We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



16

Monte Carlo Implementations of Two Sex Density Dependent Branching Processes and their Applications in Evolutionary Genetics

Charles J. Mode[1], Towfique Raj, Ph.D.[2] and Candace K. Sleeman, Ph.D.[3]
[1]*Professor Emeritus of Mathematics, Drexel University Philadelphia, PA 19104*
[2]*Postdoctoral Research Fellow, & Harvard Medical School and Brigham Women's Hospital, & Boston, MA 02115*
[3]*Manager Of Engineering, NAVTEQ Corporation, Malvern, PA 19355*
USA

1. Introduction

Branching processes are a class of stochastic processes that deals with the dynamics of evolving populations and have an extensive literature dating back one hundred years or more. Some references to this early literature may be found in the book of Harris (1963) (7) as well as in the book Mode (1971) (9) on mulitype branching process and their applications. Other books on branching processes include those on Athreya and Ney (1972) (2), Jagers (1975) (8), Asmussen and Hering (1983) (1) and Haccou et al.(2007 (6), which contains discussions of classes of stochastic processes and their applications to biology and other fields. Fisher (1958), in Dover reprint of an edition of book first published in the late 1920's, was the first to apply what is now known as a Galton-Watson process to the study of the survival of mutations in biological evolution. Mode and Gallop (2008) (11) in a review paper on the applications of Monte Carlo simulation methods in the study and analysis of the widely used Wright-Fisher process of evolutionary genetics suggested that multitype branching processes could be used to eliminate the assumptions of constant population size which characterizes most applications of this process in the study of biological evolution.

Despite this extensive literature, little attention has been paid to two sex process as well as processes that are density dependent and evolve on a time scale of discrete generations. By definition, two sex branching processes entail the formation of couples consisting of a female and male, who produce offspring of both sexes. By a density dependent process, we mean that the expected number of offspring produced by a couple as well as the survival of offspring may depend on total population size in every generation. In this chapter such processes will be referred to as self regulating branching processes. In most classes of branching processes considered in the literature heretofore, total population size increases stochastically at a geometric rate so that there is no finite bound on total population size. This property is objectionable, because in real biological populations, the size of a population cannot increase indefinitely due limitations of resources and the environment. As will be shown by examples in subsequent sections, in the class of self regulating branching processes described in this

chapter and elsewhere this objectionable property of classical branching processes has been removed.

Two sex processes also allow for the inclusion of multitypes such as three or more genotypes at some autosomal locus that provide a framework for the study of several components of natural selection. Among these components are sexual selection, different expectations of the number of offspring among couple types and the differential survival by genotypes of the offspring. Sexual selection will be characterized in terms of probabilities of a female or male selecting a mate according to genotype or phenotype of its prospective partner. Differential expectations of the number of offspring among couple types as well survival of the offspring will be expressed as functions of total population size. A preliminary version of the two sex process described in this chapter may be found in Mode (1995) and more developed versions of this process may be found in the forthcoming book Mode and Sleeman (2010-2011) (13) chapter 11. The formulation developed in this book is at a general level in an attempt to provide a framework within which many versions of the process may be considered. However, the version of the two sex self regulating process described in this chapter is a simplified version of a more general process, which entailed an almost complete rewrite of the material in the forthcoming book. This rewrite is sufficient to make this chapter self contained without laborious references to yet unpublished literature.

The class of branching processes just described would be very difficult to analyze if an investigator had to resort to only the methods of classical mathematics. But thanks to the continuing development of powerful Monte Carlo simulation methods, such processes can be described algorithmically and analyzed by conducting Monte Carlo simulation experiments and summarizing the simulated data statistically. In this chapter, an outline of the algorithms involved in the computer implementation of such processes will be given and the results of a few illustrative Monte Carlo simulation experiments will be presented as exercises in the predictive implications of the model. Another property that is included in the formulation is that of embedding of non-linear deterministic models in the stochastic process, which make it possible to compare the predictive performance of the stochastic and deterministic models in any computer experiment. It should also be mentioned that all the computer experiments reported in this chapter were motivated by the book Wells (2008) (16), which presents evidence based on mutations in mitochondrial and Y chromosome DNA that all humans existing on planet earth today are descendants of small groups of hominids that migrated out of Africa 50,000 to 60,000 years ago.

2. Parameterization of couple formation processes

In the two sex branching process under consideration, offspring of the next generation are produced by couples consisting of a female and male. The first step in the formulation of this class of stochastic processes is to describe the couple formation process, which, among other things, may involve a type of sexual selection in which females or males have preferences as to the genotype or phenotype of their prospective sexual partners. When one includes genetics in a formulation, it will also be necessary to describe the set of genotypes under consideration in some suitable notation. For example, let $\mathfrak{G} = (i, j, k, \cdots)$ be the set of gametes under consideration with respect to some autosomal locus or set of autosomal loci. A diploid genotype will be denoted by the ordered pair (i, j), where i is the gamete contributed by the female parent and j is the gamete contributed by the male parent. Let $\mathfrak{T}_f = ((i, j) \mid i \in \mathfrak{G}, j \in \mathfrak{G})$ denote the set of female genotypes and define the set $\mathfrak{T}_m$ similarly for males. Elements of these sets will be denoted by τ_f and τ_m, respectively. A couple will be

of type $\kappa = \left(\tau_f, \tau_m\right)$ if the female is of genotype τ_f and the male is of genotype τ_m and let $\mathfrak{K}$ denote the set of all couple types. Either females or males who are not members of couples will be called singles. In practice, singles in some generation will consist of all offspring of the couple types present in that generation.

For the sake of simplicity, in this paper only one autosomal locus with two alleles will be under consideration. For this case the set of gametes will be denoted by $\mathfrak{G} = (A, a)$ so that if the gametes contributed by the maternal and paternal parents are distinguished, then the set of genotypes $\mathfrak{T}_f$ for females would contain 4 elements, and, similarly, the set $\mathfrak{T}_m$ of genotypes for males would also contain 4 elements. For this case, it follows that the set $\mathfrak{K}$ of couple types would contain 16 elements.

Let the random function $X\left(n; \tau_f\right)$ denote the number of single females present in the population in generation $n = 0, 1, 2, \cdots$ of genotype τ_f, and similarly let the random function $Y\left(n; \tau_m\right)$ denote the number of single males of genotype τ_m in generation n. Given these numbers of single females and males, let the random function $N_C\left(n; \kappa\right) \geq 0$ denote the potential number of couples formed in generation n from the single females and males. A helpful way of thinking about the random function $N_C\left(n; \kappa\right)$ is that it represents the maximum number of couples that could be formed, given the interactions of single females and males in their searches for mates in generation n.

Let the random function $Z\left(n; \kappa\right)$ denote that actual number of couples of type κ formed in generation n. Then, in the formulation under consideration, it will be assumed that $Z\left(n; \kappa\right)$ is a realization of a binomial random variable with index $N_C\left(n; \kappa\right)$ and probability $p\left(\kappa\right)$ for all $\kappa \in \mathfrak{K}$. From this assumption, it follows that $0 \leq Z\left(n; \kappa\right) \leq N_C\left(n; \kappa\right)$ with probability one for all $\kappa \in \mathfrak{K}$ and generations n. As will be demonstrated subsequently, this condition plays an essential role in showing that the random functions $Z\left(n; \kappa\right)$ satisfy a set of necessary constraints imposed by the number of single females and males eligible to form couples in any generation n with probability one.

A basic component of the formulation under consideration is that of social contact probabilities among single females and males in their searches for mates. Given a single female of genotype $\tau_f \in \mathfrak{T}_f$ in generation n, let $\gamma_f\left(n; \tau_f, \tau_m\right)$ denote the conditional probability that she has contact with a single male of genotype $\tau_m \in \mathfrak{T}_m$. The conditional contact probability $\gamma_m\left(n; \tau_m, \tau_f\right)$ for single males of genotype $\tau_m \in \mathfrak{T}_m$ in generation n is defined similarly. Subsequently, it will be shown that these contact probabilities for single females and males will be constructed in every generation from, among other things, the frequencies of the genotypes within the female and male subpopulations of singles. In this construction, all probabilities will belong to the closed interval $[0, 1]$, and for females, for example, will satisfy the condition

$$\sum_{\tau_m \in \mathfrak{T}_m} \gamma_f\left(n; \tau_f, \tau_m\right) = 1 \tag{2.1}$$

for all $\tau_f \in \mathfrak{T}_f$ and generations n. An analogous condition holds for the contact probabilities of single males.

