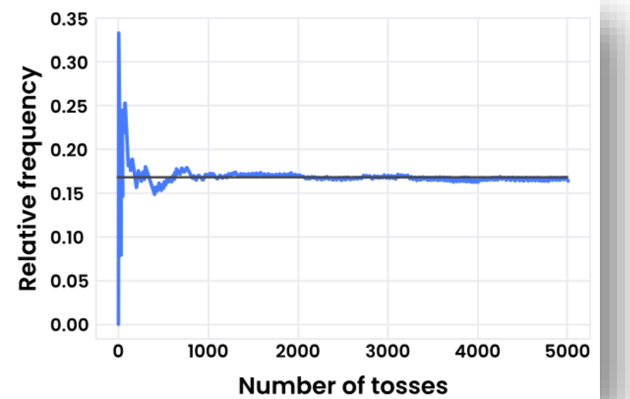
- **Note!** The relative frequency approaches the probability when the number of trials becomes very large!

**The graph shows that after approximately 1500 throws, the blue relative frequency has stabilized very close to the actual probability in black.**

**Rule**

**The Law of Large Numbers**

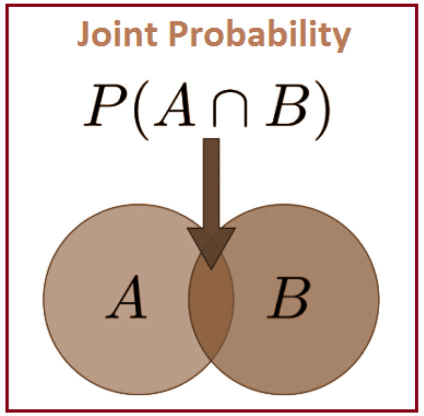
If you perform an experiment with a large number of trials, the results of the experiment will approach the expected value of the experiment.



# Joint and Conditional Probabilities

## Joint Probability

- The joint probability refers to a statistical measure that calculates the likelihood of two events occurring together at the same point in time.




- **The joint probability of the sets A and B, $\Pr(A \cap B)$ or use the notation $\Pr(A, B)$.**

- This definition and notation extends to an arbitrary number of sets. The joint probability of the sets $A_1, A_2, \ldots, A_M$, is

$$\Pr(A_1 \cap A_2 \cap \cdots \cap A_M) \quad \text{or} \quad \Pr(A_1, A_2, \ldots, A_M)$$

# How to Compute Joint Probability?

**Axiom 2.3a & Theorem 2.1 :**

- if A and B are mutually exclusive, then their joint probability is zero.
- i.e. $\mathbf{Pr(A \cap B) = Pr(A, B) = Pr(\varnothing)}$

- ***In the general case when A and B are not necessarily mutually exclusive, calculating the joint probability of A and B***
- Using the classical approach, Both events (sets) A and B can be expressed in terms of atomic outcomes.
- We then write A ∩ B as the set of those atomic outcomes that is common to both and calculate the probabilities of each of these outcomes.
- Alternatively, we can use the relative frequency approach. Let $n_{A,B}$ be the number of times that A and B simultaneously occur in n trials. Then,

$$\Pr(A, B) = \lim_{n \to \infty} \frac{n_{A,B}}{n}.$$

# Conditional Probability

- If the probability of event A depends on the occurrence of event C, i.e. the probability of A is conditional on C, and the probability of A given knowledge that the event C has occurred is referred to as the conditional probability of A given C.
- The notation Pr(A | C) is used to denote the probability of the event A given that the event C has occurred, or simply the probability of A given C.

**Definition**

- For two events A and B, the probability of A conditioned on knowing that B has occurred is

$$\boxed{\Pr(A|B) = \frac{\Pr(A,B)}{\Pr(B)}.}$$

- this formula offers a convenient way to compute joint probabilities:

$$\boxed{\Pr(A,B) = \Pr(B|A)\Pr(A) = \Pr(A|B)\Pr(B).}$$

- This idea can be extended to more than two events. Consider finding the joint probability of three events, A, B, and C:

$$\Pr(A, B, C) = \Pr(C|A, B)\Pr(A, B) = \Pr(C|A, B)\Pr(B|A)\Pr(A).$$

- In general, for M events, $A_1, A_2, \ldots, A_M$

$$\Pr(A_1, A_2, \ldots, A_M) = \Pr(A_M | A_1, A_2, \ldots, A_{M-1})\Pr(A_{M-1} | A_1, \ldots, A_{M-2}) \cdots \times \Pr(A_2 | A_1)\Pr(A_1).$$

# Conditional Probability & The Theorem of Total Probability

- THEOREM: For any events A and B such that $\Pr(B) \neq 0$,

$$\Pr(A|B) = \frac{\Pr(B|A)\Pr(A)}{\Pr(B)}.$$

- THEOREM: (Theorem of Total Probability)
- Let B1 , B2 , . . . , B n be a set of mutually exclusive and exhaustive events. That is, Bi ∩ B j = Ø for all $i \neq j$ and

$$\bigcup_{i=1}^{n} B_i = S \Rightarrow \sum_{i=1}^{n} \Pr(B_i) = 1.$$

- Then

$$\Pr(A) = \sum_{i=1}^{n} \Pr(A|B_i)\Pr(B_i)$$

# The Theorem of Total Probability

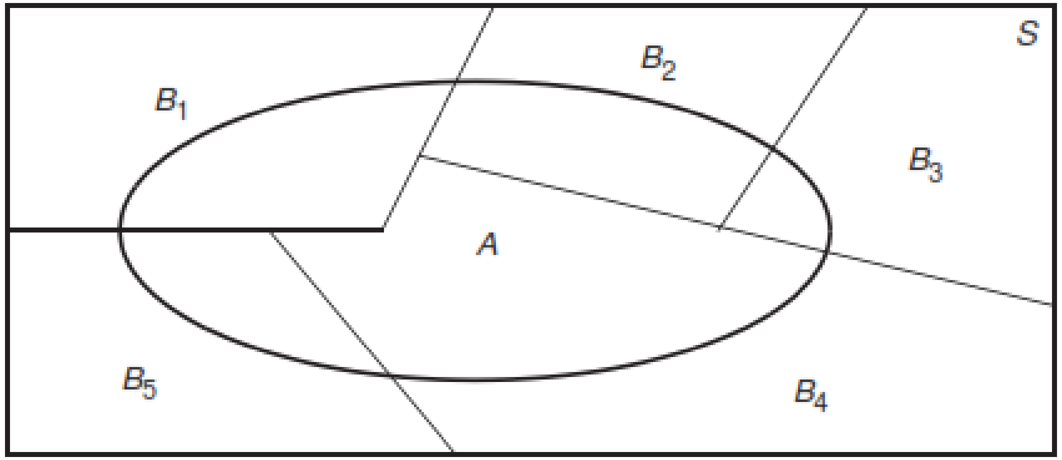
## Proof

- From the Venn diagram, it can be seen that the event A can be written as

$$A = \{A \cap B_1\} \cup \{A \cap B_2\} \cup \cdots \cup \{A \cap B_n\}$$
$$\Rightarrow \Pr(A) = \Pr(\{A \cap B_1\} \cup \{A \cap B_2\} \cup \cdots \cup \{A \cap B_n\})$$

- Also, since the $B_i$ are all mutually exclusive, then the $\{A \cap B_i\}$ are also mutually exclusive, so that

$$\Pr(A) = \sum_{i=1}^{n} \Pr(A, B_i) \qquad \text{(by Corollary 2.3)},$$
$$= \sum_{i=1}^{n} \Pr(A|B_i)\Pr(B_i) \quad \text{(by Theorem 2.4)}.$$

