Picard Approximation of Stochastic Differential Equation and its Application On Gene Pool Model

Mary Chriselda A

Abstract—. In this paper the Hardy-Weinberg equation on the frequency of alleles with its contrivances serve as a stochastic model which has been used along with the Markovian Property to prove the frequency of the alleles at the locus of a particular population in genetics.
This Gene Pool Model has been converted to a stochastic differential equation whose existence and uniqueness of the solution has been proved using Picard-Lindelof Theorem.
This provides an altogether a new way of delving into the solution which is globally consistent. The aim of this research paper is to is to firstly implicate the Markovian Property in one of the laws which is the cornerstone of population genetics. Secondly, converting the Markov process to a stochastic differential equation proposing a new global solution to the diffusion approximation as well as proving the convergence of the sequence. Thirdly, the results are validated and computed using the chi- square goodness of fit. The proposed statistical methodologies used for validation of the result is the R-software.

Index Terms— BROWANIAN STOCHASTIC DIFFERENTIAL EQUATIONS, PICARD APPROXIMATION, LIPSCHITZ CONDITION, OPTIMUM SOLUTION, MARKOVIAN PROPERTY, LINEAR REGRESSION, ITO INTEGRAL

1 INTRODUCTION

Real world phenomenas can be observed as a pattern which can serve as a model when elaborated on can bring about a change to the society. The Markov model under the stochastic processes in probability theory was named after the Russian mathematician Andrey Markov.
How is the stochastic process visualized in nature? This requires translating assumptions relating to the relationship between the causes of its behaviour and effect into contrivances with its limits to determine its nite dimensional distributions.
In the field of genetics, where this property has wide application can be served as one model. Taking the Hardy Weinberg principle which states that in large randomly mating population, allelic frequencies will remain constant barring any sort of evolutionary forces. Now if this principle with its contrivances as mentioned can be made into a model one can easily predict the frequency of the recessive allele and hence the probability of population frequency of the carriers of the recessive diseases can be predicted.

1.1 LITERATURE SURVEY

The commonly referred mathematics textbooks (e.g Ewens 2004,Burger 2000,Rice 2004) mention the evolutionary forces involved in the gene pool model.The most basic of them all is the Wright Fischer Model which I have considered for the random genetic drift without any evolutionary forces .I have modelled the Gene Pool Model generalizing the Wright Fischer Model for two alleles at same locus and hence, predict the correlation between the allelic frequency and the population. I have also used Picard Approximation to the Brownian Stochastic Differential Equation. In the present paper I have tried to introduce a new global approach. This approach is mathematically more transparent than Kimura's scheme.
For a basic two allele case proposed by Kimura (1955), which he later tried to implicate on several alleles (1955,1956). His solution being local, the probability distribution not summing to 1 made it very difficult to read all the quantifying properties of this process.
This paper has been organized as follows: Section 1 is the Introduction which is subdivided as 1.1 :Literature Survey ,Section 2 deals with Markov Process subdivided as Gene Pool Model ,Section 3 deals with the Diffusion Approximation, Section 4 deals with the Stochastic Differential Equation subdivided as 4.1 Existence and Uniqueness of the Solution 4.2 Picard Iteration Section 5 deals with the Condition for Solution Existence and Uniqueness, Section 6 deals with the proof using Picard's Methods subdivided as 6.1,6.2,6.3 -proof for drift coefficient to be Lipschitz continuous, existence of solution, proving both the drift and diffusion coefficient continuous, Section 7 deals with Result Analysis where subsection 7.1 deals with the Global Sensitivity Analysis using Regression,7.2 deals with the statistical methods employed to verify the result, Section 8 is Conclusion ,Section 9 is Acknowledgement followed by the References.

