Welcome To

BASIC CONCEPT OF PROBABILITY

Probability– Models for random phenomena

E-mail: mesbahru@gmail.com                           Tel: +88-0721-750041/4122
Dr. Md. Mesbahul Alam, Associate Professor

Random phenomena
- Unable to predict the outcomes, but in the long-run, the outcomes exhibit statistical regularity.

Examples
1. Tossing a coin – outcomes $S = \{\text{Head, Tail}\}$
   Unable to predict on each toss whether is Head or Tail. In the long run we can predict that approx 50% of the time heads will occur and approx 50% of the time tails will occur.

2. Rolling a die – outcomes $S = \{1, 2, 3, 4, 5, 6\}$
   Unable to predict outcome but in the long run one can determine that each outcome will occur 1/6 of the time.
   Use symmetry. Each side is the same. One side should not occur more frequently than another side in the long run. If the die is not balanced this may not be true.

3. Rolling a two balanced dice – 36 outcomes

Introduction
- Probability is the study of randomness and uncertainty.
- In the early days, probability was associated with games of chance (gambling).
History

- Games of chance: 300 BC
- 1565: first formalizations
- 1654: Fermat & Pascal, conditional probability
- Reverend Bayes: 1750’s
- 1950: Kolmogorov: axiomatic approach
- Objectivists vs subjectivists
  - (frequentists vs Bayesians)
- Frequentist build one model
- Bayesians use all possible models, with priors

Concerns

- Future: what is the likelihood that a student will get a CS job given his grades?
- Current: what is the likelihood that a person has cancer given his symptoms?
- Past: what is the likelihood that Monte committed suicide?
- Combining evidence.
- Always: Representation & Inference

Probability versus Statistics

- Probability is the field of study that makes statements about what will occur when a sample is drawn from a known population.
- Statistics is the field of study that describes how samples are to be obtained and how inferences are to be made about unknown populations.

Simple Games Involving Probability

Game: A fair die is rolled. If the result is 2, 3, or 4, you win Tk. 100; if it is 5, you win Tk. 200; but if it is 1 or 6, you lose Tk. 300.

Should you play this game?

Random Experiment

- A random experiment is a process whose outcome is uncertain.

Examples:
- Tossing a coin once or several times
- Picking a card or cards from a deck
- Measuring temperature of patients
- ...

Assessing Probability (Classical Approach)

- Theoretical/Classical probability—based on theory (a priori understanding of a phenomena)

For example:
- Theoretical probability of rolling a 2 on a standard die is 1/6.
- Theoretical probability of choosing an ace from a standard deck is 4/52.
- Theoretical probability of getting heads on a regular coin is ½.
The sample Space, $S$ or $\Omega$

The **sample space**, $S$, for a random phenomena is the set of all possible outcomes.

The sample space $S$ may contain
1. A finite number of outcomes.
2. A countably infinite number of outcomes, or
3. An uncountably infinite number of outcomes.

Note. For cases 1 and 2 we can assign probability to every subset of the sample space. But for case 3 we can not do this. This situation will be discussed in the later workshops.

A countably infinite number of outcomes means that the outcomes are in a one-one correspondence with the positive integers 

$\{1, 2, 3, 4, 5, \ldots\}$

This means that the outcomes can be labeled with the positive integers.

$S = \{O_1, O_2, O_3, O_4, O_5, \ldots\}$

A uncountably infinite number of outcomes means that the outcomes can not be put in a one-one correspondence with the positive integers.