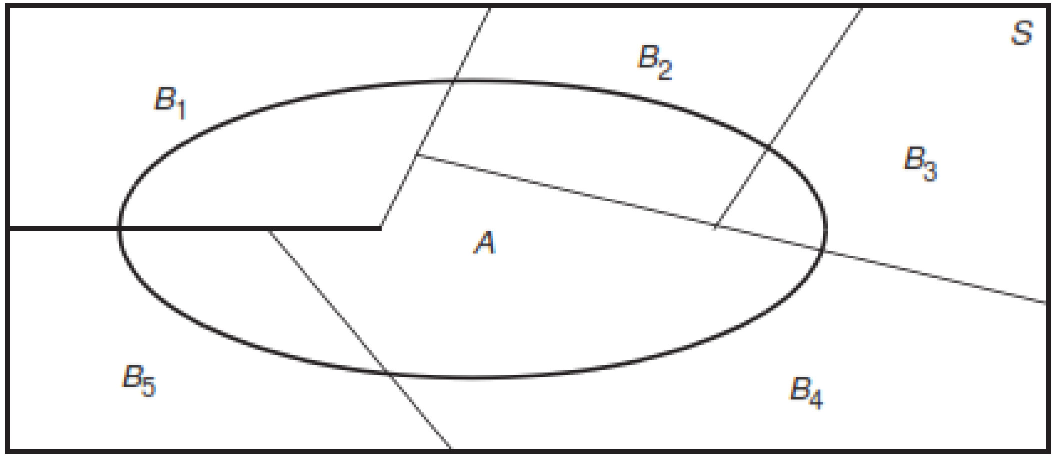


# The Theorem of Total Probability
Proof

As with Theorem 2.1, a Venn diagram (shown in Figure 2.2) is used here to aid in the visualization of our result. From the diagram, it can be seen that the event $A$ can be written as

$$A = \{A \cap B_1\} \cup \{A \cap B_2\} \cup \cdots \cup \{A \cap B_n\}$$
$$\Rightarrow \Pr(A) = \Pr(\{A \cap B_1\} \cup \{A \cap B_2\} \cup \cdots \cup \{A \cap B_n\})$$

Also, since the $B_i$ are all mutually exclusive, then the $\{A \cap B_i\}$ are also mutually exclusive, so that




$$\Pr(A) = \sum_{i=1}^{n} \Pr(A, B_i) \qquad \text{(by Corollary 2.3)},$$
$$= \sum_{i=1}^{n} \Pr(A|B_i)\Pr(B_i) \qquad \text{(by Theorem 2.4)}.$$

# Bayes's Theorem

- Bayes' Theorem can be obtained by combining Theorems of Conditional probability and Total Probability.

- **Bayes' Theorem**
- Let $B_1$ , $B_2$ , . . . , $B_n$ be a set of mutually exclusive and exhaustive events. Then

$$\Pr(B_i|A) = \frac{\Pr(A|B_i)\Pr(B_i)}{\sum_{i=1}^{n} \Pr(A|B_i)\Pr(B_i)}.$$

- $\Pr(B_i)$ is often referred to as the a **priori probability** of event $B_i$,

- while $\Pr(B_i \mid A)$ is known as the a **posteriori probability** of event $B_i$ given A.

# Bayes' Theorem

# Example

- A certain auditorium has 30 rows of seats. Row 1 has 11 seats, while Row 2 has 12 seats, Row 3 has 13 seats, and so on to the back of the auditorium where Row 30 has 40 seats. A door prize is to be given away by randomly selecting a row (with equal probability of selecting any of the 30 rows) and then randomly selecting a seat within that row (with each seat in the row equally likely to be selected). Find the probability that Seat 15 was selected given that Row 20 was selected
- Find the probability that Row 20 was selected given that Seat 15 was selected.

$$\Pr(\text{Seat } 15) = \sum_{k=5}^{30} \frac{1}{k+10}\frac{1}{30} = 0.0342.$$

$$\Pr(\text{Row } 20|\text{Seat } 15) = \frac{\frac{1}{30}\frac{1}{30}}{0.0342} = 0.0325.$$

# Independence

- Mathematically, two events A and B are independent if $\Pr(A \mid B) = \Pr(A)$.
- That is, the a priori probability of event A is identical to the a posteriori probability of A given B.
- Note that if $\Pr(A \mid B) = \Pr(A)$, then the following two conditions also hold
  - $\Pr(B \mid A) = \Pr(B)$
  - $\Pr(A, B) = \Pr(A)\Pr(B)$

- Furthermore, if $\Pr(A|B) \neq \Pr(A)$ then the other two conditions also do not hold. We can thereby conclude that any of these three conditions can be used as a test for independence and the other two forms must follow.

Definition:

- Two events are statistically independent if and only if

$$\Pr(A, B) = \Pr(A)\Pr(B)$$

# Independence…

- Definition
- The events A, B, and C are mutually independent if each pair
- of events is independent; that is,
  - $Pr(A, B) = Pr(A)Pr(B)$,
  - $Pr(A, C) = Pr(A)Pr(C)$,
  - $Pr(B, C) = Pr(B)Pr(C)$
- In addition:
  - $Pr(A, B, C) = Pr(A)Pr(B)Pr(C)$.
- Definition
- The events $A_1, A_2, \ldots, A_n$ are independent if any subset of $k <$
- $n$ of these events are independent, and in addition

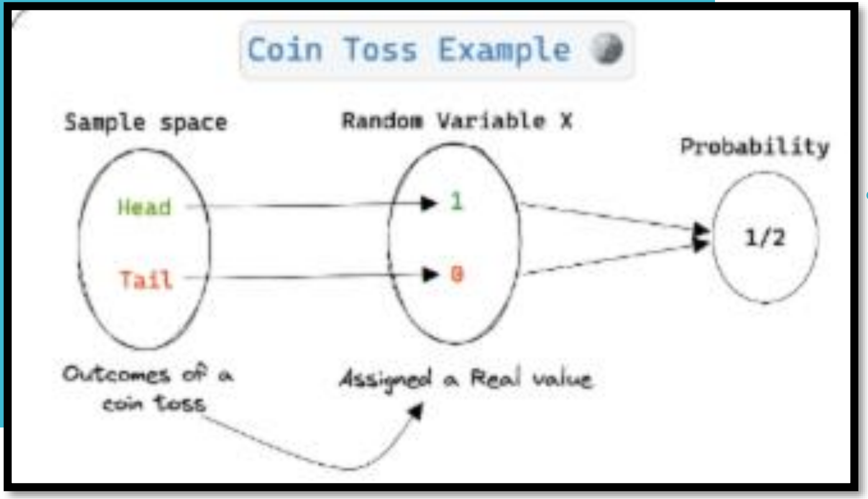
$$Pr(A_1, A_2, \ldots, A_n) = Pr(A_1)Pr(A_2) \ldots Pr(A_n).$$

# Random Variables and Probability Distributions

- Random Variables, Distributions, and Density Functions
- Cumulative Distribution Function
- Probability Density Function
- Gaussian (Normal) Random Variable
- Other Random Variables (Uniform, Exponential, Gamma, Chi-Squared)
- Conditional Distribution and Density Functions

# Random Variable

- A random variable (rv) is a real-valued function of the elements of a sample space, $S$.
- Given an experiment, $E$, with sample space, $S$, the random variable $X$ maps each possible outcome, $\xi \in S$, to a real number $X(\xi)$ as specified by some rule.
- **Discrete rv:** If the mapping $X(\xi)$ is such that the random variable $X$ takes on a finite or countably infinite number of values, then we refer to $X$ as a discrete random variable.
- **Continuous rv:** If the range of $X(\xi)$ is an uncountably infinite number of points, we refer to $X$ as a continuous random variable.



