CHAPTER I

INTRODUCTION.
1.1. INTRODUCTORY REMARKS:

Any positive number $x$ may be expressed as an $r$-adic representation

$$x = A_1 A_2 \cdots A_{s+1} \cdot \alpha_1 \alpha_2 \alpha_3 \cdots$$

Signifying

$$x = A_1 r^s + \cdots + A_{s+1} + \alpha_1 r^{-1} + \alpha_2 r^{-2} + \cdots$$

where

$$0 \leq A_i < r, \quad 0 \leq A_2 < r, \quad \ldots, \quad 0 \leq \alpha_n < r, \ldots$$

not all $A$ and $\alpha$ are zero, and an infinity of the $\alpha_n$ are less than $r - 1$. If $x \geq 1$, then $A_1 > 0$.

The $r$-adic representation for a rational number $\frac{p}{q}$ is terminating or recurring and conversely. The $r$-adic representation for an irrational number is non-recurring infinite and conversely. If each number of the sequence $0, 1, 2, \ldots, r - 1$ is denoted by a special symbol, the symbols are called the digits.

Now investigations may be made about the frequencies of all possible sequences of any number of digits in the $r$-adic representation of a real. The real numbers may be classified into various types of numbers corresponding to the frequencies of these sequences, in the $r$-adic representations. In this dissertation, several notions associated with these classifications and basic properties of the classes will be discussed by methods of Abstract Mathematics, Probability Theory and Information Theory.
1.2. PRELIMINARY NOTIONS.

Let us consider a random event with \( r \) possible outcomes which are indexed by numbers 0, 1, 2, \ldots, \( r-1 \). Let us consider repeated trials and let us assume that the trials are independent with respect to the event. Then the outcomes of the trials, repeated for an arbitrary number of times will be represented by the digits of a decimal representation of real numbers between 0 and 1 in the scale \( r \). In the case of Bernoulli's trials, evidently the outcomes will be represented by the decimals in binary scale where 1 stands for success and 0 for failure. If the trials are repeated finitely we shall get finite decimals; otherwise we shall get infinite decimals. For convenience, a finite decimal representation of a number will be looked upon as a representation in infinite decimals in which zero is repeated as an infinite number of times. On the assumption of the principle of statistical regularity (Cramer, 1949), we can speak of the probability of occurrence of each of the digits in the decimal representation.

As probability of occurrence of an event can be defined as the limit of the frequency of occurrence of an event, so for a real in its \( r \)-adic representation the probabilities of the occurrence of the digits of the real introduced as the limits
of the frequency of occurrence of digits and in general these limits exist. Therefore, after Shannon, entropy can be defined for almost all reals.

1.2.1. Shannon's definition of entropy and entropy of a real number.

If \( \gamma \) be a discrete random variable, taking on the different values \( \gamma_k \) with the probabilities \( q_k, (k=1, 2, \ldots) \), i.e., \( P(\gamma = \gamma_k) = q_k \) where \( q_k > 0 \), \( \sum_k q_k = 1 \), then the entropy of \( \gamma \) (which may also be called the entropy of the probability-distribution of \( \gamma \)), as defined by Shannon, will be denoted by \( H_0(\gamma) \), i.e., we put

\[
H_0(\gamma) = \sum_k q_k \log \frac{1}{q_k}
\]

provided that the series on the right converges. (If the series is divergent, we say that the entropy of \( \gamma \) does not exist).

The logarithms in \( H_0(\gamma) \) are taken to an arbitrary but fixed base and we always take

\[
q_k \log q_k = 0 \quad \text{if} \quad q_k = 0
\]
In the $r$-adic representation of irrational numbers, the occurrences of particular digits may be looked upon as independent random events; and hence entropy can be defined for real numbers after Shannon.

Let $H_n(p)$ be the entropy of the space of all messages of the first $n$ letters:

$$H_n(p) = -\sum_{i_k \in A} p^i x_1(\omega) = i_1, \ldots, x_n(\omega) = i_n \log p^i x_1(\omega) = i_1, \ldots, x_n(\omega) = i_n$$

$$= \sum_{k=1}^{n} \left( -\sum_{i=0}^{r-1} p_{ki} \log p_{ki} \right)$$

(This $p_{ki}$ is a limit. "It is naturally not necessary that such a limit should exist; this limit may oscillate and one might expect that usually it would. The theorem which follow prove that contrary to our expectation, there is usually a definite frequency. The existence of the limit is in a sense the ordinary event".) - (Hardy and Wright).

If $A$ and $B$ be two independent schemes and $AB$ be the product of the schemes $A$ and $B$ and if $H(A)$, $H(B)$, $H(AB)$ be the corresponding entropies of the schemes $A$, $B$, $AB$; then

$$H(AB) = H(A) + H(B).$$
If the schemes $A$ and $B$ are mutually dependent, we have

$$H(AB) = H(A) + \sum_K p_K H_K(B)$$

where $H_K(B)$ is the conditional entropy of the scheme $B$ calculated on the assumption that the event $A_K$ of the scheme $A$ occurred and $p_K$ is the probability of occurrence of $A_K$ in scheme $A$. We designate

$$\sum_K p_K H_K(B) \text{ by } H_A(B)$$

Also, we have $H_A(B) \leq H(B)$

In the most general case, we have

$$H(AB) = H(A) + H_A(B)$$

1.2.2. Source, Stationarity and Source-entropy.

If $A$ is alphabet, $\mu$ is probability measure defined over Borel-field $F_A$, the stochastic process $[A^I, F_A, \mu]$ is an information-source, where $A^I$ denote class of infinite sequences

$$\chi = (\ldots, \ldots, \chi_{-1}, \chi_0, \chi_1, \ldots)$$
where each $x_t$ belongs to $A$, $t$ belonging to the set of integers.

