CHAPTER ONE
INTRODUCTION

1.1 Background to the Study

Markov chain is a mathematical system that experiences transitions from one state to another according to certain probabilistic rules. The defining characteristic of a Markov chain is that no matter how the process arrived at its present state, the possible future states are fixed. In other words, the probability of transitioning to any particular state is dependent solely on the current state.

When a system that can change over time is studied, a way to keep track of those changes is needed. A Markov chain is a particular model for keeping track of systems that change according to given probabilities. Markov chain allows one to predict future events, but the predictions become less useful for events farther into the future. It can thus be used for describing systems that follow a chain of linked events, where what happens next depends only on the current state of the system. Markov chains are a fairly common and relatively simple way to statistically model random processes.

Markov chain models have been the most widely used in the study of random fluctuations in the genetics compositions of a population over generations. Besides being a convenient theoretical tool, Markov chains have provided rather satisfactory theoretical explanations to some observed long run phenomena related to the genetic structure of population.

Genetics, the study of heredity, is a system that changes over time and in order to keep tracking of this system, a Markov chain may be employed**.**

Heredity is the passing on of traits from parents to their offspring, either through asexual reproduction or sexual reproduction in which case, the offspring cells or organisms acquire the genetic information of their parents. Through heredity, variations between individuals can accumulate and cause species to evolve by natural selection. The study of heredity in biology is known as genetics.

In humans, eye colour is an example of an inherited characteristic: an individual might inherit the "brown-eye trait" from one of the parents. Inherited traits are controlled by genes and the complete set of genes within an organism's genome is called its genotype.

The complete set of observable traits of the structure and behavior of an organism is called its phenotype. These traits arise from the interaction of genotype with the environment. As a result, many aspects of an organism's phenotype are not inherited. For example, suntanned skin comes from the interaction between a person's phenotype and sunlight; thus, suntans are not passed on to children. However, some people tan more easily than others, due to differences in genotypes. Another striking example is people with the inherited trait of albinism, such as the skin, eye and hair colour which may change over time.

Heritable traits are known to be passed from one generation to the next via deoxyribonucleic acid (DNA), a molecule that encodes genetic information. DNA is a long polymer that incorporates four types of bases, which are interchangeable. The sequence of bases along a particular DNA molecule specifies the genetic information. This is comparable to a sequence of letters spelling out a passage of text. Before a cell divides through mitosis, the DNA is copied, so that each of the resulting two cells will inherit the DNA sequence. A

portion of a DNA molecule that specifies a single functional unit is called a gene and different genes have different sequences of bases. Within cells, the long strands of DNA form condensed structures called chromosomes. Organisms inherit genetic materials from their parents in the form of homologous chromosomes, containing a unique combination of DNA sequences that code for genes. The specific location of a DNA sequence within a chromosome is known as a locus. If the DNA sequence at a particular locus varies between individuals, the different forms of this sequence are called alleles. DNA sequences can change through mutations, producing new alleles. If a mutation occurs within a gene, the new allele may affect the trait that the gene controls; thus altering the phenotype of the organism. However, while this simple correspondence between an allele and a trait works in some cases, most traits are more complex and are controlled by multiple interacting genes within and among organisms. Developmental biologists suggest that complex interactions in genetic networks and communication among cells leads to heritable variations that may underlie some of the mechanics in developmental plasticity and canalization. (Robinson *et.al* 2018)

Recent findings have confirmed important examples of heritable changes that cannot be explained by direct agency of the DNA molecule. These phenomena are classed as epigenetic inheritance systems that are causally or independently evolving over genes. Research into modes and mechanisms of epigenetic inheritance is still in its scientific infancy. However, this area of research has attracted much recent activity as it broadens the scope of heritability and evolutionary biology in general. Bollati and Baccarelli (2010)

A gene is the basic physical and functional unit of heredity. Genes, which are made up of DNA, act as instructions to make molecules called proteins. In humans, genes vary in size from a few hundred DNA bases to more than 2 million bases. The Human Genome Project has estimated that humans have between 20,000 and 25,000 genes. Chail (2008)
Every person has two copies of each gene, one inherited from each parent. Most genes are the same in all people, but a small number of genes (less than 1 percent of the total) are slightly different between people. Alleles are forms of the same gene with small differences in their sequence of DNA bases. These small differences contribute to each person's unique physical features. (Robinson *et.al* 2018)

Trait is formed by genes which come in different varieties, called alleles. Somatic cells contain two alleles for every gene, with one allele provided by each parent of an organism. Often, it is impossible to determine which two alleles of a gene are present within an organism's chromosomes, based solely on the outward appearance of that organism. However, an allele that is hidden, or not expressed by an organism, can still be passed on to that organism's offspring and expressed in a later generation. A trait in one generation is inherited, but or not be outwardly apparent before two or more generations. Gregor Mendel (1866) was the first person to describe the manner in which traits are passed on from one generation to the next (and sometimes skip generations). Through his breeding experiments with pea plants, Mendel established three principles of inheritance that described the transmission of genetic traits before genes were even discovered. These are Law of Segregation, Law of Independent Assortment and Law of Dominance. Today, scientists use the word "phenotype" to refer to what Mendel termed an organism's "external

resemblance," and the word "genotype" to refer to what Mendel termed an organism's "internal nature." Indeed, Mendel's experiments revealed that phenotypes could be hidden in one generation, only to re-emerge in subsequent generations. Mendel thus wondered how organisms preserved the hereditary materials associated with these traits in the intervening generation, when the traits were hidden from view.(O'Neil 2012)

Traits are the ways in which family members are alike. A family trait is a genetic marker that is passed through parents' genes to their children. Most specific traits are passed directly from one or both parents. Genetic disorders are also traits that are passed from a parent to a child.

According to science, traits that are passed down to children may be dominant or recessive. A dominant trait is a likeness that comes from a more powerful gene. For example, a female parent with blue eyes might pass a blue-eyed gene down to her child, but if her husband has brown eyes, the child may have brown eyes as well, because brown-eyed genes are more dominant than blue-eyed genes. Recessive traits can be passed down by both parents. Green eyes and blue eyes are considered recessive traits; So, a child with one blue-eyed parent and one green-eyed parent can have either eye colour. Genetic disorders can also be dominant or recessive. Examples of genetic disorders from dominant genes include Huntington disease, Down syndrome while examples of recessive genetic disorders include cystic fibrosis, sickle cell anemia and albinism. (Science clarified 2018)

A genetic marker is a gene or DNA sequence with a known location on a chromosome that can be used to identify individuals or species. It can be described as a variation (which may arise due to mutation or alteration in the genomic loci) that can be observed.

Genetic markers can be used to study the relationship between an inherited disease and its genetic cause (for example, a particular mutation of a gene that results in a defective protein). It is known that pieces of DNA that lie near each other on a chromosome tend to be inherited together. This property enables the use of a marker, which can then be used to determine the precise inheritance pattern of the gene that has not yet been exactly localized. Genetic markers are employed in genealogical DNA testing to determine genetic distance between individuals or populations.

The external resemblance, Phenotype, is the observable physical or biochemical characteristics of an individual organism, determined by both genetic make-up and environmental influences, for example, height, weight and skin colour. A phenotype results from the expression of an organism's genetic code, its genotype, as well as the influence of environmental factors and the interactions between the two.

The interaction between genotype and phenotype has often been conceptualized by the following relationship:

genotype (G) + environment (E) → phenotype (P)

The genotype–phenotype distinction is drawn in genetics. "Genotype" is an organism's full hereditary information. "Phenotype" is an organism's actual observed properties, such as

morphology, development, or behavior. This distinction is fundamental in the study of inheritance of traits and their evolution.(Laird and Lange 2011)

1.2 Statement of the Problem

Genetic testing, deoxyribonucleic acid is a type of medical test that identifies changes in chromosomes, genes, or proteins. The results of a genetic test confirm or rule out a suspected genetic condition or help determine a person's chance of developing or passing on a genetic disorder. Therefore, Genetic testing is recommended. However, the test should be accessible, the cost of which should be reasonable and not exorbitant. Furthermore, results of the test should provide a sense of relief from uncertainty and within a moderate time.

On the contrary, this test is not easily accessible as well as being very costly. Consider a situation where a set of traits common to a family members that are visible or not visible in succeeding generations. In the absence or lack of access to laboratory test such as DNA, there is a need to determine the presence of traits and therefore link the family members.

In response to these problems, this study proposes an alternative test (a non-laboratory test) that is accessible, less costly, and provides relief from uncertainties and within a moderate time.

1.3 Justification of the Study

Genetic screening uses a variety of laboratory procedures to determine when a person has a genetic condition or disorder or is likely to develop a disease based on his or her genetic

makeup. This laboratory procedure is faced with a host of challenges such as cost, accessibility and poor understanding of the process by patients.

The main contribution of this research work includes, but not limited to the following: development of a probabilistic (non laboratory test) procedure that performs well and can be used as an alternative test in the absence of a laboratory test such as deoxyribonucleic acid (DNA) test to identify trait and to solve some of the challenges faced by the need of Laboratory test.

1.4 Aim and Objectives

The aim of this research work is to use Markov Chain to develop identification of a marker to link traits.

Objectives are to:

i identify traits, the state spaces and classify them using probability transition matrix

ii. determine the properties of the state space for each trait
iii. determine which of the traits are the markers
iv. demonstrate the use of the procedure using real life data

CHAPTER TWO

LITERATURE REVIEW

In recent years, several researchers have worked on heredity (the study of genetics) the heritable traits are known to be passed from one generation to the next. The approach used has been via deoxyribonucleic acid DNA. The DNA test, which is a laboratory test has been sparingly used because of the difficulty associated with it. Therefore, attempts were made to employ some non laboratory methods as alternatives.

An investigation of immigrants and their descendants undertaken by Boas (1910) in a paper on "Family Traits as Determined by Heredity and Environment" showed certain differences between those born abroad and those born in America employing some statistical parameter estimates. The first study was based on a comparison of gross series of individuals born abroad and of others born in America. It seemed possible that these differences were due to a different composition of the two series; that for some reasons, those born abroad originated from localities or from social strata not identical with those to which the American series belonged. Further investigation by Boas was based on a comparison of parents and their own children, and of fraternities, some of whose members were born abroad while others were born in America. As a result of these, he stated that under new environment, stature, head form and transversal diameter of the face undergo changes. He also showed that correlation of head index between parents and children are affected by hereditary similarities. Boas (1910) considered the relationship between fraternal variability and the variability of family lines which show that the coefficient of fraternal correlation depends solely upon the variability of family lines and the general comparison of the

population.

When the family lines are alike, the coefficient of fraternal correlation is low and when they are diverse, it is high. He assumes that all families consist of an infinite number of members of a fraternity.

The total variability σ of the population was

$$\sigma^2 = \sigma_1^{\ 2} + \sigma_{fr}^{\ 2} \qquad 2.1$$

And $$r_{fr} = \frac{[(x_1 + y^{'})(x_1 + y^{'})}{\sigma^2} \qquad 2.2$$

Since the number of members of the fraternity is taken as infinite, products yy are negligible and

$$r_{fr} = \frac{\sigma^2_1}{\sigma^2} \qquad 2.3$$

where the family mean designated by x; individual member deviation designated by y. The variability of family mean is designated by σ_i and that of fraternity by σ_{fr}. Developing lines of research on family interaction patterns suggests that delinquent behaviors of children and adolescents maybe related to, if not a direct function of, separate disordered interpersonal interaction within the family unit. But although researchers have been reasonably successful in identifying certain recurrent behavioral patterns in family interaction, overall progress has been constrained by limitations of the methods typically employed.

Holmes (1943) in his study of family fixtures used 234 figures, most of which are photographic reproductions illustrating the occurrence of similar characteristics in two or more members of a family. The selection was confined to normal traits such as stature,

obesity, head form, shape of ears, peculiarities of the eyes, shape and jointing fingers and toes, form colour and arrangement of hair, shape of nose, and many other features except those which are obviously pathological. Through, the wealth of photographic reproductions, most of which are very clear, the facts of heredity make a direct appeal to the eye.

Tweedie(1975) gave a sufficient condition for ergodicity and recurrence of Markov chains on a general state space. Let $\{X_n\}$ be a $\emptyset$-irreducible Markov chain on an arbitrary space. Sufficient conditions are given under which the chain is ergodic or recurrent. In particular, it is shown that if the space is a normed topological space, then under some continuity conditions on the transition probabilities of $\{X_n\}$ the conditions for ergodicity will be met if there is a compact set K and an $\epsilon > 0$ such that whenever x lies outside K and is bounded, $x \in K$; whilst the conditions for recurrence will be met if there exists a compact K with for all x outside K.

A reviewed work by Clifford (1980) on voice identification by listeners using data obtained by experiment from a four-year research program. The reviewed work elucidated, support, and in some cases contradict published work on effects of voice identification on such factors as speech sample size and quality, voice disguise, delay in holding voice identification sessions, incidental as opposed to intentional memory for voices and effect of age of witness. He concluded that the caution and suspicion currently accorded to visual identification must be extended also, and perhaps more so, to voice identification.

Lonardo *et.al* (1984) introduced laboratory analysis of genetic markers in human blood which enables the calculation of an “index of grand paternity”. This index reflects the probability that a child shares genes with a specified set of grandparents because he is their

grandchild, although other genetic markers including blood groups, red cell enzymes, plasma proteins, and DNA polymorphism can be tested to provide adequate information in family with only one or two living grandparents.

Lichterberg and Powel (1984) presented three methods of discrete sequential analysis which appear to hold promise for the study of family interaction processes and delinquency. These are Markov Chain analysis, Lag Sequential analysis, and information theory. Because of the increasing instances of identity theft and terrorism incidences in past few years, biometrics based security system has been an area of quality research. Modern day biometrics is a cutting edge technology which enables the automated system to distinguish between a genuine person and an imposter. Automated face recognition is one of the areas of biometrics which is widely used because of the uniqueness of one human face to other human face. Automated face recognition has basically two parts one is face detection and other one is recognition of detected faces. To detect a face from an online surveillance system or an offline image, the main component that should be detected is the skin areas.

In studies of serial cancer markers or disease states and their relations to survival, data on the marker or state are usually obtained at infrequent time points during follow-up. Kay (1986) developed a Markov model to assess the dependence of risk of death on marker level or disease state and inferences within this model are based directly on data collected in the haphazard way.

Genetic abnormalities of the melanin pigment system in which the synthesis of melanin is reduced or absent are called albinism. Richard and Gail (1988) paper show that reduction

in melanin synthesis can involve the skin, hair follicle, and eye, resulting in oculocutaneous albinism, or can be localized primarily to the eye, resulting in ocular albinism.

Hordijk and Spieksma (1992) gave an overview of recurrence and ergodicity properties of a Markov chain. Two new notions for ergodicity and recurrence were introduced. They are called μ -geometric ergodicity and μ -geometric recurrence respectively. The first condition generalises geometric as well as strong ergodicity. Their key theorem shows that μ -geometric ergodicity is equivalent to weak μ -geometric recurrence. The latter condition is verified for the time-discretised two-centre open Jackson network. Hence, the corresponding two-dimensional Markov chain is μ -geometrically and geometrically ergodic, but not strongly ergodic. A consequence of μ -geometric ergodicity with μ of product-form was the convergence of the Laplace-Stieltjes transforms of the marginal distributions. Consequently all moments converge.

Saunders *et al* (1994) assessed the accuracy and reliability of 17 individual morphological traits of the pelvis frequently used to determine the sex of human skeletal remains. A sample of 49 right and left adult hip bones and sacra of documented individuals were available from an historic church cemetery dating from the 19th century. A hypothetical ranking of the accuracy of traits was drawn from the literature, individual traits were evaluated for precision and accuracy of observations, and combinations of two and three traits were evaluated for their collective effectiveness as sex indicators. The effect of age on the accuracy of traits for sex determination was also examined. Precision of traits was generally good. Several combinations of three criteria produced higher levels of accuracy than the trait list as a whole. A total of six traits were judged to be most effective as sex

discriminators because of low intra observer error levels and better than 83% accuracy rates. It was discovered that there was no indication of an age effect on the precision or accuracy of these traits although sample sizes are small.

Kulp *et.al* (1996) presented a statistical model of genes in DNA.A Generalized Hidden Markov Model (GHMM) provides the framework for describing the grammar of a legal parse of a DNA sequence. Probabilities were assigned to transitions between states in tile GHMM and to the generation of each nucleotide base given a particular state. Machine learning techniques were applied to optimize these probabilities using a standardized training set. Given a new candidate sequence, the best parse is deduced from the model using a dynamic programming algorithm to identify the path through the model with maximum probability.

Salzberg *et.al* (1998) described a new system, GLIMMER, for finding genes in microbial genomes. In a series of tests on Haemophilus influenza, Helicobacter pylori and other complete microbial genomes, this system has proven to be very accurate at locating virtually all the genes in these sequences, outperforming previous methods. GLIMMER uses Interpolated Markov models (IMMs) as a framework for capturing dependencies between nearby nucleotides in a DNA sequence. An IMM-based method makes predictions based on a variable context; i.e., a variable-length oligomer in a DNA sequence. The context used by GLIMMER changes depending on the local composition of the sequence. As a result, GLIMMER is more flexible and more powerful than fixed-order Markov

methods, which have previously been the primary content-based technique for finding genes in microbial DNA.

Lynn (2001) builds on the theory of human identification proposed by Professor Roger Clarke and uses the product as the basis for a proposed solution to the identity theft problem. The expanded theory holds that all human identification fits a single model. The identifior matches the characteristics of a person observed in a first observation with the characteristics of a person observed in a second observation to determine whether they are the same person. From the theory it follows that a characteristic used for identification in the credit reporting system, such as social security number, mother's maiden name and date of birth, must be known to all entities participating in that system. Because those characteristics - and any substitute for them - must be distributed so widely, it is unrealistic to think they can at the same time remain secret. Hence the current efforts to curb identity theft by keeping personal information secret are doomed to failure.

As an alternative solution to the identity theft problem, this paper proposes a system by which persons concerned about identity theft can register their identities through a government agency that will make their names, social security numbers, and non-sensitive contact information publicly available on an open-access website. Credit grantors and credit reporting agencies would have the option to contact the registrant to verify that he or she is in fact the credit applicant. Creditors who opted to use the system to identify a borrower would retain their current exemption from legal liability for misidentification. Those who did not would be liable for misidentification under common law principles, including

theories of defamation, invasion of privacy, and negligence. In cases in which credit grantors and credit reporting agencies used the system, the effect would be to give the individual person control over the process of his or her own identification in credit transactions, with no meaningful loss of privacy.