2 MARKOV PROCESS

A stochastic process say $\{X_n\}$where nεN takes values of a countable set namely the state space.
Definition: If the conditional probability distribution C_{i+1} is independent of the states of the system in steps 0 to i-1 and is only dependent on t_n
$p[X_{n+1}=i_{n+1} \mid X_n=i_n, X_{n-1}=i_{n-1}.....X_1=i_1, X_0=i_0]=p[X_{n+1}=i_{n+1} \mid X_n=i_n]$

1.The condition $p[X_n=i_n ...X_0=i_0]>0$ will be assumed each time a conditional probability will be considered.
2. All the models will be homogenous i.e. Conditional probability $p[X_{n+1}=j \mid X_n=i]$ is independent on time such that $n \varepsilon N_0$ i.e $p[X_{n+1}=j \mid X_n=i]=p[X_{m+1}=j \mid X_m=i]$ for all m,n ε N

The specifications are as follows:
1. Initial Probability $a^0=a^0{}_{i:i\varepsilon s}$ $a^0{}_{i:}=p[X_0=i]$
2. Transition Probability $p_{ij}=p[X_{n+1}=j \mid X_n=i]$

Theorem 2.1

if there is σ-algebra F and probability p on R such that
1. Each of g_n where n belongs to N is a random variable.
2. Sequence $\{g_n\}$ where n belongs to N is independent.

2.1 Gene Pool Model

Definition 2.2

The Hardy Weinberg law states that the allelic frequencies over a finite population obtained by random mating is constant barring any evolutionary forces. This can be seen with the help of a result computed using the R software where g is the genotype of the alleles.

```
data:  g
N11 = 1, N12 = 1, N22 = 1, N1 = 3, N2 = 3, p-value = 1
```

Fig 1 {HWE TEST}

$p^2+q^2+2pq=1$ ……(2.1)

$p+q=1$

where p=frequency of dominant allele
q=frequency of recessive allele

proof:

Let X(t) define a's in allele pool of generation t in a finite population. If the definition can be assumed as
1.Size of the population is constant
2.Each generation is obtained by independent random mating
3. No evolutionary forces (selection, migration, mutation)
The above three postulates derived from Hardy Weinberg law can be converted to a Markov Model.
If X(t) is a non -negative integer.
Let N denote the finite size of the population which is always less than 2N.
Assumptions:
the state space={0,1,2.....2N}
t denote current time
$\{X_0=i_0.....X_t=i\}$ which denotes i allele of genotype A in generation t
From the assumptions generation(t+1) is created by N independent random mating of generation t parents. It is known that the genotype is obtained by random mating of two independent samples from the allele pool of generation t. Thus, allele pool (t+1) is created by 2N independent random samples. The frequency of the recessive allele is obtained by finding the probability
$X_t=i$
frequency of the allele a=i/2N
frequency of the allele A=1-i/2N
Hence, the transition probability can be expressed as follows:

$p[X_{t+1}=j \mid X_t=i]= ({}^{2N}C_J)(\frac{i}{2N})^J (1- \frac{i}{2N})^{2N-J}$ ……(2.2)

X_t can take values from 0 to 2N i.e if either A or a will disappear or will remain there for a future period of time.This basic model is therefore expressed in terms of the binomial distribution which verifies the Hardy Weinberg Law

```
> HWE.test(g)

        ----------------------------------
        Test for Hardy-Weinberg-Equilibrium
        ----------------------------------

Call:
HWE.test.genotype(x = g)

Raw Disequlibrium for each allele pair (D)

           A          a
A          -0.08333333
a -0.08333333

Scaled Disequlibrium for each allele pair (D')

           A          a
A          -0.3333333
a -0.3333333

Correlation coefficient for each allele pair (

           A         a
A          0.3333333
a 0.3333333

Observed vs Expected Allele Frequencies

        Obs  Exp     Obs-Exp
A/A 0.3333333 0.25  0.08333333
a/A 0.1666667 0.25 -0.08333333
A/a 0.1666667 0.25 -0.08333333
```

