Sukkur Institute of Business Administration University

Department of Electrical Engineering.

Probability and Random Variables

Course Handouts

Instructor

Dr. Muhammad Asim Ali

Fall 2024

Contents

Chapter 1

Introduction

The objective of this chapter / course is to introduce audience with algebra of set theory, axioms of probability, conditional probability, Baye's rule and independent events

- Ⓐ Set Theory
- Ⓑ Events & Sample Space
- Ⓒ Sample Space / Events
- Ⓓ Axioms of Probability
- Ⓔ Other Features
- Ⓔ Conditional Probability

Definition

The study of probability stems from the analysis of certain games of chance, and it has found applications in most branches of science and engineering. In this chapter the basic concepts of probability theory are presented. Probability Model: three components are (a) the sample space, (b) events, and (c) probabilities of events.

Set Theory

A set is a collection of objects, e.g. A set of students in a probability class. A set is defined by enumeration

$$A = \{\text{Ali, Aliya, Akhtar, Aasia}\}$$

or by description

$$A = \{\text{Students: each student enrolled in probability class.}\}$$

some more examples

$$B = \{1, 2, 3, \dots\} \qquad \text{enumeration}$$

$$B = \{\text{I:I is an integer and} \quad I \geq 1\} \quad \text{description}$$

Each object in the set is called *element* each element is *distinct*. *Ordering* within the set is not important. i.e. $\{1, 2, 3\}$ and $\{1, 3, 2\}$ are equivalent. Sets are said to be equal if they contain same elements.

Consider the set of all outcomes of tossing a die. This is

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

The numbers $1, 2, 3, 4, 5, 6$ are its elements, that are distinct. The set of integer numbers from 1 to 6 or $B = \{I : 1 \leq I \leq 6\}$ is equal to A. The set A is also the *universal* set $\mathcal{S}$ since it contains all the outcomes.

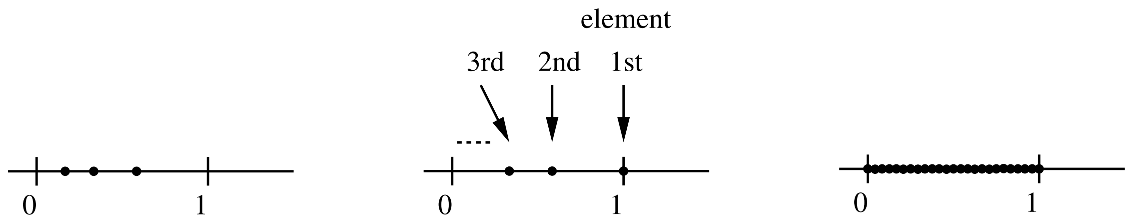
Size of Sets

Three different types of sets

$$A = \left\{\frac{1}{8}, \frac{1}{4}, \frac{1}{2}, 1\right\} \qquad \text{Finite Set- Discrete}$$

$$B = \left\{1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots\right\} \qquad \text{Countably infinite set - discrete}$$

$$C = \left\{x : 0 \leq x \leq 1\right\} \qquad \text{Infinite set - discrete}$$



The possible outcomes comprise the elements of set $\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$. We may be interested in particular outcome of the die tossing experiment; other times we might be not interested in a particular outcome of the die tossing experiment. Simplest type of events are the ones that contain only a single outcome such as $E_1 = \{1\}$, $E_2 = \{2\}$; a complex event contains multiple events $E_{\text{even}} = \{2, 4, 6\}$. Disjoint sets such as $\{1, 2\}$ and $\{3, 4\}$ are said to be *mutually exclusive*.

set theory	Probability theory	Probability Symbol
Universal Set	Sample space	$\mathcal{S}$
Element	Outcome (sample point)	$\mathscr{s}$
Subset	Event	E
disjoint set	mutually exclusive events	$E_1 \cap E_2 = \emptyset$
null set	impossible event	$\emptyset$
simple set	simple event	$E = \{\mathscr{s}\}$

Theory of probability assigns probabilities to event. What is the probability that the tossed die will produce an even outcome? denoting this probability by $P[E_{\text{even}}]$, we would intuitively say it is 1/2 since there are 3 chances out of 6 to produce an even outcome. P is a *probability function* or a function that assigns a number between 0 and 1 to sets.

The probability function must assign a number to every event, or every set. Examples:

- For tossing a coin:

$E_1 = \{H\}$	$E_2 = \{T\}$
$E_3 = \emptyset$	$E_4 = \{\mathcal{S}\}$

- For tossing a fair die:

$E_0 = \emptyset$	$E_1 = \{1\}$	$E_2 = \{2\}$	$E_3 = \{3\}$	$E_4 = \{4\}$
$E_5 = \{5\}$	$E_6 = \{6\}$			
$E_{12} = \{1,2\}$	$E_{13} = \{1,3\}$	$E_{14} = \{1,4\}$	$E_{15} = \{1,5\}$	$E_{16} = \{1,6\}$
	$\ldots$	$E_{123} = \{1,2,3\}$	$\ldots$	
	$\ldots$	$\ldots$		$E_{1\cdots6} = \mathcal{S}$

There are a total of 64 events, we must be table to assign probabilities to all of these events.

Example

For the case of a fair die. Let us determine the probability that die comes up either less than or equal to 2 or equal to 3.

$$
\begin{aligned}
P\{\{1,2\} \cup \{3\}\} &= P[\{1,2\}] + P[\{3\}] \\
&= \frac{2}{6} + \frac{1}{6} = \frac{1}{2}
\end{aligned}
$$

1.1 Experiment,Sample Spaces and Events

1. Experiment (experiment of chance, not necessarily a designed experiment), a process whose outcomes are uncertain. i.e. A non-deterministic process.
   Examples: Roll of a dice, tossing a coin, drawing a card from a deck.

2. Sample Space: $\mathcal{S}$ is a set that contains all possible outcomes from the experiment. In some cases, $\mathcal{S}$ contains outcomes that are not possible.

   (a) The number of outcomes in the sample space can be finite or infinite.

   (b) Infinite sample space can be countable or uncountable. A sample space is countable if the outcomes can be associated with the integers 1,2,$\cdots$.

Example Find sample space for the experiment of tossing coin (a) once (b) twice.

a. There are two possible outcomes, heads or tails. Thus

$$S = \{T, H\}$$

b. There are four possible outcomes. They are pair of heads and tails. Thus

$$S = \{HH, HT, TH, TT\}$$

Example Some examples with infinite sample space.

a. Tossing a coin repeatedly and counting the number until first head appears

$$S = \{1, 2, 3, \dots\}$$

b. Sample space for experiment measuring (in hours) the lifetime of a transistor.

   clearly possible outcomes are all non-negative numbers.

$$S = \{t\, 0 \leq t \leq \infty\}$$

3. Event: A subset of sample space.

   (a) A simple event contains only one outcome. A simple event is denoted by $\mathfrak{s}$.

   (b) A compound event contains two or more outcomes . Compound events are denoted by capital letters.

Very simply the probability is defined as

$$\text{Probability of an Event} = \frac{\text{No. of Favorable Outcomes}}{\text{No. of Total Outcomes}}$$

1.2 Algebra of Events

1. Relationships and definitions

   (a) Inclusion $A \subset B$ means $\mathfrak{s} \in A \Rightarrow \mathfrak{s} \in B$, where $\mathfrak{s}$ is an outcome in $\mathcal{S}$.

   (b) Equality: $A = B$ means that $A \subset B$ and $B \subset A$.

   (c) Complement: $\overline{A}$ is the set $\{s \in \mathcal{S}, \mathcal{S} \notin A\}$

2. Set operations

   (a) Union: $A \cup B$ is the set of all the outcomes that are in A and/or B.

   $$A \cup B = \{\mathfrak{c} : \mathfrak{c} \in A \quad \text{or} \quad \mathfrak{c} \in B\}$$

   (b) Intersection: $A \cap B$ is the set of all the outcomes that are in A and B.

   $$A \cup B = \{\mathfrak{c} : \mathfrak{c} \in A \quad \text{and} \quad \mathfrak{c} \in B\}$$

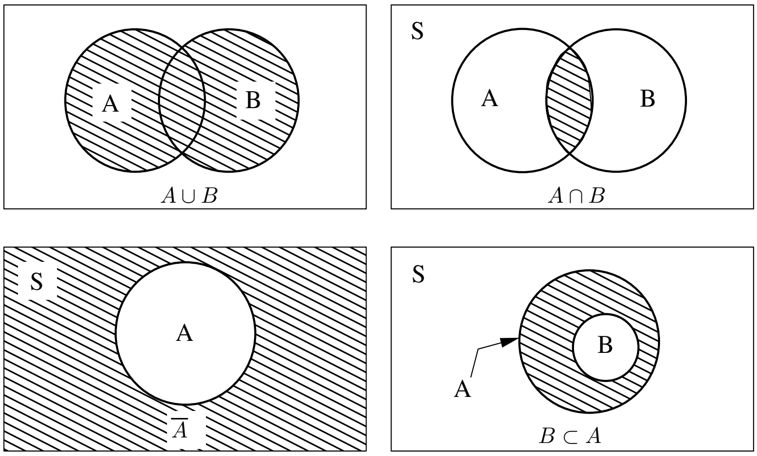
   (c) Disjoint Sets: Two sets A and B are called *disjoint* or *mutually exclusive* if they contain no common element that is if $A \cap B = \emptyset$.

   (d) Commutative Operations: $A \cup B = B \cup A$ and $A \cap B = B \cap A$.

   (e) Subtractive Operation:$A \backslash B$ consists of all the outcomes of A except B

   (f) Associative Operation:$(A \cup B) \cup C = A \cup (B \cup C)$ and $(A \cap B) \cap C = A \cap (B \cap C)$.

   (g) Distributive Operations:$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ and $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

