# CHAPTER 2
# Probability

# Contents

# What is Probability?

- The concept of probability is necessary in works with physical, biological, or social mechanisms that generate observations that cannot be predicted with certainty:
  - ✓ The blood pressure of a person at a given point in time
  - ✓ We never know the exact load that a bridge will endure before collapsing into a river
  - ✓ Events possessing this property are called random, or stochastic events
- The stable long-term relative frequency provides an intuitively meaningful measure of our belief in the occurrence of a random event if a future observation is to be made

# What is Probability?

- $\theta$ = relative frequency of head in a "large number" of "identical flip"
- Statistical results assume that data were from a controlled experiment
- Nothing is more important than repeatability (e.g., same experimental conditions)

**Activity**

1. Estimate the parameter $\theta$ by conducting experiments.
2. Give me estimates.

FIGURE 9.2
Values of $\hat{p} = Y/n$ for 50 sequences of 1000 Bernoulli trials, $p = 0.5$

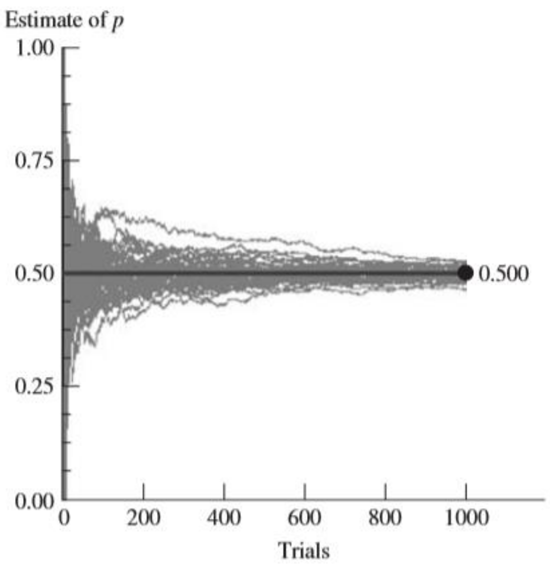
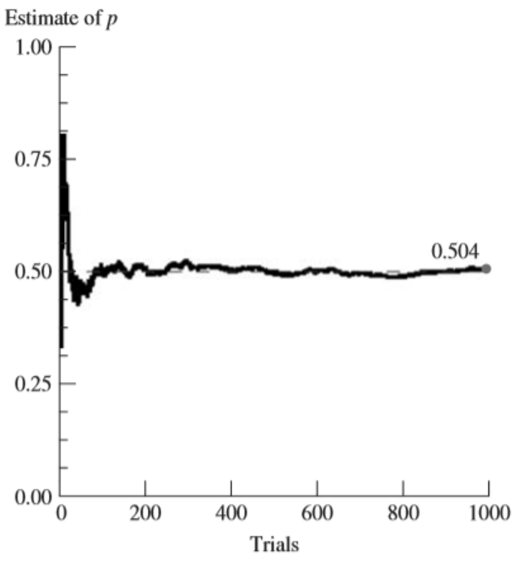



FIGURE 9.1
Values of $\hat{p} = Y/n$ for a single sequence of 1000 Bernoulli trials, $p = 0.5$



# Two Different Views on Probability?

**Frequentists :**

- Probability only has a meaning in terms of a limiting case of repeated measurements
- Probabilities are fundamentally related to frequencies of events

**Bayesian :**

- Degrees of certainty about statements
- Probabilities are fundamentally related to our own knowledge about an event

# Different Approaches to Statistics

**Frequentists :**

- Data are a repeatable random sample→ there is a frequency of occurrence
- Underlying parameters remain constant during this repeatable process
- Parameters are fixed and unchanging under all realistic circumstances

The true parameters are fixed, and a subset of data are realized from these parameters.
Then, we randomly sample a subset of data (varying) to estimate the fixed parameters.

**Bayesian :**

- Data are observed from the realized sample
- Parameters are unknown and described probabilistically (View the world probabilistically)
- Data are fixed

We collected data, thus the data is given to us (fixed).
Then, we try to estimate model parameters that can best describe the collected data.

# Different Approaches to Statistics

**Frequentists :**

- Data are a repeatable random sample→ there is a frequency of occurrence
- Underlying parameters remain constant during this repeatable process
- Parameters are fixed and unchanging under all realistic circumstances

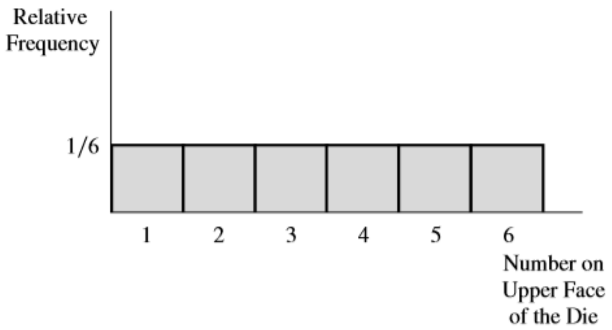
The true parameters are fixed, and a subset of data are realized from these parameters. Then, we randomly sample a subset of data (varying) to estimate the fixed parameters.

In this lecture, we accept an interpretation based on ***a relative frequency*** as a meaningful measure of our belief in the occurrence of an event.

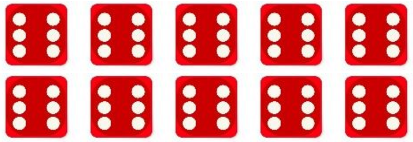
# What is Probability?

- Consider a gambler who wishes to make an inference concerning the *balance of a die*.

If die is balanced




The result is possible
but highly ***improbable***
given that the hypothesis is true
(probability of occurrence is very low)

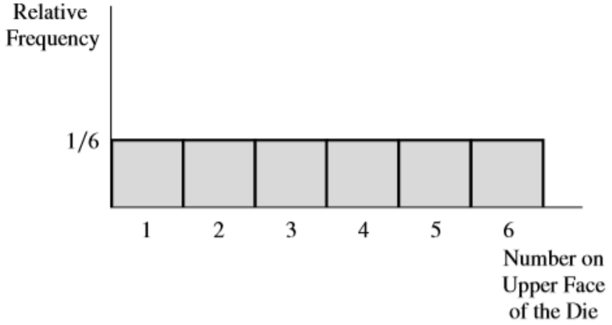


10 random samples:
All 6s were observed from ten random tosses

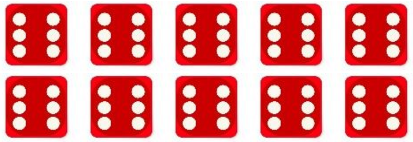
# What is Probability?

- Consider a gambler who wishes to make an inference concerning the *balance of a die*.

