Probability and Statistics

Arohan Hazarika(22B3948)
Indian Institute of Technology Bombay

July 2023

Chapter 1

Introduction to Probability

1.1 Sample Space

The modern day probability theory is based on axioms, random events are represented by sets and probability is just a normed measure defined on these sets.

1.1.1 Definition 1

A random(or statistical) experiment is an experiment in which

- all outcomes of the experiment are known in advance
- any performance of the experiment in an outcome that is not known in advance, and
- the experiment can be repeated under identical conditions

1.1.2 Definition 2

The sample space of a statistical experiment is a pair (Ω, Ψ)

- Ω is the set of all possible outcomes of the experiment and
- Ψ is a σ-field of subsets of Ω.

The elements of Ω are called sample points. Any set $A \in \Psi$ is known as an event. We say that an event A happens if the outcome of the experiment corresponds to a point in A. Each one-point set is known as a simple or an elementary event.

The choice of Ψ is an important one, and some remarks are in order. If Ω contains at most a countable number of points, we can always take Ψ to be the class of all. If Ω has uncountably many points, the class of all subsets of Ω is still a σ-field, but it is much too large a class of sets to be of interest.

1.2 Probability Axioms

We will be definiting the probability set function and study some of its properties.

1.2.1 Definition 1

Let (Ω,Ψ) be a sample space. A set function P defined on Ψ is called a probability measure if it satisfies the following conditions:

- $P(A) \geq 0$ for all $A \in \Psi$.
- $P(\Omega)=1$.
- Let A_j, $A_j \in \Psi$, j=1,2,3..., be a disjoint sequence of sets, that is, $A_j \cap A_k = \Phi$ for $j \neq k$ where Φ is the null set. Then

  $$P(\sum_{j=1}^{\infty} A_j) = \sum_{j=1}^{\infty} P(A_j)$$

  This is also known as countable additivity.

1.2.2 Definition 2

The triple(Ω,Ψ,P) is called a probability space.

1.2.3 Theorems

There are certain theorems which are as follows:

Theorem 1

P is monotone and subtractive; that is, if $A,B \in \Psi$ and $A \subseteq B$, then $P(A) \leq P(B)$ and $P(B-A) = P(B)-P(A)$, where $B-A = B \cap A^c$, A^c being the complement of the event A. (Proof is being omitted but intuitively its very simple)

Theorem 2

If A, $B \in \Psi$, then

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

(Proof is being omitted but can be understood very easily with venn diagram)

Theorem 3(The Principle of Inclusion and Exclusion)

Let $A_1,A_2,...,A_n \in \Psi$. Then

$$P(\cup_{k=1}^{n} A_k) = \sum_{k=1}^{n} P(A_k) - \sum_{k1<k2}^{n} P(A_{k1} \cap A_{k2}) + \sum_{k1<k2<k3}^{n} P(A_{k1} \cap A_{k2} \cap A_{k3}) + ... + (-1)^{n+1} P(\cap_{k=1}^{n} A_k)$$

Theorem 4(Bonferroni's Inequality)

Given n($\geq$2) events A_1,A_2,...,A_n,

$$\sum_{k=1}^{n} P(A_k) - \sum_{k1<k2}^{n} P(A_{k1} \cap A_{k2}) \leq P(\cup_{k=1}^{n} A_k) \leq \sum_{k=1}^{n} P(A_k)$$

(Proof is done by using Mathematical Induction and intuitively it can be understood by using Theorem 3)

Theorem 5(Boole's Inequality)

For any two events, A and B,

$$P(A \cap B) \geq 1 - P(A^c) - P(B^c)$$

(Proof is simple and can be easily given by using De Morgan Laws)

Theorem 6

Let $\{A_n\}$ be a nondecreasing sequence of events in Ψ, that is, $A_n \in \Psi$, n = 1,2,..., and

$$A_n \supseteq A_{n-1}, n = 2, 3...$$

Then

$$\lim_{n\to\infty} P(A_n) = P(\lim_{n\to\infty} A_n) = P(\cup_{k=1}^{\infty} A_k)$$

1.2.4 A Special Problem - Bertrand's Paradox

Problem Statement

A chord is drawn at random in the unit circle. What is the probability that the chord is longer than the side of the equilateral triangle inscribed in the circle?

Actually there are three solutions to this problem, depending on how we interpret the phrase “at random” which give different answers. In Classical Probability, it is called a paradox because of different values of probability which shows lack of rigour as the ”random” event is rigorously not defined. The paradox is resolved once we define the probability spaces carefully which is done in Axiomatic Probability.