Let

$$\boldsymbol{\gamma}_f\left(n; \tau_f\right) = \left(\gamma_f\left(n; \tau_f, \tau_m\right) \mid \tau_m \in \mathfrak{T}_m\right) \tag{2.2}$$

denote a vector of contact probabilities for single females of genotype τ_f in generation n and let $\boldsymbol{\gamma}_m\left(n; \tau_m\right)$ denote a similar vector for single males in generation n. For single females of

genotype τ_f in generation n, let the random function $Z_f\left(n;\tau_f,\tau_m\right)$ denote the number of single males of genotype τ_m selected as potential sexual partners, and let

$$\boldsymbol{Z}_f\left(n;\tau_f\right)=\left(Z_f\left(n;\tau_f,\tau_m\right)\mid\tau_m\in\mathfrak{T}_m\right) \tag{2.3}$$

denote a vector of these random functions. Given the number $X\left(n;\tau_f\right)$ of single females of genotype τ_f in generation n, it will be assumed that the vector $\boldsymbol{Z}_f\left(n;\tau_f\right)$ has a conditional multinomial distribution with index $X\left(n;\tau_f\right)$ and probability vector $\boldsymbol{\gamma}_f\left(n;\tau_f\right)$. In symbols,

$$\boldsymbol{Z}_f\left(n;\tau_f\right)\sim CMultinom\left(X\left(n;\tau_f\right),\boldsymbol{\gamma}_f\left(n;\tau_f\right)\right)\ . \tag{2.4}$$

Similarly, let $\boldsymbol{Z}_m\left(n;\tau_m\right)$ denote the corresponding vector of random functions for single males of genotype τ_m in generation n. Then, it will also be assumed that

$$\boldsymbol{Z}_m\left(n;\tau_m\right)\sim CMultinom\left(Y\left(n;\tau_m\right),\boldsymbol{\gamma}_m\left(n;\tau_m\right)\right)\ . \tag{2.5}$$

Having defined the distributions of the vectors $\boldsymbol{Z}_f\left(n;\tau_f\right)$ and $\boldsymbol{Z}_m\left(n;\tau_m\right)$ in (2.4) and (2.5), it is now possible to state how realizations of the random function $N_C\left(n;\kappa\right)$ in any generation n and couple type $\kappa=\left(\tau_f,\tau_m\right)$ will be computed. Because the potential number of pair-wise contacts of type $\left(\tau_f,\tau_m\right)$ cannot exceed the number of single females of genotypes τ_f seeking single males of genotype τ_m and similarly the number of single males of genotype τ_m seeking singly females of genotype τ_f, it follows that a plausible choice for the random function $N_C\left(n;\kappa\right)$ is

$$N_C\left(n;\kappa\right)=\min\left(Z_f\left(n;\tau_f,\tau_m\right),Z_m\left(n;\tau_m,\tau_f\right)\right)\ , \tag{2.6}$$

for all $\kappa\in\mathfrak{K}$ and generations n, where $\kappa=\left(\tau_f,\tau_m\right)$. As shown above, if the random function $Z\left(n;\kappa\right)$ is the actual number of couple of type κ realized in generation n, then $Z\left(n;\kappa\right)\leq N_C\left(n;\kappa\right)$ for all $\kappa\in\mathfrak{K}$ and generations n. Also note that from (2.4) it follows that

$$\sum_{\tau_m\in\mathfrak{T}_m}Z_f\left(n;\tau_f,\tau_m\right)=X\left(n;\tau_f\right) \tag{2.7}$$

with probability one for all genotypes $\tau_f\in\mathfrak{T}_f$ and generations n. Therefore, from (2.6) is can be seen that

$$\sum_{\tau_m\in\mathfrak{T}_m}N_C\left(n;\tau_f,\tau_m\right)\leq\sum_{\tau_m\in\mathfrak{T}_m}Z_f\left(n;\tau_f,\tau_m\right)=X\left(n;\tau_f\right) \tag{2.8}$$

for all generations n and genotypes $\tau_f\in\mathfrak{T}_f$. Therefore, the total number of females of genotype τ_f in couples of the types $\left(\tau_f,\tau_m\right)$ such the $\tau_m\in\mathfrak{T}_m$ will not exceed the number of single females of this genotype $X\left(n;\tau_f\right)$ present in the population prior to the time the couple formation process occurred. By a similar argument, it can be shown that

$$\sum_{\tau_f\in\mathfrak{T}_f}N_C\left(n;\tau_f,\tau_m\right)\leq\sum_{\tau_f\in\mathfrak{T}_f}Z_m\left(n;\tau_m,\tau_f\right)=Y\left(n;\tau_m\right) \tag{2.9}$$

with probability one for all $\tau_m \in \mathfrak{T}_m$ and generations n. Thus, the number of couples with males of genotype $\tau_m \in \mathfrak{T}_m$ will never exceed the number of single males $Y(n;\tau_m)$ of this genotype present in the population before the couple formation processes occurred in every generation n with probability one.

The last step in the formulation of the couple formation process is that of defining a procedure for calculating the contact probabilities for single females and males, which will depend on acceptance probabilities for both females and males. For example, given a single female of genotype $\tau_f \in \mathfrak{T}_f$, let $\alpha_f\left(\tau_f,\tau_m\right)$ denote the conditional probability that she finds a single male of genotype $\tau_m \in \mathfrak{T}_m$ acceptable as a sexual partner. Similarly, for a single male of genotype τ_m, let $\alpha_m\left(\tau_m,\tau_f\right)$ denote the conditional probability that he finds a single female of genotype τ_f acceptable as a sexual partner. These acceptance probabilities will be discussed in more detail subsequently.

All contact probabilities will depend on the frequencies of the genotypes in the single populations of females and males prior to the beginning of the couple formation process. By definition, the frequency of genotype τ_f in the single female population in generation n is

$$U_f\left(n;\tau_f\right) = \frac{X\left(n;\tau_f\right)}{X(n;\circ)}\,, \tag{2.10}$$

where

$$X(n;\circ) = \sum_{\tau_f \in \mathfrak{T}_f} X\left(n;\tau_f\right) \tag{2.11}$$

and $X(n;\circ) > 0$. If $X(n;\circ) = 0$, then $U_f\left(n;\tau_f\right) = 0$. Similarly, let $U_m(n;\tau_m)$ denote the frequency of genotype $\tau_m \in \mathfrak{T}_m$ in the male population of singles in generation n.

By the law of total probability, the probability that a single female of genotype $\tau_f \in \mathfrak{T}_f$ has contact with some single male is

$$\sum_{\tau_m \in \mathfrak{T}_m} U_m(n;\tau_m)\,\alpha_f\left(\tau_f,\tau_m\right)\,. \tag{2.12}$$

By an application of Bayes' formula, it follows that conditional probability that a single female of genotype τ_f in generation n has contact with a single male of genotype τ_m is

$$\gamma_f\left(n;\tau_f,\tau_m\right) = \frac{U_m(n;\tau_m)\,\alpha_f\left(\tau_f,\tau_m\right)}{\sum_{\tau_m \in \mathfrak{T}_m} U_m(n;\tau_m)\,\alpha_f\left(\tau_f,\tau_m\right)}\,. \tag{2.13}$$

Similarly, the conditional probability that a single male of genotype τ_m has contact with a single female of genotype τ_f in generation n is

$$\gamma_m\left(n;\tau_m,\tau_f\right) = \frac{U_f\left(n;\tau_f\right)\alpha_m\left(\tau_m,\tau_f\right)}{\sum_{\tau_f \in \mathfrak{T}_f} U_f\left(n;\tau_f\right)\alpha_m\left(\tau_m,\tau_f\right)}\,. \tag{2.14}$$

From these equations, it can be seen that if there are positive constants such that $\alpha_f\left(\tau_f,\tau_m\right) = a$ and $\alpha_m\left(\tau_m,\tau_f\right) = b$ for all types of contacts, $\left(\tau_f,\tau_m\right)$ and $\left(\tau_m,\tau_f\right)$, then

$$\gamma_f\left(n;\tau_f,\tau_m\right) = U_m(n;\tau_m) \tag{2.15}$$

and

$$\gamma_m\left(n;\tau_m,\tau_f\right)=U_f\left(n;\tau_f\right)\ . \tag{2.16}$$

When these conditions hold, the mating system is, by definition, random.
For a case of one autosomal locus with two alleles under consideration, the defining and numerical assignments of the acceptance probabilities for both single females and males reduce to considering 9 acceptance probabilities for each sex. Let

$$\mathfrak{T}=(AA,Aa,aa)=(1,2,3) \tag{2.17}$$

denote the set of genotypes for each sex. Then, for the case of single females, the matrix of acceptance probabilities takes the form

$$\boldsymbol{A}_f=\begin{pmatrix}\alpha_f(1,1)&\alpha_f(1,2)&\alpha_f(1,3)\\ \alpha_f(2,1)&\alpha_f(2,2)&\alpha_f(2,3)\\ \alpha_f(3,1)&\alpha_f(3,2)&\alpha_f(3,3)\end{pmatrix}\ . \tag{2.18}$$

There is also a similar matrix of acceptance probabilities for single males. At first sight, it may strike a reader that assigning numerical values of 18 probabilities may be an insurmountable task. But, in subsequent sections, it will be shown that meaningful genetic examples may be considered in computer simulation experiments by assigning numerical values to as few as two parameters for each sex.
Another version of a couple formation process described in this section for the case of sexually transmitted diseases may be found in the book Mode and Sleeman (2000) (12).

3. Gamete and offspring distribution for couples

In the class of two sex processes under consideration, describing the evolution of a diploid population on a discrete time scale of generations, genes are passed on in the offspring of those couples who reproduce. For, example consider a couple of type $\kappa=\left(\tau_f,\tau_m\right)$, where the genotype of the female is τ_f and that of the male is τ_m. For the autosomal locus with two alleles under consideration each of these genotypes will belong the to set (AA,Aa,aa). From now on for the sake of simplicity, the parental origin of the gametes making up a genotype will not be distinguished as to whether it is maternal or paternal. Given a couple of type $\kappa=\left(\tau_f,\tau_m\right)$ in some generation, let $p(\kappa;\tau)$ denote the conditional probability that the couple produces an offspring of genotype $\tau\in(AA,Aa,aa)$. Initially, the sex of the offspring will be ignored, but, subsequently, it will be taken into account in the formulation.
One of the significant processes underlying evolution is that of mutation. For the sake of simplicity, it will be assumed that mutations occur only among the two alleles in the set $(A,a)=(1,2)$. Let μ_{12} denote the conditional probability per meiosis that allele A mutates to allele a, and, similarly, let μ_{21} be the conditional probability that allele a mutates to allele A per meiosis. In what follows, it will be helpful to represent mutation probabilities in the form of the 2×2 matrix

$$\mathfrak{M}=\begin{pmatrix}\mu_{11}&\mu_{12}\\ \mu_{21}&\mu_{22}\end{pmatrix} \tag{3.1}$$

where the diagonal elements are chosen such that each row of this matrix sums to one. Observe that μ_{11} is the conditional probability that allele A does not mutate per meiosis, and, similarly, μ_{22} is the conditional probability that allele a does not mutate per meiosis.