Since the alphabet $A$ and the probability measure $\mu$ completely characterize the statistical nature of the source, we can denote the source by the symbol $[A, \mu]$.

If $\mu(TS) = \mu(S)$ for any set $S \in \mathcal{F}_A$, then the source is called stationary; $T$ being the coordinate shift-transformation defined as

$$T \mathcal{x} = (\ldots, x_{-1}', x_0', x_1', x_2', \ldots)$$

where $x_K' = x_{K+1}$, $(-\infty < K < +\infty)$.

If $H_n$ be the entropy of an $n$-term sequence, then the quantity

$$H = \lim_{n \to \infty} \frac{H_n}{n}$$

is called source-entropy.

If the above process $\mathcal{x} = \{x_n(u)\}_{n \geq 1}$ is stationary, then there exists the limit

$$\lim_{n \to \infty} \frac{1}{n} H_n(p) = H(p)$$

$H(p)$ is the entropy per symbol of the information source $\mathcal{x}$.

Obviously,

$$0 < H_n(p) < n \log \omega$$

and hence

$$0 < H(p) < \log \omega.$$
1.2.3. **Channel, Noiseless Channel, Stationarity and Connection of the Channel to the Source.**

A channel is characterized by the following three elements

1) the input-alphabet $A$; i.e., the list of signals, which the channel, in question, can transmit,

2) the output-alphabet $B$ i.e., the list of signals (letters), which the channel can omit,

3) the probability $\gamma_x(s)$ that the $\gamma$ received, when a given $x$ is transmitted, belongs to the set $S \in F_B$.

We shall denote the channel specified by these elements by $[A, \gamma_x, B]$.

If every transmitted signal $'a'$ (a letter of the input alphabet $A$) gives at the output a unique letter $b = b(a)$ of the output alphabet $B$, the channel is called a noiseless channel.

We shall call the channel $[A, \gamma_x, B]$ stationary if, for all $x \in A^*$ and $S \in F_B$,

$$\gamma_{T^x} (TS) = \gamma_x (S),$$

where $T$ is co-ordinate shift-transformation as defined before.
Connecting a stationary source \([A, \mu]\) with a stationary channel \([A, \nu_x, B]\), a new stationary source \([C, \omega]\) can be obtained where the alphabet of this source \(C\) is the direct-product \(A \times B\) and the set \(C^i\) of elementary events \((x, y)\) is the direct-product \(A^i \times B^i\) and the probability measure \(\omega(s)\)

\[
\omega(s) = \omega(M \times N) = \int \nu_x(N) \, d\mu(x)
\]

Putting \(M = A^i\),

\[
\nu_x(N) = \omega(A^i \times N) = \int \nu_x(N) \, d\mu(x)
\]

If \(H(x), H(y), H(x, y)\) are respectively the entropies of \([A, \mu]\), \([B, \nu]\), \([C, \omega]\), we define the entropy \(H(x)\) as follows:

Let \(H_n(x)\) be the entropy of the finite space, the elementary events of which are all the \(n\)-term sequences \(x_0, x_1, \ldots, x_{n-1}\) (cylinders) emitted by the source \([A, \mu]\), with corresponding probabilities determined by the distribution \(\mu\); then

\[
H(x) = \lim_{n \to \infty} \frac{1}{n} H_n(x)
\]
In complete analogy,

\[ H(Y) = \lim_{n \to \infty} \frac{1}{n} H_n(Y) \]

\[ H(x, y) = \lim_{n \to \infty} \frac{1}{n} H_n(x, y) \]

where correspondingly \( H_n(Y) \) and \( H_n(x, y) \) denote the entropies of the finite space of sequences \( y_0, y_1, \ldots, y_{n-1} \) from the source \([B, \gamma]\) and \( x_0, x_1, x_{11}, x_{12}, \ldots, x_{n-1}, x_{n-1} \) from the source \([C, \omega]\).

Also,

\[ H_n(x, y) = H_n(x) + H_{n|x}(y) \]

\[ = H_n(y) + H_{n|y}(x) \]

where \( H_{n|x}(y) \) denotes the average conditional entropy of the space of sequences \( y_0, y_1, \ldots, y_{n-1} \) for a given sequence \( x_0, x_1, \ldots, x_{n-1} \) and \( H_{n|y}(x) \) has an analogous meaning.

It follows that

\[ H_{n|x}(y) = H_n(x, y) - H_n(x) \]

\[ H_{n|y}(x) = H_n(x, y) - H_n(y) \]
1.3. **OBJECT AND MOTIVATION OF THE PRESENT INVESTIGATION**

**AND FORMULATION OF PROBLEM.**

The study of $r$-adic representations of real numbers, in particular, of numbers with various types of normality and different degrees of degeneracy from simply-normality have not been considered in details up to now from mathematical, probabilistic and information-theoretic points of views. Application of Information Theory in Theory of Normal Numbers is new. As per advantage of applying information-theoretic technique, in addition to the solution of known problems, this new technique leads to some new problems for investigation and new aspects of the subject. The main aim of the present dissertation is to solve some such problems known previously and also some new problems mentioned above.