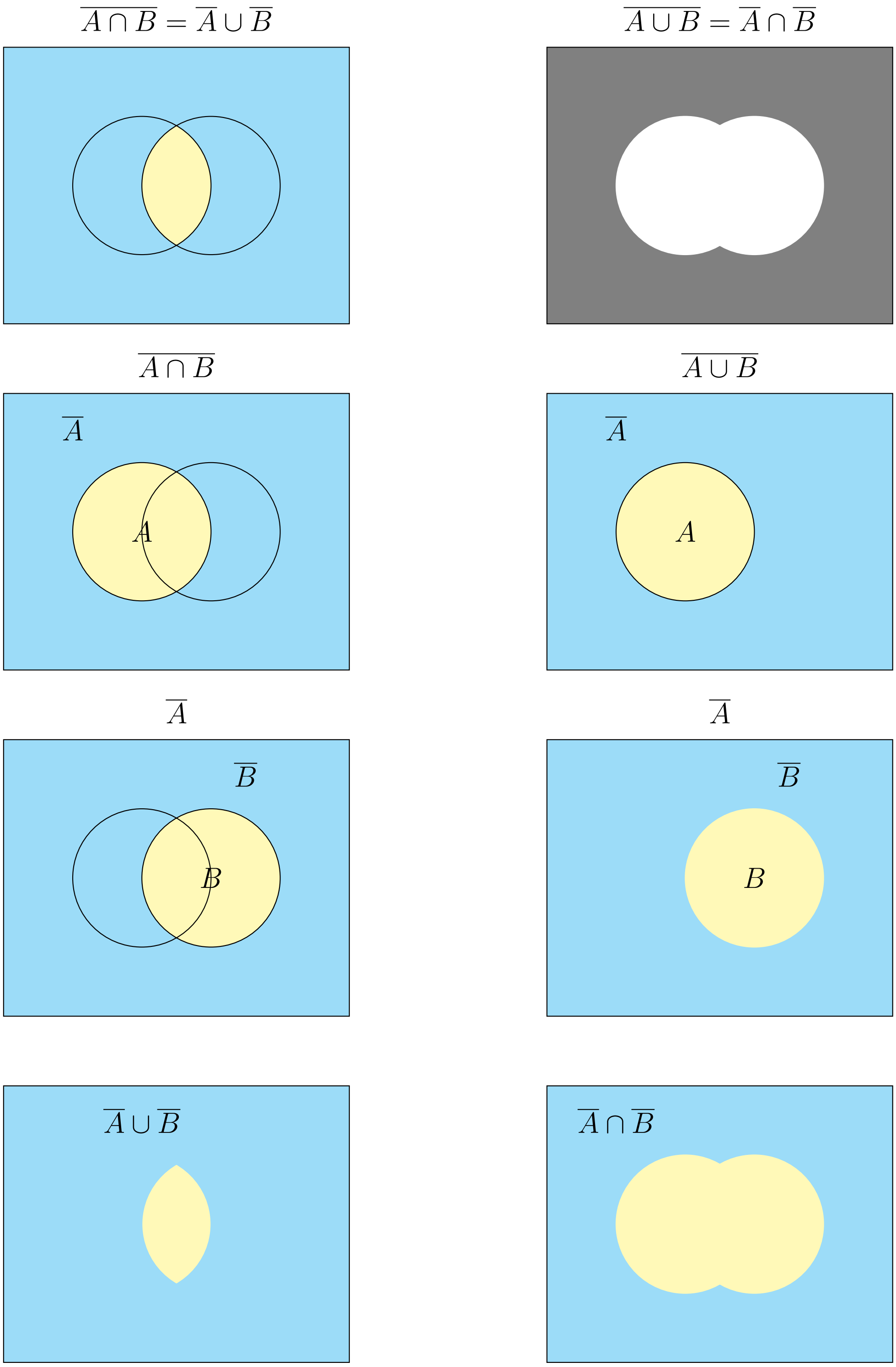


The definition of union and intersection of two sets can be extended to an finite number of sets as follows

$$
\begin{aligned}
\bigcup_{i=1}^{n} A_i &= A_1 \cup A_2 \cup \cdots \cup A_n \\
&= \{\mathfrak{c} : \mathfrak{c} \in A_1 \quad \text{or} \quad \mathfrak{c} \in A_2 \quad \text{or} \ldots \mathfrak{c} \in A_n\} \\
\bigcap_{i=1}^{n} A_i &= A_1 \cap A_2 \cap \cdots \cap A_n \\
&= \{\mathfrak{c} : \mathfrak{c} \in A_1 \quad \text{and} \quad \mathfrak{c} \in A_2 \quad \text{and} \ldots \mathfrak{c} \in A_n\}
\end{aligned}
$$

(h) Demorgan's Laws

$$
\begin{aligned}
(\overline{A \cup B}) &= \overline{A} \cap \overline{B} \quad \text{equivalently} \\
(\overline{A} \cap \overline{B}) &= (\overline{A} \cup \overline{B})
\end{aligned}
$$



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

and generally speaking

$$A \cap \left(\bigcup_{i=1}^{n} B_i\right) = \bigcup_{i=1}^{n}(A \cap B_i)$$
$$A \cap \left(\bigcup_{i=1}^{n} B_i\right) = \bigcap_{i=1}^{n}(A \cap B_i)$$

The extension of DeMorgan's Law

$$\overline{\left(\bigcup_{i=1}^{n} A_i\right)} = \bigcap_{i=1}^{n}(\bar{A}_i)$$
$$\overline{\left(\bigcap_{i=1}^{n} A_i\right)} = \bigcup_{i=1}^{n}(\bar{A}_i)$$

3. Identities

Key Properties of Sets

1. $\overline{S} = \emptyset$
2. $\overline{\emptyset} = S$
3. $\overline{\overline{A}} = A$
4. $S \cup A = S$
5. $S \cap A = A$
6. $A \cup \overline{A} = S$
7. $A \cap \overline{A} = \emptyset$

A relative frequency definition of probability:

Suppose that the random experiment is repeated n times. If event A occurs $n(a)$ times, then the probability of event A, denoted $P(A)$ is defined as

$$P(A) = \lim_{n \to \infty} \frac{n(A)}{n}$$

where $n(A)/n$ is called the relative frequency of event A. Note that limit may not exist, and in addition, there are many situations in which the concept of repeatability may not be valid. It is clear that for any event A, the relative frequency of A will have the following properties:

i. $0 \leq n(A)/n \leq 1$ '0' if no occurrences and '1' if occurs all the time.

ii. If A and B are mutually exclusive events, then

$$n(A \cup B) = n(A) + n(B)$$

and

$$\frac{n(A \cup B)}{n} = \frac{n(A)}{n} + \frac{n(B)}{n}$$

1.3 Axioms of Probability

(a) $0 \leq P(A) \leq 1$ for any event A.

(b) $P(\mathcal{S}) \equiv 1$

(c) Additivity: If $A_1, A_2 \cdots$ are pairwise disjoint events,i.e.

$$A_i \cap A_j = \emptyset \quad \forall\, i \neq j$$

then,

$$P(A_1 \cup A_2 \cup \cdots) = P(A_1) + P(A_2) + \cdots$$

In case if there are finitely many events $A_1, A_2, \cdots, A_n$ then we have

$$P(A_1 \cup A_2 \cup \cdots A_n) = \sum_{i=1}^{n} P(A_i)$$

Implication of the Axioms

- $P(\overline{A}) = 1 - P(A)$

$$P(\mathcal{S}) = 1 = P(A \cup \overline{A}) = P(A) + P(\overline{A}) \Rightarrow P(\overline{A}) = 1 - P(A).$$

Corollary-1:$P(A) \leq 1$.

Corollary-2:$P(\emptyset) = 0$.

Corollary-3:If A and B are disjoint, then $P(A \cap B) = 0$.

- If $A \subseteq B$, then $P(A) \leq P(B)$.

  Proof $A \subset B \Rightarrow B = (A \cap B) \cup (\overline{A} \cap B) \Rightarrow P(B) = P(A) + P(\overline{A} \cap B) \geq P(B)$ because A is disjoint from $\overline{A} \cap B$.

- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ for any events A and B

  Proof:

  Note that $A = (A \cap B) \cup (A \cap \overline{B})$ and $B = (A \cap B) \cup (\overline{A} \cap B)$. Accordingly $P(A) = P(A \cap B) + P(A \cap \overline{B})$ and $P(A \cap \overline{B}) = P(A) - P(A \cap B)$ because $A \cap B$ is disjoint from $A \cap \overline{B}$ Similarly, $P(B) = P(A \cap B) + P(\overline{A} \cap B)$ and $P(\overline{A} \cap B) = P(B) - P(A \cap B)$. Lastly note that $A \cup B = (A \cap B) \cup (A \cap \overline{B}) \cup (\overline{A} \cap B)$. These three events are mutually disjoint, so

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

  Corollary-4: If $A \subseteq B$, then $P(A) \leq P(B)$ Assume $A_1 = A, A_2 = B \backslash A$ Again we have $A_1 \cap A_2 = 0$(since the elements of $B \backslash A$ are by definition not in A), and $A_1 \cup A_2 = B$. So by Axiom 3,

  $$P(A_1) + P(A_2) = P(A_1 \cup A_2) = P(B)$$

  In other words,$P(A) + P(B \backslash A) = P(B)$ Now $P(B \backslash A) \geq 0$ by Axiom 1; so

  $$P(A) < P(B)$$

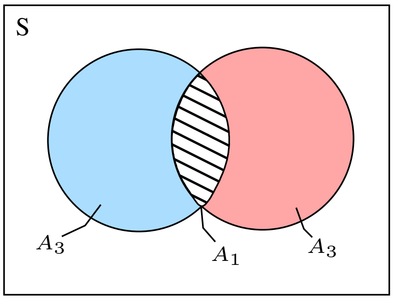
1.4 Inclusion Exclusion Principle

Proposition:

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

This proposition can be proved from the axioms using Venn Diagram as a guide, see that $A \cup B$ is made up of three parts, namely

$$A_1 = A \cap B, \qquad A_2 = A \backslash B, \qquad A_3 = B \backslash A$$



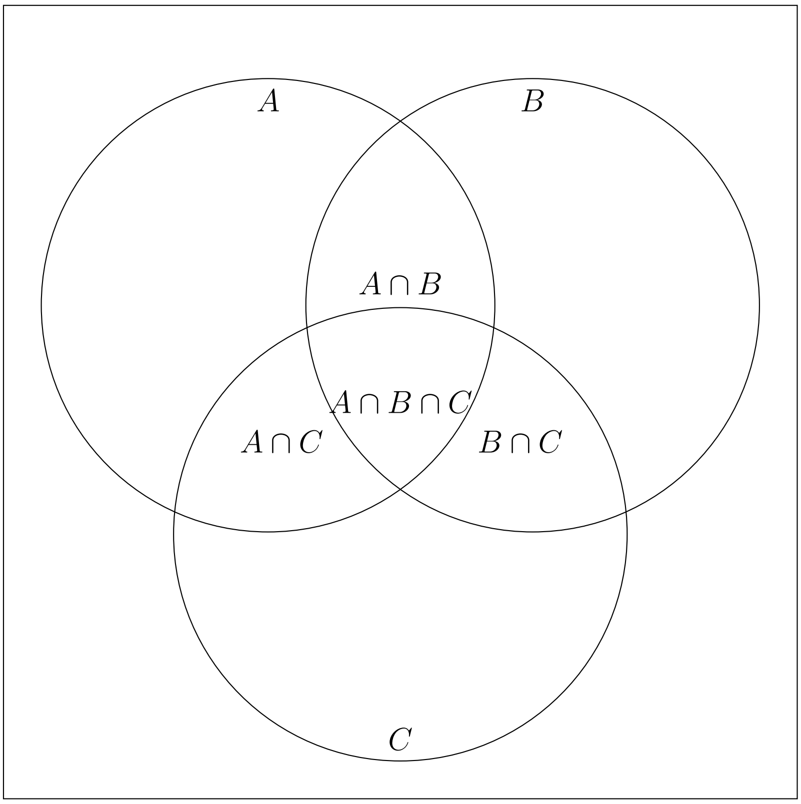
Indeed we do have $A \cup B = A_1 \cup A_2 \cup A_3$, since anything in $A \cup B$ is in both these sets are just the first of just the second. Similarly we have $A_1 \cup A_2 = A$ and $A_1 \cup A_3 = B$. The sets A_1, A_2, A_3 are mutually disjoint.

$$\begin{aligned} P(A) &= P(A_1) + P(A_2), \\ P(B) &= P(A_1) + P(A_3), \\ P(A \cup B) &= P(A_1) + P(A_2) + P(A_3), \end{aligned}$$

from this we obtain

$$\begin{aligned} P(A) + P(B) - P(A \cap B) &= (P(A_1) + P(A_2)) + (P(A_1) + P(A_3)) - P(A_1), \\ &= P(A_1) + P(A_2) + P(A_3), \\ &= P(A \cup B) \end{aligned}$$

