Stochastic Process in Evolutionery Genetics ; Fokker Plank Diffusion Model-

* N. John Britto, Asst. Prof. in Statistics, Research scholar in statistics Annamalai University,King Nandhivarman College of Arts and Science college, Thellar, 604 406 TV Dt,

** Dr. V. Rajagopalan, Professor, Department of Statistics, Annamalai University, Tamilnadu, India

*** Dr. S. Sasikala, Asst. Prof. Department of Statistics, Periyar EVR College, Trichy , Tamilnadu, India

Abstract

In Evolution, the genetic make up of species and population may change over time, some traits may be lost , other new one arises, while some persist unchanged, Ewens Sampling when the Sample Size is small, monocious population of varying size that reproduces in each generation according Wright-Fisher Model, if Natural selection, Hardy Weinber Law and a Markov Chain application in Genetics. then Weinner process for speciation is discussed.

Key Words: Wright Fisher Model, Hardy Weinberg Equlibrium, Ewens Sampling Branching Process, Natural Selection, Fokker Plank Diffusion Model.

(1) Introduction

1.1 Stochastic Treatment: Discrete Process: Fisher Wright Model

Consider a diploid population having in each generation exactly N individuals, so there are 2N genes altogether at the locus in question. Suppose that, in some ways to be choosen at out discussion, the population reproduce itself to form a daughter generation. Suppose there also the onece the daughter generation is formed , no further reproduction is possible for the parent generation. If there is no mutation, Selection gene frequencies tend to remain steady. It is reasonable to the number of A1 genes is X(t) then the number X(t+1) is a binomial variate with index 2N and parameter X(t)/2N, explicitly given that , X(t) = i, the probability pij that X(t+1) =j is given Pij =(2N)(i/2N)j(1-i/2N)2N-j , X(i) is a Markovian Variable with transition Matrix j

P = {pij}

Wright Fisher (1931) introduced the concept of effective population number , (Ne) for a population composed of equal number of males and females which would result in equivalent inbreeding or variances due to genetic drift, the object of this paper is population is is not panmix, Ewens(1972) Sampling procedure consequently and Watterson(1976) diffusion model Approximation ,.

1.2 Hardy Weinberg Law and a Markov Chain in genetics

(1) According G. H. Hardy, Mendelian Population in a mixed population, Letter to the Editor, Science , N,S, Vol.28, (1908) pp:49-50, For a better under standing of the inequlibrium population , let us consider Hardy Weinberg Equilibrium (I,e) let us fix the gene frequencies, P and q (P+q=1) , gene frequencies of A and a , In each of the two selection an A genes is selected with probability of an offspring being AA is p2, the genotype Aa can occur in two ways is 2pq, aa with q2 probabilities.

Suppose the three genotypes AA, Aa. aa occur among males and females in the same ratio, u:2v;w; we shall suppose u+2v+w=1 say u;2v ;w the genotype frequencies ; p=u+v, q=v+w . The genotype AA, Aa. aa with probabilities u1=p2, 2v1=2pq, w1=q2

Hardy-Weinberg Equilibrium explics, Whatever the composition of the parent population produce an approximately stationary genotype distribution with unchanged gene frequencies. However e quliburium deviates.

1.3 :Wright Fisher and Moran Model

Evolution of the frequency of an allele A in a population which has the size N in each generation.

Yt ; Number of A allels in generation 't'

State Space E= { 0,1,2,…… 2N}

Since the possible numbr of allels lies between 0 to 2N.

The frequencies of the three genotypes in the nth generation are three random variables whose expected values are not in the ratio u1;2v1;w1 is p2,2pq,q2 i.e the actual values will vary generation to generation lead for Stocahstic Model in g .

