Ergodic Markov Chain: A Proper Marker for Identification of a Trait

S.B. Olarinoye and E.T. Jolayemi

Abstract--- *In this paper we present the use of Ergodic Markov Chain as a marker in identification of a trait. We tried to show the Ergodicity properties of a Markov chain and its application on the three markers (Hair Color, Skin Color and Eye Color). The Marker (Eye Color) shows the highest correct classification of 75% and 89% respectively for father to child and mother to child inheritance case and with a smaller % misclassification of 25% and11% respectively. And also the expected number of steps (generations) to return to the given state (Present of Eye Color) is approximately one (1) an indication of Aperiodic state. Furthermore, the Markov chain (Eye Color) consider show that all states communicate, the chains are finite, irreducible, only one class exist and positive recurrent. Hence, the Markov chains (Eye Color) satisfies Ergodic properties which is an indication of good marker for identification of a trait.*

Keywords--- *Ergodic, Generation, Inheritance, Markers, Markov Chain, Trait.*

I. INTRODUCTION

Qualities are available in numerous assortments, known as alleles. Physical cells contain two alleles for each pleasant, with one allele gave via each discern of a dwelling being. Regularly, its miles tough to discern out which two alleles of a best are available inner a creature's chromosomes dependent on the outward look of that living being. In any case, an allele that is included up, or not communicated by way of a creature, can in any case take delivery of to that living being's posterity and communicated in a later age. A feature in a single age can be received, but now not ostensibly obvious earlier than additional ages.

Mendel, Gregor. (1866) become the principal man or woman to depict the way in which features are given starting with one age then onto the subsequent (and once in a while skip ages). Through his rearing investigations with pea flowers, Mendel constructed up 3 standards of legacy that portrayed the transmission of hereditary traits earlier than qualities have been even determined. These are Law of Segregation, Law of Independent Assortment and Law of Dominance. Today, researchers utilize "phenotype" to allude to what Mendel named a existence form's "outside likeness," and "genotype" to allude to what Mendel named a lifestyles shape's "inward nature." Indeed, Mendel's investigations exposed that phenotypes can be included up in a single age, simply to reappear in ensuing ages. Mendel in this manner thought approximately how lifestyles bureaucracy safeguarded the genetic cloth related with these attributes in the interceding age, when the traits had been escaped see. (Dennis, 2012).

A hereditary marker is a first-class or DNA association with a recognised region on a chromosome that may be utilized to distinguish people or species. It tends to be depicted as a spread (which can also emerge because of transformation or adjustment inside the genomic loci) that may be watched.

S.B. Olarinoye, Dept. of Statistics, Kwara State Polytechnic Ilorin.
E.T. Jolayemi, Dept. of Statistics, University of Ilorin, Ilorin.

Received: 10 Feb 2019 | Revised: 09 Mar 2019 | Accepted: 30 Mar 2019

Hereditary markers can be applied to bear in mind the connection between an obtained disorder and its hereditary reason (as an instance, a particular alternate of a high-quality that outcomes in an inadequate protein). It is realized that bits of DNA that lie close to one another on a chromosome will in preferred be acquired together. This belongings empowers the utilization of a marker, which could then be capable of be applied to determine the exact legacy example of the excellent that has not but been truly confined. Hereditary markers are applied in genealogical DNA testing for hereditary parentage to determine hereditary separation between human beings or populaces Autosomal markers are utilized for all circle of relatives.

Phenotype is the detectable physical or biochemical traits of an person dwelling being, dictated through each hereditary makeup and herbal influences, as an instance, tallness, weight and pores and skin shading. A phenotype effects from the announcement of a creature's hereditary code, its genotype, just because the effect of natural components and the collaborations between the 2.

The connection amongst genotype and phenotype has often been conceptualized via the accompanying courting:

$$\text{genotype (G)} + \text{circumstance (E)} \rightarrow \text{phenotype (P)}$$

The genotype–phenotype differentiation is attracted hereditary features. "Genotype" is a residing being's complete inherited facts. "Phenotype" is a lifestyles shape's genuine watched properties, for example, morphology, development, or behavior. This qualification is fundamental within the investigation of legacy of attributes and their advancement. (Laird and Lange 2011).

The difficulty of unaided relative revelation given an assortment of own family images; acquire the size of the family, just as the visual appearance and social process of each relative changed into tended to with the aid of Dai et al. (2015) utilising a non-lab take a look at . Therefore, that they had the option to perceive a similar person throughout diverse pictures. They proposed a solo EM-fashion joint derivation calculation with a probabilistic method that fashions, character and activity assignments for each single recognized face, along associated pairwise connections among them. Their trials show how joint deduction of both character and activity (over all photographs all the even as) beat self-sufficient critiques of every. The joint induction moreover improves the capacity to understand a similar person across diverse photos.