This rule can be extended to multiple variables

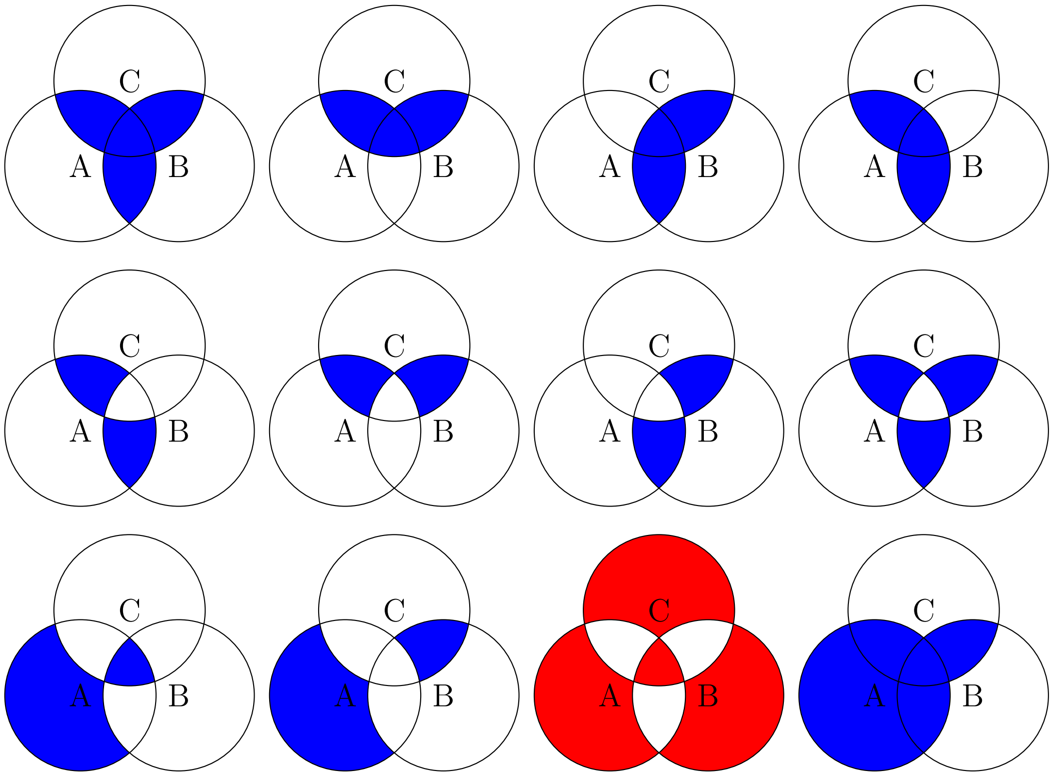


The inclusion exclusion principle for 3 variables is defined as

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

generalizing this to n-variable case we obtain.

$$\begin{aligned} P\Big(\bigcup_{i=1}^{n} A_i\Big) &= P(A_1) + \cdots + P(A_n) - \sum_{i \neq j} P(A_i \cap A_j) \\ &+ \sum_{i \neq j \neq k} P(A_i \cap A_j \cap A_k) - \ldots (-1)^{n-1} P(A_i \cap \cdots \cap A_n) \end{aligned}$$


C
A
B
C
A
B
C
A
B
C
A
B
C
A
B
C
A
B
C
A
B
C
A
B
C
A
B
C
A
B
C
A
B
C
A
B

Disjoint or Mutually Exclusive Events

Sets A and B in $\mathcal{S}$ are said to be mutually exclusive or disjoint if $A \cap B = \emptyset$. Sets $A_1, A_2 \cdots, A_n$ are said to be disjoint if $A_i A_j = \emptyset$ for every $i \neq j$. Further more if A and B are in $\mathcal{S}$, then

1. $A = AB \cup A\bar{B}$.
2. $AB \cap A\bar{B} = \emptyset$.
3. $A \subset B$, then $AB = A$ and $A \cup B = B$

1.4.1 Sampling with and without replacement

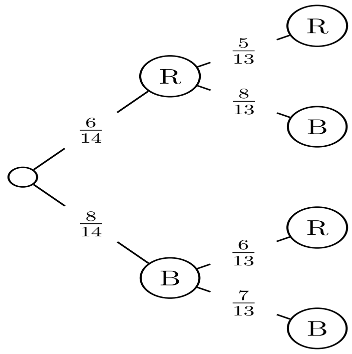
Note! The probability for sampling with/without replacement are fundamentally different.

We draw an object from a Box, we have a choice to replace or not replace the object back in the box before a redraw. In the first case a particular object can come up again and again, whereas in second case it can come up only once.
Sampling with replacement is system with infinite population. Theoretically speaking a large population can be considered as infinite population.

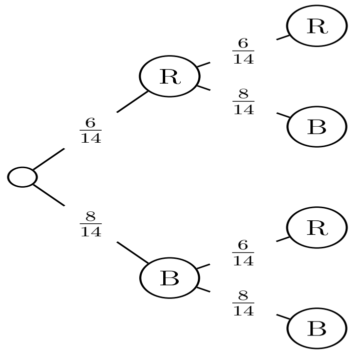
Example: Drawing from Urn (with / without replacement)

2 balls are selected at random with replacement from the urn containing 8 blue balls and 6 red balls.
Determine the probability of drawing 0, 1 or 2 red balls

Without Replacement



With Replacement



Insert the example of urn to establish the two different sets of sample spaces. -Ordered Pairs

-Unordered Pairs

Probability in terms of Odds

(i) If event A has probability $P(A)$ and event B has probability $P(B)$ then

$$\text{Odds of A to B} = \frac{P(A)}{P(B)}$$

Provided that $P(B) > 0$. Note that odds is not a probability. Rather, it's a ratio of probabilities.

(ii) Odds of a single event, A, is defined as the odds of A to $\overline{A}$. That is

$$\text{Odds of A to B} = \frac{P(A)}{P(\overline{A})} = \frac{P(A)}{1-(P(A))}$$

Provided that $P(A) < 1$

1.5 Experiment with Symmetries

1. Equally Likely Outcomes

   (a) If the outcomes in a finite sample space, $\mathcal{S} = \{\mathfrak{s}_1, \cdots, \mathfrak{s}_N\}$ are equally likely, then each has probability $1/N$ for $i = 1, \cdots, N$.

(b) Definition:If an object is drawn at random from a finite population of N objects, then the objects are equally likely to be selected.

(c) If an event A consists of a subset of k outcomes in a sample space of N equally likely outcomes, then $P(A) = k/N$.

1.6 Conditional Probability

We are often interested in the probability of an event given another event because one might be able to learn about the former by observing the latter. This is called the conditional probability and is defined below.

Definition: If $P(B) > 0$, then the conditional probability that the event A occurs given that the event B has occurred is defined as

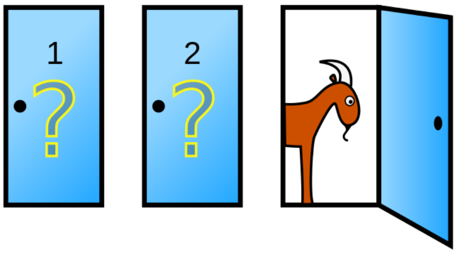
$$\begin{aligned} P(A|B) &= \frac{P(A \cap B)}{P(B)} \qquad P(B) > 0 \\ \Rightarrow P(A \cap B) &= P(A|B) \cdot P(B) \end{aligned}$$

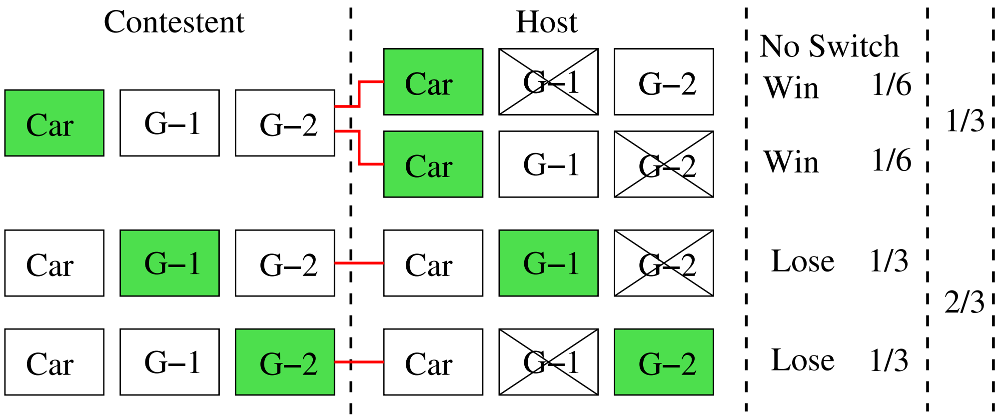
Similarly we can define $P(B|A)$.

The conditional probability can be very tricky as the following examples show.

Example

1. (Monty Hall) In an American game show, called Let's Make a Deal, the shows host, Monty Hall, shows a player three closed doors; behind one is a car, and behind each of the other two is a goat. The player is allowed to open one door, and will win whatever is behind the door. However, after the player selects a door but before opening it, Monty Hall (who knows whats behind the doors) opens another door, revealing a goat. Monty Hall then offer the player an option to switch to the other closed door. Does switching improve the players chance of winning the car?





2. (Gender of Twins) A couple is expecting twins. In a ultrasound examination, the technician was only able to determine that one of the two was boy. What is the probability that both are boys? During the delivery, the baby that was born first was a boy. What is the probability that both are boys?

1st Child

2nd Child	♂	♀
♂	$\frac{1}{2} \cdot \frac{1}{2}$	$\frac{1}{2} \cdot \frac{1}{2}$
♀	$\frac{1}{2} \cdot \frac{1}{2}$	$\frac{1}{2} \cdot \frac{1}{2}$

3. (Simpsons Paradox) A political scientist has performed a randomized experiment to determine the relative efficacy of two get-out-of-the vote strategies, with the following results.

	Partisans	Non-Partisans
	Visit−Phone	Visit−Phone
Voted	200−10	19−1000
Did not vote	1800−190	1−1000

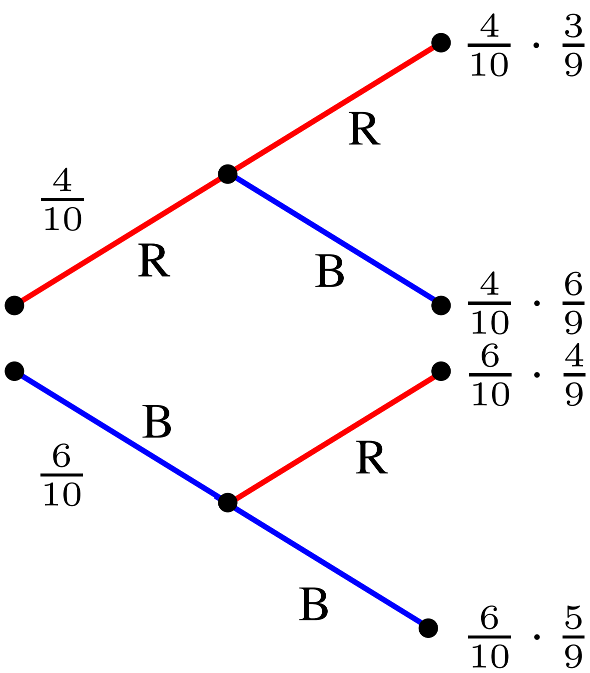
And some more examples:

Example 1

We have 10 marbles; 4 red and 6 blue, and take two of them randomly. We define the events A the 1st marble is red and B the 2nd marble is red . What is the probability that both marbles are red $P(A \cap B)$?