~~If die is balanced~~




The result is possible
but highly ***improbable***
given that the hypothesis is true
(probability of occurrence is very low)

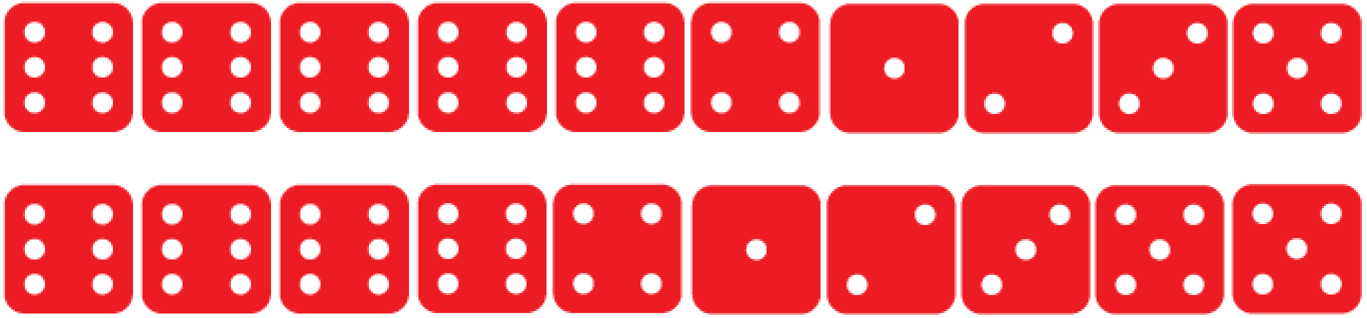


His decision was based on the probability of the observed sample.

10 random samples:
All 6s were observed from ten random tosses

# Role of Probability in Inference

- How about the following cases; do we reject or not?



- If we must rely solely on experience and intuition to make our evaluation, it is not so easy to decide whether the probability of five 6s or four 6s in ten tosses is large or small.

- We need a theory of probability that will permit us to calculate the probability (or a quantity proportional to the probability) of observing specified outcomes, assuming that our hypothesized model is correct.

# Set Theory Basics

- We will use capital letters, $A$, $B$, $C$ …, to denote sets.
- If the elements in the set $A$ are $a_1, a_2$ and $a_3$, we will write

$$A = \{a_1, a_2, a_3\}$$

- $S$ denotes the set of all elements under consideration; $S$ is the *universal set*
- A set is a collection of objects, which are its *elements*
  - ✓ $\omega \in A$ means that $\omega$ is an element of the set $A$
  - ✓ A set with no elements is called the empty set, denoted by $\emptyset$
- Types of sets:
  - Finite: $A = \{\omega_1, \omega_2, \dots, \omega_n\}$
  - Countably infinite: $A = \{\omega_1, \omega_2, \dots\}$, e.g., the set of integers
  - Uncountable: A set that takes a continuous set of values, e.g., the $[0,1]$ interval, the real line, etc.

# Set Theory Basics

- A set can be described by all $\omega$ having a certain property, e.g., A = [0,1] can be written as $A = \{\omega : 0 \leq \omega \leq 1\}$
- A set $B \subset A$ means that every element of $B$ is an element of $A$
- A universal set $S$ contains all objects of particular interest in a particular context, e.g., sample space for random experiment.

# Set Operation

- Assume a universal set $S$
- Three basic operations:
  - ✓ Relative complementation: A complement of a set $A$ with respect to $S$ is

$$A^c = \bar{A} = \{\omega \in S: \omega \notin A\}, \quad \text{so } S^c = \emptyset$$

  - ✓ Intersection:

$$A \cap B = AB = \{\omega: \omega \in A \text{ and } \omega \in B\}$$

  - ✓ Union:

$$A \cup B = \{\omega: \omega \in A \text{ or } \omega \in B\}$$

- Notation:
  - $\bigcup_{i=1}^{n} A_i = A_1 \cup A_2 \ldots \cup A_n$
  - $\bigcap_{i=1}^{n} A_i = A_1 \cap A_2 \ldots \cap A_n$
- A collection of sets $A_1, A_2, \ldots, A_n$ are *disjoint* or *mutually exclusive* if $A_i \cap A_j = \emptyset$ for all $i \neq j$ (no two of them have a common element)
- A collection of sets $A_1, A_2, \ldots, A_n$ *partition* $S$ if they are disjoint and $\bigcup_{i=1}^{n} A_i = S$

# Algebra of Sets

- Distributive laws
  - $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
  - $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

- *DeMorgan's law*
  - $\overline{(A \cap B)} = \bar{A} \cup \bar{B}$
  - $\overline{(A \cup B)} = \bar{A} \cap \bar{B}$

# Experiment

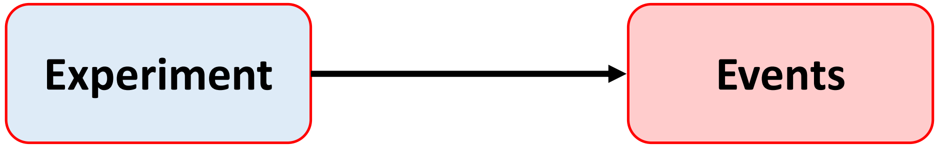
- We will use the term *experiment* to include observations obtained from completely uncontrollable situations:
  - ✓ Tossing a coin
  - ✓ Measuring the IQ score of an individual
  - ✓ Determining the number of bacteria per cubic centimeter in a portion of processed food
  - ✓ Observations on the daily price of a particular stock
  - ✓ Observations made under controlled laboratory conditions

**DEFINITION 2.1**

An *experiment* is the process by which an observation is made.

# Event

- When an experiment is performed, it can result in one or more outcomes, which are called *events*




- For example, if the experiment consists of counting the number of bacteria in a portion of food, some events of interest could be

  ✓ *A* : Exactly 110 bacteria are present.
  ✓ *B* : More than 200 bacteria are present.
  ✓ *C* : The number of bacteria present is between 100 and 300.

# Event

- Some events associated with a single toss of a balanced die are these:
  - ✓ $E_1$: Observe a 1.
  - ✓ $E_2$: Observe a 2.
  - ✓ $E_3$: Observe a 3.
  - ✓ $E_4$: Observe a 4.
  - ✓ $E_5$: Observe a 5.
  - ✓ $E_6$: Observe a 6.
  - ✓ $A$: Observe an odd number.
  - ✓ $B$: Observe a number less than 5.
  - ✓ $C$: Observe a 2 or a 3.