The next step in the formulation is to derive the gametic distribution for each genotype under the assumption that mutation occurs. Let $p_g(\tau;\nu)$ denote the conditional probability that a genotype $\tau \in (AA, Aa, aa) = (1,2,3)$ produces a gamete of type $\nu \in (A, a) = (1,2)$ during meiosis. To illustrate the procedure for deriving these probabilities, suppose $\tau = AA = 1$ and $\nu = A = 1$. Then the probability that the left allele in the genotype AA is contributed to the gene pool of the population without mutation is $\mu_{11}/2$. Similarly the right allele in the genotype AA is contributed to the gene pool of the population without mutation is $\mu_{11}/2$. Thus, $p_g(1;1) = \mu_{11}$ and by a similar argument it may be shown that $p_g(1;2) = \mu_{12}$. By using this argument repeatedly, it can be shown that $p_g(2;1) = (\mu_{11} + \mu_{21})/2$, $p_g(2;2) = (\mu_{12} + \mu_{22})/2$, $p_g(3;1) = \mu_{21}$ and $p_g(3;2) = \mu_{22}$.

Next consider a couple of type $\kappa = \left(\tau_f, \tau_m\right)$ and let $p_c(\kappa;\tau)$ denote the conditional probability that a couple of type κ produces an offspring of genotype $\tau = (\nu, \nu')$. Then, under the assumption that female and male gametes unite independently to form a zygote, it follows that

$$p_c(\kappa;\tau) = p_g\left(\tau_f;\nu\right) p_g\left(\tau_m;\nu'\right) \tag{3.2}$$

for all couple types $\kappa \in \Re$ and genotypes $\tau = (\nu, \nu') \in (AA, Aa, aa)$. The collection of probabilities in (3.2) as τ varies over the set (AA, Aa, aa) will be referred to as the offspring distribution for a couple type $\kappa \in \Re$. Note that for each couple type $\kappa \in \Re$ there corresponds an offspring distribution. For the case under consideration in which the parental origin of the gametes making up an individual is not accounted for, it is easy to see that there would be 9 types of couples. In what follow, it will be helpful to let the vector

$$\boldsymbol{p}_c(\kappa) = (p_c(\kappa;\tau) \mid \tau \in (AA, Aa, aa)) \tag{3.3}$$

denote the offspring distribution for a couple of type $\kappa \in \Re$.

Up to now in our formulation, the only component of natural selection that has been accommodated in the formulation is that of sexual selection. Another component of natural selection that will be accommodated in the model is that of reproductive success of each couple type $\kappa \in \Re$. Let $N(\kappa)$ denote a random variable, taking values in the set $(m \mid m = 0,1,2,\cdots)$ of nonnegative integers, be defined for each couple type $\kappa \in \Re$, and let $W(\kappa;\tau)$ denote a random variable representing the number of offspring of genotype τ produced by a couple of type $\kappa \in \Re$ in any generation. To simplify the notation, it will also be helpful to let

$$\boldsymbol{W}(\kappa) = (W(\kappa;\tau) \mid \tau \in (AA, Aa, aa)) \tag{3.4}$$

denote a vector representing the number of offspring of each of the three genotypes produced by a couple of type κ under consideration in some generation. It will be assumed that, given a realization of the random variable $N(\kappa)$, the random vector $\boldsymbol{W}(\kappa)$ has a multinomial distribution with an index $N(\kappa)$ and probability vector $\boldsymbol{p}_c(\kappa)$. In symbols,

$$\boldsymbol{W}(\kappa) \sim CMultinom\left(N(\kappa), \boldsymbol{p}_c(\kappa)\right) . \tag{3.5}$$

To complete the formulation of this module of the model, it will be assumed that each random variable $N(\kappa)$ has a Poisson distribution with parameter $\lambda(\kappa)$, where $\kappa \in \Re$. For the case under consideration, there are three genotypes $(AA, Aa, aa) = (1,2,3)$ and 9 couple types. Let $\lambda(i,j)$ denote the Poisson parameter for a couple of type $\kappa = (i,j)$. Then, the nine Poisson

parameters under consideration may be represented in the form of a 3×3 matrix

$$\mathbf{\Lambda} = \begin{pmatrix} \lambda(1,1) & \lambda(1,2) & \lambda(1,3) \\ \lambda(2,1) & \lambda(2,2) & \lambda(2,3) \\ \lambda(3,1) & \lambda(3,2) & \lambda(3,3) \end{pmatrix} . \tag{3.6}$$

When faced with models with many parameters, it is always of interest to find ways in which the number of parameters to be considered may be reduced. For the case of the matrix in (3.6) it seems reasonably to assume that the matrix $\mathbf{\Lambda}$ is symmetric. That is $\lambda(i,j) = \lambda(j,i)$ for all pairs. In terms of the genotypes of the female and male making up a couple, it seems reasonable to assume that couples of type (AA, aa) or (aa, AA) would on average produce the same number of offspring per generation. Of course, it may also be plausible to assume that mothers of genotype AA are more successful in producing offspring than those of genotype aa, but the exploration of such questions in computer experiments will be left to readers.

4. A Self regulating stochastic population process

In this section some algorithms will be described for computing realizations of the two sex stochastic process under consideration. The aim of these algorithms is set up a procedure such that, given some number of generations G such as four to six thousand, a number $M \geq 1$ of replications of the stochastic evolution of the process for G generations may be computed. This sample of M realizations of the process will then be summarized statistically and will be used as a graphic description of the variability among the realizations of the process for G generations of evolution. In any generation n, for $n = 0, 1, 2, 3, \cdots$, let $X\left(n; \tau_f\right)$ and $Y(n; \tau_m)$ denote, respectively, the number of single females and males in the population with genotypes τ_f and τ_m. In any simulation experiment, the initial numbers $X\left(0; \tau_f\right)$ and $Y(0; \tau_m)$ will be assinged by an experimenter. Given these numbers of single females and males in some generation n, let the random function $Z(n; \kappa)$ denote the number of couples of type $\kappa = \left(\tau_f, \tau_m\right)$ formed in generation n from the single females and males.

In generation $n \geq 1$, let the random function $T(n; \kappa; \tau)$ denote the total number of individuals of genotype $\tau \in \mathfrak{T} = (AA, Aa, aa)$ produced by couples of type κ in generation. For, given $Z(n; \kappa) \geq 0$, let $(W_\nu(\kappa; \tau) \mid \nu = 1, 2, \cdots, Z(n; \kappa))$ be a collection of conditionally independent random variables whose common distribution is that of the random variable $W(\kappa; \tau)$ in (3.5). Then, $T(n; \kappa; \tau)$ is the random sum

$$T(n; \kappa; \tau) = \sum_{\nu=1}^{Z(n;\kappa)} W_\nu(\kappa; \tau) , \tag{4.1}$$

where $T(n; \kappa; \tau) = 0$ if $Z(n; \kappa) = 0$. Let the random function $V(n; \tau)$ denote the total number of offspring of genotype τ produced in generation n. Then,

$$V(n; \tau) = \sum_{\kappa \in \mathfrak{K}} T(n; \kappa; \tau) . \tag{4.2}$$

The sex of each genotype will be considered subsequently. The total number of offspring produced by couples in generation n is given by the random function

$$T(n) = \sum_{\tau \in \mathfrak{T}} V(n; \tau) , \tag{4.3}$$

which will play a role in the survivability of each offspring in generation n.
At this point in the discussion, the sex of each offspring will be taken into account. Let p_f denote the probability that an offspring is female and let p_m denote the probability that an offspring is male. Then, given $V(n;\tau)$, the total number of females offspring of genotype $\tau_f = \tau$, say $X_T(n;\tau)$, is given by a realization of a conditional binomial random variable with index $V(n;\tau)$ and probability p_f. In symbols,

$$X_T(n;\tau) \sim CBinom\left(V(n;\tau), p_f\right) \tag{4.4}$$

and the number of male offspring of genotype $\tau_m = \tau$ is given by $Y_T(n;\tau) = V(n;\tau) - X_T(n;\tau)$.
If at this point in the description of the algorithms, a reader may concerned as to the validity of these computations, because it would make more sense to simulate whether an offspring is female or male for each couple type. In this connection, it is suggested that the following property of a sum of independent binomial random variables X_k with respective indices N_k for $k = 1, 2, \cdots, M$ with a common probability p_f be recalled. For in this case, the random variable $Y = X_1 + X_2 + \cdots X_M$ has a binomial distribution with index $N = N_1 + N_2 + \cdots + N_M$ and probability p_f. Hence, the computational procedure indicated in (4.4) is justified, and, moreover, it is more efficient from the computational point of view than doing the computation for each couple type separately.
At this point in the formulation of the model, another component of natural selection will be taken into account; namely, the ability of the female and male offspring of generation n to survive, form couples and contribute offspring to generation $n+1$. For a female of genotype $\tau_f \in \mathfrak{T}$ in generation n, let $s\left(n;T(n),\tau_f\right)$ denote the conditional probability that she survives to reproduce in generation $n+1$, given $T(n)$. Then, it will be assumed that $s\left(n;T(n),\tau_f\right)$ has a parametric form of a Weibull survival function

$$s_f\left(n;T(n),\tau_f\right) = \exp\left(-\left(\beta_f\left(\tau_f\right)T(n)\right)^{\alpha_f\left(\tau_f\right)}\right) \tag{4.5}$$

for every $\tau_f \in \mathfrak{T}$, where $\beta_f\left(\tau_f\right)$ and $\alpha_f\left(\tau_f\right)$ are positive parameters. The corresponding survival function $s_m(n;T(n),\tau_m)$ for males of genotype τ_m is defined similarly with parameters $\beta_m(\tau_m)$ and $\alpha_m(\tau_m)$.
Let the random functions $X\left(n+1;\tau_f\right)$ and $Y(n+1;\tau_m)$, respectively, denote the number of single females and males in generation n of genotypes τ_f and τ_m who survive to produce the offspring of generation $n+1$. Then, it will be assumed that

$$X\left(n+1;\tau_f\right) \sim CBinom\left(X_T\left(n;\tau_f\right), s_f\left(n;T(n),\tau_f\right)\right) \tag{4.6}$$

for every $\tau_f \in \mathfrak{T}$. Similarly, it will be assumed that

$$Y(n+1;\tau_m) \sim CBinom\left(Y_T(n;\tau_m), s_m(n;T(n),\tau_m)\right) \tag{4.7}$$

for every $\tau_m \in \mathfrak{T}$.
Given realizations of the random functions in (4.6) and (4.7), the algorithms developed in section 2 may be applied to compute realization of the collection of random functions $(Z(n+1;\kappa) \mid \kappa \in \mathfrak{K})$ for generation $n+1$. It should also be observed that, because the

parameters in the Weibull survival function depend on sex and genotype, values of these parameters may be assigned by sex, genotype or a combination of sex and genotype. Thus, in implementation of the formulation under consideration, the α and the β parameters may differ among genotypes in each sex. In subsequent computer experiments, examples of these differences will be given.