Since we can take the marbles out one at a time, the probability of 1^{st} marble being red is 4/10. Getting two read marbles then can be seen as the conditional probability of getting a second red marble $P(B|A)$, given the first marble is red. After removal of the first marble, the sample space has changed: we now have 3 red and 6 blue marbles, so the probability of getting red one now is $P(B|A) = 3/9$.

$P(A \cap B) = P(A) \cdot P(B|A) = 4/10 \times 3/9 = 2/15$



Example 2

A typical student appearing in a MCQ exam, generally, knows correct answers to 75 questions out of 100. For the remaining 25 questions he picks randomly 1 of the 5 choices. What is the probability for event that student guesses the answer correctly.

cond2-eps-converted-to.pdf

1.7 Product Rule of Probability

$$P(A \cap B) = P(B|A) \cdot P(A)$$

Generalization of the product Rule of Probability

If $P(A_1 \cap \cdots \cap A_{n-1}) \neq 0$, then:

$$P(A_1 \cap \cdots \cap A_n) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap \cdots \cap A_{n-1})$$

The multiplication rule implies that the probability of all the events $(A_1 \cap \cdots \cap A_n)$ happening is equal to the probability of event A_1 times the probability of event A_2 given A_1 and so forth.

$\Rightarrow$ Also known as Chain-rule of probability, Iterative conditional probability.

1.8 Law of Total Probability

For events $A_1, A_2, \cdots, A_n$ form a partition of the sample space if the following two conditions hold

(a) The events are pairwise disjoint, that is $A_i \cap A_j = \emptyset$ and $i \neq j$, for any pair of events A_i and A_j

(b) $A_1 \cup A_2 \cup \cdots \cup A_n = \mathcal{S}$.

Let $A_1, A_2, \cdots, A_n$ form a partition of the sample space with $P(A_i) \neq 0\, \forall i$, and let B be any event. Then

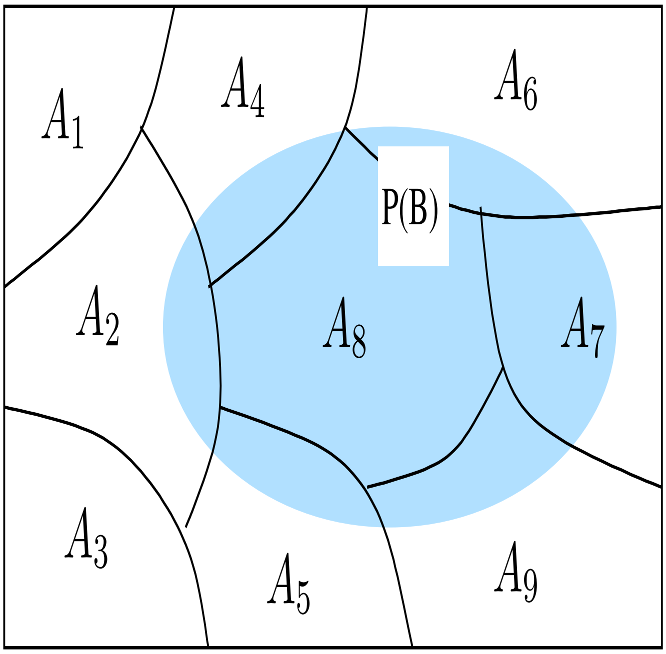
$$P(B) = \sum_{i=1}^{n} P(B \cap A_i) = \sum_{i=1}^{n} P(B|A_i) \cdot P(A_i)$$

Proof: By definition, $P(B|A_i) = P(B \cap A_i)/P(A_i)$. Multiplying up, we find that

$$P(B \cap A_i) = P(B|A_i) \cdot P(A_i)$$

Now consider the events $B \cap A_1, B \cap A_2, \cdots, B \cap A_n$. These events are pairwise mutually disjoint: for any outcome lying in both $B \cap A_i$ and $B \cap A_j$ would lie both in A_i and A_j, and by assumption there are no such outcomes. Moreover, the union of all these events is B, since every outcomes lies in one of the A_i. So, by Axiom 3 we conclude that

$$\sum_{i=1}^{n} P(B \cap A_i) = P(B)$$



Example Consider the ice-cream salesman has to decide whether to order more stocks for Eid holiday. He estimates that if weather is sunny, he has 90% chance of selling all his stocks, if it is cloudy he has 60% chance of selling all his ice creams, if it is rainy he has 20% chance of selling all his stocks. According to weather forecast, the probability of sunshine is 30% the probability of cloud is 45% and probability of rain is 25%.

Let

$$P(A_1) = 0.3, P(A_2) = 0.45, P(A_3) = 0.25$$

let B is the event 'salesman sells all his stock', then according to the information we have

$$P(B|A_1) = 0.9, P(B|A_2) = 0.6, P(B|A_3) = 0.2$$

Then according to the theorem of total probability

$$P(B) = (0.9 \times 0.3) + (0.6 \times 0.45) + (0.2 \times 0.25) = 0.59$$

Further more,

$$P(A) = P(A|B)P(B) + P(A|\overline{B})P(\overline{B})$$

1.9 Baye's Rule

We prove Baye's theorem starting from the definition of conditional probability. Let A and B be two events

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\overline{A})P(\overline{A})}$$

Proof:

$$\begin{aligned} P(A|B) &= \frac{P(A \cap B)}{P(B)} \quad \text{By definition of conditional probability} \\ &= \frac{P(B|A)P(A)}{P(B)} \quad \text{By multiplication rule} \\ &= \frac{P(B|A)P(A)}{P(B \cap A) + P(B \cap \overline{A})} \quad \text{By Law of total probability} \\ &= \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\overline{A})P(\overline{A})} \quad \text{By multiplication rule} \end{aligned}$$

as required ♣.

1.10 Independent Events

It may happen that knowing that an event occurs does not change the probability of another event.

Definition

Motivated by the discussion above, we say that two events A and B are independent if

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

Note that the independence is a notion that depends on the probability. Independent and disjoint should not be confused. If two events are disjoint, then they are independent only if atleast one of them has probability 0. Indeed, if they are disjoint $P(A \cap B) = P(\emptyset) = 0$ so that $P(A \cap B) = P(A)P(B)$ only if $P(A) = 0$ or $P(B) = 0$. We may extend the definition of independence to more than three events. The events $A_1, A_2, \cdots, A_n$ are independent if and only if for every subset $\{A_{i_1}, A_{i_2}, \cdots, A_{i_k}\}(2 \leq k \leq n)$ of these events,

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_k})$$

i.e., we define an infinite set of events to be independent if and only if every finite subset of the events is independent.

To distinguish between the mutual exclusiveness (or disjointness) and independence of a collection of events we summarize as follows:

(a) If $\{A_i = 1, 2, \cdots, n\}$ is a sequence of mutually exclusive events, then

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

(b) If $\{A_i, i = 1, 2, \cdots, n\}$ is a sequence of independent events, then

$$P\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} P(A_i)$$

This implies that $P(A_i \cap A_2 \cap \cdots A_n) = P(A_1)P(A_2) \cdots P(A_n)$ and vice versa

Properties of Independence

Proposition-1: If A and B are independent, then A and $\overline{B}$ are independent.

we are given that $P(A \cap B) = P(A) \cdot P(B)$ and asked to prove that $P(A \cap \overline{B}) = P(A) \cdot P(\overline{B})$ from a few pages ago, we know that $P(A) = P(A \cap B) + P(A \cap \overline{B})$. Thus,

$$\begin{aligned} P(A \cap \overline{B}) &= P(A) - P(A \cap B) \\ &= P(A) - P(A) \cdot P(B) \qquad \text{Since A,B are independent} \\ &= P(A)(1 - P(B)) \\ &= P(A) \cdot P(\overline{B}) \end{aligned}$$

Proposition-2: If A and B are independent, then $\overline{A}$ and $\overline{B}$ are independent.

Proposition-3: Let events A, B, C be mutually independent. Then A and $B \cap C$ are independent, and A and $B \cup C$ are independent

Chapter 2

Random Variables

In this chapter, the concept of random variables is introduced. The main purpose of using a random variables is so that we can define probability functions that make it both convenient and easy to compute the probabilities of various events.

2.1 Random Variables

Definition:

Consider an experiment with sample space S.

1. Definition: A Random Variable is a characteristic of the outcome of an experiment.

2. Notation: Use capital letters to denote random variable (rvs.). Example: $X(s)$ is a rv. Use small letters to denote the realization of the random variable.

3. A random variable $X(s)$ is a single-valued real function that assigns a real number called value of $X(s)$ to each point s of S. Often we use single letter X for this function in space of $X(s)$.

Note!
$X()$ is rv and
x it's value

4. The sample space $\mathcal{S}$ is termed the domain of the r.v. X, and the collection of all numbers (values of X) is termed as range of r.v. X.

5. Example: Consider the experiment of choosing a student at random from a classroom. Then $\mathcal{S} = \{Jack, Dolores, \cdots\}$. Let $X(s)$ be a characteristics of student s. Then $X(\mathfrak{s})$ is a rv.

6. Types of random variables

   (a) Catagorical versus Numerical

   - $X(s) =$ gender of selected student is catagorical random variable and $X(s_2) = x_2 =$"Female" is a realization of the random variable.
   - $Y(s) =$ age of selected student is a numerical random variable and $Y(s_1) = Y_1 = 19.62$ is a realization of the random variable.

   (b) Continuous versus Discrete

   - If the possible value of a rv are countable then the rv is discrete.
   - if possible value of a rv are contained in open subsets (or half open subsets) of the real line, then the rv is continuous.

Physical Examples

Noise voltage at a given time and place, temperature at a given time and place, height of the next person to enter the room, and so on.

Examples of function

- We pick a ball randomly from a bag and we note its weight X and its diameter Y.
- We observe the temperature at a few different locations.
- We measure the noise voltage at different times.
- We track the evolution over time of the value of Cisco shares and we want to forecast the future values.

- A transmitter sends some signal and receiver observes the signal it receives and tries to guess which signal the transmitter sent.

Events defined by Random Variables

If X is a r.v. and x is a fixed real number, we can define the event $(X = x)$ as

$$(X = x) = \{s : X(s) = x\}$$

Similarly, for fixed numbers x,x_1 and x_2, we can define the following events.

$$\begin{aligned}(X \leq x) &= \{s : X(s) \leq x\} \\ (X > x) &= \{s : X(s) > x\} \\ (x_1 < X \leq x_2) &= \{s : x_1 < X(s) \leq x_2\}\end{aligned}$$

These events have probabilities that are denoted by

$$\begin{aligned}P(X = x) &= P\{s : X(s) = x\} \\ P(X \leq x) &= P(s : X(s) \leq x) \\ P(X > x) &= P(s : X(s) > x) \\ P(x_1 < X \leq x_2) &= P(s : x_1 < X(s) \leq x_2)\end{aligned}$$

2.2 Discrete Probability Distributions

1. Components of a Discrete Probability Model
   (a) A countable sample space, $S = \{s_1, s_2, \cdots\}$
   (b) A non-negative number $P(s)$ assigned to each outcome such that $\sum P(s_i) = 1$.

2.3 Distribution Function

1. Definition: The distribution function or Cumulative Distribution Function (cdf) of X is defined by

$$F_X(x) = P(X \leq x) \qquad -\infty < x < \infty$$

Most of the information about a random experiment described by the r.v. X is determined by the behavior of $F_X(x)$.