# Event

- Some events associated with a single toss of a balanced die are these:

***Simple Event***

- ✓ $E_1$: Observe a 1. corresponding sample point $s_1$ in sample space $S$ then $E_1 = \{s_1\}$
- ✓ $E_2$: Observe a 2. corresponding simple point $s_2$ in sample space $S$ then $E_2 = \{s_2\}$
- ✓ $E_3$: Observe a 3.
- ✓ $E_4$: Observe a 4.
- ✓ $E_5$: Observe a 5.
- ✓ $E_6$: Observe a 6.

- ✓ $A$: Observe an odd number. $A = E_1 \cup E_3 \cup E_5$ or $A = \{s_1, s_3, s_5\}$
- ✓ $B$: Observe a number less than 5. $B = E_1 \cup E_2 \cup E_3 \cup E_4 \cup E_5$
- ✓ $C$: Observe a 2 or a 3. $C = E_2 \cup E_3$

***Compound Event***

# Simple Event : Sample Point

**DEFINITION 2.2**

A *simple event* is an event that cannot be decomposed. Each simple event corresponds to one and only one *sample point*. The letter $E$ with a subscript will be used to denote a **simple event** or the corresponding **sample point**.

- We can think of a simple event as a set consisting of a single point—namely, the single sample point associated with the event.
- However, to prevent any notational confusion, an event will be in capital letter, while an element will be given in lower cases in the slides, unless otherwise specified.

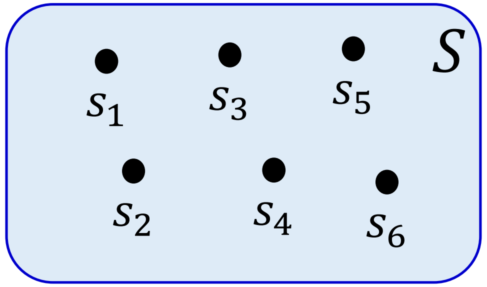
# Sample Space

**DEFINITION 2.3**

The *sample space* associated with an experiment is the set consisting of all possible sample points. A sample space will be denoted by $S$.

- Sample space $S$ associated with the die-tossing experiment consist of six sample points

- Counting bacteria in a food specimen, let $s_0$ sample points correspond to observing 0 bacteria, $s_1$ correspond to observing 1 bacterium, and so on.




$$S = \{s_1, s_2, s_3, s_4, s_5, s_6\}$$

**finite distinct sample points**

$$S = \{s_0, s_1, \ldots\}$$

**infinite countable sample points**

# Discrete Sample Space

**DEFINITION 2.4**

A *discrete sample space* is one that contains either a finite or countable number of distinct sample points.

- Thus, the single sample point $s_1$ associated with observing a 1 and the single sample point $s_2$ associated with observing a 2 are distinct, and the sets $\{s_1\}$ and $\{s_2\}$ are mutually exclusive sets.

- Thus, events $E_1 = \{s_1\}$ and $E_2 = \{s_2\}$ are mutually exclusive events. Similarly, all distinct simple events correspond to mutually exclusive sets of simple events and are thus mutually exclusive events.

- For experiments with discrete sample spaces, **compound events** can be viewed as collections (sets) of sample points or, equivalently, as unions of the sets of single sample points corresponding to the appropriate simple events.

set of sample points

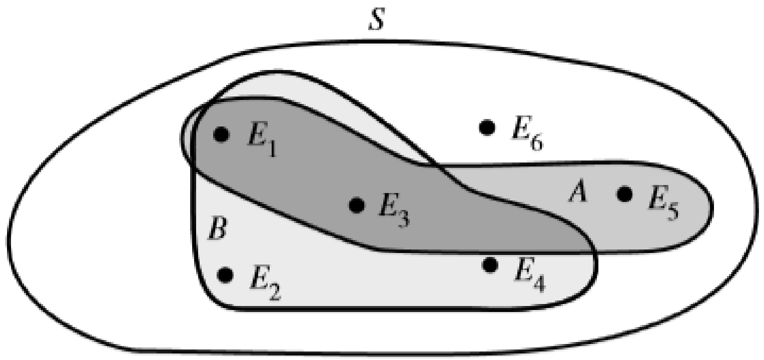
$$A = E_1 \cup E_2 = \{s_1\} \cup \{s_2\} = \{s_1, s_2\}$$

union of the sets of single sample points

# Event Defined as a Set

- For experiments with discrete sample spaces, compound events can be viewed as collections (sets) of sample points or, equivalently, as unions of the sets of single sample points corresponding to the appropriate simple events.
- For example, the die tossing event $A$ (observe an odd number) will occur if and only if one of the simple events $E_1, E_3$, or $E_5$ occurs. Thus,

$$A = E_1 \cup E_3 \cup E_5$$




**DEFINITION 2.5**

An *event* in a discrete sample space $S$ is a collection of sample points – that is, any subset of $S$.

# Probabilistic Model for an Experiment

- A probabilistic model for an experiment with a discrete sample space can be constructed by assigning *a numerical proba*bility to *each simple event* in the sample space S.
  (or sample point)

- We will select this number, a measure of our belief in the event's occurrence on a single repetition of the experiment, in such a way that it will be consistent with the relative frequency concept of probability.

- Although relative frequency does not provide a rigorous definition of probability, any definition applicable to the real world should agree with our intuitive notion of the relative frequencies of events.

# Axioms of Probability

**DEFINITION 2.6**

Suppose $S$ is a sample space associated with an experiment. To every event $A$ in $S$ ($A$ is a subset of $S$), we assign a number, $P(A)$, called the *probability* of $A$, so that the following axioms hold:

Axiom 1: $P(A) \geq 0$.
Axiom 2: $P(S) = 1$.
Axiom 3: If $A_1, A_2, A_3, \ldots$ form a sequence of pairwise **mutually exclusive** events in $S$ (that is, $A_i \cap A_j = \phi$ if $i \neq j$), then

$$P(A_1 \cup A_2 \cup A_3 \ldots) = P(\cup_{i=1}^{\infty} A_i) = \Sigma_{i=1}^{\infty} P(A_i)$$

- Notice that the definition states only the conditions an assignment of probabilities must satisfy;
  - it does not tell us how to assign specific probabilities to events.