Fig 2 {CHI-SQUARE TEST FOR ALLELES}

3. DIFFUSION APPROXIMATIONS

From equation 2.2 obtained in the above section I can derive for the expected time A or a to be either fixed at $X_t=2N$ or extinct at $X_t=0$ given its initial condition X_0.Now using the idea proposed by Wright(1945) to rescale time and population size.
$t=\frac{n}{2N}$ and $[Y_t=\frac{Xt}{2N}]$ then considering N tends to infinity. Once I rescale it I end up getting the discrete Markov Chain Y_t as { 0,1/2,.....1 } with t=1 corresponding to 2N.
Similarly,I can obtain $X_0=p=\frac{i0}{2N}$ .Now, a basic idea of approach that I have used here is to find the kth central moments of this binomial distribution using the recurrence formula.
By definition, the kth central moment u_K is defined by

$u_K=\{E(Y-E(Y)\}^K$ .....(3.1)

For the binomial distribution B(n,p) I have Y=r, E(Y)=np,q=1-p

$u_k=\sum_{r=0}^{n}(r-np)^k\binom{n}{r}p^r q^{n-r}$ ..... (3.2)

Differentiating equation with respect to p where q=1-p ,I get

$\frac{duk}{dp}=\sum_{r=0}^{n}\binom{n}{r}[(-nk)(r-np)^{k-1} p^r q^{n-r}]+(r-np)^k[rp^{r-1} q^{n-r}-p^r(n-r)q^{n-1}]$ (3.3)

$=-nku_{k-1}+\frac{1}{pq}\sum_{r=0}^{n}\binom{n}{r}p^r q^{n-r}(r-np)^{k+1}$ (3.4)

$=-nku_{k-1}+\frac{1}{pq}u_{k+1}$ (3.5)

Therefore, $\frac{duk}{dp}=-nku_{k-1}+\frac{1}{pq}u_{k+1}$ (3.6)

Using the recurrence formula

P(t+1)-P(t) P(t) np-(1-p)-t 1-p+(1-p)t

$$\frac{P(t+1)-P(t)}{P(t)} = \frac{np-(1-p)-t}{1-p+(1-p)t} \quad (3.7)$$

Now, partitioning the above equation t to t+1 and k portions.Applying k tending to infinity I get the following result

$$\frac{1}{z}\frac{dz}{dx} = \frac{r+x}{b0+b1x} \quad (3.8)$$

Now the following differential equation is obtained where x=-t,r=np-(1-p) ,b_0=(1-p),b_1=(p-1)

4. STOCHASTIC DIFFERENTIAL EQUATION

Definition:

Let a(x): $R^n \rightarrow R^n$ and b(x): $R^n \rightarrow R^{mn}$ be measurable functions (vector and matrix valued respectively),
W an m-dimensional Winer Process and X_0 and L_2 random variable in R^n ,independent of W.Then on R^n-valued stochastic process X on R^+ is a solution to the autonomous stochastic differential equation
dX=a(X)dt+b(X)dW ,X(0)=X_0 ……..(4.1)
when with probability 1, it is equal to the corresponding Ito process,
X(t) =X_0+ $(\int_0^t a(X(s))ds))+\int_0^t b(X(s))dW$……(4.2)
The a term is called the drift coefficient and the b term is the diffusion coefficien

4.1 EXISTENCE AND UNIQUENESS OF SOLUTION

From the equation (3.8) I can form the stochastic differential equation as defined in the section above as follows
From the definition,
$\frac{dz}{dx}$=kz ……..(4.3)
dz=a(z,t)dt+b(z,t)dW$_t$……(4.4)
The intial conditions
z(0)=z_0 ……..(4.5)
when W_t ,b≥0
The above equation represents a stochastic differential equation

Theorem 4.1:

Let X_0,a,b and W be as defined in definition 4.1 and let a and b be uniformly Lipschitz continuous.Then there exists a square-integrable non-anticipating X(t) which solves
dX=a(X)dt+b(X)dW with the initial conditions X_0 and this solution is unique

Theorem 4.2:

This theorem holds true for multi-dimensional stochastic differential equations, provided a and b are uniformly Lipschitz inthe Euclidean norms.

