B. SRIKRISHNA

EE 511

Aug - Nov 2004

COMMUNICATION THEORY

Lecture 1:

The Theory of Probability: (Papoulis: Preface to the first edition)

- Theory should be developed axiomatically.
- Theory should be viewed as a conceptual structure and its conclusions should ~~not~~ rely not on intuition but on logic.
- Motivating sections should be separated from the deductive part of the theory.
- Theory must be mathematical (deductive) in form but without the generality or rigour of mathematics.
- Need for an axiomatic approach and a clear distinction between assumptions and logical conclusions.
- Introduce special topics as illustrations of the general theory.

Introduction

Probability theory studied in order to

- model and analyze "unpredictability", "randomness", "something not deterministic".
- describe and predict average behaviour of non-deterministic systems.

Eg. Communication systems

Information needs to be random to be meaningful

Channel is also random

Receiver needs to model "information", channel and extract information from the received signal.

Control systems

Model for plant parameters (to be controlled)

Noisy observations

Axiomatic theory based on tools from set theory

Steps in application of probability to real problems:

- Observation (physical)
- Deduction (conceptual) — I
- Prediction (physical).

Experiment → Outcome

Event as a set of outcomes.

Classical definition of probability

The probability $P(A)$ of an event A is determined as

$$P(A) = \frac{N_A}{N}$$

where N is the number of possible outcomes & N_A is the number of outcomes that are favourable to the event A.

Problem: Assume that all ~~events~~ outcomes are equally likely.

Eg: Where it works

(1) Die experiment : outcomes $\{1, 2, 3, 4, 5, 6\}$.

$$P(\text{even}) = \frac{3}{6}$$

where it does not work

(2) Two-Die experiment : outcomes $\{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$

$$P(7) \neq \frac{1}{11}.$$

If outcomes are $(1,1), (1,2), \dots, (1,6)$

$\vdots$

$(6,1), (6,2), \dots, (6,6)$

Then

$$P(7) = \frac{6}{36} \checkmark$$

Relative frequency definition

The probability $P(A)$ of an event A is the limit

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

where n_A is the number of occurences of A
n is the number of trials

(Theory founded on observation)

In a physical experiment, the numbers n_A and n might be large but they are only finite; their ratio cannot be equated, even approximately, to a limit. Therefore, $P(A) = \lim_{n \to \infty} \frac{n_A}{n}$ must be accepted as a "hypothesis", not as a number that can be determined experimentally.

Axiomatic definition. (Kolmogorov 1933)

S: Certain event that occurs in every trial (sample space).

$$P(A) \geq 0$$

$$P(S) = 1$$

If A and B are mutually exclusive, then

$$P(A+B) = P(A) + P(B)$$

This axiomatic definition will be discussed in more detail and used to develop probability theory.

Occasionally, the classical definition is used in step 1 of (I) in page 2.

Relative frequency interpretation is used to show the link between theory and applications (step 3 of (I)).

Lecture 2:

Preliminaries: Set Theory: Definitions.

* A set is a collection of objects called elements.

* A subset B of a set A is another set whose elements are also elements of A.

* All sets under consideration will be subsets of S which we shall call space.

* The empty or null set is the set that contains no elements. (ϕ)

Set operations:

* Unions

The sum or union of two sets A and B is a set whose elements are all elements of A or of B or of both. ($A+B$ or $A \cup B$)

* Intersections

The product or intersection of two sets A and B is a set consisting of all elements that are common to the sets A and B. (AB or $A \cap B$)

* Complements

The complement $\bar{A}$ of a set A is the set consisting of all elements of S that are not in A.

Mutually exclusive sets:

Two sets A and B are called mutually exclusive or disjoint if they have no common elements, i.e., if $AB = \phi$. Several sets $A_1, A_2, \ldots$ are mutually exclusive if $A_i A_j = \phi$ for every i and $j \neq i$.

Partitions:

A Partition of a set S is a collection of mutually exclusive subsets A_i of S whose union equals S.

$$A_1 + \cdots + A_n = S, \quad A_i A_j = \phi \text{ for } i \neq j$$

De Morgan's law:

$$\overline{A+B} = \bar{A}\,\bar{B}, \quad \overline{AB} = \bar{A} + \bar{B}$$

If in a set identity we replace all sets by their complements, all unions by intersections, and all intersections by unions, the identity is preserved.

Eg. $A(B+C) = AB + AC$

$$\Rightarrow \bar{A} + \bar{B}\bar{C} = (\bar{A}+\bar{B})(\bar{A}+\bar{C}) \quad \text{— (I)}$$

In an identity like (I), we can replace $\bar{A}$ by A, $\bar{B}$ by B and $\bar{C}$ by C and the identity is still preserved.