# Probability Distribution

- **Probability Distribution**

A probability distribution is a statistical function that describes all the possible values and likelihoods that a random variable can take within a given range.

- This range will be bounded between the **minimum** and **maximum** possible values.

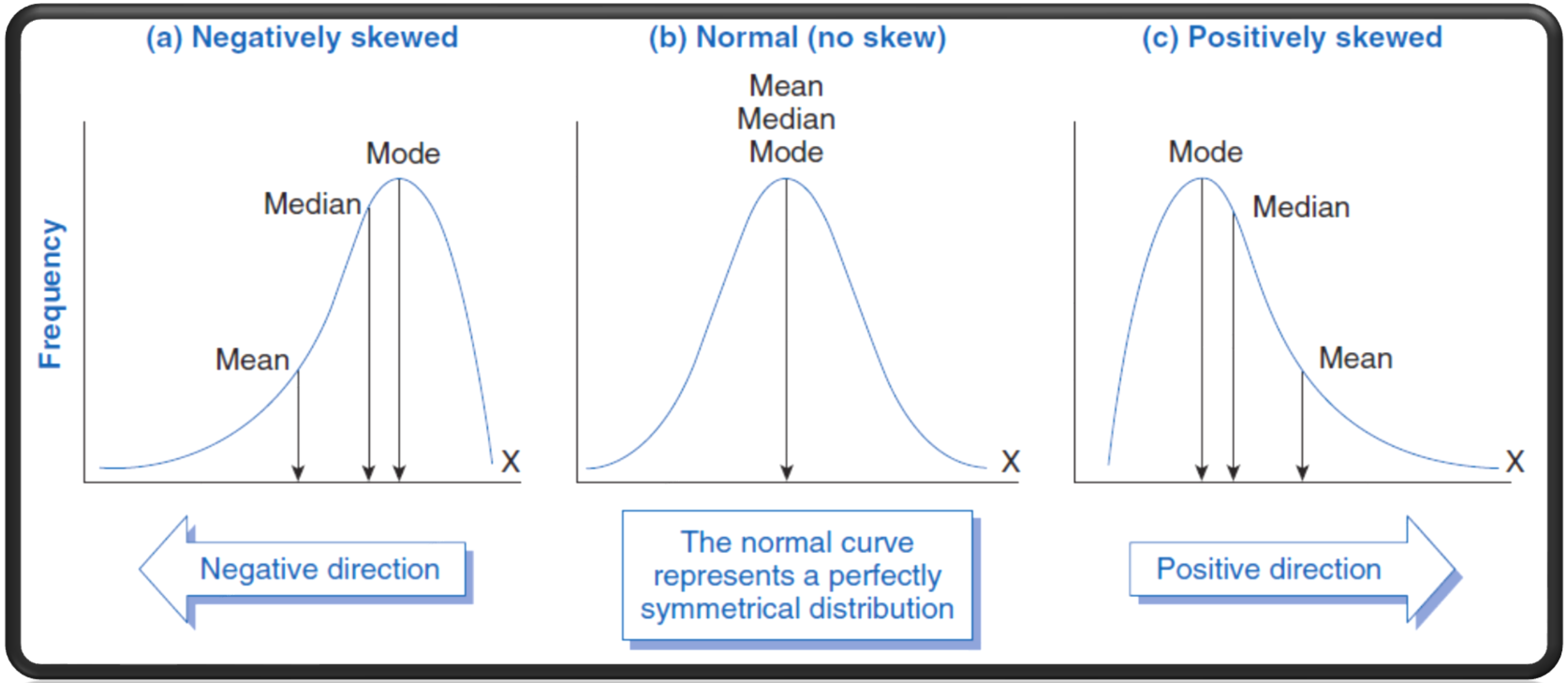
- The possible value is likely to be plotted on the probability distribution depends on several factors: (Four Moments of Distribution)
  - distribution's mean (average),
  - standard deviation,
  - skewness, (skewness will be unit-free)
    - measure of the asymmetry of the probability distribution of a real-valued random variable about its mean
  - Kurtosis (kurtosis will be unit-free)
    - Kurtosis measures the peakedness or flatness of the data
    - There are three kurtosis categories: mesokurtic (normal), platykurtic (less than normal), and leptokurtic (more than normal).

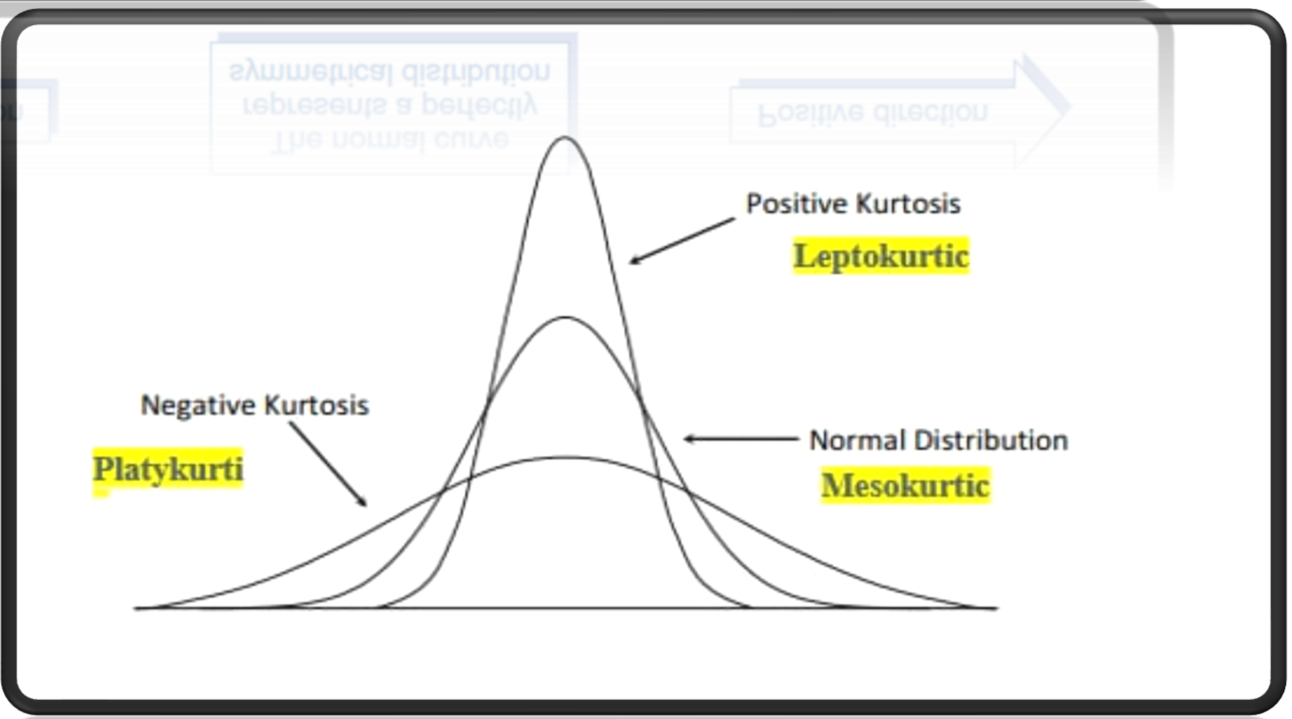
| Moment number | Name | Measure of | Formula |
|---|---|---|---|
| 1 | Mean | Central tendency | $\bar{X} = \frac{\sum_{i=1}^{N} X_i}{N}$ |
| 2 | Variance (Volatility) | Dispersion | $\sigma^2 = \frac{\sum_{i=1}^{N}(X_i - \bar{X})^2}{N}$ |
| 3 | Skewness | Symmetry (Positive or Negative) | $Skew = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{(X_i - \bar{X})}{\sigma}\right]^3$ |
| 4 | Kurtosis | Shape (Tall or flat) | $Kurt = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{(X_i - \bar{X})}{\sigma}\right]^4$ |

Where X is a random variable having N observations (i = 1,2,…,N).