Chapter 2

Conditional Probability and Bayes Theorem

2.1 Definition

Let (Ω,Ψ,P) be a probability space, and let H $\in \Psi$ with P(H) $>$ 0. For an arbitrary A $\in \Psi$ we shall write

$$P(A|H) = P(A \cap H)/P(H)$$

and call the quantity so defined the conditional probability of A, given H. Conditional probability remains undefined when P(H) = 0.

2.1.1 Theorem 1

Let (Ω,Ψ,P) be a probability space, and let H$\in$S with P(H) $>$ 0. Then (Ω,Ψ,$\mathrm{P_H}$), where $\mathrm{P_H}$(A)=P(A|H) for all A $\in \Psi$, is a probability space.

2.1.2 Theorem 2(The Multiplication Rule)

Let(Ω,Ψ,P) be a probability space and A1,A2,...,An$\in$S, then

$$P(\cap_{k=1}^{n} A_k) = P(A_1)P(A_2|A_1)P(A_3|A_2 \cap A_1)...P(A_n| \cap_{j=1}^{n-1} A_j)$$

provided that

$$P(A_n| \cap_{j=1}^{n-1} A_j) > 0$$

2.1.3 Theorem 3(Bayes Theorem)

Let $\mathrm{H_n}$ be a disjoint sequence of events such that P($\mathrm{H_n}$) $>$ 0, n = 1,2,... and $\sum_{n=1}^{\infty} H_n = \Omega$ Let B $\in \Psi$ with P(B) $>$ 0. Then

$$P(H_j|B) = \frac{P(H_j)P(B|H_j)}{\sum_{i=1}^{\infty} P(H_i)P(B|H_i)}, j = 1, 2...$$

Chapter 3

Random Variables and Their Probability Distributions

3.1 Random Variable

3.1.1 Definition

Let (Ω,Ψ) be a sample space. A finite, single-valued function which maps Ω into R is called a random variable (RV) if the inverse images under X of all Borel sets in R are events, that is, if

$$X^{-1}(B) = (\omega : X(\omega) \in B) \in \Psi \ \forall \ B \in \mathcal{B}$$

3.1.2 Probability Distribution of a Random Variable

Theorem 1

The RV X defined on the probability space (Ω,Ψ,P) induces a probability space ($\Re$,$\mathcal{B}$,Q) by means of the correspondence

$$Q(B) = P(X^{-1}(B)) = P(\omega : X(\omega) \in B) \forall \ B \in \mathcal{B}$$

We call Q the probability distribution of X.

Definition 1

A real-valued function F defined on (-∞,∞) that is nondecreasing, right continuous, and satisfies

$$F(-\infty) = 0 \ and \ F(+\infty) = 1$$

is called a distribution function.

Theorem 2

The set of discontinuity points of a DF F is at most countable.

Definition 2

Let X be an RV defined on (Ω,Ψ,P). Define a point function F(.) on R by using Theorem 1, namely

$$F(x) = Q(-\infty, x] = P(\omega : X(\omega) \leq x) \ \forall \ x \in \Re$$

The function F is called the distribution function of RV X.

3.2 Continuous vs Discrete Random Variables

3.2.1 Discrete Random Variables

An RV X defined on (Ω,Ψ,P) is said to be of the discrete type, or simply discrete, if there exists a countable set E $\subseteq$ R such that P(X $\in$ E) = 1. The points of E which have positive mass are called jump points or points of increase of the DF of X, and their probabilities are called jumps of the DF.

The collection of numbers $\{p_i\}$ satisfying $\mathrm{P(X = x_i) = p_i} \geq 0$, for all i and $\sum_{i=1}^{\infty} p_i = 1$, is called the probability mass function (pmf) of RV X.

3.2.2 Continuous Random Variables

Let X be an RV defined on (Ω,Ψ,P) with DF F. Then X is said to be of the continuous type (or, simply, continuous) if F is absolutely continuous, that is, if there exists a nonnegative function f(x) such that for every real number x we have

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

The function f is called the probability density function (PDF) of the RV X.

3.3 Functions of a Random Variable

Theorem 1

Let X be an RV defined on (Ω,Ψ,P). Also, let g be a Borel-measurable function on R. Then g(X) is also an RV.

Theorem 2

Given an RV X with a known DF, the distribution of the RV Y = g(X), where g is a Borel-measurable function, is determined.