Probability Theory: Terminology

Sample space S is the set of all experimental outcomes. S is called the certain event, its elements outcomes, and its subsets events. The empty set ϕ is the impossible event and the event consisting of a single element is an elementary event.

In the applications of probability theory to physical problems, the identification of experimental outcomes is not always unique. We will assume that the outcomes are clearly identified.

The Axioms of Probability:

The probability of each event A, $P(A)$ is assigned so as to satisfy the following three conditions:

(1) $P(A) \geq 0$

(2) $P(S) = 1$

(3) If $AB = \phi$, then

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

Some simple properties follow:

1) $P(\phi) = 0$

$$\left[A\phi = \phi \Rightarrow P(A+\phi) = P(A) + P(\phi) \right.$$

$$\left. \Rightarrow P(\phi) = P(A) - P(A) = 0 \right]$$

These two properties also imply that

* If $A \in \mathcal{F}$ and $B \in \mathcal{F}$, then $AB \in \mathcal{F}$.
* $S \in \mathcal{F}$, $\phi \in \mathcal{F}$.
* All sets that can be written as unions & intersections of finitely many sets in $\mathcal{F}$ are also in $\mathcal{F}$. (not necessarily the case for infinitely many sets).

$\boxed{\text{Lecture 3:}}$

Borel fields:

Suppose that $A_1, \ldots, A_n, \ldots$ is an infinite sequence of sets in $\mathcal{F}$. If the union and intersection of these sets also belongs to $\mathcal{F}$, then $\mathcal{F}$ is called a Borel field.

* The class of all subsets of a set S is a Borel field.
* Suppose $\mathfrak{S}$ is a class of subsets of S that is not a field. We can add other subsets of S to form a field with $\mathfrak{S}$ as its subset.

There exists a smallest Borel field containing all the elements of $\mathfrak{S}$.

Eg: $S = \{a, b, c, d\}$

$\mathfrak{S}: \{a\}, \{b\}$.

Smallest Borel field containing $\mathfrak{S}$,

$$\{\phi\}, \{a\}, \{b\}, \{a, b\}, \{c, d\}, \{b, c, d\}, \{a, c, d\}, S.$$

Extension of axiom III:

$$P(A_1 + \cdots + A_n) = P(A_1) + \cdots + P(A_n)$$

Axiom of infinite additivity:

If the events $A_1, A_2, \cdots$ are mutually exclusive, then

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

Summary:

An experiment is specified in terms of:

(a) The set S of experimental outcomes.

(b) The Borel field of all events of S.

(c) The probabilities of these events.

Examples:

(1) Countable spaces:

If the space S consists of N outcomes and N is a finite number, then the probabilities of all events can be expressed in terms of the probabilities of the N elementary events. $(p_1, p_2, \cdots, p_N)$.

From the axioms it follows that we need

$$p_i \geq 0 \quad \text{for all } i$$

$$\text{and } \sum_{i=1}^{N} p_i = 1.$$

Similar statements can be made if S consists of an infinite but countable number of elements.

(2) Real line:

Suppose that S is the set of all real numbers.

- It can be shown that it is impossible to to satisfy the axioms.

* To construct a Borel field on the real line, we consider as events all intervals $x_1 \le x \le x_2$ and their countable unions and intersections.

This field

- is the smallest Borel field that includes all half-lines $x \le x_i$ where x_i is any number.
- contains all open and closed intervals, all points, every set of points on the real line that is of interest in the applications.
- does not include all subsets of S.

* To complete the specification of the experiment, it suffices to assign probabilities to the events $\{x \le x_i\}$. All other probabilities can then be determined from the axioms.

* We will return to this example when we discuss random variables.

Conditional Probability:

The conditional probability of an event A assuming an event B, denoted by $P(A/B)$, is by definition

$$P(A/B) = \frac{P(AB)}{P(B)}$$

where we assume that $P(B)$ is not 0.

It directly follows that

* if $B \subset A$, then $P(A/B) = 1$

* if $A \subset B$, then $P(A|B) = \dfrac{P(A)}{P(B)} \geq P(A)$.

We can show that the above definition of conditional probabilities satisfies the axioms of probability.

(1) $P(A|B) \geq 0$

since $P(AB) \geq 0$ and $P(B) > 0$.

(2) Since $B \subset S$, $P(S|B) = 1$.

(3) If A and C are mutually exclusive, AB and CB are also mutually exclusive.

$$\Rightarrow P(A+C|B) = \frac{P(AB+CB)}{P(B)}$$

$$= \frac{P(AB)}{P(B)} + \frac{P(CB)}{P(B)}$$

$$= P(A|B) + P(C|B).$$

Frequency interpretation: $P(A|B) \approx \dfrac{n_{AB}}{n_B}$.