Modern methods of authorship attribution are reviewed. Despite the huge variety of methods, none of those described in either paper has been applied to a large number of texts. Often such methods require an element of human intervention which makes their application to large numbers of texts almost impossible. Khmelev *et al* (2001) presented a technique for authorship attribution, based on a simple Markov chain of letters (i.e., just letter bigrams are used). Many proposed methods of authorship attribution are illustrated on small examples. They show that this technique provides excellent results when applied to over 380 texts, as well as to two previously published data sets.

The large number of Web pages on many Web sites raised some navigational problems. Markov chains have recently been used to model user navigational behavior on the World Wide Web (WWW). Zhu *et.al* (2002) proposed a method for constructing a Markov model of a Web site based on past visitor behavior. The use of Markov model to make link predictions that can assist a new user to navigate the Web site was employed. An algorithm for transition probability matrix compression was used to cluster Web pages with similar transition behaviors and compress the transition matrix to an optimal size for efficient probability calculation in link prediction. A maximal forward path method was also used to further improve the efficiency of the link prediction.

The comparison of the means of two treatments or populations when more than one variable is measured may be done using Hotelling's T^2 statistic. In many real world situations the data obtained are dichotomous, and assumption of multivariate normality upon which Hotelling's T^2 is based is no longer valid. Singh *et.al* (2003) proposed an approximation of Hotelling's T^2 test for bivariate dichotomous data and empirically evaluated in term of Type 1 error rate.

Kimmel and Shamir (2005) presented a new stochastic model for genotype generation. The model offers a compromise between rigid block structure and no structure altogether: It reflects a general blocky structure of haplotypes, but also allows for "exchange" of haplotypes at non boundary SNPsites. It also accommodates rare haplotypes and mutations. They used a hidden Markov model and inferred its parameters by an expectation-maximization algorithm. The algorithm was implemented in a software package called HINT (haplotype inference tool) and tested on58 datasets of genotypes. To evaluate the utility of the model in association studies, they used biological human data to create a simple disease association search scenario. When comparing HINT to three other models, HINT predicted association most accurately.

Tangrot *et.al* (2006) proposed Family Identification of Sequence Homologues (FISH) using structure anchored hidden Markov models (saHMM) server which was highly accurate in identifying the family membership of domains in a query protein sequence, even in the case of very low sequence identities to known homologues. A performance test showed that 99.3% of the top hits are to the correct family saHMM. In addition, the FISH server allows

users to upload and search their own protein sequence collection or to query public protein sequence data bases with individual saHMMs.

One of the main goals of analyzing DNA sequences is to understand the temporal and positional information that specifies gene expression. An important step in this process is the recognition of gene expression regulatory elements. Experimental procedures for this are slow and costly. Anizova *et.al* (2006) presented a computational non-supervised algorithm that facilitates the process by statistically identifying the most likely regions within a putative regulatory sequence. A probabilistic technique is presented, based on the approximation of regulatory DNA with a Markov chain, for the location of putative transcription factor binding sites in a single stretch of DNA. Hereto developed a procedure to approximate the order of Markov model for a given DNA sequence that circumvents some of the prohibitive assumptions underlying Markov modeling. Application of the algorithm to data from 55 genes in five species shows the high sensitivity of this Markov search algorithm. The algorithm does not require any prior knowledge in the form of description or cross-genomic comparison. It is context sensitive and takes DNA heterogeneity into account.

In recent years, a very large variety of statistical methodologies, at various levels of complexity, have been put forward to analyze genotype data and detect genetic variations that may be responsible for increasing their susceptibility to diseases. Giovanni (2006) review provides a concise account of a number of selected statistical methods for

population-based association mapping, from single-marker tests of association to multi-marker data mining techniques for gene–gene interaction detection.

Genetic regulatory networks are families of biochemically interacting genes that regulate most functions of a living cell via the synthesis of proteins and other essential molecules. Cinquemani *et.al* (2008) using a stochastic hybrid approach introduce a piecewise deterministic model of genetic network and devise a systematic procedure for the identification of the model parameters from experimental observations of the protein concentration dynamics. Numerical results on simulated data were presented to show the effectiveness of the method.

In many cases human identification biometrics systems are motivated by real-life criminal and forensic applications. Some methods, such as fingerprinting and face recognition, proved to be very efficient in computer vision based human recognition systems. Michal (2008) focus on novel methods of human identification motivated by the forensic and criminal practice. His goal is to develop computer vision systems that would be used to identify humans on the basis of their lips, palm and ear images.

Brighenti *et.al* (2009) studied the input design problem for system identification where time domain constraints have to be considered. A finite Markov chain was used to model the input of the system. This allows to directly include input amplitude constraints in the input model, by properly choosing the state space of the Markov chain. The state space is so defined that the Markov chain will generate a binary sequence. The probability distribution of the Markov chain will be shaped in order to minimize the cost function considered in the input design problem. Stochastic approximation was used to minimize that cost function.

With this approach, the input signal to apply to the system can be easily generated by extracting samples from the optimal distribution.

Oculocutaneous albinism (OCA) is a genetically heterogeneous group of disorders characterized by the absence or reduced pigmentation of the skin, hair and eyes. Mohamed *et. al.* (2010) assess the clinic-epidemiologic features of different forms of OCA among Egyptian patients, they performed a retrospective study to determine the frequency, types,clinical presentation and associated genomic errors in albino patients and their relatives consulting the Genetics clinic. The clinical findings (clinical phenotypes) show that Oculocutaneous albinism was detected in one hundred and four (92.04%) patients from the one hundred and thirteen(113) patients observed with hair, skin and eyes colour affected. Ocular albinism (OA) was also detected which affected only the eyes. The epidemiological profile indicates that fifty-two (52) out of the one hundred and thirteen (113) patients (46.01%) had appositive family history of albinism.

TABLE 2.0 Summary table of clinico-epidemiological features of 113 patients with albinism

Clinico-epidemiological data	Complete OCA No. &%	Partial OCA No. &%	OA No. &%
Hair colour	55(100%)	47(100%)	
Skin colour	55(100%)		
Eye colour	55(100%)	47(100%)	2(100%)

A.F. Mohamed et al. (2010)

The clinical findings show that Oculocutaneous albinism was detected in one hundred and four (92.04%) patients and eyes colour affected all types (Table 2.0).

Gomez *et.al* (2010) evaluates the possibility of identifying people through their personality traits. The study was conducted using the answers of a population of 734 individuals to a

collection of 206 items. These items aimed at measuring five common different personality traits usually called the big five. These five traits are neuroticism, extraversion, agreeableness, conscientiousness and openness. The traits are estimated using the widely used Samejima's model, a non-laboratory procedure which is later used to discriminate the individuals.

Plomin and Daniels (2011) showed the important findings that emerged from human behavioral genetics which involves the environment rather than heredity, providing the best available evidence for the importance of environmental influences on personality, psychopathology and cognition. The research also converges on the remarkable conclusion that these environmental influences make two children in the same family as different from one another as are pairs of children selected randomly from the population.

Perrachione *et.al* (2011) tested voice-recognition abilities of dyslexic and control listeners for voices speaking listeners' native language or an unfamiliar language. Individuals with dyslexia exhibited impaired voice-recognition abilities compared with controls only for voices speaking their native language. Their results demonstrated the importance of linguistic representations for voice recognition.

Hopwood *et.al*(2011) investigate the patterns and origins of personality trait changes of some selected age intervals using three (3) waves of multidimensional personality Questionnaire data provided by twins. The results suggested that trait changes were more profound in first relative to the second half of the transition to adulthood and tend to be stable during the second half of the transition.

Glazner and Thompson (2011) presented a method for improving the power of linkage analysis by detecting chromosome segments shared by individuals not known to be related. Existing Markov chain Monte Carlo methods sample descent patterns on pedigrees condition on observed marker data was employed. These patterns can be stored as IBD graphs, which express shared ancestry only, rather than specific family relationships. A model for IBD between unrelated individuals allows the estimation of co ancestry between individuals in different pedigrees. IBD graphs on separate pedigrees can then be combined using these estimates. They show that when families share a gene for a trait due to shared ancestry on the order of tens of generations, the method can detect a linkage signal when independent analyses of the families do not.

Gilissen *et.al* (2012) reviewed work on disease gene identification strategies for exome sequencing done with next generation sequence to search for Mendelian disease genes in an unbiased manner by sequencing the entire protein. In the review, current strategies for Mendelian disease gene identification by exome resequencing were discussed. They concluded that exome strategies are successful and identify new Mendelian disease genes in approximately 60% of the projects. Improvements in bioinformatics as well as in sequencing technology will likely increase the success rate even further. They also established that exome sequencing is likely to become the most commonly used tool for Mendelian disease gene identification for the coming years.

Maji and Garg (2012) performed a mini review on Hidden Markov Models (HMMs) which is recently important and popular among bioinformatics researchers Large software tools

were based on this technique. The mathematical foundations of HMMs were considered first in a brief manner and then the gene identification application was considered later. In the case of gene identification process, HMMs basically resolve three problems: First was the evaluation problem, in which it computes the probability that a particular HMM will generate a given sequence of observations. Second is Decoding problem, in which it will uncover the most likely hidden state and Third is Learning problem, which is used to adjust the model parameter and train the HMM to find an optimal model.

Surbhi and Vishal (2012) presented a Face Identification Technique for Human Facial Image. Face is a primary focus of attention in social intercourse, playing a major role in conveying identity and emotion. Human face recognition plays an important role in many user authentication applications in the modern world. Facial expression recognition can be utilized for automated analysis of human emotion. The system commenced on convolving a face image after preprocessing the image at different scales and orientations. A given image of a face identifies or verifies the emotion of person in the scene using a stored database of facial image properties. Available collateral information such as race, age, gender, facial expression, or speech may be used in narrowing the search and enhance recognition. The solution to the problem involves detection of emotion from a given image, feature extraction from the face regions and recognition of emotion.

Correa *et.al* (2012) proposed a robust system for enabling robots to detect and identify humans in domestic environments. Robust human detection was achieved through the use of thermal and visual information sources that are integrated to detect human-candidate

objects, which are further processed in order to verify the presence of humans and their identity, using face information in the thermal and visual spectrums. Face detection is used to verify the presence of humans, and face recognition to identify them. Active vision mechanisms are employed in order to improve the relative pose of a candidate object/person in case direct identification is not possible.

Prediction of human behavior from his/her traits has long been sought by cognitive scientists. Human traits are often embedded in one's writings. Although some work has been done on identification of traits from essays, very little work can be found on extracting personality traits from written texts. Psychological studies suggest that extraction and prediction of rules from a data has been long pursued, and several methods have been proposed. Gupta and Chatterjee (2013) used Rough sets to extract the rules for prediction of personality traits. Rough Set is a comparatively recent method that has been effective in various fields such as medical, geological and other fields where intelligent decision making is required. Their experiments with rough sets in predicting personality traits produced encouraging results.

Bhavani and Rani(2013) combined biometric system based on texture of hand knuckles, namely finger-knuckle-print (FKP) and IRIS, to extract the image local texture information and represent FKPs features, by Linear Discriminate Analysis (LDA) and IRIS by the 2D Block based Gabor method. The results show that LDA is the best performance for identification of FKPs and Gabor is the best performance for IRIS and it is able to provide an excellent recognition rate and provide higher security.

Human identification is critical in today's society for enhancing public safety and privacy protection. Shape of human body is unique from one person to another. Rashid *et.al* (2013) presented an intelligent system approach for human identification at a distance using human body shape information. The body features used were the head, shoulder, and trunk. Image processing techniques for detection of these body features were developed in this work. Then, the features are recognized using the same fuzzy logic approach as inputs to a recognition system, based on a multilayer neural network. The developed system is only applicable for recognizing a person from his frontal view and specifically constrained to male gender to simplify the algorithm. In this research, the accuracy for human identification using the proposed method is 77.5%. Thus, it is proved that human can be identified at a distance using body shape information.

The perception of oneself as absorbed in the thoughts, feelings and happenings of a fictive character (e.g. in a novel or film) as if the character's experiences were one's own is referred to as identification. Cheetham *et.al* (2014) investigated whether individual variation in the personality trait of identification is associated with individual variation in the structure of specific brain regions, using surface and volume-based morphometry. The hypothesized regions of interest were selected on the basis of their functional role in subserving the cognitive processing domains considered important for identification (i.e. mental imagery, empathy, theory of mind and merging) and for the immersive experience called 'presence'. Controlling for age, sex, whole-brain volume and other traits, identification covaried significantly with the left hippocampal volume, cortical thickness in the right anterior insula and the left dorsal medial prefrontal cortex, and with gray matter

volume in the dorsolateral prefrontal cortex. These findings show that trait identification is associated with structural variation in specific brain regions. The findings are discussed in relation to the potential functional contribution of these regions to identification.

Cho et.al (2014) introduced a new characterization of the equality of two positive-definite matrices A and B, and used this to propose several new computationally convenient statistical tests for the equality of two unknown positive-definite matrices. Their primary focus is on testing the information matrix equality. They characterize the asymptotic behavior of new trace-determinant information matrix test statistics under the null and the alternative and investigate their finite-sample performance for a variety of models: linear regression, exponential duration, probit, and Tobit. This test can only be applied to positive-definite matrices (a symmetric and positive eigen values matrices).

Markov chain models at the cellular level fail to explain their underlying molecular mechanisms. Jia *et.al* (2014) proposed a nonlinear stochastic model of multistable bacterial systems at the molecular level. It turns out that the model not only provides a clear description of stochastic phenotype switching and bet-hedging within isogenic
bacterial populations, but also provides a deeper insight into the analysis of multidimensional experimental data. Moreover, they also used some deep mathematical theories to show that the proposed stochastic model and traditional Markov chain models are essentially consistent and reflect the dynamic behavior of the bacterial system at two different time scales.

Speaker identification is a biometric process. The objective of speaker identification is to extract, characterize and recognize the information about speaker identity. Speaker Recognition technology has recently been used successfully in a vast number of commercial areas such as in voice based biometrics; voice controlled appliances, security control for confidential information, remote access to computers and many more interesting areas. Alam *et.al* (2014) used Hidden Markov Model (HMM) method as the training/recognition algorithm which makes the final decision about the specification of the speaker by comparing unknown features to all models in the database and selecting the best matching model. i, e. the highest scored model. The speaker who obtains the highest score is selected as the target speaker.

The problem of unsupervised family member discovery given a collection of family photos; infer the size of the family, as well as the visual appearance and social role of each family member was addressed by Dai *et. al*. (2015) using a non-laboratory test. As a result, they were able to recognize the same individual across many different photos. They proposed an unsupervised EM-style joint inference algorithm with a probabilistic approach that models, identity and role assignments for all detected faces, along with associated pair wise relationships between them. Their experiments illustrate how joint inference of both identity and role (across all photos simultaneously) outperform independent estimates of each. The joint inference also improves the ability to recognize the same individual across many different photos.

Kanchan (2015) proposes a skin based segmentation algorithm for face detection in colour images with detection of multiple faces and skin regions. Skin colour has proven to be a

useful and robust cue for face detection, localization and tracking. Biometric system has been actively emerging in various industries for the past few years, and it is continuing to roll to provide higher security features for access control system. In the recent years, hand based biometrics is extensively used for personal recognition.

Identification of Personality is a complex process. Personality traits are stable over time .Individual's behavior naturally varies from occasion to occasion. But there is a core consistency which defines the true nature. Asra and Shubhangi's(2015) paper addressed this issue of behavior using Graphology a technique used to identify personality Trait using Unconstrained Cursive and Mood Invariant Handwritten Text. Accuracy of this technique depends on how skilled the analyst is. Although human intervention in handwriting analysis has been effective, it is costly and prone to fatigue. An automation of handwritten text was proposed and it basically considered three important features in the direction of orientation of the lines: (i) up hill (ii) down hill (iii) constant line.

The structure of blood vessels in the sclera- the white part of the human eye, is unique for every individual, hence it is best suited for human identification. However, this is a challenging research because it has a high insult rate (the number of occasions the valid user is rejected). Vanita and Patil(2015) proposed simplified method for sclera segmentation, a new method for sclera pattern enhancement based on histogram equalization and line descriptor based feature extraction and pattern matching with the help of matching score between the two segment descriptors. They attempt to increase the awareness about this topic, as much of the research is not done in this area.

Voice Identification System refers to a system which comprises hardware, software and it is used to identify voice for several applications. Das *et.al* (2016) developed a system based on the technique of Hidden Markov Model. The Hidden Markov Model is a stochastic approach which models the algorithm as a double stochastic process in which the observed data is thought to be the result of having passed a hidden process through second process. A database of voice information was created. To extract features from voice signals, Mel-Frequency Cepstral Coefficients (MFCC) technique has been applied to produce a set of feature vectors. Subsequently, the system uses The Vector Quantization (VQ) for features training and classification. Speech recognition based on Hidden Markov Model achieved successfully for the conversion of speech to text.

Olarinoye and Jolayemi (2016) presented a paper on use of transition probability matrix in studying traits. The paper show how traits can be model into stochastic process and also show the various states of the traits with the use of transition probability matrix. Dimosthenis *et.al* (2017) presented a large-scale heritability study of face geometry that aims to address these issues. High-resolution, three-dimensional facial models have been acquired on a cohort of 952 twins recruited from the Twins UK registry, and processed through a novel land marking workflow, GESSA (Geodesic Ensemble Surface Sampling Algorithm). The algorithm places thousands of landmarks throughout the facial surface and automatically establishes point-wise correspondence across faces. These landmarks enabled them to intuitively characterize facial geometry at a fine level of detail through curvature measurements, yielding accurate heritability maps of the human face.

Hajj visitors (known as pilgrims) arrive yearly to the holy city of Makkah to perform Hajj rituals. According to reports, more than ten million pilgrims (foreign and local) arrive to the holy city of Makkah every year to perform Hajj and Umrah. Due to the large crowd gathering in the two holy cities of Makkah and Medina, the issues arise of identifying the missing, dead and found people, and collection and distribution of missing and found objects. Salah and Adnan (2017) presented a paper on an integrated recognition system for identifying missing-and- found objects as well as missing, dead and found people during Muslim religious Hajj and Umrah seasons. This system can be deployed in the two holy cities of Makkah and Medina in the Kingdom of Saudi Arabia. The proposed system integrates facial recognition and object identification solutions during the Hajj and Umrah rituals. The missing-and-found computerised system is part of the proposed crowd-sensing system for Hajj and Umrah crowd estimation, management and safety.