3.4 Some Standard Useful Random Variable Distributions

3.4.1 Bernoulli Distribution

The Bernoulli distribution is a discrete distribution having two possible outcomes labelled by n=0 and n=1 in which n=1 ("success") occurs with probability p and n=0 ("failure") occurs with probability q=1-p, where 0<p<1. It therefore has probability density function

$$P(n) = \begin{cases} 1-p & \text{n} = 0 \\ p & \text{n} = 1 \end{cases}$$

3.4.2 Binomial Distribution

Suppose a random experiment has the following characteristics.

- There are n identical and independent trials of a common procedure.
- There are exactly two possible outcomes for each trial, one termed "success" and the other "failure."
- The probability of success on any one trial is the same number p

Then the discrete random variable X that counts the number of successes in the n trials is the binomial random variable with parameters n and p . We also say that X has a binomial distribution with parameters n and p. If X is a binomial random variable with parameters n and p , then

$$P(x) = {}^{n}C_{x} = \frac{n!}{x!(n-x)!} p^{x}(1-p)^{n-x}$$

3.4.3 Poisson Distribution

A Poisson distribution is a discrete probability distribution, meaning that it gives the probability of a discrete (i.e., countable) outcome. For Poisson distributions, the discrete outcome is the number of times an event occurs, represented by k. The Poisson distribution has only one parameter, λ (lambda), which is the mean number of events. The probability mass function of the Poisson distribution is:

$$P(X = k) = \frac{e^{-\lambda}\lambda^{k}}{k!}$$

3.4.4 Gaussian Distribution

A Gaussian distribution, also referred to as a normal distribution, is a type of continuous probability distribution that is symmetrical about its mean; most observations cluster around the mean, and the further away an observation is from the mean, the lower its probability of occurring. The probability density function (pdf) of a Gaussian distribution is a Gaussian function that takes the form:

$$f(x) = \frac{e^{-0.5\frac{(x-\mu)^2}{\sigma^2}}}{\sigma\sqrt{2\pi}}$$

where μ and σ are the mean and standard deviation.

Chapter 4

Moments and Generating Functions

4.1 Moment of a Distribution Function

4.1.1 Mean or Mathematical Expectation Value of a Random Variable X

Let X be a random variable of the discrete type with probability mass function $p_k = P\{X = x_k\}, k = 1, 2....$ If

$$\sum_{k=1}^{\infty} |x_k| p_k$$

we say that the expected value or the mean of X exists and write

$$\mu = EX = \sum_{k=1}^{\infty} x_k p_k$$

If X is of the continuous type and has PDF f , we say that EX exists and equals $\int xf(x)dx$, provided that

$$\int |x| f(x) dx < \infty$$

4.1.2 Theorem 1

Let X be an RV, and g be a Borel-measurable function on $\Re$. Let $Y = g(X)$. If X is discrete type then

$$EY = \sum_{j=1}^{\infty} g(x_j) P\{X = x_j\}$$

If X is of the continuous type with PDF f, let $h(y)$ be the PDF of $Y = g(X)$. Then, according to Theorem 1,

$$EY = \int g(x)f(x)dx = \int yh(y)dy$$

provided that $E|g(X)| < \infty$.

4.1.3 Theorem 2

If the moment of order t exists for an RV X, moments of order $0 < s < t$ exist.

4.1.4 Theorem 3

Let X be an RV on a probability space (Ω,Ψ,P). Let $E|X|^k < \infty$ for some $k > 0$. Then

$$n^k P\{|X| > n\} \to 0 \; as \; n \to \infty$$

The converse of the theorem does not hold in general.

Lemma 1

Let X be a nonnegative RV with distribution function F. Then

$$EX = \int_0^\infty [1 - F(x)]dx,$$

in the sense that, if either side exists, so does the other and the two are equal.

4.1.5 Theorem 4

Let X be an RV with a distribution satisfying $n^\alpha P\{|X| > n\} \to 0 \; as \; n \to \infty$ for some $\alpha > 0$. Then $E|X|^\beta < \infty$ for $0 < \beta < \alpha$.

4.1.6 Theorem 5

Let X be an RV satisfying

$$\frac{P\{|X| > cx\}}{P\{|X| > x\}} \to 0 \;\; as \; x \to \infty \;\; for \; all \; c > 1;$$

then X possesses moments of all orders.