Total Probability:

If $A_1, A_2, \ldots, A_n$ is a partition of S and B is an arbitrary event, then

$$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots + P(B|A_n)P(A_n)$$

Proof:

$$B = BS = B(A_1 + A_2 + \cdots + A_n)$$

$$= BA_1 + BA_2 + \cdots + BA_n.$$

Since $A_1, A_2, \ldots, A_n$ are mutually exclusive,

$$\Rightarrow P(B) = P(BA_1) + P(BA_2) + \cdots + P(BA_n)$$

$$= P(B/A_1)\,P(A_1) + \cdots + P(B/A_n)\,P(A_n).$$

$$\left(\text{since } P(BA_i) = P(B/A_i)\,P(A_i)\right)$$

Lecture 4:

Bayes' Theorem:

$$P(A_i/B) = \frac{P(B/A_i)\,P(A_i)}{P(B)}$$

$$= \frac{P(B/A_i)\,P(A_i)}{\sum_{i=1}^{n} P(B/A_i)\,P(A_i)}$$

$$\left(\text{Since } P(BA_i) = P(B/A_i)\,P(A_i) = P(A_i/B)\,P(B)\right).$$

Independence:

Two events A and B are called independent if $P(AB) = P(A)\,P(B)$.

Frequency interpretation: $P(A) \approx \frac{n_A}{n}$ $P(B) \approx \frac{n_B}{n}$ $P(AB) \approx \frac{n_{AB}}{n}$

$$\frac{n_A}{n} = \frac{n_{AB}}{n_B}$$

The relative frequency of occurence of A in the original sequence of n trials equals the relative frequency of occurence of A in the subsequence in which B occurs.

$$P(A/B) = P(A).$$

If the events A and B are independent, then

* the events $\bar{A}$ and B are independent.
* the events $\bar{A}$ and $\bar{B}$ are independent.

Proof:

We know

- AB and $\bar{A}B$ are mutually exclusive
- $B = AB + \bar{A}B$
- $P(\bar{A}) = 1 - P(A)$

Therefore

$$P(\bar{A}B) = P(B) - P(AB)$$
$$= P(B) - P(A)P(B)$$
$$= [1 - P(A)]P(B)$$
$$= P(\bar{A})P(B).$$

$\Rightarrow \bar{A}$ and B are independent.

$\Rightarrow \bar{A}$ and $\bar{B}$ are independent, A and $\bar{B}$ are independent.

Independence of three events:

The events A_1, A_2, A_3 are called (mutually) independent events if they are

(i) independent in pairs

$$P(A_i A_j) = P(A_i)P(A_j) \text{ for } i \neq j$$

and (ii) $P(A_1 A_2 A_3) = P(A_1)P(A_2)P(A_3)$

It follows from independence that:

intersection of the other two.

(2) If we replace one or more of these events with their complements, the resulting events are also independent.

(3) Any one of them is independent of the union of the other two.

Independence of n events:

Suppose that we have defined independence of k events for every $k < n$. Then, we say that the events $A_1, \ldots, A_n$ are independent if ~~for~~ any $k < n$ of them are independent and

$$P(A_1 A_2 \ldots A_n) = P(A_1)\, P(A_2) \ldots P(A_n)$$

Combined experiments:

We are given two experiments: The first experiment is the rolling of a fair die

$$S_1 = \{f_1, \ldots, f_6\} \qquad P_1(f_i) = \frac{1}{6}\,.$$

The second experiment is the tossing of a fair coin

$$S_2 = \{h, t\} \qquad P_2(h) = P_2(t) = \frac{1}{2}\,.$$

Now, what is the probability of "two" on the die and "heads" on the coin?

If we assume that the two events are independent, the result is $\frac{1}{6} \times \frac{1}{2}$. However, this independence is not {"two" on the die} and {"heads" on the coin} are not subsets of the same space. We must first construct a space S which has both these events as subsets.

Cartesian products:

Given two sets S_1 and S_2 with elements s_1 and s_2 respectively, we form all ordered pairs $s_1 s_2$ where s_1 is any element of S_1 and s_2 is any element of S_2.

$$S = S_1 \times S_2.$$

If A is a subset of S_1 and B is a subset of S_2, then the set $C = A \times B$ is a subset of S.

Also, $A \times B = (A \times S_2) \cap (S_1 \times B)$

The cartesian product of two experiments S_1 and S_2 is a new experiment $S = S_1 \times S_2$ whose events are cartesian products of the form $A \times B$ where A is an event of S_1 and B is an event of S_2, and their unions and intersections. In this experiment, $P(A \times S_2) = P_1(A)$

and $P(S_1 \times B) = P_2(B)$.

