'Markov Chain'- The Most Invaluable Contribution of A.A.Markov Towards Probability Theory And Modern Technology: A Historical Search

Akhil Goswami
Department of Mathematics
Gauhati University, Guwahati-14, Assam, India

Gautam Choudhury
Department of Mathematical Sciences Division
Institute of Advanced Studies in Science and Technology, Paschim Boragaon, Guwahati-35, Assam, India

Hemanta Kumar Sarmah
Department of Mathematics
Gauhati University, Guwahati-14, Assam, India

Anjana Begum
Department of Mathematical Sciences Division
Institute of Advanced Studies in Science and Technology, Paschim Boragaon, Guwahati-35, Assam, India

Abstract

The Russian Mathematician Andrey Andreyevich Markov (1856–1922) contributed immensely towards the development of Applied Probability Theory. Markov extended certain sequences of dependent random variables to form special classes of 'chains' which, later on, are termed after his name as Markov Chains. The process, represented by models, in which the effect of past on future is summarized by a state which changes over time according to given probabilities is called a Markov Process. Through this paper, we are presenting a series of chronological events which led to the development of Markov chain. This might help the researchers of the current domain to get some sources to look in to for their future work. Applications of Markov chain in the fields of Biological Sciences, Physical Sciences and Modern Telecommunication System are also mentioned towards the end of the paper.
Keywords: Markov Chain, Stochastic Process, Transition matrix

---

I. Introduction

Great Russian mathematician Andrey Andreyevich Markov was significantly influenced by Chebyshev as a teacher and a researcher. At the age of 30, Markov became a Professor at St. Petersburg University and a member of St. Petersburg Academy of Sciences. He published more than 120 scientific research papers on many fields of Modern Mathematical Sciences viz. Number theory, Continuous fraction theory, Differential equations, Probability theory and Statistics. His classic textbook '*Calculus of Probabilities*' was published in Russian and was also translated into German.

Many of his papers were devoted to creating a new field of research, Markov Chain. The solution of many fundamental problems of modern science and technology could not have been possible without his contribution. It is a fitting tribute to him that these chains bear his name, acknowledging his pioneering role in the development of Markov Chains. In fact, as early as 1926, just twenty years after his initial discoveries, a paper by Russian mathematician S.N. Bernstein used the phrase 'Markov Chain'.

The present paper is divided into four main sections. In the first section, we have presented a brief sketch of Markov's personal life. In the second section, an attempt has been made to provide a glimpse of Markov's contribution towards Probability Theory. The third section is devoted to Markov's academic (research) work, with special emphasis on Markov chain. Finally, in the fourth section, we have mentioned some ongoing applications of Markov Chain in modern science and technology.

II. Andrey Andreyevich Markov's personal life:

Andrey Andreyevich Markov was born on June 14th, 1856 in the town of Ryazan, Russia. His father Andrei Grigorievich Markov was a public officer at the Forestry Department. In his early life, he was rather poor in many subjects, but not in mathematics. In 1874, Markov graduated and entered as a Faculty of Mechanics and Mathematics of St. Petersburg University. In 1877, he was awarded a gold medal for his research work entitled '*On Solution of Differential Equations with the Help of Continued Fractions*'. In 1880, Markov defended his Master Degree thesis, '*On the Binary Square Forms with Positive Determinant*'. In 1884, he received his Doctorate Degree for the thesis entitled '*On Certain Applications of the Algebraic Continuous Fractions*'. At the age of 30, Markov became a professor at St. Petersburg University, and a member of St. Petersburg Academy of Sciences. In 1900 Markov's classical textbook, '*Calculus of Probabilities*' was published. It was published four times in Russian (in the years 1900, 1908, 1913 and 1924). In 1912, it was translated into German.

After his initial entry in 1874 as a Faculty member of Mechanics and Mathematics of St. Petersburg University he was elevated to the post of an Associate Professor in 1880. In 1883 Chebyshev resigned from the same university and Markov started teaching Probability Theory. In 1886, he was elected as an Extraordinary Professor.