5. Non-linear deterministic equations embedded in the stochastic population process

Some stochastic formulations start with a deterministic set of equations, such as difference or differential equations in continuous or discrete time, and then proceed to extend them to stochastic model by adding a stochastic "error" term or some other random components. In this section, however, a different approach will be followed in that, given a stochastic process, a set of recursive non-linear difference equations will be embedded in the stochastic process discussed in the previous section by a procedure that entails the idea of estimating the sample functions of the process. To illustrate the idea of estimating the sample function of a stochastic process, let the sequence of random variables $X_1, X_2, X_3, \cdots$ denote some Markov process in discrete time and suppose each random variable in this sequence has a finite expectation.
It is well known that if one is considering a Markovian sequence of random variables, then the best estimate, in the sense of mean square error, of the sample function X_{n+1} based on some function of the random variable X_n is to find some function of X_n, say $f(X_n)$, such that the conditional expectation

$$E\left[(X_{n+1} - f(X_n))^2 \mid X_n\right] \tag{5.1}$$

is a minimum. By using well known properties of conditional expectations, it can be shown that

$$f(X_n) = E[X_{n+1} \mid X_n] \tag{5.2}$$

Thus, if we let $\widehat{X}_{n+1}$ denote the best estimate of X_{n+1}, then

$$\widehat{X}_{n+1} = E[X_{n+1} \mid X_n] \ . \tag{5.3}$$

Observe that the conditional expectation on the right is a random variable, since it is a function of the random variable X_n. Quite often, the conditional expectation on the right is a non-linear function of the random variable X_n so that finding a useful expectation of this random variable if often difficult or nearly impossible. But, if one knows that the value of an initial random variable X_0, then the estimate

$$\widehat{X}_1 = E[X_1 \mid X_0] \tag{5.4}$$

is known exactly. Next consider the estimate

$$E[X_2 \mid X_1] \ , \tag{5.3}$$

which is also a non-linear function of the X_1. But, an estimate of X_1 is given in (5.4). Thus, it seems reasonable, to choose

$$\widehat{X}_2 = E\left[X_2 \mid \widehat{X}_1\right] \tag{5.4}$$

as an estimate of X_2. By continuing this procedure, one arrives at the recursive system

$$\widehat{X}_{n+1} = E\left[X_{n+1} \mid \widehat{X}_n\right] \tag{5.5}$$

for $n = 0, 1, 2, \cdots$. In subsequent sections of this chapter, trajectories that have been computed using a non-linear difference equation or equations of form (5.5) will be compared graphically with a statistical summarization of a sample of realizations of the stochastic population process discussed in the preceding section.

As a first step in applying the ideas just discussed, let the symbol $\mathfrak{B}(n)$ denote a realization of the sample functions of the population process in generation n. Then, from (4.6) and (4.7) it follows that for any arbitrary genotype $\tau \in \mathfrak{T}$

$$E\left[X(n+1;\tau) \mid \mathfrak{B}(n)\right] = s_f\left(n; T(n), \tau\right) X_T(n;\tau) \ . \tag{5.6}$$

Similarly,

$$E\left[Y(n+1;\tau) \mid \mathfrak{B}(n)\right] = s_f\left(n; T(n), \tau\right) Y_T(n;\tau) \ . \tag{5.7}$$

From these equations, it can be seen that the random variables $T(n)$, $X_T(n;\tau)$ and $Y_T(n;\tau)$ will have to be estimated. From (4.4) it follows that

$$E\left[X_T(n;\tau) \mid \mathfrak{B}(n)\right] = p_f V(n;\tau) \ , \tag{5.8}$$

and

$$E\left[Y_T(n;\tau) \mid \mathfrak{B}(n)\right] = p_m V(n;\tau) \ . \tag{5.9}$$

Thus, the random function $V(n;\tau)$, which denotes the total number of offspring of genotype τ produced by couples in generation n, will also have to be estimated.

From (4.1) and (4.2) it can be seen that the first step in estimating the random function $V(n;\tau)$ is that of estimating the random function

$$T(n;\kappa;\tau) = \sum_{\nu=1}^{Z(n;\kappa)} W_\nu(\kappa;\tau) \ . \tag{5.10}$$

Observe that for every $\nu = 1, 2, \cdots, Z(n;\kappa)$, it follows from (3.2) and (3.5) that

$$E\left[W_\nu(\kappa;\tau)\right] = \lambda(\kappa)\, p_c(\kappa;\tau) \ . \tag{5.11}$$

Therefore, from (5.10) it can be seen that

$$E\left[T(n;\kappa;\tau) \mid \mathfrak{B}(n)\right] = Z(n;\kappa)\, \lambda(\kappa)\, p_c(\kappa;\tau) \ . \tag{5.12}$$

From this equation, it follows that the random function $Z(n;\kappa)$ needs to be estimated.

To estimate this random function, section 2 should be consulted. From this section, it can be seen that

$$E\left[Z(n;\kappa) \mid \mathfrak{B}(n)\right] = N_c(n;\kappa)\, p(\kappa) \ , \tag{5.13}$$

where $p(\kappa)$ is the probability that a couple of type $\kappa \in \mathfrak{K}$ is formed during any generation n. From (2.6) it can be seen that the random function $N_c(n;\kappa)$ is computed from the random functions $Z_f\left(n;\tau_f,\tau_m\right)$ and $Z_m\left(n;\tau_m,\tau_f\right)$ for couples of type $\kappa = \left(\tau_f,\tau_m\right)$. But, these random functions are, respectively, elements of random vectors of multinomial distributions with probability vectors $\gamma_f\left(n;\tau_f\right)$ and $\gamma_m(n;\tau_m)$ and indices $X\left(n;\tau_f\right)$ and $Y(n;\tau_m)$, see $(2.1) - (2.5)$ for details. From the properties of multinomial distribution, it follows that

$$E\left[Z_f\left(n;\tau_f,\tau_m\right) \mid \mathfrak{B}(n)\right] = X\left(n;\tau_f\right) \gamma_f\left(n;\tau_f,\tau_m\right) \tag{5.14}$$

and

$$E\left[Z_m\left(n;\tau_m,\tau_f\right)\mid\mathfrak{B}\left(n\right)\right]=Y\left(n;\tau_m\right)\gamma_m\left(n;\tau_m,\tau_f\right)\ . \tag{5.15}$$

Observe that the probabilities $\gamma_f\left(n;\tau_f,\tau_m\right)$ and $\gamma_m\left(n;\tau_m,\tau_f\right)$ are computed from the realized random functions in every generation n, see $(2.10)-(2.15)$ for details. In particular, in the initial generation 0 integers are assigned to the initial values $X\left(0;\tau\right)$ and $X\left(0;\tau\right)$ for all $\tau\in\mathfrak{T}$. Given these initial values, $\gamma_f\left(0;\tau_f,\tau_m\right)$ and $\gamma_m\left(0;\tau_m,\tau_f\right)$ could be calculated. Thus, the estimate of the random function $Z_f\left(n;\tau_f,\tau_m\right)$ in generation 0 would be computed according to the formula

$$\widehat{Z}_f\left(0;\tau_f,\tau_m\right)=X\left(0;\tau_f\right)\gamma_f\left(0;\tau_f,\tau_m\right)\ , \tag{5.16}$$

and similarly

$$\widehat{Z}_m\left(0;\tau_m,\tau_f\right)=Y\left(0;\tau_m\right)\gamma_m\left(0;\tau_n,\tau_f\right)\ . \tag{5.17}$$

Finally, the random function $N_c\left(n;\kappa\right)$ would be estimated in generation $n=0$ by using the formula

$$\widehat{N}_C\left(0,\kappa\right)=\min\left(\widehat{Z}_f\left(0;\tau_f,\tau_m\right),\widehat{Z}_m\left(0;\tau_m,\tau_f\right)\right) \tag{5.18}$$

for every couple of type $\kappa=\left(\tau_f,\tau_m\right)\in\mathfrak{K}$.

Given these estimates, the random function $V\left(n;\tau\right)$ denoting the total number of offspring of genotype τ produced in generation $n=0$ would be estimated using the formula

$$\widehat{V}\left(0;\tau\right)=\sum_{\kappa\in\widehat{\mathfrak{K}}}\widehat{N}_C\left(0,\kappa\right)p\left(\kappa\right)\lambda\left(\kappa\right)p_c\left(\kappa;\tau\right)\ . \tag{5.19}$$

Therefore, from $(5.6)-(5.9)$ it follows that

$$\widehat{X}\left(1;\tau\right)=s_f\left(0;\widehat{T}\left(0\right),\tau\right)p_f\widehat{V}\left(0;\tau\right)\ , \tag{5.20}$$

and

$$\widehat{Y}\left(1;\tau\right)=s_m\left(0;\widehat{T}\left(0\right),\tau\right)p_m\widehat{V}\left(0;\tau\right) \tag{5.21}$$

for all $\tau\in\mathfrak{T}$.