4.1.7 Theorem 6

If $h_1, h_2, ..., h_n$ are Borel-measurable functions of an RV X and $Eh_i(X)$ exists for $i = 1, 2, ..., n$, then $E\{\sum_{i=1}^n h_i(X)\}$ exists and equals $\sum_{i=1}^n Eh_i(X)$.

4.1.8 Variance of a Random Varibale X

If EX^2 exists, we call $E\{X-\mu\}^2$ the variance of X, and we write $\sigma^2 = var(X) = E\{X-\mu\}^2$. The quantity σ is called the *standard deviation* (SD) of X.

4.1.9 Theorem 7

Var(X)=0 if and only if X is degenerate.

4.1.10 Theorem 8

$Var(X) < E(X-c)^2$ for any $c \neq EX$.

4.2 Generating Functions

4.2.1 Definition 1

The function defined by

$$P(s) = \sum_{k=0}^{\infty} p_k s^k$$

which surely converges for $|s| \leq 1$, is called the probability generating function (PGF) of X.

4.2.2 Definition 2

Let X be an RV defined on (Ω,Ψ,P). The function

$$M(s) = Ee^{sX}$$

is known as the moment generating function (MGF) of the RV X if the expectation on the right side exists in some neighborhood of the origin.

4.2.3 Theorem 1

The MGF uniquely determines a DF and, conversely, if the MGF exists, it is unique.

4.2.4 Theorem 2

If the MGF $M(s)$ of an RV X exists for s in $(-s_0, s_0)$ say, $s_0 > 0$, the derivatives of all order exist at $s = 0$ and can be evaluated under the integral sign, that is,

$$M^{(k)}(s)|_{s=0} = EX^k \textit{ for positive integral } k.$$

4.2.5 Definition 3

Let X be an RV. The complex-valued function Φ defined on $\Re$ by

$$\Phi(t) = E(e^{itX}) = E(\cos tX) + iE(\sin tX), \quad t \in \Re$$

where $i = \sqrt{-1}$ is the imaginary unit, is called the characteristic function (CF) of RV X.

4.3 Some Moment Inequalities

4.3.1 Theorem 1

Let $h(X)$ be a nonnegative Borel-measurable function of an RV X. If $Eh(X)$ exists, then, for every $\varepsilon > 0$,

$$P\{h(X) \geq \varepsilon\} \leq \frac{Eh(X)}{\varepsilon}$$

Corollary

Let $h(X) = |X|^r$ and $\varepsilon = K^r$, where $r > 0$ and $K > 0$. Then

$$P\{|X| \geq K\} \leq \frac{E|X|^r}{K^r}$$

which is $Markov's$ *inequality*. In particular, if we take $h(X) = (X-\mu)^2, \varepsilon = K^2\sigma^2$, we get *Chebychev–Bienayme inequality*:

$$P\{|X-\mu| \geq K\sigma\} \leq \frac{1}{K^2}$$

where $EX = \mu, var(X) = \sigma^2$

4.3.2 Theorem 2(Lyapunov Inequality)

Let $\beta_n = E|X|^n < \infty$. Then for arbitrary $k, 2 \leq k \leq n$, we have

$$\beta_{k-1}^{\frac{1}{k-1}} \leq \beta_k^{\frac{1}{k}}$$

Chapter 5

Multiple Random Variables

5.1 Definition of Multi Dimensional Random Variable

5.1.1 Definition 1

The collection $\mathbf{X} = (X_1, X_2, ..., X_n)$ defined on (Ω,Ψ,P) into $\Re_n$ by

$$\mathbf{X}(\omega) = (X_1(\omega), X_2(\omega), ..., X_n(\omega)), \quad \omega \in \Omega$$

is called an n-dimensional RV if the image inverse of every n-dimensional interval

$$I = \{(x_1, x_2, ..., x_n : -\infty < x_i \leq a_i, ai \in \Re, i = 1, 2, ..., n\}$$

is also in Ψ, that is, if

$$\mathbf{X}^{-1}(I) = \{\omega : X_1(\omega) \leq a_1, ..., X_n(\omega) \leq a_n\} \in \Psi \ \ for \ a_i \in \Re$$

5.1.2 Theorem 1

Let $X_1, X_2, ..., X_n$ be n RVs on (Ω,Ψ,P). Then $\mathbf{X} = (X_1, X_2, ..., X_n)$ is an n-dimensional RV on (Ω,Ψ,P).

