Probability and Statistics

MAT 271E

Ahmet H. KAYRAN

Section 1

(Probability Theory)

February 2019

Probability and Statistics

1

CHAPTER I

1. Basic Probability Concepts

Everyone accepts that the probability of an event is a number between 0 and 1.

$$0 \leq Pr\{A\} \leq 1$$

The probability theory gives us the basic tools for analyzing mathematical models for random phenomena.

A random experiment is specified by an experimental procedure and a set of measurements or observations.

(D) • Random Experiments:

An experiment that can result in different outcomes, even though it is repeated in the same manner every time, is called a RANDOM EXPERIMENT.

(D) • An Outcome:

of an experiment is any possible observation of the random experiment. It may be different even when it is repeated under the same conditions, unlike deterministic experiments. Only one outcome occurs and it is indecomposable.

misoro

(D) • **The Sample Space**

of an experiment is mutually-exclusive set of all possible outcomes.

(D) • **An Event**

is a set of outcomes of an random experiment.

• **$Pr\{A\}$**

The probability of an event A is a number $Pr\{A\}$ assigned to this event.

Remark:

If the random experiment is performed n times and the event A occurs n_A times, then, with a high degree of certainty, the realative frequency $\frac{n_A}{n}$ of the occurance of A is close to $Pr\{A\}$:

$$Pr\{A\} \cong \frac{n_A}{n}$$

provided that n is sufficiently large.

misoro

Example 1:

- Suppose we roll a six-sided die and observe the number of dots on the side facing upwards.

- We can label these outcomes $i = 1, 2, \ldots, 6$ where i denotes the outcome that i dots appear on the face up.

- The sample space of this random experiment is

$$S_1 = \{1, 2, 3, 4, 5, 6\}.$$

- Each subset of S_1 is an event. Examples of events are given as follows:

- The event $A = \{\text{The roll is even}\}$
$= \{2, 4, 6\}$

- The event $B = \{\text{the roll is odd}\}$
$= \{1, 3, 5\}$

- The event $C = \{\text{the roll is the square of an integer}\}$
$= \{1, 4\}$

misoro

- The event $D = \{$ Roll 4 or higher $\}$

$$= \{4, 5, 6\}$$

- The event $E = \{$ Roll 6 $\}$

$$= \{6\}.$$

Question:

- How many events can be defined for this sample space?

- The set of events is

$$\Omega = \{A, B, C, D, \dots\}$$

and the number of events at this set is

$$n(\Omega) = 2^6 = 64.$$

Notation:

The capital letters $A, B, C, \Phi, \Omega, \dots$ shall be used to denote events, and lower-case letters $a, b, c, \phi, w, \dots$ to denote their elements.

misoro

Example 2:

Flip a coin and record the results, head or tail. The sample space is

$$S_2 = \{h, t\}.$$

Example 3:

- Flip a coin three times and observe the squence of heads and tails:
- The sample space is

$$S_3 = \{hhh, hht, hth, thh, htt, tht, tth, ttt\}.$$

- The same random experimental procedure, but, we can make different measurements.

Example 4:

- Flip a coin three times and observe the number of heads.

$$S_4 = \{0, 1, 2, 3\}.$$

We can classify sample spaces and therefore random experiments with respect to its number of outcomes. Indeed, a sample

space can be finite, countably infinite or uncountably infinite outcomes.

(D) • Discrete Sample Space

has a finite number of outcomes or countably infinite number of outcomes.

• Toss a coin until a head occurs and recond the number of the toss. $S' = \{1, 2, \dots\}$ has a countably infinite ocutcomes.

(D) • Continuous Sample Space

has uncountable outcomes in the sample space.

Example 5:

Wait for someone to make a phone call and observe the duration of call in minutes. An outcome x is a nonnegative real number. The sample space is

$$S_5 = \{ x \mid x \geq 0 \}.$$

- An example of an event for this example;

$$A = \left\{ \begin{array}{l} \text{The Phone call lasts} \\ \text{longer than five minutes} \end{array} \right\}$$

$$= \{ x \mid x > 5 \}.$$

- This is an example for the Continuous sample space.

D) • Event Space

- We know that a collection of outcomes, namely, a subset of S, is called an "event".
- The event space is a set of events, while the sample space is a set of outcomes.
- For the example 3, the set $B = \{B_0, B_1, B_2, B_3\}$ is an event space. For example, $B_i = \{\text{outcomes with } i \text{ heads}\}$ and $B_1 = \{tth, tht, htt\}$.

misoro

(D) PROBABILITY SPACES

- In order to form a probability model, the final step, for each element of the sample space x_n, we correspond a probability value,

$$P_n = Pr\{x_n\}.$$

- Indeed, if we have a discrete sample space, $S = \{x_1, x_2, \dots\}$, we correspond $P_n = Pr\{x_n\}$ number for each element of S. These probabilities must satisfy the following conditions:

$$0 \leq P_n \leq 1 \quad \text{and} \quad P_1 + P_1 + \dots = 1$$

where the summation consists of all outcomes of S.

<u>Example:</u> Uniform Probability Space.

Has equally likely outcomes.

If $P_1 = P_2 = \dots = P_N$, for a sample space with a finite N points, if each outcome x_i equally likely,

$$S = \{ x_1 , x_2 , \dots x_N \}$$

$$\downarrow \qquad \downarrow \qquad \downarrow$$

$$P_1 = \frac{1}{N} , P_2 = \frac{1}{N} , \dots , P_N = \frac{1}{N}$$

and hence

$$P_r\{x_1\} + P_r\{x_2\} + \dots + P_r\{x_n\} = 1$$

Condition is satisfied.

(D) Probability of an Event:

The probability of an event

$$A = \{ s_1 , s_2 , \dots , s_m \}$$

is the sum of all probabilities of the outcomes in the event:

$$P_r\{A\} = \sum_{i=1}^{m} P_r\{s_i\} .$$

Example:

The event $A = \{$ the roll even $\}$

$$= \{ 2 , 4 , 6 \}$$

For $Pr\{s_i\} = \frac{1}{6}$, $i = 1, \dots, 6$,

$$Pr\{A\} = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2} .$$

- The event $D = Pr\{\text{Roll 4 or higher}\}$

$$= Pr\{4\} + Pr\{5\} + Pr\{6\}$$

$$= \frac{1}{2} .$$

There are important events:

- <u>Certain (Definite) Event:</u> $(A = S)$
  - If we observe the same event when we repeat the random experiment, we call this event definite event.
  - At this situation, event $A = S$ is the sample space, universal set or the set of all outcomes.

- <u>Impossible (Null) Event:</u> $(A = \phi)$
  - If we never observe one event when we repeat the random experiment, we call this event impossible event. This is the empty set, Φ. S always occurs, ϕ never occurs.

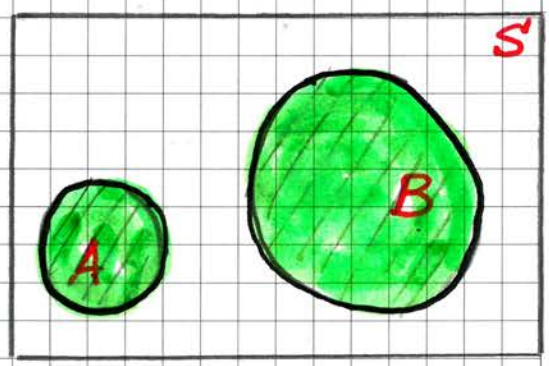
- If the event $A = \phi$,

$$Pr\{A\} = Pr\{\phi\} = 0.$$

One can form other events using basic <u>set operations</u>:

- **Mutually Exclusive Events (or Disjoint Events)**

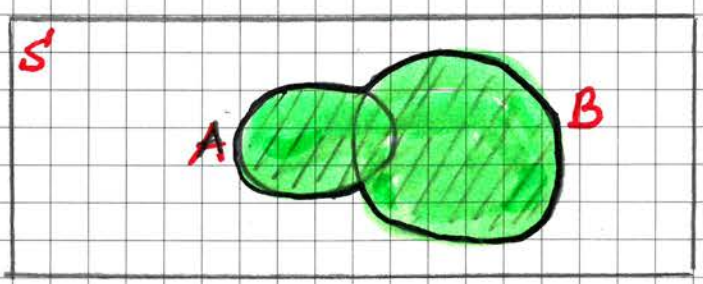
- Two events, denoted as A and B, such that $A \cap B = \phi$ are said to be mutually exclusive.



- $A \cap B$ is impossible event, $Pr\{A \cap B\} = 0$, and $Pr\{A \cup B\} = Pr\{A\} + Pr\{B\}$.
- Mutually exclusive events can not occur simultaneously.

- **The Union of two events A and B**

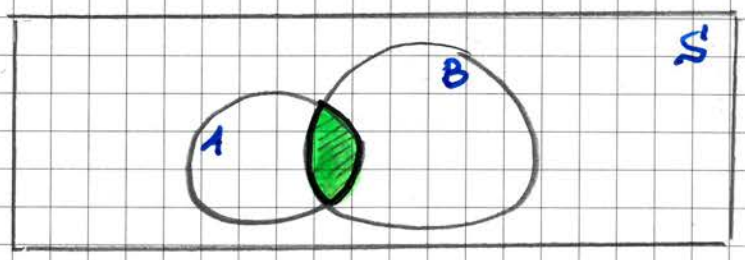
It is defined by $A \cup B$.
It occurs if either A, or B, or both A and B occur.



misoro

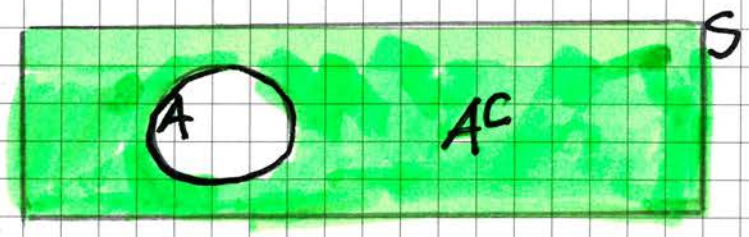
The intersection of two events A and B

- The event $A \cap B$ occurs when both A and B occur.



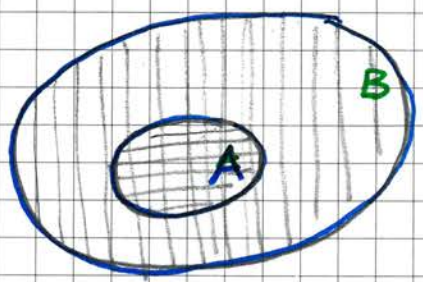
The complement of an event A

- The Complement of A is denoted A^c. A^c occurs when A does not occur.



An event A is a subset of an event B

- That is, $A \subset B$, A is a subset of B. If A occurs than B also occurs.