(2) Sampling

2.1 Ewens(1972) Sampling Formula

Data on variability in samples are of a type that are might call configurational. That is to say, we can encourage that a sample of r genes is taken from among the 2N that are such that variablility doesn't necessarly implies that a particular small samples from a population will contain several allels at substantial frequencies . If a sample of such a structure is found a question arises about where the presence of some favourably frequent allels is due to selection or is merely what is to be expected from random sampling, that were present in the zygote and that it contains K allels, which are respectivitely represented n1,n2,….nk times in the sample collection {n1,n2,…nk} as a configuration, whenever it is to be assumed that n1>n2>….nk, probability that a random samples of size ,r, contains the configuration {n1,n2,….nk} is equal to

$$P\{r; k; n1,n2,,,,nk\} = r!/n1.n2\ldots..nk.\ l1!....lk!\ Qk/Q(Q+!)\ldots.Q+r-1) \quad \ldots(1)$$

where lj is the number of frequencies n1,…..nk that are equal to j, Q =4Nu and N is the population size. What is the proper effective population Nuber Ne to substitute for N in(1)

Following Karlin(1968) and chia and Pollak(1974), we shall assume that the possible population sizs are the finite numbers N(1) ….N(s) and that the sequence of sizes {Nt, t= 0,1,…} is finite irreducible Markov Chain

3. Analysis

The assicated transition Matrix is C =[civ], where civ = P[Nt+1 =N(v)/Nt =N(i)]

3.1 Estimating the transion probabilityes;

Let X= {x1,x2,……} be a random process in the discrete state space S. the conditional probability {xt+1=xt+1/, xt= xt, xt-1= xt-1, …x1=x1} = Px1x2, px2x3…px;-1xl the conditional probability of a path conditioned on the first value is the product of the transition probabilities between successive states of the path

3.2 Estimating the changes in gene frequency

In polymorphism, a gene with allels there n(n+1)/2 possible genotypes, the relationship between gene frequency and genotype frequency for a single gene at the population level can be used to infer the genetic states of the genes in population, if the genes and genotypic frequencies are constnt from generation to generation.

Population genetics, is the study of the distribution and change in Frequency of allels within population is knowns Evolutionery Biology. the four Process of Evolution

(i) Fishers' Natural Selection (2) Genetic Drift (3) Gene flow(Diffusion/ Weiener process) and (4) Mutation , (Continuous time ,continuous state Fokker Plank Diffusion Model) is widely used.

(i) Three State Random Walk; p{xi=1} = p and p{xi=-1} =1-p

Markov chain with transition probabilities

$$P_{ij} = \begin{cases} p & \text{if } j=i+1 \\ 1-p & \text{if } j=i-1 \\ 0 & \text{if o therwise} \end{cases}$$

Then xt+1=xt+ei+t , while ei+t and xt are independent

Consider the transition probabilities as follows

Xn \ xn+1	-1	0	1
-1	p-1-1	p-10	p-11
0	p01	p00	p01
1	p1-1	p10	p11

Assume all frequency is pn1= [1- mi]pn0+ mipi

=pno+ (mi(pi-pno)

Change in allele frequency is pn1-pn0 = (mi(pi-pn0)

(ii) Selection:

Consider the following Lif e cycle stage, their corresponding effect

Stage	Type of effect
Zygote forma tion to Adult	differential survival
Gametic	differential output, survival
Mating	Non random mating
Incompatability between mates And parent and progeny	

Selection includes differential Fertility, Fecundity, viability all ages and differential emigration.

$$Pn1 = \frac{f1pn02+f2pn0(1-pn0)}{F1p2n0+2f2pn0(1-pn0)+f3(1-pn0)2}$$

Where pn0 is the alleleic freguency for A in the population before selection and coefficient for selection is the measure of the disadvantage, those have least I.Q. has least chance of Survial.

Assuming random mating among AA and Aa and an appear in the total population be u,2v,w. The corresponding frequencies for parents are the

$$U^* = u/1-w, \quad 2v^* = 2v/1-w, \quad w^* = u \qquad ...(1)$$

$$P=u+v/1-w, \quad q=2v/1-w, \quad w^*=0 \qquad ...(2)$$

The probabilities of the three genotypes in first filal generation are p1=p2,2v1=2pq,w1=q2

In general pn= un+vn/1-wn , qn=1-w1/1-wn ...(3)

And un+1 =p2n ; 2vn+2 = 2pnqn; wn+1= qn2 ...(4)

From (3) and (4). Pn+1=un+1+un+1/1-wn+1 = pn/1-qn2 = 1/1+qn ...(5)

Qn+1 = vn+1/1-wn+1 =qn/!+qn ...(6)

From (6) we can calculate qn explicitly taking reciprocal we get

qn+1 -1 = 1+qn-1 ...(7)

hence substituting. Q1-1 = 1+q-1, q2-1 =2+q-1 ...(8)

qn = q/1+nq, wn+1=((q/1+qn)2 ...(9)

unproductive (or undesirable genes) genotype gradually drops out . the selection may be

(i) systematic effect, in which both the size and direction of the change are in principle

determinants

(ii) Dispensive effect, for which the size is determinant,in principle but the direction of the change is not

(iii) Non –recurrent Event, for which neither, size nor direction of change is determinate.