2. Properties of $F_X(x)$:

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

   (ii) $F_X(x_1) \leq F_X(x_2) \quad \text{if } x_1 < x_2$

   (iii) $\lim_{x\to\infty} F_X(x) = F_X(\infty) = 1$

   (iv) $\lim_{x\to-\infty} F_X(x) = F_X(-\infty) = 0$

   (v) $\lim_{x\to a^+} F_X(x) = F_X(a^+) = F_X(a^+) \qquad a^+ = \lim_{0<\epsilon\to 0} a + \epsilon$

   Property (i) follows because $F_X(x)$ is a property. Property (ii) shows that $F_X(x)$ is a nondecreasing function. Properties (iii) and (iv) follow from elementary definition of probability.

3. Determination of Probability from Distribution Function: From the definition of distribution a few lines above, we can compute other probabilities such as $P(a < X \leq b)$,$P(x > a)$, and $P(x < b)$.

$$\begin{aligned} P(a < X \leq b) &= F_X(b) - F_X(a) \\ P(X > a) &= 1 - F_X(a) \\ P(X < b) &= F_X(b^-) \qquad b^- = \lim_{0<\epsilon\to 0} b - \epsilon \end{aligned}$$

2.4 Discrete Random Variables

1. Definition: Let X be a r.v. with *cdf* $F_X(x)$ changes values only in jumps (at most a countable number of them) and is constant between jumps – that is

$F_X(x)$ is a staircase – then X is called a discrete random variable. Alternatively, X is a discrete r.v. only if its range contains a finite or countably infinite number of points.

2. Probability Mass Function: Suppose that the jumps in $F_X(x)$ of a discrete r.v. X occur at the points $x_1, x_2, \cdots$, where the sequence may be either finite or countably infinite, and we assume that $x_i < x_j$ if $i < j$. Then

$$F_X(x_i) - F_X(x_{i-1}) = P(X \leq x_i) - P(X \leq x_{i-1}) = P(X = x_i)$$

we define

$$p_X(x) = P(X = x)$$

The function $p_X(x)$ is called the *probability mass function* (pmf) of the discrete r.v. X.

3. Properties of $p_X(x)$

   (i) $0 < p_X(x_k) \leq 1 \quad k = 1, 2, \cdots$

   (ii) $p_X(x) = 0 \quad \text{if } x \neq x_k (k = 1, 2, \cdots)$

   (iii) $\sum_k p_X(x_k) = 1$

   The cdf $F_X(x)$ of a discrete r.v. X can be obtained by

$$F_X(x) = P(X \leq x) = \sum_{x_k \leq x} p_X(x_k)$$

This function is also known as *Probability mass function* (pmf).

(a) The pmf can be an equation, a table or a graph that shows how probability is assigned to possible values of a random variable.

(b) The distribution of probabilities across all possible values is called probability distribution may be displayed as (a) a table (b) a graph (c) an equation.

2.5 Continuous Random Variables and Probability Density Functions

1. Let X be a r.v. with cdf $F_X(s)$ is a continuous and also has a derivative $dF_X(x)/dx$ which exists everywhere except at possible a finite number of points and is piecewise continuous, then X is called a continuous random variable. Alternatively, X is a continuous r.v. only if its range contains an interval (either finite or infinite) of real numbers. Thus, if X is a continuous r.v. then

$$P(X = x) = 0$$

2. Probability Density Function let

$$f_X(x) = \frac{F_X(x)}{dx}$$

   The function $f_X(x)$ is called the *probability density function* (pdf) of the continuous r.v. X.

3. Properties of $f_X(x)$

   (a) $f_X(x) \geq 0$.

   (b) $\int_{-\infty}^{\infty} f_X(x)dx = 1$.

   (c) $f_X(x)$ is piecewise continuous.

   (d) $P(a < X \leq b) = \int_a^b f_X(x)dx$.

The cdf $F_X(x)$ of a continuous r.v. X can be obtained by

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(\xi)d\xi$$

if X is a continuous r.v., then

$$\begin{aligned} P(a < X \leq b) &= P(a \leq X \leq b) = P(a < X < b) \\ &= \int_a^b f_X(x)dx = F_x(b) - F_x(a) \end{aligned}$$

2.6 Median, Quartiles, Percentiles

Another measure commonly used for continuous random variables is *median*; this is the value m such that "half of the distribution lies to the left of the m and half to the right". More formally, m should satisfy $F_X(m) = 1/2$. It is not the same as the mean or expected value.
In other words if a data sequence is sorted in ascending order and value at the middle of the sequence is the 'median'. In case if the sequence has even length, the median is not unique, the median in this case is the average of the two center points. If there is a value m such that the graph of $y = f_X(x)$ is symmetric about $x = m$, then both the expected value and median of X are equal to m.
Quartile is value that represents 1/4th of the data. The lower quartile l and the upper quartile u are similarly defined by

$$F_X(l) = 1/4, \qquad F_X(u) = 3/4$$

Thus, the probability that X lies between l and u is $3/4 - 1/4 = 1/2$, so quartiles give an estimate of how spread-out the distribution is. More generally, we define nth percentile of X to be the value of x_n such that

$$F_X(X_n) = n/100,$$

i.e. The probability that X is smaller then x_n is $n\%$
$25th$ percentile is the lower quartile, $50th$ percentile is the median
Reminder: If the c.d.f. of X is $F_X(x)$ and the p.d.f is $f_X(x)$,then

- Differentiate F_X to get f_X and integrate f_X to get F_X;
- Use f_X to calculate $E(x)$ and $Var(X)$;
- Use f_X to calculate $P(a \leq X \leq b)$ (this is $F_X(b) - F_X(a)$), and the median and the percentiles of X.

2.7 Expected Value and Variance and Moments

Let X be a discrete random variable which takes the values $a_1, \cdots, a_n$. The expected value or mean of X (denoted as μ)is the value $E(X)$ given by the formula

$$E(X) = \sum_{i=1}^{n} a_i P(X = a_i)$$

That is we multiply each value of X by the probability that X take that value, and some these terms. The expected value is a kind of 'generalized average': if each of the values is equally likely, so that each has probability $1/n$, then $E(X) = (a_1 + \cdots + a_n)/n$, which is just the average of the values.
There is another interpretation of the expected value in terms of mechanics. If we put a mass P_i on the axis at position a_i for $i = 1, \cdots, n$, where $P_i = P(X = a_i)$, the center of mass of all these masses is at the point $E(x)$. If the random variable X takes infinitely many value, say $a_1, a_2, a_3 \cdots$, then we define the expected value of X to be infinite sum

$$\mu_x = E(X) = \sum_{i=1}^{\infty} a_i P(X = a_i)$$

Usually in practice we have random variables with finitely many values.
The variance of X is the number $Var(X)$ (denoted as σ_x^2) is given by

$$\sigma_x^2 = Var(X) = E(x^2) - [E(x)]^2$$

Here, X^2 is just the random variable whose values are squares of the values of X.Thus

$$E(X^2) = \sum_{i=1}^{n} {a_i}^2 P(X = a_i)$$

(or an infinite sum, if necessary). The next theorem shows that, if $E(X)$ is a kind of average of values of X, then $Var(X)$ is a measure of how spread-out the values are around their average.

Proposition Let X be a discrete random variable with $E(X) = \mu$. Then

$$Var(X) = E((X-\mu)^2) = \sum_{i=1}^{n}(a_i - \mu)^2 P(X = a_i)$$

$$\begin{aligned} Var(X) = E((X-\mu)^2) &= \sum_{i=1}^{n}(a_i - \mu)^2 P(X = a_i) \\ &= \sum_{i=1}^{n}({a_i}^2 - 2\mu a_i + \mu^2) P(X = a_i) \\ &= \left(\sum_{i=1}^{n}({a_i}^2 P(X = a_i)\right) - 2\mu\left(\sum_{i=1}^{n}(a_i P(X = a_i)\right) + \mu^2\left(\sum_{i=1}^{n} P(X = a_i)\right) \end{aligned}$$

Notice that the terms in the second terms contain just the definition of expected value and third term contains a summation over all events, applying the definitions to simplify we find,

$$\begin{aligned} E((X-\mu)^2) &= E(X^2) - 2\mu E(X) + \mu^2 \\ &= E(X^2) - E(X)^2 \end{aligned}$$

The *nth moment* of a r.v. X is defined by

$$E(X^n) = \begin{cases} \sum_k X_k^n p(x_k) & X : discrete \\ \int_{-\infty}^{+\infty} x^n f_x(x)dx & X : continuous \end{cases}$$

Remarks

- The expected value of X always lies between the smallest and largest values of X.
- The variance of X is never negative. From the definition of variance each of the term in the summation form $(a_i - \mu)^2$ (a square,hence non-negative) times $P(X = a_i)$(a probability hence non-negative).
- μ and σ^2 are 1st and 2nd order moments of a random variable.

Proposition: *Let C be a constant random variable with value c.Let X be any random variable*

(a) $E(C) = c, Var(C) = 0$

(b) $E(X+c) = E(X) + c, Var(X+c) = Var(X)$

(c) $E(cX) = cE(X), Var(cX) = c^2 Var(X)$

Proof:

(a) The random variable C takes the single value c with $P(C = c) = 1$. So $E(C) = c \cdot 1 = c$. And also

$$Var(C) = E(C^2) - E(C)^2 = 0$$

(b) This follows from previous lecture, we observe that the constant random variable C and any random variable X are independent.(This is true because $P(X = a, C = c) = P(X = a) \cdot 1$). Then

$$\begin{aligned} E(X+c) &= E(X) + E(C) = E(X) + c \\ Var(X+c) &= Var(X) + Var(C) = Var(X) \end{aligned}$$

(c) If $a_1, \cdots, a_n$ are the values of X, then $ca_1, \cdots, ca_n$ are the values of cX,and $P(cX = ca_i) = P(x = a_i)$ So

$$\begin{aligned} E(cX) &= \sum_{i=1}^{n} ca_i P(cX = ca_i) \\ &= c \sum_{i=1}^{n} a_i P(X = a_i) \\ &= cE(X) \end{aligned}$$

then

$$\begin{aligned} Var(cX) &= E(c^2X^2) - E(cX)^2 \\ &= c^2 E(X^2) - (cE(X))^2 \\ &= c^2 (E(X^2) - E(x)^2) \\ &= c^2 Var(X) \end{aligned}$$

Probability generating function

Let $G_X(x)$ be the *Probability generating function* of a random variable X. Then

1. $[G_X(x)]_{x=1} = 1$

2. $E(X) = [\frac{d}{dx} G_X(x)]$

3. $Var(X) = [\frac{d^2}{dx^2} G_X(x)] + E(X) - E(X)^2$

Probability generating function is also known as *characteristic equation* of the probability distribution function. [not discussed here further]

1	$\mu = E\{X\}$	mean
2	$\sigma^2 = E\{X^2\} - E^2\{X\}$	variance
3	$m_3 = E\left\{\left(\frac{X-\mu}{\sigma}\right)^3\right\}$	skewness
4	$m_4 = E\left\{\left(\frac{X-\mu}{\sigma}\right)^4\right\}$	kurtosis

2.8 Counting and Permutation

If the sample space of the experiment S is a finite set, one can compute the probabilities associated with each event by counting. In particular, consider an experiment in which each of the finite number of outcomes is equally likely. In such a situation, the probability of an event is equal to the number of outcomes that belong to the event of interest divided by the total number of outcomes. The most basic method of counting is so called the *multiplication rule*. Consider an experiment consisting of two completely separate experiments. The sample space of the first experiment has m outcomes i.e. $S_1 = \{a_1, a_2, \cdots, a_m\}$, whereas that of the second experiment has n outcomes i.e. $S_2 = \{b_1, b_2, \cdots, b_n\}$. Then, the sample space of the overall experiment space S has $m \times n$ Extending this method to more complicated situations, we have the following three basic rules for counting when sampling without replacement.