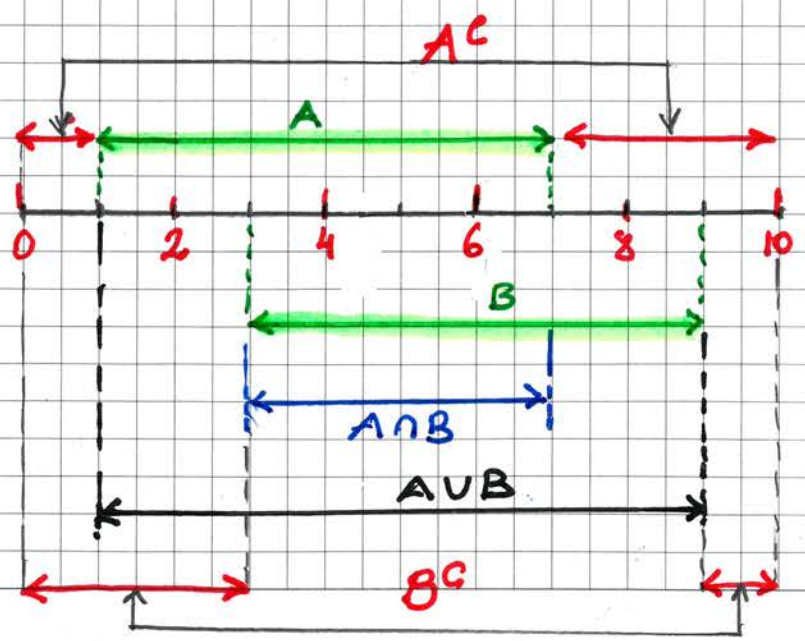
- $Pr\{B|A\} = 1$, the probability of an event B under the knowledge the the outcome will be an event A. B is an inevitable definite event!

2. SET THEORY OPERATIONS [T.T. Soong, 2005]

- **Notation:**
  - The intersection or product of A and B, written as $A \cap B$, or simply AB.
  - If $AB = \phi$, sets A and B contain no common elements, and we call A and B disjoint. The symbol "+" will be reserved to denote the union of two disjoint sets when it is advantageous to do so.

- **Example:**

  Let S be the space consisting of a real-line segment from 0 to 10 and let A and B sets of the real-line segments from 1-7 and 3-9, respectively.



- **Observation:**

  A set and its complement are always disjoint.

- The Generalization of union and intersection:

- $$A_1 \cup A_2 \cup \cdots \cup A_n = \bigcup_{j=1}^{n} A_j$$

all elements belonging to one or more of sets A_j, $j=1,2,\ldots,n$.

- $$A_1 A_2 \cdots A_n = \bigcap_{j=1}^{n} A_j$$

is the set of all elements common to all A_j, $j=1,2,\ldots,n$.

- If $A_i A_j = \phi$ for every i,j $(i \neq j)$, the sets A_j, $j=1,2,\ldots,n$, are disjoint.

- Properties of Set Operations:

(operations are associative, commutative, and distributive)

- $(A \cup B) \cup C = A \cup (B \cup C) = A \cup B \cup C$
- $A \cup B = B \cup A$,
- $(AB)C = A(BC) = ABC$,
- $AB = BA$,
- $A(B \cup C) = (AB) \cup (AC)$.

- We have the following relations:
  - $A \cap A = AA = A$
  - $A \cup \phi = A$,
  - $A\phi = \phi$,
  - $A \cup S = S$,
  - $AS = A$,
  - $A \cup A^c = S$,
  - $AA^c = \phi$.

- The following useful relations can be easily verified using Venn diagrams:
  - $A \cup (BC) = (A \cup B)(A \cup C)$
  - $A \cup B = A \cup (A^c B) = A + (A^c B)$,
  - $(A \cup B)^c = A^c B^c$,
  - $(AB)^c = A^c \cup B^c$,
  - $\left(\bigcup_{j=1}^{n} A_j\right)^c = \bigcap_{j=1}^{n} A_j^c$,
  - $\left(\bigcap_{j=1}^{n} A_j\right)^c = \bigcup_{j=1}^{n} A_j^c$.

  De Morgan's law.

misoro

Remark:

The description of sample space, sample points, and events fit nicely into the framwork of set theory. All relations between events in the probability theory can be described by sets and set operations. The corresponding statements in the set theory and probability theory are given in the following table.

SET THEORY	Probability Theory
Space, S	Sample Space, definite event
Empty set, ϕ	Impossible event
Elements $a, b, \ldots$	Sample points $a, b, \ldots$ (or simple events)
Sets $A, B, \ldots$	Events $A, B, \ldots$
A	Event A occurs
A^c	Event A does not occur
$A \cup B$	At least one of A and B occurs
AB	Both A and B occur
$A \subset B$	A is a subevent of B (namely, the occurrence of A necessarily implies the occurrence of B)
$AB = \phi$	A and B are mutually exclusive. (They cannot occur simultaneously).

misoro

3. Axioms of Probability

We will build our entire theory of probability on the following three axioms: A probability measure $Pr\{\cdot\}$ is a function that maps events in the sample space to real numbers such that

Axiom 1: For any event A, $Pr\{A\} \geq 0$.

Axiom 2: $Pr\{S\} = 1$.

Axiom 3: For any mutually exclusive events A and B, $A \cap B = \phi$,

$$Pr\{A \cup B\} = Pr\{A\} + Pr\{B\}.$$

Observations:

- Axioms 1 and 2 simply establish a probability as a number between 0 and 1.
- Axiom 3 states that probability of the union of mutually exclusive events is the sum of individual probabilities.

misoro

Some Consequences of the Axioms:

Here we list some properties of probabilities that follow directly from the three axioms. You can find the proofs in the text book.

Theorem 1:

The complement of an event A, denoted by A^c,

$$Pr\{A^c\} = 1 - Pr\{A\}.$$

hint: $A \cap A^c = \phi$
$A \cup A^c = S$

Theorem 2:

The probability of the impossible event is zero;

$$Pr\{\Phi\} = 0.$$

hint: $A = S$
$A^c = \phi$

Theorem 3:

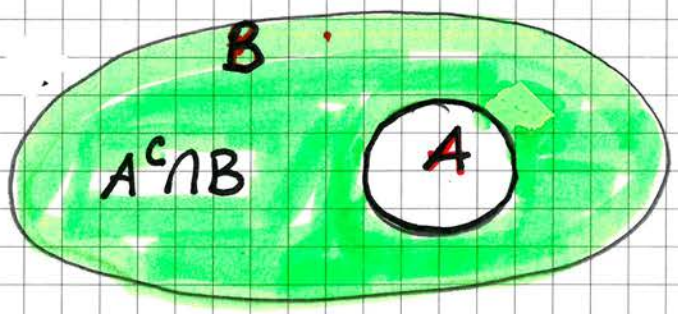
The probability of an event always is less than or equal to one,

$$Pr\{A\} \leq 1.$$

hint: Theorem 1 and Axiom 1

Theorem 4:

If $A \subset B$, then $Pr\{A\} \leq Pr\{B\}$.



Proof:

$$Pr\{B\} = \underbrace{Pr\{A^c \cap B\}}_{\geq 0 \text{ from Axiom 1}} + Pr\{A\}$$

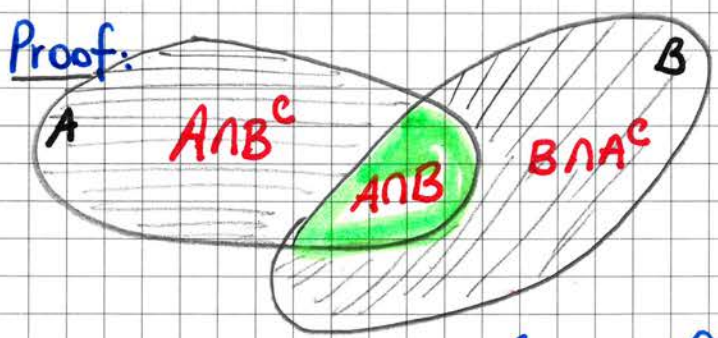
Therefore, $Pr\{B\} \geq Pr\{A\}$.

Theorem 5:

For any events A and B, not necessarily disjoint, namely $A \cap B \neq \phi$,

$$Pr\{A \cup B\} = Pr\{A\} + Pr\{B\} - Pr\{A \cap B\}.$$

Proof:



hint:

- $Pr\{A\} = Pr\{A \cap B^c\} + Pr\{A \cap B\}$
- $Pr\{B\} = Pr\{B \cap A^c\} + Pr\{A \cap B\}$

$$Pr\{A \cup B\} = Pr\{A \cap B^c\} + Pr\{A \cap B\} + Pr\{B \cap A^c\}$$

misoro

Example [T.T. Soong]:

- Consider an experiment of counting the number of left-turning cars at an intersection in a group of 100 cars.

- The possible outcomes, namely, possible numbers of left-turning cars are $0, 1, 2, \ldots, 100$. Then, the sample space S is

$$S = \{0, 1, 2, \ldots, 100\}.$$

Each element of S is a sample point or a possible outcome.

- The subset

$$A = \{0, 1, 2, \ldots, 50\}$$

is the event that there are 50 or fewer cars turning left.

- The subset

$$B = \{40, 41, \ldots, 60\}$$

is the event of 60 or fewer cars turning left.

- The set

$$A \cup B = \{1, 2, \ldots, 60\}$$

is the event of 60 or fewer cars turning left.

- The set

$$A \cap B = \{40, 41, \ldots, 50\}$$

misoro

is the event that the number of left-turning cars is between 40 and 60 (inclusive).

- Let

$$C = \{80, 81, \ldots, 100\}.$$

- Events A and C are <u>mutually exclusive</u> since they cannot occur simultaneously. Hence disjoint sets are mutually exclusive events in probability theory.

- $Pr\{A\}$, $Pr\{B\}$, $Pr\{C\}$: are assumed to be known. Now we wish to determine $Pr\{A \cup B\}$ and $Pr\{A \cup C\}$.

- $Pr\{A \cup C\}$, the probability of having either 50 or fewer cars turning-left or between 80 to 100 cars turning left is

$$Pr\{A \cup C\} = Pr\{A\} + Pr\{C\}.$$

This follows from Axiom 3, since A and C are mutually exclusive events.

- However, $Pr\{A \cup B\}$, the probability of having 60 or fewer cars turning cars turning left, is found from

misoro

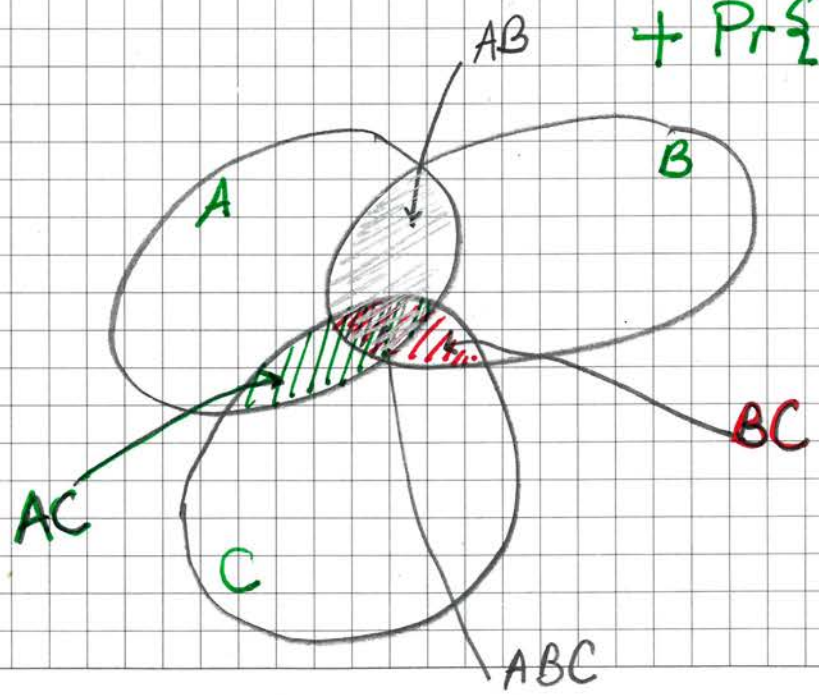
$$Pr\{A \cup B\} = Pr\{A\} + Pr\{B\} - Pr\{AB\}.$$

This probability cannot be calculated by the given information. We need an additional information, $Pr\{AB\}$, which is the probability of having between 40 and 50 cars turning left.