Now suppose the procedure just described in continued recursively so that in generation $n\geq 2$ we arrive at the collection of estimates

$$\widehat{\mathfrak{B}}\left(n\right)=\left(\widehat{X}\left(n;\tau\right),\widehat{Y}\left(n;\tau\right)\mid\tau\in\mathfrak{T}\right)\ . \tag{5.22}$$

Then the estimates of these random functions in generation $n+1$ would be computed using the equations

$$\widehat{X}\left(n+1;\tau\right)=s_f\left(n;\widehat{T}\left(n\right),\tau\right)p_f\widehat{V}\left(n;\tau\right)\ , \tag{5.23}$$

and

$$\widehat{Y}\left(n+1;\tau\right)=s_m\left(n;\widehat{T}\left(n\right),\tau\right)p_m\widehat{V}\left(n;\tau\right)\ . \tag{5.24}$$

As can be seen form the derivation just completed, the pair of equations (5.23) and (5.24) are purely deterministic and depend on the numerical values that are assigned to the parameters of the model in any computer experiment as well as the initial conditions. Given these

parameter assignments in generation $n = 0$ and the initial conditions, trajectories based on the embedded deterministic model could be computed for generations $n = 1, 2, \cdots, G$, where G is some preassigned integer. It would also be possible to compute a sample of $M \geq 2$ of Monte Carlo replications of realizations of the process for G generations, which could then be statistically summarized in the form of quantile and mean trajectories over G generations. In such experiments, an interesting question to ask is: do the predictions for the evolution of a population for G generations computed by using the embedded deterministic model compare in some favorable or unfavorable sense with the predictions of the stochastic process? In subsequent sections of this chapter, the results of a number of computer experiments will be reported in an attempt to provide some answers to this question.

6. Failure of deterministic model to predict the evolution of the stochastic process

Among the conceptual schemes of evolution is the idea that many existing populations evolved from a small founder population and beneficial mutations among them and their descendants led to significant genetic transformations as the population evolved. For the case of the present of the world wide human population, a reader may wish the consult the interesting book by Wells (2006) (16) in which evidence is presented for the idea that various sub-populations of man making up the present world population evolved from a small founder population that migrated out of Africa in successive waves about 50,000 to 60,000 years ago. Another related theme is that humans also evolved in genetically significant ways with the development of agriculture during the last 10,000 years, which provided a means to gather in communities in which further cultural evolution occurred at various geographic regions around the world. Cultural evolution in turn, which progressed on a shorter time scale than biological evolution, made it possible for populations to grow to unprecedented sizes, which increased the probability that beneficial mutations would emerge and become established in a population. For an extensive account of this view that human populations have undergone significant genetic changes during the last 10,000 years, the interesting book of Cochran and Harpending (2009) (3) with the title, "The 10,000 Thousand Year Explosion" may be consulted.

The views just discussed motivated the computer experiment that will be reported in this section, using the stochastic framework described in the preceding sections of this chapter. To implement the embedded deterministic model described in section 5, 4,000 generations of evolution were considered. If one assumes that in humans, the length of a generation is 15 years, the age most human females become fertile, then 4,000 generations would represent about $4,000 \times 15 = 60,000$ years, which falls within the "out of Africa" hypothesis. If, however, a generation is assumed to be 20 years, then 4,000 generations would represent $4,000 \times 20 = 80,000$ years. Thus, under both assumptions as to the time length of a generation, 4,000 generations are adequate to cover the evolutionary time periods under consideration. As indicated in previous sections of this chapter, the genetic evolution of a population will be considered with respect to an autosomal locus with two alleles with the set $\mathfrak{T} = (AA, Aa, aa) = (1, 2, 3)$ of genotypes common to both sexes.

In the experiment reported in this section, two components of natural selection were considered. The first was that sexual selection was in force among males so that the matrix of acceptance probabilities for males were given the assinged values

$$\boldsymbol{A}_m = \begin{pmatrix} 0.1 & 0.1 & 0.9 \\ 0.1 & 0.1 & 0.9 \\ 0.1 & 0.1 & 0.9 \end{pmatrix} . \tag{6.1}$$

Observe that according to this matrix, males of all genotypes 1, 2, and 3, prefer females of genotype $3 = aa$ as sexual partners. By assumption sexual selection among females was neutral in the sense that each element in the matrix of acceptance probabilities $\boldsymbol{A}_f$ was assigned the value 1. It was also assumed that the probability per generation that the probability $p(\kappa)$ that a pair of type κ formed a couple was $p(\kappa) = 0.9$ for all $\kappa \in \mathfrak{K}$.
Another component of natural selection considered in this experiment was that of differential reproductive success as expressed in terms the matrix of expected values

$$\boldsymbol{\Gamma} = \begin{pmatrix} 3 & 3 & 3 \\ 3 & 3 & 3 \\ 4 & 4 & 4 \end{pmatrix} \tag{6.2}$$

for Poisson distributions. According to these assignments of values, all those couples in which the females were of genotype $3 = aa$ were most successful reproductively in the sense that such couples produced an expected number of 4 offspring per generation; whereas the other couple types produced on average only 3 offspring per generation.
Among the driving forces of evolution is that of mutation among the two alleles under consideration. To take mutation into account is was assumed that the mutation matrix had the numerical form

$$\mathfrak{M} = \begin{pmatrix} \mu_{11} & 10^{-6} \\ 10^{-7} & \mu_{22} \end{pmatrix}, \tag{6.3}$$

where the diagonal elements μ_{11} and μ_{22} were chosen such that the sum of each row the matrix $\mathfrak{M}$ was 1. Observe that by assumption the mutation $A \to a$ occurred with probability 10^{-6} per generation, but the back mutation $a \to A$ occurred with probability 10^{-7} per generation.
With respect to the survival of the offspring for each generation, it was assumed that natural selection was neutral in the sense that all parameters in the Weibull survival functions were the same, see formula (4.5). Let $\boldsymbol{\beta}_f = \left(\beta_{f1}, \beta_{f2}, \beta_{f3}\right)$ denote the vector of alpha parameters for each of the three female genotypes and let $\boldsymbol{\beta}_m$ denote the corresponding vector of parameters for males. Throughout of all experiments reported in this chapter, the beta parameters for females and males were assigned the common values $(2, 2, 2)$. Let alpha parameters for females and males be denoted by the vectors $\boldsymbol{\beta}_f$ and $\boldsymbol{\beta}_m$, respectively. For the experiment reported in this section, both the beta vectors of the Weibull survival function were assigned the common values $\boldsymbol{\beta}_f = \boldsymbol{\beta}_m = \left(10^{-7}, 10^{-7}, 10^{-7}\right)$. Another way of viewing these parameter assignments was that the environment in which the population evolved could support about 10,000,000 individuals of each sex in any generation.
It was also assumed that the evolution of the hypothetical population under consideration evolved from a small founder population with the initial number of 100 females of genotypes $1 = AA$ as symbolized by the vector $\boldsymbol{X}(0) = (100, 0, 0)$ and that the initial population vector for males had the form $\boldsymbol{Y}(0) = (105, 0, 0)$, indicating that only genotype $3 = AA$ was present in the initial female and male populations. Finally, it was assumed the p_f, the probability an offspring was female, had the value $p_f = 100/205$ and that for males had the value $p_m = 105/205$. Throughout all experiments reported in this chapter, it was assumed that the initial conditions and values of parameters stated in this paragraph were in force.
Presented in Figure 1 are the graphs of the trajectories of the three genotypes, which were computed using the embedded deterministic model described in section 5. In this figure, the horizontal axis is expressed in generations and the vertical axis denotes the estimated number of individuals of each genotype present in the population in each of the 4,000 generations.

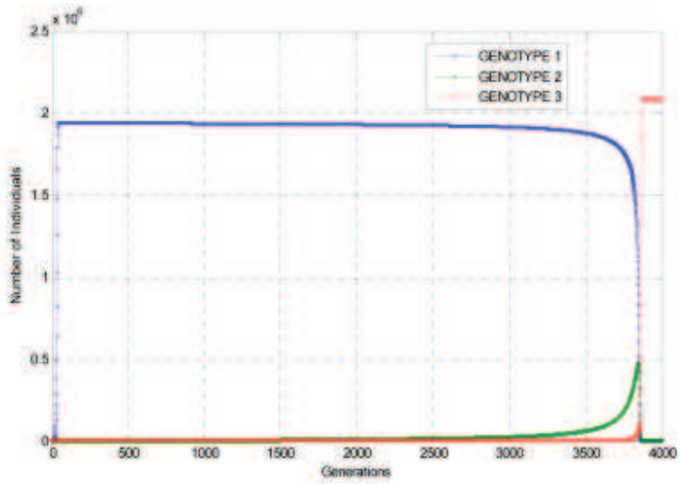


Fig. 1. Graphs of the Trajectories for the Three Genotypes as Computed by the Deterministic Model for the Female Population

As can be seen from this figure, the trajectory for genotype $1 = AA$ rises to a little less that 2×10^6 individuals in less than 500 generations into the evolutionary time period considered in the experiments; whereas the heterozygote, genotype $2 = Aa$, did not rise to noticeable numbers until about 2,000 generations of evolution. But, on the other hand, genotype $3 = aa$, which had selective advantages from the point of view of sexual selection and reproductive success, did not rise to noticeable numbers in the population until a little less than 4,000 generations when it rose to predominance among the three genotypes. By generation 4,000 the numbers of individuals of this genotype appear to have converged to a constant with a value greater than 2×10^6, while the numbers of the other two genotypes fell to values near zero. It should also be mentioned that the deterministic graphs for the male population were similar to those presented in 1.