Markov grew up in a large family. He was one of the four sons and two daughters from the side of his mother Nadezhda Petrovna, first wife of his father. The second wife of his father Anna Josephovna had a son and two daughters. This son Vladimir Andreevich Markov also became a mathematician of repute who in the year 1882 extended Markov's inequality for algebraic polynomials to consider all derivatives. Sadly, Vladimir died at a very young age (at the age of twenty-six) of tuberculosis. In 1897, Markov completed and published his brother Vladimir's unfinished master's thesis.

When Markov's father was a manager in the estate of E.A. Valvatiev, Markov was invited to become a tutor in mathematics for Valvatiev's daughter Maria. In 1883 Markov married Maria. Their son, Andrei Andrey Andreyevich Markov (junior) (1903-1979) also became a prominent Russian mathematician. He worked in the fields of Algebra, Topology, Mechanics and Mathematical logic. From 1959 to 1979, Junior Markov chaired the Department of Mathematical logic at Moscow State University. After his death, the Department was chaired by A.N. Kolmogorov.

Markov led a very active life. He was an active political activist too. Honorary membership to St. Petersburg Academy of Sciences was often bestowed on members of the royal families and distinguished members of the society. Markov strictly opposed such honorary membership for royals. Markov refused to accept 'tsarist' awards in protest against the exclusion of the famous writer A.M. Gorky from the academy. Many people believe that it was for Markov's protest that A.M. Gorky was re-admitted into the academy. Markov wrote more than 20 letters to the various newspapers about burning social and educational issues. For these letters, the press referred him by the colorful nicknames 'Andrey the Furious' and 'The Militant Academician'.

On July 20^{th}, 1922, A. A. Markov died of sepsis. He was buried in the Mytrophany Cemetery in St. Petersburg. For detailed life history of Markov, one can go through Basharin et al. [2] and for contributions of other Russian Mathematicians to the Probability Theory, Goswami et al. [9].

III. CONTRIBUTION OF MARKOV IN PROBABILITY THEORY:

A.A. Chuprov (1874-1926), a Russian by birth, was a pioneering researcher in the field of statistics. Through his book '*Essays on the theory of Statistics*' [6], he left an indelible influence on study and research of Statistics in the Russian Empire. In the said book, he wrote about ideological conflicts in the processes of thinking about the Law of Large Numbers, especially in Russia. He observed long term stability of the proportion of successes in independent binomial trials but did not explain its cause.

During the celebration of bi-centenary of Jacob Bernoulli's theorem, a serious academic discussion between Markov and Chuprov took place and, this did not pass un-noticed [6]. Markov went on to discuss Poisson's notion on the Law of Large Numbers (LLN) as an approximating procedure but for some reason discarded it by mentioning '*....not bounding the error in an appropriate way*', and so he continued with Chebyshev's (1846) proof in Crelle's Journal[3].

Chuprov's bi-centennial talk was based on two methods of knowledge: the study of the individual entity and the study of the collective via averages, '*based on the Laws of Large Numbers*. For an illustration of the success of the latter method, Mendel's laws of heredity was cited. What seems to be of essences here is the goodness of fit to a probability model of repeated statistical observations under uniform conditions.

Markov perceived in such arguments the vexed question of statistical regularity being interpreted as the LLN, which to him was a mathematical theorem, under specific mathematical conditions, only reflects statistical regularity and does not explain it. In the year 1913 Markov published '*Examples of various methods of calculation of Probability*' (in the 3^{rd} Bicentenary edition [13]) where he discussed '*The Law of Large Numbers*'. The innovative ideas discussed were very important, as Markov himself declared in its preface. As a result of these innovations, he made some important contributions to the theory of probability. Some of these contributions are as follows:

A. *Markov's Inequality:*

Markov made seminal contributions towards the development of the Theory of Modern Probability. In 1889, Markov proposed an inequality which later on became famous as Markov's inequality and in fact, it is the most elementary tail bound in the Probability Theory. The basic idea that Markov proposed through this inequality is that if the mean of a positive random variable is small then it is unlikely to be too large too often i.e., the probability that it is too large is too small. While Markov's Inequality on its own is fairly crude, it became the foundation stone for many of the much more refined tail bounds. In mathematical language, it can be defined as- for a positive random variable X and a positive number α,

$$Pr\left(X \geq \alpha\right) \leq \frac{E(X)}{\alpha}.$$

B. Markov Theorem 1:

Through this theorem, Markov stated that $\frac{Var\ (X_1 + X_2 + ... + X_n)}{n^2} \to 0$ as $n \to \infty$ is the sufficient condition for the Weak Law of Large Number to hold for arbitrary summands $\{X_1 + X_2 + ...\}$[15]. The main significance of the theorem is that it dropped the assumption of independence, although the assumption of finite individual variance was still retained. According to Seneta [15], it is mentioned as 'Markov Theorem 1'in Russian Literature. Later on, this theorem was used to prove 'Bienaym´e–Chebyshev Inequality'.