# Probabilistic Model for an Experiment

- For discrete sample spaces, it suffices to assign probabilities to each simple event.
- If a balanced die is used for the die-tossing example, it seems reasonable to assume that all simple events would have the same relative frequency in the long run.
  - ✓ We will assign a probability of 1/6 to each *simple event*:
    - ➢ $P(E_i) = 1/6, \quad \text{for } i = 1,2,\ldots,6.$
    - ➢ This assignment of probabilities agrees with **Axiom 1**.
  - ✓ To see that **Axiom 2** is satisfied, write

$$P(S) = P(E_1 \cup E_2 \cup \cdots \cup E_6) = P(E_1) + P(E_2) + \cdots + P(E_6) = 1.$$

    - ➢ The second equality follows because **Axiom 3** must hold.
  - ✓ Axiom 3 also tells us that we can calculate the probability of any event by summing the probabilities of the simple events contained in that event (recall that distinct simple events are mutually exclusive).
    - ➢ Event $A$ was defined to be "observe an odd number." Hence,

$$P(A) = P(E_1 \cup E_3 \cup E_5) = P(E_1) + P(E_3) + P(E_5) = 1/2.$$

# Overview

**EXAMPLE 2.1**

A manufacturer has five seemingly identical computer terminals available for shipping. Unknown to her, two of the five are defective. A particular order calls for two of the terminals and is filled by randomly selecting two of the five that are available.

a. List the sample space for this experiment.
b. Let $A$ denote the event that the order is filled with two non defective terminals. List the sample points in $A$.
c. Construct a Venn diagram for the experiment that illustrates event $A$.
d. Assign probabilities to the simple events in such a way that the information about the experiment is used and the axioms in Definition 2.6 are met.
e. Find the probability of event $A$.

# Overview

## SOLUTION 2.1

# The Sample-Point Method

- Finding the probability of an event defined on a sample space that contains a finite or denumerable (countably infinite) set of sample points can be approached in two ways,
  - ✓ the sample-point method
  - ✓ the event-composition methods (**Chapter 2.9**)

# The Sample-Point Method

**DEFINITION**

The sample-point method is outlined in Section 2.4. The following steps are used to find the probability of an event:

1. Define the experiment and clearly determine how to describe one simple event.
2. List the *simple events* associated with the experiment and test each to make certain that it cannot be decomposed. This defines the sample space $S$.
3. Assign reasonable probabilities to the sample points in $S$, making certain that $P(E_i) \geq 0$ and $\sum_i P(E_i) = 1$.
4. Define the event of interest, $A$, as a specific collection of sample points. (A sample point is in $A$ if $A$ occurs when the sample point occurs. Test *all* sample points in $S$ to identify those in $A$.)
5. Find $P(A)$ by summing the probabilities of the sample points in $A$.

# The Sample-Point Method

- Recall that sample space $S$ is said to be discrete if it is countable

- The probability measure $P$ can be simply defined by first assigning probabilities to outcomes, i.e., elementary events $\{\omega\}$, such that:

$$P(\{\omega\}) \geq 0, \text{for all } \omega \in S, \text{and}$$

$$\sum_{\omega \in S} P(\{\omega\}) = 1$$

- The probability of any other event $A$ (by the additivity axiom) is simply

$$P(A) = \sum_{\omega \in A} P(\{\omega\})$$

# The Sample-Point Method

- Examples:
  - ✓ For the coin flipping experiment, assign

$$P(\{H\}) = p \text{ and } P(\{T\}) = 1 - p, \qquad \text{for } 0 \leq p \leq 1$$

Note: $p$ is the bias of the coin, and a coin is fair if $p = \frac{1}{2}$

  - ✓ For the die rolling experiment, assign

$$P(\{i\}) = \frac{1}{6}, \text{for } i = 1,2, \dots, 6$$

- The probability of the event "the outcome is even", $A = \{2,4,6\}$, is

$$P(A) = P(\{2\}) + P(\{4\}) + P(\{6\}) = \frac{1}{2}$$

# The Sample-Point Method

- If $S$ is countably infinite, we can again assign probabilities to elementary events

- Example: Assume $S = \{1, 2, \dots\}$, assign probability $2^{-k}$ to event $\{k\}$. The probability of the event "the outcome is even"

$$
\begin{aligned}
P(\text{ outcome is even }) &= P(\{2,4,6,8,\dots\}) \\
&= P(\{2\}) + P(\{4\}) + P(\{6\}) + \dots \\
&= \sum_{k=1}^{\infty} P(\{2k\}) \\
&= \sum_{k=1}^{\infty} 2^{-2k} = \frac{1}{3}
\end{aligned}
$$

# Example

**EXAMPLE 2.3**

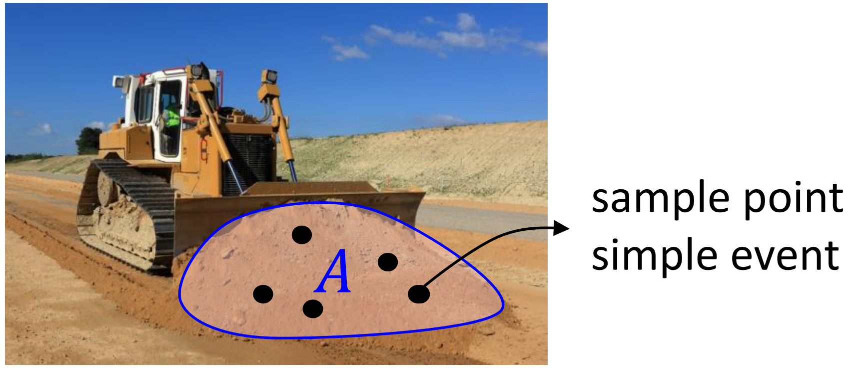
A balanced coin is tossed three times. Calculate the probability that exactly two of the three tosses result in heads.

# Example

## SOLUTION 2.3

# Limitations

- The sample-point method for solving a probability problem is direct and powerful and in some respects is a bulldozer approach.




- **Limitations:**
  - ✓ it is not resistant to human error.
  - ✓ Occasionally, very large number of sample points and a complete itemization is tedious and time consuming.

- Tools that reduce the effort and error associated with the sample-point approach for finding the probability of an event include *orderliness*, *a computer*, and *the mathematical theory of counting, called combinatorial analysis*.

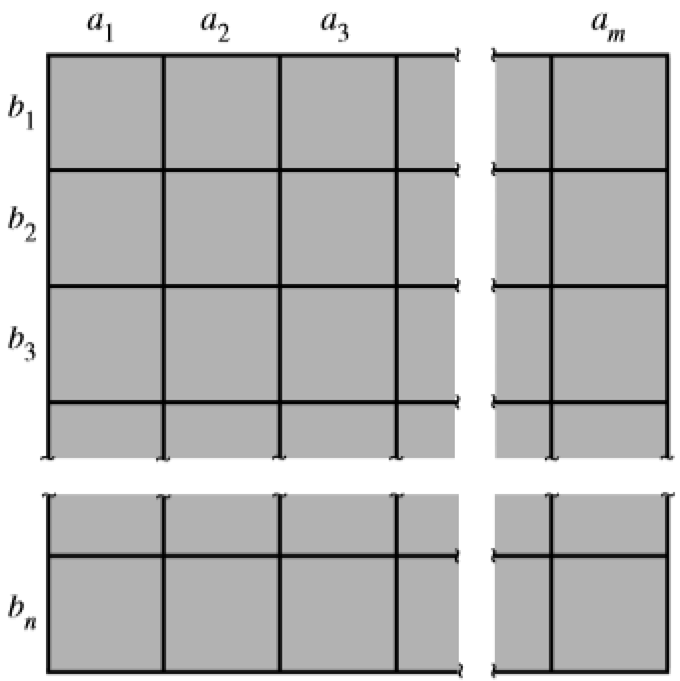
# Overview

- Presents some useful results from the *theory of combinatorial analysis* and illustrates their application to the sample-point method for finding the probability of an event.
- In many cases, these results enable you to count the total number of sample points in the sample space $S$ and in an event of interest.