Theorem(Sampling without Replacement)

1. (Permutation) Consider an experiment in which k outcomes are sampled without replacement from the pool of n distinct outcomes ($k \leq n$). An ordered arrangement of k outcomes of such an experiment is called permutation, and the number of permutations of n distinct outcomes taken k at a time without replacement is given by

$$^nP_k = \frac{n!}{(n-k)!}$$

Further

If event A can occur in m ways, event B can occur in n ways, event C can occur in r ways, The ABC can occur in $m \times n \times r$ ways.

Examples: If you have 5 Shalwar Kameez, 3 waist coats, 3 hats; How many combinations of dress possible?

$$5 \times 4 \times 3$$

Examples: repetition of events, possible sequence of outcomes by rolling a dice

- 1 times 6^1
- 2 times 6^2

$\cdots \quad \cdots$

- r times 6^r

How many number plates can be made with 3 letters and 3 digits

$$26 \times 26 \times 26 \times 10 \times 10 \times 10$$

How many number plates begin with ABC

$$1 \times 1 \times 1 \times 10 \times 10 \times 10$$

If a plate is chosen at random the probability that it starts with ABC $\dfrac{10^3}{26^3 \times 10^3}$

How many ways 6 people can be arranged in a row.

$$6 \times 5 \times 4 \times 3 \times 2 \times 1 = 6!$$

How many arrangement are possible if only 3 are chosen.

$$6 \times 5 \times 4$$

Distinctly ordered sets are called permutation. The number of permutation of n objects taken r at a time

How many ways a team of 4 ppl can be arranged to take picture	How many ways in team of 4 ppl Captain and Vice Captain be chosen
$4 \times 3 \times 2 \times 1$	4×3
${}^4P_4$	${}^4P_2$

In how many ways can 5 seniors and 4 juniors be arranged on a bench?

$$\text{Seniors\&Juniors} \qquad \text{Juniors\&Seniors}$$

$$5! \times 4! \quad + \quad 4! \times 5!$$

$$2 \times {}^5P_5 {}^4P_4$$

Two guys want to sit together $2 \times 8!$ or $2 \cdot {}^8P_8$

Arrangement with Repetition

If we have n-element of which x are first type

y are second type

z are third type

Then the number of ordered selection or permutation is given by

$$\frac{n!}{x!y!z!}$$

How many different arrangements of word PARRAMATTA are possible

10 letters 4A's 2R's 2 T's

$$\frac{10!}{4!2!2!} = 37800$$

How many arrangement of letter of word REMAND are possible

- No Restriction ${}^6P_6 = 720$ or $6!$
- Begins with RE R E _ _ _ _ $= {}^4P_4 = 24$
- Not begins with RE Total - (Previous part)= 6! - 4! = 696

Arrangement with restrictions

- From digital 2,3,4,5,6; How many numbers can be greater than 4000 ${}^5P_5$.
- 4 digits (start with digit ≥4)=${}^3P_1 \times {}^4P_3$

$${}^5P_5 + {}^3P_1 \times {}^4P_3$$

- How many 4 digit numbers would be even? Even (ends with 2,4, or 6)= _ _ _${}^3P_1$

$$= {}^4P_3 \times {}^3P_1$$

2. (Combination) Consider an experiment in which k outcomes are sampled without replacement from the pool of n distinct outcomes $k \leq n$. An unordered arrangement of k outcomes of such an experiment is called a combination, and the number of combinations of n outcomes taken k at a time without replacement is given by

$${}^nC_k = \binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{{}^nC_k}{k!}$$

How many ways can a basketball team of 5 players be chosen from 8 players ${}^8C_5$.

A committee of 5 ppl is to be chosen from group of 6 men and 4 women.

- there are no restriction ${}^{10}C_5$
- one particular person must be chosen one committee $1 \times {}^9C_4$
- one particular women must be excluded from committee ${}^9C_5$

A committee of 5 ppl is to be chosen from group of 6 men and 4 women, how many committees are possible

- there are 3 men and 2 women ${}^6C_3 \times {}^4C_2$
- there are men only ${}^6C_5$
- there is majority of women i.e. 3W+2M or 4W+1M ${}^6C_2 \times {}^4C_3 + {}^6C_1 \times {}^4C_4$

3. (Partition) Consider an experiment in which the total of n distinct outcomes will be partitioned into k distinct events (i.e. sampling without replacement) containing $n_1, \cdots, n_k$ outcomes, respectively, where $n_i \geq 0$ for $i = 1, 2, \cdots, k$ and $\sum_{i=1}^{k} n_i$. The total number of partitions of n distinct outcomes into k distinct events given by

$$\binom{n}{n_1}\binom{n-n_1}{n_2}\binom{n-n_1}{n_2}\binom{n-n_1-n_2}{n_3}\cdots\binom{n-n_1-n_2\cdots n_{k-1}}{n_k}$$

which equals to

$$\frac{n!}{n_1!(n-n_1)!}\cdot\frac{(n-n_1)!}{n_2!(n-n_1-n_2)!}\cdots\frac{(n-n_1\cdots-n_{k-1})!}{n_k!(n-n_1\cdots-n_{k-1}-n_k)!}$$
$$\binom{n}{n_1,\cdots,n_k} = \frac{n!}{n_1!\cdots n_k!}$$

For combinations, it is important to note that the number of ways of selecting k outcomes from n distinct outcomes is the same as the number of ways of avoiding $n-k$ outcomes. That is $C_{n,k} = C_{n,n-k} \cdot C_{n,k}$ is also called *binomial coefficients* are the number of partitions is called *multinomial coefficient* because of the following theorems

Examples

1.By the permutation of the letters *abc* we mean all of their possible arrangement

$$abc, acb, bac, bca, cab, cba$$

There are 6 permutations of different things. As the number of things increases, their permutation grows astronomically.

2. Imagine putting the letters a, b, c, d into a hat, and then drawing two of them in succession. We can draw the first in 4 different ways, either a or b or c or d. After that has happened, there are 3 ways to choose the second, three ways to choose the third and 1 way to choose the last letter. Therefore the number of permutations is

$$4 \cdot 3 \cdot 2 \cdot 1 = 24$$

Permutation is commonly represented as ${}^{n}P_k$ means permutation of n different things taken k at a time.

3. The number of combinations of two out of the four letters A,B,C and D is found by letting $n = 4$ and $k = 2$. It is

$$\binom{4}{2} = \frac{4!}{2!2!} = 6$$

4. In counting permutations we consider *abc* as different from *bca*. But in combinations we are concerned only that *a*,*b* and *c* have been selected. *abc* and *bca* are the same combination. All the combinations of *abcd* taken three at a time are represent as follows:

$$abc, abd, acd, bcd$$

Number of combinations and permutations are related to each other via following expression

$$^nC_k = \frac{^nP_k}{k!}$$

Theorem(Sampling with Replacement)

1. (Permutation) Consider an experiment in which k outcomes are sampled with replacement from the pool of n distinct outcomes. The number of permutations of n outcomes takes k at a time with replacement is given by n^k

2. (Combination) consider an experiment in which k outcomes are sampled with replacement from the pool of n distinct outcomes. The total number of combinations of n outcomes k at a time with replacement is given by

$$\binom{n+k-1}{k} = \frac{(n+k-1)!}{k!(n-1)!}$$

Theorem(Binomial and Multinomial Theorems)

1. (Binomial) For any real number x and y and a non-negative integer n

$$(x+y)^n = \sum_{i=0}^{n} \binom{n}{i} x^i y^{n-i}$$

2. (Multinomial) let k and n be positive integers and A be a set of vectors $a = (n_1, n_2 \cdots, n_k)$ such that each n_i is a non-negative integer and $\sum_{i=1}^{k} n_i = n$. Then, for any real number $x_1, x_2, \cdots, x_k$

$$(x_1 + x_2 + \cdots + x_k)^n = \sum_{a \in A}^{n} \binom{n}{n_1 n_2 \cdots n_k} x_1^{n_1} x_2^{n_2} \cdots x_k^{n_k}$$

2.9 Some Discrete Random Variables

In this section we describe random variables each depending on one or more parameters. We describe their pmf , mean and variance.

Let X be a random variable taking values $a_1, a_2, \cdots, a_n$. We assume that these are arranged in the ascending order $a_1 < a_2 < \cdots < a_n$. The commutative distributions function or cdf of X is given by

$$F_X(a_i) = P(X \leq a_i)$$

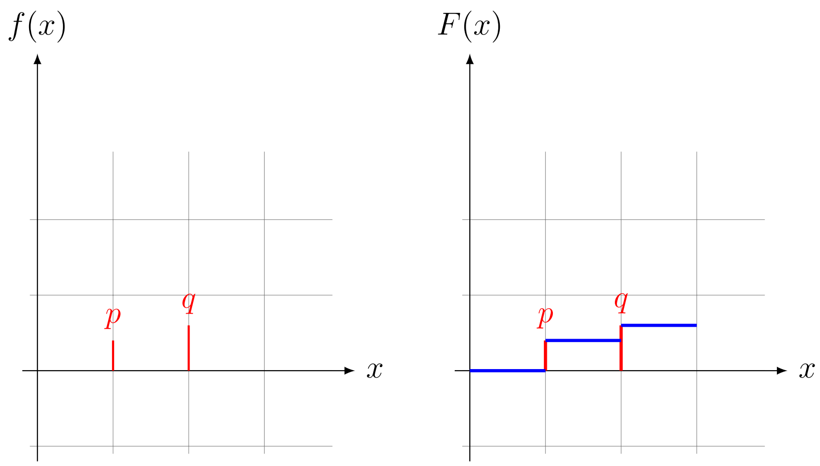
we see that it can be expressed in terms of the pmf of X as follows:

$$F_x(a_i) = P(X = a_1) + \cdots + P(X = a_i) = \sum_{j=1}^{i} P(X = a_j)$$

In the other direction, we can recover the pmf from the cdf

$$P(X = a_i) = F_x(a_i) - F_x(a_{i-i})$$

We won't use the cdf of a discrete random variable except for looking up the tables.It is much more important for continuous random variables



Bernoulli random variable Bernoulli(P)

A Bernoulli random variable is the simplest type of all. It takes two values 0 and 1.

The Probability mass function looks like

$$p_x(k) = P(X = k) = p^k(1-p)^{1-k} \quad k = 0, 1$$

where $0 \leq p \leq 1$, the pmf in table form is given below

x	0	1
$P(X = x)$	p	q

Here,P is the probability that $X = 1$, it can be any number between 0 and 1. Necessarily q (the probability that $X = 0$) is equal to $1 - p$. So p determines everything.

For Bernoulli random variable X, we sometimes describe the experiment as a 'trial', the event $X = 1$ as 'success', and the even $X = 0$ as a 'failure'.

For example, if a biased coin has probability p of coming down heads, then the number of heads that we get when we toss the coin once is Bernoulli(p) random variable.

The cumulative distribution function is given as follows

$$F_X(x) = \begin{cases} 0 \quad if\ x < 0; \\ 1-p \quad if\ 0 \leq x \leq 1 \\ 1 \quad if x \geq 1 \end{cases}$$

The random variable I_A is called the indicator variables of A, because its value indicates whether or not A occurred. It is Bernoulli(P) random variable, where $p = P(A)$.