The experiment just described was also carried out with the same parameter values and initial numbers as those used for the deterministic projections displayed in 1, using the two sex stochastic population stochastic process described in foregoing sections of this chapter. In this Monte Carlo simulation experiment 100 replications of 4,000 generations of evolution were computed and summarized statistically as described in Mode and Gallop (2008). Furthermore, the random number generators described in that paper were also used in this and all other Monte Carlo simulation experiments that are reported in this and subsequent sections of this chapter. Curiously, in this Monte Carlo experiment, it was observed that none or very few individuals of genotype $3 = aa$ arose in either the female or male sub-populations among the 100 replications of 4,000 generations of evolution.

This observation came as a surprise, because in other Monte Carlo simulation experiments conducted with this model, it was observed that deterministic predictions based on the embedded deterministic model were reasonable predictors for the stochastic process in the long run, even though it would miss entirely the high levels of variation or stochasticity that were observed among the 100 realizations of the process, during the beginning generations of experiments. Similar very high levels of stochasticity in Monte Carlo simulation experiments were also observed during those time periods of evolution when one genotype started its evolution to predominance over the others. Consequently, the contrary observations made in this experiment motivated a search of the parameter space of the model to find examples of

chosen parameter values such that the embedded deterministic model was a better predictor for the trajectories of the stochastic process in the long run. It is also important to mention that without a formulation in which deteministic and stochastic approaches to modeling were considered simultaneously within one structure, this discrepancy between the predictions of a deteministic and a stochastic process would never have been realized. It is also very interesting to note that if Monte Carlo simulation techniques had not been developed with the help of many investigators over several decades, it would have been impossible to carry out the computer experiments reported in this chapter.

7. A Case in which the Deterministic model was a better predictor of the stochastic process

In the computer experiment reported in this section, all parameter values assigned in the experiment reported in the preceding section were in force except that the vectors of beta parameters $\boldsymbol{\beta}_f$ and $\boldsymbol{\beta}_m$ were assigned different values. In particular, the values chosen for these vectors were $\boldsymbol{\beta}_f = \boldsymbol{\beta}_m = \left(10^{-9}, 10^{-9}, 10^{-9}\right)$ so that in this experiment the carrying capacity of the environment for each sex and genotype was about 10^9 or one billion individuals. The value of 10^{-9} was chosen, because in the experiment reported in the preceding section, where the beta value 10^{-7} was used for all genotypes in both sexes, it was thought that in the Monte Carlo simulation experiment the number of individuals of genotype $1 = AA$ in the female population did not become sufficiently large to ensure that genotype $3 = aa$ would occur with sufficiently high probability. As will be seen subsequently, the beta value 10^{-9} was sufficient to ensure that the genotype $3 = aa$ arose in the early generations of the experiment and in the long run became the predominant genotype in the population. It should be mentioned that, even by today's standards, a population of a billion people or more people would be considered large and may be distributed over a geographic area which would encompass several countries.

Presented in Figure 2 are the trajectories for the three genotypes, which were computed using the deterministic model embedded in the stochastic process for the first 200 generations of a computer simulation experiment simulating 4,000 generations of evolution.

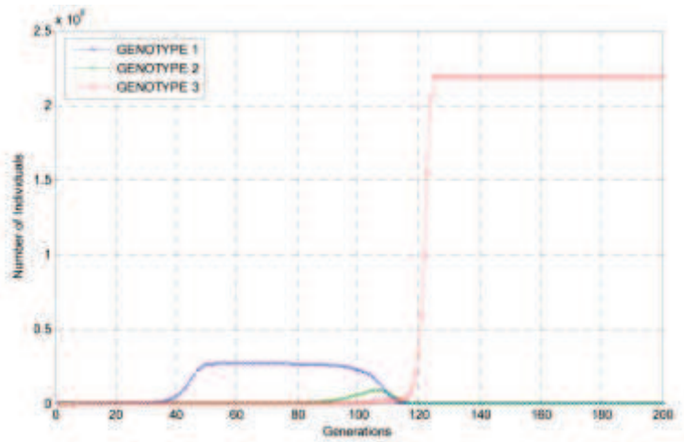


Fig. 2. Deterministic Trajectories for the Three Genotypes in the Male Population for the First 200 Generations

As can be seen from 2, genotype $3 = aa$ rose to predominance in the male population in many fewer generations than that shown in 1 for the population of females. In this figure, it can be seen that by about 50 generations of evolution, the number of individuals of genotype $1 = AA$ had risen to about $0.4 \times 10^8 = 40,000,000$ individuals, which was sufficiently large to ensure that the genotype $3 = aa$ did appear the in population after of about 120 generations of evolution. After appearing in the population, this genotype quickly rose to predominance and reached a constant value of over 2^8 individuals in the male population. It should also be mentioned that, had the graphs for the female population been plotted, they would have been very similar to those for the male population displayed in 2. In this experiment, the numbers of individuals of each of the three genotypes converged to constants, which has been the case for most experiments conducted with the embedded deterministic model.
Presented in Figure 3 are the $Min, Q50, Max, Mean$ and SD of the trajectories for the population of individuals of genotype $3 = aa$, which were estimated from the first 200 generations of a sample Monte Carlo realizations of the stochastic process based on 100 replications of evolution for 4,000 generations. By way of definitions, the symbols $Min, Q50, Max, Mean$ and SD represent, respectively, the minimum, median, maximum, mean and standard deviation of 100 realizations of the process at each of the 200 generations shown in 3.

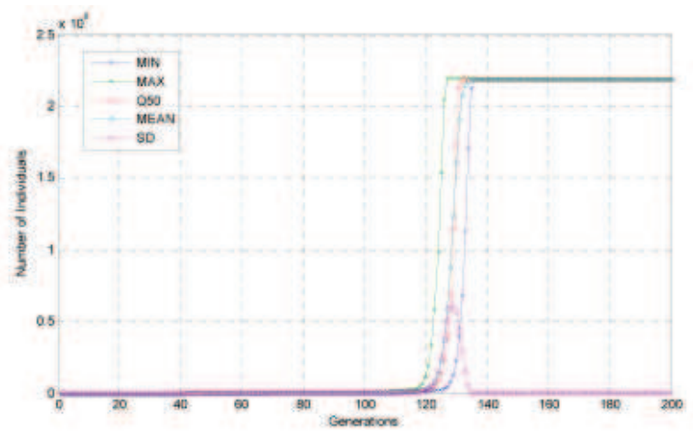


Fig. 3. Estimated Statistical Trajectories for Individuals of Genotype aa **in the Male Population in the First 200 Generations of Evolution**

As can be seen from this figure, summarizing the results of a Monte Carlo simulation experiment, genotype $3 = aa$ did not begin to appear in the population until about 100 generations into the projection as indicated by the Max trajectory, in contrast the experiment reported in the previous section in which a significant number of the genotype did not appear during 4,000 generations of evolution. At somewhat more than 120 generations, the Min trajectory for the genotype $3 = aa$ begins its ascent to higher values and the $Q50$ and $Mean$ trajectories, which lie between the Min and the Max also begin their ascent to higher values. At about 140 generations of evolution, all trajectories are close to each other and settle in to values over 2×10^8 individuals. It is also very interesting to note that during of the period between about 120 and 140 generations the trajectory for the standard deviation, SD, rises to a maximum of about 0.5×2^8 individuals before it declined to small values at about 140 generations of evolution. This rise and fall of the SD trajectory is a signature of the presence of high levels of stochasticity during the evolutionary period that genotype $3 = aa$ was rising

to predominance in the population. If one relied only on the trajectories computed using the deterministic model, this high level of stochasticity would have been missed entirely. In should also be mentioned in passing, that if trajectories for the female population had been plotted, they would have been very similar to those in 3.

It is also of basic interest to mention that the close proximity of the Min, Q50, Max and Mean trajectories after about 140 generations of evolution is indicative of a stochastic process converging to a quasi-stationary distribution for which the level of variation among the realizations of the process is low as indicated by small values of the SD trajectory after about 140 generations of evolution. No formal treatment of this statement will be given here, except to mention that the stochastic process under consideration may be viewed as a Markov chain with a state space that will not be described here. But, if a reader is interested pursuing the subject of quasi-stationary distributions further, it is suggested that this phrase be typed into a search engine for the internet, where many references to this concept may be found.

If a generation time in human population is in the range of 15 to 20 years, then 200 generations would consist of $200 \times 15 = 3,000$ to $200 \times 20 = 4,000$ years of evolution. The notion that a case for sexual selection in humans may be made was the thought that in some populations males preferred female sexual partners with blue eyes, which are inherited as a recessive homozygous autosomal genotype aa. Of course, if it were assumed that females or both females and males prefer sexual partners with blue eyes, the results of such computer experiments would not have differed much from the one reported in this section. Consequently, the experiment reported in this section lends credence to the notion advanced by Cochran and Harpending (2009) (3) and others that during the last 10,000 years as agriculture developed and populations became large, blue eyes may have risen by mutation and rose to significant frequencies in some human populations due to the process of sexual selection or perhaps some other form of selection. Computer simulation experiments, such as the one reported in this section, are supportive of this notion.