# mn rule

**THEOREM 2.1**

With $m$ elements $a_1, a_2, \dots, a_m$ and $n$ elements $b_1, b_2, \dots, b_n$ it is possible to form $mn = m \times n$ pairs containing one element from each group.



# Permutation

**DEFINITION 2.7**

An ordered arrangement of $r$ distinct objects is called a *permutation*. The number of ways of ordering $n$ distinct objects taken $r$ at a time will be designated by the symbol $P_r^n$.

- The total number of sample points equals the number of distinct ways that the respective symbols can be arranged in sequence.

**THEOREM 2.2**

$$P_r^n = n(n-1)(n-2)\dots(n-r+1) = \frac{n!}{(n-r)!}$$

# Partitioning

**THEOREM 2.3**

The number of ways of partitioning $n$ distinct objects into $k$ distinct groups containing $n_1, n_2, \dots, n_k$ objects, respectively, where each object appears in exactly one group and $\sum_{i=1}^{k} n_i = n$, is

$$N = \binom{n}{n_1 \; n_2 \; \dots \; n_k} = \frac{n!}{n_1!\, n_2! \dots n_k!}$$

- The terms $\binom{n}{n_1 \; n_2 \cdots n_k}$ are often called multinomial coefficients because they occur in the expansion of the multinomial term $y_1 + y_2 + \cdots y_k$ raised to the $n$th power:

$$(y_1 + y_2 + \cdots y_k)^n = \sum \binom{n}{n_1 \; n_2 \; \dots \; n_k} y_1^{n_1} y_2^{n_2} \cdots y_k^{n_k},$$

where this sum is taken over all $n_i = 0, 1, \dots, n$ such that $n_1 + n_2 + \cdots n_k = n$

# Combination

**DEFINITION 2.8**

The number of *combinations* of $n$ objects taken $r$ at a time is the number of subjects, each of size $r$, that can be formed from the $n$ objects. This number will be denoted by $C_r^n$ or $\binom{n}{r}$.

**THEOREM 2.4**

The number of unordered subsets of size $r$ chosen (without replacement) from $n$ available objects is

$$\binom{n}{r} = C_r^n = \frac{P_r^n}{r!} = \frac{n!}{r!\,(n-r)!}.$$

- The terms $\binom{n}{r}$ are generally referred to as binomial coefficients because they occur in the binomial expansion

$$\begin{aligned}(x+y)^n &= \binom{n}{0}x^n y^0 + \binom{n}{1}x^{n-1}y^1 + \binom{n}{2}x^{n-2}y^2 + \cdots + \binom{n}{n}x^0 y^n \\ &= \sum_{i=0}^{n}\binom{n}{i}x^{n-i}y^i\end{aligned}$$

# Conditional Probability

- Probability of rain tomorrow (*unconditional probability*)

- Probability of rain tomorrow
  given that it has rained two proceeding days and that a tropical storm is predicted
  (*conditional probability*)

Which one is larger?

# Conditional Probability

**DEFINITION 2.9**

The *conditional probability of an event* $A$, given that an event $B$ has occurred, is equal to

$$P(A|B) = \frac{P(A \cap B)}{P(B)},$$

provided $P(B) > 0$. [The symbol $P(A|B)$ is read "probability of $A$ given $B$".]

# Consistency with Relative Frequency Concept

| | $A$ | $\overline{A}$ | |
|---|---|---|---|
| $B$ | $n_{11}$ | $n_{12}$ | $n_{11}+n_{12}$ |
| $\overline{B}$ | $n_{21}$ | $n_{22}$ | $n_{21}+n_{22}$ |
| | $n_{11}+n_{21}$ | $n_{12}+n_{22}$ | $N$ |

$$P(A) \approx \frac{n_{11}+n_{21}}{N}$$

$$P(B) \approx \frac{n_{11}+n_{12}}{N}$$

$$P(A \cap B) \approx \frac{n_{11}}{N}$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \approx \frac{\left(\frac{n_{11}}{N}\right)}{\left(\frac{n_{11}+n_{12}}{N}\right)} = \frac{n_{11}}{n_{11}+n_{12}}$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \approx \frac{\left(\frac{n_{11}}{N}\right)}{\left(\frac{n_{11}+n_{21}}{N}\right)} = \frac{n_{11}}{n_{11}+n_{21}}$$

# Example

**EXAMPLE 2.14**

Suppose that a balanced die is tossed once. Use Definition 2.9 to find the probability of a 1, given that an odd number was obtained.

# Example

## SOLUTION 2.14

# Independence

**DEFINITION 2.10**

Two events $A$ and $B$ are said to be *independent* if any one of the following holds:

$$P(A|B) = P(A),$$

$$P(B|A) = P(B),$$

$$P(A \cap B) = P(A)P(B).$$

Otherwise, the events are said to be *dependent*.

- The probability of the occurrence of an event $A$ is unaffected by the occurrence or nonoccurrence of event $B$.
  - ✓ "smoking" and "contracting lung cancer"
  - ✓ "rain today" and "rain a month from today"

# Independence : Example

**EXAMPLE 2.15**

Consider the following events in the toss of a single die:

$A$: Observe an odd number.

$B$: Observe an even number.

$C$: Observe a 1 or 2.

a. Are $A$ and $B$ independent events?
b. Are $A$ and $C$ independent events?