4.2 PICARD'S ITERATION

Definition:

Obtaining the initial condition as $z_0(t)=z_0$.Then computing approximations $z_1(t)$, $z_2(t)$ and so on via the following recursion:

$$z_{n+1}(t)=z_0+\int_{t0}^{t} f(z_n(\tau),\tau)d\tau \quad (4.6)$$

This iteration which can be used to find a solution for linear differential equation can be shown to converge to unique solution

$$\lim_{n\to\infty} z_n(t)=x(t) \quad (4.7)$$

provided f(x,t) is continuous in both arguments and Lipschitz continuous in the first argument.

5. CONDITION FOR SOLUTION EXISTENCE AND UNIQUENESS

Let us consider the stochastic differential equation
dz=a(z,t)dt+b(z,t)dW$_t$ (5.1)
Now using the Picard-Lindelof theorem in the above section on this stochastic differential equation.
The assumptions involving this :

- Continuity in both aruguments
- Lipschitz continuity in the first argument

6. PROOF USING PICARD'S ITERATIVE METHOD

Theorem 6.1

Assume that a : R→ R and b:R →R+ are uniformly Lipschitz ,that is, there exists a constant C<∞
such that for all x,y ε R,
|u(x)-u(y)| ≤ C|x-y| and (6.1)
|b(x)-b(y)| ≤ C|x-y| (6.2)
Then the stochastic differential equation has strong solutions.In particular for any standard Brownian Motion {W_t} t≥0, any admissible filtration F={F_t} t≥0 , and any initial value x ε R there exists a unique adapted process X_t=X_t^x with continuous paths such that I obtain equation (4.2)

6.1 PROVING THE DRIFT COEFFICIENT TO BE LIPSCHITZ CONTINUOUS

Proof:

Proving diffusion coefficient is Lipschitz continuous.
For this purpose ,considering b(X)=0 as constant treating equation (4.4) as an ordinary differential equation z′=a(z,t).
Now proving the uniqueness by taking two solutions which are continuous
\Now proving the uniqueness by taking two solutions which are continuous
Z_t= z_0+ $\int_0^t a$ (z(t)dt)+$\int_0^t b$dW and (6.3)

y_t= y_0+ $\int_0^t a$(y(t)dt)+$\int_0^t b$Dw (6.4)

Now finding the difference between the two solutions I obtain

y_t - z_t=$\int_0^t a\,(y(t)-z(t))dt$ (6.5)

$|y_t - z_t| = C\int_0^t |(y(t) - z(t))dt|$ (6.6)

This holds true stating the drift coefficient is Lipschitz continuous for all t≤0 for some constant C ≤ 0

6.2 PROVING THE EXISTENCE OF SOLUTION}

Proof:

Proving the existence of solutions using Picard's approximations

Now defining the sequence of the stochastic process obtained intially

$z_0(t)=z$ (6.7)

and

$Z_{n+1}(t)= z+ \int_0^t a\ (z(t)dt)+bW(t)$ (6.8)

To prove $z_n(t)$ converges uniformly on compacted time intervals

$|z_n - z_{n+1}| \leq MK^{n-1} h^n/n!$ for all n (6.9)

Using this ,the infinite series

$z_0+(z_1-z_0)+(z_2-z_1)+....\leq z_0+Mh+\frac{1}{2!}MKh^2.....\leq z_0+ \frac{M}{K}[e^{hk}-1]$ (6.10)

This equation is known to be convergent for all K,h,M.Thus by Iir strass-M test ,the series converges uniformly for[0,T]

6.3{PROVING FUNCTIONS OF BOTH THE DRIFT AND DIFFUSION COEFFICIENT CONTINUOUS}

Now using the same iterative method as in case of proving the drift constant to generate the approximate solutions.

$z_0\ (t)=z$ (6.11)

and

$z_{n+1}(t)=z+\int_0^t a\ (z_n(s)ds+\int_0^t b(z_n(s)ds)$ (6.12)

By the principle of induction the process $X_n(t)$ is defined and has continuous paths.