Remark:

The theorem 5 can be extended as follows:

$$Pr\{A \cup B \cup C\} = Pr\{A\} + Pr\{B\} + Pr\{C\} - Pr\{A \cap B\} - Pr\{A \cap C\} - Pr\{B \cap C\} + Pr\{A \cap B \cap C\}.$$



hint:
$Pr\{(A \cup B) \cup C\}$

4. CONDITIONAL PROBABLITY

Definition:

The conditional probability of the event A given the occurrence of the event B is

$$\Pr\{A \mid B\} = \frac{\Pr\{A \cap B\}}{\Pr\{B\}}.$$

outcome is included in $A \cap B$

"Conditional Probability of A given B"

Normalizing Factor

Observations:

(1) Conditional probability is defined only when $\Pr\{B\} > 0$.

$\Pr\{B\} = 0$ means that it is certain that B never occurs. In this case, it is illogical to speak of the probability of A given that B occurs.

(2) Conditional probability satisfies the axioms of probability:

Axiom 1: $\Pr\{A \mid B\} \geq 0$

Axiom 2: $\Pr\{S \mid B\} = 1$

Axiom 3: $\Pr\{A_1 \cup A_2 \mid B\} = \Pr\{A_1 \mid B\} + \Pr\{A_2 \mid B\}$

with $A_1 \cap A_2 = \phi$.

misoro

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

and

$$Pr\{B|A\} = \frac{Pr\{A \cap B\}}{Pr\{A\}} .$$

Therefore, we can write as follows

$$Pr\{A \cap B\} = Pr\{A|B\}\, Pr\{B\}$$
$$= Pr\{B|A\}\, Pr\{A\} .$$

(4) For $Pr\{A\} > 0$ and $Pr\{B\} > 0$,

$$\frac{Pr\{A|B\}}{Pr\{A\}} = \frac{Pr\{B|A\}}{Pr\{B\}}$$

or

$$Pr\{B|A\} = \frac{Pr\{A|B\}\, Pr\{B\}}{Pr\{A\}}$$ Bayes' theorem.

In many situations, we have advance information about $Pr\{A|B\}$ and need to calculate $Pr\{B|A\}$. To do so, we have this formula, known as Bayes' theorem.

misoro

(5) For events A, B and C,

$$\underbrace{Pr\{A \cap B \cap C\}}_{\text{joint probability}} = \underbrace{Pr\{C \mid A \cap B\}}_{\text{Conditional Probablities}} \underbrace{Pr\{B \mid A\}}_{} \underbrace{Pr\{A\}}_{\text{Marginal Probability.}}$$

- This relationship is called the <u>Probability chain rule</u>.

- The joint probability is determined by conditional probabilities and the marginal probability.

<u>Example:</u> [T.T. Soong]

In a card game, determine the probability of drawing, without replecement, two aces in succession.

Let A_1 be the event that the first card drawn is an ace, and similarly for A_2. We wish to compute $Pr\{A_1 \cap A_2\}$:

$$Pr\{A_1 \cap A_2\} = Pr\{A_2 \mid A_1\} \, Pr\{A_1\}$$

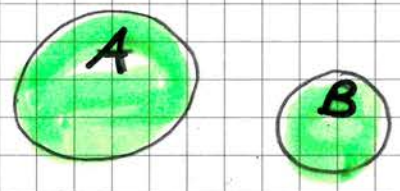
Now, $Pr\{A_1\} = 4/52$ and $Pr\{A_2 \mid A_1\} = 3/51$ (there are 51 cards left and three of them are aces).

Therefore, $Pr\{A_1 \cap A_2\} = \frac{3}{51}\left(\frac{4}{52}\right) = \frac{1}{221}$.

misoro

(6) If A and B are mutually exclusive events, $A \cap B = \phi$.
For this case,

$$Pr\{A \mid B\} = 0$$

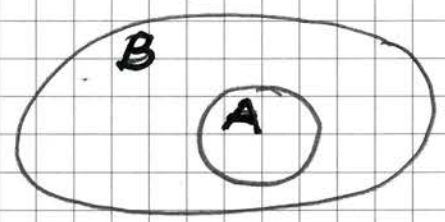


and

$$Pr\{B \mid A\} = 0.$$

Another words, if event A occurs, B is an impossible event and vice versa.

(7) If event A is a subset of event B, $A \subset B$,



$$Pr\{B \mid A\} = 1.$$

Indeed, when A occurs B is a definite event.

Example:

- In a digital communication system a "0" or "1" are transmited to a receiver.
- Typically, either bit is equally likely to occur so that a prior probability of 1/2 is assumed. Namely,

$$Pr\{0\} = Pr\{1\} = \frac{1}{2}.$$

misoro

- At the receiver, due to the channel noise, a decoding error can be made. A "0" may be mistaken for a "1" and vice versa.

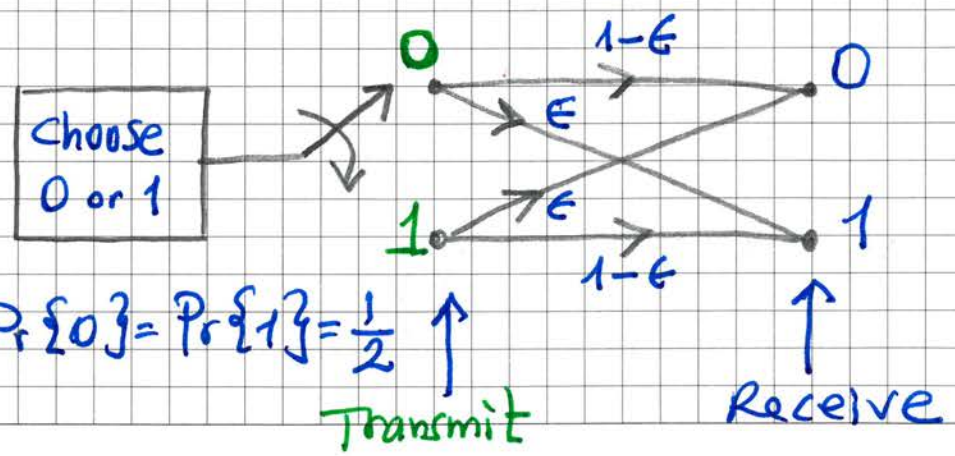
- Defining the probability of decoding a "1" when a "0" transmitted as ϵ and a "0" when a "1" is transmitted also as ϵ.

$$\Pr\{\text{error} \mid \text{"1" transmitted}\} = \epsilon$$

$$\Pr\{\text{error} \mid \text{"0" transmitted}\} = \epsilon .$$

- We are interested in the overall probability of an error.

$$\Pr\{\text{error}\} = \Pr\{\text{error} \mid \text{"1" transmitted}\}\Pr\{\text{"1" transmitted}\}$$
$$+ \Pr\{\text{error} \mid \text{"0" transmitted}\}\Pr\{\text{"0" transmitted}\}.$$
$$= \epsilon\frac{1}{2} + \epsilon\frac{1}{2} = \epsilon.$$



WARNING !

- Conditional probabilities can be quite tricky, they sometimes produce counterintuitive results.

- A famous instance of this is the "Monty Hall Problem" or "Let's Make a Deal Problem".

MONTY HALL PROBLEM

- About 30 years ago there was a TV game show "Let's make a deal". The game show host, Monty Hall, would present the contestant three closed doors. Behind one door was a new car, while the others had less desirable prizes.

- The contestant would first have the opportunity to choose a door, but it would not open it. Monty would then choose one of the remaining doors and open it. Since Monty would have knowledge of which door lead to the car, he would always choose a door with a less desirable price.

- At this point contestant was faced with two close doors, one of which led to the car and other to less desirable prizes.

- The contestant was given the option of either opening the door he had orginally chosen or deciding to open the other door

- What should the contestat do?

- The answer is suprisingly, is that by choosing to switch doors he has a probability of 2/3 of winning the car.

- If he stays with his original choise, then the probability is only 1/3.

- However, most people would say that regardless of which strategy she decided upon, his probability of winning the car 1/2.

- You can read the proof in the textbook. [kayran]

misoro

STATISTICALLY INDEPENDENT EVENTS

(D) Two events A and B are said to be statistically independent (or sometimes just independent):

If $\Pr\{A|B\} = \Pr\{A\}$.

- If this is true, than

$$\Pr\{A|B\} = \frac{\Pr\{A \cap B\}}{\Pr\{B\}} = \Pr\{A\}.$$

This equation gives the condition for statistical independence,

$$\underbrace{\Pr\{A \cap B\}}_{\text{joint probability}} = \Pr\{A\}\Pr\{B\}.$$

Marginal probabilities

Example:

Independence does not mean one event does not affect another event.

- If a fair die is tossed,

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

- The probability of a 2 or a 3 is

misoro

$$Pr\{A = \{2,3\}\} = \frac{1}{3}.$$

Now, assume we know that the outcome is an even number or $B = \{2,4,6\}$.

Recomputing the probability,

$$Pr\{A|B\} = \frac{Pr\{A \cap B\}}{Pr\{B\}} = \frac{Pr\{2\}}{Pr\{2,4,6\}}$$

$$= \frac{1/6}{1/2} = \frac{1}{3} = Pr\{A\}.$$

Hence, two events A and B are independent.

- In this example, the knowledge of B occurring has affected the possible outcomes.

- In particular, the event $A \cap B = \{2\}$ has half as many elements as A, but reduced sample space $S' = B = \{2,4,6\}$ also has half as many elements.

- Indeed,

misoro

- $$\Pr\{A \mid S\} = \frac{\Pr\{A \cap S\}}{\Pr\{S\}} = \frac{\Pr\{A\}}{1}$$
$$= \Pr\{A\}$$

- $$\Pr\{A \mid B\} = \Pr\{A \mid S'\} = \Pr\{A\}.$$

REMARKS:

(1) The condition for the event A is the independent of the event B is

$$\Pr\{A \cap B\} = \Pr\{A\}\Pr\{B\}.$$

Therefore, we need only know the marginal probabilities or $\Pr\{A\}$ and $\Pr\{B\}$ to determine the joint probability $\Pr\{A \cap B\}$.

(2) In practice, this property turns out to be very useful. It is important to observe that the statistical independence has symmetry property, as we might expect. If A is independent of B, then B must be independent of A.

Since

$$Pr\{B \mid A\} = \frac{Pr\{B \cap A\}}{Pr\{A\}} \leftarrow \text{definition}$$

$$= \frac{Pr\{A \cap B\}}{Pr\{A\}} \leftarrow \text{Commutative Property}$$

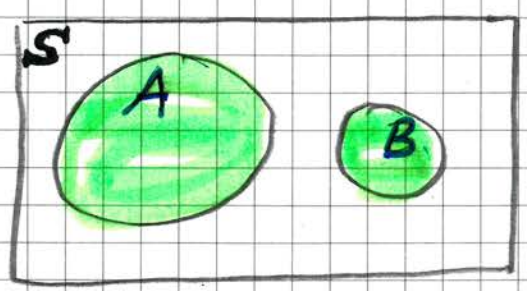
$$= \frac{Pr\{A\}\, Pr\{B\}}{Pr\{A\}} \leftarrow \text{A is independent of B}$$

$$= Pr\{B\}.$$

Therefore, we can say that the events A and B are statistically independent of each other.

(3) Statistically Independent events are different than mutually exclusive events.

- If A and B are mutually exclusive events, when B occurs, the event A cannot occur.