# Independence : Example

## SOLUTION 2.15

# The Multiplicative Law of Probability

**THEOREM 2.5 (The Multiplicative Law of Probability)**

The probability of the intersection of two events $A$ and $B$ is

$$\begin{aligned} P(A \cap B) &= P(A)P(B|A) \\ &= P(B)P(A|B). \end{aligned}$$

If $A$ and $B$ are independent, then

$$P(A \cap B) = P(A)P(B).$$

- **Proof:**

# The Multiplicative Law of Probability

**THEOREM 2.5 (The Multiplicative Law of Probability)**

The probability of the intersection of two events $A$ and $B$ is

$$\begin{aligned} P(A \cap B) &= P(A)P(B|A) \\ &= P(B)P(A|B). \end{aligned}$$

If $A$ and $B$ are independent, then

$$P(A \cap B) = P(A)P(B).$$

- The multiplicative law can be extended to find the probability of the intersection of *any number* of events:

$$\begin{aligned} P(A \cap B \cap C) = P[(A \cap B) \cap C] &= P(A \cap B)P(C|A \cap B) \\ &= P(A)P(B|A)P(C|A \cap B) \end{aligned}$$

- For the intersection of any number of, say, $k$ events:

$$P(A_1 \cap A_2 \cap A_3 \cap \cdots \cap A_k) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \ldots P(A_k|A_1 \cap A_2 \cap \cdots \cap A_{k-1})$$

# The Additive Law of Probability

**THEOREM 2.6 (The Additive Law of Probability)**

The probability of the union of two events $A$ and $B$ is

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

If $A$ and $B$ are mutually exclusive events, $P(A \cap B) = 0$ and

$$P(A \cup B) = P(A) + P(B).$$

- **Proof :**

# The Additive Law of Probability

**THEOREM 2.6 (The Additive Law of Probability)**

The probability of the union of two events $A$ and $B$ is

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

If $A$ and $B$ are mutually exclusive events, $P(A \cap B) = 0$ and

$$P(A \cup B) = P(A) + P(B).$$

- The additive law can be extended to find the probability of the intersection of *any number* of events:

$$\begin{aligned} P(A \cup B \cup C) &= P[A \cup (B \cup C)] \\ &= P(A) + P(B \cup C) - P[A \cap (B \cup C)] \\ &= P(A) + P(B) + P(C) - P(B \cap C) - P[(A \cap B) \cup (A \cap C)] \\ &= P(A) + P(B) + P(C) - P(B \cap C) - P(A \cap B) - P(A \cap C) + P(A \cap B \cap C) \end{aligned}$$

$$\because (A \cap B) \cap (B \cap C) = A \cap B \cap C$$

# Complementary

**THEOREM 2.7**

If $A$ is an event, then

$$P(A) = 1 - P(\bar{A}).$$

- **Proof:**

- It is sometimes easier to calculate $P(\bar{A})$ than to calculate $P(A)$. In such cases, it is easier to find $P(A)$ by the relationship $P(\bar{A}) = 1 - P(A)$ than to find $P(A)$ directly.

# The Event-Composition Method

- Sets (events) can often be expressed as unions, Intersections, or complements of other sets.
- The event-composition method for calculating the probability of an event, $A$
  - ✓ expresses $A$ as a composition involving unions and/or intersections of other events.
  - ✓ The laws of probability are then applied to find $P(A)$.

# The Event-Composition Method

**DEFINITION**

A summary of the steps used in the event-composition method follows:

1. Define the experiment.
2. Visualize the nature of the sample points. Identify a few to clarify your thinking.
3. Write an equation expressing the event of interest – say, $A$ – as a composition of two or more events, using unions, intersections, and/or complements. (Notice that this equates point sets.) Make certain that event $A$ and the event implied by the composition represent the same set of sample points.
4. Apply the additive and multiplicative laws of probability to the compositions obtained in step 3 to find $P(A)$.

- The event-composition approach does *not require listing the sample points in $S$*, but it does require a clear understanding of the nature of a typical sample point.

# The Event-Composition Method : Example

**EXAMPLE 2.21**

Observation of a waiting line at a medical clinic indicates the probability that a new arrival will be an emergency case is $p = \frac{1}{6}$. Find the probability that the $r$th patient is the first emergency case. (Assume that the conditions of arriving patients represent independent events.)

# The Event-Composition Method : Example

**SOLUTION 2.21**

# The Law of Total Probability

**DEFINITION 2.11**

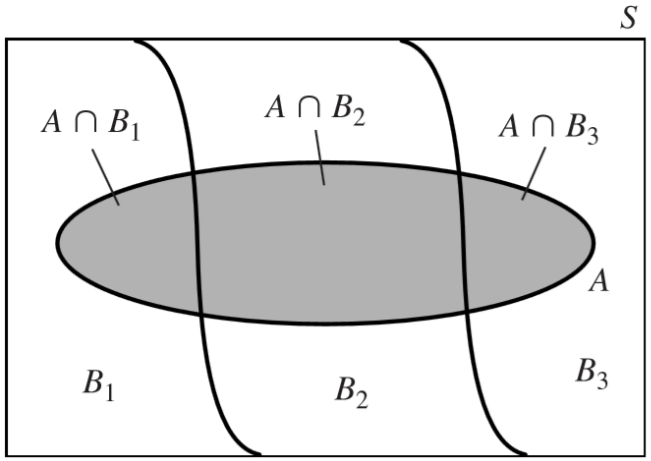
For some positive integer $k$, let the sets $B_1, B_2, \ldots B_k$ be such that

1. $S = \bigcup_{i=1}^{k} B_i$
2. $B_i \cap B_j = \phi$, for $i \neq j$.

Then, the collection of sets $\{B_1, B_2, \ldots B_k\}$ is said to be a *partition* of $S$.

- If $A$ is any subset of $S$ and $\{B_1, B_2, \ldots, B_k\}$ is a partition of $S$, $A$ can be decomposed as follows:

$$A = (A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_k)$$




(When $k = 3$)

# The Law of Total Probability

**THEOREM 2.8**

Assume that $\{B_1, B_2, \ldots, B_k\}$ is a partition of $S$ (see Definition 2.11) such that $P(B_i) > 0$, for $i = 1,2, \ldots, k$. Then for any event $A$

$$P(A) = \sum_{i=1}^{k} P(A|B_i)P(B_i)\,.$$

- **Proof:**

# Bayes' Rule

**THEOREM 2.9 (Bayes' Rule)**

Assume that $\{B_1, B_2, \dots B_k\}$ is a partition of $S$ (see Definition 2.11) such that $P(B_i) > 0$, for $i = 1,2,\dots,k$. Then

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^{k} P(A|B_i)P(B_i)}.$$

- **Proof:**

# Bayes' Rule : Example

**EXAMPLE 2.23**

An electronic fuse is produced by five production lines in a manufacturing operation. The fuses are costly, are quite reliable, and are shipped to suppliers in 100-unit lots. Because testing is destructive, most buyers of the fuses test only a small number of fuses before deciding to accept or reject lots of incoming fuses.

All five production lines produce fuses at the same rate and normally produce only 2% defective fuses, which are dispersed randomly in the output. Unfortunately, production line 1 suffered mechanical difficulty and produced 5% defectives during the month of March. This situation became known to the manufacturer after the fuses had been shipped. A customer received a lot produced in March and tested three fuses. One failed. What is the probability that the lot was produced on line 1? What is the probability that the lot came from one of the four other lines?