II. METHODOLOGY

In a deterministic world, it's far acceptable to comprehend that now and again haphazardness can at present happen. A stochastic procedure is the unique inverse of a deterministic one, and is an abnormal technique which could have several consequences as time advances. This implies at the off chance that we know an underlying circumstance for the manner and the capability through which it's far characterised, we will discuss likely effects of the method. One of the maximum commonly pointed out stochastic techniques is the Markov chain. Markov chain models were the maximum typically applied ones in the research of arbitrary variances inside the hereditary characteristics pieces of populaces over a long time. Other than being a useful hypothetical device, Markov chains have given as an alternative agreeable hypothetical clarifications to some watched because a long time in the past run marvels diagnosed with the hereditary shape of populaces.

Received: 10 Feb 2019 | Revised: 09 Mar 2019 | Accepted: 30 Mar 2019

Ergodicity Theorem

The status quo of Markov chain hypothesis is the Ergodicity Theorem. It units up the conditions beneath which a Markov chain may be tested to determine its steady country conduct. An Ergodic Markov chain homes are:

- Very last (All states impart),
- Fine intermittent, and
- Aperiodic

Markov Chain

Stochastic process (discrete time): $\{x_1, x_2, \ldots.,\}$. Consider a discrete time stochastic process with space. $X_n \varepsilon [0,1,2...]$

Markovian Property

$P\{X_{n+1}=j|X_n=i, X_{n-1}=i_{n-1},\ldots X_o=i_o\}$

$= P\{X_{n+1}=j|X_n=i\} = P_{ij}$ (1)

P_{ij} is the transition probability, the probability of marking a transition from i to j

$$P_{ij} = \begin{bmatrix} p_{00} & p_{01} & p_{02} & - & - \\ p_{10} & p_{11} & p_{12} & - & - \\ - & - & - & - & - \\ p_{i0} & p_{i1} & p_{i2} & - & - \\ - & - & - & - & - \end{bmatrix}$$

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

More generally, we define the n-step transition probabilities

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

and the n-step transition matrix

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

We can now generalize (2).

Chapman-Kolmogorov Equations

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

$$P_{ij}^{m+n} = P[X_{m+n} = j \mid X_0 = i] \\ = \sum_{k \varepsilon S} P_{ik}^{(m)} P_{kj}^{(n)} \qquad (3)$$

The above equation is called the **Chapman-Kolmogorov Equations:**

Def.1 A state is said to be recurrent if, any time that we leave the state, we will return to that state in future with probability one.

Def. 2 A Markov Chain is irreducible if all the states communicate with each other (there is only one class)

Def.3 State i and j communicate if they are accessible from each other, written as $i \leftrightarrow j$

Def.4 Ergodic State is a state that is positive recurrent and aperiodic

Def. 5 A Markov Chain is ergodic if all states are ergodic.

Def. 6 State i is aperiodic if it can only return to itself with transition number equal one

Def. 7 A Positive recurrent state is defined by f_i =1 and $m_i < \infty$ (expected time until the process returns to state i is finite)

Def. 8 A Null recurrent state is defined by f_i =1 and $m_i = \infty$ (expected time until the process returns to state i is infinite)

Recurrent State property

Positive recurrent and Null recurrent states are called distinguishing property of recurrent states. Positive recurrent state i is 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 defined 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 and stay there forever. We stay at j for n steps with probability

$$[P_{ij}]^n \to 0, \qquad (5)$$

as $n \to \infty$, 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 we first visit j is given by

$$\mu_{ij} = 1 + \sum_{r \neq j} p_{ir} \mu_{rj}, \qquad (6)$$

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

Table 1

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

Transforming into Markov Chain

Let X_n be the appearance of a trait in generation n and suppose whether a trait will appear in next generation(Children) depends on whether it will appear current generation(father) and past generations (Grandparent) then we have $P(X_{n+1} \mid X_{n,} X_{n-1,}......, X_{1)} = P(X_{n+1} \mid X_{n,} X_{n-1})$ This process is not a first other Markov Chain.

Geometric Distribution

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

$$P_j\{N_j<\infty\}=\begin{cases}1 & \text{if } F_{jj}<1,\\ 0 & \text{if } F_{jj}=1.\end{cases} \qquad (7)$$

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 i, and defined by

$$R_{ij} = E_i(N_j) \qquad (8)$$

We have

$$R_{jj} = \frac{1}{(1-F_{jj}),} \qquad (9)$$

(here, 1/0= ∞, 0 ∞=0)

and

$$R_{ij} = F_{ij} R_{jj}, \quad \text{if } i \neq j \qquad (10)$$

Case 1 (j is a recurrent state)

Suppose j is a recurrent state, then F_{jj}=1; then, (9) implies R_{jj}= +∞ also if j can be reached from i, then F_{ij}>0 and R_{ij}= ∞ again. If, on the other hand, j cannot be reached from i, then F_{ij} =0 and R_{ij} =0. Hence, for j recurrent,

$$R_{ij} = \begin{cases}0 & \text{if } F_{ij}=0\\ +\infty & \text{if } F_{ij}>0\end{cases} \qquad (11)$$

Case 2 (j is a transient state)

Suppose j is transient, then $F_{ij}=0$; then, (11) implies $R_{ij}=0$, j cannot be reached from i; therefore,

$$F_{ij}=0$$

$$R_{ij}=0 \qquad 2$$

III. APPLICATION

The method of identification of a trait(s) using Markov chain approached is flexible and thus applicable to all living organisms. The data used was collected from the Albino foundation with two generations of the selected families observed which includes the colour of the eye, skin and hair.