The representation of any real number in any scale $r$ is unique where an infinity of the digits are less than $r-1$. Since probability of occurrence of any one of them is not affected by supplementary knowledge concerning the materialization of any number of remaining digits, so occurrence of these digits may be looked upon as independent random events. Thus, entropy can be defined for real numbers.
This introduction of the concept of entropy for real numbers makes some notions like the definitions associated with normality of real very simple and the proofs of some well-known properties easy, straight-forward and neat.

Again, as entropy can be defined in all measurable sets, so entropy can be defined for a real K-tuple. Normality in K-dimension can be related with its entropy, specially with respective entropies of its components. Hence, various results of normal K-tuples can be proved by Information-theoretic technique and some new notions and results can be introduced.

Moreover, it provides some measure of normality. Hence, notions of degeneracy (weak and strong) from normality, almost normal numbers, almost non-normal numbers have been introduced and these yield some new interesting problems in Theory of Numbers.

1.4. HISTORICAL REMARKS.

Notion of Normal Numbers were introduced by E. Borel (1909) [4] who proved the central theorem that almost all real numbers (in the sense of Lebesgue) are normal numbers. According to Borel, x is normal to base r if each of \( x_r, \sqrt{r}x_r, \sqrt{r^2}x_r \ldots \) is simply-normal to all of the
S. S. Pillai (1940) proved that Borel's definition of normal numbers is equivalent with the statement that $x$ is normal to base $r$ if $x$ is simply-normal to bases $v_2, v_3, \ldots$. That Borel's definition of normal numbers is equivalent to the statement that a normal number is a real number expressed in scale $r$, in which every sequence of $K$ digits of scale $r$, occurs with the asymptotic frequency $v_2^{-K}$, was first proved by I. Niven and H.S. Zuckermann (1951) and latter a very-simple proof was given by J.W.S. Cassels (1952). J.E. Maxfield (1952) also devoted his attention to problems connected with Borel's definition of Normal numbers. Calvin T. Long (1957) proved that the real number $\alpha$ is normal to base $r$ if and only if there exists a set of positive integers $m_1 < m_2 < m_3 < \ldots$ such that $\alpha$ is simply-normal to base $v_2^{m_i}$, for each $i \geq 1$ and no finite set of $m$'s will suffice.

Borel exhibited a normal number in a rather complicated way. D.G. Champernowne (1933) first exhibited the simplest explicit example of a normal number. One typical example of normal number written in the scale of 2, constructed by D.G. Champernowne is
According to Champernowne, 

.1 2 3 4 5 6 7 8 9 10 11 12 13 14 . . . .

i.e. the number formed by writing the natural numbers in succession after the decimal point, is normal to base 10. He formulated the conjecture that the number whose decimal is obtained by writing 0 for the integral part and the consecutive prime numbers (instead of consecutive natural numbers) to the right of decimal point i.e. the real number

0 2 3 5 7 11 13 17 . . . . . .

is normal in the scale of 10. The conjecture and a more general theorem have been proved by A.H. Copeland and P. Erdos (1946)

\[ 10 \]

D. G. Champernowne also conjectured that

a) The infinite decimal

.4 6 8 9 10 12 14 . . . .

(i.e. sequence of composites) is normal with respect to base 10.

b) If \( a \) is any positive number and \( a_{\gamma} \) denotes the integral part of \( a_{\gamma} \), then the infinite decimal

\[ a_1 a_2 \ldots \ a_\gamma \ldots \]

is normal in scale of 10.
c) If \( \text{Lyt} = \text{the integral part of } \sqrt[n]{\log n} \), then the infinite decimal
\[ L_1 L_2 \ldots \]
is normal in scale of 10.

D. D. Wall (1949) [51] proved that the number \( x \) is normal to base \( r \) if and only if the numbers
\[ x, nx, n^2 x, \ldots \]
are uniformly-distributed module 1. D. D. Wall also proved that if \( x \) is normal in scale \( r \), then \( \frac{x}{S} \) is also normal in scale \( r \), for every positive integer \( S \). The last result is known as D. D. Wall's second result in the literature.

Niven [37] also stated that if \( x \) is normal to base \( r \), so is \( jx \) for any non-zero rational \( j \).

E. Borel (1909) also defined an absolutely normal number as one that is normal to every base and the existence of such numbers was proved by E. Borel. His proof is based on the measure-theory and being purely existential, it does not provide any method for constructing such a number. The first effective example of an absolutely normal number was given by W. Sierpinski (1916) [48]. Though according to the theorem of Borel, almost all numbers are absolutely normal, it was by no means easy to construct an example of an absolutely normal number. Examples of such numbers are indeed fairly complicated.

W. Schmidt (1960) [46] first proved that there exists numbers which are normal to one base but not absolutely normal. Schmidt
proved that if \( \forall^n = S^n \) (\( n, m \) are integers) then any real normal to base \( r \) is also normal to base \( S \). Schmidt also proved that the set of reals which are normal to one base but not even simply-normal to another base has power of the continuum.

Tibor Salat (1966) \([45]\) proved that the set of all simply-normal numbers and the set of all absolutely normal numbers are of the first Baire category.

The expression "Normality of Order \( K \)" is due to I. J. Good \([25]\) who gave a method for constructing decimals of period \( \forall^K \) having normality of order \( K \) for any \( K \geq 1 \). The problem was also studied by D. Rees (1946) \([42]\), de Brujin (1946) \([13]\) and Korobov (1956) \([36]\) who gave a variety of methods of constructing such decimals. Later on, Calvin T. Long (1966) \([6]\) established another construction for a periodic decimal having normality of order \( K \). In addition to the problem of constructing numbers having normality of order \( K \), Calvin T. Long (1966) \([6]\) introduced the notion of \( C \)-uniform distribution modulo One and proved that a real number \( \alpha \) has normality of order \( K \) if and only if \( \alpha \forall^K \) is \( \forall^K \)-uniformly distributed modulo One.