C. Markov Theorem 2:

After initiation of his first theorem, Markov, later on, proposed an advanced version of the Weak Law of Large Number which came to be known as 'Markov Theorem 2'.
The theorem in its modern form, states as:

$$\frac{S_n}{n} - E\left(\frac{S_n}{n}\right) \xrightarrow{\Pr} 0,$$

where $S_n = \sum_{i=1}^{n} X_i$ and $\{X_i, i = 1.2,...\}$ are independent and satisfy $E\left(|X_i|^{1+\delta}\right) < c < \infty$ for some constants $\delta > 0$ and c. In this context, it is to be noted that the particular case of the theorem for $\delta = 1$ was named as "Chebyshev's Theorem" in Russian literature [15].
Seneta[15], mentioned that in a paper of Markov entitled 'O ZadacheYakova Bernoulli' (i.e. On the Problem of Jacob Bernoulli) which was published in 1914, Markov modified De' Moivre's approximation formula

$$\frac{1}{\sqrt{\pi}}\int_{z}^{\infty} e^{-x^2}dx \quad \text{for} \quad \Pr\left(X > np + z\sqrt{2npq}\right)$$

to the form

$$\frac{1}{\sqrt{\pi}}\int_{z}^{\infty} e^{-x^2}dx + \frac{(1-2z^2)(p-q)e^{-z^2}}{6\sqrt{2npq}},$$

and named it as Chebyshev's formula. This paper motivated other Russian Mathematicians to a great extent.
In addition to these, in 1906, Markov developed the basic concept of the Weak Law of Large Number and Central Limit Theorem to a certain sequence of dependent random variables [2]. In 1910, he proved that the independence of random variables was not a necessary condition for the validity of the Weak Law of Large Number and Central Limit Theorem. Markov mentioned about the applications of the method of Mathematical Expectations, method of Moments and also about the proof of the Second Limit Theorem of the calculus of Probabilities.

IV. MARKOV'S CONTRIBUTION ON STOCHASTIC PROCESS WITH SPECIAL EMPHASIS ON MARKOV CHAIN:

The concept of 'chains' was first introduced in Markov's paper [12] which he communicated in 1906, in which he considered chains with only two states 0 and 1. The journal article, however, was not published until 1907. By that time, Markov enriched the manuscript by adding two more sections, introducing the general concept of a chain. Though Markov discussed the chain, he did not name it as 'Markov chain'. In 1920, S.N. Bernstein used the word 'Markov's chain' for the first time.
In this paper, Markov defined the simple chain as '*an infinite sequence* $x_1, x_2, ..., x_k, x_{k+1}, ...$ *of variables connected in such a way that* x_{k+1} *for any* $'k'$ *is independent of* $x_1, x_2, ..., x_{k-1}$, *in case* $'x_k'$ *is known*. Markov called the chain is homogenous if conditional probability distributions of x_{k+1} forgiven x_k were independent of $'k'$. Many systems have this property that given the present state, the past states have no influence on the future. This property of chain is called Markov property. It is mathematically stated as,

$$\Pr\{X_{k+1} = i_{k+1} \mid X_0 = i_0, X_1 = i_1, ..., X_k = i_k\} = \Pr\{X_{k+1} = i_{k+1} \mid X_k = i_k\} \qquad (4.1)$$

For every choice of the non –negative integer $'k'$ and states $i_0, i_1, ..., i_{k+1} \in S$ for which left-hand side is defined and S represents state-space, which is, in fact, a collection of all possible states that can arise in the chains.