# Bayes' Rule : Example

**SOLUTION 2.23**

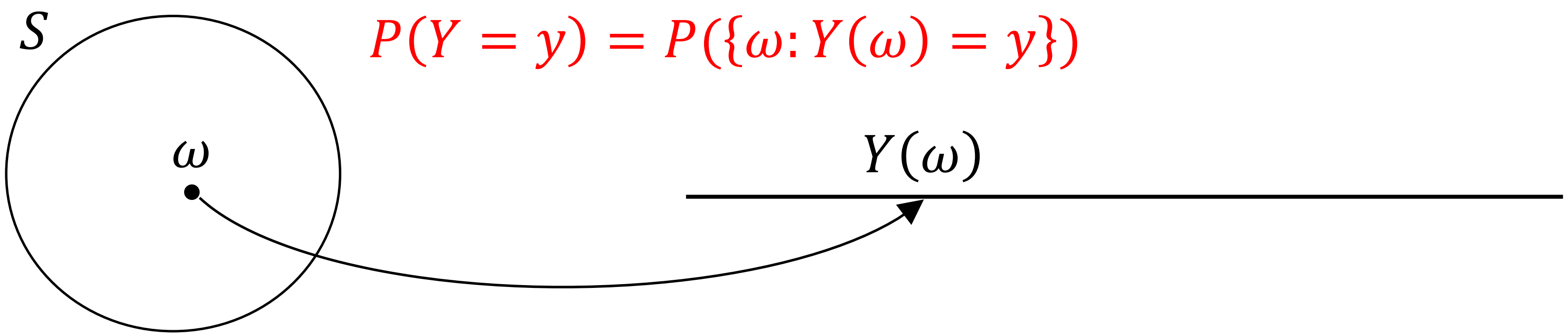
# Random Variable

- Events of major interest to the scientist, engineer, or businessperson are those identified by numbers, called *numerical events*.
  - ✓ The physician is interested in the event that ten of ten treated patients survive an illness;
  - ✓ the business person is interested in the event that sales next year will reach $5 million.
- Let $X$ denote a variable to be measured in an experiment. Because the value of $X$ will vary depending on the outcome of the experiment, it is called a *random variable*.

# Random Variable

**DEFINITION 2.12**

A *random variable* (r.v.) $Y$ is a real-valued function $Y(\omega)$ over the sample space $S$ of a random experiment, i.e., $Y : S \to \mathbb{R}$

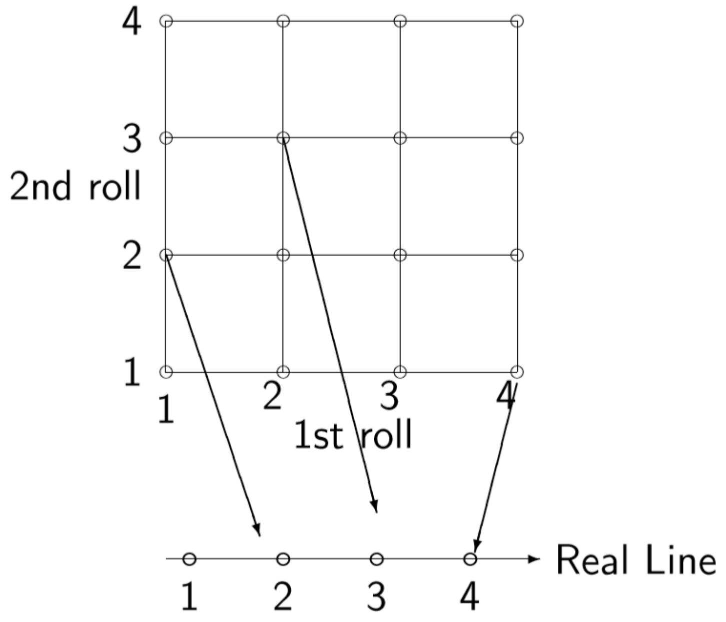



$P(Y = y) = P(\{\omega : Y(\omega) = y\})$

- Randomness comes from $\omega$ ($Y(\omega)$ is a deterministic function)
- Notations:
  - ✓ Always use upper case letters for random variables ($X, Y, \ldots$)
  - ✓ Always use lower case letters for values of random variables: $Y = y$ means that the random variable $Y$ takes on the value $y$

# Random Variable : Examples

1. Flip a coin $n$ times. Here $S = \{H, T\}^n$. $\{H, T\}^n = \{(x_1, \ldots, x_n) : x_i \in \{H, T\}, i = 1, \ldots, n\}$
2. Define the random variable $Y \in \{0,1,2, \ldots, n\}$ to be the number of heads.
3. Roll a 4-sided die twice.
   (a) Define the random variable $Y$ as the maximum of the two rolls




   (b) Define the random variable $X$ to be the sum of the outcomes of the two rolls
   (c) Define the random variable $Z$ to be 0 if the sum of the two rolls is odd and 1 if it is even

# Random Variable : Examples

3. Flip coin until first heads shows up. Define the random variable $Y \in \{1,2 \dots.\}$ to be the number of flips until the first heads

4. Let $S = \mathbb{R}$. Define the two random variables

   (a) $X(\omega) = \omega$

   (b) $Y(\omega) = \begin{cases} +1 & \text{for } \omega \geq 0 \\ -1 & \text{otherwise} \end{cases}$

5. $n$ packets arrive at a node in a communication network. Here $S$ is the set of arrival time sequences $(t_1, t_2, \dots, t_n) \in (0, \infty)^n$ $\quad : (0,\infty)^n = \{(x_1, \dots, x_n) : 0 < x_i < \infty, i = 1, \dots, n\}$

   (a) Define the random variable $N$ to be the number of packets arriving in the interval $(0,1]$