# Skewness

# Kurtosis



## Probability Mass Function

## Probability Density Function

- **Probability Mass Function (PMF):** A PMF, denoted by P(X = x), describes the probability of a discrete random variable X taking on a specific value x.
  - It's a step function, where the probability mass is accumulated at each discrete value.

- **Probability Density Function (PDF):** A PDF, denoted by f(x), describes the probability density of a continuous random variable X falling within a particular interval.
  - It's a continuous function, where the probability density is spread across the range of values.

- The probability density function (PDF) of the random variable $X$ evaluated at the point $x$ is

$$f_X(x) = \lim_{\varepsilon \to 0} \frac{\Pr(x \le X < x + \varepsilon)}{\varepsilon}.$$

- As the name implies, the probability density function is the probability that the random variable X lies in an infinitesimal interval about the point X = x, normalized by the length of the interval.

# Relationship between PDF & CDF

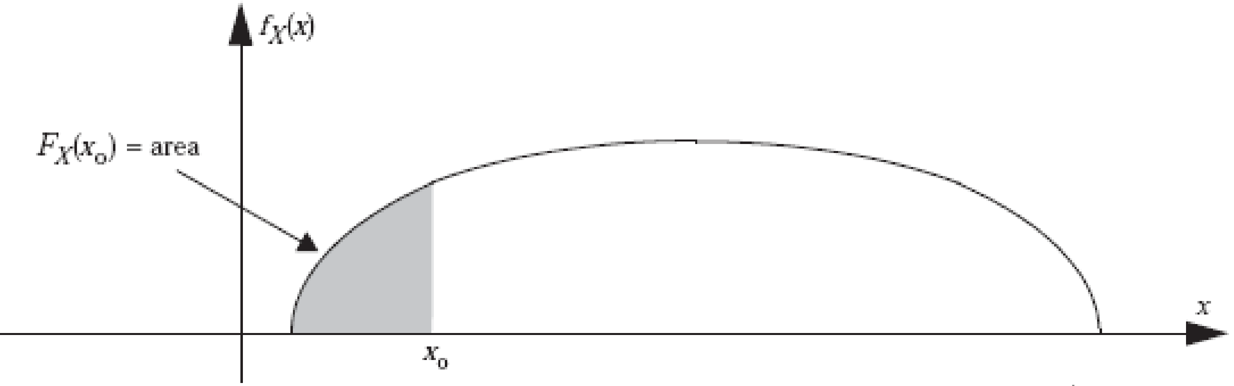
- The probability of a random variable falling in an interval can be written in terms of its CDF.

$$\Pr(x \le X < x+\varepsilon) = F_X(x+\varepsilon) - F_X(x)$$

- For continuous random variables,
- So that

$$f_X(x) = \lim_{\varepsilon \to 0} \frac{F_X(x+\varepsilon) - F_X(x)}{\varepsilon} = \frac{dF_X(x)}{dx}.$$

- Hence, it is seen that the PDF of a random variable is the derivative of its CDF. Conversely, the CDF of a random variable can be expressed as the integral of its PDF.
- It is apparent that the PDF is a nonnegative function, although it is not restricted to be less than unity as with the CDF.
- Some properties of PDFs are




(1) $f_X(x) \ge 0$;

(2) $f_X(x) = \frac{dF_X(x)}{dx}$;

(3) $F_X(x) = \int_{-\infty}^{x} f_X(y)\,dy$;

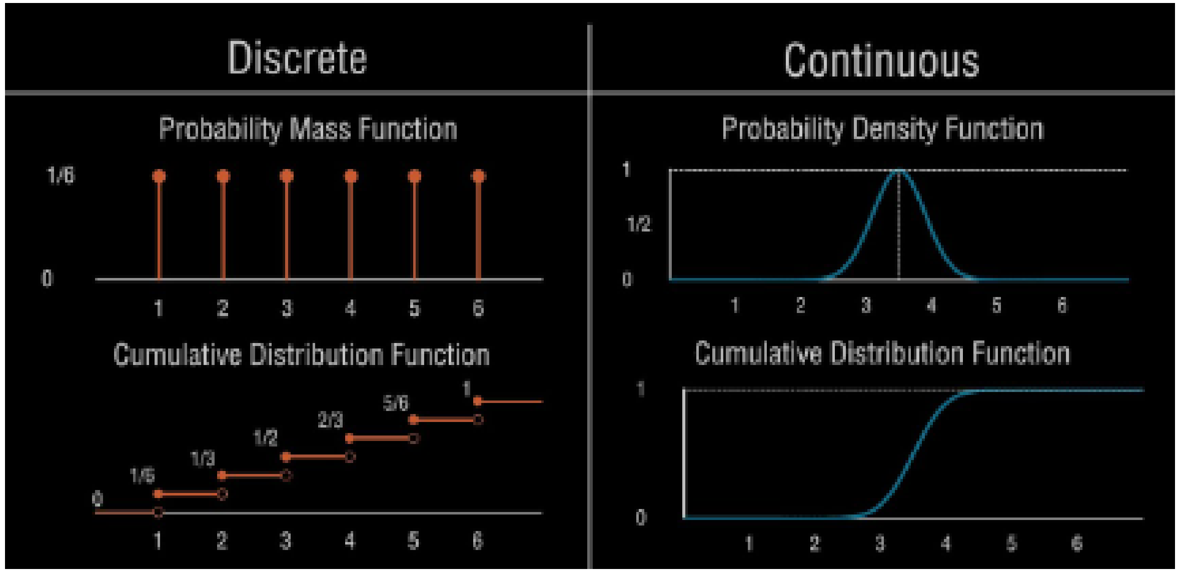
(4) $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$;

(5) $\int_{a}^{b} f_X(x)\,dx = \Pr(a < X \le b)$.