A. Transition Probability Matrix:

In the context of (4.1), let us use the following notation,

$$Pr\{X_{k+1}=i_{k+1}\mid X_k=i_k\}=P^{(k,k+1)}_{i_k,i_{k+1}} \qquad (4.2)$$

In the new notation system, clearly, the right- hand side depends on the present state, future state and '*k'*. If $P^{(k,k+1)}_{i_k,j_{k+1}}$ is independent of *k* then we say that Markov Chain is time homogeneous and probabilities are stationary transition probabilities,

$$Pr\{X_{k+1}=i_{k+1}\mid X_k=i_k\}=Pr\{X_{k+1}=j\mid X_k=i\}=P_{ij} \qquad (4.3)$$

The probability $P_{i,j}$ is called one –step transition probability and the square matrix $P=\left(P_{i,j}\right)$ one-step transition probability matrix when the row sum is added up to1. In simple language, the expression (4.3) represents the transition probability of moving from state '*i'* in the k^{th} trial to the state '*j*' in the (k+1)th trial.

1) Example:

Suppose that tomorrow's weather conditions will be the same as it is today. For simplicity, we assume that there are only two kinds of weather: rain or dry. Suppose it is also assumed that today's weather is rainy. Then there will be rain tomorrow will have probability '*p*' and if we assume today's weather is dry then it will be dry, with probability 1-*p*. Then the weather condition will form a two-state Markov Chain with state-space $S=\{0,1\}$ (where 0 denotes rainy day and 1 denotes dry day) , and consequently, the transition probability matrix will be,

$$P=\begin{pmatrix} p & 1-p \\ 1-p & p \end{pmatrix}$$

Most of Markov's works were devoted to simple homogeneous chains. By determining the probability P_{ij} of an event $x_k=i$ are connected by the simple formula

$$P_j^{(k+1)}=\sum_{i\in s} p_i^{(k)}p_{i,j} \qquad (4.4)$$

Markov also deduced the following equalities for the mathematical expectations,

$$a_i=E\left(x_i\right),\ A_j^m=E\left(X_{k+m}\mid x_k=j\right),\quad a_{k+m}=\sum_i p_i^{(m)}A_i^{(m)},\ A_i^{(m)}=\sum_j p_{i,j}A_j^{(m-1)} \qquad (4.5)$$

In the same paper [12], Markov studied the correctness of the Week Law of Large Number for homogeneous chains with positive transition probability matrices. From (4.5) he established that a_{k+m} is located between the extreme values $\alpha^{(m)}=\min_{i\in s} A_i^{(m)}$ and $\beta^{(m)}=\max_{i\in s} A_i^{(m)}$. Further, he showed that these latter numbers are located between $\alpha^{(m-1)}$ and $\beta^{(m-1)}$. Markov proved that if ' m ' increases, the difference $\gamma^{(m)}=\beta^{(m)}-\alpha^{(m)}$ tends to 0, so that mathematical expectations of a_{k+m} and $A_i^{(m)}$ have the same limit ' a '. His elegant proof of this theorem, in modern terminology, is called the Ergodic theorem. In fact, this is one of the founding pillars of Markov's chain theory, and it establishes the conditions under which a Markov chain can be analyzed to determine its steady-state behavior. The statement of the theorem is presented below:

2) Theorem: (Markov [12])

For a chain with a positive transition matrix all the numbers a_{k+m} and $A_i^{(m)}$ have the same limit, which they differ from by numbers less than $\gamma^{(m)}$. At the same time, $\gamma^{(m)}<CH^m$, where C and H are constants and $0<H<1$.

In the year 1908, in his paper [14], Markov dropped the assumption that the transition probability matrix' P ' be positive and described what we call irreducible chains in the following way: 'we consider only those chains $x_1, x_2, \ldots, x_n, \ldots$, where the appearance of some numbers (say), i, j, k do not rule out the possibility of the ultimate appearance of the others.' In brief, we can say that for irreducible chains, every state is reachable from every other state. Markov produced the following criterion which was later on termed as the condition of the irreducibility of a homogeneous chain:

$$\begin{vmatrix} u & p_{j,i} & p_{k,i} & \cdots \\ p_{i,j} & v & p_{k,i} & \cdots \\ p_{i,k} & p_{j,k} & w & \cdots \\ \vdots & \vdots & \vdots & \vdots \end{vmatrix}$$