The probabilities of events of the form $A \times B$ and of their unions and intersections cannot, in general, be expressed in terms of P_1 and P_2. To determine them, we need additional information about the experiments S_1 and S_2.

Independent experiments:

In many applications, the events $A \times S_2$ and $S_1 \times B$ of the combined experiment S are independent for any A and B. Since $A \times B = (A \times S_2) \cap (S_1 \times B)$, we get

$$P(A \times B) = P(A \times S_2)\, P(S_1 \times B)$$
$$= P_1(A)\, P_2(B).$$

Back to the example:

Resulting space of combined experiment is $\{f_1 h, f_2 h, \ldots, f_6 h, f_1 t, f_2 t, \ldots, f_6 t\}$.

event {"two" on the die} is $\{f_2 h, f_2 t\}$

event {"heads" on the coin} is $\{f_1 h, f_2 h, \ldots, f_6 h\}$

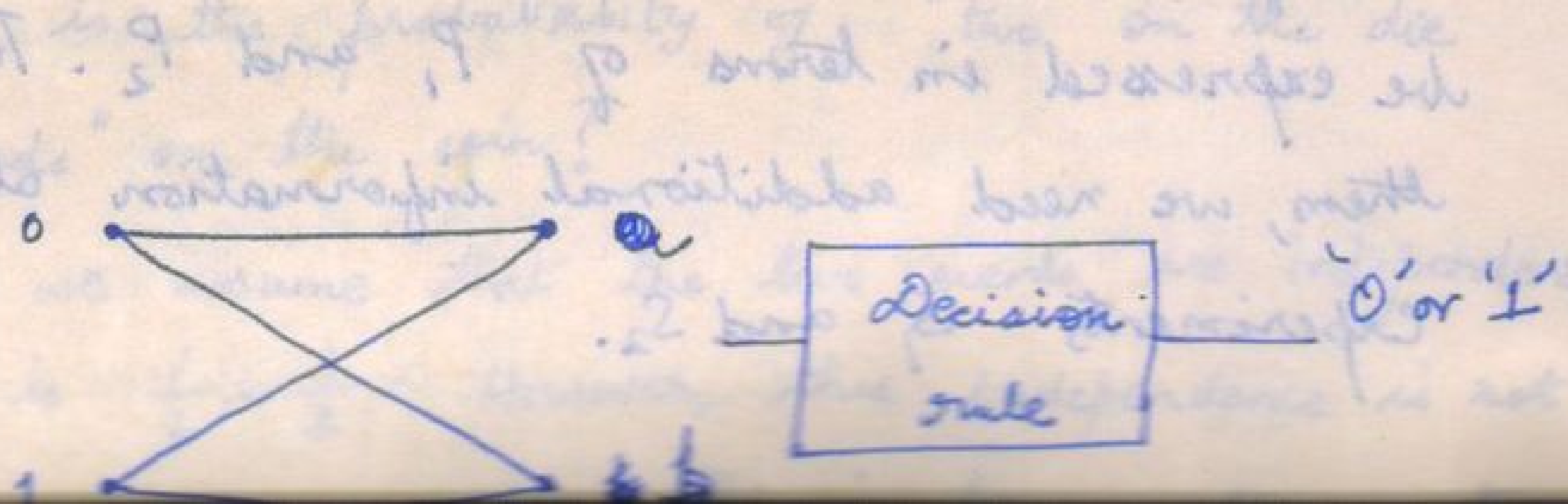
P {"two" on the die} $= \frac{1}{6}$

P {"heads" on the coin} $= \frac{1}{2}$.

The intersection of the two events is $\{f_2 h\}$. Assuming that the events are independent, we get the required probability as $\frac{1}{6} \times \frac{1}{2}$.

Lecture 5:

Example:



(18) (19)

Sample space $S_1 \{0,1\}$ at tx   Sample space $S_2 \{a,b\}$ at rx input.

Combined experiment

$$S = S_1 \times S_2 = \{0a, 0b, 1a, 1b\}$$

Suppose that the probabilities of each of these outcomes is given. At the receiver, once we receive 'a' or 'b' we need to decide whether '0' or '1' was transmitted. What should the decision rule be?

Decision criterion: Minimize the probability that the decoded bit $\neq$ transmitted bit. (or) Maximize the probability that the "decoded bit = transmitted bit." (correct decision)

$$S = \{0a, 0b, 1a, 1b\} = \underbrace{\{0a, 1a\}}_{\text{"a received"}} \cup \underbrace{\{0b, 1b\}}_{\text{"b received"}}$$

Partition

"Error" event or "correct decision" event.