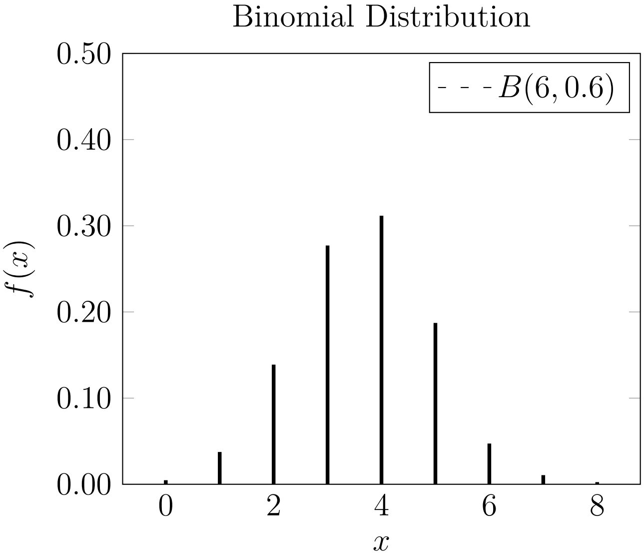
Calculation of the expected value and variance of a Bernoulli random variable is easy. Let $X \sim$Bernoulli(p) i.e. X has the same pmf as Bernoulli(p).

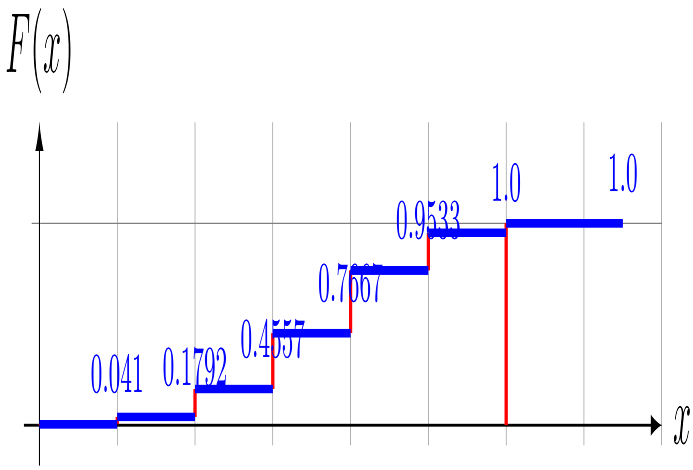
$$E(X) = 0 \cdot q + 1 \cdot p = p$$
$$Var(X) = 0^2 \cdot q + 1^2 \cdot p - p^2 = p - p^2 = pq$$

(Remember that $q = 1 - p$)

Binomial Random variable Bin(n,p)

Remember that for Bernoulli random variable, we describe the even $X = 1$ as a 'success'. Now a Binomial random variable counts the number of successes in independent trials each associated with Bernoulli(p) random variable.





For example, suppose that we have a biased coin for which the probability of heads is p. We toss the coin n times and count the number of heads obtained. This number is a Bin(n,p) random variable.

A Bin(n,p) random variable X takes the values $0, 1, 2, \cdots, n$ and the pmf of X is given by

$$P(X = k) = {}^nC_k \, q^{n-k} p^k$$

For $k = 0, 1, 2, \cdots, n$, where $q = 1 - p$. ${}^nC_k$ is known as the binomial coefficient. This is because there are ${}^nC_k$ different ways of obtaining k heads in a sequence of n throws (the number of choices of the positions in which the heads occur), and the probability of getting k heads and $n - k$ tails in a particular order is $q^{n-k}p^k$.
The corresponding cdf of X is

$$F_X(x) = \sum_{k=0} n^nC_k p^k q^{n-k} \quad \underline{n \leq x < n+1}$$

The table of probabilities for outcome of heads and tails from four Bernoulli's trials

i.e. Bin(4,p)

k	0	1	2	3	4
$P(X = k)$	q^4	$4q^3p$	$6p^2q^2$	$4qp^3$	p^4

Note: When we add all the probabilities in the table we get

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

As it should be by the definition of binomial theorem

$$(x+y)^n = \sum_{k=0}^{n} {}^nC_k x^{n-k} y^k$$

This argument explains the name of the binomial random variable

If X Bin(n,p), then

$$E(X) = np, \quad Var(X) = npq$$

Proof:

Let us consider a random variable X Bin(n,p). We have

$$p_k = P(X = k) = {}^nC_k q^{n-k} p^k$$

So the probability generating function is

$$\sum_{k=0}^{n} {}^nC_k q^{n-k} p^k = (q + px)^n$$

By the Binomial Theorem,Putting $x = 1$ gives $(q + p)^n = 1$, Differentiating once, using the Chain Rule, we get $np(q + px)^{n-1}$. Putting $x = 1$ we find that

$$E(X) = np$$

Differentiating again we get $n(n-1)p^2(q + px)^{n-2}$. Putting $x = 1$ gives $n(n-1)p^2$. Not adding $E(X^2) - E(X)^2$,we get

$$Var(X) = n(n-1)p^2 + np - n^2p^2 = np - np^2 = npq$$

Example: Grade of Service: An internet service provider has installed c modems to serve a population of n costumers. It is estimated that at any given time, each customer will need a connection with probability of p, independently with others. What is the probability that there are more customers needing a connection than there are modems?

We are interested in the probability that more than c customer simultaneously need a connection. It is equal to

$$\sum_{k=c+1}^{n} p(k)$$

where

$$p(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

are the binomial probabilities. For instance, if $n = 100$, $p = 0.1$ and $c = 15$, the desired probability turns out to be 0.0399.

Poisson Random Variable $Poisson(\lambda)$

The Poisson random variable,unlike the ones we have seen before, is very closely connected with continuous things.

Suppose that incidents occur at random times, but at a steady rate overall. The best example is radioactive decay: atomic nuclie decay randomly, but the average number λ which will decay in a given interval is constant. The Poisson random variable X counts the number of 'incidents' which occur in a given random interval. So if, on average, there are 2.4 nuclear decays per second, then the number of decays in one second starting now is Poisson(2.4) random variable.

Another example might be the number of telephone calls a minute to a busy telephone number.

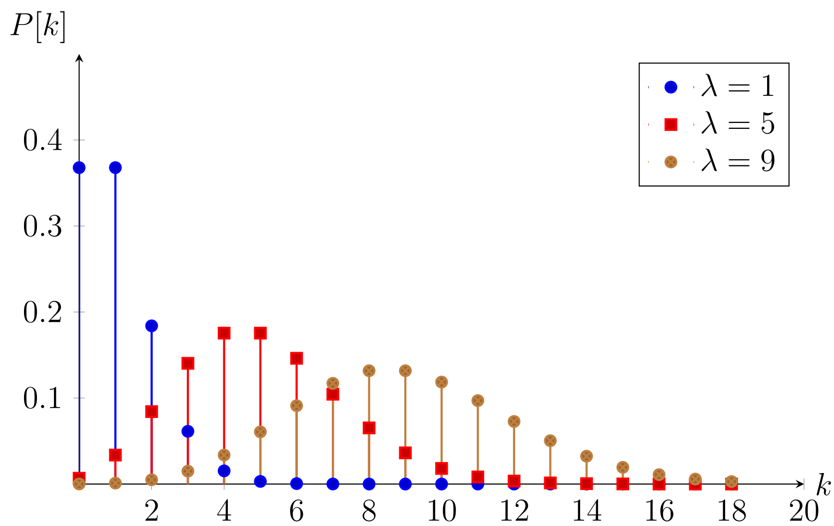
$$P(X = k) = \frac{\lambda^k}{k!}e^{-\lambda} \quad k = 0, 1, \cdots$$

Lets check that all these possibilities add up to one. We get

$$\left(\sum_{k=0}^{\infty} \frac{\lambda^k}{k!}e^{-\lambda}\right) = e^{\lambda} \cdot e^{-\lambda} = 1$$

The corresponding cdf of X is

$$F_X(x) = P(X = k) = e^{-\lambda} \sum_{k=0}^{n} \frac{\lambda^k}{k!} \quad n \leq x < n + 1$$



Since the expression in the bracket is sum of exponential series.
Drawing analogies with Binomial random variable. The Expected value and variance of a $Poisson(\lambda)$ random variable are given by

$$E(X) = Var(X) = \lambda$$

If $X \sim Poisson(\lambda)$, then

$$G_X(x) = \sum_{'k=0}^{\infty} \frac{(\lambda x)^k}{k!} e^{-\lambda} = e^{\lambda(x-1)}$$

Differentiation gives $\lambda e^{\lambda(x-1)}$,so $E(X) = \lambda$. Differentiating again gives $\lambda^2 e^{\lambda(x-1)}$,so

$$Var(X) = \lambda^2 + \lambda - \lambda^2$$

An other example of Poisson distribution is that if a fisher catches fishes at the average rate of 2.4 fishes an hour, then the probability that fisher can catch no fish at all in the next hour is 0.0907, while the probability that he catches five fishes or fewer fishes is 0.9643, so the probability of catching six or more fishes is 0.0357.
Another situation where Poisson distribution arises. Suppose that we are looking for a very rare event which only occurs once in a 1000 trials on average. So we conduct 1000 independent trials. How many occurrences of the event do we see? This number is really a Binomial random variable $Bin(1000, 1/1000)$. But it turns out to be Poisson(1),to be a very good approximation. So, for example the probability that the event doesn't occur is about $1/e$.
The general rule is:
If n is large, p is small, and $np = \lambda$, then $Bin(n, p)$ can be approximated by $Poisson(\lambda)$

2.10 Continuous Random Variables

Continuous random variable is a set or real numbers, or perhaps the non-negative real numbers or just an interval. The crucial property is that , any real number a, we have $(X = a) = 0$ that is, the probability that the height of a random student or the time I have to wait for a bus, is precisely a, is zero. So we can't use the probability mass function for continuous random variables;it would be zero and give no information.

We use the cumulative distribution function, as discussed before

$$F_X(x) = P(X \leq x)$$

Proposition: *The c.d.f is an increasing function(this means* $F_X(x) \leq F_X(y)$ *if* $x < y$*),and approaches the limits* 0 *a* $x \to -\infty$ *and* 1 *as* $x \to \infty$ The function is increasing because,if $x < y$,then

$$F_X(y) - F_X(x) = P(X \leq y) - P(X \leq x) = P(x < X \leq y) \geq 0$$

Also $F_X(\infty) = 1$ because X must certainly take some finite value; and $F_X(-\infty) = 0$ because no value is smaller than $-\infty$

Another important function is the probability density function F_X. It is obtained by differentiating the c.d.f.:

$$f_X(x) = \frac{d}{dx} F_X(x)$$

Now $F_X(x)$ is a non-negative, since it is the derivative of an increasing function. If we know $f_X(x)$, then F_X is obtained by integrating. Because $F_X(-\infty) = 0$, we have

$$f_X(x) = \int_{-\infty}^{x} f_X(t)dt$$

Note the use of dummy variable" t in this integral. Note also that

$$P(a \leq X \leq b) = F_X(b) - F_X(a) = \int_{a}^{b} f_X(t)dt$$

The pmf is the probability that the value of X lies in a very small interval from x to $x + h$ is approximately $f_X(x) \cdot h$. So, although the probability of getting exactly the value x is zero, the probability of being close to x is proportional to $f_X(x)$.

There is mechanical analogy which simply illustrates the scenario. We have modeled the discrete random variable X by placing at each value a of X a mass equal to $P(X = a)$. Then the total mass is one, and the expected value of X is the center of mass. For a continuous random variable, imagine we have a wire of variable thickness, so that the density of the wire (mass per unit length) at the point x is equal to $f_X(x)$. Then again the total mass is one: the mass to the left of x is $F_X(x)$; and again it will hold that the center of mass is at $E(X)$.