In recent times, the study of various representations of real numbers from the probabilistic point of view has gained considerable attention of Mathematicians \( \{ \text{Cf} . \text{Erdos and Renyi (1958), Renyi (1961, 1962), Galambos (1970, 1971) etc.} \} \). In their work, they found that these new approaches lead to a
much simpler roof of some known results and these points of view have led to the formulation and solution of a quite surprising number of questions which have not been investigated beforehand.

For $\beta$-adic representations of $x$, $x = \sum_{n=0}^{\infty} \frac{E_n}{\beta^n}$ where the digits $E_n$ can be obtained by the recursion formulae

$$E_0 = \lfloor x \rfloor, \quad (x) = \gamma_0$$

$$E_{n+1} = \left[ \beta \cdot \gamma_n \right], \quad \gamma_{n+1} = \left( \beta \cdot \gamma_n \right), \quad n = 0, 1, 2, \ldots$$

[Here $\lfloor \ \rfloor$ denotes integral part and $\ (\ )$ denotes fractional part $\]$

A. Renyi (1957) [43] proved some theorems on the ergodic properties of digits $E_n(x)$ and the remainders $\gamma_n(x)$

The notion of entropy was first introduced by Clausius, in thermodynamics in 1865 (Zeisse (1944)). Boltzmann's work (1923), particularly his discussions of the H-theorem and of the principle named after him, demonstrated the statistical significance of entropy. The fundamental role of entropy in classical thermodynamics and also its statistical significance became evident in subsequent developments due to Gibbs (1902), Planck (1887, 1927), Dutta (1966) and Others.
The main contributions, which really gave birth to the so-called Information Theory, came shortly after the Second World War from the mathematicians C. E. Shannon and N. Wiener. Wiener's mathematical contributions to the field of Fourier Series and later to time-series plus his genuine interest in the field of communication, led to the foundation of communication theories in general. Shannon made the first integrated mathematical attempt to deal with the new concept of the amount of Information and its main consequences. Shannon's papers laid the foundation for the new science to be named Information Theory. Shannon's earlier contribution may be summarized as follows:

1. Definition of the amount of Information from the Semi-axiomatic point of view.

2. Study of the flow of information for discrete messages, in channels with and without noise.

3. Defining the capacity of a channel i.e. the highest rate of transmission of information for a channel with or without noise.

4. In the light of 1, 2 and 3, Shannon gave some fundamental encoding Theorems. These theorems stated roughly that for a given source and a given channel one can always devise an encoding procedure leading to the highest possible rate of transmission of information.
(5) Study of the flow of information for continuous signals in the presence of noise, as a logical extension of discrete case.

Investigations with the aim of setting Information Theory on a solid mathematical basis have begun to appear only in recent years. First of all, we must mention the work of B. McMillan (1953) \[35\] in which the fundamental concepts of the theory of discrete sources (source, channel, code etc.) were first given precise mathematical definitions.

The most important result of this work must be considered to be the proof of the remarkable theorem that any discrete ergodic source has the property which Shannon attributed to sources of Markov-type and which underlies almost all the asymptotic calculations of Information Theory. This circumstance permits the whole theory of discrete information to be constructed without being limited, as was Shannon, to Markov-type sources. In the rest of his paper McMillan tries to put Shannon's fundamental Theorem On Channels with noise on a rigorous basis. In doing so, it becomes apparent that the sketchy proof given Shannon contains gaps which remain even in the case of Markov sources. The elimination of these gaps is begun in McMillan's paper, but it is not completed.
Next, it is necessary to mention the work of A. Feinstein (1954). Like McMillan, Feinstein considers the Shannon Theorem on channels with noise to be pinnacle of the general theory of discrete information and he undertakes to give a mathematically rigorous proof of this theorem. Accepting completely McMillan's mathematical apparatus, he avoids following Shannon's new and apparently very fruitful idea of a "distinguishable set of sequences".

C.E. Shannon formulated his theorem as channels with noise in two different ways. One was in terms of a quantity called "equivocation", and the other was in terms of the "probability of error". McMillan's analysis leads to the conclusion that these two formulations are not equivalent and that the second gives a more exact result than the first. Feinstein's more detailed investigation proved that although the first formulation is implied by the second, a rigorous derivation of this implication is not only non-trivial but fraught with considerable additional difficulties.

A Ya Khinchin (1956) established a complete detailed proof of both of these Shannon theorems, assuming any ergodic source and any stationary channel with a finite memory. Khinchin followed the path indicated in the works of McMillan and Feinstein, deviating from them only in few cases.
1.5. **REPORT OF THE RECENT WORK RELEVANT TO THE PRESENT INVESTIGATION.**

H.A. Hanson (1954) \[26\] first introduced the notion of Quasi-normal numbers in literature. Hanson established how to construct Quasi-normal numbers out of normal numbers. For these Quasi-normal numbers, the asymptotic frequency of any sequence of \(K\) digits was determined and it was established that every normal number is a quasi-normal number but converse is not necessarily true.