The determinant with variables $u, v, w, \cdots$ cannot be transformed into the product of several determinants of the same type. A formal definition of the concept of irreducibility did not come until 1912, the year in which a famous paper [8] by Frobenius got published. It appears that we can translate Markov's condition for irreducibility into today's language as: 'no symmetric permutation can transform P into the block form

$$\begin{pmatrix} X & Y \\ 0 & Z \end{pmatrix}.$$

After the development of Markov chain, it was seen that many earlier ideas of classical probability can be connected to the new idea of a Markov chain. It was verified that some of the urn problems studied by Laplace, D. Bernoulli and Ehrenfests are special cases of Markov chain. Random walks, such as the gambler's ruin problem, studied by Bunyakovsky, also have a clear connection to Markov chains. Similarly, Bachelier's lesser known study of stock exchange had ties to Markov's work. It is also to be noted here that Markov chains have a direct relationship to Brownian motion. However, most of these earlier studies were either unknown or unrecognized by Markov and were certainly not his motivation for developing the chains, which latter on carried his name. His long-time fellow mathematician Chuprov pointed out some of the fore-runners of this field of study and about their work in a letter to Markov in the year 1910. In a reply letter to Chuprov, in November 11, 1910 Markov wrote '*I most humbly beg you to point out to me those articles of Bohlmann and Bachelier to which you refer. Up to now I had thought that I was the first to dwell on the application of the law of large numbers to dependent variables…*'.Chuprov's earlier mentioned letter which prompted this remark acknowledged the connections of Markov's predecessors. After carefully reading the articles mentioned by Chuprov, Markov found his work on chains more general, complete and correct. In November 15th, 1910, Markov in a letter again to Chuprov wrote '*I, of course, have seen Bachelier's article but strongly dislike it. I do not attempt to judge its significance for statistics, but with respect to mathematics, it has no importance in my opinion. In any case, it does not contain an extension of Bernoulli's theorem to dependent variables*'. A few days later, on November 18th, he continued '*The cases I indicated are not included in Bohlmann's cases, but contains them as special cases. There is a huge difference. I am prepared to admit that Bohlmann gave an elegant special formula but he did not point out even one new (after my article) case of the generalization of Bernoulli's theorem*.' Such language hints at Markov's research style and character. He was very exacting, thorough and complete in his own work and also was not shy about spelling out what he deemed rightful criticism of the work of his colleagues. In fact, no colleague was excluded from Markov's judgmental eye. Not even his mentor Chebyshev, whom Markov implicitly accused of plagiarism in 1891[2,16].

V. APPLICATIONS OF MARKOV CHAINS IN MODERN SCIENCE AND TECHNOLOGY:

Every action or incidence in our life that takes place around us consists of random components in its performance. The random components invariably depend on time, space, volume etc. These situations trigger modern research to look into some unexpected phenomenon to bring into the ambit of mathematical functions and analysis. The primary concern of this section is to provide some information to the readers about notions of Markov Chain as a foundation stone for mathematical modeling with random components and its possible applications in Biological Science, Physical Science and Modern Telecommunications Technology.

A. Application of Markov Chain to Biological Sciences:

The theory of heredity, originated by G.Mendel (1822-1884), provided instructive illustrations for the applicability of simple probability models. We shall restrict ourselves to indications concerning the most elementary problems. In describing the biological background, we shall necessarily simplify and concentrate on such facts as are pertinent to mathematical treatment.

Inheritable characters depend on special carries, called genes. All cells of the body, except the reproductive cells or gametes, carry exact replicas of the same gene structure. The salient fact is that genes appear in pair. We can visualize them as a vast collection of beads or short pieces of string, the chromosomes. These also appear in pairs. The paired genes occupy the same position on paired chromosomes. In the simplest case, each gene of a particular pair can assume two forms (alleles), A, a. Then three different pairs can be formed, and with respect to this particular pair, the organism belongs to one of the three genotypes AA, Aa, aa (i.e., there is no distinction between Aa and aA).

The genotypes of offspring depend on a chance process. At every occasion, each parental gene has probability $\frac{1}{2}$ to be transmitted, and successive trials are independent. In other words, we conceive of the genotypes of 'n' offspring as the result of n independent trials, each of which corresponds to the tossing of two coins. For example, the genotypes of descendants of an $Aa \times Aa$ pairing are AA, Aa, aa with respective probabilities $\frac{1}{4}, \frac{1}{2}, \frac{1}{4}$.