Now for each t≥0 the sequence of random variables converges to a random variable in z_t.

$E(z_1(t) - z_0(t))^2 \leq M$ and t≤T, M≤∞ (6.13)

and t≤ T , M ≤∞

for t bounded in [0,T]

\Now by hypothesis, the functions of a and b are uniformly Lipschitz for C≤0

7. RESULT ANALYSIS

Hence, I have proved using the equation (6.6),(6.10),(6.13) the following criteria that must be adhered to

while following the assumptions in section 5 to use Picard-Lindelof theorem for the stochastic differential equations framed using the Gene Pool Model.The iteration converges and thus,the stochastic differential equation has a unique strong solution with the following conditions met as

- Functions of a and b grow at most linearly in z
- They are Lipschitz continuous in z

Having proved the above two postulates I can therefore say that the following stochastic differential equation has a unique strong solution.The problem that can usually be faced is the white noise formulation of the stocastic differential equation with the chain rule,non-linearities and even the existence of the solution. The integral cannot be defined as Riemann,Lesbesgue as white noise is unbounded and discontinuous so I had to employ the the Ito stochastic integral.I have used the Picard Iteration method to prove the existence and the uniqueness of the solution which was used only for Ordinary Differential Equations.

7.1 GLOBAL SENSITIVITY ANALYSIS USING LINEAR REGRESSION

Given the stochastic process z_t ,t ε [0,T] I get the stochastic differential equation

$dz_t=a(z,t)dt+b(z,t)dW_t$ (7.1)

such that the expectation of the square error random variable is obtained

$\int_0^t z_t - E(z_t)^2 dt$ (7.2)

is minimum.

Suppose I denote ˆz as the best fitting process.For this purpose I use the linear regression principle of least squares.Let us now replace ˆz by a line y=ct+d with c and d such that sum of the squares of errors is minimized.

$f(z,t)=E[\int_0^t z_t - \hat{z}_t)^2 dt$ (7.3)

The minimum obtained from this equation will provide the goodness of fit.

The proof is not explained in detail here as only the computated result is mentioned.

The following result is given below

$f(z,t)=\int_0^t \sigma^2 - \frac{1}{t} COV^2(z_t,W_t)]dt \geq 0$ (7.4)

To prove that

$COV^2(z_t,W_t)] \leq \sigma^2\ t$ (7.5)

I can use the Cauchy inequality

if u,v are two random variables

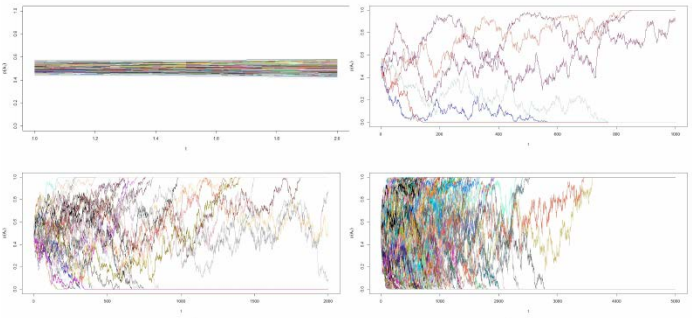
$COV^2(z_t,W_t)] \leq Var(u).Var(v)$ (7.6)

This implies that here the goodness of fit is always non-negative.Here I have used the R software to compute the following and as seen I get p=1 which accepts the hypothesis made.Here,I can conclude the positive correlative between z_t and W_t proving the strong correlation between our gene pool stochastic process and the Wiener(Brownian) Process.

7.2 STATISTICAL METHODOLOGIES TO VERIFY THE RESULTS

In the genetic model formulated, where the evolutionary forces are barred the allelic frequencies remain constant over a

finite population can be seen in the diagram below created using the R software. These graphs show the relation between the allelic frequencies and the time period of a population. From the graphs it can be observed as how the genetic drift vary for different population as explained below



GENETIC DRIFT FOR THE GENETIC MARKOV MODEL

Fig. 3: GRAPH OF THE ALLELIC FREQUENCIES VERSUS TIME

The above figure provides a bird's eye view of the genetic drift for a finite population barring evolutionary forces in accordance with the Genetic Model following the Markovian Property .As stated in the Hardy Weinberg law when the population sample becomes large and finite the genetic drift is constant as seen in the following diagrams as I vary the population size i.e for a small population size the genetic drift is very prominent in comparison to a larger sample where the drift is negligible.

Figure 1 HWE TEST AND Figure 2 CHISQ FOR ALLELES mentioned in section 1 of the Gene Pool Model computed using the R software verifies our result hypothesis.

8.CONCLUSION

In conclusion, I have tried to model a stochastic process Gene Pool Model and using the binomial distribution I have calculated the kth central moments resulting in a stochastic differential equation. The problem raised during this paper was to apply integral on the Gaussian white noise which being unbounded and discontinuous, could not satisfy the condition of Riemann or Lesbesgue integral letting me define the Itô integral now completely treating it as Brownian Stochastic Differential Equation. Using Picard's Iterative method

I could prove the existence and uniqueness of the solution obtained for a stochastic differential equation

Lastly for the Global sensitivity of our solution I have used the linear regression to compute the goodness of fit between the stochastic process I generated the Brownian motion. My answer obtained finally confirmed the relation between the two process, and hence the dependence of the both the drift and diffusion coefficient was directly proved. I have used the R-software to generate the graphs between the allelic frequencies and the population sample for a better understanding.

The main objective of this research paper is to find the probability of the allele frequency that could lead to genetic disorders like haemophilia, colour-blindness, sickle-cell anaemia. This paper satisfies the condition of the Hardy -Weinberg Law, propelling us to delve deep into the various aspects of how genetically caused diseases can be eradicated for the benefit of the society at large.

9. ACKNOWLEDGEMENT

It is with great humility and gratitude that I thank the Almighty for providing me with the prudence to write this research paper. I thank my parents for all the support they have rendered to me throughout this process. I also offer my gratitude to my alma mater, teachers, friends and well-wishers for all their encouragement and confidence in me.

REFERENCES

[1] Kimura, Motoo, "Solution of a process of random genetic drift with a continuous model",Proceedings of the National Academy of Science
[2] Rabiner, Lawrence R,"A tutorial on hidden Markov models and selected applications in speech recognition",Proceedings of the IEEE
[3] "Markov processes and learning models",Academic Press New York
[4] Griffiths, Robert C,"Lines of descent in the diffusion approximation of neutral Wright-Fisher models",Theoretical population biology
[5] Protter, Philip E,"Stochastic differential equations",Stochastic integration and differential equations,Springer
[6] Oksendal, Bernt,"Stochastic differential equations",Stochastic differential equations,Springer
[7] Talay, Denis,"Numerical solution of stochastic differential equations",1994,"Taylor Francis
[8] Stochastic differential equations and application, Mao, Xuerong,2007, Elsevier
[9] Picard approximation of stochastic differential equations and application to LIBOR models,"Papapantoleon, Antonis and Skovmand, David",
[10] Evolutionary game dynamics in a Wright-Fisher process,"Imhof, Lorens A and Nowak, Martin A",Journal of mathematical biology,Springer
[11] Patrick Alfred Pierce,Random processes in genetics,Mathematical Proceedings of the Cambridge Philosophical Society},
[12] Kimura, Motoo,"Solution of a process of random genetic drift with a continuous model",Proceedings of the National Academy of Sciences
[13] A First Course in Stochastic Processes,"By Samuel Karlin"
[14] Wagner Principles On Operational Research,"Wagner"
[15] Computer ALgorithm On Operation Research,"Johny Gillbet"
[16] "Ordinary and Partial Differential Equations",M d Raisighania

Mary Chriselda A
First Year B.Sc(EMS) Student
Email:chriseldaoliver23@gmail.com