8. Competitive advantage for one genotype is a sufficient condition for evolutionary predominance

In the computer experiment reported in this section, all acceptance probabilities in the matrices $\boldsymbol{A}_f$ and $\boldsymbol{A}_m$ were chosen as 1 so that, by definition, the mating system was random. It was also assumed that each couple type, consisting of a female and male of given genotypes, produced 4 offspring per generation; thus no genotype was had reproductive advantage over the others. All other parameters of the model were the same as those used in the experiment in section 7 except for the beta parameters that were assigned to each genotype for both sexes. In the experiment reported in section 6, it was observed that when the probability of the mutation $A \rightarrow a$ was assigned the value $\mu_{12} = 10^{-6}$, the reverse probability of mutation was assigned the value $\mu_{21} = 10^{-7}$ and all beta parameters were assigned the value 10^{-7}, then the mutant genotype $3 = aa$ rose to predominance in the population when the embedded deterministic model was used to compute the evolutionary trajectories for each of the three genotypes. In the corresponding Monte Carlo simulation experiment, however, only very small numbers of the mutant genotype $3 = aa$ appeared among the 100 replications of 4,000 generations of evolution.

Similar results were observed in a preliminary experiment in which the beta parameters were assigned the values $\boldsymbol{\beta}_f = \boldsymbol{\beta}_m = \left(10^{-7}, 10^{-7}, 10^{-8}\right)$ for both sexes, which, by assumption, gave the mutant genotype $3 = aa$ a competitive edge over the other genotypes. In an attempt to avoid the case in which the deterministic model was not a good predictor for the

evolutionary behavior of stochastic model, the beta parameters for both sexes were chosen as $\boldsymbol{\beta}_f = \boldsymbol{\beta}_m = \left(10^{-8}, 10^{-8}, 10^{-9}\right)$ in the experiment reported in this section. As will be shown in what follows, these choices of parameter values were sufficient to show that after a rather large number of generations of evolution, the mutant genotype $3 = aa$ rose to predominance in the population according to both the deterministic and stochastic models.
Presented in Figure 4 are the evolutionary trajectories for the three genotype in the male population as computed using the deterministic model. As can be seen form this figure, the competitive advantage for genotype $3 = aa$ was sufficient for this genotype to become predominant in the population at about 1,800 generations into the projection. In this deterministic projection, the number of individuals of genotype $3 = aa$ converged to a constant with a value greater than 2×10^8.

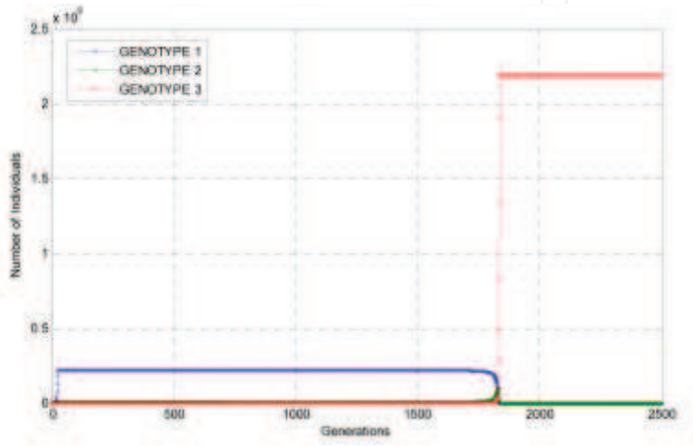


Fig. 4. Trajectories for the Three Genotypes in the Male Population for the First 2,500 Generations in an Experiment with 4,000 Generations

In Figure 5, the trajectories *Min*, *Q*50, *Max*, *Mean* and *SD* are plotted using for the first 2,500 generations of Monte Carlo simulation data consisting of 100 replications of 4,000 generations of evolution.

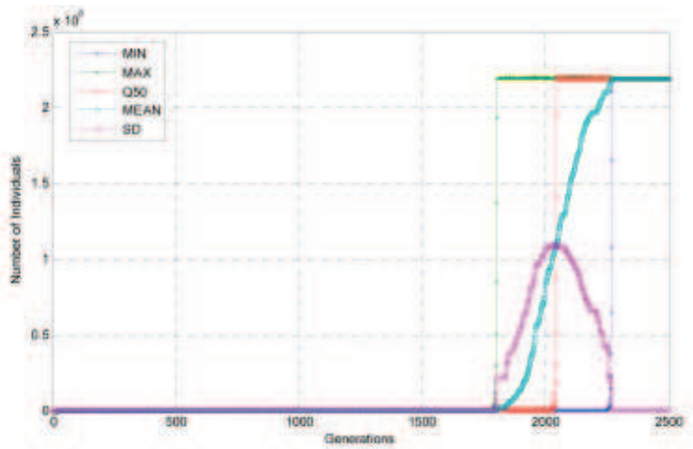


Fig. 5. Estimated Statistical Trajectories for Individuals of Genotype aa **in the Male Population in the First 2,500 Generations of Evolution**

As can be seen from this figure, significant numbers of genotype $3 = aa$ did not appear in the population until about 1,800 generations into a stochastic projection of 4,000 generations as can be seen from the *Max* trajectory as it begins its nearly vertical rise for a level greater than 2×10^8. Unlike the experiment reported in section 7, the level of variability or stochasticity was higher in this experiment in that the *Min* trajectory did not begin its steep rise until about 2,250 to 2,300 generations into the stochastic projection. Interestingly, the *Q*50 trajectory begins its steep rise at a little more than 2,000 generations, which is approximately the midpoint between the steep rise of the *Max* and *Min* trajectories. That the level of stochasticity among the realizations of the process was high can be seen by observing that the *SD* trajectory had a maximum greater than 1×10^8 at about 2,000 generations of evolution. Another interesting feature of 5 is that, unlike previous experiments with the stochastic model, the *Q*50 and *Mean* trajectories are distinct, during the evolutionary period when genotype $3 = aa$ was rising to predominance in the population. The coalescence of all trajectories in this figure at about 2,500 generations is indicative of a stochastic process converging to a quasi-stationary distribution with relatively low levels of variability among the realizations of the stationary process.
When compared with the experiment reported in section 7, the pace of evolution in this experiment was rather slow. For example, if in humans a generation time is in the interval 15 to 20 years, then the period of evolutionary time covered in the above figures was in the range $2,500 \times 15 = 37,500$ to $2,500 \times 20 = 50,000$ years. It is interesting to note that this range is within the estimated range of the "Out of Africa Hypothesis", in which it is postulated that all humans that inhabit the earth today are descendants of waves of human migration out Africa about 50,000 to 60,000 years ago.

9. An Example of slow evolution when mutation probabilities are small

In the experiment reported in this section, all values of parameters used in the experiment in section 8 were retained except that the probabilities of mutation per generation were chosen to be much smaller and the capacities to compete with other individuals for resources, which depended on the genotype, were chosen to be orders of magnitude smaller. In particular, the mutation probabilities per generation were assigned the values $\mu_{12} = 10^{-8}$ and $\mu_{21} = 10^{-9}$ and the beta parameters for both sexes were assigned the values $\boldsymbol{\beta}_f = \boldsymbol{\beta}_m = \left(10^{-10}, 10^{-10}, 10^{-12}\right)$. Observe in this experiment, the beta values 10^{-10} were two orders of magnitude smaller than the mutation probability $\mu_{12} = 10^{-8}$ of the mutation $A \rightarrow a$, which was analogous to the values used in the experiment reported in section 8. In terms of actual population size, these choices of beta parameters indicate, by assumption, that individuals of genotypes $1 = AA$ and $2 = Aa$ of either sex could compete for resources only in populations of sizes up to about 10^{10} individuals; whereas individuals of genotype $3 = aa$ of either sex could compete for resources successfully in populations of sizes up to 10^{12} individuals. It is suggested that reader compare these numbers with the estimated size of the present world population of humans, which is thought to lie in the interval 6 to 7 billion individuals.
In preliminary experiments with the deterministic model, it was observed that no individuals of genotype $3 = aa$ appeared in the population after 10,000 generations of evolution. Consequently, it was decided to do an experiment which simulated 20,000 generations of evolution based on the deterministic model. As can be seen from 6 below, mutant genotype $3 = aa$ did not rise to predominance in the population until about 18,000 generations of evolution when its numbers rose almost vertically to a constant number of individuals with a value of over 2×10^{11}.

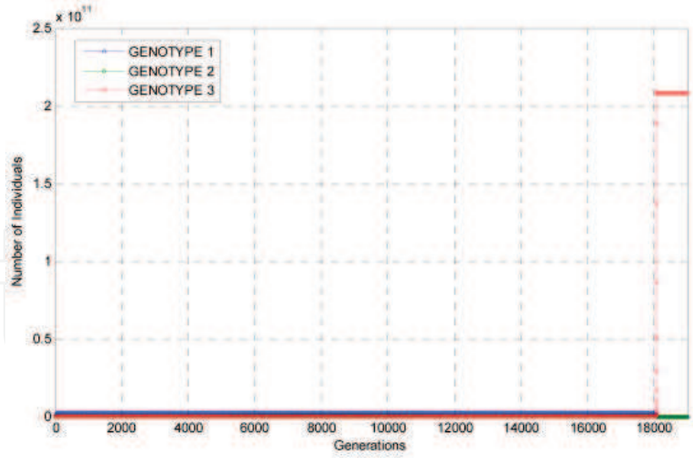


Fig. 6. Evolutionary Trajectories for the Three Female Genotypes as Computed Using the Deterministic Model

In terms of human evolution with generation times of 15 to 20 years, the time period of evolution covered in this experiment was in the range $20,000 \times 15 = 300,000$ to $20,000 \times 20 = 400,000$ years. It is thought that Homo erectus, a hominid species ancestral to man, evolved in Africa about 1.8 million years ago and spread out of Africa to as far as China. For further information on this extinct species, it is suggested that the internet be consulted. When compared to 1.8 million years of evolution time periods of 300,000 to 400,000 years are relatively short. Even though in terms of recent evolutionary history of the hominid line, it is doubtful whether population sizes ever attained those considered in this experiment, but it is plausible that other combinations of numerical values of parameters could be found such that genotype $3 = aa$, with a superior ability to compete for resources, could rise to predominance in a population within 300,000 to 400,000 years. With reference to human evolution, it is plausible that a superior ability to compete for resources would be associated with increased cognitive abilities, which may be in turn associated with enhanced neural networks in the brains of individuals.