Prediction of human physical traits and demographic information from genomic data challenges privacy and data deidentification in personalized medicine. To explore the current capabilities of phenotype-based genomic identification, Venter *et.al* (2017) applied whole-genome sequencing, detailed phenotyping, and statistical modeling to predict biometric traits in a cohort of 1,061 participants of diverse ancestry. Individually, for a large fraction of the traits, their predictive accuracy beyond ancestry and demographic information is limited. However, they developed a maximum entropy algorithm that integrates multiple predictions to determine which genomic samples and phenotype measurements originate from the same person. This work challenges current conceptions of personal privacy and may have far-reaching ethical and legal implications.

Shahin(2017) centres on enhancing the performance of text-dependent and speaker dependent talking condition identification systems using second-order hidden Markov models (HMM2s). His results show that the talking condition identification performance based on HMM2s has been improved significantly compared to first-order hidden Markov models (HMM1s). The talking conditions in the work are neutral, shouted, loud, angry, happy, and fear.

Human identification is critical in today's society for enhanced safety and privacy. Olarinoye and Jolayemi (2017) presented a paper on Albinism markers using Markov Chain where each marker(s) and their behavioral state pattern was observed.

The idea that beauty is hereditary seems commonsensical, considering that other facial features have been shown to be affected by genes. Sasaki *et.al* (2018) presented both the results of an experiment on the heredity of facial ratios and discussed the impact of these ratios on real world. Thus, facial ratios were used as a measure of attractiveness. The facial ratio of 19 families, each containing six family members was calculated by making six measurements for each face and creating an average facial ratio (AFR) for each individual. The results were then statistically analyzed using an Analysis of Variance test (ANOVA). The test suggested that there was a significant difference between the facial ratio of family members and non-family members. The result also suggested that facial ratio and thus attractiveness are hereditary. The human face is a complex trait under strong genetic control, as evidenced by the striking visual similarity between twins. Nevertheless, heritability estimates of facial traits have often been surprisingly low or difficult to

replicate. Furthermore, the construction of facial phenotypes that correspond to naturally perceive facial features remains largely a mystery.

In all the above literature, the use of non-laboratory traditional method of identification using pictures, the use of face identification techniques using human emotion, also the study of shape using distance and finally the use of deoxyribonucleic acid (DNA) test were the available methods employed.

In this research work, a probabilistic method (Markov Chain) which is flexible, accessible and less cost alternative test (non-laboratory) procedure was introduced in the identification of a Trait marker.

CHAPTER THREE
METHODOLOGY

Mendelian gene inheritance distribution is normally considered in terms of Markov models and the trait is controlled by a single locus in an inheritance pattern. In such cases, a mutation in a single gene can cause a disease that is inherited according to Mendel's laws.

The combination of genes that are inherited from parents contributes to our physical and psychological traits. Defects in one or more genes may cause or predispose to disease. In man, many of inherited diseases are transmitted by recessive genes.

A trait, or simply phenotypic trait, is a distinct variant of a phenotypic characteristic of an organism; it may be either inherited or determined environmentally, but typically occurs as a combination of the two. A phenotypic trait is an obvious and observable trait; it is the expression of genes in an observable way.

A trait could be individual characteristic or inherited characteristic that is genetically determined. The assumption here is that a trait follows hereditary links and can be of k types. In a generation, a step then ($X_{n+1} = j \mid X_n = i$) is a possible transition from state i at n^{th} generation to state j at the next generation, where states i and j are any of the k types. Consider a trait, where trait is observed as either present or absent in an individual in a given generation, there is a need to determine the presence of traits and therefore link the family members.

3.1 MARKOV CHAIN

In studying a system that can change over time, there is need to keep track of those changes. A Markov chain is a particular model for keeping track of systems that change according to given probabilities. As it will be seen, a Markov chain allows one to predict future events, but the predictions become less useful for events farther away from the current into the future. It can thus be used for describing systems that follow a chain of linked events, where what happens next depends only on the current state of the system. Markov chains are a fairly common, and relatively simple, way to statistically model random processes.

Markov chains are a fundamental part of stochastic processes. They are used widely in many different disciplines. A Markov chain is a stochastic process that satisfies the Markov property, which means that the past and future are independent when the present is known. This means that if one knows the current state of the process, then no additional information of its past states is required to make the best possible prediction of its future. This simplicity allows for great reduction of the number of parameters when studying such a process.

Markov chain is a sequence of random variables such that for any n, X_{n+1} is conditionally independent of $X_0, \ldots, X_{n-1}$ given X_n. That is, the "next" state X_{n+1} of the process is independent of the "past" states $X_0, \ldots, X_{n-1}$ provided that the "present" state X_n be known. It is required to possess the property that is usually characterized as being "memoryless": the probability distribution of the next state depends only on the current state and not on the sequence of events that precedes it. This specific kind of "memorylessness" is called the Markov property. Markov chains have many applications as statistical models of real-world processes. The stochastic process $X=[X_n;\ n \in N]$ is called a Markov chain provided that $P[X_{n+1} = j \mid X_0, \ldots, X_n] = P[X_{n+1} = j \mid X_n = i]$ for all $j \in E$ and $n \in N$.

Model

Formally, a Markov chain is a probabilistic automaton. The probability distribution of state transitions is typically represented as the Markov chain's *transition matrix.* If the Markov chain has K possible states, the matrix will be an K x K matrix, such that entry (i, j) is the probability of transitioning from state i to state j. Additionally, the transition matrix must be a stochastic matrix, a matrix whose entries in each row must add up to exactly 1. This makes complete sense, since each row represents its own probability distribution. A Markov Model is a stochastic model that models random variables in such a manner that the variables follow the Markov property.

3.2 BUILDING MARKOV MODEL FROM GENERATIONS

The need to construct a link structure that represents family generations and the structure is then used to build a Markov model. Let **P** and **A** implies the Present or Absent of a Trait. See fig 3.1

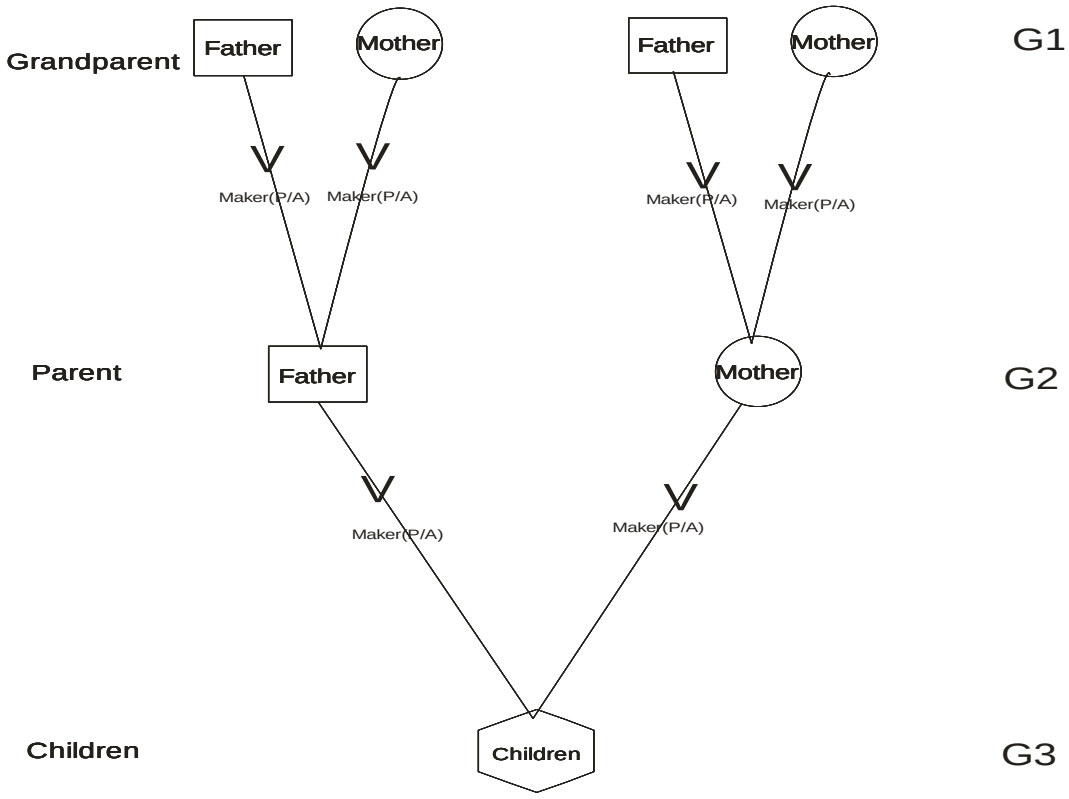


Figure3.1 Inheritance of Traits over three generations

Each row on the link graph is viewed as a generation and with the likely state of Trait (Present /Absent), where the state space containing all the rows likely states, P_{ij} is the probability transition matrix containing one-step transition probability between rows. The link in the generations can be seen as a stochastic process $[X_n]$, which has S as the collection state space.

If the conditional probability of linking generation j together in the next step is dependent only on the last generation {m}, is called a m-order Markov chain.

$$P_{ij}^{\ m} = P[x_{n+m} = j \mid x_0 = i_o, \ldots, x_n, x_{n+1}, \ldots x_{n+m-1} = i] \qquad 3.1$$

$$p_{ij} = p(x_{n+1} = j \mid x_0 = i_0, \ldots, x_n = i) = p(x_{n+1} = j \mid x_n = i) \qquad 3.2$$

When m=1, X_{n+1} is dependent only on the current state X_n and is a one-order Markov chain, where P_{ij} is the probability that transition is made from i to state j in one step. Let n_{ij} be the observation matrix. The initial state X_o has an arbitrary probability distribution.

Let n_{ij} be the observation starting from state i and move to j, where $P_{ij} = \frac{n_{ij}}{n_{i.}}$,

$$n_{i.} = \sum_j n_{ij}$$

$$n_{cxc} = \begin{bmatrix} n_{11} & n_{12} & - & - & - & n_{1c} \\ n_{21} & n_{22} & - & - & - & n_{2c} \\ - & - & - & - & - & - \\ n_{i1} & n_{i2} & - & n_{ij} & - & n_{ic} \\ - & - & - & - & - & - \\ n_{c1} & n_{c2} & - & - & - & n_{cc} \end{bmatrix} \qquad 3.3$$

$$p_{ij}= \begin{bmatrix} p_{11} & p_{12} & - & - & - & p_{1c} \\ p_{21} & p_{22} & - & - & - & p_{2c} \\ - & - & - & - & - & - \\ p_{i1} & p_{i2} & - & p_{ij} & - & p_{ic} \\ - & - & - & - & - & - \\ p_{c1} & p_{c2} & - & - & - & p_{cc} \end{bmatrix} \quad - \quad - \quad - \quad - \quad - \quad - \quad - \quad 3.4$$

From Figure3.1, let P_{ij} be the transition probability matrix of Grandparent to Parent and also C_{ij} be the transition probability matrix of Parent to Children.

Generally,

$$p_{ij} = P_{ij}(x_n = j \mid x_{n-1} = i) \text{ and } c_{ij} = C_{ij}(x_{n+1} = j \mid x_n = i)$$

$$P_{ij} = \begin{bmatrix} p_{11} & p_{12} & - & - & p_{1c} \\ p_{21} & p_{22} & - & - & p_{2c} \\ - & - & - & - & - \\ - & - & - & - & - \\ p_{c1} & p_{c2} & - & - & p_{cc} \end{bmatrix} and \qquad C_{ij} = \begin{bmatrix} c_{11} & c_{12} & - & - & c_{1c} \\ c_{21} & c_{22} & - & - & c_{2c} \\ - & - & - & - & - \\ - & - & - & - & - \\ c_{c1} & c_{c2} & - & - & c_{cc} \end{bmatrix} \quad 3.5$$

P_{ij} and C_{ij} are the transition probabilities, the probabilities of transition from i to j

$$P_{ij} = \begin{bmatrix} p_{00} & p_{01} & p_{02} & - & - \\ p_{10} & p_{11} & p_{12} & - & - \\ - & - & - & - & - \\ p_{i0} & p_{i1} & p_{i2} & - & - \\ - & - & - & - & - \end{bmatrix}, \quad P^2{}_{ij} = \begin{bmatrix} p_{00} & p_{01} & p_{02} & - & - \\ p_{10} & p_{11} & p_{12} & - & - \\ - & - & - & - & - \\ p_{i0} & p_{i1} & p_{i2} & - & - \\ - & - & - & - & - \end{bmatrix} \begin{bmatrix} p_{00} & p_{01} & p_{02} & - & - \\ p_{10} & p_{11} & p_{12} & - & - \\ - & - & - & - & - \\ p_{i0} & p_{i1} & p_{i2} & - & - \\ - & - & - & - & - \end{bmatrix} 3.6$$

More generally, define the n-step transition probabilities

$$P^n{}_{ij} = P[X_n = j \mid X_0 = i], \quad \text{for n=0,1,2,..,} \qquad 3.7$$

and the n-step transition matrix

$$P^n{}_{ij} = \begin{bmatrix} p^n{}_{00} & p^n{}_{01} & p^n{}_{02} & - & - \\ p^n{}_{10} & p^n{}_{11} & p^n{}_{12} & - & - \\ - & - & - & - & - \\ p^n{}_{i0} & p^n{}_{i1} & p^n{}_{i2} & - & - \\ - & - & - & - & - \end{bmatrix} \qquad 3.8$$

(3.7) can now be generalized.

Chapman-Kolmogorov Equation

Let m and n be two positive integers and assume X_0=i. In order to get to state j in (m+n) steps, the chain will be at some intermediate state k after m steps. To obtain $P_{ij}^{(m+n)}$, we sum over all possible intermediate states

$$P_{ij}{}^{m+n} = P[X_{m+n} = j \mid X_0 = i]$$
$$= \sum_{k \in S} P_{ik}{}^{(m)} P_{kj}{}^{(n)} \qquad 3.9$$

The above equation is called the Chapman-Kolmogorov Equation.

Definition 1. A state is said to be recurrent if, any time that the process leaves the state, it returns to that state in future with probability one.

Definition 2. State i and j communicate if they are accessible from each other, written as

i ⟷ j

Definition 3. Ergodic State is a state that is positively recurrent and aperiodic

Definition 4. A Markov Chain is ergodic if all states are ergodic.

Definition 5. A Markov Chain is irreducible if all the states communicate with each other (there is only one class)

Lemma 1 For an irreducible ergodic Markov chain, $\lim_{n\to\infty} P^n{}_{ij}$ exists and is independent of i.

Let $\pi_j = \lim_{n\to\infty} P^n{}_{ij}$, j ≥ 0, then π_j is the unique nonnegative solution of

$$\pi_j = \sum_{i=0}^{\infty} \pi_i P_{ij} \quad , \quad j \geq 0 \qquad 3.10$$

$$\sum_{j=o}^{\infty} \pi_j = 1 \qquad 3.11$$

in which π_j denotes the long run proportion of time that the chain spends in state j:

if and only if the Markov chain is positively recurrent.
If a solution exist then

1. it will be unique, and
2.

$$\pi_j = \begin{cases} \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} p_{ij}^{(k)} & \text{if the chain is periodic} \\ \lim_{n\to\infty} p^n{}_{ij} & \text{if the chain is aperiodic} \end{cases}$$

Definition 6. State i is aperiodic if it can only return to itself with transition number equal one and periodic if it can only return to itself with transition number greater than one.

Definition 7. Probability that the process returns to state i (eventually) given that it starts from i is f_i

Definition 8. A Positive recurrent state is defined by f_i =1 and $\mu_i < \infty$ (expected time until the process returns to state i is finite)

Definition 9. A Null recurrent state is defined by f_i =1 and $\mu_i = \infty$ (expected time until the process returns to state i is infinite)

Definition 10. The first –passage time probability, $f_{ij}(n)$, of a Markov chain is the probability, condition on $X_0 = i$, that the first subsequent entry to state j occurs at discrete epoch n. That is, $f_{ij}(1) = P_{ij}$ and for n≥2,

$$f_{ij}^{(n)} = \Pr\{X_n = j, X_{n-1} \neq j, X_{n-2} \neq j, ..., X_1 \neq j / X_0 = i\} \qquad 3.12$$

The distinction between $f_{ij}^{(n)}$ and $P_{ij}^{n} = \Pr\{X_n = j / X_0 = i\}$ is that $f_{ij}^{(n)}$ is the probability that the first entry to j (after time 0) occurs at time n, whereas P_{ij}^{n} is the probability that any entry to j occurs at time n, both conditional on starting in state i at time 0. The definition 3.12 also applies for j=i.

Thus

$$f_{ii}^{1} = p_{ii}^{1}$$
$$f_{ii}^{2} = p_{ii}^{2} - f_{ii}^{1} p_{ii}^{1}$$
$$f_{ii}^{3} = p_{ii}^{3} - f_{ii}^{1} p_{ii}^{2} - f_{ii}^{2} p_{ii}^{1}$$
$$--------$$
$$--------$$
$$f_{ii}^{n} = p_{ii}^{n} - \sum_{k=1}^{n-k} f_{ii}^{k} p_{ii}^{n-k} \qquad 3.13$$

To calculate f_{ii}^{n} for n=1,2,…, we need to have the corresponding values of p_{ii}^{n} according to equation (3.13)

Let $F_{ij}^{(n)}$, for n ≥ 1, be the probability, given $X_0 = i$, that state j occurs at some time between 1 and n inclusive. Thus,

$$F_{ij}^{(n)} = \sum_{m=1}^{n} f_{ij}^{(m)} \qquad 3.14$$

Definition11. The mean recurrent time at state i is the expected return time μ_i

$$\mu_i = E(Ti) = \sum_{n=1}^{\infty} n f_{ii} \qquad 3.15$$

state i is Positively recurrent(non null Persistent) if μ_i is finite and

f_{ii} = Probability that the process returns to state i (eventually) given that it starts from i.

$f_{ii}^{n} = p_{ii}$ if n=1, were f_{ii} is the return times and p_{ii} return probabilities

Connection between Mean return times and stationary distribution

$$\lim_{n\to\infty} P^{n}_{ii} = \frac{1}{\sum_{n=0}^{\infty} n f_{ii}^{(n)}} = \frac{1}{\mu_i} \qquad 3.16$$

μ_i =mean return times

Recall, every finite irreducible Markov Chain has a unique stationary distribution and if the chain is aperiodic then

$\lim_{n\to\infty} P^{n}_{ii} = \pi_i$ for each i 3.17

So, for a finite irreducible aperiodic Markov Chain

$$\pi_i = \frac{1}{\mu_i},$$

So it is possible to find the mean return times from the stationary distribution.

e.g

$$P = \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix},$$ (with a,b ≠0 0r 1)

Stationary distribution is

$$\pi = \begin{bmatrix} \pi_0 & \pi_1 \end{bmatrix} = \begin{bmatrix} \frac{b}{a+b} & \frac{a}{a+b} \end{bmatrix}$$

$$\Rightarrow \mu_0 = \frac{a+b}{b}, \mu_1 = \frac{a+b}{a}$$

Recurrent State property

Positive recurrent and Null recurrent states are called distinguishing property of recurrent states. Positive recurrent state implies that when the system leaves state i, it is certain to return eventually to i; however, if i is null recurrent then expected time to re-visit is infinite.
From def. 1, we define recurrent state as:

f_{ii} = P (ever reenter i|X_0=i)

=1

but in order never to return to i, we need to go to state j, j≠i and stay there forever. We stay at j for n steps with probability

$(p_{ij})^n \to 0,$ i≠j 3.18

as n→∞, so the probability of staying at j forever is 0 and consequently $f_{ii}=1$

Computations of Recurrent State property

If the probability of eventually visiting state j given that we start in i is 1(need to take at least one step) then the expected number of step until the first visit j is given by

$$\mu_{ij} = 1 + \sum_{r \neq j} p_{ir}\mu_{rj}, \quad 3.19$$

for i= 0,1 . . .,m-1, m is number of equations

Table 3.0(Recurrent Condition)

Recurrent Property	
Positive recurrent	$\mu_{ij} < \infty$
Null recurrent	$\mu_{ij} = \infty$

Recurrence is define on a bases of mean return (μ_{ij}) which may be finite or infinite

Transforming into Markov Chain

Let X_n be the appearance of a trait in generation n and suppose that a trait appears in next generation(Children) depends on whether it appears in current generation(father) and past generations (Grandparent) then we have

$$P(X_{n+1} \mid X_{n,} X_{n-1,}.....,X_{1)} = P(X_{n+1} \mid X_n) \quad 3.20$$

3.3 Methods of Data Analysis

The methods of data analysis are outlined in the following sections in view of the objectives stated in chapter one.

3.3.1 State spaces and classification to probability transition matrix.

A Markov chain may be visualized as a process which moves from one state to another, as time progresses. We will develop this method with respect to two systems of inheritance, with respect to phenotypes.

Table 3.1 Kind of Mating

Mating type	Kind of mating with respect to phenotype
1	Father = Yes ,Mother = Yes
	Father = N0, Mother = N0
2	Father = Yes, Mother = N0
	Father = N0, Mother = Yes

Thus, in any generation there are two mating types: mating type 1 and mating type 2. Let S_1 denote Yes (Present) and S_2 denote No (Absent).

Let p_{ij} denote the probability of transition from state i in generation n-1 to state j in generation n and c_{ij} denote transitional probability from state i in generation n to state j in generation n+1, each transitional probability is a conditional probability which takes on the following values for a 2x2 matrix,

p_{11} = p(S_1 in generation n| S_1 in generation n-1)

p_{12} = p(S_2 in generation n | S_1 in generation n-1)

p_{21} = p(S_1 in generation n | S_2 in generation n-1)

p_{22} = p(S_2 in generation n | S_2 in generation n-1)

$$(0 \leq p_{ij} \leq 1) \qquad 3.21$$

c_{11} = c(S_1 in generation n+1 | S_1 in generation n)

c_{12} = c(S_2 in generation n+1 | S_1 in generation n)

c_{21} = c(S_1 in generation n+1 | S_2 in generation n)

c_{22} = c(S_2 in generation n+1 | S_2 in generation n)

where S_1= Trait Present and S_2=Trait Absent

$$(0 \leq c_{ij} \leq 1) \qquad 3.22$$

These p_{ij} 's and c_{ij}'s can be arranged in a matrix form

$$P = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}$$

Matrix P and C are known as transition matrices or sometimes more specifically as stochastic matrices. The transitional probabilities in each row sum to one, i.e.

$$\sum_{j=1}^{2} p_{ij} = 1 \text{ for i=1,2} \qquad 3.23$$

$$\sum_{j=1}^{2} c_{ij} = 1 \text{ for i=1, 2} \qquad 3.24$$

$P_{s_1 s_1}$ = P(S_1 in generation n | S_1 in generation n -1)
$P_{s_1 s_2}$ = P(S_2 in generation n |S_1 in generation n -1)
$P_{s_2 s_1}$ = P(S_1 in generation n | S_2 in generation n-1)
$P_{s_2 s_2}$ = P(S_2 in generation n | S_2 in generation n -1)

Generation n

P= generation n – 1 $\begin{bmatrix} p_{s_1 s_1} & p_{s_1 s_2} \\ p_{s_2 s_1} & p_{s_2 s_2} \end{bmatrix}$

Generation n+1

C= generation n $\begin{bmatrix} c_{s_1 s_1} & c_{s_1 s_2} \\ c_{s_2 s_1} & c_{s_2 s_2} \end{bmatrix}$

where P and C are the probabilities of being in state j given that you are in state i with the properties 3.21, 3.23 and 3.22, 3.24

3.3.2 Link generations together.

Similarity between distribution of generation (X_n) and generation (X_{n+1})

Let the generations ($X_0, X_1, X_2, \ldots X_n$) be a Markov Chain with N x N transition matrix P. If the probability distribution of X_n is given by the 1xN row vector λ^T ($X_0 \sim \lambda^T$) then the probability distribution of X_{n+1}

$X_n \sim \lambda^T \Rightarrow \quad X_{n+1} \sim \lambda^T P = \lambda^T$

Then

$X_n \sim \lambda^T$

$\Rightarrow \quad X_{n+1} \sim \lambda^T P \quad = \quad \lambda^T$

$\Rightarrow \quad X_{n+2} \sim \lambda^T P \quad = \quad \lambda^T$

$\Rightarrow \quad X_{n+3} \sim \lambda^T P \quad = \quad \lambda^T$

$\Rightarrow \ldots$

In other words, if $\lambda^T P \quad = \quad \lambda^T$, and $X_n \sim \lambda^T$ then

$X_n \sim X_{n+1} \sim X_{n+2} \sim X_{n+3} \sim \ldots$

If $\lambda^T P \quad = \quad \lambda^T$, we say the distribution λ^T is an equilibrium distribution, implies the generation distributions of $X_n \sim X_{n+1} \sim X_{n+2} \sim X_{n+3} \sim \ldots$ are the same. Equilibrium means level position does not imply the value of $X_n \sim X_{n+1} \sim X_{n+2} \sim X_{n+3} \sim \ldots$ .are equal.

3.3.3 Equality of Matrices

In testing the equality of two transition probability matrices using the row or the column vectors of the matrices, we consider the problem of comparing two transition probability matrices with r by c dimensions(r=c). Let the transition probability matrices be

$$P_{ij} = \begin{bmatrix} p_{11} & p_{12} & - & - & p_{1c} \\ p_{21} & p_{22} & - & - & p_{2c} \\ - & - & - & - & - \\ - & - & - & - & - \\ p_{c1} & p_{c2} & - & - & p_{cc} \end{bmatrix} and \quad C_{ij} = \begin{bmatrix} c_{11} & c_{12} & - & - & c_{1c} \\ c_{21} & c_{22} & - & - & c_{2c} \\ - & - & - & - & - \\ - & - & - & - & - \\ c_{c1} & c_{c2} & - & - & c_{cc} \end{bmatrix} \qquad 3.25$$

If the system is consistent, let b_{ij} be a Markov Chain and $b_{ij} = P^2_{ij}$ and for this study c=2.

Testing hypothesis about the difference between two population proportion

i. Testing difference between cells i.e.

$H_0 : b_{11} = c_{11}$ vs $H_1 : b_{11} \neq c_{11}$ $H_0 : b_{21} = c_{21}$ vs $H_1 : b_{21} \neq c_{21}$

$$Z = \frac{C_{11} - b_{11}}{\sqrt{C_{11}(1-C_{11})/n_1 + b_{11}(1-b_{11})/n_{2.}}} \qquad Z = \frac{C_{21} - b_{21}}{\sqrt{C_{21}(1-C_{21})/n_1 + b_{21}(1-b_{21})/n_{2.}}}$$

Which is approximately N (0, 1)

ii. Testing difference between vectors.

vectors equality test for the two Markov chains needs to be carried out i.e.

$$H_0 : \begin{bmatrix} b_{11} \\ b_{21} \end{bmatrix} = \begin{bmatrix} c_{11} \\ c_{21} \end{bmatrix} \text{vs } H_1 = \begin{bmatrix} b_{11} \\ b_{21} \end{bmatrix} \neq \begin{bmatrix} c_{11} \\ c_{21} \end{bmatrix}$$

$$V(b_{11} - c_{11}) = V(b_{11}) + V(c_{11})$$

$$V(b_{21} - c_{21}) = V(b_{21}) + V(c_{21})$$

$$V(b_{ij}) = S^2{}_{bij} = \begin{bmatrix} \frac{b_{11}(1-b_{11})}{n_{1.}} & 0 \\ 0 & \frac{b_{22}(1-b_{22})}{n_{2.}} \end{bmatrix}, \quad V(c_{ij}) = S^2{}_{cij} = \begin{bmatrix} \frac{c_{11}(1-c_{11})}{n_{1.}} & 0 \\ 0 & \frac{c_{22}(1-c_{22})}{n_{2.}} \end{bmatrix}$$

and estimated variance–covariance matrix defined by

$$S^2 = V(b_{ij}) + V(c_{ij})$$

The test statistic for the hypothesis for large sample is given by

$$Q = \begin{pmatrix} b_{11} - c_{11} \\ b_{21} - c_{21} \end{pmatrix}^T S^{-2} \begin{pmatrix} b_{11} - c_{11} \\ b_{21} - c_{21} \end{pmatrix} \qquad 3.26$$

Approximately χ_c^2 we carry out a single test and we reject H_0 if Q> χ_c^2

3.3.4 Identify a proper marker

In identifying a proper maker, the marker must reflect the status of the clinical manifestations in patients, the strength of association between the marker and traits and also the reoccurrence behavior.

In doing this, we try to have some deep insight into Markov chain and look into the behavior of chain.

Absorbing Markov Chains

State i of a Markov chain is an absorbing state if $P_{ii}=1$

Using the idea of an absorbing state, we can define an absorbing Markov chain.

Definition1. A Markov chain is an absorbing chain if and only if the following two conditions are satisfied:

i. the chain has at least one absorbing state; and

ii. It is possible to go from any non-absorbing state to an absorbing state (perhaps in more than one step).

Definition2. In a absorbing Markov chain, a state which is not absorbing called transient state

Canonical form

To obtain the canonical form of absorbing Markov chain

- renumber the states so that the transient states come first.
- if there are r absorbing states and t-transient states the transition matrix will have the following canonical form

$$P= \begin{array}{c} \\ TR \\ \\ ABS \end{array} \begin{array}{c} \begin{array}{cc} TR & ABS \end{array} \\ \left(\begin{array}{c|c} Q & R \\ \hline 0 & I \end{array}\right) \end{array}$$

where Q is a t-by-t matrix, R is a nonzero t-by-r matrix, $\mathbf{0}$ is an r-by-t zero matrix, and I_r is the r-by-r identity matrix. Thus, Q describes the probability of transitioning from some transient state to another while R describes the probability of transitioning from some transient state to some absorbing state.

The P^n_{ij} is the probability of being in a state S_j after n-steps when the chain started in state S_i

$$P^n_{ij}= \begin{array}{c} \\ TR \\ \\ ABS \end{array} \begin{array}{c} \begin{array}{cc} TR & ABS \end{array} \\ \begin{pmatrix} Q^n & R^n \\ 0 & I \end{pmatrix} \end{array}$$

Definition 3. In Absorption Markov chain, the probability that process will be absorbed is 1 (i.e$Q^n \longrightarrow 0$ as $n \longrightarrow \infty$)

Definition4. if i is transient and j absorbing , j can be reached from i but not vice vasa (not communicate)

Transient and Recurrent Markov Chains

In any Markov chain, we define

f_i= P(Eventually return to state i|x_0=i) = P (X_n= i for some n ≥ 1|X_0 = i).

If f_i = 1, then we say that state i is recurrent. Otherwise, if f_i < 1,then we say that state i is transient. In Markov chain, every state must be either transient or recurrent. There is a pretty important applied reason why we care whether a state is transient or recurrent. Whether a state is transient or recurrent determines the kinds of questions we ask about the process. Also limiting probabilities $\lim_{n\to\infty} P_{ij}(n)$ tell us about the "long-run" or "steady-state" behavior of the process, and give us a way to predict how the system our process is modeling will tend to behave.

Canonical form for Transient and Recurrent Markov Chains

To obtain the canonical form of Transient and Recurrent Markov chain

- renumber the states so that the transient states come first.

  we have the following canonical form

$$P_{ij} = \begin{array}{c} \\ TR \\ Rc \end{array} \begin{array}{c} \begin{array}{cc} TR & Rc \end{array} \\ \left(\begin{array}{c|c} Q & L \\ \hline 0 & K \end{array}\right) \end{array}$$

  where *Q* is a *t*-by-*t* matrix, *L* is a nonzero *t*-by-*r* matrix, **0** is an *r*-by-*t* zero matrix, and *K* is the *r*-by-*r* matrix. Thus, *Q* describes the probability of transitioning from some transient state to another while *L* describes the probability of transitioning from some transient state to some recurrent state and K describes the probability of transitioning from some recurrent state to another.

  The P^n_{ij} is the probability of being in a state S_j after n-steps when the chain started in

state S_i

$$P^n_{ij} = \begin{array}{c} \\ TR \\ Rc \end{array} \begin{array}{c} \begin{array}{cc} TR & Rc \end{array} \\ \begin{pmatrix} Q^n & R \\ 0 & K \end{pmatrix} \end{array}$$

Definition 5. If i is recurrent and j is recurrent , then j communicate i

Definition 6. If i is transient and j is recurrent, j can be reached from i but not vice vasa (not communicate)

Definition 7. If J is transient, then for any $i \in E$,

$$\lim_{n\to\infty} P_{ij}^{\,n} = 0$$

$$LimP_{ij} = \begin{bmatrix} P_{11} & P_{12} \\ 0 & 1 \end{bmatrix}^n = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}$$

Definition 8. If J is recurrent, then for any $i \in E$,

$$\lim_{n\to\infty} P_{ij}^{\,n} \neq 0$$

$$LimP_{ij} = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}^n = \begin{bmatrix} P_{11}^{\,n} & P^n{}_{12} \\ P^n{}_{21} & P^n{}_{22} \end{bmatrix}$$

3.3.5 Ergodic Markov chain

The foundation of Markov chain theory is the Ergodicity Theorem. It establishes the conditions under which a Markov chain can be analyzed to determine its steady state behavior. A Markov chain is said to be ergodic if it is possible to eventually get from every state to every other state with positive probability.
If a Markov chain with transition matrix P is ergodic, then there is a unique vector λ such

that, for any probability vector λ , and for large n, the balance equation

λ .P= λ 3.27

- **Use of linear equation**

If a Markov chain with transition matrix P* is ergodic, then there is a unique vector λ such that, for any probability vector λ , and for large n, the balance equation

λ .P*= λ 3.28

and

$P^* = \lim_{n\to\infty} P^n$

Vector λ is called the equilibrium vector or the fixed vector of the Markov Chain.

Vector λ from equation 3.27 can be determined by finding P^n for larger and larger value of n, and then looking for a vector that the product $\lambda .P^n$ approaches.

λ P*= λ 3.29

[$\lambda_1, \lambda_2, \ldots \lambda_n$]P*= [$\lambda_1, \lambda_2, \ldots \lambda_n$] 3.30

To find the value of $\lambda_1, \lambda_2, \ldots \lambda_n$ using 3.29 and3.30, so that the normalization condition

$\lambda_1 + \lambda_2 + \ldots + \lambda_n = 1$ 3.31

Solving the system

[$\lambda_1, \lambda_2, \ldots \lambda_n$]P*= [$\lambda_1, \lambda_2, \ldots \lambda_n$]

$\lambda_1 + \lambda_2 + \ldots + \lambda_n = 1$

The result obtained will show that if there was no intervention of any kinds on the trait from generation to generation, then $\lambda_1, \lambda_2, \ldots \lambda_n$ implies the stationary distribution of the states.

- **<u>Induction Method</u>**

  Consider a Markov Chain with two possible states, suppose that the transition matrix is given by

  $P = \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix}$, where 0≤a,b≤1 and $P = E[\pi]$

  Suppose that a+b≠0. Then

$$P^n = \frac{1}{a+b}\begin{bmatrix} b & a \\ b & a \end{bmatrix} + \frac{(1-a-b)^n}{a+b}\begin{bmatrix} a & -a \\ -b & b \end{bmatrix}$$

By induction

$\lim_{n \to \infty} P^n = \frac{1}{(a+b)}\begin{bmatrix} b & a \\ b & a \end{bmatrix}$ and also

$$\lim_{n \to \infty} \pi^n = \begin{bmatrix} \frac{b}{a+b} & \frac{a}{a+b} \end{bmatrix}$$

$$\pi = [\pi_0 \quad \pi_1] = \begin{bmatrix} \frac{b}{a+b} & \frac{a}{a+b} \end{bmatrix}$$ Is called the limiting distribution

The probability distribution $\pi = [\pi_0 \quad \pi_1 \quad \pi_2 .....,.]$ is called the limiting distribution of the Markov chain X_n if

$\pi_j = \lim P(X_n = j \mid X_o = i)$ for all i,j ϵS, and we have

$$\sum_{j \in S} \pi_j = 1$$

Ergodic properties

An Ergodic Markov chain is

- irreducible (All states communicate),

- recurrent, and aperiodic

Most of the systems in which we are interested are modeled with ergodic Markov chains, because this corresponds to a well-defined steady state behavior.

3.3.6 Mean Recurrence

Let X be a Markov chain with state space S and transition matrix P and let us consider a sequence of trials where each trial has only two possible outcome(Present and Absent) of a trait. The probability of success (Present) is assumed to be the same for each trial. In such a sequence of trials, the geometric distribution is useful to model the distribution.

let F_{jj} be the probability of ever reaching j starting from j

$$P_j\{N_j<\infty\}= \begin{cases} 1 & \text{if } F_{jj}< 1, \\ 0 & \text{if } F_{jj} =1. \end{cases} \qquad 3.32$$

$F_{jj}<1$, N_j has the geometric distribution with success probability $P=1- F_{jj}$ starting at j; then $E_j[N_j] = \frac{1}{p} = \frac{1}{(1-F_{jj}),}$ and R_{ij} be the expected number of visits to state j starting at j, and defined by

$$R_{ij} = E_i(N_j) \qquad 3.33$$

We have

$$R_{jj} = \frac{1}{(1-F_{jj}),} \qquad 3.34$$

(here, $1/0= \infty$, $0 \infty=0$)

and

$$R_{ij} = F_{ij}R_{jj}, \quad \text{if } i\neq j \qquad 3.35$$

Case 1 (j is a recurrent state)

Suppose j is a recurrent state, then $F_{jj}=1$; then, (3.34) implies $R_{jj}= +\infty$
also if j can be reached from i, then $F_{ij}>0$ and $R_{ij}= \infty$ again. If, on the other hand, j cannot be reached from i, then $F_{ij} =0$ and $R_{ij} =0$. Hence, for j recurrent,

$$R_{ij} = \begin{cases} 0 & \text{if } F_{ij} =0 \\ +\infty & \text{if } F_{ij}>0 \end{cases} \qquad 3.36$$

Case 2 (j is a transient state)

Suppose j is transient, then $F_{ij} =0$; then, (3.36) implies $R_{ij} = 0$, j cannot be reached from i ; therefore,

$$F_{ij} =0$$
$$R_{ij} =0 \qquad 3.37$$

CHAPTER FOUR

VALIDATION AND DISCUSSION OF RESULTS WITH DATA

The newly introduced non-laboratory method of identification of a trait(s) using Markov chain approached is flexible and thus applicable to all living organisms. The data was a life data collected from the Albino Foundation Lagos, Nigeria (2018). The data was classified using the present and absent of visible trait of an albino (See Appendix A) and this was further classified by Hair, Skin and Eye colour (See Appendix B). Three generation was observed using the middle generation to get the data for the above and below generation (see figure 3.1). The state of the trait was classified as present and absent (P|A) with the study limited to three generations only (Grandparent, Parent and Child). The summary of the data were presented in a 2 by 2 observation matrices (n_{ij}).

4.1 Long run distribution

Using a trait

Children

n_{ij}=Parent	A	NA	
A	21	23	44
NA	24	13	37
	45	36	81

$$C_{ij}=\begin{pmatrix}0.48 & 0.52\\ 0.65 & 0.35\end{pmatrix}$$

Parent

n_{ij}=Grand Parent	A	NA	
A	9	13	22
NA	35	24	59
	44	37	81

$$P_{ij}=\begin{pmatrix}0.41 & 0.59\\ 0.59 & 0.41\end{pmatrix}$$

using Canonical presentation P_{ij} and C_{ij} are Recurrent

$$\lambda C_{ij} = \lambda$$

$$\lambda\begin{pmatrix}0.48 & 0.52\\ 0.65 & 0.35\end{pmatrix} = \lambda$$

$$[\lambda_1, \lambda_2,]\begin{pmatrix}0.48 & 0.52\\ 0.65 & 0.35\end{pmatrix} = [\lambda_1, \lambda_2] \qquad 4.1$$

$$\lambda_1 + \lambda_2 = 1 \qquad 4.2$$

expanding (4.1) become

$$0.48\lambda_1 + 0.65\lambda_2 = \lambda_1 \qquad 4.3$$

$$0.52\lambda_1 + 0.0.35\lambda_2 = \lambda_2 \qquad 4.4$$

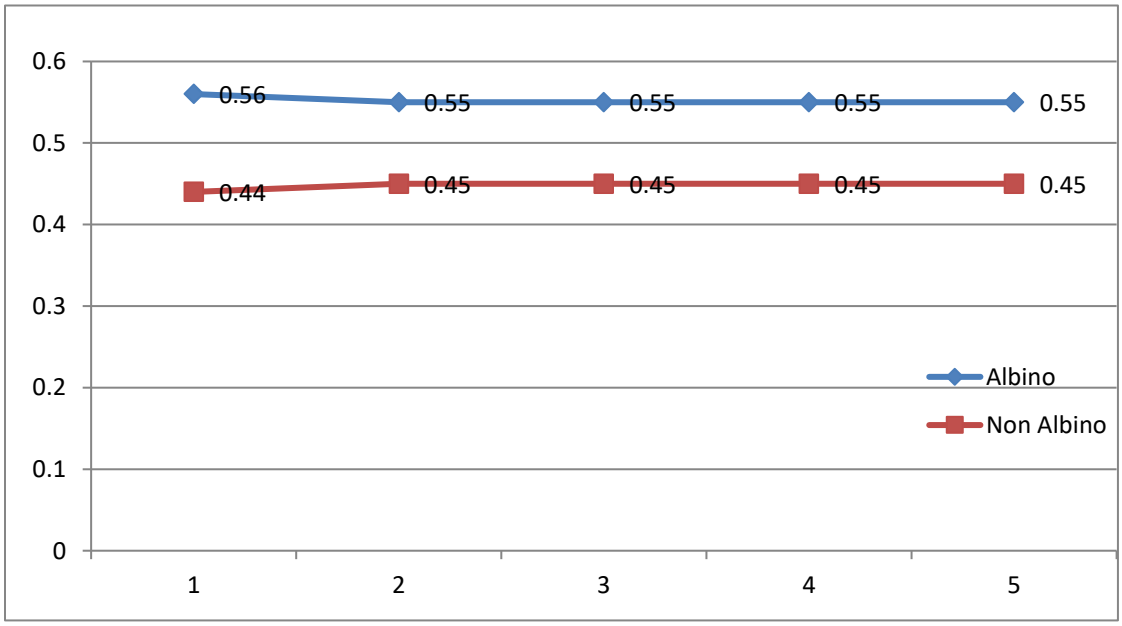
Solving the system using equation 4.2, 4.3 and 4.4

$\lambda_1 = 0.55$ and $\lambda_2 = 0.45$

Limiting vector = (0.55, 0.45)

The result show that if there was no intervention of any kinds on the trait from generation to generation, then implies the stationary distribution of the states.

$$C_{ij} = \begin{pmatrix} 0.48 & 0.52 \\ 0.65 & 0.35 \end{pmatrix}, \quad C^4{}_{ij} = \begin{pmatrix} 0.55 & 0.45 \\ 0.55 & 0.45 \end{pmatrix}$$



Graph 4.0 Line chart showing chance of Albino and Non-Albino generations from parent to children

Table 4.0 Table of Equilibrium of parent to children

			Equilibrium
n		$\lambda^T P^n$	λ^T
4		$\lambda^T P^4$	(0.55 0.45)
		The chain has reached	

		equilibrium of its own accord.	

n_{ij}=Grand parent

	Parent A	Parent NA	
A	9	13	22
NA	35	24	59
	44	37	81

$$P_{ij}=\begin{pmatrix}0.41 & 0.59\\ 0.59 & 0.41\end{pmatrix}$$

$$P_{ij}^2=\ b_{ij}=\begin{pmatrix}0.52 & 0.48\\ 0.48 & 0.52\end{pmatrix}$$

$$b_{ij}=\begin{pmatrix}0.52 & 0.48\\ 0.48 & 0.52\end{pmatrix},\quad \lambda=(\lambda_1,\lambda_2)$$

$$[\lambda_1,\lambda_2,]\begin{pmatrix}0.52 & 0.48\\ 0.48 & 0.52\end{pmatrix}=[\lambda_1,\lambda_2] \qquad 4.5$$

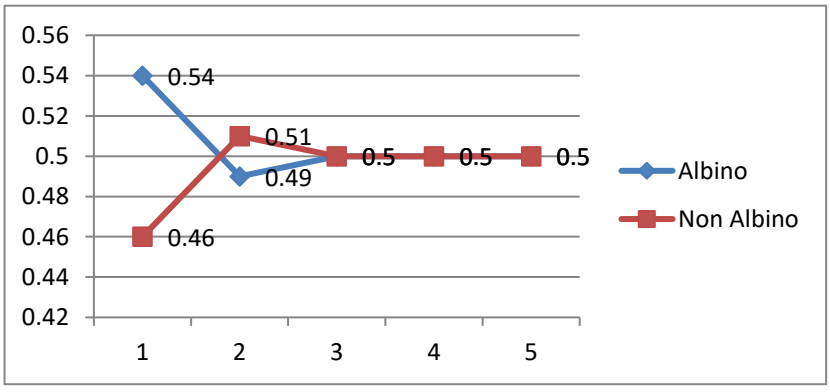
$$\lambda_1+\lambda_2=1 \qquad 4.6$$

Solving the system

$\lambda_1=0.50$ and $\lambda_2=0.50$

Limiting vector = (0.50, 0.50)

The result show that if there was no intervention of any kinds on the trait from generation to generation, then implies the stationary distribution of the states.



Graph 4.1 Line chart showing chance of Albino and Non-Albino generations from Grandparent to Parent

Grandparent to Parent

$$P_{ij}^2 = \begin{pmatrix} 0.52 & 0.48 \\ 0.48 & 0.52 \end{pmatrix}, \quad - \quad - \quad -, \quad [P_{ij}^2]^4 = \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}$$

Table 4.1 Table of Equilibrium of Grandparent to Parent

		Equilibrium
n	$\lambda^T P^n$	λ^T
4	$\lambda^T P^4$	(0.5 0.5)
	The chain has reached equilibrium of its own accord.	

4.2Testing equality of two Markov Chains

State Space = (Present, Absent)

Father to child Grandparent to Parent

$$C_{ij} = \begin{pmatrix} 0.48 & 0.52 \\ 0.65 & 0.35 \end{pmatrix} P_{ij} = \begin{pmatrix} 0.41 & 0.59 \\ 0.59 & 0.0.41 \end{pmatrix}, \text{ Let } b_{ij} = P^2_{ij} \quad , \quad b_{ij} = \begin{pmatrix} 0.52 & 0.48 \\ 0.48 & 0.52 \end{pmatrix}$$

$$H_0 : \begin{bmatrix} b_{11} \\ b_{21} \end{bmatrix} = \begin{bmatrix} c_{11} \\ c_{21} \end{bmatrix} \text{ vs } H_1 \begin{bmatrix} b_{11} \\ b_{21} \end{bmatrix} \neq \begin{bmatrix} c_{11} \\ c_{21} \end{bmatrix}$$

$$Q = \begin{pmatrix} b_{11} - c_{11} \\ b_{21} - c_{21} \end{pmatrix}^T S^{-2} \begin{pmatrix} b_{11} - c_{11} \\ b_{21} - c_{21} \end{pmatrix}$$

$$S^2{}_{cij} = \begin{bmatrix} 0.0057 & 0 \\ 0 & 0.0048 \end{bmatrix}, \quad S^2{}_{bij} = \begin{bmatrix} 0.0113 & 0 \\ 0 & 0.0042 \end{bmatrix}$$

$$S^2 = \begin{bmatrix} 0.0170 & 0 \\ 0 & 0.009 \end{bmatrix}$$

$$Q = \begin{bmatrix} 0.04 \\ -0.17 \end{bmatrix}^T \begin{bmatrix} 58.760 & 0 \\ 0 & 110.242 \end{bmatrix} \begin{bmatrix} 0.04 \\ -0.17 \end{bmatrix}$$

$$= 3.279$$

$P(\chi^2_{2,0.05} > 3.279) = 0.194$, since the Pvalue (0.194) > 0.05 we accept H_0 and conclude that the two transition probability matrices are equal and come from the same generation.

Comment

Stationary distribution exists for the two Markov Chain with limiting values at the fourth generation. The existence of stationary values indicates that the states were recurrent (Transient states cannot be made stationary).Furthermore, the generations were linked together testing the equality of the two transition matrices from the two generations (parent – children, Grandparent - parent), the result of the analysis shows that the transition probability matrices are unique.

4.3 Identify a proper marker

Hair

Father & Child E =(Trait Present, Trait Absent)

$$n_{ij}=\begin{bmatrix}7 & 24\\ 2 & 31\end{bmatrix}, \quad P_{ij}=\begin{bmatrix}0.23 & 0.77\\ 0.06 & 0.94\end{bmatrix}, \quad P^2{}_{ij}=\begin{bmatrix}0.10 & 0.90\\ 0.07 & 0.93\end{bmatrix}, - \; - \; - , \quad P^6{}_{ij}=\begin{bmatrix}0.07 & 0.93\\ 0.07 & 0.93\end{bmatrix}$$

Mother & Child

$$n_{ij}=\begin{bmatrix}5 & 17\\ 5 & 37\end{bmatrix}, \quad P_{ij}=\begin{bmatrix}0.23 & 0.77\\ 0.12 & 0.88\end{bmatrix}, \quad P^2{}_{ij}=\begin{bmatrix}0.14 & 0.86\\ 0.13 & 0.87\end{bmatrix}, - \; - \; - , \quad P^6{}_{ij}=\begin{bmatrix}0.13 & 0.87\\ 0.13 & 0.87\end{bmatrix}$$

Skin

Father & Child

$$n_{ij}=\begin{bmatrix}5 & 26\\ 3 & 30\end{bmatrix}, \quad P_{ij}=\begin{bmatrix}0.16 & 0.84\\ 0.09 & 0.91\end{bmatrix}, \quad P^2{}_{ij}=\begin{bmatrix}0.10 & 0.90\\ 0.07 & 0.93\end{bmatrix}, - \; - \; - , \quad P^6{}_{ij}=\begin{bmatrix}0.10 & 0.90\\ 0.07 & 0.93\end{bmatrix}$$

Mother & Child

$$n_{ij}=\begin{bmatrix}3 & 20\\ 4 & 37\end{bmatrix}, \quad P_{ij}=\begin{bmatrix}0.13 & 0.87\\ 0.10 & 0.90\end{bmatrix}, \quad P^2{}_{ij}=\begin{bmatrix}0.10 & 0.90\\ 0.10 & 0.90\end{bmatrix}, - \; - \; - , \quad P^6{}_{ij}=\begin{bmatrix}0.10 & 0.90\\ 0.10 & 0.90\end{bmatrix}$$

Eye

Father & Child

$$n_{ij}=\begin{bmatrix}28 & 9\\ 15 & 12\end{bmatrix},\quad P_{ij}=\begin{bmatrix}0.76 & 0.24\\ 0.56 & 0.44\end{bmatrix},\quad P^{2}{}_{ij}=\begin{bmatrix}0.71 & 0.29\\ 0.67 & 0.33\end{bmatrix},\text{- - - ,}\quad P^{6}{}_{ij}=\begin{bmatrix}0.70 & 0.30\\ 0.70 & 0.30\end{bmatrix}$$

Mother & Child

$$n_{ij}=\begin{bmatrix}28 & 4\\ 27 & 5\end{bmatrix},\quad P_{ij}=\begin{bmatrix}0.88 & 0.12\\ 0.84 & 0.16\end{bmatrix},\quad P^{2}{}_{ij}=\begin{bmatrix}0.87 & 0.13\\ 0.87 & 0.13\end{bmatrix},\text{- - - ,}\quad P^{6}{}_{ij}=\begin{bmatrix}0.87 & 0.13\\ 0.87 & 0.13\end{bmatrix}$$

Table 4.2 Computation of F_{jj} and R_{jj} for Hair

	Father To Child		Mother To Child	
n-step	F_{jj}	R_{jj}	F_{jj}	R_{jj}
1	$\begin{bmatrix}0.23\\ 0.94\end{bmatrix}$	$\begin{bmatrix}1\\ 17\end{bmatrix}$	$\begin{bmatrix}0.23\\ 0.88\end{bmatrix}$	$\begin{bmatrix}1\\ 8\end{bmatrix}$
2	$\begin{bmatrix}0.10\\ 0.98\end{bmatrix}$	$\begin{bmatrix}1\\ 50\end{bmatrix}$	$\begin{bmatrix}0.14\\ 0.87\end{bmatrix}$	$\begin{bmatrix}1\\ 8\end{bmatrix}$
3	-	-	-	-
4	-	-	-	-
5	-	-	-	-
6	$\begin{bmatrix}0.07\\ 0.93\end{bmatrix}$	$\begin{bmatrix}1\\ 14\end{bmatrix}$	$\begin{bmatrix}0.13\\ 0.87\end{bmatrix}$	$\begin{bmatrix}1\\ 8\end{bmatrix}$

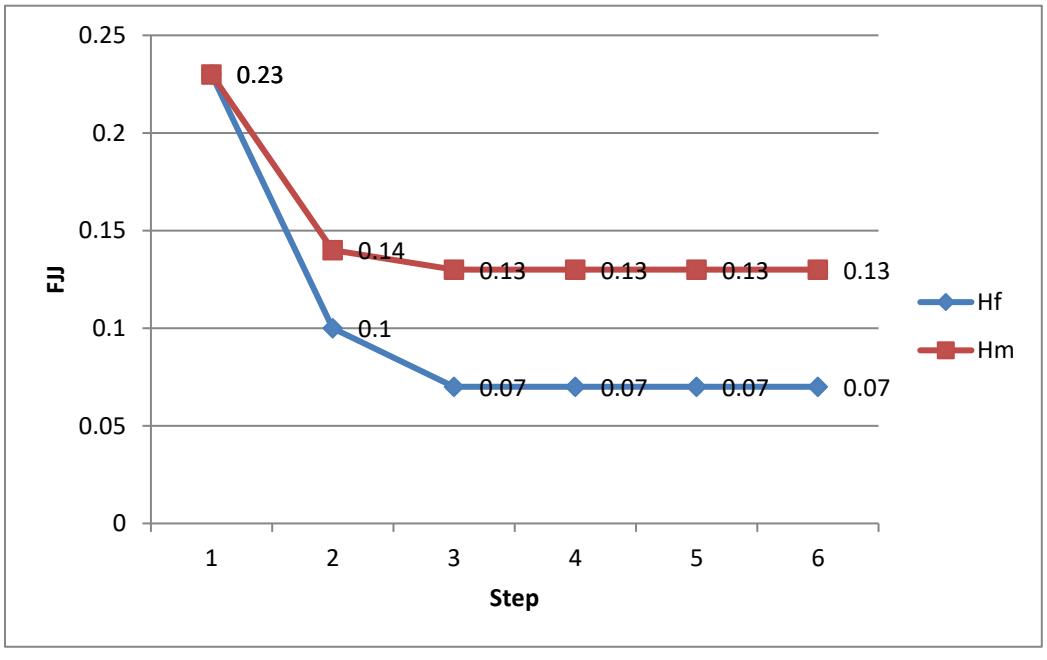
Table 4.3 Computation of F_{jj} and R_{jj} for Skin

	Father To Child		Mother To Child	
n-step	F_{jj}	R_{jj}	F_{jj}	R_{jj}
1	$\begin{bmatrix}0.16\\ 0.91\end{bmatrix}$	$\begin{bmatrix}1\\ 11\end{bmatrix}$	$\begin{bmatrix}0.13\\ 0.90\end{bmatrix}$	$\begin{bmatrix}1\\ 10\end{bmatrix}$
2	$\begin{bmatrix}0.10\\ 0.90\end{bmatrix}$	$\begin{bmatrix}1\\ 10\end{bmatrix}$	$\begin{bmatrix}0.10\\ 0.90\end{bmatrix}$	$\begin{bmatrix}1\\ 10\end{bmatrix}$
3	-	-	-	-
4	-	-	-	-
5	-	-	-	-

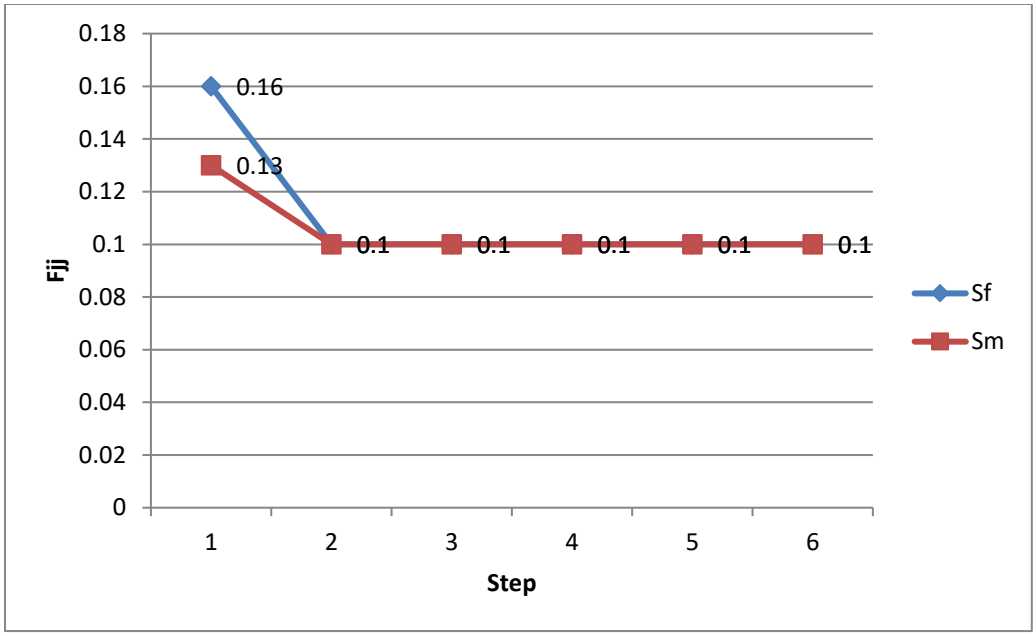
6	$\begin{bmatrix} 0.10 \\ 0.90 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 10 \end{bmatrix}$	$\begin{bmatrix} 0.10 \\ 0.90 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 10 \end{bmatrix}$

Table 4.4 Computation of F_{jj} and R_{jj} for Eye

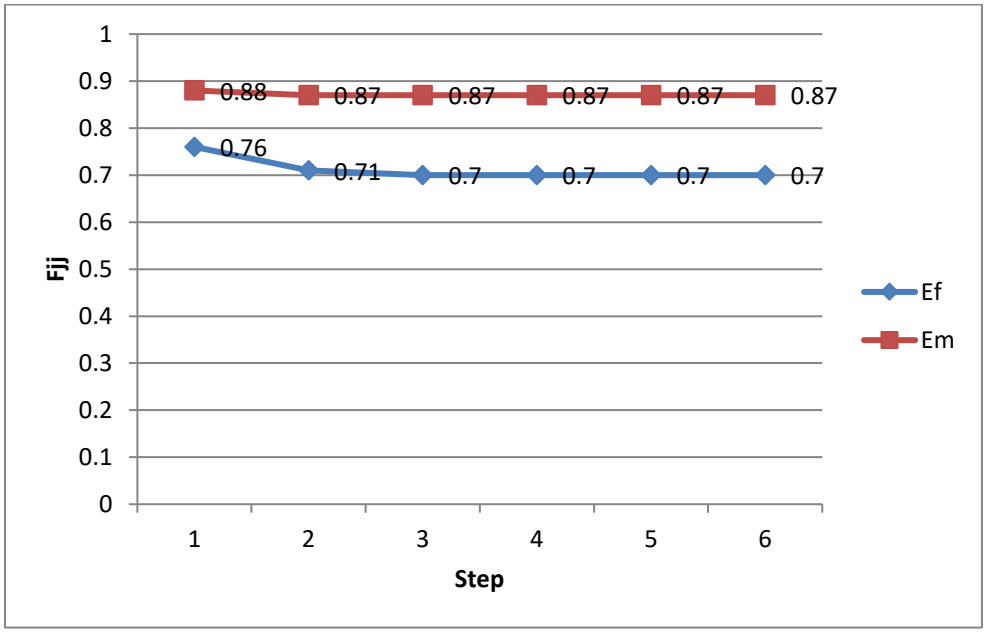
	Father To Child		Mother To Child	
n-step	F_{jj}	R_{jj}	F_{jj}	R_{jj}
1	$\begin{bmatrix} 0.76 \\ 0.44 \end{bmatrix}$	$\begin{bmatrix} 4 \\ 2 \end{bmatrix}$	$\begin{bmatrix} 0.88 \\ 0.16 \end{bmatrix}$	$\begin{bmatrix} 8 \\ 1 \end{bmatrix}$
2	$\begin{bmatrix} 0.71 \\ 0.33 \end{bmatrix}$	$\begin{bmatrix} 3 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 0.87 \\ 0.13 \end{bmatrix}$	$\begin{bmatrix} 8 \\ 1 \end{bmatrix}$
3	-	-	-	-
4	-	-	-	-
5	-	-	-	-
6	$\begin{bmatrix} 0.70 \\ 0.30 \end{bmatrix}$	$\begin{bmatrix} 3 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 0.87 \\ 0.13 \end{bmatrix}$	$\begin{bmatrix} 8 \\ 1 \end{bmatrix}$



Graph 4.2 Fjj line chart for Father-Child and Mother-Child on Hair



Graph 4.3 Fjj line chart forFather-Child and Mother-Child for Skin



Graph 4.4 Fjj line chart for Father-Child and Mother-Child for Eye

Linear Equation Application

If a Markov chain with transition matrix P_{ij} is ergodic, then there is a unique vector λ [Proportion of the Marker (λ_1) present and proportion of the Marker (λ_0) absent] such that,for any probability vector λ , and for large n, the balance equation

$$\lambda P = \lambda \qquad 4.7$$

$$[\lambda_0, \lambda_1,]P = [\lambda_0, \lambda_1,] \qquad 4.8$$

Utilizing the probability condition that

$$\lambda_0 + \lambda_1 = 1 \qquad 4.9$$

To find the value of $\lambda_{0, \text{ and }} \lambda_1$ for the above matrices, solving the system we have following with the expected number of time (generations) to return to the given state (m_1 marker present and m_o marker absent).

Table 4.5 Showing the Proportion of marker present,absent and the mean recurrent generations

Marker		Invariant Distribution		Mean Recurrence generation	
		λ_1	λ_0	m_1	m_o
Hair Colour	F&C	0.11	0.89	9.09	1.12
	M&C	0.11	0.89	9.09	1.12
Skin Colour	F&C	0.27	0.73	3.70	1.37
	M&C	0.10	0.90	10.00	1.11
Eye Colour	F&C	0.75	0.25	1.33	4.00
	M&C	0.89	0.11	1.12	9.09

4.3 Calculating Chances of Hidden and Reemerge of Trait

From the transition probabilities matrices on eye, looking at the chances of hidden and reemerge of trait with 0=Trait Present and 1=Trait Absent.

Father & Child (Eyes)

P(Father=0, Child =0|Grand parent =0)

$$= P[X_n = 0, X_{n+1} = 0 \mid X_{n-1} = 0]$$
$$= P[X_n = 0 \mid X_{n-1} = 0]P[X_{n+1} = 0 \mid X_n = 0, X_{n-1} = 0]$$
$$= P[X_n = 0 \mid X_{n-1} = 0]P[X_{n+1} = 0 \mid X_n = 0]$$
$$= P_{00}P_{00}$$
$$= 0.71$$

P(Father=0, Child =0|Grand =1)

$$= P[X_n = 0, X_{n+1} = 0 \mid X_{n-1} = 1]$$
$$= P[X_n = 0 \mid X_{n-1} = 1]P[X_{n+1} = 0 \mid X_n = 0, X_{n-1} = 1]$$
$$= P[X_n = 0 \mid X_{n-1} = 1]P[X_{n+1} = 0 \mid X_n = 0]$$
$$= P_{01}P_{00}$$
$$= 0.18$$

P(Father=1,Child =1|Grand parent =0)

$$= P[X_n = 1, X_{n+1} = 1 \mid X_{n-1} = 0]$$
$$= P[X_n = 1 \mid X_{n-1} = 0]P[X_{n+1} = 1 \mid X_n = 1, X_{n-1} = 0]$$
$$= P[X_n = 1 \mid X_{n-1} = 0]P[X_{n+1} = 1 \mid X_n = 1]$$
$$= P_{10}P_{11}$$
$$= 0.25$$

P(Father =0, Child =1 |Grand parent =1)

$$= P[X_n = 0, X_{n+1} = 1 \mid X_{n-1} = 1]$$
$$= P[X_n = 0 \mid X_{n-1} = 1]P[X_{n+1} = 1 \mid X_n = 0, X_{n-1} = 1]$$
$$= P[X_n = 0 \mid X_{n-1} = 1]P[X_{n+1} = 1 \mid X_n = 0]$$
$$= P_{01}P_{10}$$
$$= 0.13$$

P(Father=0, Child=1 |Grand parent =0)

$= P[X_n = 0, X_{n+1} = 1 \mid X_{n-1} = 0]$
$= P[X_n = 0 \mid X_{n-1} = 0]P[X_{n+1} = 1 \mid X_n = 1, X_{n-1} = 0]$
$= P[X_n = 0 \mid X_{n-1} = 0]P[X_{n+1} = 1 \mid X_n = 1]$
$= P_{00}P_{11}$
$= 0.35$

P(Father =1,Child =0|Grand parent =1)

$= P[X_n = 1, X_{n+1} = 0 \mid X_{n-1} = 1]$
$= P[X_n = 1 \mid X_{n-1} = 1]P[X_{n+1} = 0 \mid X_n = 1, X_{n-1} = 1]$
$= P[X_n = 1 \mid X_{n-1} = 1]P[X_{n+1} = 0 \mid X_n = 1]$
$= P_{11}P_{01}$
$= 0.11$

Mother & Child (Eyes)

P(Mother=0,Child =0|Grand parent =0)

$= P[X_n = 0, X_{n+1} = 0 \mid X_{n-1} = 0]$
$= P[X_n = 0 \mid X_{n-1} = 0]P[X_{n+1} = 0 \mid X_n = 0, X_{n-1} = 0]$
$= P[X_n = 0 \mid X_{n-1} = 0]P[X_{n+1} = 0 \mid X_n = 0]$
$= P_{00}P_{00}$
$= 0.87$

P(Mother=0,Child =0|Grand parent =1)

$= P[X_n = 0, X_{n+1} = 0 \mid X_{n-1} = 1]$
$= P[X_n = 0 \mid X_{n-1} = 1]P[X_{n+1} = 0 \mid X_n = 0, X_{n-1} = 1]$
$= P[X_n = 0 \mid X_{n-1} = 1]P[X_{n+1} = 0 \mid X_n = 0]$
$= P_{01}P_{00}$
$= 0.11$

P(Mother=1,Child=1|Grand parent =0)

$= P[X_n = 1, X_{n+1} = 1 \mid X_{n-1} = 0]$
$= P[X_n = 1 \mid X_{n-1} = 0]P[X_{n+1} = 1 \mid X_n = 1, X_{n-1} = 0]$
$= P[X_n = 1 \mid X_{n-1} = 0]P[X_{n+1} = 1 \mid X_n = 1]$
$= P_{10}P_{11}$
$= 0.13$

P(Mother=0, Child =1 |Grand parent =1)

$$= P[X_n = 0, X_{n+1} = 1 \mid X_{n-1} = 1]$$
$$= P[X_n = 0 \mid X_{n-1} = 1]P[X_{n+1} = 1 \mid X_n = 0, X_{n-1} = 1]$$
$$= P[X_n = 0 \mid X_{n-1} = 1]P[X_{n+1} = 1 \mid X_n = 0]$$
$$= P_{01}P_{10}$$
$$= 0.10$$

P(Mother =0,Child=1|Grand parent=0)

$$= P[X_n = 0, X_{n+1} = 1 \mid X_{n-1} = 0]$$
$$= P[X_n = 0 \mid X_{n-1} = 0]P[X_{n+1} = 1 \mid X_n = 1, X_{n-1} = 0]$$
$$= P[X_n = 0 \mid X_{n-1} = 0]P[X_{n+1} = 1 \mid X_n = 1]$$
$$= P_{00}P_{11}$$
$$= 0.14$$

P(Mother=1,Child=0|Grand parent =1)

$$= P[X_n = 1, X_{n+1} = 0 \mid X_{n-1} = 1]$$
$$= P[X_n = 1 \mid X_{n-1} = 1]P[X_{n+1} = 0 \mid X_n = 1, X_{n-1} = 1]$$
$$= P[X_n = 1 \mid X_{n-1} = 1]P[X_{n+1} = 0 \mid X_n = 1]$$
$$= P_{11}P_{01}$$
$$= 0.02$$

The chances of hidden and reemerge of trait using three generations are given in the table below.

Table 4.6 chances of hidden and reemerge of trait using three generations

Father - Child		Prob that					
		X_n		X_{n+1}			
		0	1	0	1		
Given	0					P_{00} P_{00}	0.71

X_{n-1}	1					P_{01} P_{00}	0.18
	0					P_{10} P_{11}	0.25
	1					P_{01} P_{10}	0.13
	0					P_{00} P_{11}	0.35
	1					P_{11} P_{01}	0.11

Mother - Child		Prob that					
		X_n		X_{n+1}			
		0	1	0	1		
Given X_{n-1}	0					P_{00} P_{00}	0.87
	1					P_{01} P_{00}	0.11
	0					P_{10} P_{11}	0.13
	1					P_{01} P_{10}	0.10
	0					P_{00} P_{11}	0.14
	1					P_{11} P_{01}	0.02

Comment Markov chain concepts, techniques and how to apply them to the identification of a trait that can be used as marker using a 2x2 Markov chain was demonstrated. Using both the Mean Recurrence and the Linear equation, the results indicate that the three traits (Hair, Skin and Eye) are recurrent in nature with probability of the trait present in both children and parent varies. The third trait (eye) was the best recurrent state with least mean recurrence time -approximately one.

4.4 Marker properties and application on albinism

Ergodicity property

(Eyes Colour)

Father & Child Mother & Child

$$P_{ij}=\begin{bmatrix}0.76 & 0.24\\ 0.56 & 0.44\end{bmatrix} \quad P_{ij}=\begin{bmatrix}0.88 & 0.12\\ 0.84 & 0.16\end{bmatrix}$$

Class [0,1] Class [0,1]

Irreducibilty

The chains are irreducible since the two states are reachable from each other (no redundant or dead-end state) and with a single class of communication for each.

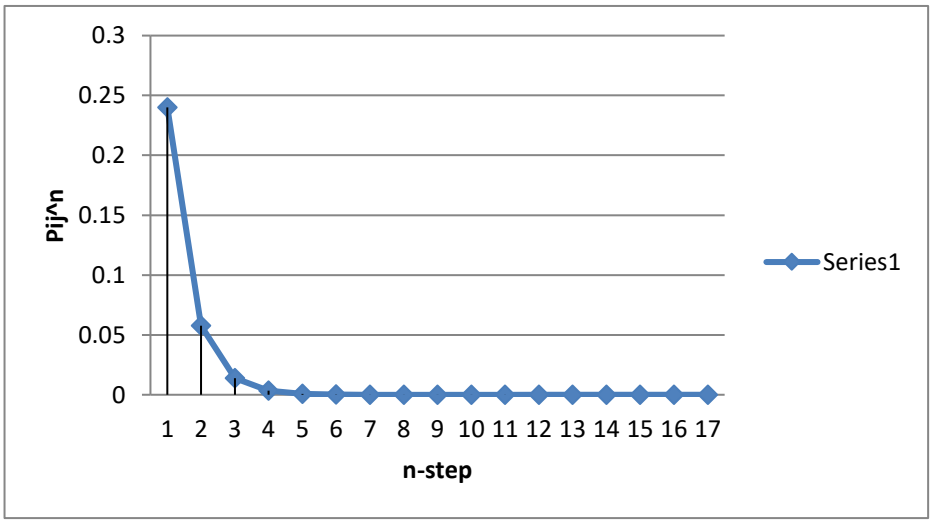
Periodicity

The chains are aperiodic since the two states can only return to itself with transition number equal one

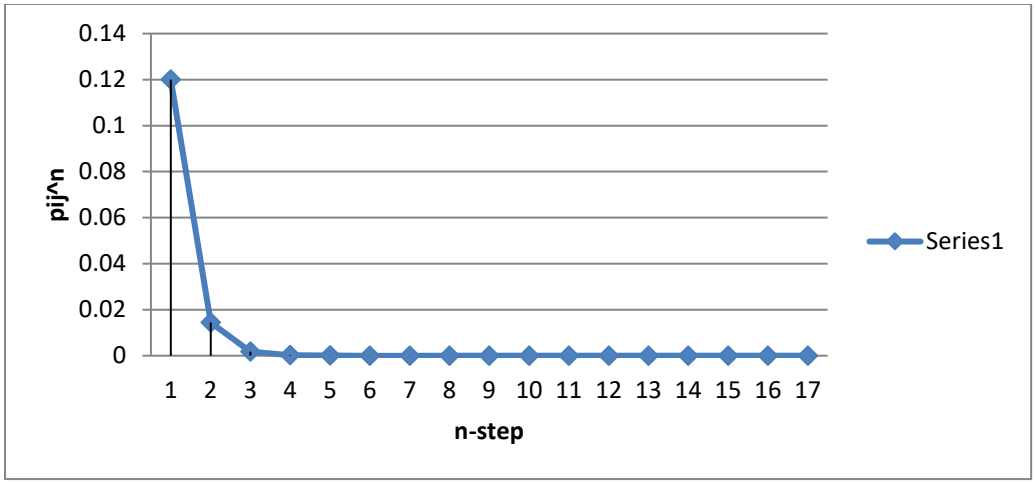
Recurrent State property

from (3.9) , $[P_{ij}]^n \to 0, \text{as} \quad n\to\infty$

we have the following graphs for the two chains,



Graph 4.5 line chart for Father-Child on recurrent state



Graph 4.6 line chart for Mother-Child on recurrent state

as $n \rightarrow \infty$, so the probability of staying at j forever is 0 and consequently $f_{ii}=1$, which is an indication that state i(present of eyes colour) is a recurrent state.

Positive Recurrent States Computation

A Markov chain is called recurrent if it returns back in a finite time with probability 1. That means you can always expect it evolves to its origin. However, this cannot guarantee that the mean time of return is also finite. If it is, then the chain is positive-recurrent, otherwise null-recurrent.
A positive recurrent state (in a finite state MC) has a finite expected return time. Otherwise, the state is null recurrent. A positive recurrent chain is a chain where all states are positive recurrent.

Prove to show that state is positive recurrent
(Calculating expected number of transition to return to given state)

There is need to make the claim about the positive recurrence of a state, through calculating expected number of transitions to return to the given state if starting from the state.

Father & Child Mother & Child

$$P_{ij}=\begin{bmatrix}0.76 & 0.24\\ 0.56 & 0.44\end{bmatrix} \quad , \quad P_{ij}=\begin{bmatrix}0.88 & 0.12\\ 0.84 & 0.16\end{bmatrix}$$

(Father and Child)

Denote n (k) = the number of transitions to return to state k when starting from k.

k= (0, 1)

we want to calculate n(k) and show that n(k) $<\infty$. We have the following systems

$$n(0) = 1 + 0.76n(0) + 0.24n(1)$$
$$n(1) = 1 + 0.56n(0) + 0.44n(1)$$

Solving the above yields n(0)= -1 and n(1)= -1, so n(0) = -1 $< \infty$ and n(1)= -1 $< \infty$ and therefore state 0 and 1 are positive recurrent

(Mother and Child)

$$n(0) = 1 + 0.88n(0) + 0.14n(1)$$
$$n(1) = 1 + 0.84n(0) + 0.16n(1)$$

Solving the above yields |n(0)|= $\frac{50}{29}$ and |n(1)|= $\frac{25}{29}$, so |n(0)| $< \infty$ and

|n(1)| $< \infty$, and therefore state 0 and 1 are positive recurrent

Comment. The Markov chains show that all states communicate, the chains are finite, irreducible, only one class exists and also the states are positive recurrent. Hence, from the results of the work it shows that Ergodic Markov chain (properties) is the only good marker that can be used in the identification of a trait.

4.4General Remark

With the introduction of Markov Chain concepts and techniques and the applications with also the use of charts on the three traits-Hair, Skin and eye. It was observed that the Ergodicity property of Markov Chain hold for any trait to be a proper marker and the eye colour seem to be the dominant trait among the three traits with mean recurrent generation to be one.

CHAPTER FIVE

SUMMARY, CONCLUSION AND RECOMMENDATION

In this chapter, we have the summary, conclusion, recommendations, and contributions to knowledge in this research work.

5.1 Summary

Traits are the ways in which family members are alike, a genetic maker that is passed through parents' genes to their children. Most specific traits are passed directly from one or both parents. Genetic disorders are also traits that can be passed from a parent to a child. Deoxyribonucleic Acid (DNA) test is the only known method of identification of genetics. Deoxyribonucleic Acid (DNA) test can be helpful in determining many different family relationships and help expand the family tree. DNA test is a valuable tool to help people who are interested in tracing their genealogy or ancestry. For the purpose of finding alternative to this Deoxyribonucleic Acid (DNA) test when not readily available or accessible, a probabilistic procedure (non- laboratory test)was proposed using Markov Chain methodology that is less costly, accessible and provides a sense of relief for proper classification.

5.2 Conclusion

In this study, probabilistic procedure was introduced using Markov chain Monte Carlo (MCMC) concepts and techniques and how to apply them to classify cases based on probability transition matrix. The generations was link together using the matrices from the three generations (Grandparent – parent - Child), the result of the analysis shows the uniqueness of the system when three successive time points exist. Markov chain approach was successfully applied with all matrices found to be recurrent but with different mean recurrence. The Markov chain modeling introduced accessibility to an alternative to laboratory test and conclude that useful trait must possess a marker whose categories are recurrent and with periodicity of close to one.

The application was done using the traits: Hair, Skin and Eye colour of Albinos with the results indicating that the three traits (Hair ,Skin and Eye) are recurrent. The third trait (Eye colour) was the best marker with mean recurrence equal one. The clinical epidemiologic results also show this features (See table 2.0, 4.4 and 4.5).

5.3 Contributions to Knowledge

The research work introduced a probabilistic (non- laboratory test) procedure using Markov Chain approach in identification of trait that can be used as a proper marker. Rather than trying to look for laboratory test in identification of a trait that can be used as a proper marker, this test is available and less expensive.

5.4 Suggestion for further studies

Although we have developed the model for the identification of a inheritable trait using the probabilistic procedure – the test can also be apply to other traits not specific to albinism, two or more traits identification can be studied and the test can be applied to study other genetic traits with appropriate modification.

Reference

Abnizova,. I; Rust, A.G.; Robinson.M; TeBoekhorst. R and Gilks, W.R.(2006):Transcription binding site predicting using Markov Models, *Journal of Bioinformatics and Computational Biology, 01 Apr 2006, 4(2):425-441.*

Alam.M.M;Uddin.S. and Mahmud.N.U(2014): Text Dependent Speaker Identification using Hidden Markov Model and Mel Frequency Cepstrum Coefficient, *International Journal of computer Application 104(14): 33-37.*

Arie .H and Spickman,F(1992): On Ergodicity and Recurrence Properties of a Markov Chain with an Application to an Open Jackson Network ,*Advances in Applied Probability Journal 24,(2), 343-376.*

Asra S.,and Shubhangi ,D. C. (2015): Personality Trait Identification Using Unconstrained Cursive and Mood Invariant Handwritten Text, *International Journal of Education and Management Engineering 5(5),20-31.*

Bhasin M.K. and Walter H. (2013) Genetic Markers in Human Blood, *Journal of Life Sciences 5(2), 71-121.*

Bhavani,M and Shoba Rani(2013): Human Identification Using Finger and Iris Images, *International Journal of Computer Trends and Technology-4(3),258-263.*

Boas Franz (1910*): Family Traits as Determinant by Heredity and Environment,* Department of Anthropology, Columbia University. *American Anthropologist Journal 105(1):125—138.*

Bollati ,V and Baccarelli,A (2010) : Heredity, *Environmental epigenetics Journal 105, 105–112*

Brian R. Clifford (1980): Voice identification by human listeners: On ear witness reliability, *Journal of Law and Human Behavior December 1980, 4(4), 373–394.*

Brighenti.C,Wahlberg.B and Rojas.C.R.(2009):"*Input design using Markov chains for system identification" 48th IEEE Conference on Decision and Control(CDC) 110-124.*

Chial, H. (2008): *DNA sequencing technologies key to the Human Genome Project. Nature Education 1(1):219*

Cho,J.S and White,H (2014): *Testing the equality of two positive-Definite Matrices with Application to Information Matrix Testing,* KEA International Conference *Korea University,* 34-76.

Christopher J. Hopwood, M. Brent Donnellan, Daniel M. Blonigen, Robert F. Krueger, Matt McGue, William G. Iacono, and S. Alexandra Burt(2011): Genetic and environmental influences on personality trait stability and growth during the transition to adulthood. *J PersSoc Psychol. 100(3): 545–556.*

Cinlar, Erhan (1975): *Introduction to Stochastic processes, Prentice-Hall, Inc., Englewood cliffs, N.J: Prentice-Hall, [1974, c1975] U.S.A. 25-76*

Cinquemani.E, Argeitis.M.A and Lygeros.J(2008):Identification of Genetic Regulatory Networks: A Stochastic Hybrid Approach, 17th Word Congress International Federation of Automatic Control Seoul, Korea 237-341 .

Dai,Q., Carr,P., Sigal,L and Hoiem,D(2015): Family Member Identification from Photo Collection, IEEE Winter Conference on Applications of Computer VisionWaikoloa,IEEE Hl, USA 124-154

Das,.K.T. and Nahar,K.M.O (2016): A Voice Identification System using Hidden Markov Model, *Indian Journal of Science and Technology, 9(4), 1-6*

David Delgado-GómezFedericoSuknoDavidAguadoCarlosSantacruzAntonioArtés-Rodriguez(2011) : Individual identification using personality traits, *Journal of Network and Computer Applications33(3),293-299*

Gales, Mark and Young, Steve (2007):The Application of Hidden Markov Models in Speech Recognition, *Foundation and Trends in signal processing 1(3), 195–304.*

Gamez Wakiza, Frank Schmidt and David Watson (2010):Psychological Bulletine 136 (5), 768-821

George, A.W. and Thompson, E.A.(2003*:" Discovering Disease genes: Multipoint Linkage Analysis via a New Markov Chain Monte Carlo Approach*" *Journal of Statistical Science 18 (4)*: *92-96*

Gilissen, C., Hoischen, A., Brunner, H.G. and Veltman, J.A (2012):Disease gene identification strategies for exome sequencing, *European Journal of Human Genetics* 20(5*):*490-497.

Glazner, Chris and Thompson, Elizabeth Alison (2011); Improving Pedigree-based Linkage Analysis by Estimating Co ancestry Among Families, *University of Washington Berkeley Electronic Stat Appl Genet Mol Biol. 6;11 (2),1544-6115*

Goldman.R.N. and Weinberg J.S. (1985): *Statistics An Introduction*, ,Englewood cliffs, N.J *Englewood Cliffs, NJ*: *Prentice Hall*. 334–353. U.S.A.

Grønskov. K., Jakob, E.k and Nielsen, K.B (2007): Oculocutaneous albinism, *Orphanet Journal of Rare Diseases 2:43. doi:10.1186/1750-1172-2-43*

Giovanni, Montana (2006): Statistical methods in genetics, *Bioinformatics Journal 7(3):297-308*

Gregor Mendel (1866) http://www.esp.org/foundations/genetics/classical/gm-65.pdf

Holmes, S.J.(1943) :Family Fixture, *Journal of Heredity*, 34(1), 3–7, https://doi.org/10.1093/oxfordjournals.

Jia.C,Qian.M,Kang.Y and Jiang.D (2014): Modelling Stochastic Phenotype switching and bet-heding in bacteria: Stochastic nonlinear dynamics and critical state identification *Quantitative Biology,2(3): 110-125.*

Jonathan R. Pleban, D. Scott Mackay, Timothy L. Aston, Brent E. Ewers, and Cynthia Weinig (2018):*Phenotypic Trait Identification Using a Multimodel Bayesian Method, Front Plant Science.* 9: 448 | https://doi.org/10.3389/fpls

Kanchan Wani (2015) *International Journal of Computer Science and Networking IJCSN Journal 4, (4), 650 - 654*

Kimmel,.G and Shamir.R (2005): "A Block-Free Hidden Markov Model for Genotypes and Its Application to Disease Association" *Journal of Computational Biology 12(10):1243-60*

Khmelev,.D.V. and Tweedie,F.J.(2001): Using Markov Chains for Identification of Writers, *Literary and Linguistic Computing, 16 (4), 299-307*

Kulp.D, Haussler.D,Reese.M.G. and Eeckman.F.H (1996): A Generalized Hidden Markov Model for the Recognition of Human Genes in DNA, *SystMol Biol. Journal 4:134-42. Califonia*

Laird,Nan.M and Lange Christopher (2011): The fundamentals of modern statistical genetics, *Department of Biostatistics Harvard University 1-6*

Lichtenberg. J.W and Powel. L.K (1984): Methods of Sequential Analysis for Studying Family Interactions, *Annual Convention of the Association for Behavior Analysis Journal Anxiety Research 4(2), 109-124.*

Lonardo Ana Maria Di; Darlu, Pierre; Baur, Max; Orrego, Cristian; King andMary-Claire (1984): Identifying the families of kidnapped children, *The American Journal of Forensic Medicine and Pathology: 5 (4), 339-347*

Lynn M. LoPucki (2001): Human Identification Theory and the Identity Theft Problem, *Texas Law Review, 80, 89-134.*

Maji,S. and Garg,.D (2012): Mini Review on gene finding using hidden Markov Model, *Journal of Applied Sciences, 12(15), 1518-1525.*

Marcus CheethamJürgen HänggiLutz Jancke (2014) *Social Cognitive and Affective Neuroscience Journal*, 9(11),4, 1836–1844, https://doi.org/10.1093/scan/nst179

Mauricio Correa,Gabriel Hermosilla,Rodrigo Verschae and Javier Ruiz-del-Solar(2012): Human Detection and Identification by Robots Using Thermal and Visual Information in Domestic Environments. *Journal of Intelligent & Robotic Systems, 66(1–2), 223–243*

Mendel, Gregor. (1866). http://www.esp.org/foundations/genetics/classical/gm-65.pdf

Michal Choras (2008) https://ieeexplore.ieee.org/document/5967615/metrics#metrics

Mohamed, F.A.,El-Sayed,N,and Seifeldin,N.S (2010) :Clinico-epidemiologic features of albinism in northeast section of Cairo-Egypt. *The Egyptian Journal of Medical Human Genetics 11,167-172*

Nisha (2017): Voice Recognition Technique, *International Journal for Research in Applied Science and Engineering Technology (IJRASET) 5 (V), 262-268*

O'Neil, Dennis(2013). Basic Principles of Genetics: Mendel's Genetics. *Basic Principles of Genetics: Mendel's Genetics*.N.p., n.d. Web. <http://anthro.palomar.edu/mendel/mendel_1.htm>.

Olarinoye,S.B. and Jolayemi,E.T (2016): The use of transition probability Matrix in Studying Traits. Nigerian Statistical Association (NSA) annual Conference Abuja State 460 -462

Olarinoye,S.B and Jolayemi E.T.(2017*)*- Identification of An Albinism Marker Using Markov Chain, *ICASTOR Journal of Mathematical Sciences11(2) 3-24 India*

Olarinoye, S.B. and Jolayemi, E.T (2018*):* Linking generations together using Markov Chain Approach. 52nd Annual Conference of Science Association of Nigeria (SAN) Lagos 72-83

Priyanka Srikanth (2015) Using Markov Chains to predict the natural progression of diabetic retinopathy, *International Journal Ophthalmology*. 8(1): 132–137.

Rashid,N ,Yahya ,.M and Shafie,A.A(2013): Human identification at a Distance using Body shape information, 5th international conference on Mechatronic university of Malaysia 312-387.

Riccardo.S, Alesandro.F and Matteo, M (2019): Agent Behavioral Analysis Based on Absorbing Markov Chain, International Conference on Autonomous Agents and Multi Agent Systems (AAMAS) 647-655.

Richard A.J and Dean W.W(1988)- *Applied Multivariate Statistical analysis 5th edition prentice-*Hall, Inc. Englewood Cliff, New Jersey, USA 221-278.

Richard A. King and C. Gail Summers (1988) : Albinism, *Dermatologic Clinics, Journal 6,(2), 217-228*

Richard Kay (1986): A Markov Model for Analysing Cancer Markers and Disease States in Survival Studies , *International Biometric Society Journal 42, (4), 855-865*

Richard L.Tweedie(1975):*Sufficient conditions for ergodicity and recurrence of Markov chains on a general state space, Stochastic Processes and their Applications3(4), 385-403*

Richard Mayeux(2005): Mapping the new frontier: complex genetic disorders. *Journal Clinical Investigation. 115(6): 1404–1407.*

Robert Plominand Denise Daniels (2011) :Why are children in the same family so different from one another? *International Journal of Epidemiology*, 40(3), 563–582.

Robinson, A , Dobzhansky, T and Griffiths, A. J.F (2018): Heredity. *Encyclopedia Britannica https://www.britannica.com/science/heredity-genetics.*

SalahAly and AdnanGutub (2017): intelligent Recognition System for Identifiying Items and Pilgrims.*NED University Journal of Research 94, 78-124.*

Sasaki,M,Abdel,D.A, and Dinu,E.(2018) :Heredity of Beauty, *Journal of Young Investigator,* 34 (4),1-6.

Saunders, S and Rogers, T (1994): Accuracy of Sex Determination Using Morphological Traits of the Human Pelvis, *Journal of Forensic Sciences* 39(4),1047-1056.

Sazberg,S.L., Delcher,A.L, Kasif,S. and White,O.(1998): Microbial gene identification using interpolated Markov Models. National Library of Medicine *National Institutes of Health Nucleic Acids Journal 26(2):544-8.*

SeemaG, VandanaS.K, Sandeep .V, Huma K, Bhatt, Nuzhat.H, Mahendra .P, Negi, V V,B, and Ashim .G,(2016): Understanding molecular markers in recurrent oral squamous cell carcinoma treated with chemoradiation, *Journal of Heliyon 2(12): e00206*

Science clarified (2018)http://www.scienceclarified.com/Ga-He/Genetics.html

Shahin ,Ismail (2017):Talking Condition identification Using Second-Order Hidden Markov Models, 3rd international conference on Information and Communication Technologies Syria 542-583.

Singh, P. ,Khamis,I and Higgins, J.(2003) : Hotelling's T2 approximation for bivariate Dichotomous data, 15th Annual Conference on Applied Statistics in Agriculture Kensas State University 218-223

Surbhi and Vishal,A.(2012):A Face Identification Technique for Human Facial image, *International Journal of Computer Science and Information Technologies 3(6).5390-5393*

SyedaAsra ,D. C. Shubhangi(2018) *International Journal of Education and Management Engineering 5(5):20-31*

Tangrot,.J, Wang,.L., Kagstrom, B. and Sauer,U.H(2006): FISH- family identification of sequence homologues using structure anchored hidden Markov models, Pub Med Us National Library of Medicine National Institutes of Health, *Nuleric Acid Res. 34(2), 10–14*

Thomas H. Maugh II (2009) Microsoft ® Encarta ® 2009. © 1993-2008 Microsoft Corporation.

Tsagkrasoulis ,D; Hysi ,P; Spector, T and Montana (2017):*Scientific Reports volume 7, (45885)Macmillan Publishers Limited, part of Springer Nature. (www.heritabilitymaps.info).*

Tyler K. Perrachione, Stephanie N. Del Tufo and John D. E. Gabriel (2011): Human Voice Recognition Depends on Language, *Ability* Science *Journal 333(6042):595. doi: 10.1126/science.1207327*

Umang Gupta and Niladri Chatterjee (2013): International Symposium on Computational and Business Intelligence Indian Institute of Technology Delhi India.https://ieeexplore.ieee.org/document/6724349/

VanitaPatil, and PatilA M. (2015): Survey Paper on Human Identification Using Sclera, *International Journal of Computer Science and Network(IJCSN), 4(6),*

Venter J. Craig, Jean-Pierre Hubaux, Bradley Adam Malin, and EffyVayena(2017) PNAS https://doi.org/10.1073/pnas.1711125114

www.science clarified.com/…He/Genetic-Disorder html 2018 Advameg, Inc.

Yingbo ZhouandAjay Kumar(2011):Palm-Vein based identification system, *IEEE Transactions on Information Forensics and Security Journal ,6(4*):*1259-1274 ·,*

Zhu, J., Hong. and Hughes,J.G.(2002) :*Using Markov Chains for Link Prediction in Adaptive Web Sites* School of Information and Software Engineering, University of Ulster 145-174 {jh.zhu, j.hong, jg.hughes}@ulster.ac.uk

Zyang,K., Yang,.Y,Devanarayan,.V,Xie,.L, Deng,.Y. and Donald,.S.(2011):A Hidden Markov model-based algorithm for identifying tumor subtype using array CGH data" Zhang et al. license BMC Genomics 12 (5). doi: 10.1186/1471-2164-12-S5-S10.

APPENDIX A

	GRAND PARENT	STATUS	PARENT	STATUS	CHILDREN
1	FATHER	NO			
	MOTHER	NO	FATHER	No	
	FATHER	NO			**YES**
	MOTHER	NO	MOTHER	No	
2	FATHER	YES			
	MOTHER	NO	FATHER	YES	
	FATHER	NO			YES
	MOTHER	NO	MOTHER	No	
3	FATHER	YES			
	MOTHER	YES	FATHER	YES	
	FATHER	YES			NO
	MOTHER	YES	MOTHER	YES	
4	FATHER	NO			
	MOTHER	NO	FATHER	No	
	FATHER	NO	MOTHER		YES
	MOTHER	NO		No	
5	FATHER	NO	FATHER		
	MOTHER	YES		YES	
	FATHER	YES			YES
	MOTHER	NO	MOTHER	No	
6	FATHER	YES			
	MOTHER	YES	FATHER	YES	
	FATHER	YES			NO
	MOTHER	YES	MOTHER	YES	
7	FATHER	YES			
	MOTHER	NO	FATHER	No	
	FATHER	YES			YES
	MOTHER	YES	MOTHER	YES	
8	FATHER	NO			
	MOTHER	NO	FATHER	No	
	FATHER	NO			YES
	MOTHER	NO	MOTHER	No	
9	FATHER	NO			
	MOTHER	NO	FATHER	No	
	FATHER	NO			NO
	MOTHER	NO	MOTHER	No	
10	FATHER	YES			
	MOTHER	YES	FATHER	YES	
	FATHER	YES			YES
	MOTHER	YES	MOTHER	No	

11	FATHER	NO			
	MOTHER	YES	FATHER	No	
	FATHER	NO			NO
	MOTHER	NO	MOTHER	YES	
12	FATHER	YES			
	MOTHER	NO	FATHER	No	
	FATHER	NO			YES
	MOTHER	NO	MOTHER	No	
13	FATHER	YES			
	MOTHER	NO	FATHER	No	
	FATHER	YES			YES
	MOTHER	NO	MOTHER	YES	
14	FATHER	NO			
	MOTHER	YES	FATHER	YES	
	FATHER	NO			YES
	MOTHER	YES	MOTHER	No	
15	FATHER	YES			
	MOTHER	YES	FATHER	YES	
	FATHER	NO			YES
	MOTHER	YES	MOTHER	No	
16	FATHER	YES			
	MOTHER	NO	FATHER	YES	
	FATHER	NO			NO
	MOTHER	YES	MOTHER	YES	
17	FATHER	YES			
	MOTHER	YES	FATHER	YES	
	FATHER	NO			YES
	MOTHER	NO	MOTHER	YES	
18	FATHER	NO			
	MOTHER	YES	FATHER	YES	
	FATHER	NO			YES
	MOTHER	NO	MOTHER	No	
19	FATHER	NO			
	MOTHER	YES	FATHER	YES	
	FATHER	NO			NO
	MOTHER	YES	MOTHER	YES	
20	FATHER	NO			
	MOTHER	YES	FATHER	No	
	FATHER	YES			YES
	MOTHER	NO	MOTHER	YES	
21	FATHER	YES			
	MOTHER	YES	FATHER	YES	
	FATHER	NO	MOTHER		NO

	MOTHER	YES		YES	
22	FATHER	NO			
	MOTHER	YES	FATHER	No	
	FATHER	YES			YES
	MOTHER	NO	MOTHER	YES	
23	FATHER	NO			
	MOTHER	NO	FATHER	No	
	FATHER	NO			NO
	MOTHER	NO	MOTHER	No	
24	FATHER	YES			
	MOTHER	YES	FATHER	No	
	FATHER	NO			YES
	MOTHER	YES	MOTHER	No	
25	FATHER	NO			
	MOTHER	NO	FATHER	YES	
	FATHER	NO			YES
	MOTHER	NO	MOTHER	No	
26	FATHER	NO			
	MOTHER	NO	FATHER	No	
	FATHER	YES			YES
	MOTHER	YES	MOTHER	YES	
27	FATHER	NO			
	MOTHER	NO	FATHER	No	
	FATHER	NO			
	MOTHER	NO	MOTHER	No	YES
28	FATHER	YES			
	MOTHER	YES	FATHER	YES	
	FATHER	NO			
	MOTHER	YES	MOTHER	No	YES
29	FATHER	YES			
	MOTHER	YES	FATHER	No	
	FATHER	YES			
	MOTHER	NO	MOTHER	YES	YES
30	FATHER	NO			
	MOTHER	YES	FATHER	No	
	FATHER	YES			
	MOTHER	YES	MOTHER	YES	NO
31	FATHER	YES			
	MOTHER	YES	FATHER	No	
	FATHER	YES			
	MOTHER	YES	MOTHER	No	YES
32	FATHER	YES			
	MOTHER	YES	FATHER	YES	

	FATHER	YES			
	MOTHER	YES	MOTHER	YES	NO
33	FATHER	YES			
	MOTHER	NO	FATHER	No	
	FATHER	YES			
	MOTHER	YES	MOTHER	YES	NO
34	FATHER	YES			
	MOTHER	YES	FATHER	YES	
	FATHER	NO			
	MOTHER	YES	MOTHER	No	NO
35	FATHER	NO			
	MOTHER	NO	FATHER	No	
	FATHER	YES			
	MOTHER	YES	MOTHER	YES	YES
36	FATHER	YES			
	MOTHER	YES	FATHER	YES	
	FATHER	NO			
	MOTHER	YES	MOTHER	YES	YES
37	FATHER	YES			
	MOTHER	YES	FATHER	No	
	FATHER	YES			
	MOTHER	YES	MOTHER	No	NO
38	FATHER	NO			
	MOTHER	YES	FATHER	No	
	FATHER	NO			
	MOTHER	YES	MOTHER	YES	NO
39	FATHER	NO			
	MOTHER	NO	FATHER	YES	
	FATHER	YES			YES
	MOTHER	YES	MOTHER	No	
40	FATHER	YES			
	MOTHER	NO	FATHER	YES	
	FATHER	NO			NO
	MOTHER	YES	MOTHER	No	
41	FATHER	YES			
	MOTHER	YES	FATHER	YES	
	FATHER	NO			YES
	MOTHER	YES	MOTHER	YES	
42	FATHER	YES			
	MOTHER	YES	FATHER	No	
	FATHER	YES			YES
	MOTHER	NO	MOTHER	No	
43	FATHER	YES	FATHER		

	MOTHER	YES		No	
	FATHER	NO			NO
	MOTHER	YES	MOTHER	YES	
44	FATHER	NO			
	MOTHER	NO	FATHER	No	
	FATHER	NO			NO
	MOTHER	NO	MOTHER	No	
45	FATHER	NO			
	MOTHER	YES	FATHER	YES	
	FATHER	NO			YES
	MOTHER	NO	MOTHER	YES	
46	FATHER	YES			
	MOTHER	YES	FATHER	YES	
	FATHER	YES			NO
	MOTHER	NO	MOTHER	No	
47	FATHER	YES			
	MOTHER	YES	FATHER	No	
	FATHER	NO			
	MOTHER	NO	MOTHER	No	YES
48	FATHER	NO			
	MOTHER	NO	FATHER	YES	
	FATHER	NO			NO
	MOTHER	YES	MOTHER	YES	
49	FATHER	YES			
	MOTHER	NO	FATHER	YES	
	FATHER	NO			NO
	MOTHER	NO	MOTHER	YES	
50	FATHER	YES			
	MOTHER	YES	FATHER	No	
	FATHER	NO			NO
	MOTHER	YES	MOTHER	No	
51	FATHER	NO			
	MOTHER	NO	FATHER	YES	
	FATHER	YES			YES
	MOTHER	YES	MOTHER	YES	
52	FATHER	NO			
	MOTHER	NO	FATHER	YES	
	FATHER	NO			NO
	MOTHER	NO	MOTHER	YES	
53	FATHER	NO			
	MOTHER	NO	FATHER	No	
	FATHER	YES			YES
	MOTHER	NO	MOTHER	No	

54	FATHER	YES			
	MOTHER	NO	FATHER	YES	
	FATHER	NO			YES
	MOTHER	NO	MOTHER	No	
55	FATHER	YES			
	MOTHER	NO	FATHER	No	
	FATHER	YES			NO
	MOTHER	YES	MOTHER	No	
56	FATHER	NO			
	MOTHER	NO	FATHER	YES	
	FATHER	YES			YES
	MOTHER	YES	MOTHER	YES	
57	FATHER	NO			
	MOTHER	NO	FATHER	YES	
	FATHER	YES			YES
	MOTHER	NO	MOTHER	No	
58	FATHER	NO			
	MOTHER	YES	FATHER	YES	
	FATHER	NO			NO
	MOTHER	NO	MOTHER	YES	
59	FATHER	NO			
	MOTHER	NO	FATHER	YES	
	FATHER	NO			YES
	MOTHER	NO	MOTHER	No	
60	FATHER	NO			
	MOTHER	YES	FATHER	YES	
	FATHER	YES			NO
	MOTHER	YES	MOTHER	No	
61	FATHER	NO			
	MOTHER	NO	FATHER	YES	
	FATHER	NO			YES
	MOTHER	NO	MOTHER	No	
62	FATHER	YES			
	MOTHER	NO	FATHER	No	
	FATHER	NO			YES
	MOTHER	NO	MOTHER	YES	
63	FATHER	NO			
	MOTHER	YES	FATHER	YES	
	FATHER	YES			NO
	MOTHER	NO	MOTHER	No	
64	FATHER	NO			
	MOTHER	NO	FATHER	YES	
	FATHER	YES	MOTHER		NO

	MOTHER	YES		YES	
65	FATHER	YES			
	MOTHER	YES	FATHER	No	
	FATHER	NO			YES
	MOTHER	YES	MOTHER	No	
66	FATHER	NO			
	MOTHER	YES	FATHER	No	
	FATHER	NO			NO
	MOTHER	YES	MOTHER	YES	
67	FATHER	YES			
	MOTHER	NO	FATHER	No	
	FATHER	NO			YES
	MOTHER	NO	MOTHER	No	
68	FATHER	NO			
	MOTHER	YES	FATHER	No	
	FATHER	YES			YES
	MOTHER	YES	MOTHER	No	
69	FATHER	YES			
	MOTHER	NO	FATHER	No	
	FATHER	YES			YES
	MOTHER	YES	MOTHER	No	
70	FATHER	NO			
	MOTHER	NO	FATHER	YES	
	FATHER	NO			YES
	MOTHER	NO	MOTHER	YES	
71	FATHER	NO			
	MOTHER	NO	FATHER	YES	
	FATHER	YES			YES
	MOTHER	YES	MOTHER	No	
72	FATHER	YES			
	MOTHER	NO	FATHER	No	
	FATHER	NO			YES
	MOTHER	NO	MOTHER	No	
73	FATHER	NO			
	MOTHER	NO	FATHER	No	
	FATHER	NO			YES
	MOTHER	NO	MOTHER	YES	
74	FATHER	NO			
	MOTHER	NO	FATHER	YES	
	FATHER	NO			YES
	MOTHER	NO	MOTHER	YES	
75	FATHER	NO			
	MOTHER	YES	FATHER	No	YES

	FATHER	YES			
	MOTHER	YES	MOTHER	No	
76	FATHER	NO			
	MOTHER	NO	FATHER	No	
	FATHER	YES			YES
	MOTHER	YES	MOTHER	YES	
77	FATHER	YES			
	MOTHER	YES	FATHER	No	
	FATHER	NO			NO
	MOTHER	YES	MOTHER	No	
78	FATHER	NO			
	MOTHER	YES	FATHER	No	
	FATHER	YES			YES
	MOTHER	NO	MOTHER	YES	
79	FATHER	YES			
	MOTHER	YES	FATHER	YES	
	FATHER	YES			NO
	MOTHER	YES	MOTHER	No	
80	FATHER	YES			
	MOTHER	YES	FATHER	YES	
	FATHER	YES			NO
	MOTHER	YES	MOTHER	No	
81	FATHER	YES			
	MOTHER	YES	FATHER	No	
	FATHER	NO			NO
	MOTHER	YES	MOTHER	No	

APPENDIX B

HAIR		**SKIN**		**EYES**	
FC	**MC**	**FC**	**MC**	**FC**	**MC**
PP	PP	PA	PA	pp	PP
AP	PP	AP	PP	PP	PP

PA	AA	PA	AA	AA	AP
PA	PA	PA	PA	pp	PP
PA	PA	PA	PA	pp	PP
AA	AA	AA	AA	pp	PP
AA	PA	AA	PA	PA	PA
PP	PP	PA	PA	pp	PP

PA	AP	PA	PA	PA	AA
AA	PA	AA	PA	PP	PP
AA	AA	AA	AA	AA	AP
PA	PA	PA	PA	PA	PA
PA	AA	PA	AA	AA	AP
AP	PP	AP	PP	PP	PP

AA	AA	AA	AA	AA	AA
AA	PA	AA	PA	PA	PP
PA	PA	PA	PA	PA	PP
AA	**AA**	AA	**AA**	**AA**	**AP**
AA	PA	AA	PA	PA	PA
PA	AA	PA	AA	PP	AP

AA	AA	AA	AA	AP	AP
PA	PA	PA	PA	PP	PP
PA	AA	PA	AA	AA	AA
AA	AA	AA	AA	AP	AP
PA	PA	PA	PA	PP	PP

PP	AP	PP	AP	PP	AP
AA	AA	AA	AA	AA	AP
AA	AA	AA	AA	AA	AA
PA	PA	PA	PA	PP	PP
PA	AA	PA	AA	AA	AA

AA	AA	AA	AA	AA	AP
PP	AP	PP	AP	PP	AP
AA	AA	AA	AA	AA	AP

PA	PA	PA	PA	PP	PP
AA	AA	AA	AA	AP	PP
AA	AA	AA	AA	PP	PP
PA	AA	PA	AA	AP	AP

AA	AA	AA	AA	AP	AP
PP	AP	PP	AA	PP	AP
AA	AA	AA	AA	AP	AP
PA	PA	PA	PA	PP	PP
PA	AA	PA	AA	PP	AP

AA	AA	AA	AA	PP	PP
AA	AA	AA	AA	AP	AP
PA	AA	PA	AA	AP	AP
PP	PP	PP	PP	PP	PP
PA	AA	PA	AA	PP	PP

AA	AA	AA	AA	AA	AP
AA	AA	AA	AA	AP	AP
PA	PA	PA	PA	PP	PP
AA	AA	AA	AA	AP	AP

PA	AA	PA	AA	PP	AP
AA	PA	AA	PA	AP	PP
AA	AA	AA	AA	AP	AP
AA	AA	AA	AA	AP	AP
PA	PA	PA	PA	PA	PP
AA	AA	AA	AP	PA	AP

PA	PA	PA	PA	PA	PA
AA	AA	AA	AA	AP	AP
PP	AP	PP	AP	AP	PP
AA	AA	AA	AA	PP	PP
PA	PA	PA	PA	PP	PP

AA	AA	AP	AA	PP	PP
PA	AA	PA	AA	PP	PP
AA	AA	AA	AA	PP	PP

M= Mother F= Father C= Children P=Present A=Absent

Source: Albino foundation Lagos, Nigeria 2018