(iv) Genetic Drift

Genetic drift refers to the chance changes in frequency of allels from one generation to the next

3.3 Derivation of the two fundamental equations of Ewens Sampling

We shall first derive a recurrence equation that relates the probabilities of configurations in generations t and t+1 . We assume that the Wright Fisher Model holds . Thus if N(t) =N(i) and Nt+1 = N(v) the 2N(1) genes among the zygote of the offspring generation are obtained by repeated sampling with replacement of the 2N(i) genes of the parent generation. We also assume that as each gene is passed from the parent to the offspring it has a probability V of being is mutant to a type that didn't preiously exist in the population. Then if a sample of r+1 genes is taken from among the offspring and we denote by q(r+1,m/N(t) = N(i), N(t+1) = N(v)), the probability that this sample was transmitted from exactly m distinct parent genes,

q(r+1,r+1/Nt=N(i), Nt+1 =N(v)), …(2)

= 2N(i)(2N(i)-1)…….(2N(l)-r)/{2N(i)]r+1

= 1-r(r+1)/4N(i) +0(N(i))-2] …(3)

q(r+1, r/Nt c= N(i), Nt+1 = N(v)

= r+1C22N(I)(2n(I)-1)……(2n(I)……-R+1)/(2n(I))R+1

= R(R+1)/4n(I) +0[9n(I))-2) …(4)

as N(i) _ & , we obtain expression (4) because r+1C2 is the number of ways to whose 2 genes from among r+1 that were derived from the same parent and 2N(i) that were derived from the same parent and 2N(i)(2N(i)-1) ….(2N(i)-r+1) is the number of ways to select r distinct parental genes and a specified repeated parent from among 2N(i),

Now let P(t+1(r+1:k;n1,,,,,nk)/m,N(i),N(v)) and Pt+1(r+1; k; n1,….nk/N(i),N(v)) respectively denote conditional probabilities of the offspring configuration {n1,…nk} given m, N(i)… N(v) and N(i), N(v).

It follows from (3) and (4) that, P(t+1(r+1; k; n1,n2…nk/N(i), N(v))

= {1-r(r+1)/4N(i)}Pt+1(r+1;k; n1.n2…..nk/r+1), N(i),…N((v))+r(r+1)/4N(i) Pt+1(r+1; k; n1…..nk/r, N(i), N((v)) …(5)

if the possible population sizes are all large. we shall now calculate the probabilities on the right side of (5), if l=0. In this case the r+1 offspring genes must be unmated couples of the parental genes. Thus Pt+1(r+1; k; n1,….nk/r+1, N(i), N(v))

= (1-u)r+1pt(r+1; k; n1,…nk) …(6)

The second conditional configuration, probability of {A1, …Ak} is the probability of an event that, can occur in several mutually exclusive ways. Each of these is associated with r parental genes one of which is used as parent twice, on way is to have the parental configuration { n1,….nj-1,…nj+1….nk}. the probability of this is { l(nj))-1 times as large P(r;k; n1,,,,nj-1,nj-1,nj+1..nk} if there are l(nj) offspring genes in a sample of size of size r+1 that are represented nj times. Now, if the offspring genes represented nj-1 times are l(nj-1) in nuber, there are l(nj-1)+1 if such allels among the parents.

Hence, some parental allels represented nj-1 times is choosen to produce two copies with probability (nj-1)(l(nj-1)+1)/r, if parental genes are randomly choosen to be replicated twice. Hence

$$P\ t+1(r+1; k;n1,\ldots nk/r, N(i), N(v))$$

$$= (nj-1)(l(nj-1)+1/rl(nj)XPt(r;k;n1,\ldots.nj-1,nj+1\ldots nk/N(i) \quad \ldots(7)$$