# Cumulative Distribution Function

- **Cumulative Distribution Function:** It is defined as the probability that a random variable X takes on a value less than or equal to x. In other words, it represents the area under the probability density function (PDF) of X up to x.
- **The cumulative distribution function (CDF) of random variable X is defined as**

$$F_X(x) = P(X \leq x), \text{ for all } x \in \mathbb{R}.$$



# Cumulative Distribution Function

- These properties of cumulative distribution functions

  (1) $F_X(-\infty) = 0, F_X(\infty) = 1,$

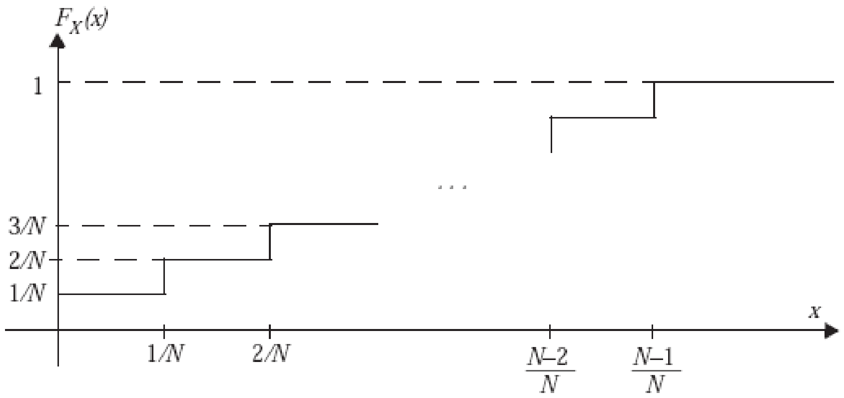
  (2) $0 \le F_X(x) \le 1,$

  (3) For $x_1 < x_2, F_X(x_1) \le F_X(x_2),$

  (4) For $x_1 < x_2, \Pr(x_1 < X \le x_2) = F_X(x_2) - F_X(x_1).$

- The functional form of this CDF is

$$F_X(x) = \begin{cases} 0 & x \le 0 \\ x & 0 < x \le 1. \\ 1 & x > 1 \end{cases}$$



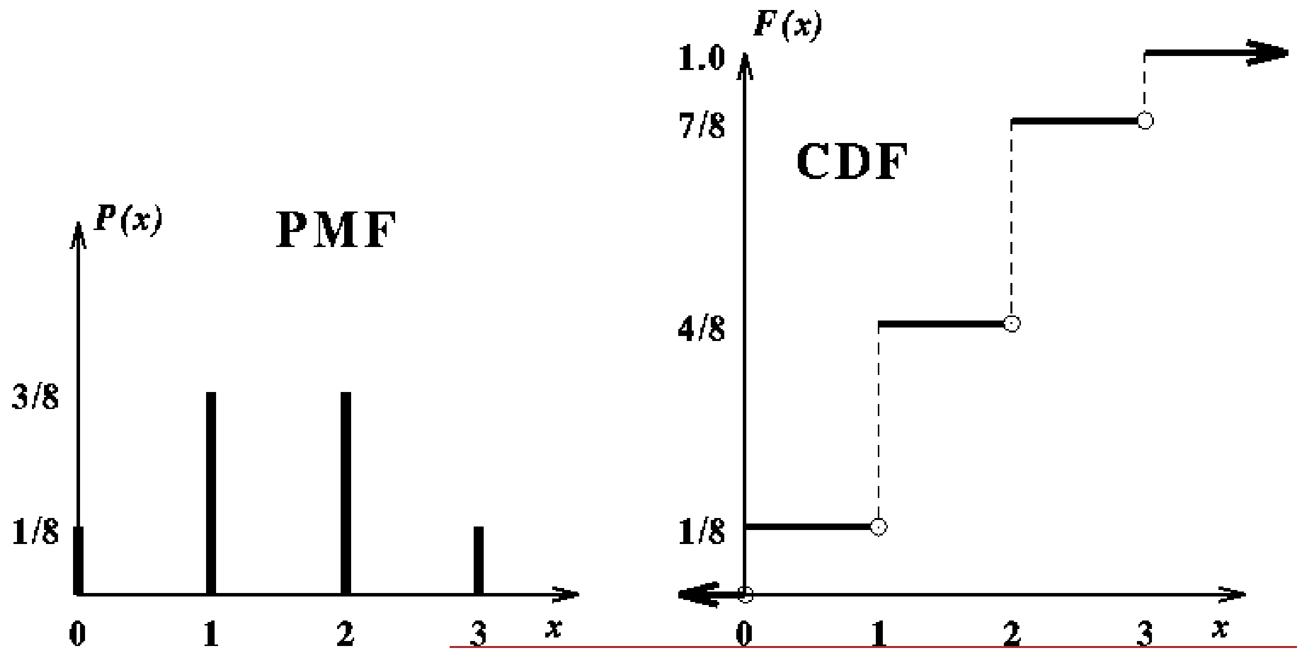

General CDF of the random variable $X$.

- For discrete random variables, the CDF can be written in terms of the probability mass function. Consider a general random variable, $X$, which can take on values from the discrete set $\{x1, x2, x3, \ldots\}$. The CDF for this random variable is

$$F_X(x) = \sum_{i=1}^{k} P_X(x_i), \quad \text{for } x_k \le x < x_{k+1}.$$

# Example

- A discrete random variable may be defined for the random experiment of flipping a coin.
- The sample space of outcomes is $S = \{H, T\}$.
- Defining the random variable $X$ to be $X(H) = 0$ and $X(T) = 1$.
- That is, the sample space $H$, $T$ is mapped to the set $\{0, 1\}$ by the random variable $X$.
- Assuming a fair coin, the resulting probability mass function is $PX(0) = 1/2$ and $PX(1) = 1/2$.



# Discrete Random Variables

- Bernoulli Random Variable
- Binomial Random Variable
- Poisson Random Variable
- Geometric Random Variable

# Continuous Random Variables

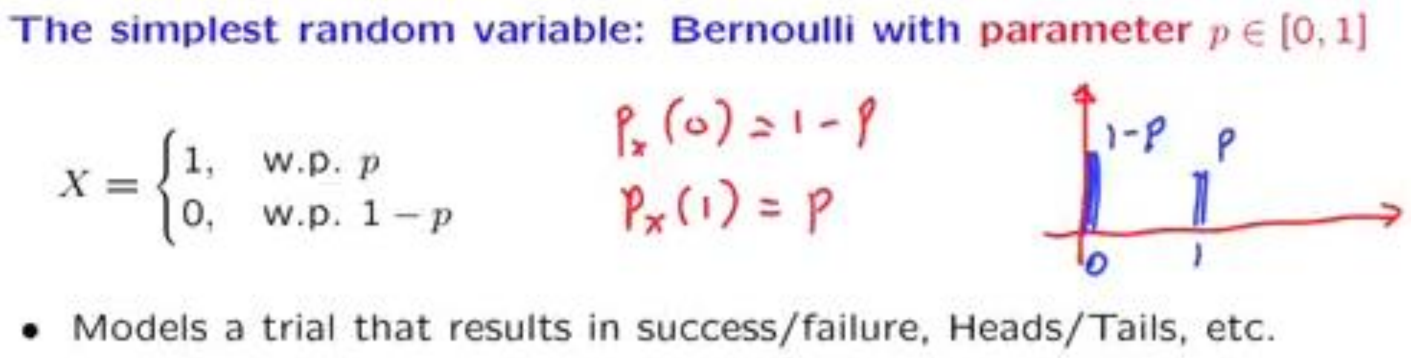
- Gaussian (Normal) Random Variable
- Uniform Variable
- Exponential Random Variable
- Gamma Random Variable
- Chi-Squared Random Variable

Bernoulli Random Variable

### *Bernoulli Random Variable*

- This is the simplest possible random variable and is used to represent experiments that have two possible outcomes.
- These experiments are called Bernoulli trials and the resulting random variable is called a *Bernoulli random variable*.
- It is most common to associate the values {0,1} with the two outcomes of the experiment.
- If $X$ is a Bernoulli random variable, its probability mass function
- is of the form

$$P(X(0)) = 1 - p,\ P(X(1)) = p.$$



# Bernoulli Random Variable

- The coin tossing experiment would produce a Bernoulli random variable.
- In that case, we may map the outcome $H$ to the value $X = 1$ and $T$ to $X = 0$.
- Also, we would use the value $p = 1/2$ assuming that the coin is fair.
- Examples of engineering applications might include
- 
    - radar systems where the random variable could indicate the presence ($X = 1$) or absence ($X = 0$) of a target
    - a digital communication system where $X = 1$ might indicate a bit was transmitted in error while $X = 0$ would indicate that the bit was received correctly.