Most facts about continuous random variables are obtained by replacing the pmf by the pdf and replacing sums by integrals. Thus expected value of X is given by

$$E(X) = \int_{-\infty}^{+\infty} x f_X(x) dx$$

and the variance (as before)

$$Var(X) = E(X^2) - E(X)^2$$

where

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x) dx$$

It is also true that $Var(X) = E((X - \mu)^2)$,where $\mu = E(X)$.

$$f_X(x) = \begin{cases} 2x \text{ if } 0 \leq x \leq 1 \\ 0 \; otherwise. \end{cases}$$

The support of X is the interval $[0, 1]$. We check the integral

$$\int_{-\infty}^{\infty} f_X(x) dx = \int_0^1 2x dx = [x^2]_{x=0}^{x=1} = 1$$

The cumulative distribution function of X is

$$F_X(x) = \int_{-\infty}^{x} f_X(t) dt = \begin{cases} 0 \text{ if } x < 0 \\ x^2 \text{ if } 0 \leq x \leq 1 \\ 1 \text{ if } x > 1 \end{cases}$$

We have

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f_X(x) dx = \int_0^1 2x^2 dx = \frac{2}{3} \\ E(X) &= \int_{-\infty}^{\infty} x^2 f_X(x) dx = \int_0^1 2x^3 dx = \frac{1}{2} \\ Var(X) &= \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{18} \end{aligned}$$

2.11 Some Continuous Random Variables

In this section we examine three important continuous random variables: the uniform, exponential and normal.

Uniform random variable U(a,b)

Let a and b be real numbers with $a < b$. A uniform random variable on the interval $[a, b]$ is roughly speaking,"equally likely to anywhere in the interval". In other words, it probability density is function is constant on the interval $[a, b]$ (and zero outside the interval). The integral of the p.d.f. is the area of a rectangle of height c and base $b - a$: this area must be 1, so $c = 1/(b - a)$. This, the p.d.f. of the random variable $X \sim U(a, b)$ is given by

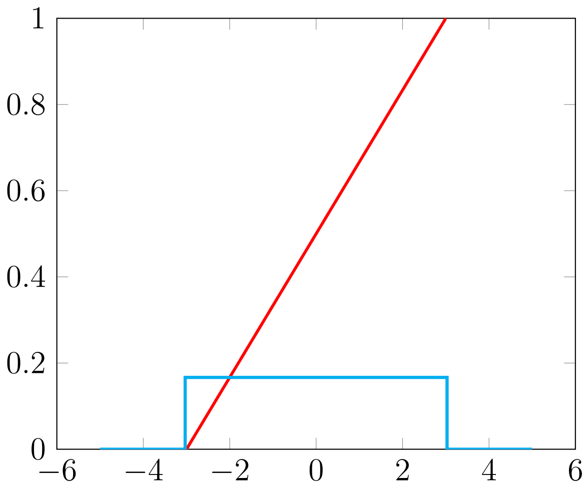
$$f_X(x) = \begin{cases} 1/(b-a) \text{ if } a \leq x \leq b, \\ 0 \; otherwise. \end{cases}$$

By integration, we find that the c.d.f. is

$$F_X(x) = \begin{cases} 0 \text{ if } x < a, \\ (x-a)/(b-a) \text{ if } a \leq x \leq b, \\ 1 \;\; x > b. \end{cases}$$

Further calculation (or symmetry of the p.d.f.) show that the expected value and the median of X are both given by $(a + b)/2$ (the midpoint interval), while $Var(x) = (b - a)^2/12$.

The uniform random variable really doesn't arise in practicle situations. Most computer systems include a random number generator, which apparently produces independent values of a uniform random variable on the interval $[0, 1]$.Of course, they are not really random, since the computer is a deterministic machine; but there should be no obvious pattern to the numbers produced, and in a large number of trials they should be distributed uniformly over the interval.



Exponential random variable $Exp(\lambda)$

The exponential random variable arises in the same situation as the Poisson: be careful not to confuse them. We have events which occur randomly but at a constant average rate of λ per unit time. (e.g radioactive decays, fish catch). The Poisson random variable, which is discrete, counts how many events will occur in the next time unit. The exponential random variable, which is continuous measures exactly how long from now it is until next event occurs. Not that it takes non-negative real numbers are values.

If $X \sim Exp(\lambda)$, the pdf of X is

$$f_X(x) = \begin{cases} 0 \text{ if } x < 0, \\ \lambda e^{-\lambda x} \text{ if } x \geq 0 \end{cases}$$

By integration, we find c.d.f. to be

$$F_X(x) = \begin{cases} 0 \text{ if } x < 0, \\ 1 - \lambda e^{-\lambda x} \text{ if } x \geq 0 \end{cases}$$

Further calculations yield

$$E(X) = 1/\lambda, \quad Var(X) = 1/\lambda^2$$

Normal random variable $N(\mu, \sigma^2)$

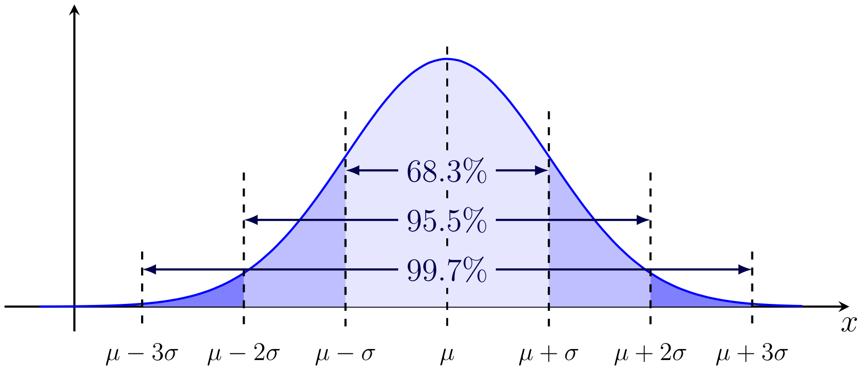
The normal random variable is the common-most of all applications, and the most important. It is also popularly known as Gaussian Distribution, If you take sum (or the average) of n independent random variables with the same distribution as X, the result will be approximately normal, and will become more and more like normal variable as n grows. This partly explains why a random variable affected by many independent factors, like a man's height, has an approximately normal distribution.

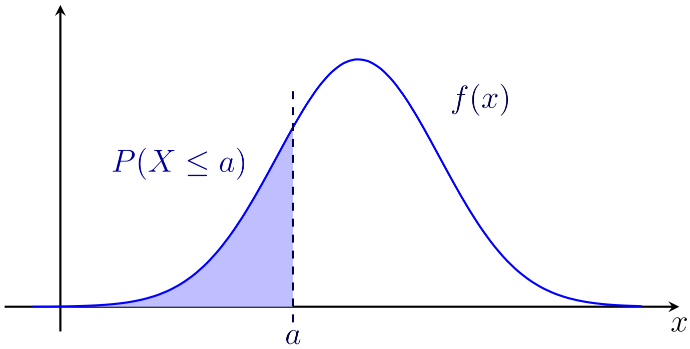
2.12 Normal Distribution

Also known as the Gaussian distribution, a well-known and widely applicable random variable distribution.The probability density function is defined as

$$P_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}(\frac{(x-\mu)}{\sigma})^2}$$

where μ is the mean and σ^2 is the variance of the distribution. The distribution is typically specified as $\underline{\text{n}}\text{orm}(\mu,\sigma^2)$. The plot of the gaussian distribution is illustrated below





For the Gaussian rv the probability can be easily calculated using the expression $z = \dfrac{x-\mu}{\sigma}$, where z is the scaled parameter which assumes value as provided in the table below:

Few exercises have been considered in the class lectures.

	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

Chapter 3

Joint Probability Functions

3.1 Joint Pmf of two random variables

Let X be a random variable taking $a_1, \cdots, a_n$ and let Y be random variable take the value $b_1, \cdots, b_m$. We say that X and Y are independent if for any possible values i and j, we have

$$P(X = a_i, Y = b_j) = P(X = a_i) \cdot P(Y = b_j)$$

Here $P(X = a_i, Y = b_j)$ means the probability of the event take the value a_i and Y takes the value b_j. So we could re-state the definition as follows:

The random variables X and Y are independent if, for any value a_i of X and for any value b_j of Y,the events $X = a_i$ and $Y = b_j$ are independent (events)

some examples should be inserted for clearing up the concept of independent events an

Theorm X and Y be random variables

(a) $E(X + Y) = E(X) + E(Y)$.

(b) If X and Y are independent, then $Var(X + Y) = Var(X) + Var(Y)$.

I have two red pens,one green pen and one blue pen, I choose two pens without replacement.Let X be the number of red pens that I choose and Y the number of green pends. Then the pmf of X and Y is given in the following table

	0	1
0	0	$\frac{1}{6}$
1	$\frac{1}{3}$	$\frac{1}{3}$
2	$\frac{1}{6}$	0

The row and column sums give us pmf for X and Y

a	0	1	2
$P(X=a)$	$\frac{1}{6}$	$\frac{2}{3}$	$\frac{1}{6}$

b	0	1
$P(Y=b)$	$\frac{1}{2}$	$\frac{1}{2}$

Consider the joint pmf of X and Y. The random variable $X+Y$ takes the values a_i+b_j for $i=1\cdots,n$ and $j=1\cdots,m$. Now the probability that it takes a give value c_k is the sum of probabilities $P(X=a_i,Y=b_j)$ over all i and j such that $a_i+b_j=c_k$. Thus

$$\begin{aligned} E(X+Y) &= \sum_k c_k P(X+Y=c_k) \\ &= \sum_{i=1}^n \sum_{j=1}^m P(X=a_i,Y=b_j) \\ &= \left(\sum_{i=1}^n a_i \sum_{j=1}^m P(X=a_i,Y=b_j)\right) + \left(\sum_{j=1}^m b_j \sum_{i=1}^n P(X=a_i,Y=b_j)\right) \end{aligned}$$

now $\sum_{j=1}^m P(X=a_i,Y=b_j)$ is a row sum of the joint pmf table so is equal to $P(X=a_i)$ and similarly $\sum_{i=1}^n P(X=a_i,Y=b_j)$ is a column sum and is equal to $P(Y=b_j)$. So

$$\begin{aligned} E(X+Y) &= \sum_{i=1}^n a_i P(X=a_i) + \sum_{j=1}^m b_j P(Y=b_j) \\ &= E(X)+E(Y) \end{aligned}$$

The variance is bit trickier. We start by calculating

$$E((X+Y)^2) = E(X^2+2XY+Y^2) = E(X^2)+2E(XY)+E(Y^2)$$

Now we have to consider the term $E(XY)$. For this we have to make the assumption that X and Y are independent, that is

$$P(X=a_i,Y=b_j) = P(X=a_i)\cdot P(Y=b_j)$$

we have

$$\begin{aligned}
E(XY) &= \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j P(X=a_i, Y=b_j) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j P(X=a_i)P(Y=b_j) \\
&= \left(\sum_{i=1}^{n} a_i P(X=a_i)\right)\cdot\left(\sum_{j=1}^{m} b_j P(Y=b_j)\right) \\
&= E(X)\cdot E(Y)
\end{aligned}$$

So

$$\begin{aligned}
Var(X+Y) &= E((X+Y)^2) - (E(X+Y))^2 \\
&= (E(X^2) + 2E(XY) + E(Y^2)) - (E(X)^2 + 2E(X)E(Y) + E(Y)^2) \\
&= Var(X) + Var(Y)
\end{aligned}$$