Therefore, if we combine(2),(5),(6) and (7) we obtain

$$pt+1(r+1;k,n1\ldots\ldots nk/N(i),N(v))P(Nt=N(i))C1v$$

$$= Pt+1(r+1;k, n1\ldots\ldots nk/N(i), N(v)\ 0((Nt=N(i))Civ$$

$$P(Nt = N(i))\ Civ(1-(r+1)(r(r+1/4N(i) \quad nj-1\ Xl(nj-1)+1]/rl(nj)XPt(r;k; n1\ldots.nj-1, nj+1..nk)/N(i)$$

it applies if li =0

Now we shall now derive the econd if the fundamental equation , which only refer to configuration withtin one generation. To do this, a sample of r+1 genes in generation t will be looked upon as consisting of two parts. first 'v' genes are drawn and next the resulting subsample is supplement by the drawing of one more gene.with random sampling the configuration {n1, ..nk} among the first r genes is 1/(r+1) times as probable all sets of r+1 genes containg this as a subset of v. also one of the ways to have {n1,…nk} among the first r genes is to have it followed by a gene represented nj+1 times among r+1 genes and nj times any r. thus, if l(ni) is equal to the number of allels represented nj times among the r genes, each of these ways is 1/l(nj) times s probable as all configurations of the type {n1,n2,..nj-1,nj+1, ..nk}

finally, there are (nj+1)[l(nj+1)+1](nj+1)/l(nj) ways to pick a game times in the sample, therefore,

$$Pt(r;k,n1,\ldots nk) = \{l/nj+1)+1(nj+1)/l(nj)(r+1)Xpt(r+1;k, ,,,,,,nj-1,nj,nj+1,,,,nk)+l1+1/r+1pt(r+1,k+1,n..nk,l)$$

whee Ri is the number of allels represented once aomong the first 'r' genes.

Ewens sampling when the sample size is small in comparison with that of the population slection plays a negligible role, Mutation is nonrecurrent and the population is at equilibrium under mutuation random drift.

Evwens Sampling Formula: Q= 4Na , where N is the population size

To determine the overall frequency P for a given generation,

P= ¼(pmm+pmf+pfm+pff)

Where pym represent the allel frequencies transmitted from the y sex of parents to the z sex of offspring choosen to be used as parents with m=male and f=female when the chance occurrences along these pathways are independent, we can represent by

Vp =1/16(Vpmm+vpmf+vpfm+vpff)

Where vpyz is the variance of the frequency along each path. In an idealized reference situation of random sampling with replacement of allels the variances would be binomial with

Vpyz=pyz(1-pyz)/2Neyz

And Vp = p(1-p)/2Ne

Where Ne yz is the effective population number for specific path ,sex y tosex Z and we is the overall effective number. Using the combinatorial mathematics,what is the proper effective population number Ne=[1/d IIni/N(i)I-1).

If there is a nonoceious population of varying size that reproduce in each generation according to the Wright Fisher Model, Karlin(1968) and Coir and Pollak(1974), we shall assume . The population sizes are the fineite numbers N(1)…N(s) and that the queue of sizes{Nt, t=0,1,..} is a finite irreducible Markov chan it , the associative transiion matrix is C= {civ} where

Civ + P{N9+1 = N(v)/NN(t) =N(1)}

Which apply to theortical population as well as Wright –Fisher Reproduction population sturtcure.

Effective Breeding Population Size , For Structural Random Mating with Random or directional Selection pioneer work of Wright (1931) introduced the concept of effective population number to reflect the magnitude of expected random variation and fixation in gene frequency due to finite size of population known as effective popupulation number(Nw) for a population composed of Nm breeding males and Nf breeding female was 1/Ne = 1/4Nm+1/4/Nf

Four pathways allel pass between generation (i) male to male (ii) Male to female

(iii) Female to male (iv) female to female , which leads to four types of Stochastic process

3.4 Fisher's Natural Selection

The essence of the theory of evolution through the selection is that in any population there will exist genetic variation between individuals and that those genotypes which are better suited to the the environment than other will contribute rather more than their fair share of offspring to the following generation. Thus the geneetical make-up of the following generation will differ somewhat from that of the parent generation leading to substantial changes over large number of generation. such evolution depends as the genetical variation in the population, so that it might be expected that the greater variation, the greter will be the changes which occur. Further it appears that in some sense the process leads to the improvement in the population. Selection differences among genotypes generally leads to changes in gene frequencies.