# *Binomial Random Variable*

- Consider repeating a Bernoulli trial $n$ times, where the outcome of each trial is independent of all others.
- The Bernoulli trial has a sample space of $S = \{0, 1\}$ and we say that the repeated experiment has a sample space of $Sn = \{0, 1\}n$, which is referred to as a *Cartesian space*.
- That is, outcomes of the repeated trials are represented as $n$ element vectors whose elements are taken from $S$.
- Consider, for example, the outcome

$$\xi_k = (\overbrace{1, 1, \dots, 1}^{k \text{ times}}, \overbrace{0, 0, \dots, 0}^{n-k \text{ times}})$$

- The probability of this outcome occurring is

$$\Pr(\xi_k) = \Pr(1, 1, \dots, 1, 0, 0, \dots, 0) = \Pr(1)\Pr(1) \dots \Pr(1)\Pr(0)\Pr(0) \dots \Pr(0)$$
$$= (\Pr(1))^k (\Pr(0))^{n-k} = p^k (1-p)^{n-k}.$$

# *Binomial Random Variable*

- The random variable $X$ represent the number of times the outcome 1 occurred in the sequence of $n$ trials.
- This is known as a *binomial random variable* and takes on integer values from 0 to $n$.
- To find the probability mass function of the binomial random variable, let $Ak$ be the set of all outcomes that have exactly $k$ 1s and $n - k$ 0s.
- Note that all outcomes in this event occur with the same probability.
- Furthermore, all outcomes in this event are mutually exclusive. Then

$$P_X(k) = \Pr(A_k) = (\#\text{ of outcomes in } A_k) * (\text{probability of each outcome in } A_k)$$

$$= \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, 2, \ldots, n.$$

$$\boxed{\binom{n}{k} = \frac{n!}{k!(n-k)!}.}$$

- binomial expansion

$$\boxed{(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}.}$$

$$\boxed{\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = (p + 1 - p)^n = 1^n = 1}$$

# *Binomial Random Variable*

# *Applications*

- Binomial random variables occur in practice any time Bernoulli trials are repeated.
- For example, in a digital communication system, a packet of $n$ bits may be transmitted and we might be interested in the number of bits in the packet that are received in error.
- A bank manager might be interested in the number of tellers who are serving customers at a given point in time.
- A medical technician might want to know how many cells from a blood sample are white and how many are red.
- The coin tossing experiment was repeated $n$ times and the random variable $Y$ represented the number of times heads occurred in the sequence of $n$ tosses.
- This is a repetition of a Bernoulli trial, and hence the random variable $Y$ should be a binomial random variable with $p = 1/2$ (assuming the coin is fair).

# Poisson Random Variable

- Consider a binomial random variable, $X$, where the number of repeated trials, $n$, is very large.
- In that case, evaluating the binomial coefficients can pose numerical problems.
- If the probability of success in each individual trial, $p$, is very small, then the binomial random variable can be well approximated by a *Poisson random variable*.
- That is, the Poisson random variable is a limiting case of the binomial random variable. Formally, let $n$ approach infinity and
- $p$ approach zero in such a way that $\lim_{n\to\infty} np = a$.
- Then the binomial probability mass function converges to the form

**Poisson Distribution Formula**

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

where

$x$ = 0, 1, 2, 3, …

$\lambda$ = mean number of occurrences in the interval

$e$ = Euler's constant $\approx$ 2.71828

# *Poisson Random Variable*

# Applications

- It is most commonly used in queuing theory and in communication networks.
- The number of customers arriving at a cashier in a store during some time interval may be well modeled as a Poisson random variable.
- As may the number of data packets arriving at a given node in a computer network.

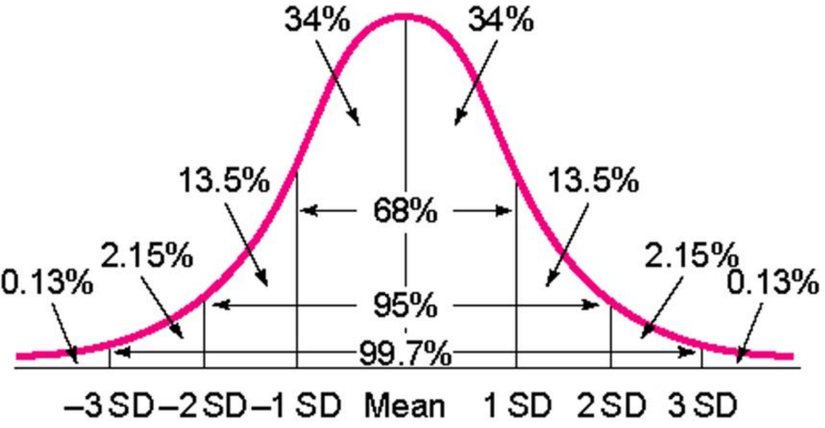
# *Geometric Random Variable*

- Consider repeating a Bernoulli trial until the first occurrence of the outcome $\xi 0$.

- If $X$ represents the number of times the outcome $\xi 1$ occurs before the first occurrence of $\xi 0$, then $X$ is a *geometric random variable* whose probability mass function is

$$P_X(k) = (1-p)p^k, \quad k = 0, 1, 2, \ldots .$$

# The Gaussian Random Variable




- In the study of random variables, the Gaussian random variable is clearly the most commonly used and of most importance.

- A Gaussian random variable is one whose probability density function can be written in the general form

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) \mu$$

- The PDF of the Gaussian random variable has two parameters, $\mu$ and $\sigma$, which have the interpretation of the mean and standard deviation respectively.
- The parameter $\sigma^2$ is referred to as the variance.

- In general, the Gaussian PDF is centered about the point $x = \mu$ and has a width that is proportional to $\sigma$.

- This random variable is referred to as a "normal" random variable. Furthermore,
- for the special case when $\mu = 0$ and $\sigma = 1$, it is called a "standard normal" random variable.

# The Gaussian Random Variable

- Gaussian random variable, $X \sim N(\mu, \sigma^2)$. This is read "$X$" is distributed normally (or Gaussian) with mean, $\mu$, and variance, σ2."
- The CDF of a Gaussian random variable is written as

$$F_X(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-m)^2}{2\sigma^2}\right) dy.$$

- For general Gaussian random variables that are not in the normalized form, the CDF can be expressed in terms of a –function using a simple transformation.
- Making the transformation $t = (y - \mu)/\sigma$, resulting in

$$F_X(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-m)^2}{2\sigma^2}\right) dy = \int_{-\infty}^{\frac{x-m}{\sigma}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt = \Phi\left(\frac{x-m}{\sigma}\right)$$

- The $Q$-function is more natural for evaluating probabilities of the form $\Pr(X > x)$.
- Following a line of reasoning identical to the previous paragraph, it is seen that if $X \sim N(m, \sigma 2)$, then

$$\Pr(X > x) = \int_{\frac{x-m}{\sigma}}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt = Q\left(\frac{x-m}{\sigma}\right)$$

- Furthermore, since we have shown that $\Pr(X > x) = Q((x-m)/\sigma)$ and $\Pr(X \le x) =((x - m)/\sigma)$, it is apparent that the relationship between the $\Phi$-function and the $Q$-function is