Total probability Thm.

$$P(\text{"correct decision"}) = P(\text{"correct decision"} / \text{"a received"})\, P(\text{"a received"}) + P(\text{"correct decision"} / \text{"b received"})\, P(\text{"b received"})$$

The choice of decision rule cannot change P("a received") or P("b received"). If we maximize P("correct decision"/"a received") & P("correct decision"/"b received"), it is sufficient to maximize P("correct decision").

Given that "a received"

Two decision options: "1 transmitted" —(1)

or

"0 transmitted". —(2)

Suppose we choose option (1)

$P(\text{"correct decision"} / \text{"a received"})$

$$= P(\text{"1 transmitted"} / \text{"a received"}) = \frac{P(\text{"1a"})}{P(\text{"a received"})} \Bigg\} \text{ Definition of conditional probability.}$$

Suppose we choose option (2)

$P(\text{"correct decision"} / \text{"a received"})$

$$= P(\text{"0 transmitted"} / \text{"a received"})$$

$$= \frac{P(\text{"0a"})}{P(\text{"a received"})}$$

Depending on whether $P(1a) > P(0a)$ or $P(0a) > P(1a)$ we need to make the decisions. (choose the decision rule).

Random Variables:

We are given an experiment specified by

(1) space S,

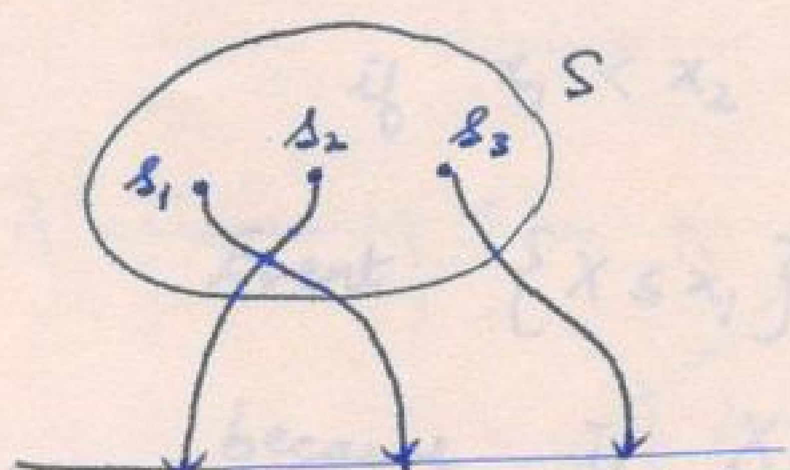
(2) field of subsets of S called events, and

(3) probabilities assigned to these events.

A random variable is a function from the domain S to range R (real line). Each experimental outcome in S is mapped onto R by the random variable.

Remark:

Strictly speaking, $X: S \to R$ is neither random nor a variable; it is simply a mapping. The strange name is simply a reminder of the randomness associated with the elements of the sample space S.



Events generated by random variables:

Some questions of interest: What is the probability that the RV X is less than a given number x? What is the probability that X is between the numbers x_1 and x_2?

Meaning of $\{X \le x\}$: A subset of S consisting of all outcomes s such that $X(s) \le x$. (a set of experimental outcomes not numbers)

If A is a set of numbers on the x axis, then $\{x \in A\}$ represents the subset of S consisting of all outcomes s such that $X(s) \in A$.

Formal definition:

A random variable X is a mapping to a number $X(s)$ of every outcome s. The resulting function must satisfy the following two conditions:

1. The set $\{X \leq x\}$ is an event for every x.
2. The probabilities of the events $\{X = \infty\}$ and $\{X = -\infty\}$ equal 0.

The second condition states that, although we allow X to be $+\infty$ or $-\infty$ for some outcomes, we demand that these outcomes form a set with zero probability.

Summary:

* RV : mapping of each experimental outcome to a numerical value
* If we choose a mapping such that $\{X \leq x\}$ is an event for all x, specifiying $F(\cdot)$ of x specifies $P[\ ]$ of all events.
* Any measurable subset of $\mathbb{R}$ can be represented using sets of the form $(-\infty, x]$.
* Any experimental outcome converted to a numerical outcome. In some cases, experimental outcomes are therefore numerical (noise voltage, etc).

Lecture 6 :

The Cumulative Distribution Function (CDF) of a random variable X is the function

$$F_X(x) = P\{X \leq x\} \text{ defined for every } x \text{ from } -\infty \text{ to } +\infty$$

(could also be $F_X(\alpha) = P\{X \leq \alpha\}$).

Properties of Distribution functions:

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

2) $F_X(+\infty) = 1$ and $F_X(-\infty) = 0$.

i.e. $\lim_{x \to +\infty} F_X(x) = 1$, $\lim_{x \to -\infty} F_X(x) = 0$

$\downarrow$ $P(S)$ $\qquad$ $\downarrow$ $P(\phi)$

3) $F_X(x)$ is a non-decreasing function of x.

if $x_1 < x_2$ then $F_X(x_1) \leq F_X(x_2)$.