Therefore,

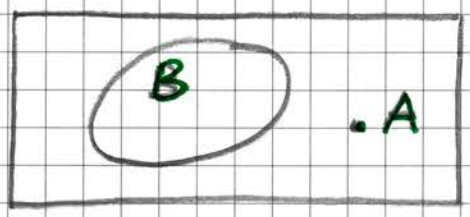
$$Pr\{A \mid B\} = 0.$$

A and B are not independent!

- If A and B are statistically independent and B occurs, then

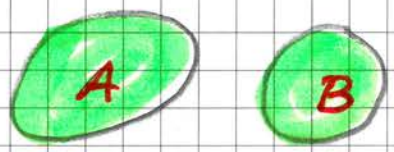
$$Pr\{A|B\} = Pr\{A\}.$$

- If $Pr\{A\}=0$, events are mutually exclusive (since $A \cap B = \phi$) and independent (since $Pr\{A \cap B\} = 0$ and $Pr\{A\}\,Pr\{B\} = 0 \cdot Pr\{B\} = 0$).



$Pr\{A\}=0$, $Pr\{B\} \neq 0$
$A \cap B = \phi$

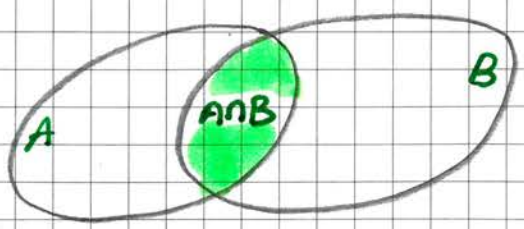
- Finally, we can state that,
  - Mutually exclusive events :



$$Pr\{A \cup B\} = Pr\{A\} + Pr\{B\}.$$

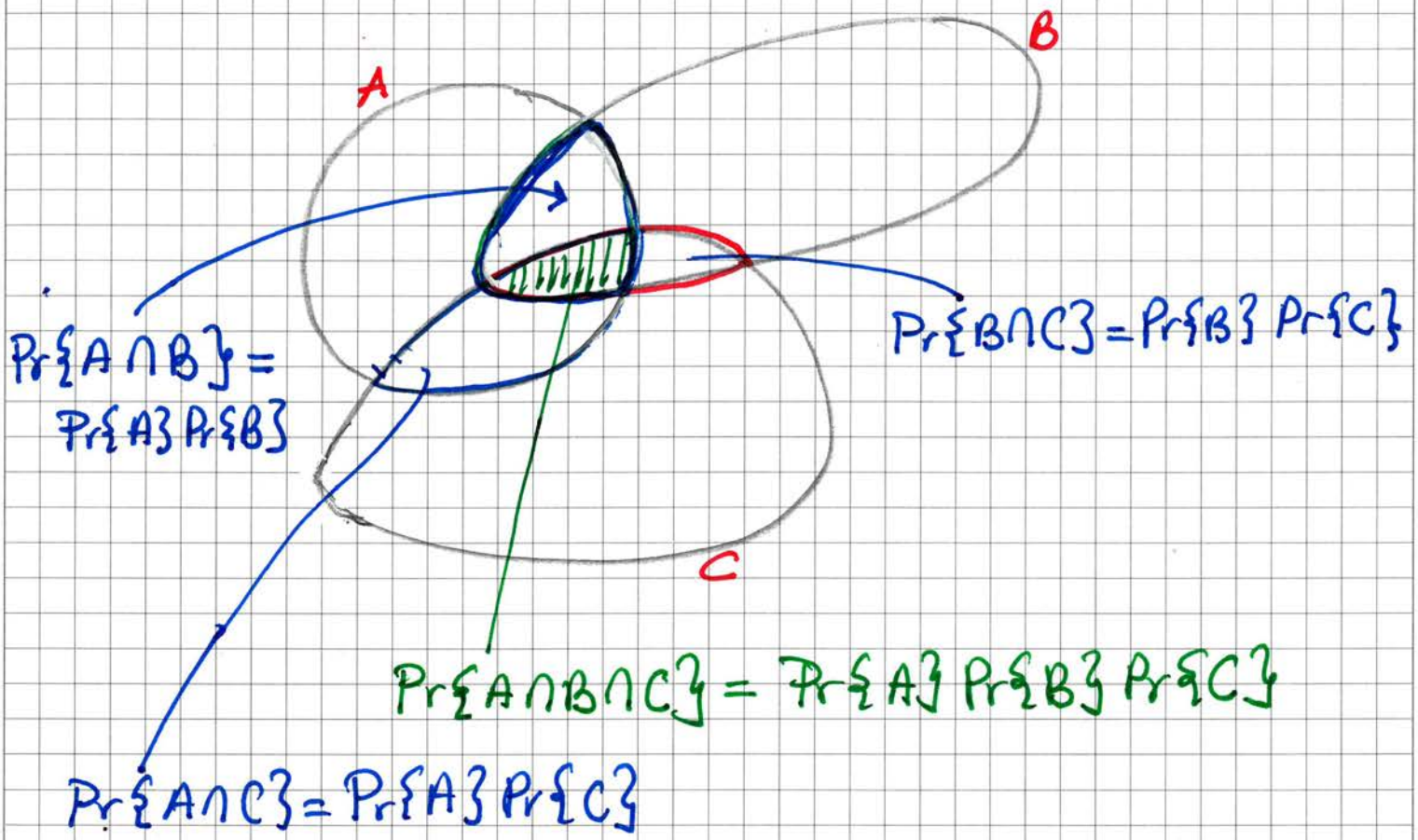
  - Statistically independent events:

$$Pr\{A \cap B\} = Pr\{A\}\,Pr\{B\}.$$



misoro

(4) Extension of the idea of statistical independence to three events:



If we satisfy these 4 conditions, we have

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

- The Generilization of statistical indepence for K events requires

$$2^K - K - 1$$

condition must be satisfied. Indeed, for $K=2$ events, we have one conditions and for $K=3$ we have 4 conditions.

Example [T. T. Soong]:

- In a large number of trails of a random experiment, let n_A and n_B be, respectively, the numbers of occurrences of two outcomes A and B, and let n_{AB} be the number of times both A and B occur.

- Using relative frequency interpretation, ratios

$$\frac{n_A}{n} \xrightarrow{\text{tends to}} Pr\{A\}$$

and

$$\frac{n_B}{n} \xrightarrow{\text{tends to}} Pr\{B\} \quad , \text{ as } n \text{ becomes large.}$$

- Similarly

$$\frac{n_{AB}}{n} \xrightarrow{\text{tends}} Pr\{AB\}.$$

- Assume the event A is realized. If A and B are independent events, we expect that the ratio

$$\frac{n_{AB}}{n_A} \xrightarrow{\text{also tends to}} Pr\{B\} \cong \frac{n_B}{n}$$

misoro

This then gives,

$$\frac{n_{AB}}{n} \cong \left(\frac{n_A}{n}\right)\left(\frac{n_B}{n}\right)$$

or, in the limit as n becomes large,

$$Pr\{A \cap B\} = Pr\{A\}\, Pr\{B\}.$$

This is the definition of independence.

MULTIPLE EXPERIMENTS

Let us consider, a random experiment can be viewed as a succession of subexperiments, $E_1, E_2, \ldots, E_n$. If we assume that the subexperiments are independent. We would expect the corresponding events $B_1, B_2, \ldots, B_n$ are also independent events. Therefore the joint event

$$B = B_1 \cap B_2 \cap \cdots \cap B_n$$

has the probability,

$$Pr\{B\} = Pr\{B_1 \cap B_2 \cap \cdots \cap B_n\}$$
$$= Pr\{B_1\}\, Pr\{B_2\} \cdots Pr\{B_n\}.$$

misoro

Example [A. Leon Garcia]:

Suppose that 10 numbers are selected at random from the interval [0,1]. Find the probability that the first 5 numbers are less than 1/4 and the last 5 numbers are greater than 1/2. Let $x_1, x_2, \ldots, x_{10}$ be the sequence of 10 numbers, then the events of interest are

$$A_k = \left\{ x_k < \frac{1}{4} \right\} \quad \text{for} \quad k = 1, \ldots, 5$$

$$A_k = \left\{ x_k > \frac{1}{2} \right\} \quad \text{for} \quad k = 6, \ldots, 10.$$

If we assume that each selection of a number is independent of the other selections, then

$$Pr\{A_1 \cap A_2 \cap \cdots \cap A_{10}\} = Pr\{A_1\} Pr\{A_2\} \cdots Pr\{A_n\}$$

$$= \left(\frac{1}{4}\right)^5 \left(\frac{1}{2}\right)^5.$$

• PARTIONS, TOTAL PROBABILITY AND BAYES' THEOREM

A partition of the sample space may be defined as follows:

Definitions:

If $B_1, B_2, \ldots, B_k$ are disjoint, mutually exclusive subsets of a sample space S,

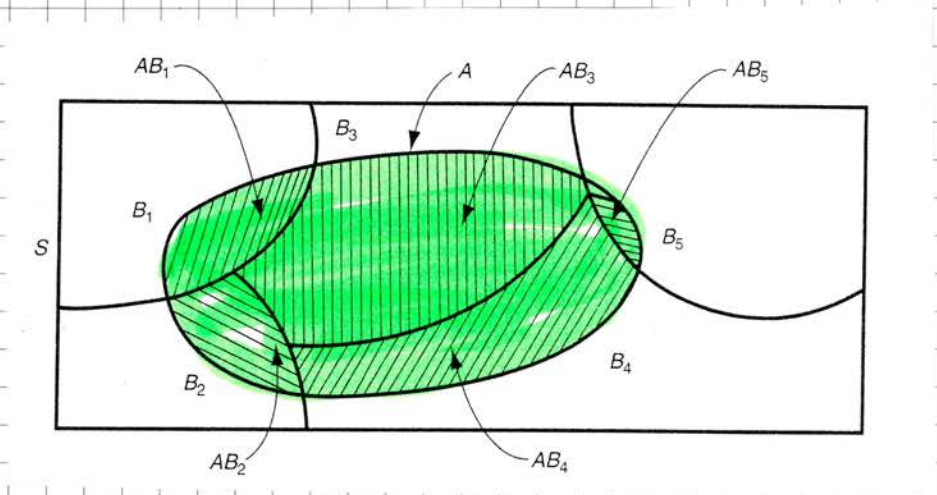
$$B_i \cap B_j = \phi \quad , \quad \text{if } i \neq j$$

and if the union of all these subsets form the sample space

$$B_1 \cup B_2 \cup \cdots \cup B_k = S.$$

We call these events as a partition of the sample space. When the experiment is performed, one and only one of the event B_i, occurs if we have a partition of S.

An event A at the subspace S:



An event A at the subspace S.

- A is an arbitrary event with respect to the sample space S. Then, we may write,

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

So that

$$Pr\{A\} = Pr\{A \cap B_1\} + Pr\{A \cap B_2) + \cdots + Pr\{A \cap B_k\},$$

From this figure, we can see this situation. Then the total probability is given by the conditional probabilities and the marginal probabilities:

$$Pr\{A\} = Pr\{B_1\} Pr\{A|B_1\} + \cdots + Pr\{B_k\} Pr\{A|B_k\}$$

$$= \sum_{i=1}^{k} Pr\{B_i\} Pr\{A|B_i\}.$$

TOTAL PROBABILITY

misoro

- In many practical problems, the probability A, $Pr\{A\}$ cannot be computed directly.

- Another important result of the total probability low is known as Bayers' theorem:

BAYERS' THEOREM

If $B_1, B_2, \ldots, B_k$ constitute a partition of the sample space S and A is an arbitrary event on S, then for $r = 1, 2, \ldots, k$, we have

$$Pr\{B_r \mid A\} = \frac{Pr\{B_r \cap A\}}{Pr\{A\}}$$

A posteriori Probability

$$= \frac{Pr\{B_r\}\, Pr\{A \mid B_r\}}{\sum_{i=1}^{k} Pr\{B_i\}\, Pr\{A \mid B_i\}}$$

A priori information $Pr\{B_i\}$ and $Pr\{B_i \mid A\}$

for any r where events B_i represent a set of mutually exclusive events.

Example [T.T. Soong]:

We assume that, for a given binary channel, 40% of the time a "1" is transmitted; the probability that a transmitted "0" correctly received is 0.90, and the probability that a transmitted "1" is correctly received is 0.95. Determine

(a) The probability of a "1" being received, and

(b) given a "1" is received, the probability that 1 was transmitted.

Answer: Let

A = event that 1 is transmitted,

A^c = event that 0 is transmitted,

B = event that 1 is received,

B^0 = event that 0 is received.

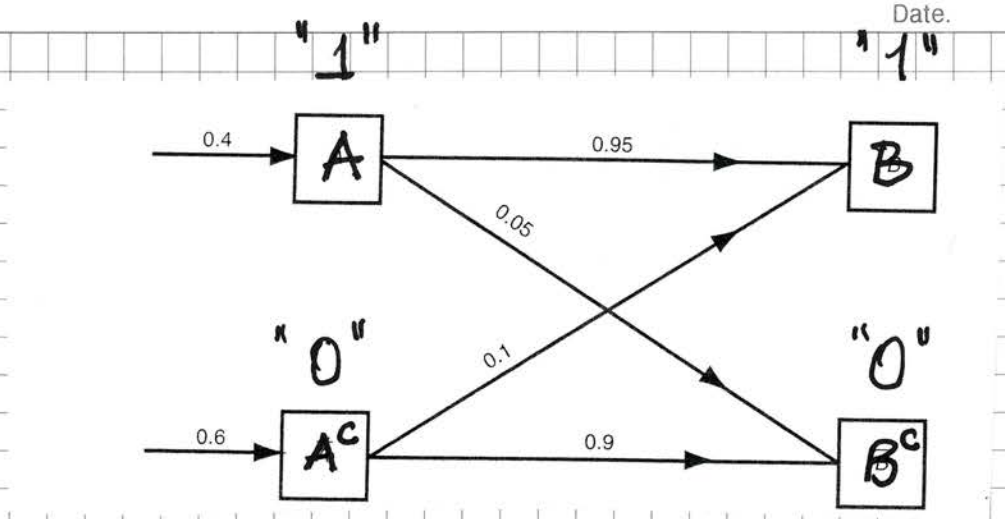
The information given in the problem statement tells us "A priori information";

$Pr\{A\} = 0.4$ $\quad$ $Pr\{A^c\} = 0.6$

$Pr\{B|A\} = 0.95$ $\quad$ $Pr\{B^c|A\} = 0.05$

$Pr\{B^c|A^c\} = 0.90$ $\quad$ $Pr\{B|A^c\} = 0.10$.

misoro



The information "a priori" are represented diagrammatically in this figure.

(A) We wish to find $Pr\{B\}$.
Since A and A^c are mutually exclusive, from the total probability,

$$Pr\{B\} = Pr\{B|A\}\,Pr\{A\} + Pr\{B|A^c\}\,Pr\{A^c\}$$

$$= 0.95\,(0.4) + 0.1\,(0.6) = 0.44.$$

(B) $Pr\{A|B\}$ The conditional probability of "1" transmitted, given a "1" received.
Using Bayes' theorem:

$$Pr\{A|B\} = \frac{Pr\{B|A\}\,Pr\{A\}}{Pr\{B\}}$$

$$= \frac{(0.95)(0.4)}{0.44} = 0.863.$$

TREE DIAGRAMS

We can add a second stage to the communication channel, with the earlier figure showing all associated probabilities.

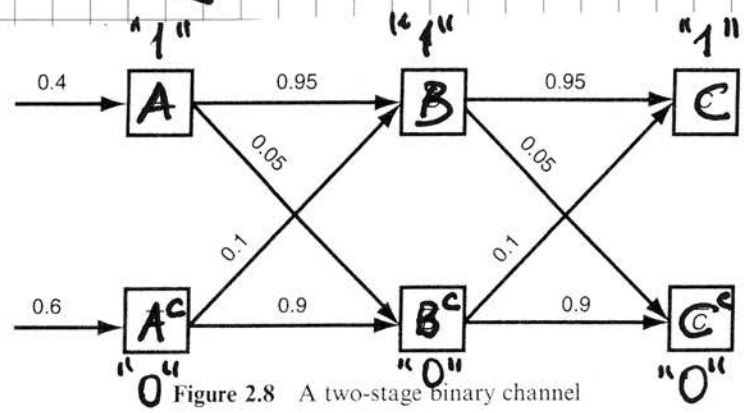


Figure 2.8 A two-stage binary channel

- We wish to determine, $Pr\{C\}$, the probability of receiving a "1" at the second stage.

- When the system has a "one-stage" memory, <u>three diagrams</u> are useful for determining the behavior of this system.
- This means that outcome of the second stage is dependent only on what happened at the first stage. and not on outcomes at stages prior to the first. It follows that,

$$Pr\{C|BA\} = Pr\{C|B\}^{=0.95}, \quad Pr\{C^c|B^cA\} = Pr\{C^c|B^c\}^{=0.9}$$

$$Pr\{C|BA^c\} = Pr\{C|B\}, \quad Pr\{C^c|B^cA^c\} = Pr\{C^c|B^c\}$$

0.95 0.9

misoro

$$Pr\{C \mid B^c A^c\} = Pr\{C \mid B^c\} = 0.1$$

$$Pr\{C \mid B^c A\} = Pr\{C \mid B^c\} = 0.1$$

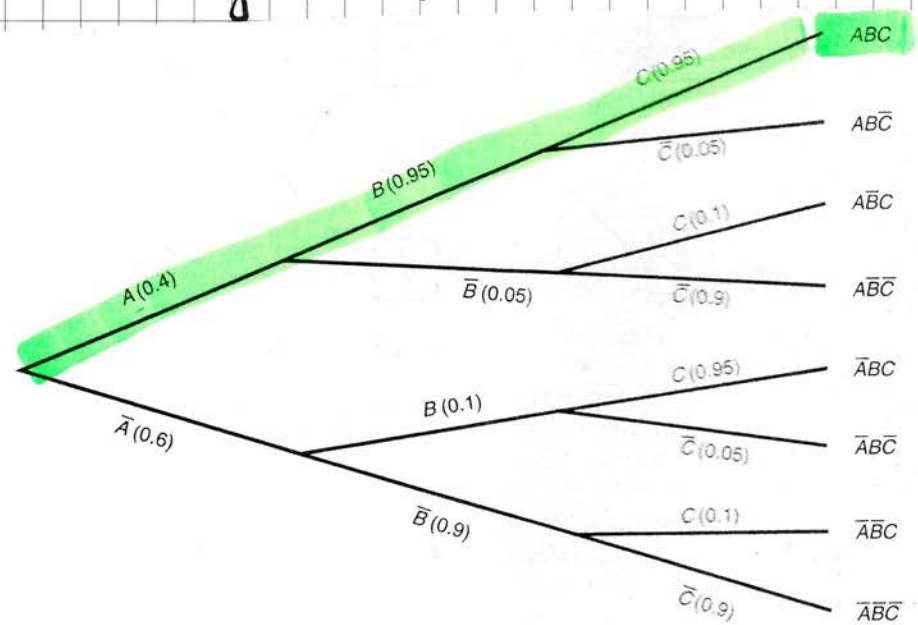
and

$$Pr\{C^c \mid BA\} = Pr\{C^c \mid B\} = 0.05$$

$$Pr\{C^c \mid BA^c\} = Pr\{C^c \mid B\} = 0.05$$

- These properties described above are commonly referred to as MARKOVIAN Properties.

- The <u>tree diagram</u> gives the flow of conditional probabilities originating from the source. Starting from the transmitter, the tree diagram for this problem given as



$\bar{A} = A^c$

$\bar{B} = B^c$

$\bar{C} = C^c$

misoro

- The top branch, for example, leads to the probability of the occurrence of event ABC,

$$Pr\{ABC\} = Pr\{A\}\,Pr\{B|A\}\,Pr\{C|BA\}$$
$$= Pr\{A\}\,Pr\{B|A\}\,Pr\{C|B\}$$
$$= 0.4 \cdot (0.95) \cdot (0.95) = 0.361$$

- The probability of C is found by summing the probabilities of all events end with C. Thus,

$$Pr\{C\} = Pr\{ABC\} + Pr\{AB^cC\}$$
$$+ Pr\{A^cBC\} + Pr\{A^cB^cC\}$$

$$= (0.95)(0.95)(0.4) + 0.1\,(0.05)(0.4)$$
$$+ (0.95)(0.1)(0.6) + 0.1\,(0.9)\,(0.6)$$
$$= 0.472$$

misoro

misoro

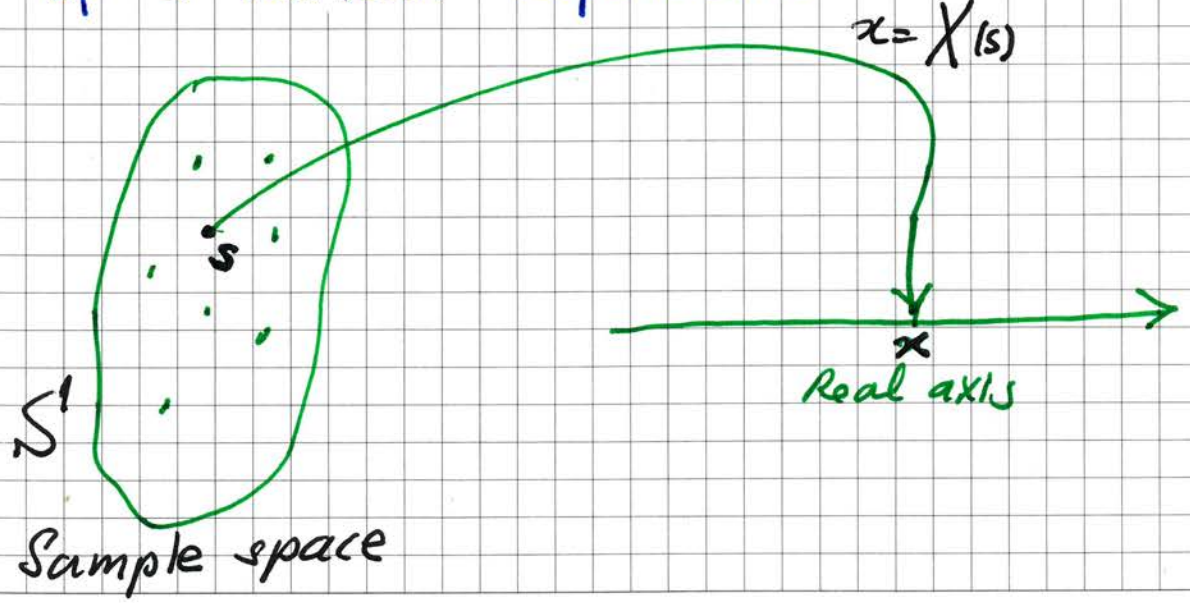
CHAPTER 3

RANDOM VARIABLES

- Our objective here at this point is to introduce of random variables.
- Outcomes of a random experiment not necessarily always numerically-valued.
- In order to construct a model for a random variable, we assume that it is possible to assign a real number $X(s)$ for each outcome s following a certain set of rules.

Definition:

A random variable X is a process of assigning a number $X(s)$ to every outcome s in the sample space S of a random experiment.

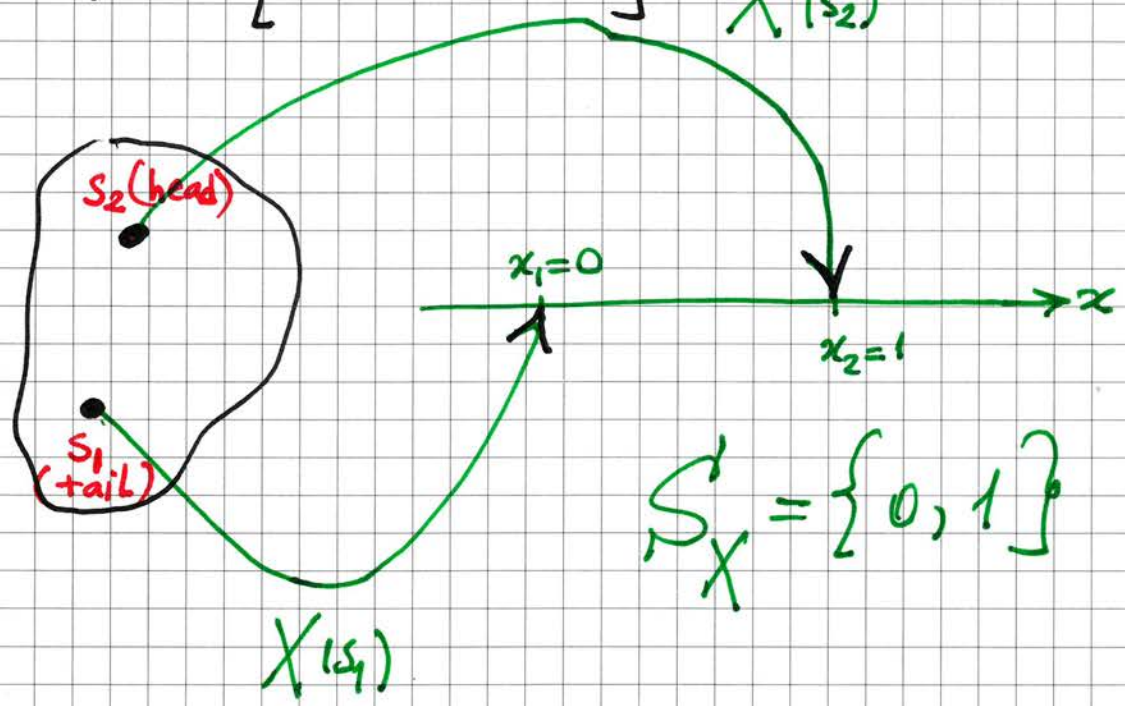


misoro

Example:

- In the case of a coin toss the outcomes form non numerical sample space

$$S = \{ \text{head}, \text{tail} \}$$



$$S_X = \{ 0, 1 \}$$

- For many applications, this is a convenient mapping of the outcome of a coin toss into set of real numbers.

$$X(s_i) = \begin{cases} 0 & , \quad s_1 = \text{tail} \\ 1 & , \quad s_2 = \text{head} \end{cases}$$

- Notationally, the dependence of random variable $X(s)$ on s will be omitted for convenience.

- If the possible outcomes of a random experiment are success and failure, we can assign number one to the event success and zero to failure. If Y is the random variable associated with this experiment, then X takes on two possible values; 1 and 0. Moreover, the following statements are equivalent:

- The outcome is success $\equiv$ The outcome is 1 $\equiv$ $X=1$,

- The outcome is failure $\equiv$ The outcome is 0 $\equiv$ $X=0$.

- The random variable X is called a discrete random variable if it is defined over a sample space having a finite or a countably infinite number of sample points,

$$R_X = \{x_1, x_2, \dots, x_k, \dots\} .$$

- In the case of a sample space having a uncountably infinite number of sample points is called continuous random variable. Its values are distributed

over one or more continuous intervals on the real line.

- In our notation, the name of all random variables are written in capital letters, $X, Y, Z, \ldots$. The value that a random variable will be denoted by corresponding lower case letters such as x, y, z or $x_1, x_2 \ldots$

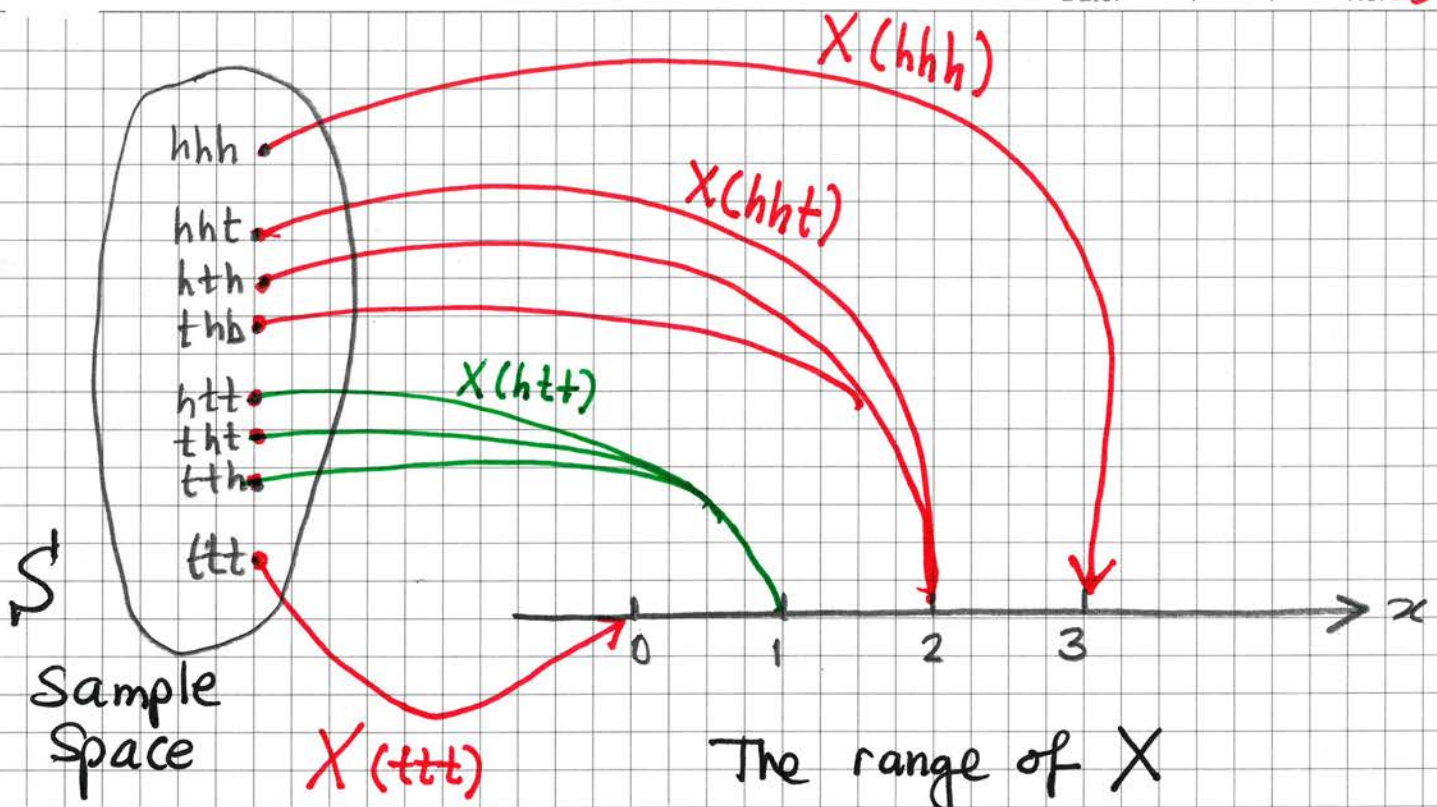
- The set of all possible values of X is the range of the random variable X. We denote the range of a random variable by the letter R with a subscript which is the name of the random variable. For example, $R_X, R_Y, R_Z, \ldots$

Example:

Consider a true (fair) coin tossed three times and the sample space is

$$S = \{hhh, hht, hth, hht, htt, tht, tth, ttt\}$$

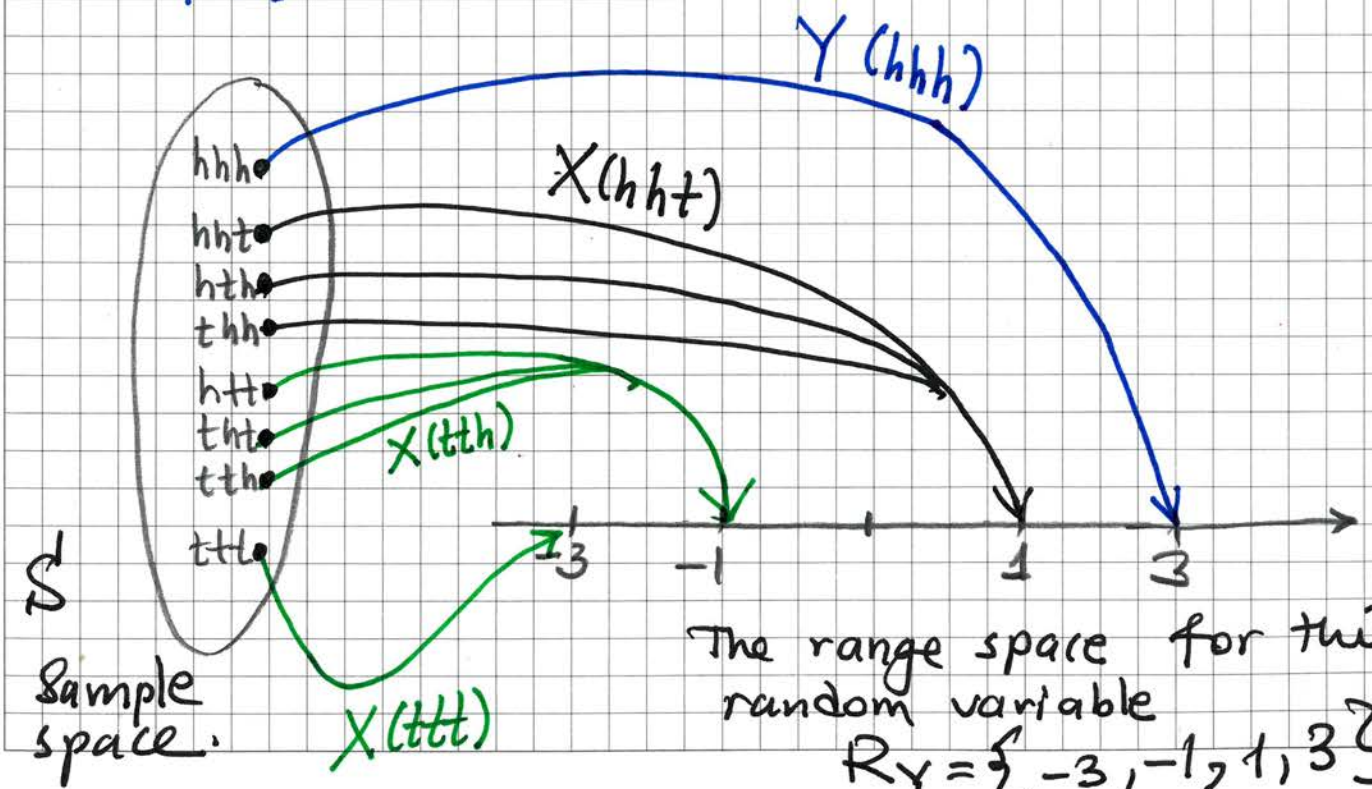
(1) If X is the number of heads showing than