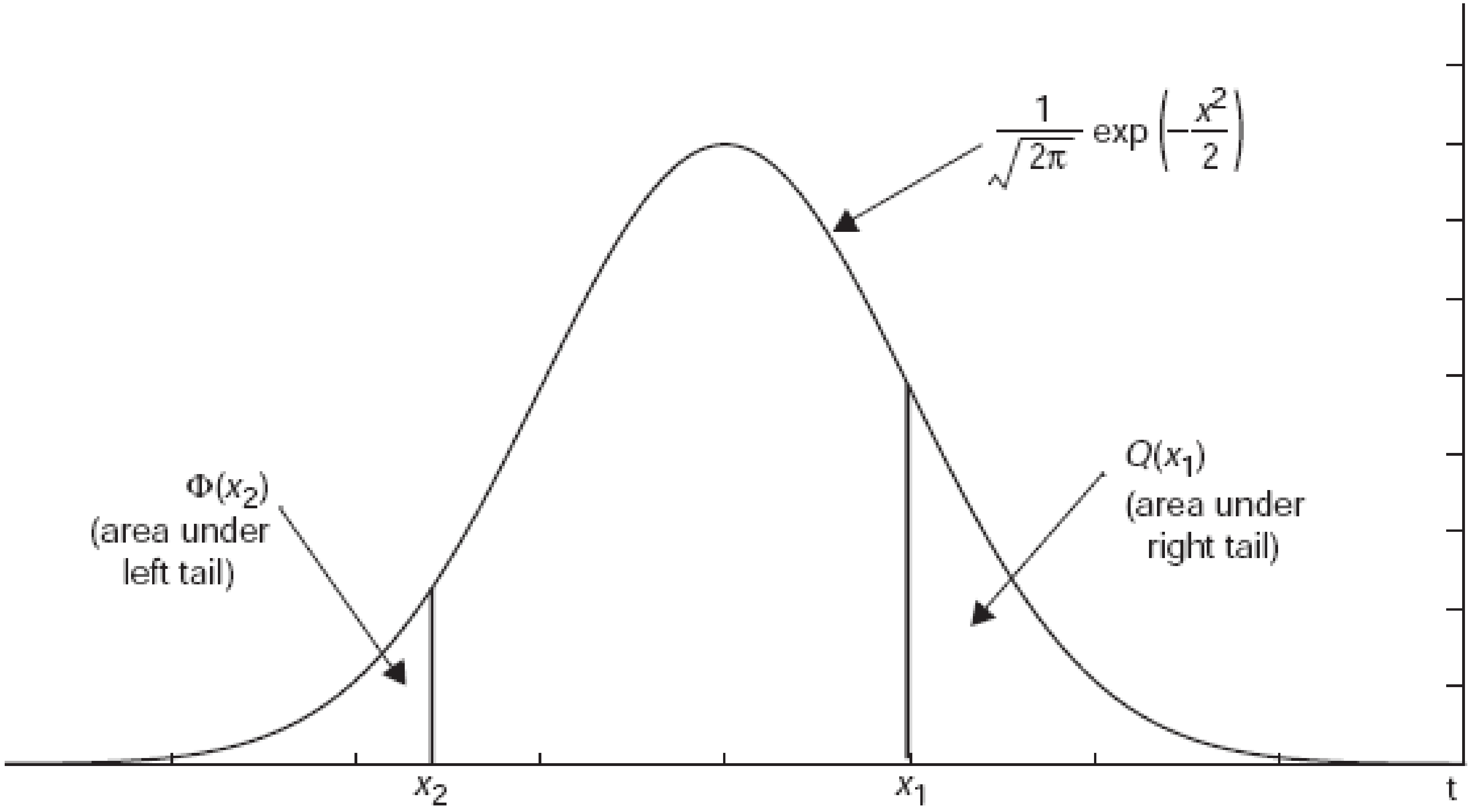
$$Q(x) = 1 - \Phi(x).$$

- The CDF of a Gaussian random variable can be written in terms of a $Q$-function as

$$F_X(x) = 1 - Q\left(\frac{x-m}{\sigma}\right)$$

- If it is required to evaluate the $Q$- function at a negative value, the relationship $Q(x) = 1 - Q(-x)$ can be used.
- That is, to evaluate $Q(-2)$ for example, $Q(2)$ can be evaluated and then use $Q(-2) = 1 - Q(2)$.

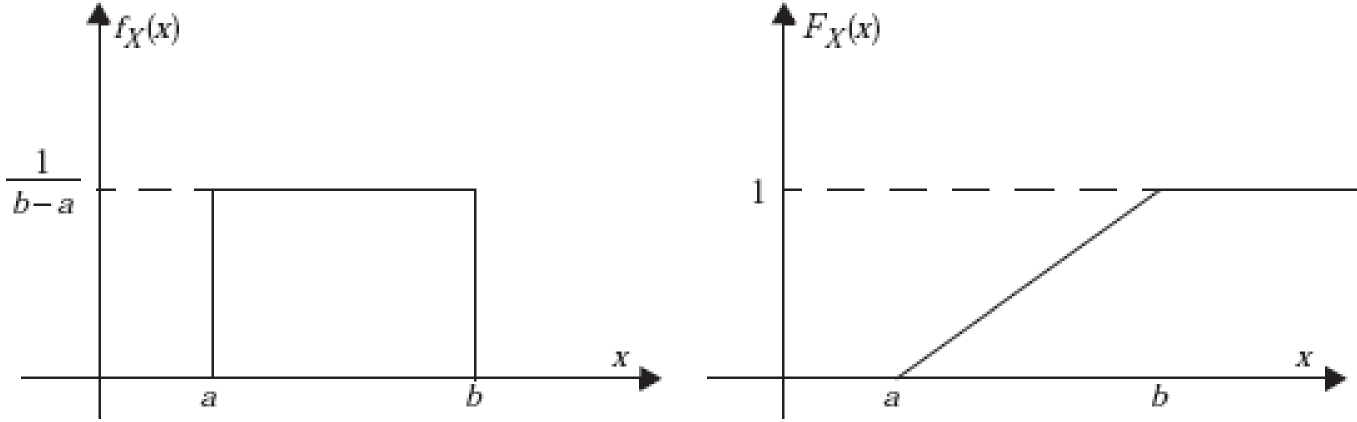
- The Standardized integrals related to the Gaussian CDF: the $\Phi(\cdot)$ and $Q(\cdot)$ functions.



# Uniform Random Variable

- The uniform probability density function is constant over an interval $[a, b)$. The PDF and its corresponding CDF are

$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & a \le x < b, \\ 0 & \text{elsewhere.} \end{cases} \qquad F_X(x) = \begin{cases} 0 & x < a, \\ \dfrac{x-a}{b-a} & a \le x < b, \\ 1 & x \ge b. \end{cases}$$




**PDF and CDF of a Uniform Random Variable**

- An example of a uniform random variable would be the phase of a radio frequency sinusoid in a communications system.
- Although the transmitter knows the phase of the sinusoid, the receiver may have no information about the phase.
- In this case, the phase at the receiver could be modeled as a random variable uniformly distributed over the interval [0, 2π).