Event $\{X \leq x_1\}$ is a subset of event $\{X \leq x_2\}$

because, if $X(s) \leq x_1$ for some s, $X(s) \leq x_2$.

$\Rightarrow P\{X \leq x_1\} \leq P\{X \leq x_2\}$.

4) If $F_X(x_0) = 0$, then $F_X(x) = 0$ for every $x \leq x_0$.

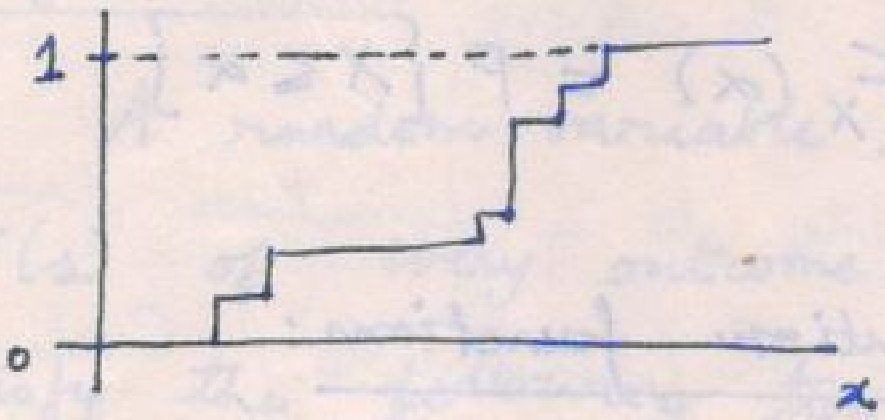
5) $P(X > x) = 1 - F_X(x)$

6) $P(x_1 < X \leq x_2) = F_X(x_2) - F_X(x_1)$

Continuous, Discrete and Mixed RVs:

Based on the CDF, random variables can be classified as continuous, discrete and mixed

A random variable X is <u>discrete</u> if and only if X maps S to a countable subset of $\mathbb{R}$. In this case the CDF looks like



and $$F_X(x) = \sum_{k \le x} P[X = k]$$

where k can take discrete values only. $P[X=k]$ is called the <u>probability mass function (pmf)</u> and is shorthand for $P[\{s \in S : X(s) = k\}]$.

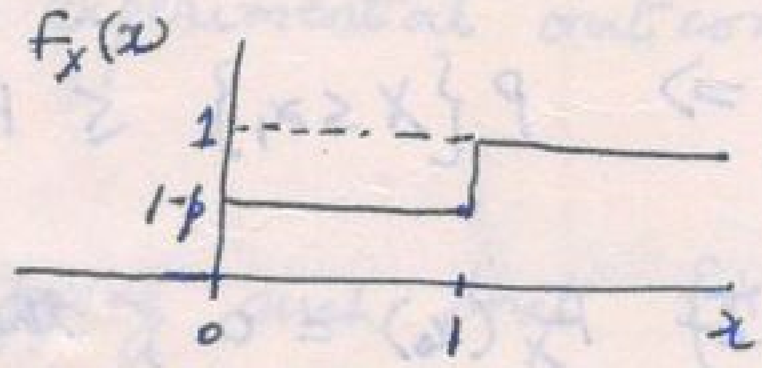
<u>Examples:</u>

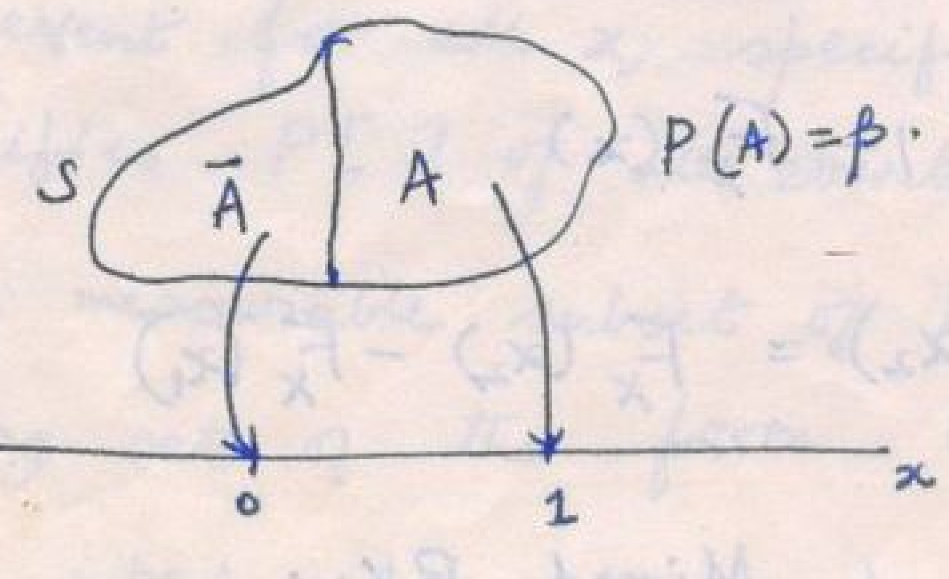
① Bernoulli Random Variable X with parameter p

X takes on values 0 and 1.

$P[X=0] = 1-p$

$P[X=1] = p$





② Poisson Random variable X with parameter λ.

$X \in \{0, 1, 2, \ldots\}$

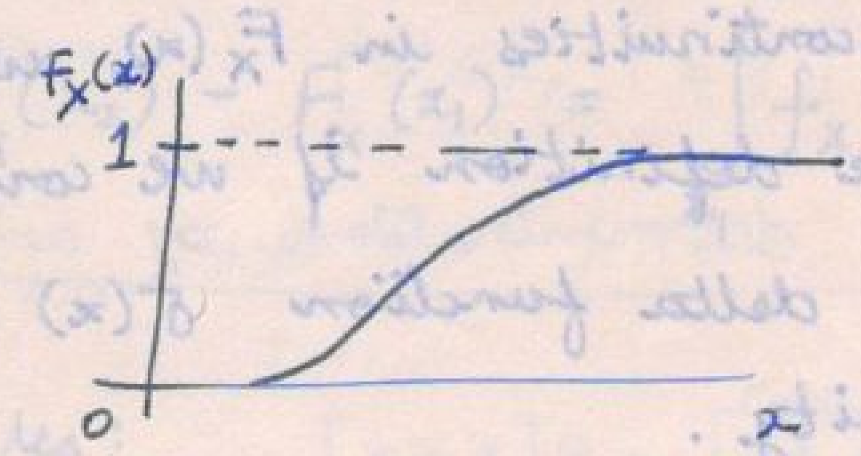
$$P[X=k] = \frac{\lambda^k}{k!} e^{-\lambda} \quad \text{for } k = 0, 1, 2, \dots$$

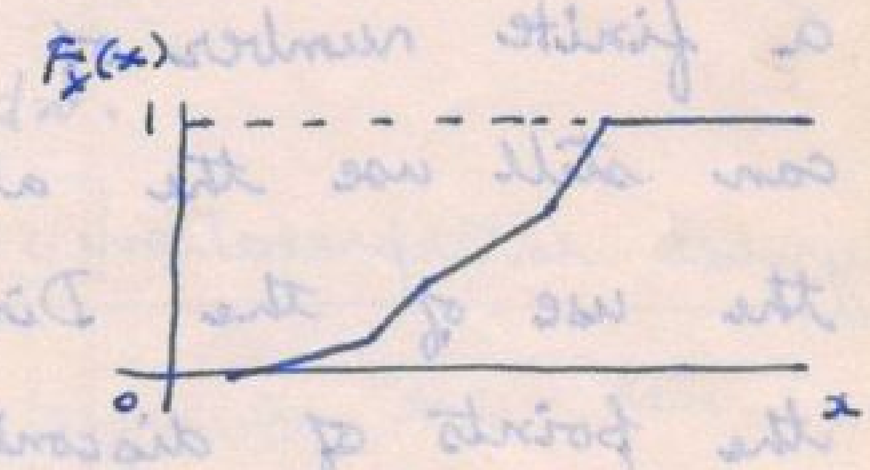
where $\lambda > 0$ (is a positive real number).

(Widely used model for arrivals in queueing systems)

A random variable X is continuous if and only if its CDF $F_X(x)$ is (absolutely) continuous for all $x \in R$.

Eg:

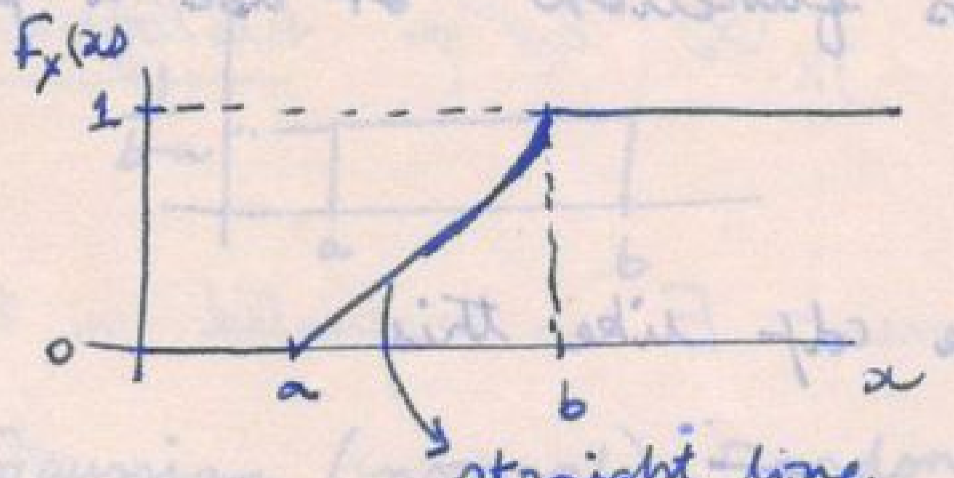




Examples:

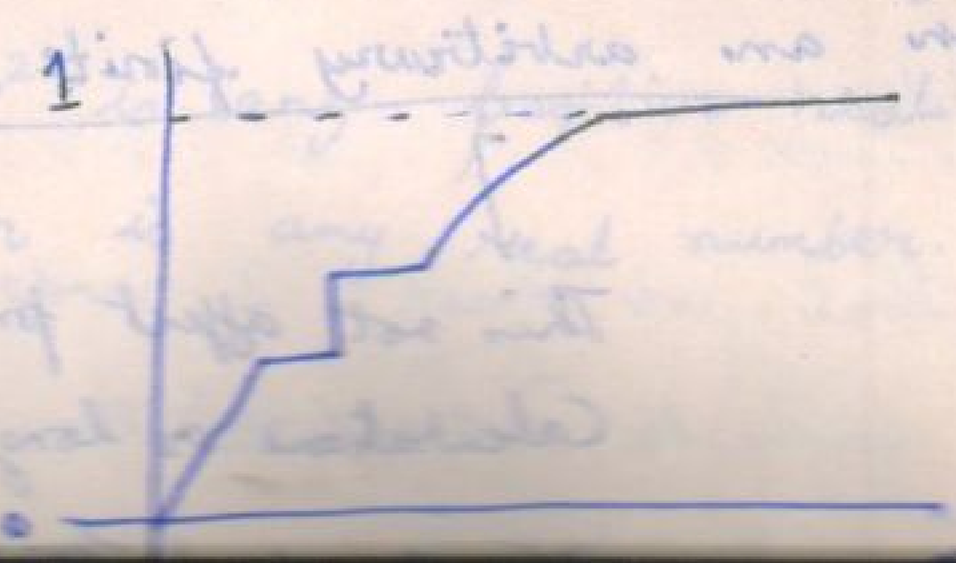
① Uniform random variable.

Uniform in the interval $[a, b]$. $a, b \in R$ & $a < b$.



A random variable X is mixed if and only if its CDF can be written as the CDF of a ~~sum~~ linear combination of a continuous and a discrete random variable.

Eg:



Probability density function: (PDF):

The probability density function (pdf) of a continuous random variable is defined as the derivative of the cdf $F_X(x)$ and is denoted by $f_X(x)$.

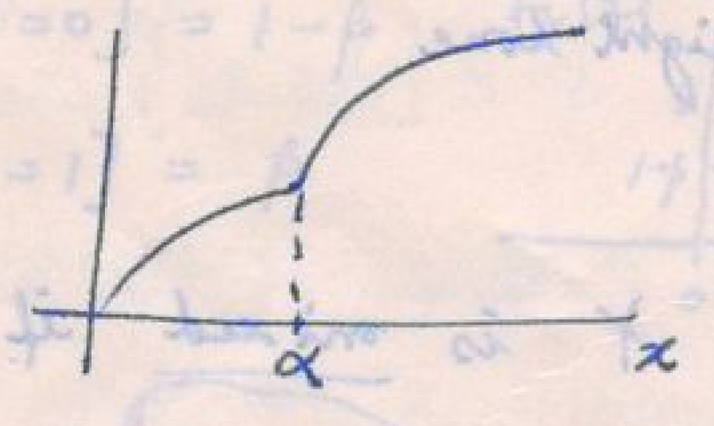
$$f_X(x) \triangleq \frac{d}{dx} F_X(x) \quad \forall\ x \in \mathbb{R}.$$

When the random variable X is mixed with a finite number of discontinuities in $F_X(x)$, we can still use the above definition if we consider the use of the Dirac delta function $\delta(x)$ at the points of discontinuity.

$$\int_{-\infty}^{\infty} \delta(x)\,dx = 1 \quad , \quad \delta(x) = 0 \text{ for all } x \neq 0.$$

For discrete random variables, we can just use the probability mass function or use impulses (delta functions).

If we have the cdf like this



at $x = \alpha$, the derivative (left and right are unequal) is not well defined. In such cases, we can assign an arbitrary finite value for $f_X(\alpha)$.

This not affect probability Calculations as long as the