5.2 Distribution Function and Calculation of Probability Density of Multiple Random Variables

5.2.1 Definition 1

The function $F(,)$, defined by

$$F(x, y) = P\{X \leq x, Y \leq y\}, \quad all \ (x, y) \in \Re$$

is known as the DF of the RV (X, Y).

5.2.2 Theorem 1

A function F of two variables is a DF of some two-dimensional RV if and only if it satisfies the following conditions:

- F is nondecreasing and right continuous with respect to both arguments;
- $F(-\infty, y) = F(x, -\infty) = 0$ and $F(+\infty, +\infty) = 1$; and
- for every $(x_1, y_1), (x_2, y_2)$ with $x_1 < x_2$ and $y_1 < y_2$ the inequality

$$F(x_2, y_2) - F(x_2, y_1) + F(x_1, y_1) - F(x_1, y_2) \geq 0$$

holds.
The above theorem can be extended for n RVs also.

5.2.3 Definition 2

A two-dimensional (or bivariate) RV (X, Y) is said to be of the *discrete* type if it takes on pairs of values belonging to a countable set of pairs A with probability 1. We call every pair (x_i, y_j) that is assumed with positive probability p_{ij} a *jump point* of the DF of (X, Y) and call p_{ij} the jump at (x_i, y_j). Here A is the support of the distribution of (X, Y).

Clearly $\sum_{ij} p_{ij}$. As for the DF of (X, Y), we have

$$F(x, y) = \sum_{B} p_{ij},$$

where $B = \{(i, j) : x_i \leq x, y_j \leq y.\}$

5.2.4 Definition 3

Let (X, Y) be an RV of the discrete type that takes on pairs of values (x_i, y_j),i = 1,2,..., and j = 1,2,.... We call

$$p_{ij} = P\{X = x_i, Y = y_j\}, \quad i = 1, 2, ..., j = 1, 2, ...,$$

the *joint probability mass function* (PMF) of (X, Y).

5.2.5 Definition 4

A two-dimensional RV (X, Y) is said to be of the *continuous type* if there exists a nonnegative function $f(,)$ such that for every pair $(x, y) \in \Re^2$ we have

$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v) dv du$$

where F is the DF of (X, Y). The function f is called the (joint) PDF of (X, Y).

5.3 Marginal and Conditional Distributions

5.3.1 Definition 1

The collection of numbers $\{p_{i.} = \sum_{j=1}^{\infty} p_{ij}\}$ is called the marginal PMF of X, and the collection $\{p_{.j} = \sum_{j=1}^{\infty} p_{ij}\}$, the marginal PMF of Y.

Same goes in case of continuous RVs where just PMF is replaced by PDF.

5.3.2 Definition 2

Let (X, Y) be an RV with DF F. Then the *marginal* DF of X is defined by

$$F_1(x) = F(x, \infty) = \lim_{y \to \infty} F(x, y)$$

A similar description is given for the marginal DF of Y.

5.3.3 Definition 3

Let (X, Y) be an RV of the discrete type. If $P\{Y = y_j\} > 0$, the function

$$P\{X = x_i | Y = y_j\} = \frac{P\{X = x_i, Y = y_j\}}{P\{Y = y_j\}}$$

for fixed j is known as the *conditional* PMF of X, given $Y = y_j$. A similar definition is given for $P\{X = x_i | Y = y_j\}$, the conditional PMF of Y, given $X = x_i$, provided that $P\{X = xi\} > 0$.

5.3.4 Definition 4

In case of Continuous Random Variables, $P\{X = x\} = P\{Y = y\} = 0$ for any x, y, hence conditional probability is defined as follows

Let $\epsilon > 0$, and suppose that $P\{y - \epsilon < Y \leq y + \epsilon\} > 0$. For every x and every interval $(y - \epsilon, y + \epsilon]$, consider the conditional probability of the event $\{X \leq x\}$, given that $Y \in (y - \epsilon, y + \epsilon]$. We have

$$P(X \leq x \mid y - \varepsilon < Y \leq y + \varepsilon) = \frac{P(X \leq x, y - \varepsilon < Y \leq y + \varepsilon)}{P(Y \in (y - \varepsilon, y + \varepsilon])}$$

5.3.5 Definition 5

The conditional DF of an RV X, given $Y = y$, is defined as the limit

$$\lim_{\varepsilon \to 0^+} P(X \leq x \mid Y \in (y - \varepsilon, y + \varepsilon])$$

provided that the limit exists. If the limit exists, we denote it by $F_{X|Y}(x|y)$ and define the conditional density function of X, given $Y = y$, as a nonnegative function satisfying

$$F_{X|Y}(x|y) = \int_{-\infty}^{x} f_{X|Y}(t|y)\,dt \quad for\ all\ x \in R$$

5.4 Independent Random Variables and Independent and Identically Distributed Random Variables

5.4.1 Definition 1

We say that X and Y are independent if and only if

$$F(x, y) = F_1(x)F_2(y) \quad for\ all\ (x, y) \in R^2$$

where $F(x, y)$ and $F_1(x), F_2(y)$, respectively, be the joint DF of (X, Y) and the marginal DFs of X and Y.