The range of X

$$R_X = \{0, 1, 2, 3\}.$$

(2) Let's define another random variable Y; for <u>each head</u> we are gaining 1 lira and loasing 1 lira for <u>each tail</u>, Then



The range space for this random variable

$$R_Y = \{-3, -1, 1, 3\}$$

misoro

If $Pr\{head\} = p$, then $Pr\{tail\} = 1-p$

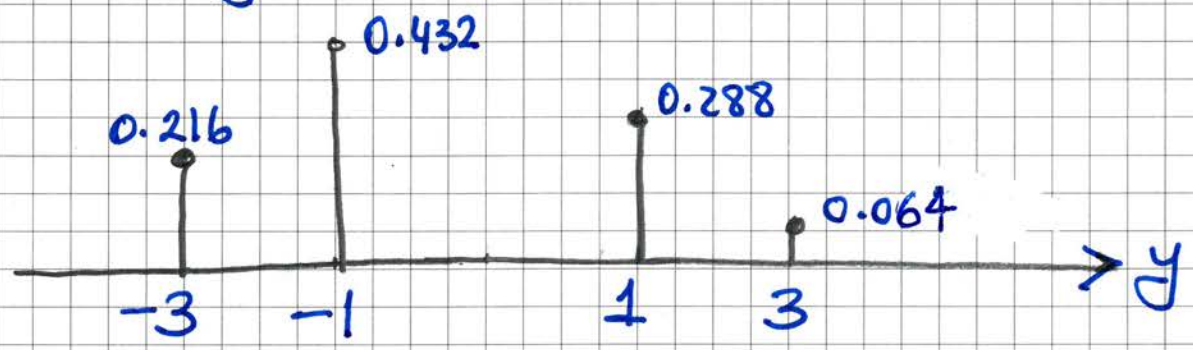
$$Pr\{Y=3\} = Pr\{hhh\} = p^3$$

$$Pr\{Y=1\} = Pr\{hht\} + Pr\{hth\} + Pr\{thh\} = 3p^2(1-p)$$

$$Pr\{Y=-1\} = Pr\{tth\} + Pr\{tht\} + Pr\{htt\} = 3p(1-p)^2$$

$$Pr\{Y=-3\} = Pr\{ttt\} = (1-p)^3$$

For $p=0.4$, we get the following probability distribution:



This figure is called as the probability mass function (pmf) of the random Variable Y.

- **Random Vector:**

We will have many occasions to consider a sequence of random variables

$$X_j, \quad j = 1, 2, \ldots, n.$$

In these cases we assume that they are defined on the same prabability space.

misoro

The random variables $X_1, X_2, \ldots, X_n$ will than map every element s of S in the probability space onto a point in the n-dimensional space $\mathbb{R}^n$,

$$\underline{X} = [X_1 \; X_2 \; \cdots \; X_n]^T .$$

An analysis involving n random variables is equivalent to considering a random vector having n random variables as its components. A random vector will be denoted by bold letters $\underline{X}, \underline{Y}, \underline{Z}, \ldots$

PROBABILITY DISTRIBUTION FUNCTIONS

• For Discrete Random Variables

For a discrete random variable X, the probability distribution, is a function that assigns a probability to each value in the range $R_X = \{x_1, x_2, \ldots, x_k, \ldots\}$.

$$P_X(x_i) = Pr\{X = x_i\}$$

is defined as probability mass function (pmf) and satisfies following conditions:

(1) $P_X(x_i) \geq 0$, for all i, (2) $\sum_{\text{all } i} P_X(x_i) = 1$.

Example:

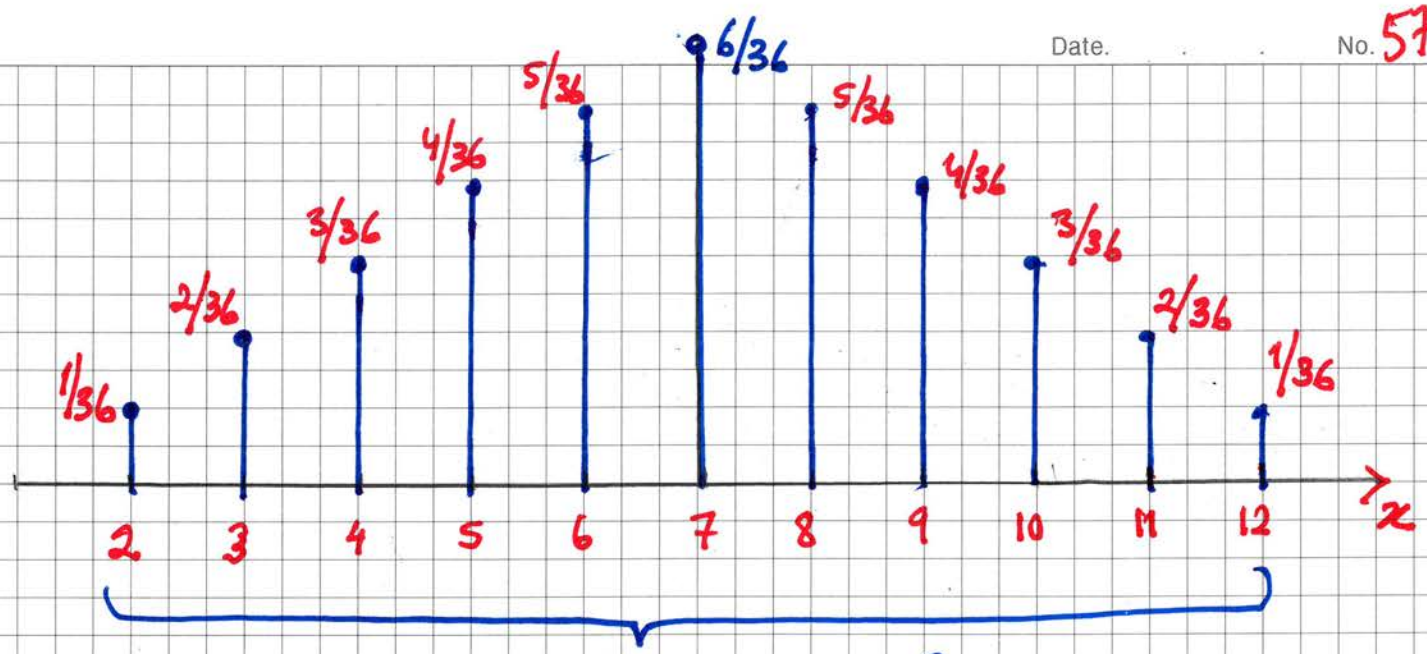
Consider the experiment of rolling two dice and observing the sum of faces that appear. There are 36 points in the sample space.

$$S = \{(1,1), (1,2), (2,1), (2,2), \ldots, (6,6)\}.$$

- There are 36 equally likely (equally probable) outcomes.
- Therefore, the probability of each simple event or outcome is 1/36.
- We tabulate X and its probability distribution below:

X Sum of the two faces	Number of ways that X can occur	$P_X(x)$ Prob. mass function
2	1	1/36
3	2	2/36
4	3	3/36
5	4	4/36
6	5	5/36
7	6	6/36
8	5	5/36
9	4	4/36
10	3	3/36
11	2	2/36
12	1	1/36

misoro



$$R_X = \{2, 3, \ldots, 12\}$$

- Satisfies conditions,

$$P_X(x_i) \geq 0$$

and

$$\sum_{\text{all } i} P_X(x_i) = 1.$$

- Another interesting and useful function is,

Probability Distribution Function (PDF),
or Distribution Function,
or Cumulative Distribution Function (CDF).

- This probability function is clearly dependent on the assigned value x:

$$F_X(x) = P_r\{X \leq x\}.$$

misoro

- For the discrete random variable X, cdf satisfies following conditions:

(1) $F_X(x) = Pr\{X \leq x\}$

$$= \sum_{x_i \leq x} P_X(x_i)$$

(2) $0 \leq F_X(x) \leq 1$

(3) If $x_1 < x_2$, then $F_X(x_1) \leq F_X(x_2)$.

- This means $F_X(x)$ nondecreasing function of the real variable x.
- It exists for discrete and continuous random variables and has values between 0 and 1.
- It is a nonnegative, continuous-to-the-left, and nondecreasing function of real variable x. Moreover, we have

$$F_X(-\infty) = 0 \quad , \quad \text{and} \quad F_X(+\infty) = 1.$$

- If a and b are two real numbers such that $a < b$, then

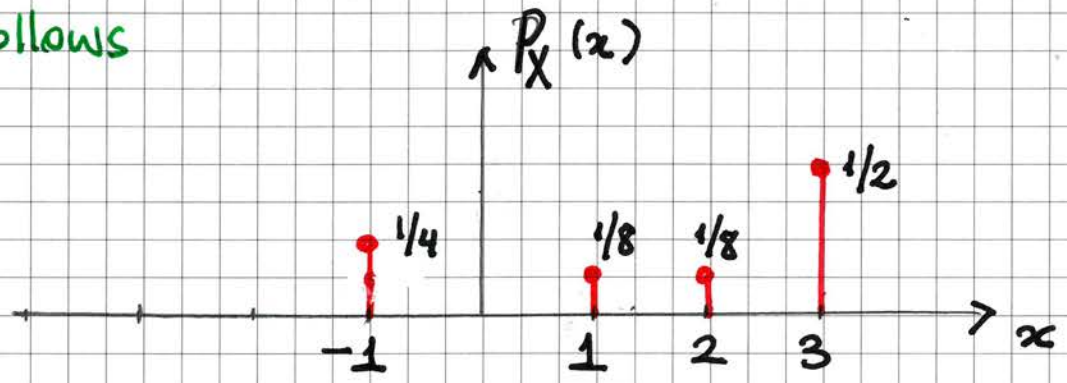
$$\Pr\{a < X \leq b\} = F_X(b) - F_X(a).$$

This relation is a direct result of the identity

$$\Pr\{X \leq b\} = \Pr\{X \leq a\} + \Pr\{a < X \leq b\}.$$

Example:

Let a discrete random variable X is a follows

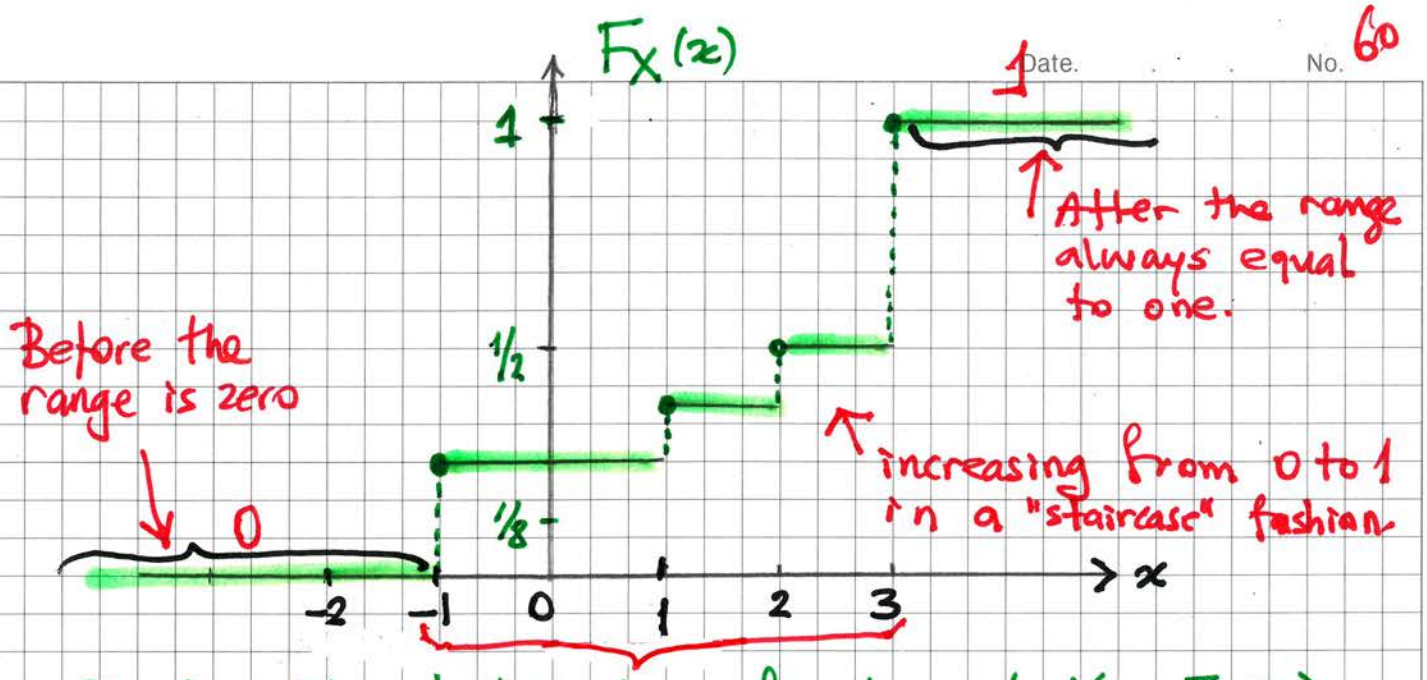


We then have

$$F_X(x) = \begin{cases} 0 \ , & \text{for } x < -1 \\ 1/4 \ , & \text{for } -1 \leq x < -1 \\ 3/8 \ , & \text{for } 1 \leq x < 2 \\ 1/2 \ , & \text{for } 2 \leq x < 3 \\ 1 \ , & \text{for } x \geq 3 \end{cases}.$$

This function is plotted in the following figure:

misoro



Probability distribution function of X, $F_X(x)$.

$R_X = \{-1, 1, 2, 3\}$

Range of X

SPECIAL VALUES OF A DISCRETE RANDOM VARIABLE

- **Mean** (or expected Value), $\mu_X = E[X]$.
- **Variance**, $VAR[X] = \sigma_X^2$.

The mean of a random variable is a weighted average of the possible values of X, with weights equal to the probablities:

$$\mu_X = E[X] = x_1 p_X(x_1) + x_2 p_X(x_2) + \cdots$$

$$= \sum_{\text{all } k} x_k p_X(x_k).$$

misoro

- Let us consider playing a certain game of chance.
- In order to play the game, we must pay a fee of a liras. As a result of the game, we receive X liras, where X is a random variable having possible values $x_1, x_2, \dots, x_r$.
- The question is:
  Should we play the game?
- If the game is to be played only once, then this question is quite dificult. However, suppose we play the game a large number of times.
- After n plays, we would pay na liras and receive $X_1 + X_2 + \dots + X_n$ liras.

  If we assume that succesive plays of the game are independent random experiments, then we can take the random variables $X_1, X_2, \dots, X_n$ as mutually independent and having the common probability mass function of X, $p_X(x_i)$:
- Let $N_n(x_i)$ denote the number of games that yielded the value x_i, another words,

the number of X_i's that assume the value x_i. Then we can write,

$$X_1 + X_2 + \cdots + X_n = \sum_{i=1}^{r} x_i \, N_n(x_i)$$

The average amount received is then

$$\frac{X_1 + X_2 + \cdots + X_n}{n} = \sum_{i=1}^{r} x_i \underbrace{\left[\frac{N_n(x_i)}{n}\right]}_{P_X(x_i)}$$

- According to relative frequency interpretation of probabilitie, if n is large; this number should be approximately equal to

$$\boxed{\mu_X = E[X] = \sum_{i=1}^{r} x_i \, P_X(x_i).}$$

Now, let's return our question, should we play the game?

- It seems reasonable to anticipate a net gain in playing the game if $\mu > a$ and we should expect a net loss if $\mu < a$. If

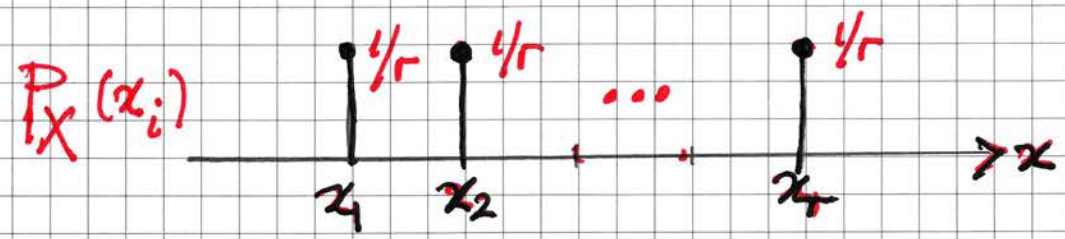
$\mu = a$ we would anticipate just about breaking even.

Remarks:

(1)
$$\mu_X = E[X]$$
$$= \sum_{i=1}^{r} x_i P_X(x_i)$$

shows a weighted average of the possible values of X; the weight attached to the ith value x_i is its probability $P_X(x_i)$. Sometimes, μ_X and $E[X]$ is called statistical average of X.

(2) Suppose X has a uniform distribution on the set $\{x_1, x_2, \dots, x_r\}$



Then
$$E[X] = \frac{1}{r}\left[x_1 + x_2 + \dots + x_r\right]$$

so in this case, $E[X]$ is just the aritmetic average of the possible values of X.

misoro

Remark 3:

The mean is one way of trying to summarize a probability distribution by a single number that is supposed to represent a "typical value" of X.

PROPERTIES OF EXPECTATION:

- Let us assume X and Y to be two random variables having finite expectation.

1. If c is a constant and

$$\Pr\{X=c\}=1 \text{ , than } E[X]=c.$$

2. If c is constant, then cX has finite expectation and

$$E[cX] = c\,E[X].$$

3. $X+Y$ has finite expectation and

$$E[X+Y] = E[X] + E[Y].$$

4. Suppose $\Pr\{X \geqslant Y\}=1$, then $E[X] \geqslant E[Y]$.
Moreover,
$E[X]=E[Y]$ if and only if $\Pr\{X=Y\}=1$.

6. $$|E[X]| \leq E[|X|].$$

VARIANCE OF A DISCRETE RANDOM VARIABLE

- The expected value $E[X]$, by itself, provides us with a very limited information about the random variable X.
- For example, if we know $E[X]=0$, then it could be that X is zero all the time. However, it could be as well that X is equally likely to take extremely large positive and negative values.
- Therefore, we are interested not only in the mean of a random variable, but also the variation about its mean.
- Let the deviation of X about its mean be

$$D = X - E[X],$$

then D can take on positive and negative values. Since we are intrested in the magnitude of the variations only, it is convenient to work with D^2, which is always positive.

- The variance of a random variable X is defined as the mean-squared variation, $E[D^2]$:

$$\sigma_X^2 = VAR[X] = E[D^2]$$
$$= E\left[(X - E[X])^2\right]$$
$$= \sum_{\text{all } k} (x_k - \mu_X)^2 P_X(x_k).$$

- This expression can be simplified as follows:

$$VAR[X] = E\left[(X - \mu_X)^2\right]$$

Varyans is also called the second moment of X about the mean.

$$= \underbrace{\sum_{\text{all } k} x_k^2 P_X(x_k)}_{E[X^2]} + \mu_X^2 \underbrace{\sum_{\text{all } k} P_X(k)}_{1} - 2\mu_X \underbrace{\sum x_k P_X(x_k)}_{E[X] = \mu_X}$$

$$= E[X^2] - \mu_X^2 - 2\mu_X^2$$

$$= \underbrace{E[X^2]}_{} - \underbrace{E[X]^2}_{}.$$

This is called as the 2nd moment about the origin.

This is called as the first moment about the origin.

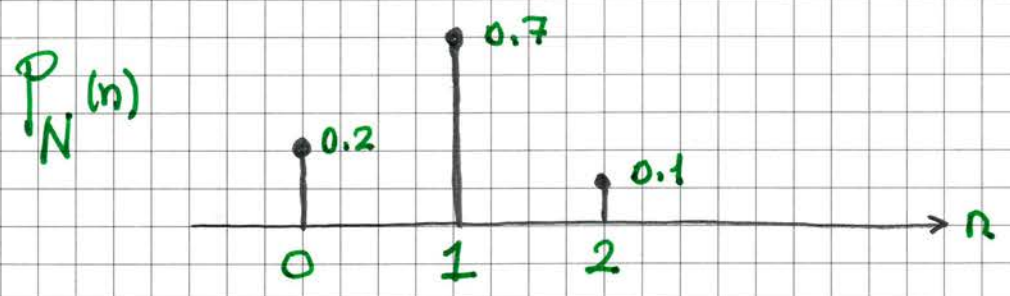
misoro

- The standard deviation of the Random variable X is defined by

$$\sigma_X = \left(VAR[X]\right)^{\frac{1}{2}} .$$

Example:

The pmf of a random variable N is given as



- $$\mu_N = E[N] = \sum_{n=0}^{2} n P_N(n)$$
$$= 0\,P_N(0) + 1\,P_N(1) + 2\,P_N(2)$$
$$= 0(0.2) + 1(0.7) + 2(0.1) = 0.9$$

- $$\sigma_N^2 = \sum_{n=0}^{2} (n-\mu_N)^2 P_N(n)$$
$$= (0-0.9)^2(0.2) + (1-0.9)^2(0.7) + (2-0.9)^2(0.1) = 0.29$$

- Standard derivation is
$$\sigma_N = \sqrt{0.29} = 0.53.$$

misoro

Estimation of μ_X and σ_X^2

If we do not know the pmf (probability mass function) of a discrete random variable, μ_X and σ_X^2 can be estimated by observing M output of the random variable X,

$$\{x_1, x_2, \ldots, x_M\}.$$

The estimation of the mean ($\hat{\mu}_X$) and the variance ($\hat{\sigma}_X^2$) are

$$\hat{\mu}_X = \frac{1}{M}\sum_{k=1}^{M} x_k$$

$$\hat{\sigma}_X^2 = E\left[X^2\right] - \mu_X^2$$

$$= \frac{1}{M}\sum_{k=1}^{M} x_k - \left(\frac{1}{M}\sum_{k=1}^{M} x_k\right)^2.$$

Properties of the variance:

1. $VAR[c] = 0$, c is a constant,
2. $VAR[X+c] = VAR[X]$,
3. $VAR[cX] = c^2\, VAR[X]$.

misoro

Remark:

VAR[X] is a nonlinear operator.
However, E[X] is a linear operator.
Therefore,

- $E[X+X] = E[2X]$
$= 2E[X]$

- $\sigma^2_{X+X} = \sigma^2_{2X}$
$= 4\sigma^2_X$.

PROBABILITY DISTRIBUTIONS
For Continuous Random Variables

For a continuous random variable, we are unable to assign a probability to sample point or experimental outcome in the same manner as we did for the discrete random variable, because the range contains an uncountably infinite number of values.