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

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

Mother & Child

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

Table 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 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: 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	$\begin{bmatrix}0.70 & \\ & 0.30\end{bmatrix}$	$\begin{bmatrix}3 & \\ & 1\end{bmatrix}$	$\begin{bmatrix}0.87 & \\ & 0.13\end{bmatrix}$	$\begin{bmatrix}8 & \\ & 1\end{bmatrix}$

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 state are reachable from each other (i.e there are no redundant or dead-end state). And with a single class of communication for each.

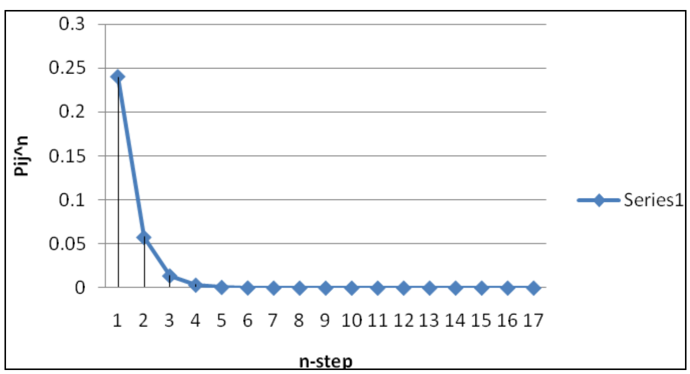
Periodicity

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

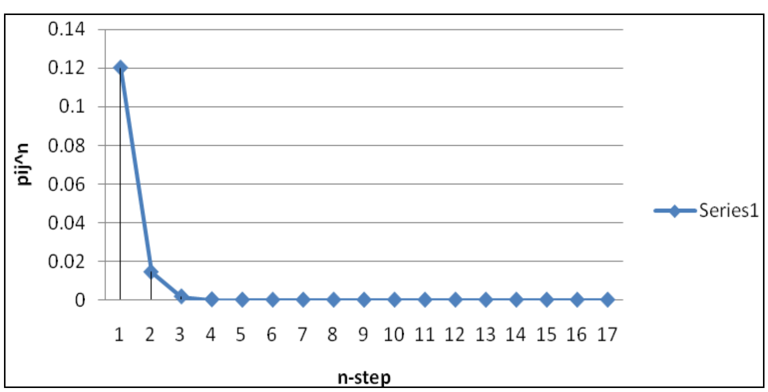
Recurrent State Property

from (7) $[P_{ij}]^n \to 0$, as $n \to \infty$ we have the following graphs for the two chains,

Father & Child



Mother & Child



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.

Expected Number of Transition to return to given State

We 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

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

(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) <∞ . We have the following system

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

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

IV. CONCLUSION

This paper considers Ergodic theory which is the foundation of Markov chain theory. We tried to show the Ergodicity properties of a Markov chain and the recurrent type that can be use as a marker. The Markov chains consider show that all states communicate, the chains are finite, irreducible, only one class exist and it also follows that the states are positive recurrent. Hence, from the results of the work its shows that Ergodic Markov chain (properties) is the only good marker that can be used in the identification of a trait.

REFERENCES

[1] Dai, Q., Carr, P., Sigal, L and Hoiem, D. (2015): "Family Member Identification from Photo Assortment" *IEEE winter conference on applications of computer imaginative and prescient Waikoloa, IEEE Hl,* u.S.A. 124-154.

[2] Cinlar, Erhan (1975): Introduction to Stochastic procedures, Prentice-hall, Inc., Englewood cliffs, N.J: Prentice-corridor, [1974, c1975] americaA. 25-seventy six

[3] Kimmel, G and Shamir. R (2005): "A Block-Free Hidden Markov mannequin for Genotypes and Its utility to ailment organization". *Journal of Computational Biology* 12(10):1243-60.

[4] Laid Nam, M and Lange Christopher (2011) "The necessities of latest actual hereditary characteristics" division of Biostatistics harvard college 1-6.

[5] Mendel, Gregor. (1866). Http://www.Esp.Org/associations/hereditary features/vintage style/gm-sixty 5.Pdf

[6] Olarinoye, S.B. What is extra, Jolayemi, E.T (2016): The utilization of alternate chance Matrix in inspecting features. Nigerian Statistical association (NSA) every year conference Abuja State 34-sixty seven

[7] O'Neil, Dennis. "Principal ideas of Genetics: Mendel's Genetics." general concepts of Genetics: Mendel's Genetics. N.P., n.D. Internet. 03 Nov. 2012 <http://anthro.Palomar.Edu/mendel/mendel_1.Htm>.

[8] Robert Plomin and Denise Daniels (2011). *Global Journal of Epidemiology*, forty (three), 563–582.

Received: 10 Feb 2019 | Revised: 09 Mar 2019 | Accepted: 30 Mar 2019