5.4.2 Lemma

If X and Y are independent and $a < c, b < d$ are real numbers, then

$$P(a < X \leq c, b < Y \leq d) = P(a < X \leq c)P(b < Y \leq d)$$

5.4.3 Theorem 1

Let X and Y be independent RVs and f and g be Borel-measurable functions. Then $f(X)$ and $g(Y)$ are also independent.

5.4.4 Theorem 2

A collection of jointly distributed RVs $\{X_1, X_2, ..., X_n\}$ is said to be mutually or completely independent if and only if

$$F(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} F_i(x_i), \quad for\ all\ (x_1, x_2, \ldots, x_n) \in R^n$$

where F is the joint DF of $(X_1, X_2, ..., X_n)$, and $F_i (i = 1, 2, ..., n)$ is the marginal DF of X1,...,Xn, which are said to be pairwise independent if and only if every pair of them are independent.

5.4.5 Theorem 3

If $X_1, X_2, ..., X_n$ are independent, every subcollection $X_{i_1}, X_{i_2}, ..., X_{i_k}$ of $X_1, X_2, ..., X_n$ is also independent.

5.4.6 Theorem 4

We say that RVs X and Y are identically distributed if X and Y have the same DF, that is,

$$F_X(x) = F_Y(x) \quad for\ all\ x \in R$$

where F_X and F_Y are the DF's of X and Y, respectively.

5.5 Covariance and Correlation

5.5.1 Definition 1

If $E\{(X - EX)(Y - EY)\}$ exists, we call it the *covariance* between X and Y and write

$$cov(X, Y) = E\{(X - E[X])(Y - E[Y])\} = E(XY) - E[X]E[Y]$$

5.5.2 Definition 2

If EX^2, EY^2 exist, we define the correlation coefficient between X and Y as

$$\rho = \frac{cov(X, Y)}{SD(X)SD(Y)} = \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X^2) - (E(X))^2}\sqrt{E(Y^2) - (E(Y))^2}}$$

where SD(X) denotes the standard deviation of RV X.

5.6 Order Statistics

Let $(X_1, X_2, \ldots, X_n)$ be an n-dimensional random variable and $(x_1, x_2, \ldots, x_n)$ be an n-tuple assumed by $(X_1, X_2, \ldots, X_n)$.Arrange $(x_1, x_2, \ldots, x_n)$ in increasing order of magnitude so that

$$x(1) \leq x(2) \leq \cdots \leq x(n)$$

where $x(1) = \min(x_1, x_2, \ldots, x_n)$, $x(2)$ is the second smallest value in $x_1, x_2, \ldots, x_n$, and so on, $x(n) = \max(x_1, x_2, \ldots, x_n)$. If any two x_i, x_j are equal, their order does not matter.

5.6.1 Definition 1

The function $X(k)$ of $(X_1, X_2, \ldots, X_n)$ that takes on the value $x(k)$ in each possible sequence $(x_1, x_2, \ldots, x_n)$ of values assumed by $(X_1, X_2, \ldots, X_n)$ is known as the kth order statistic or statistic of order k. $\{X(1), X(2), \ldots, X(n)\}$ is called the set of order statistics for $(X_1, X_2, \ldots, X_n)$.

5.6.2 Theorem 1

Let $(X_1, X_2, ..., X_n)$ be an n-dimensional RV. Let $X(k), 1 \leq k \leq n$, be the order statistic of order k. Then $X(k)$ is also an RV.

5.6.3 Theorem 2

The joint PDF of $(X_{(1)}, X_{(2)}, ..., X_{(n)})$ is given by

$$g(x(1), x(2), \ldots, x(n)) = \begin{cases} n! \prod_{i=1}^{n} f(x_{(i)}) & \mathrm{x}_{(1)} < x_{(2)} < ... < x_{(n)} \\ 0 & otherwise \end{cases}$$