S. J. Lupkiewicz and J. E. Maxfield (1969) \[32\] solved a very interesting problem on theory of normal numbers. They defined the \(\sigma\)-tags of \(\mathcal{S}\)-numbers (in scale \(r\)) as follows:

Let \(C\) be a digit (or group of \(S\) digits) and let \(\lambda\) be a number whose decimal in the scale \(r\) has \(o(N)\) non-\(c\)-digits among the first \(N\) digits (or non-\(c\)-digit groups among the first \(\sum N\) digits). The numbers \(\lambda\) constitute the \(\mathcal{S}\)-numbers. It is proved that if \(\alpha\) is normal in scale \(r\) then \(\sigma_0\) is \(\alpha + u\lambda + \nu\), for arbitrary rational numbers \(u, \nu\) and \(\lambda \in \mathcal{S}\). This gives a continuum of normal numbers associated with \(\alpha\). If \(L\) is the class of Liouville's numbers, it is shown that

\[
\mathcal{S} \cap L \neq \emptyset, \quad \mathcal{S} \subseteq L, \quad L \subseteq \mathcal{S}
\]
J. E. Maxfield (1953) has treated normal numbers in the $K$-dimensional case. Maxfield also points out the set of numbers, simply-normal to no scale, is uncountable. As S. S. Pillai proved the equivalence of two definitions of normal numbers in one dimension, Maxfield indicated, the equivalence of two similar definitions of normal numbers in $K$-dimension. After introducing the notion of correspondent number, Maxfield established a connection between normality of $K$-tuple and that of correspondent number. Maxfield also proved that the $K$-tuple $\beta = (\alpha_1, \alpha_2, \ldots, \alpha_K)$ is normal to scale $r$ if and only if the function-system defined by

$$f_{\epsilon}(x) = \alpha_\epsilon \gamma^x_\epsilon, \ (\epsilon = 1, 2, \ldots, K),$$

is uniformly-distributed modulo one.

A. S. Besicovitch (1935) first introduced the notion of $(K, \epsilon)$ normality and $\epsilon$-normality of numbers for any given integer $K$ and any given $\epsilon > 0$. Besicovitch proved that for any given integer $K$ and any given $\epsilon > 0$, almost all integers are $(K, \epsilon)$ normal and almost all squares of integers are $\epsilon$-normal. That the squares of almost all integers are also $(K, \epsilon)$ normal is a particular case of a theorem proved by Davenport and Erdos (1952). In fact, Davenport and Erdos proved that for any $K$ and $\epsilon$, almost all the numbers $f(i), f(2), \ldots$ are $(K, \epsilon)$ normal.
where \( f(x) \) is any polynomial in \( x \), all of whose values for \( x = 1, 2, \ldots \) are positive integers. They also proved that if \( f(x) \) be prescribed as above, the real \( f(1), f(2), f(3), \ldots \) is normal decimal.

H. A. Hanson (1954) \([26]\) reexamined how the problem of the \((k,\varepsilon)\) normality of almost all of an increasing sequence of integers can be reduced to the case \( k = 1 \). Also Hanson established a sufficient condition under which the real number \( x = .a_1 a_2 a_3 \cdots a_n \cdots \) formed by writing the integers \( a_n \) in order and in juxtaposition after the decimal point is a normal number, where \( \{a_n\} \) is an increasing sequence of positive integers expressed in some scale \( r \) and having the property that almost all \( a_n \) are \((k,\varepsilon)\) normal for any given integer \( k \) and any given \( \varepsilon > 0 \). According to Hanson, the sufficient condition for normality of \( x \) is \( \forall \nu_n = O(S_n) \) where \( \nu_i \) denotes the number of digits in \( a_i \), \( (i = 1, 2, \ldots) \), and \( S_n = \sum_{i=1}^{n} \nu_i \).

Recently, much attention was paid to some ergodic properties concerning the \( r \)-adic representation of real. It was recognised by J. Cigler (1961) \([9]\) that if \( x = \sum_{k=1}^{\infty} \lambda_k r^{-k} \) all \( \lambda_k \) are non-negative integers and \( \lambda_k < \gamma_0 \) and if
be the transformation defined on unit-interval by
\[ T_x = x \pmod{1}, \text{ then there exists a } T\text{-invariant measure } \mathcal{H}\text{ equivalent with Lebesgue-measure. Also, this } T\text{ is ergodic with respect to Lebesgue-measure. J. Cigler also proved that if } 1 = \sum_{K=1}^{n+1} \alpha_K \nu^{-K}, \text{ then as a generalisation of a result by A. Renyi, the } n\text{-term sequence}

\[ S_\nu = [\lambda_{\nu+1}, \ldots, \lambda_{\nu+n}], \quad (\nu = 0, 1, 2, \ldots), \]

do \nu-adic representation of real numbers form a homogeneous Markov-Chain, probabilities being induced by measure \( \mathcal{H}\) as defined above.

After Shannon's (1948) definition of entropy in terms of the probability of events without restricting the notion of events in any way, this generalisation of the notion of entropy stimulated a good deal of research among a large number of workers in various fields of knowledge.

M. Dutta and M. Sen (1967) \[ [17] \] first considered the theory of Normal Numbers from the information-theoretic point of view. Dutta and Sen defined entropy of a real number expressed in \( \nu\)-adic representation and established the property that the entropy of a real number is maximum if and only if it is simply-normal number. After defining "Standard Number" they proved that almost all standard numbers are
simply-normal numbers and the set of numbers which are not "standard" is of measure zero. Also all simply normal numbers are standard numbers in the scale $r$ was proved by Dutta and Sen. They only suggested that notion of "degree of normalcy", almost simply-normal numbers, almost non-normal numbers might be introduced by help of entropy, but there the idea has not been discussed.

1.6. SYNOPSIS OF THE RESULTS OF THE PRESENT INVESTIGATION WITH POINTS WHICH ARE CLAIMED TO BE NEW.

The main special feature of this dissertation is that properties of Normal numbers, Normal $K$-tuples and $(K, \varepsilon)$ normal numbers have been studied from Information-theoretic point of view and so far as knowledge of author of this dissertation, such an attempt is new in theory of Normal Numbers. Such a new approach in the theory of Normal numbers and that of Normal $K$-tuples leads to a much simpler proof of the original results. Also by help of Information-Theoretic technique, new notions and results regarding normality etc., have been studied.
The present dissertation consists of Five Chapters. This present Chapter I is Introductory.

In Section 2.1 of Chapter II, some topological properties of the set of uniformly distributed sequences (modulo 1) have been obtained.

In Section 2.2 and Section 2.3 of Chapter II, some well-known properties of Normal and Quasi-normal numbers have been proved by Probabilistic and Information-theoretic technique. This is an essential simplification compared with the original proofs where the same results have been obtained by rather difficult technique. As a result of this new technique some new results have been obtained.

The following problems of Normal numbers will be studied in details in Chapter II.

In Section 2.1 of Chapter II, it has been proved that if be the set of all uniformly distributed sequences (modulo 1), then,

**Theorem 2.1.1.**
\[ T \] has infinite number of limit-points.

**Theorem 2.1.2.**
\[ T \] is dense in \((0,1)\).
Theorem 2.1.3.
Every point of $T$ is a boundary point of $T$.

Theorem 2.1.4.
Every point of $CT$ (i.e., of complement of $T$) is a boundary point of $T$.

Following Theorem 2.1.1 of Section 2.1, the following result has been introduced in Theorem 2.2.1 of Section 2.2.

Let $\alpha_0, \alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_n, \ldots$ represent a number $\Theta$ in scale $r$. Also, let

$\Theta_n = \alpha_n \alpha_{n+1} \ldots \ldots \ldots \ldots (n = 1, 2, 3, \ldots)$

and $L$ be set of limit-points of $\Theta_1, \Theta_2, \ldots, \Theta_n, \ldots$.

If $\Theta$ is normal, then $L$ contains infinity of points.

Theorem 2.2.2.
Let $S$ be a set, whose elements, when arranged in increasing order of magnitude, form a normal $\mathcal{X}$ in some scale $r$. Let $T$ be the image of $S$ under strictly monotonic preserving mapping $f$ where $f$ maps integers into integers. Let $\gamma$ be the number formed by elements of $T$, arranged in increasing order of magnitude. Then
clearly, elements of $Y$ are images of corresponding elements of $X$.

To prove that $Y$ is also normal in the same scale $r$.

Theorem 2 is a generalisation of a theorem proved by Davenport and Erdos (1952) in which they proved that the real $f(0) f(n) \ldots$ is normal decimal where $f(x)$ is a polynomial in $x$, all of whose values for $x = 1, 2, \ldots$ are positive integers. Many known and new results follow as a direct corollary to Theorem 2.2.2.

**Theorem 2.2.3.**

Let $C$ be a given digit (or finite sequence of length $S$ of digits) to scale $r$. Let $\lambda$ be a number to scale $r$ such that the number of non-$c$-digits in first $N$ digits (or $SN$ digits) of $\lambda$ is $f(N) = o(N)$. We define $\mathcal{E}$ to be class of elements of type $\lambda$.

To prove that when elements of $\mathcal{E}$ are added to normal numbers, a new set of normal numbers is generated.

Theorem 2.2.3 is a known result but here it has been proved by information-theoretic technique.

As an application of Theorem 2.2.3, the following theorem follows. This problem has been formulated as follows:
Let $\alpha = .a_1 a_2 a_3 \ldots \ldots \ldots$ is a normal number to scale $r$. We arrange $a_j$'s in increasing order of magnitude and form different difference sequences by substituting

$$a_r - a_j = d_{ij} \quad (i \geq j)$$

For a fixed $j$, and subject to variation of values of $i$, we will get different $d_{ij}$'s and hence, we obtain $r$-adic representations of following forms:

$$d_{11} \, d_{21} \, d_{31} \ldots \ldots \ldots$$

$$d_{22} \, d_{32} \, d_{42} \ldots \ldots \ldots$$

The collection of all such representations, we denote by $(A)^j$.

Next, let $A^t_j$ denotes the set $\{d_{ij} : t$ is fixed and $i \geq t \}$ and $\mathcal{F}$ be the family of sets. We consider union of family sets $\mathcal{F}$ and elements of this union are re-named as $e_1, e_2, \ldots \ldots \ldots \ldots$. From this, we get $r$-adic representation

$$e_1 e_2 e_3 \ldots \ldots \ldots$$

representing a real $\beta$ (say).

We prove then

**Theorem 2.2.4.**

(i) If $\alpha$ is normal to scale $r$, then every number of set $(A^t_j)$ is normal to scale $r$ and conversely.

(ii) If $\alpha$ is normal to scale $r$, $\beta$ is also normal to scale $r$. 
Theorem 2.2.4 is a new result in the literature.

The following result has been introduced in Theorem 2.3.1 of Section 2.3.

**Theorem 2.3.1.**

Every normal number is a quasi-normal number.

Theorem 2.3.1 is a known result but it has been proved here by Information-Theoretic technique.

In Chapter III, some properties of Normal K-tuples have been proved by Information-theoretic technique. This new approach consists in introducing the entropy of the K-tuples and as a result of this, some new ideas and new results have been evolved.

The following theorems will be discussed in details in Chapter III.

After defining the entropy of a K-tuple in scale \( r \), it has been shown that the K-tuple

\[ \beta = (\alpha_1, \alpha_2, \ldots, \alpha_K) \]

**Theorem 3.1.**

The entropy of a K-tuple is maximum if, and only if, it is simply-normal number.
Theorem 3.1 relates the simply-normality of a $K$-tuple with the entropy of the $K$-tuple. It is to be remembered here that M. Dutta and M. Sen relates simply-normality of a real in one-dimension with entropy of that real. So, Theorem 3.1 is a generalisation of the result of Dutta and Sen and a new result in Theory of Normal $K$-tuples.

Theorem 3.2.

Normality of $\alpha_1^{\varepsilon_1} \alpha_2^{\varepsilon_2} \cdots \alpha_K^{\varepsilon_K}$, where all $\alpha_\varepsilon$'s are different real imply normality of $\beta$ and conversely.

Theorem 3.2 is a new result in theory of Normal $K$-tuples. This theorem will be our main tool in this Chapter. As a result of this, some new notions and new results have been introduced. In particular, notions of absolutely normality and quasi-normality in $K$-dimension have been defined here which were not discussed by J.E. Maxfield. As a corollary to Theorem 3.2, among many results, D.D. Wall's second result has been generalised in $K$-dimension which was not discussed by J.E. Maxfield.

As a direct corollary to Theorem 3.2, it has been proved that the normality of a $K$-tuple $\beta$ remains unaltered by its addition with another $K$-tuple $\gamma$, where...
all components $\alpha_i$'s of the last $K$-tuple $\Lambda$ are of such a nature that the number of non-c-digits (C is a digit to scale $r$) in the first $N$ digits of each $\alpha_i$ are of order $\theta(N)$.

Also the normality of the $K$-tuples, obtained by the method of difference sequences of the sequence of digits of components $\alpha_i$'s of the $K$-tuple $\beta$, imply the normality of the $K$-tuple $\beta$ and conversely has been proved as an application of this theorem.

Following Maxfield's definition of correspondent number, it has been shown that:

**Theorem 3.3.**
A $K$-tuple $\beta$ is normal to scale $r$ if and only if its correspondent number to scale $r$ is normal to scale $r$.

**Theorem 3.4.**
The $K$-tuple $\beta = (\alpha_1, \alpha_2, \ldots, \alpha_K)$ is normal to scale $r$, if and only if the function-system defined by

$$f_{\alpha_i}(x) = \alpha_i x^{\alpha_i} \quad (i = 1, 2, \ldots, K)$$

is uniformly-distributed modulo one.
Theorems 5.3 and 5.4 are known results but have been studied here from Information-theoretic point of view.

In Chapter IV, different cases when entropy of a real has values in the closed interval \([0, 1]\), specially when in the closed sub-interval \([1/2, 1]\), have been investigated. In these cases, some rough bounds for probability of occurrence of digits have been estimated. Here notion of degeneracy (weak and strong) from normality, almost normal numbers, almost non-normal numbers, standard weakly-degenerated normals have been introduced. It is to be remembered here that from Number-theoretic point of view, there are only two types of reals regarding normality concept, viz., Normals and Non-normals. But as entropy gives a measure of uncertainty associated with the digits of the reals, so from the information-theoretic point of view, we may classify the set of non-normals into different categories based on the degree of degeneracy from normality and this classification is necessary for it gives some measure of non-normality.

Again, the set of standard weakly-degenerated normals being a subset of set of non-normal numbers, is of measure zero. But whether these non-normals consists of equally-likely digits (obviously not all digits will appear in such cases) or not
was an open question. In more details, no problem regarding the frequencies of digits of non-normal numbers was discussed earlier in the literature. It is very interesting to note that of all reals of the set of standard weakly-degenerated normals having some arbitrary but fixed entropy \( \frac{1}{K} \) (say) (\( K \leq 2 \)), the reals having \( \frac{1}{K} \) as their maximum entropy are most probable, i.e., almost all reals of set of standard weakly-degenerated normals consists of equally-likely digits.

Notions have been extended to the discussions of K-tuples of numbers and some new results have been obtained.

In these connections, the following theorems will be studied in details in Chapter IV.

**Theorem 4.1.**
Almost all standard weakly-degenerated normals consist of equally-likely digits.

**Theorem 4.2.**
Almost all reals of set of non-simply normals are weakly-degenerated.

**Theorem 4.3.**
Weakly-degenerated normality of \( \alpha_{i}^{*} \mathcal{S}_{\mathcal{K}} (i=1,2,\ldots,K) \),
imply weakly-degenerated normality of the $K$-tuple

$$\beta = (\alpha_1, \alpha_2, \ldots, \alpha_K)$$

but converse is not, in general, true.

Following Maxfield's definition of correspondent number to a $K$-tuple $\beta$, it has been shown that:

Theorem 4.4.

If correspondent number $\alpha$ is weakly-degenerated normal to scale $r$, then the $K$-tuple $\beta$ is also weakly-degenerated normal to scale $r$, but converse is not, in general true.

Also, investigations have been made about probabilities of digits present in a weakly-degenerated simply-normal number with maximum entropy by a method similar to that of maximum likelihood estimate. Also when average and then average along with second moment of the real are given, then it has been shown in each case that the probabilities of the digits of the real possess some mathematical relation among themselves.

In this connection, the following results have been obtained.
(i) The probabilities of digits present in a weakly degenerated simply-normal number with maximum entropy satisfies the relation

\[
\frac{1}{\gamma_0^{1/2}} \geq p > \frac{1}{\gamma_0^{1/2}}
\]

where \( \gamma_0 \) is number of digits in \( r \)-adic representation of the real.

(ii) If the average \( \alpha \) of a weakly-degenerated normal in scale \( r \) (with maximum entropy) is given, then the probabilities of digits of the \( r \)-adic representation are in a geometrical progression of common ratio \( K \) and the value of \( K \) must satisfy the relation

\[
\frac{1}{\gamma_0} < \frac{K^\alpha (1-K)}{1-K^{\gamma_0}} < \frac{1}{\gamma_0^{1/2}}
\]

(iii) If the average \( \alpha \) and the second moment \( \sigma^2 \) of a weakly-degenerated normal number in scale \( r \) (with maximum entropy) is given, then
(a) \( \frac{p_t}{p_{t-1}} \) is in geometrical progression, with common ratio \( K \) where \( p_t \) is probability of \( t \)th digits, \( t = 1, 2, \ldots, r - 1 \).

(b) if \( \frac{p_t}{p_{t+1}} = p \), then values of \( p \) and \( K \) must satisfy the inequality

\[
\frac{1}{\gamma_0} \leq \frac{p^\alpha \cdot \frac{1}{K^{2(d-\alpha)}}}{\sum_{t=0}^{n-1} K^{\frac{t}{2} + \frac{t(t-1)}{2}} \cdot p^t} \leq \frac{1}{\gamma_0 \frac{1}{2}}
\]

Chapter V contains four theorems concerning \( (K, \varepsilon) \) normal numbers. In these theorems Information-theoretic methods have been applied and as a result of this, some new results have also been obtained. The following theorems have been discussed in details in Chapter V.

**Theorem 5.1.**

Given \( \varepsilon > 0 \) and \( \gamma_0 > 0 \), no matter how small, for sufficiently large number, all sequences \( (C) \) of \( K \)-numerals in \( r \)-adic representation of the number can be divided into two groups with the following properties:
(i) the probability $p(c)$ of any sequence of $K$-numerals of first group satisfies the inequality

$$\left| \log \frac{1}{p(c)} - \frac{H}{K} \right| < \nu$$

where $H$ is maximum entropy of the real number in same scale $r$

(ii) the sum of probabilities of all sequences of second group is less than $\varepsilon$.

Theorem 5.1 is a new form of Besicovitch's result concerning $(K, \varepsilon)$ normal numbers but Theorem 5.1 gives some extra property concerning sequences $C$. The proof has been obtained by probabilities technique.

Next, let us arrange all $K$-term sequences $(C)$ of scale $B$ in order of decreasing probability $p(C)$.

We select sequences from this series in the order in which we have arranged them until the sum of probabilities of
sequences selected just exceeds a pre-assigned positive number $\lambda$, $0 < \lambda < 1$. We denote by $N_k(\lambda)$ the number of sequences so selected.

The following Theorem 5.2, permits us to make the following estimate of the number $N_k(\lambda)$.

**Theorem 5.2.**

$$\lim_{K \to \infty} \frac{\log N_k(\lambda)}{K} = \log B.$$ 

Again, we arrange all $K$-term sequences $(C)$ of scale $B$ in order of increasing probability $\mathcal{P}(C)$.

We select sequences from this series in the order in which we have arranged them until the sum of probabilities of sequences selected is just less than a pre-assigned positive number $\lambda$, $0 < \lambda < 1$. We denote by $N'_k(\lambda)$ the number of sequences so selected.

The following Corollary permits us to make the following estimate of the number $N'_k(\lambda)$.

**Corollary.**

$$\lim_{K \to \infty} \frac{\log N'_k(\lambda)}{K} = \log B.$$
Theorem 5.3.

Let \( \{a_n\} \) be an increasing sequence of positive integers having property that for any given integer \( K \) and any given \( \varepsilon > 0 \), almost all \( a_n \) are \((K, \varepsilon)\) normal in scale \( r \). Let \( \nu_i \) denotes number of digits in \( a_i \), \((i = 1, 2, \ldots, \ldots, \ldots)\), and let

\[
S_n^K = \sum_{i=1}^{\nu_i} \nu_i^K
\]

To prove that a sufficient condition for normality of the real

\[
\chi = a_1 a_2 a_3 \ldots \ldots
\]

in scale \( r \) is

\[
S_n^K \geq \nu_r 2^{\nu_r}.
\]

Theorem 5.3 implies a new sufficient condition for normality of \( \chi \), other than Hanson's condition.

After introducing the notion of \((K, \varepsilon)\) normal \(n\)-tuple, it has been proved that

Theorem 5.4.

If \( \beta = (\alpha_1, \alpha_2, \ldots, \alpha_n) \) be the \( n \)-tuple where different \( \alpha_i \)'s, \((i = 1, 2, \ldots, n)\), are real, then
(i) (K, ς) normality of components αςςςςςς(ς = 1, ...n) imply that of n-tuple β.

(ii) Almost all n-tuples are (K, ς) normal.

The notion of (K, ς) normal n-tuples is new in literature and consequently, Theorem 5.4 is also new in literature.