**Example:** A spinner on a circular disc is spun and points at a value $x$ on a circular disc whose circumference is 1.

$S = \{x \mid 0 \leq x < 1\} = [0, 1)$

Examples: Sample Space

1. Tossing a coin – outcomes $S = \{\text{Head, Tail}\}$
2. Rolling a die – outcomes $S = \{1, 2, 3, 4, 5, 6\}$
3. Rolling a two balanced dice – 36 outcomes

$S = \{(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6), (3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6), (4, 1), (4, 2), (4, 3), (4, 4), (4, 5), (4, 6), (5, 1), (5, 2), (5, 3), (5, 4), (5, 5), (5, 6), (6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6)\}$

outcome $(x, y)$,

$x =$ value showing on die 1

$y =$ value showing on die 2

An Event, $E$

The event, $E$, is any subset of the sample space, $S$, i.e. any set of outcomes (not necessarily all outcomes) of the random phenomena.
Examples: Event

1. Rolling a die – outcomes

\[ S = \{ \text{1, 2, 3, 4, 5, 6} \} \]

\[ E = \text{the event that an even number is rolled} \]

\[ E = \{ 2, 4, 6 \} \]

E = the event that “sum of rolled point is 7”

\[ E = \{ (6, 1), (5, 2), (4, 3), (3, 4), (3, 5), (1, 6) \} \]

2. Rolling a two balanced dice.

Sample Space

The sample space is the set of all possible outcomes.

Simple Events

The individual outcomes are called simple events.

Event

An event is any collection

of one or more simple events.

Special Events

The Null Event, The empty event - \( \phi \)

\[ \phi = \{ \} \]

The entire event, \( S \), always occurs.

The empty event, \( \phi \), never occurs.

Set operations on Events

Union

Let \( A \) and \( B \) be two events, then the union of \( A \)
and \( B \) is the event (denoted by \( A \cup B \)) defined by:

\[ A \cup B = \{ e | e \in A \text{ or } e \in B \} \]

The event \( A \cup B \) occurs if the event \( A \) occurs or
the event \( B \) occurs.

Union Examples

- \( \{a,b,c\} \cup \{2,3\} = \{a,b,c,2,3\} \)
- \( \{2,3,5\} \cup \{3,5,7\} = \{2,3,5,3,5,7\} \)

© Md. Mesbahul Alam, 2011
Intersection
Let $A$ and $B$ be two events, then the intersection of $A$ and $B$ is the event (denoted by $A \cap B$) defined by:

$$A \cap B = \{ e | e \in A \text{ and } e \in B \}$$

The event $A \cap B$ occurs if the event $A$ occurs and the event $B$ occurs.

Intersection Examples
- $\{a,b,c\} \cap \{2,3\} = \emptyset$
- $\{2,4,6\} \cap \{3,4,5\} = \{4\}$

Complement
Let $A$ be any event, then the complement of $A$ (denoted by $\overline{A}$) defined by:

$$\overline{A} = \{ e | e \notin A \}$$

The event $\overline{A}$ occurs if the event $A$ does not occur.

Set operations on Events
In problems you will recognize that you are working with:

1. **Union** if you see the word **or**,
2. **Intersection** if you see the word **and**,
3. **Complement** if you see the word **not**.

Rules from Set Theory
- **Commutative Laws**:
  $$A \cup B = B \cup A, \quad A \cap B = B \cap A$$
- **Associative Laws**:
  $$(A \cup B) \cup C = A \cup (B \cup C)$$
  $$(A \cap B) \cap C = A \cap (B \cap C).$$
- **Distributive Laws**:
  $$(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$$
  $$(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$$

DeMoivre’s laws
1. $\overline{A \cup B} = \overline{A} \cap \overline{B}$
2. $\overline{A \cap B} = \overline{A} \cup \overline{B}$
DeMoivre’s laws (in words)

1. $\overline{A \cup B} = \overline{A} \cap \overline{B}$
   The event $A$ or $B$ does not occur if the event $A$ does not occur and the event $B$ does not occur.

2. $\overline{A \cap B} = \overline{A} \cup \overline{B}$
   The event $A$ and $B$ does not occur if the event $A$ does not occur or the event $B$ does not occur.

Another useful rule

$A = (A \cap B) \cup (A \cap \overline{B})$

In words
The event $A$ occurs if $A$ occurs and $B$ occurs or $A$ occurs and $B$ doesn’t occur.

Rules involving the empty set, $\emptyset$, and the entire event, $S$.

1. $A \cup \emptyset = A$
2. $A \cap \emptyset = \emptyset$
3. $A \cup S = S$
4. $A \cap S = A$

Definition: mutually exclusive
Two events $A$ and $B$ are called mutually exclusive if:

$A \cap B = \emptyset$

If two events $A$ and $B$ are mutually exclusive then:

1. They have no outcomes in common. They can’t occur at the same time. The outcome of the random experiment can not belong to both $A$ and $B$.

Definition: Subset/ Superset
We will use the notation $A \subseteq B$ (or $B \supseteq A$) to mean that $A$ is a subset of $B$. ($B$ is a superset of $A$.) i.e. if $e \in A$ then $e \in B$. 
Union: \[ \bigcup_{i=1}^{k} E_i = E_1 \cup E_2 \cup E_3 \cup \cdots \cup E_k \]

Intersection: \[ \bigcap_{i=1}^{k} E_i = E_1 \cap E_2 \cap E_3 \cap \cdots \]

DeMorgan’s laws
1. \[ \overline{\bigcup_{i} E_i} = \bigcap_{i} \overline{E_i} \]
2. \[ \overline{\bigcap_{i} E_i} = \bigcup_{i} \overline{E_i} \]

Probability: Classical Approach

Example: Finite uniform probability space

In many examples the sample space \( S = \{o_1, o_2, o_3, \ldots o_N\} \) has a finite number, \( N \), of outcomes. Also each of the outcomes is equally likely (because of symmetry).

Then \( P\{o_i\} = 1/N \) and for any event \( E \)

\[ P[E] = \frac{n(E)}{n(S)} = \frac{n(E)}{N} = \frac{\text{no. of outcomes in } E}{\text{total no. of outcomes}} \]

Note: \( n(A) = \text{no. of elements of } A \)

Another Example:

We are shooting at an archery target with radius \( R \). The bullseye has radius \( R/4 \). There are three other rings with width \( R/4 \). We shoot at the target until it is hit

\[ S = \text{set of all points in the target} = \{(x, y) | x^2 + y^2 \leq R^2\} \]

\( E \), any event is a sub region (subset) of \( S \).
Thus this definition of $P[E]$, i.e.

$$P[E] = \frac{\text{Area}(E)}{\text{Area}(S)} = \frac{\text{Area}(E)}{\pi R^2}$$

satisfies the properties:

1. $P[E] \geq 0$, for each event $E$.
3. $P[\bigcup_{i=1}^{k} E_i] = \sum_{i=1}^{k} P[E_i]$ if $E_i \cap E_j = \emptyset$ for all $i, j$.

Finite uniform probability space

Many examples fall into this category

1. Finite number of outcomes
2. All outcomes are equally likely
3. $P[E] = \frac{n(E)}{n(S)} = \frac{n(E)}{N} = \frac{\text{no. of outcomes in } E}{\text{total no. of outcomes}}$

Note: $n(A) = \text{no. of elements of } A$

To handle problems in case we have to be able to count. Count $n(E)$ and $n(S)$.

Properties of Probability

1. For the null event $\emptyset$, $P(\emptyset) = 0$.
2. For any event $A$, $0 \leq P(A) \leq 1$.
3. For any event $A$, $P(A^c) = 1 - P(A)$.
4. If $A \subset B$, then $P(A) \leq P(B)$.
5. For any two events $A$ and $B$,

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

For three events, $A$, $B$, and $C$,

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

Independent Events

Intuitively, we define independence as:

Two events $A$ and $B$ are independent if the occurrence or non-occurrence of one of the events has no influence on the occurrence or non-occurrence of the other event.

Mathematically, we define independence as:

Two events $A$ and $B$ are independent if

$$\Pr(A \cap B) = \Pr(A) \Pr(B)$$.
Example of Independence

Are party ID and vote choice independent in presidential elections?

Suppose \( \Pr(\text{Rep. ID}) = .4 \), \( \Pr(\text{Rep. Vote}) = .5 \), and \( \Pr(\text{Rep. ID} \cap \text{Rep. Vote}) = .35 \).

To test for independence, we ask whether:

\[ \Pr(\text{Rep. ID}) \times \Pr(\text{Rep. Vote}) = .35 \, ? \]

Substituting into the equations, we find that:

\[ \Pr(\text{Rep. ID}) \times \Pr(\text{Rep. Vote}) = .4 \times .5 = .2 \neq .35 \]

so the events are not independent.

Independence of Several Events

The events \( A_1, \ldots, A_n \) are independent if:

\[ \Pr(A_1 \cap A_2 \cap \ldots \cap A_n) = \Pr(A_1)\Pr(A_2)\ldots\Pr(A_n) \]

And, this identity must hold for any subset of events.

Independent \( \neq \) Mutually exclusive

- Events \( A \) and \( \sim A \) are mutually exclusive, but they are NOT independent.
- \( \Pr(A \cap \sim A) = 0 \)
- \( \Pr(A)\Pr(\sim A) \neq 0 \)

Conceptually, once \( A \) has happened, \( \sim A \) is impossible; thus, they are completely dependent.

Conditional Probability

Conditional probabilities allow us to understand how the probability of an event \( A \) changes after it has been learned that some other event \( B \) has occurred.

The key concept for thinking about conditional probabilities is that the occurrence of \( B \) reshapes the sample space for subsequent events.

- That is, we begin with a sample space \( S \)
- \( A \) and \( B \) \( \in S \)
- The conditional probability of \( A \) given that \( B \) looks just at the subset of the sample space for \( B \).

The conditional probability of \( A \) given \( B \) is denoted \( \Pr(A \mid B) \).

- Importantly, according to Bayesian orthodoxy, all probability distributions are implicitly or explicitly conditioned on the model.

Conditional Probability Cont.

By definition: If \( A \) and \( B \) are two events such that \( \Pr(B) > 0 \), then:

\[ \Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)} \]

Example: What is the \( \Pr(\text{Republican Vote} \mid \text{Republican Identifier}) \)?

\( \Pr(\text{Rep. Vote} \cap \text{Rep. Id}) = .35 \) and \( \Pr(\text{Rep. ID}) = .4 \)

Thus, \( \Pr(\text{Republican Vote} \mid \text{Republican Identifier}) = .35 / .4 = .875 \)

Urn Example

- Suppose there are five balls in an urn. Three are red and two are blue. We will select a ball, note the color, and, without replacing the first ball, select a second ball.

There are four possible outcomes:

- Red, Red
- Red, Blue
- Blue, Red
- Blue, Blue

We can find the probabilities of the outcomes by using the multiplication rule for dependent events.
### Factorials
- For counting numbers 1, 2, 3, ...
- ! is read "factorial"
  - So for example, 5! is read "five factorial"
- \( n! = n \times (n-1) \times (n-2) \times \ldots \times 3 \times 2 \times 1 \)
  - So for example, \( 5! = 5 \times 4 \times 3 \times 2 \times 1 = 120 \)
- \( 1! = 1 \)
- \( 0! = 1 \)

### Permutations
- Permutation: ordered grouping of objects.
- Counting Rule for Permutations

  \[
  P_r = \frac{n!}{(n-r)!}
  \]

  where \( n \) and \( r \) are whole numbers and \( n \geq r \). Another commonly used notation for permutations is \( nPr \).

### Combinations
- A combination is a grouping that pays no attention to order.
- Counting Rule for Combinations

  \[
  \binom{n}{r} = \frac{n!}{r!(n-r)!}
  \]

  where \( n \) and \( r \) are whole numbers and \( n \geq r \). Other commonly used notations for combinations include \( nCr \) and \( \binom{n}{r} \).