IV Conclusion

Fokker Plank Diffusion Model

Continuous state continuous Parameter Diffusion Approximation of Fokker plank Diffusion Model. May be studied

Consider the random variable differing over (o,1), in such a way that if at any time, t, the random variable assumes the value x, then the value x+ x assumed at time t+dt is a random variable

E(dx) =m(x)dt+0(dt)2

V(dx) = v(x)dt+0(dt)2

E(dx) = 0(dt) i>3

Here m(x) and v(x) are function of x, but not of t, and are called respectively the drift and diffusion co-efficient of the process. If f(x;t) is the probability density of the random variable at time t, the theorem of total probability shows

f(x; t+ t) = F(x-dx;t)g(dx; x-dx).d(dx)

where g(dx; x-dx) is the probability density of change in the value of the random vriable from x-dx to x in the time interval (t, t+dt)

Acknoledgment:

Dr. S. Thobias, Ph.D., Head, Department of Statistics, Loyola College(Aut), Madras, Dr.(Mrs) Nancy, Ph.D. Asst. Prof in Statistics, Presidency College, Dr. N. Felix , Ph.D., Associate Professor in Bio Statistics and Community Medicine, Annamali University,

Dr.(Miss) S. Muthalagi, Asst. Professor in Zoology, Kolanjiapar Govt, Arts College, Virudhachalm, For their valuable suggestion for preparation of this paper

Reference

(1) N. John Britto, Probability Models for Estimation of Software Reliability and Review, Pages 33-34 ISSN- 12394-9333,www.ijtrd.com

(2)N. John Britto,Stochastic Process and its Application in HIV and A IDS(VOL 03, MAY 2014, PP ; 994-997, INTERNATIONAL JOURNAL COMPUTING ALGORITHM(ISSN-2278-2349)

(3) N. JOHN BRITTO, DR(MISS) S. MUTHALAGI, INTRODUCTION TO MATHEMATICAL MODELINIG IN BIOLOGY (AWITING FOR PUBLICATION)

(4) MRS. THILIAGM(2003), FRAILTY MODELS . (UNPUBLISHED) MADRAS UNIVERSITY ,UNPUBLISHED M.PHIL THEISIS

(5) N. JOHN BRITTO(2010), APPLICATION OF STOCHASTIC PROCESS IN EVOLUTIONERY GENETICS(UNPUBLISHED)

(6) HENRY L, SELVARAJ, C. SACHITHANTHAN S. " A MODEL FOR OPTIMAL BASE STOCK STYSEM FOR PATIENT CUSTOMERS WITH LEAD TIME SATISFYING SCBZ PROPERTRY' (2012) VOL.31E(MATHS&STAT) ISSUE NO.1, P,119-128

(7) N. JOHN BRITTO , MATHEMATICAL STRUCTURES IN POPULATION GENETICS(AWAITING FOR PUBLICATION)

(8) N. JOHN BRITTO, RACE, GENES AND INTELLEIGENCE: BAYESIAN METHODOLOGY ISSN 2347-1557, INTERNATIONAL JOURNAL OF MAHEMATICS AND ITS APPLICATIONS, ONLINE HTTP;/IJMA.IN VOL 04, ISSUE 4 (2016), 123-128(SPECIAL ISSUE)

(9)KEMTHRONE OSCAR(1957) AN INTRODUCTION TO GENETIC STATISTICS, JOHN WILEY & SONS, NEWYORK

10) AN INTRODUCTION TO PROBABILITY THEORY AND ITS APPLICATIONS WILLIAM FELLER, WILEY

(11) R.C. ELANDT JOHNSON(1975) PROBABILITY MODELS AND STATISTICAL METHODS IN GENETICS, WILEY

12) C.C.U(1976) FIRST COURS E IN POPULATION GENETICS, BOXWOOD PRESS

!3) D.J. FINNEY(1978) STATISTICAL METHODS IN BIOLOGICAL ASSAYS, CHARELES GRIFFICS CO

14) A.P. GORE AND S.A PARAJPE(2000) A COURSE IN MATHEMATICAL STATISTICAL METHODS IN GENETICS, WILEY.

15. INTRODUCTION TO QUANITATIVE GENETICS, D.S. FALCONER

16. POPULATION GENETICS, CHUN, CHING

17. POPULATION GENETICS, EWENS, W.J

18. EVOLUTION AND POPULATION GENETICS, SISODIA RASHMI

19. POPULATION GENETICS, M.S. BARTLETE

20. ELEMENTARY GENETICS, W. RALPH. SINGLOH