   (b) Define the random variable $T$ to be the first inter-arrival time

# Random Variable Examples

**EXAMPLE 2.24, 2.25**

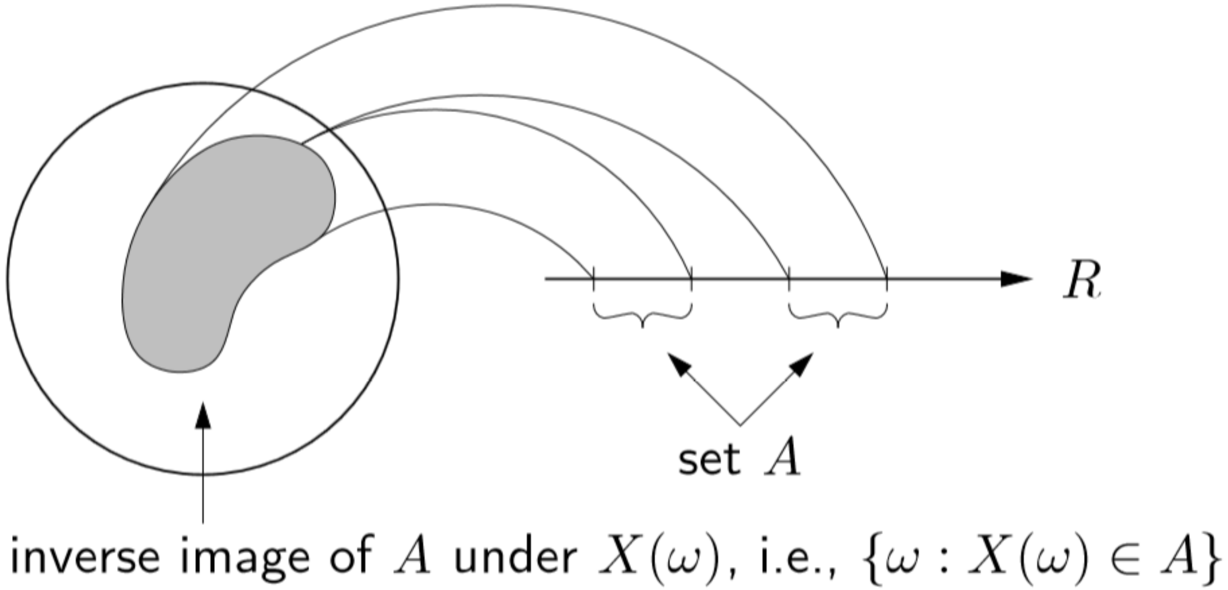
Define an experiment as tossing two coins and observing the results. Let $Y$ equal the number of heads obtained. Identify the sample points in $S$, assign a value of $Y$ to each sample point, and identify the sample points associated with each value of the random variable $Y$. Plus, compute the probabilities for each value of $Y$.

# Random Variable Examples

## SOLUTION 2.24, 2.25

# Specifying a Random Variable

- Specifying a random variable means being able to determine the probability that $X \in A$ for any interval $A \subset \mathbb{R}$
- To do so, we consider the inverse image of the set $A$ under $X(\omega), X^{-1}(A) \coloneqq \{\omega: X(\omega) \in A\}$

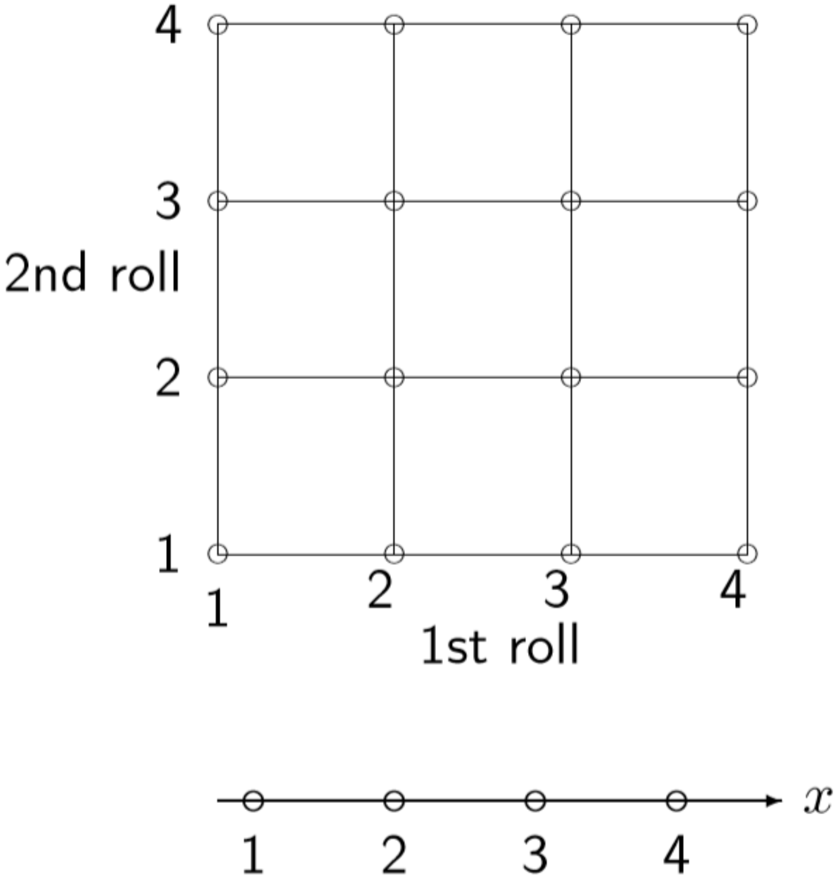



- So, $X \in A$ iff $\omega \in \{\omega: X(\omega) \in A\}$,

$$P(\{X \in A\}) = P(\{\omega: X(\omega) \in A\}) \text{ or } P\{X \in A\} = P\{\omega: X(\omega) \in A\}$$

## Specifying a Random Variable : Example

- Roll fair 4-sided die twice independently: Define the r.v. $X$ to be the maximum of the two rolls. What is the $P\{0.5 < X \leq 2\}$?



# Types of a Random Variable

- **Discrete**: $X$ can assume only one of a countable number of values. Such r.v. can be specified by a *Probability Mass Function* (pmf). **(Chapter 3)**

- **Continuous**: $X$ can assume one of a continuum of values and the probability of each value is 0. Such r.v. can be specified by a *Probability Density Function* (pdf). **(Chapter 4)**

- **Mixed**: $X$ is neither discrete nor continuous. Such r.v. (as well as discrete and continuous r.v.s) can be specified by a *Cumulative Distribution Function* (cdf). **(Chapter 4)**

# Random Sampling

- A statistical experiment involves the observation of a sample selected from a larger body of data, existing or conceptual, called a population.

- The probability of the observed sample plays a major role in making an inference and evaluating the credibility of the inference.

- Without belaboring the point, it is clear that the method of sampling will affect the probability of a particular sample outcome. For example, if all the N = 5 population elements are distinctly different,

  ✓ What is the probability of drawing a specific pair, when sampling without replacement?

  ✓ What is the probability of drawing the same specific pair, when sampling with replacement?

# Random Sampling

**DEFINITION 2.13**

Let $N$ and $n$ represent the numbers of elements in the population and sample, respectively. If the sampling is conducted in such a way that each of the $\binom{N}{n}$ samples has an equal probability of being selected, the sampling is said to be random, and the result is said to be a *random sample*.

- In many situations *the population is conceptual*, as in an observation made during a laboratory experiment.
- Here the population is envisioned to be the infinitely many measurements that would be obtained if the experiment were to be repeated over and over again.
- If we wish a sample of $n = 10$ measurements from this population, we repeat the experiment ten times and hope that the results represent, to a reasonable degree of approximation, a random sample.

# Questions?