### Useful Properties of Conditional Probabilities

#### Property 1. The Conditional Probability for Independent Events

- If \( A \) and \( B \) are independent events, then:
  \[
  \Pr(A \cap B) = \Pr(A) \Pr(B)
  \]

#### Property 2. The Multiplication Rule for Conditional Probabilities

In an experiment involving two non-independent events \( A \) and \( B \), the probability that both \( A \) and \( B \) occur can be found in the following two ways:

\[
\Pr(A \cap B) = \Pr(B)\Pr(A | B)
\]

or

\[
\Pr(A \cap B) = \Pr(A)\Pr(B | A)
\]

### Which Drug is Better?

<table>
<thead>
<tr>
<th></th>
<th>Women Drug I</th>
<th>Drug II</th>
<th>Men Drug I</th>
<th>Drug II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Success</td>
<td>200</td>
<td>10</td>
<td>19</td>
<td>1000</td>
</tr>
<tr>
<td>Failure</td>
<td>1800</td>
<td>190</td>
<td>1</td>
<td>1000</td>
</tr>
</tbody>
</table>

### Simpson’s Paradox: View I

<table>
<thead>
<tr>
<th></th>
<th>Women Drug I</th>
<th>Drug II</th>
<th>Men Drug I</th>
<th>Drug II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Success</td>
<td>200</td>
<td>10</td>
<td>19</td>
<td>1000</td>
</tr>
<tr>
<td>Failure</td>
<td>1800</td>
<td>190</td>
<td>1</td>
<td>1000</td>
</tr>
</tbody>
</table>

#### Drug II is better than Drug I

\( A = \{ \text{Using Drug I} \} \)

\( B = \{ \text{Using Drug II} \} \)

\( C = \{ \text{Drug succeeds} \} \)

\( \Pr(C | A) \sim 10\% \)

\( \Pr(C | B) \sim 50\% \)
Simpson’s Paradox: View II

Female Patient
A = {Using Drug I}
B = {Using Drug II}
C = {Drug succeeds}
Pr(C|A) ~ 20%
Pr(C|B) ~ 5%

Drug I is better than Drug II

Male Patient
A = {Using Drug I}
B = {Using Drug II}
C = {Drug succeeds}
Pr(C|A) ~ 100%
Pr(C|B) ~ 50%

Conditional Independence

- Event A and B are conditionally independent given C in case
  \[ \Pr(AB|C) = \Pr(A|C)\Pr(B|C) \]
- A set of events \( \{A_i\} \) is conditionally independent given C in case
  \[ \Pr(\bigcup_i A_i | C) = \prod_i \Pr(A_i | C) \]
- A and B are independent \( \neq \) A and B are conditionally independent

Example of conditional probability and partitions of a sample space

The set of events \( A_1, \ldots, A_k \) form a partition of a sample space if \( \bigcup_{i=1}^k A_i = S \). If the events \( A_1, \ldots, A_k \) partition S and if B is any other event in S (note that it is impossible for \( A_i \cap B = \emptyset \) for some i), then the events \( A_1 \cap B, A_2 \cap B, \ldots, A_k \cap B \) will form a partition of B.

Thus, \( B = (A_1 \cap B) \cup (A_2 \cap B) \cup \cdots \cup (A_k \cap B) \)
\[ \Pr( B ) = \sum_{i=1}^k \Pr( A_i | B ) \]

Finally, if \( \Pr( A_i ) > 0 \) for all i, then:
\[ \Pr( B ) = \sum_{i=1}^k \Pr( B | A_i ) \Pr( A_i ) \]

Example. What is the Probability of a Republican Vote?
\[ \Pr(\text{Rep. Vote}) = \Pr(\text{Rep. Vote} | \text{Rep. ID}) \Pr(\text{Rep. ID}) + \Pr(\text{Rep. Vote} | \text{Ind. ID}) \Pr(\text{Ind. ID}) + \Pr(\text{Rep. Vote} | \text{Demo. ID}) \Pr(\text{Demo. ID}) \]

Note: the definition for \( \Pr(B) \) defined above provides the denominator for Bayes’ Theorem.
Bayes’ Rule

Given two events $A$ and $B$ and suppose that $Pr(A) > 0$. Then

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

Example:

$Pr(R) = 0.8$

<table>
<thead>
<tr>
<th></th>
<th>$R$</th>
<th>$\neg R$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$W$</td>
<td>0.7</td>
<td>0.4</td>
</tr>
<tr>
<td>$\neg W$</td>
<td>0.3</td>
<td>0.6</td>
</tr>
</tbody>
</table>

R: It is a rainy day
W: The grass is wet
$Pr(R|W) = ?$

Bayes’ Rule: More Complicated

Suppose that $B_1, B_2, \ldots, B_k$ form a partition of $S$:

$$B_i \cap B_j = \emptyset, \quad \cup_{i=1}^{k} B_i = S$$

Suppose that $Pr(B_i) > 0$ and $Pr(A) > 0$. Then

$$Pr(B_i | A) = \frac{Pr(A \cap B_i)}{Pr(A)} = \frac{Pr(A | B_i) Pr(B_i)}{Pr(A)}$$

Bayes’ Rule: More Complicated

Suppose that $B_1, B_2, \ldots, B_k$ form a partition of $S$:

$$B_i \cap B_j = \emptyset, \quad \cup_{i=1}^{k} B_i = S$$

Suppose that $Pr(B_i) > 0$ and $Pr(A) > 0$. Then

$$Pr(B_i | A) = \frac{Pr(A | B_i) Pr(B_i)}{\sum_{j=1}^{k} Pr(A | B_j) Pr(B_j)}$$
A More Complicated Example

It rains
W The grass is wet
U People bring umbrella

Pr(UW|R) = Pr(U|R)Pr(W|R)
Pr(UW|¬R) = Pr(U|¬R)Pr(W|¬R)

Assessing Probability
(Frequency Approach)

Empirical probability—based on empirical data
For example:
• Toss an irregular die (probabilities unknown) 100 times and find that you get a 2 twenty-five times; empirical probability of rolling a 2 is 1/4.
• The empirical probability of an Earthquake in Bay Area by 2032 is 0.62 (based on historical data)
• The empirical probability of a lifetime smoker developing lung cancer is 15 percent (based on empirical data).

Relative Frequency Probability:
If some process is repeated a large number of times, n, and if some resulting event E occurs m times, the relative frequency of occurrence of E, m/n will be approximately equal to the probability of E. 

Pr(U|W) = ?
Pr(R) = 0.8
Subjective Probability

Probability measures the confidence that a particular individual has in the truth of a particular proposition.

Example:
In a certain population, 10% of the people are rich, 5% are famous, and 3% are both rich and famous. A person is randomly selected from this population. What is the chance that the person is
- not rich?
- rich but not famous?
- either rich or famous?

Intuitive Development

Subjective Probability

Probability measures the confidence that a particular individual has in the truth of a particular proposition.

Example:
In a certain population, 10% of the people are rich, 5% are famous, and 3% are both rich and famous. A person is randomly selected from this population. What is the chance that the person is
- not rich?
- rich but not famous?
- either rich or famous?

Why Axiomatic Approach?

Real World

Concept

Rectangle

Circle

Probability

What will be the Appropriate Axioms?

Andrey Nikolaevich Kolmogorov (Russian) (25 April 1903 – 20 October 1987)

In 1933, Kolmogorov published the book, Foundations of the Theory of Probability, laying the modern axiomatic foundations of probability theory.

The Axiomatic “Definition” of Probability

A probability distribution on a sample space $S$ is a specification of numbers $P(A_i)$ which satisfy the following three axioms, A1, A2 and A3.

A1. For any outcome $A_i$, $P(A_i) \geq 0$.
A2. $P(S) = 1$.
A3. For any infinite sequence of disjoint events $A_1, A_2, \ldots$:

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

Anomaly in Uncountable Set

- Previous plan is OK when $S$ is finite or countably infinite (discrete)
- For uncountably infinite $S$ additivity makes nonsense
  - Union of infinite events with probability 0 can make event of probability one; This makes analysis tough
  - E.g., for $\{0,1\}$ as $S$
    - Each singleton is of probability 0
    - Uncountable union of singletons constitute $S$ which is of probability 1