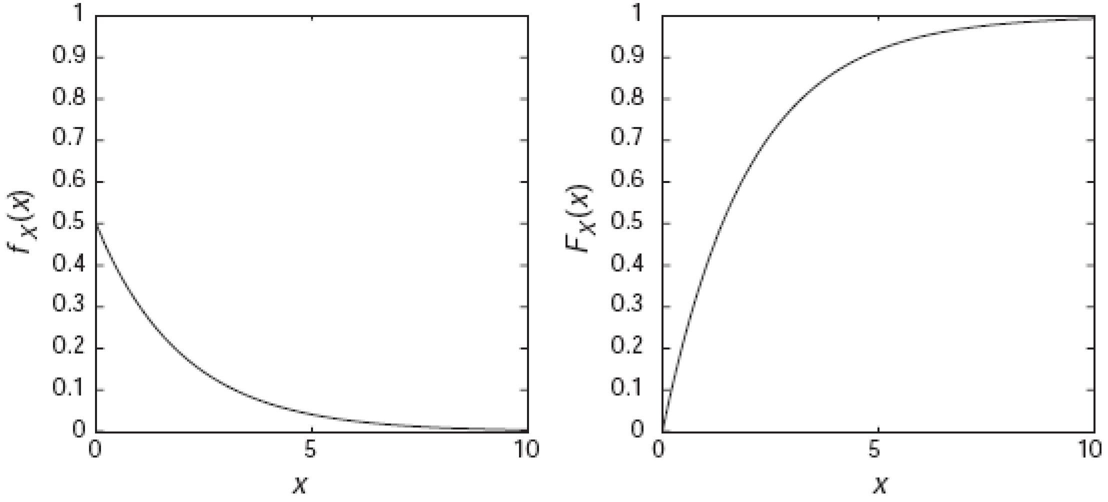
# Exponential Random Variable

- The exponential random variable has a probability density function and cumulative distribution function given (for any $b > 0$) by

$$f_X(x) = \frac{1}{b}\exp\left(-\frac{x}{b}\right)u(x),$$

$$F_X(x) = \left[1 - \exp\left(-\frac{x}{b}\right)\right]u(x).$$

- A plot of the PDF and the CDF of an exponential random variable is shown in Figure below.
- The parameter b is related to the width of the PDF and the PDF has a peak value of 1/b which occurs at x = 0. The PDF and CDF are nonzero over the semi-infinite interval (0,∞), which may be either open or closed on the left endpoint.



- Exponential random variables are commonly encountered in the study of queueing systems.

- The time between arrivals of customers at a bank, for example, is commonly modeled as an exponential random variable,

- as is the duration of voice conversations in a telephone network.

# Conditional Distribution and Density Functions

- Conditional CDF.
- The conditional cumulative distribution function of a random variable, $X$, conditioned on the event $A$ having occurred is

$$F_{X|A}(x) = \Pr(X \leq x|A) = \frac{\Pr(\{X \leq x\}, A)}{\Pr(A)}.$$

- Naturally, this definition requires the caveat that the probability of the event $A$ must not be zero.

- The properties of CDFs

$$(1)\ F_{X|A}(-\infty) = 0, F_{X|A}(\infty) = 1,$$

$$(2)\ 0 \leq F_{X|A}(x) \leq 1,$$

$$(3)\ \text{For } x_1 < x_2, F_{X|A}(x_1) \leq F_{X|A}(x_2),$$

$$(4)\ \text{For } x_1 < x_2, \Pr(x_1 < X \leq x_2|A) = F_{X|A}(x_2) - F_{X|A}(x_1).$$

# Conditional PDF

- The conditional probability density function of a random variable $X$ conditioned on some event $A$ is

$$f_{X|A}(x) = \lim_{\varepsilon \to 0} \frac{\Pr(x \le X < x + \varepsilon | A)}{\varepsilon}.$$

- Properties of Conditional PDF

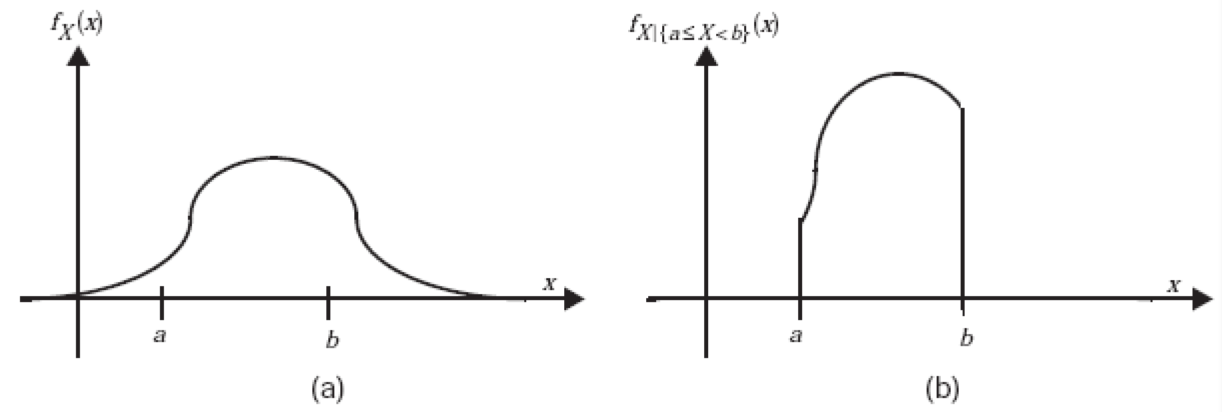
$$(1)\ f_{X|A}(x) \ge 0.$$

$$(2)\ f_{X|A}(x) = \frac{dF_{X|A}(x)}{dx}.$$

$$(3)\ F_{X|A}(x) = \int_{-\infty}^{x} f_{X|A}(y)\, dy.$$

$$(4)\ \int_{-\infty}^{\infty} f_{X|A}(x)\, dx = 1.$$

$$(5)\ \int_{a}^{b} f_{X|A}(x)\, dx = \Pr(a < X \le b|A).$$




A PDF (a) and the corresponding conditional PDF (b).

## Operations on a Single Random Variable

- Expected Value of a Random Variable
- Expected Values of Functions of Random Variables
- Moments
- Conditional Expected Values
- Transformations of Random Variables
- Characteristic Functions
- Probability Generating Functions
- Moment Generating Functions
- Evaluating Tail Probabilities

# Expected Value of a Continuous Random Variable

- **Expected Value of a Random Variable**
- This is the idea of an average or expected value of a random variable.
- The *expected value* of a random variable $X$ which has a PDF, $fX(x)$, is

$$E[X] = \int_{-\infty}^{\infty} x f_X(x)\, dx.$$

- The terms *average*, *mean*, *expectation*, and *first moment* are all alternative names for the concept of expected value.
- Furthermore, an overbar is often used to denote expected value so that the symbol $X$ is to be interpreted as meaning the same thing as $E[X]$.
- Another commonly used notation is to write $\mu_X = E[X]$.

# Expected Value of a Discrete Random Variable

- The definition of expected values for discrete random variables

$$E[X] = \sum_k x_k P_X(x_k).$$

- Hence, the expected value of a discrete random variable is simply a weighted average of the values that the random variable can take on, weighted by the probability mass of each value.

- **Expected Values of Functions of Random Variables**
- Given a random variable $X$ with PDF $fX(x)$, the expected value of a function, $g(X)$, of that random variable is given by

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx.$$

- For a discrete random variable, this definition reduces to

$$E[g(X)] = \sum_k g(x_k) P_X(x_k).$$

- **THEOREM**
- For any constants $a$ and $b$, $E[aX + b] = aE[X] + b$.
- Furthermore, for any function $g(x)$ that can be written as a sum of several other functions (i.e., $g(x) = g1(x) + g2(x) + \cdots + gN(x)$),

$$E\left[\sum_{k=1}^{N} g_k(X)\right] = \sum_{k=1}^{N} E[g_k(X)]$$

- In other words, expectation is a linear operation and the expectation operator can be exchanged (in order) with any other linear operation.