Looking at the population as a whole, we conceive of the pairing of parents as the result of a second chance process. We shall investigate only the so-called random mating, which is defined by the condition: If r descendants in the first filial generation are chosen at random, then their parents form a random sample of size 'r', with possible repetitions from the aggregate of all possible parental pairs.

Suppose that the three genotypes AA, Aa, aa occur among males and females in the same ratios, $u : 2v : w$. We shall suppose $u + 2v + w$ and we may call u, $2v$, w the genotype frequencies. Now, let us assume,

$$p = u + v, \qquad q = v + w \tag{5.1.1}$$

Clearly, the numbers of *A*-genes and *a*-genes are as $p : q$, since $p + q = 1$. We shall call p and q the gene frequencies of A and a. In each of the two selections, an *A*-genes is selected with probability p and because of the assumed independence, the probability of an offspring being AA is p. The genotype Aa can occur in two ways, and its probability is therefore $2pq$. Thus, under random mating conditions, an offspring belong to the genotypes AA, Aa or aa with probabilities,

$$u_1 = p^2 \, , \, 2v_1 = 2pq \; , \, w_1 = q^2 \qquad (5.1.2)$$

It is also observed that from the second generation on, there is no tendency toward a systematic change; the steady state is reached with the first filial generation. This difficulty in Mendelian law was first noticed by G.H. Hardy (1908)[10]. He also observed that the three genotypes in n^{th} generation are three random variables whose expected values are given by (5.1.2) and do not depend on 'n'. Their actual values will vary from generation to generation and form a stochastic process of Markov Chain type. For a detailed discussion on this topic, we may refer to Feller [7].

B. *Application of Markov Chains in Physical Sciences:*

A reaction network is a chemical system involving multiple reactions and chemical species. The simplest stochastic models of such network treat the system as a Markov Chain with state of the system being the number of molecules of species with reaction models as possible transitions of the chain. It is a common place to apply the theory of Markov Chains to the problems connected with the study of statistical thermodynamics of infinitely long polymer chains. The problems connected with polymer chains results from volume exclusion. The volume exclusion can result from non-self-intersection of polymer chains or form steric (relating to the arrangement of atoms in space) hindrances imposed on a molecule near the surface. The stochastic model of Markov Chain can be employed in connection with these problems, which describe the behavior of a random walk with reflecting barriers. In problems dealing with self- exclusion of polymer volume, a different random walk is found in the literature, which is descriptive of these polymer chains. Different polymer configurations are included in the matrix of transition probability representing absorbing Markov State i.e., random walk with absorbing barrier. It should be noted here that to deal with these types of problems, one has to take proper care of the excluded volume effect of these chains keeping an eye on the fact that stochastic nature of the process is not lost. Mazor [11] considered such a type of model in order to study the statistical thermodynamics of infinitely long polymer chains, simulated by a random flight on a lattice. To make the simulation study of such a type of polymer chain, the author considered the transition probability matrix (TPM) of the following nature in one of the models under the study.

$$P = \begin{pmatrix} 1 & (q-1)^{-1} & (1-n)/(q-1) & 0 & \cdots & 0 & 0 \\ 0 & 0 & n/(q-1) & 0 & \cdots & 0 & 0 \\ 0 & p_{21} & p_{2,2} & p_{2,3} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & p_{i-2,1} & p_{i-2,2} & p_{i-2,3} & \cdots & p_{i-2,i-1} & 0 \\ 0 & p_{i-1,1} & p_{i-1,2} & p_{i-1,3} & \cdots & p_{i-1,i-1} & p_{i-1,i} \\ 0 & p_{i,1} & p_{i,2} & p_{i,3} & \cdots & p_{i,i-1} & p_{i,i} \end{pmatrix}$$

In order to know the status of the thermodynamic functions of the polymer model, one has to compute the largest Eigen value of the TPM 'P'.

C. *Application of Markov Chain in Modern Tele-Communication System:*

Agner Krarup Erlang (1878-1929) was a Danish mathematician, statistician and engineer who worked for the Copenhagen Telephone Exchange and for the first time invented the field of traffic engineering and telecommunication system in the form of queueing theory. While he was working for the Copenhagen telephone company, Erlang proposed the classical problem of determining how many circuits were needed to provide a satisfactory service. For solving the problem, he realized that mathematics could be applied to assess how many operators were needed to handle a given volume of telephone calls. During that period of time, the telephone exchanges used human operators and cord boards to switch telephone calls by means of jack plugs. Erlang was the first person to study the problems of telephone networks. Through his investigation in a village telephone exchange, he worked out a formula, now known as Erlang's formula. Although Erlang's model is a simple one, the mathematics underlying today's complex telephone networks is still based on his work. A.A. Erlang published his first paper in 1909. Erlang's impressive contribution to tele-traffic stimulated and continue to stimulate an enormous volume of work on queueing theory.

Around the year 1950, the theory of queues entered a period of intensive investigation by workers in a verity of fields, particularly in mathematics and operations research. David Kendall (1951) was the pioneer who viewed and developed the theory from the perspective of Stochastic Processes. Kendall (1951,1953) first introduced the method of embedded Markov Chain technique, which is undoubtedly a most powerful and popular technique widely used in various branch of stochastic processes, especially in queueing theory. He applied this method to basic *M /G /1* system in 1951 and *G /M /S* system in1953 by using the concept of regeneration points, which was pointed out by Palm in 1943. The basic idea behind this technique is to simplify the descript of the state, which is from two-dimensional state to one-dimensional state space. Embedded Markov Chain is basically a homogeneous Markov Chain. In the earlier mentioned paper of 1951, Kendall considered the transition probability matrix (TPM) to be of the following type,

$$P = \begin{bmatrix} p_{00} & p_{01} & p_{02} & \cdots \\ p_{10} & p_{11} & p_{12} & \cdots \\ 0 & p_{21} & p_{22} & \cdots \\ 0 & 0 & p_{32} & \cdots \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

In 1965, Allan L. Scherr completed his thesis, '*An Analysis of time chaired Computer System*', and received his PhD in electrical engineering from M.I.T. At that time, the compatible time staring system was new to the M.I.T campus and allowed 300 users to access the computer and its software interactively. The goal of Scherr's[1] thesis was to characterize the system's usage. He conducted simulation studies to predict the system's usage and wanted to compare this with real data from similar systems. He found that no such data existed, so he conducted his own comprehensive measurements of system performance. Scherr declared his analysis of time-shared systems complete after he obtained his own real data and compared this with his simulation results based on a single server queueing system.

Scherr's single-server queue Markov Chain captured the dominating aspect of the system. The tri-diagonal transition rate matrix A appearing in Scherr's thesis is shown below. The states of the chain are the number of users interacting with the system and thus, are labeled $(0,1,2,\cdots,n)$

$$A = \begin{pmatrix} \frac{-n}{T} & \frac{n}{T} & . & . & . & . & . & . \\ \frac{1}{p} & -\left(\frac{1}{p}+\frac{n-1}{T}\right) & \frac{n-1}{T} & . & . & . & . & . \\ . & \frac{1}{p} & -\left(\frac{1}{p}+\frac{n-2}{T}\right) & \frac{n-2}{T} & . & . & . & . \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ . & . & . & . & \frac{1}{p} & -\left(\frac{1}{p}+\frac{2}{T}\right) & \frac{2}{T} & \\ . & . & . & . & . & \frac{1}{p} & -\left(\frac{1}{p}+\frac{1}{T}\right) & \frac{1}{T} \\ . & . & . & . & . & . & \frac{1}{p} & -\frac{1}{p} \end{pmatrix}$$

Scherr assumed 'n' as number of users in the system.

Almost ten years later, the 'performance evaluation committee' noted Scherr's ground breaking contribution through his thesis [1] of 1965. In 1975, his thesis was awarded the Grace Murray Hopper Award by the Association for Computing Machinery.

More recently, Choudhury et.at [4,5] have applied Embedded Markov Chain technique to establish the stochastic decomposition result for two types of service queueing system to include several kinds of generalizations for vacation queueing model. Vacation models are characterized by the fact that the idle time of the server may be utilized for some other secondary jobs, for instance, to serve the customers in other systems. Allowing the server to take vacations makes the queueing model more realistic and flexible in studying real-world queueing situations. Applications of such models arise naturally in call centers with multitasking employees, modern telecommunication and computer networks, production and quality control problems etc.

VI. CONCLUSION:

In this paper, we have presented the important contributions of A.A. Markov towards Probability and Statistics in general and Markov chain in particular. We have presented the chronological developments which led to the introduction of the Markov chain, which has become a fundamental tool in solving many modern technology related problems. We have cited some situations from Biological and Physical sciences also where the technique of Markov chain can be used to solve problems arising in those fields. The chronology of events and information cited in the paper may stimulate some readers to go for fruitful research in the field of stochastic processes, which we feel, will continue to be the backbone of many more technological developments in the years to come.

REFERENCES

[1] Allan Lee Sherr: 'An Analysis of Time-shared Computer System', Ph.D. thesis, Massachusetts Institute of Technology, 1962.
[2] Basharin G. P., Langville A. N.and Naumov V. A. : 'The Life and Work of A. A. Markov, Linear Algebra and its Applications', 2004, 386: 3-26.
[3] Chebyshev P.L.: 'An Experience in Elementary Analysis of the Probability Theory', Crelle's Journal fur die reine und angewandte Mathematik, 1846, 33, 259-267.
[4] Choudhury G, Goswami A., Begum A. and Sarmah H.K. : ' Stochastic decomposition result of an unreliable queue with two types of sevices', Mathematics and Statistics, 2020, 8 (2); 225-232.
[5] Choudhury G., Goswami A., Begum A. and Sarmah H.K. : 'A note on Stochastic decomposition Results and its Application.', Thailand Statistician (in press),2020.
[6] Chuprov A.A.: 'Ocherkipob Teoriistatistiki' [Essays on the theory of Statistics] Sankt Peterburg [2nd edition of 1910 was reprinted in 1959, by Gosstatizelat: Moskva].
[7] Feller.W.: 'An introduction to probability theory and its applications', volume-I, Jhon Wiley and Sons, Inc, New York, 1959.
[8] FrobeniusG.F.: 'Uber Matrizen aus nicht negutiven Elementeen', Sitzungaber, Preuss. Akad.Wiss., Berl.,1912, 456-477.
[9] Goswami A., Choudhury G. and Sarmah H.K.: 'Contribution of Russian Mathematicians in the development of Probability: A Historical Search', International journal of statistics and systems, 2019, 14(1), 1-27.
[10] Hardy. G.H.: ' Mendelian proportions in mixed population', Letter to the Editor, Sciences, N.S.;1908, 28,49-50.
[11] Major, J.: 'Application of the theory of absorbing Markov Chains to Statistical Thermodynamics of Polymer Chains in a Lattice'. The Journal of Chemical Physics, 41(8), 1964, 2256-2266.
[12] Markov A.A.: 'Rasprostranenie zakona bol'shih chisel na velochiny', zavisyaschine drug ot druga, Izvestiya Fiziko-matematicheskogo obschestva pri kazanskom universityete, 2-ya seriya, tom 15, 9, 4,1906, 94; 135-156.
[13] Markov A.A.: 'Ischislenie Veroiatnostei' [The calculus of Probabilities], 3rd ed., Sankt Peterburg; Tipografia Imperatorskoi Akademii Nauk, 1913.
[14] Markov A.A.: ' Rasprostranenie predel'nyh teorem ischisleniya veroyatnostej na summu velichin svyazannyh v cep', zapiski Akademii Nauk po Fiziko-mathemticheskomu otdeleniyu, VIII seriya, tom 25(3), 1908.[Translated into English, Extension of limit theorems of Probability Theory to a sum of variables connected in a chain (translated by S. Petelin) in R.A. Howard, ed. Dynamic Probabilities Systems, vol.1, wiley, New York, 1971, 552-576].
[15] Seneta Eugene : 'A Tricentenary history of the Law of Large Numbers', Bernoulli 19(4) ,2013, 1088-1121.
[16] Seneta Eugene: 'Markov and the creation of Markov chains'. Markov Anniversary Meeting 2006. Boson books, pp.1-20.