No attempt was made to do a stochastic version of the experiment reported in this section by computing Monte Carlo simulations of the process for 100 replications of 20,000 years of evolution, because it was doubtful that such an experiment could be completed in an acceptable period of time. It was also doubtful if the desk top computers available to do such an experiment had sufficient memory capacity to store the rather large arrays that would have been generated.

10. Discussion

From the experiments reported in the preceding sections on the evolution of an autosomal locus with two alleles by mutation and selection, a general conclusion that may be drawn is that whether a beneficial mutation becomes established in a population depends not only the probability of mutation per generation that an ancestral form of a gene mutates to a more beneficial form but also on the ability of a mutant genotype to compete with other individuals for available resources. At the molecular level many mechanisms may underlie the ability of a mutant genotype to compete with other individuals, but in this section attention will be focused only on two mechanisms that may be involved in this component of natural selection.

One mechanism that comes to mind is that a mutant genotype may be able to process available sources of food more efficiently, which may, for example, involve some change in the chemical mechanisms of digestion such that an individual of a mutant genotype can extract more calories from the same amount food than as an individual with an ancestral genotype and would thus have more energy to expend in the search for food and other activities. Another mechanism that may be involved in a beneficial mutation is that individuals of a mutant genotype may have increased cognitive abilities, which result in the invention of technologies and methods to greatly expand the amount of food that can be extracted from a given geographical region. Another beneficial impact of increased cognitive abilities is that such individuals may be more adept at passing on information from generation to generation and would thus increase the survivability of individuals with such genotypes over many generations. It is beyond the scope of this chapter to further pursue the evolution of the human brain, but, it is suggested that an interested series of lectures on this and other subjects by Sapolsky (2005) (14) be consulted.

One of the limitations of the class of branching processes described in this chapter is that attention has been focused only on discrete generations of individuals and not on accommodating the existence of many birth cohorts in a population at any time, which provides a milieu for the passing of information from the older to younger individuals. This passing of information and techniques from older to younger individuals is among the defining characteristics of human populations, and there is little doubt that this educational process has played a significant role in the development of civilization as we know it today. An example of the outcome of this type of activities in human populations are the inventions of techniques and the passing on of information from generation to generation in man as agriculture developed during the last 10,000 years. It is also beyond the scope of this chapter to give a detailed account of the evolution of agriculture and the rise of urban communities, but it is suggested that an interested reader consult a series of lectures by Sojka (2009) (15), where this and other related subjects are treated in more detail.

In chapter 12 of the forthcoming book, Mode and Sleeman (2010-2011) (13), the two sex branching process described in this chapter has been extended to an age dependent process in a preliminary formulation that accommodates over lapping generations or the presence of many birth cohorts in a population at any time. Within the framework of this formulation, many more components of natural selection can be considered than in the formulation described in this chapter. Among these components is a parametric survival function for long lived species such as humans, which provide a means for calculating conditional probabilities of death or survival for each sex, given an age group to which an individual belongs. In principle, these parameterized conditional probabilities provide a means for changing the conditional probabilities of death or survival for infants and the elderly as a population evolves in a simulation experiment as conditions for survival either improve or decline over time. Within this age dependent framework, components of natural selection involving altruism may also be included in the formulation in terms of parental care of the young, which increases their probabilities of survival. This age dependent formulation also has the self regulating property in the sense that whether an individual of any age group or sex survives to reproduce also depends of total population size as described in this chapter and elsewhere. Although the case that age dependence should be included in mathematical models of human evolution that accommodate age dependence is easy to justify, including age groups in a formulation as well as a set of possible genotypes leads to the necessity of processing large arrays in a computer, which can be problematic even if networks of computers are available to

an investigator or team of investigators. Among the problems that arise is the need to consider many couple types, which give rise to very large arrays, particularly when sexual selection according to age as well as genotype or phenotype are included in a formulation. These problems become particularly acute when computing samples of Monte Carlo realizations of the process and inventing procedures for informative statistical summarizations of Monte Carlo simulation data of high dimension. For these reasons, in the preliminary computer experiments reported in chapter 12 of the forthcoming book (13) attention has been confined to the deterministic model embedded in the stochastic process. Even in this deterministic case, a special sub-case in which females choose male sexual partners without couple formation was implemented to avoid the necessity of processing large arrays that arise when all possible couple types are considered. Some cases were also implemented when the evolution of a hypothetical species may be described in terms of only a few age groups. These cases served as a "proof of principle" that the formulation could be implemented on a computer even though much more work needs to be done on the age dependent formulation. It is hoped that in the future other investigators will join in the quest to develop stochastic age dependent self regulating population processes that may be implemented on computers. Even though these age dependent formulations involve high dimensional structures, the implementation of Monte Carlo simulation methods will play an essential role in developing an understanding of these models.

There is one other property of the embedded deterministic model considered in this chapter as well as those described in the forthcoming book Mode and Sleeman (2010-2011) (13) that should be mentioned. Without exception, when some combinations of numerical values of parameters are considered in computer experiments, the trajectories, which were computed using a deterministic model, become chaotic after a small number of generations of evolution. If a reader in interested in the subject chaos in deterministic systems, it is suggested that the book Gulick (1992) (5) and others be consulted. In chapter 9 of the forthcoming book, Mode and Sleeman (2010-2011) [13], examples of chaotic behavior of the embedded deterministic model are presented along with statistical trajectories estimated from samples of Monte Carlo simulation for a case of only one type of individual. As it turned out, in some cases the behavior of the statistical trajectories displayed more regularity than the chaotic deterministic trajectories. It is suggested that an interested reader consult chapter 9 of the forthcoming book for more details on cases of chaotic behavior in one type self regulating branching processes. For the sake of simplicity, no cases of chaotic behavior in the embedded deterministic model have been included in this chapter.

11. References

[1] Asmussen, S. and Hering, H. (1983), Branching Processes. Birkhauser, Boston, Basel, and Stuttgart.
[2] Athreya, K. B. and Ney, P. (1972), Branching Processes. Springer-Verlag, Berlin, Heidelberg, New York.
[3] Cochran, G. and Harpending, H. (2009) The 10,000 Year Explosion - How Civilization Accelerated Evolution Basic Books. New York.
[4] Fisher, R. A. (1958) The Genetical Theory of Natural Selection - Second Revised Edition. Dover Publications, New York, N.Y.
[5] Gulick, D. (1992), Encounters With Chaos. McGraw-Hill, New York, London

[6] Haccou, P., Jagers, P. and Vatutin, V. A. (2007) Branching Processes - Variation, Growth and Extinction of Populations. Cambridge University Press, The Edinburgh Building, Cambridge, CB2 8Ru, UK
[7] Harris, T. E. (1963), The Theory of Branching Processes. Springer-Verlag, Berlin, Heidelberg, New York.
[8] Jagers, P. (1975), Branching Processes with Biological Applications. John Wiley and Sons, London, New York, Sydney, Toronto.
[9] Mode, C. J. (1971), Multitype Branching Processes - Theory and Applications. American Elsevier, New York, Toronto.
[10] Mode, C. J. (1995) An Extension of the Galton-Watson Process to a Two Sex Density Dependent Genetic Model. Proceeding of the First World Conference on Branching Processes. Edited by C. Heyde, Springer Lecture Notes in Statistics. 99:152-168.
[11] Mode, C. J. and Gallop, R. J. (2008) A Review on Monte Carlo Simulation Methods as They Apply to Mutation and Selection as Formulated in Wright-Fisher Models of Evolutionary Genetics. Mathematical Biosciences 211: 205-225.
[12] Mode, C. J. and Sleeman, C. K. (2000) Stochastic Processes in Epidemiology - HIV/AIDS, Other Infectious Diseases and Computers. World Scientific, New Jersey, London, Hong Kong.
[13] Mode, C. J. and Sleeman, C. K. (2010-2011) Stochastic Processes in Genetics and Evolution - Computer Experiments in the Quantification of Mutation and Selection. World Scientific, New Jersey, London, Hong Kong.
[14] Sapolsky, R. (2005) Biology and Human Behavior: The Neurological Origins of Individuality The Teaching Company. Chantilly, Virginia, USA. www.teach12.com.
[15] Sojka, G. A. (2009) Understanding the Human Factor: Life and its Impact The Teaching Company. Chantilly, Virginia, USA. www.teach12.com.
[16] Wells, S. (2006) Deep Ancestry - Inside the Genographic Project. National Geographic Society, Washington, D. C. , U.S.A.

IntechOpen



Applications of Monte Carlo Methods in Biology, Medicine and Other Fields of Science

Edited by Prof. Charles J. Mode

ISBN 978-953-307-427-6
Hard cover, 424 pages
Publisher InTech
Published online 28, February, 2011
Published in print edition February, 2011

This volume is an eclectic mix of applications of Monte Carlo methods in many fields of research should not be surprising, because of the ubiquitous use of these methods in many fields of human endeavor. In an attempt to focus attention on a manageable set of applications, the main thrust of this book is to emphasize applications of Monte Carlo simulation methods in biology and medicine.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Charles J. Mode, Towfique Raj and Candace K. Sleeman (2011). Monte Carlo Implementations of Two Sex Density Dependent Branching Processes and their Applications in Evolutionary Genetics, Applications of Monte Carlo Methods in Biology, Medicine and Other Fields of Science, Prof. Charles J. Mode (Ed.), ISBN: 978-953-307-427-6, InTech, Available from: http://www.intechopen.com/books/applications-of-monte-carlo-methods-in-biology-medicine-and-other-fields-of-science/monte-carlo-implementations-of-two-sex-density-dependent-branching-processes-and-their-applications-

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2011 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License, which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen