Mathematical Models for Population Dynamics: Randomness versus Determinism

Jean Bertoin

Mathematical Models for Population Dynamics: Randomness versus Determinism

Jean Bertoin

Abstract. Mathematical models are used more and more frequently in Life Sciences. These may be deterministic, or stochastic. We present some classical models for population dynamics and discuss in particular the averaging effect in the setting of large populations, to point at circumstances where randomness prevails nonetheless.

2010 Mathematics Subject Classification. Primary 92D25; Secondary 60J85.
Keywords. Population models, determinism, randomness.

1. Introduction

Without any doubt, Biology is amongst the sciences in which advances accomplished during the last century have been the most spectacular. For mathematicians, it is both a source of inspiration (for instance, genetic algorithms mimic natural selection to solve optimization problems), and raises formidable challenges, notably in the field of modeling. Actually, most Life Sciences require pertinent mathematical models, which should not only fit experimental data, but more importantly, should enable practitioners to make reliable analysis and predictions (for instance, one wishes to predict the outbreak of an epidemic and prevent its occurence by an appropriate vaccination program). These mathematical models may be *deterministic*, in the sense that the outputs are entirely determined by the values of the parameters of the model, or *stochastic*, when the model incorporates inherent randomness and outputs then depend not only on the parameters but also on some additional stochastic elements.

The study of the dynamics of populations is a tool of fundamental importance in this area, notably in Genetics, Ecology, and Epidemiology, to name just a few. In general, deterministic models in this field concern global or averaged features of the population, typically the size of certain sub-populations, or the proportion of individuals sharing certain characteristics. That is, the features of the population are averaged and the model aims at depicting the evolution of those averaged quantities as time passes. They are based on the implicit assumption that, roughly speaking, all individuals in a given sub-population behave essentially the same. Dynamics are usually modeled in discrete times through some difference equations, and through differential equations in continuous times. The reader is referred to the textbooks by Allman and Rhodes [1], Edelstein-Keshet [10] or Hofbauer and Sigmund [15] for some basic models in this framework and their biological motivations, and to the monograph by Bürger [5] for a comprehensive overview.

In turn, stochastic models are built either by adding noise terms to deterministic evolution equations, in order to take random fluctuations into account, or more interestingly, by considering individual behaviors which are then viewed as stochastic processes. Individual-based models permit in particular to consider how individuals collaborate or compete with each other for resources, or interact with their environment. Stochastic models of population dynamics rely essentially on Markov chains in discrete times, Markov jump processes and stochastic differential equations in continuous times, including notably branching processes and coalescent processes. We refer in particular to the monographs by Durrett [9], Haccou, Jagers and Vatutin [13], and Hein, Schierup, and Wiuf [14], and the lecture notes by Dawson [8] and Etheridge [11].

"All models are wrong but some are useful", George Box used to say. The mechanisms driving dynamics of populations in nature are extremely intricate and involve a number of diverse features, whereas mathematical models must remain tractable and thus can only incorporate a few of them. In general, mathematical models for highly complex phenomena focus on a few key variables, and view the effects of the remaining ones as small (possibly random) perturbations of the simpler model. In this respect, deciding whether to opt for a deterministic versus a stochastic model may be a delicate issue. Deterministic models are simpler to solve analytically or numerically; random models can be considerably more complicated, in particular in the individual-based case, but it is generally admitted that they may be also more realist. One may wonder whether it is useful to handle more complex random models when a deterministic answer is expected anyway, and at the opposite, one may be concerned with the risk of missing some important consequences of randomness by making an oversimplified deterministic analysis.

Of course, a first key question is whether a given mathematical model accurately describes a phenomenon of interest, which is usually answered by checking the agreement with experimental measurements. Once the scope of a model has been validated and the model is applied in concrete situations, another fundamental problem for practitioners is the comparison with available data in order to estimate its parameters, and then to be able to make reliable predictions about the future (or inferences about the past) of the population. Thus statistical analysis plays a crucial role in this area; see for instance the books by Allman and Rhodes [1] and Turchin [20]. However here we shall not discuss applied statistical aspects and rather focus on more theoretical issues.

In this text, we shall first briefly present some of the simplest, best known and widely used models for population dynamics, both in the determinist and the random frameworks. Starting from the most elementary models, which were introduced in the 18th and 19th centuries, and in which the reproduction mechanism is assumed to independent of the characteristics of the population, we shall then discuss how models have been thereafter modified and complexified in order to incorporate some more realistic features. We further address the question of possible averaging effects, which may suggest that for large populations, determinist models should prevail, and then point at simple situations for which naive intuitions may fail.

2. Some classical population models

In this section, we briefly review some classical population models, deterministic or random, both for discrete and continuous times. We shall mainly focus the simple case where each individual has a single parent, which corresponds to haploid populations (i.e. individuals have only one set of chromosomes). However this can be also relevant for sexual reproduction, either by considering the subpopulations of individuals of the same sex (in most diploid populations, mitochondria DNA are inherited exclusively from the mother), or by viewing a diploid population with N individuals each carrying a pair of chromosomes as a haploid population with $2N$ individuals each having a single chromosome.

2.1. Exponential growth model and branching processes. The simplest of all population models was considered by T.M. Malthus at the very end of the 18th century. If the size of a population at date t is measured by $P(t)$, where $t \geq 0$ is either an integer or a real number, then one assumes that $P(t)$ grows at constant rate $r \in \mathbb{R}$ in time. That is, in discrete times, the increment of the population has the form

$$\Delta P(t) := P(t+1) - P(t) = rP(t), \tag{1}$$

whereas in continuous times,

$$\frac{\mathrm{d}P(t)}{\mathrm{d}t} = rP(t).$$

In terms of the initial population size $P(0)$, one thus gets

$$P(t) = (1+r)^t P(0) \quad \text{for the discrete time version,}$$

and

$$P(t) = \mathrm{e}^{rt} P(0) \quad \text{for the continuous time version.}$$

An individual-based stochastic counterpart of the Malthus growth model in discrete time was introduced first in the middle of 19th century by I.J. Bienaymé, and then re-discovered nearly 20 years later by F. Galton and H.W. Watson. Originally, F. Galton was motivated by the study of the extinction of family names; the model can also be useful for describing, for instance, the initiation of a nuclear chain reaction, or the early stages of the spread of contagious diseases.

The building block is an integer valued random variable ξ, which represents the number of children of a typical individual. The probability distribution of ξ is called the reproduction law. One imagines that at each generation, each individual i is replaced by a random number ξ_i of individuals (the children of i), where ξ_i has the same law as ξ, and for different individuals, the ξ_i's are given by independent random variables. If $Z(n)$ denotes the number of individuals at the n-th generation, the chain $(Z(n))_{n\in\mathbb{N}}$ is known as a *Bienaymé-Galton-Watson process* (thereafter in short, BGW process), and its dynamics are depicted at each generation by the identity

$$Z(n+1) = \xi_1^N + \cdots + \xi_{Z(n)}^N \tag{2}$$

where $\xi_1^N, \ldots$ denote independent copies of the variable ξ.

BGW processes are merely elementary prototypes of more sophisticated branching processes, which can be used to model more accurately a variety of dynamics. In short, multi-type branching processes cover the situation where individuals in the population may have different types, and the reproduction law of each individual depends on its type. Branching processes can also be defined in continuous time, possibly with values in $[0, \infty)$ rather than merely in $\mathbb{N}$. They can incorporate phenomena such as immigration, spatial displacements, migration, mutations, random environments, etc. The so-called Crump-Mode-Jagers processes deal with situations where the rate of reproduction of an individual may depend on a number of characteristics of that individual, including for instance, its age, and in particular the reproduction rate varies as time passes. This gives access to probabilistic models of age-structured populations. We refer to the book by Haccou, Jagers and Vatutin [13] for much more on this topic.

The connection with the Malthus growth model in discrete times is easy to explain. Assume that the variable ξ has a finite mathematical expectation $\mathbb{E}(\xi)$. Then it follows from (2) that there is the identity

$$\mathbb{E}(Z(n+1)) = \mathbb{E}(\xi) \times \mathbb{E}(Z(n)),$$

so that, if we set $P(n) = \mathbb{E}(Z(n))$, then we get (1) with $r = \mathbb{E}(\xi) - 1$.

BGW processes fulfill the fundamental *branching property*, which can be viewed as the stochastic analogue of the elementary additivity property of the Malthus growth model. If $(Z(n))_{n\in\mathbb{N}}$ and $(Z'(n))_{n\in\mathbb{N}}$ are two independent BGW processes with the same reproduction law, then their sum, $S(n) := Z(n) + Z'(n)$ is again a BGW process, of course with the same reproduction law. Combining the branching property with the Law of Large Numbers, one sees that the Malthus growth model can be viewed as the limit of BGW process started with a large initial population. Indeed, if $(Z_1(n))_{n\in\mathbb{N}}, (Z_2(n))_{n\in\mathbb{N}}, \ldots$ denote independent BGW processes with the same reproduction law, each started with a single ancestor, then $S_k(n) := Z_1(n) + \cdots + Z_k(n)$ is a BGW process started with k ancestors, and the Law of Large Numbers entails that, provided that $\mathbb{E}(\xi) < \infty$,

$$\lim_{k\to\infty} \frac{1}{k} S_k(n) = \mathbb{E}(Z(n)) = P(n).$$

This suggests that, on average, when the number of ancestors is large, the population should increase exponentially fast when $\mathbb{E}(\xi) > 1$ (the super-critical case), should decay exponentially fast when $\mathbb{E}(\xi) < 1$ (the sub-critical case), and should remain roughly stable when $\mathbb{E}(\xi) = 1$ (the critical case). However, this is not entirely correct; in fact a critical BGW process always become eventually extinct, despite of the fact that the mathematical expectation of $Z(n)$ remains constant, and no matter how large the initial population is. The analysis of the extinction is a cornerstone of the theory of branching processes; see in particular Haccou, Jagers and Vatutin [13].

The fact that in the super-critical case, the size of the population increases indefinitely exponentially fast is clearly unrealistic. Thus modeling the dynamics

of a population by a BGW process can only be pertinent at early stages of its development, and one needs different models to describe the long term behaviors.

2.2. Models with regulated growth. The fact that exponential growth of populations is unrealistic for a large time horizon yield P.-F. Verhulst to introduce in 1838 the so-called *logistic equation*[1] to describe the dynamics of populations with self-limiting growth. Roughly speaking, the underlying idea is that the rate of growth should be proportional to both the existing population and the amount of available resources, and informally, the effect of the latter is to slow down the growth of the population when it is already large. The equation, in continuous time, has the form

$$P'(t) = \frac{\mathrm{d}P(t)}{\mathrm{d}t} = rP(t)(1 - P(t)/K)$$

where $r \geq 0$ should be thought of as the growth rate in absence of self-regulation (typically when the population is small), and $K > 0$ is known as the carrying capacity.

The logistic equation can also be written in the form

$$P'(t)/P(t) = r(1 - P(t)/K);$$

observe that the rate of growth $P'(t)/P(t)$ is positive when $P(t) < K$, negative when $P(t) > K$, and approaches 0 when $P(t)$ is close to K. The carrying capacity K corresponds to the limiting size of the population when times goes to infinity. Indeed, the logistic equation can be solved,

$$P(t) = \frac{KP(0)\mathrm{e}^{rt}}{K + P(0)(\mathrm{e}^{rt} - 1)},$$

so in particular $\lim_{t\to\infty} P(t) = K$.

During the early 20th century, A.J. Lotka proposed a related equation for predator-prey systems, which was then later on re-derived independently by V. Volterra. Typically, consider a population of prey, whose size at time t is denoted by $N(t)$, and a population of predators, whose size at time t is denoted by $P(t)$. Imagine that the population of prey grows naturally at constant rate and that the presence of predators induces an additional rate of decay proportional to the size of the population of predators. That is,

$$\frac{\mathrm{d}N(t)}{\mathrm{d}t} = (a - bP(t))\, N(t) \tag{3}$$

where a and b are two positive constants. In turn, the population of predators declines naturally at a constant rate, and only increases in the presence of preys with rate proportional to the size of the population of prey:

$$\frac{\mathrm{d}P(t)}{\mathrm{d}t} = (-c + dN(t))\, P(t), \tag{4}$$

[1]In turn, the equation was re-derived several times in the sequel, notably by R. Pearl, and is sometimes also known as the Verhulst-Pearl equation.

where c and d are also two positive constants. Informally, the growth of the prey population is regulated by the predator population, and vice-versa. The system formed by (3) and (4) is known as the *Lotka-Volterra equations*; although it has no simple explicit solution, it can be proved that solutions are always periodic.

The Lotka-Volterra equations can be modified to incorporate a further logistic growth element reflecting the competition between individuals of the same species (these are known as the competitive Lotka-Volterra equations), or to any number of species competing against each other. We refer the interested reader to Chapter 2 in Hofbauer and Sigmund [15]. We also mention that there exist other differential equations for modeling regulated growth named after, among others, Gompertz, von Bertalanffy and Weibull.

In turn, there exist stochastic versions of the logistic growth equation and of the Lotka-Volterra equations, which mainly rely on stochastic calculus. In particular, Lambert [19] introduced branching processes with logistic growth, which, in the simpler case of continuous processes, are viewed as solutions to the stochastic differential equation

$$\mathrm{d}Z(t) = aZ(t)\mathrm{d}t - bZ^2(t)\mathrm{d}t + c\sqrt{Z(t)}\mathrm{d}W(t)$$

with $(W(t))_{t\geq 0}$ a Brownian motion. Providing a detailed account of the meaning of such stochastic differential equation would drift us too far away from our purpose, let us simply mention that, in comparison with the deterministic logistic equation, the additional stochastic term $c\sqrt{Z(t)}\mathrm{d}W(t)$ is meant to take into account random fluctuations of the model.

Somewhat similarly, the stochastic version of the Lotka-Volterra equations takes the form

$$\begin{cases} \mathrm{d}X(t) &= (aX(t) - bY(t)X(t))\mathrm{d}t + \sigma_1\sqrt{X(t)}\mathrm{d}W_1(t) \\ \mathrm{d}Y(t) &= (cY(t) - dX(t)Y(t))\mathrm{d}t + \sigma_2\sqrt{Y(t)}\mathrm{d}W_2(t) \end{cases}$$

where $(W_1(t))_{t\geq 0}$ and $(W_2(t))_{t\geq 0}$ are two (possibly correlated) Brownian motions. Let us simply observe that, since stochastic integrals have (usually) zero mathematical expectation, $\mathbb{E}(Z(t))$ (respectively, $\mathbb{E}(X(t))$ and $\mathbb{E}(Y(t))$) solve the deterministic logistic (respectively, Lotka-Volterra) equation.

In Section 4, we shall further see that the logistic growth model is also related to other random individual-based models with large constant size population, which describe the evolution of the size of a sub-population carrying a favorable allele in the regime of strong selection.

2.3. Constant size populations. We shall now present some basic (stochastic) population models with constant total size $N \geq 2$, where subpopulations of different types can be distinguished. One is then interested in the evolution of these subpopulations as time passes. First, the *Wright-Fisher model* was introduced around 1930 for studying the transmission of genes for diploid populations, but for the sake of simplicity, we shall consider here a haploid version. Imagine thus a population with fixed size N and non-overlapping generations, such that for each individual i at generation $n+1$, the parent of i is an individual chosen uniformly at random amongst the N individuals at generation n, independently of the other

individuals. Because for a given individual j at generation n, the event that j is the parent of i has probability $1/N$, and is independent of the event that j is the parent of another individual i' at generation $n+1$, the number of children ν_j of j is given by the sum of N independent Bernoulli variables with parameter $1/N$, and has therefore the binomial distribution with parameter $(N, 1/N)$, that is

$$\mathbb{P}(\nu_j = k) = \binom{N}{k} N^{-k}(1-1/N)^{N-k} \quad \text{for } k = 0, 1, \ldots, N.$$

Note that the sequence of the numbers of children for individuals at the same generation, $(\nu_j : j = 1, \ldots, N)$, must fulfill the identity $\nu_1 + \cdots + \nu_N = N$, and thus does not consist of independent variables.

In this setting, one often supposes that individuals have types, or that their chromosomes carry certain alleles, which are transmitted to their descend; and one is interested in the propagation of those types or alleles. Assume for simplicity, that one gene has just two alleles, a and A, which are neutral for the reproduction, in the sense that the reproduction laws are the same for all individuals, no matter which allele they carry. If $P_N(n)$ denotes the number of individuals carrying allele a at generation n, then $(P_N(n) : n \geq 0)$ is a Markov chain with values in $\{0, 1, \ldots, N\}$. That is to say, roughly speaking, that conditionally on $P_N(n)$, the statistics of $P_N(n+1)$ are independent of the values of the chain P_N for the preceding generations. Specifically, conditionally on $P_N(n) = j$, as each individual at generation $n+1$ has probability j/N of having a parent that carries allele a, independently of the other individuals at the same generation, one has thus

$$\mathbb{P}(P_N(n+1) = k \mid P_N(n) = j) = \binom{N}{k} \left(\frac{j}{N}\right)^k \left(1 - \frac{j}{N}\right)^{N-k}$$

for $k = 0, 1, \ldots, N$. The states $j = 0$ and $j = N$ are called absorbing, as once the chain reaches state 0 (respectively, N), there are no more individuals carrying allele a (respectively, A), and allele a (respectively, A) has therefore disappeared forever in the population. One then says that fixation occurred for allele A (respectively, a). It is easy to check, that the probability that allele a eventually fixates is simply given by the proportion of individuals carrying allele a in the initial population.

In 1958, P. Moran introduced a somewhat even simpler model, now in continuous time, that can be described as follows. Each individual lives for an exponential time, say with fixed rate $r > 0$ (that is the probability that the lifetime is larger than t equals e^{-rt}), and then dies and is instantaneously replaced by a clone of an individual sampled uniformly at random in the remaining population. Using elementary properties of independent exponential variables, we can also reformulate the evolution as follows. Starting from a population with N individuals, after an exponential time with parameter $2rN$, we select a pair of individuals uniformly at random and then replace one of them by a copy of the other.

If one assumes, just as in the Wright-Fisher model, that the population has a fixed size N and that individuals carry a neutral allele a or A, and if $P_N(t)$ denotes the number of individuals carrying allele a at time t, then the process in

continuous time $(P_N(t))_{t\geq 0}$ is Markovian, and more precisely is a so-called *birth and death process*. Its dynamics are determined by the transition semigroup

$$T_{N,t}f(j) = \mathbb{E}_j(f(P_N(t))), \qquad j = 0, 1, \ldots, N,$$

where $f : \{0, 1, \ldots, N\} \to \mathbb{R}$ denotes a generic function on the state space and $\mathbb{E}_j$ the mathematical expectation given $P_N(0) = j$ (that is, there are j individuals carrying allele a at the initial time $t = 0$). However, the transition semigroup is not so simple to express explicitly, and one rather works with its time differential. Specifically, the infinitesimal generator of the process is defined as

$$\mathcal{A}_N f(j) = \lim_{t\to 0+} \frac{T_{N,t}f(j) - f(j)}{t},$$

and fulfills the Kolmogorov's forward and backward equations

$$\frac{\mathrm{d}T_{N,t}f(j)}{\mathrm{d}t} = T_{N,t}(\mathcal{A}_N f)(j) = \mathcal{A}_N(T_{N,t}f)(j).$$

This forms a systems of linear differential equations, and enables to recover the transition semigroup as the exponential of a matrix:

$$T_{N,t} = \exp(t\mathcal{A}_N).$$

Roughly speaking, the infinitesimal generator provides an analytic description for the evolution of a Markov process in terms of its rates of jumps. In the setting of the Moran process, one has for every $j = 1, \ldots, N-1$, that

$$\mathcal{A}_N f(j) = r\left(f(j+1) + f(j-1) - 2f(j)\right)\frac{j(N-j)}{N-1}. \tag{5}$$

Indeed, $j(N-j)/(N-1)$ is both the rate at which an individual is picked amongst the j individuals carrying the allele a and is replaced by a clone of one of the $N-j$ individuals carrying allele A, resulting in a decay of one unit for P_N, and the rate at which an individual is picked amongst the $N-j$ individuals carrying the allele A and is replaced by a clone of one of the j individuals carrying allele a, resulting in an increase of one unit for P_N. Finally, the states 0 and N are absorbing, therefore $\mathcal{A}_N f(0) = \mathcal{A}_N f(N) = 0$. It is easy to check, that, just as for the Wright-Fisher model, the probability that allele a eventually fixates when j individuals carry allele a in the initial population equals j/N.

Both the Wright-Fisher model and the Moran model can be considerably generalized by considering several alleles, or by changing the reproduction laws (which yields the Cannings model [6]), or by incorporating further phenomena such as mutations, selection, competition, etc. In this direction, we shall see in the forthcoming Section 4.2 that introducing a strong selection mechanism in Moran's model yields the deterministic logistic growth model of Section 2.2 in the limit when the size of the population goes to infinity.

2.4. Genealogies. In the early 1980's, J.F.C. Kingman [17, 18] formalized the idea of building the genealogical tree of a population by tracing backwards the ancestral lineages, and since then, his new approach has had a considerable impact on the way genealogies are viewed and studied. Roughly speaking, the object of interest is the process obtained by letting time run backward and observing the partition of the present population into the sub-populations, called blocks thereafter, which have the same ancestors at time $-t < 0$ as t increases.

To describe both its mechanism and its purposes as simply as possible, we start by considering the Moran population model which was presented in the preceding section. Recall that for a population with size N, a pair of individuals is picked uniformly at random after an exponential time with parameter rN, one of them is replaced by a copy of the other, and thereafter the process continues to evolve according to the same dynamics, independently of its past. We now assume that $r = (N-1)/2$, which induces no loss of generality since we can simply rescale time as a function of the size of the population. Loosely speaking, this means that one unit of time corresponds in average to $(N-1)/2$ generations. Observe that then $rN = \binom{N}{2}$ is the number of pairs of individuals at any given time. Now imagine that the present time is used as the origin of times, and that we follow the ancestral lineages of individuals backward in time. Specifically, we label the individuals in the present population by $\{1, \ldots, N\}$ uniformly at random, and for every $t \geq 0$, we obtain a partition $\Pi_N(t)$ of $\{1, \ldots, N\}$ into subpopulations which stem from the same ancestor at time $-t$; see Figure 1 below. The process $(\Pi_N(t))_{t\geq 0}$ is called the N-coalescent. Plainly, as t increases, these partitions get coarser, and $(\Pi_N(t))_{t\geq 0}$ evolves by coalescent events which are related to certain reproduction events in the past of the Moran process. More precisely, $\Pi_N(0)$ is the partition into singletons, and the first instant $t > 0$ at which $\Pi_N(t)$ does not only consists of singletons, corresponds to the last reproduction event before the present time. It has an exponential distribution with parameter $\binom{N}{2}$. At this instant, the ancestral lineages of two individuals chosen uniformly at random coalesce (i.e. merge).

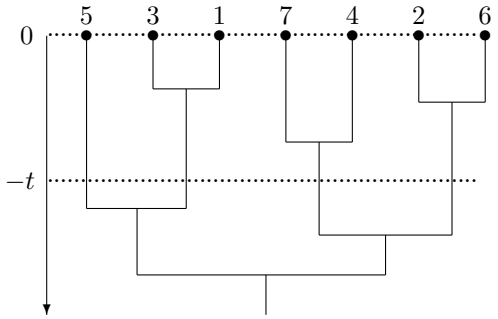

Figure 1: N-coalescent for $N = 7$

$\Pi_N(t) = (\{1, 3\}, \{2, 6\}, \{4, 7\}, \{5\})$

By iteration, one can check that the process $(\Pi_N(t))_{t\geq 0}$ is Markov and its rates of jumps are given as follows. When the state of the process is given by some partition of $\{1,\ldots,N\}$, say with blocks $B_1,\ldots,B_k$, where $k\geq 2$, then it stays at this state during an exponential time with parameter $\binom{k}{2}$, and then two blocks chosen uniformly at random amongst $B_1,\ldots,B_k$ (and independently of the waiting time) merge into a single block of the new partition. This corresponds to focussing on the k ancestors in the population at time $-t$ who generate the k subpopulations that form the entire present population, and letting time run further backward until a pair of their ancestral lineages meets. Equivalently, we may equip each pair of blocks $\{B_i,B_j\}$ with an independent standard (i.e., with mean 1) exponential variable, say $\mathbf{e}_{i,j}$, so that the minimum of those exponential variables over all possible pairs, $\mathbf{e}=\min_{1\leq i<j\leq k}\mathbf{e}_{i,j}$, has the exponential distribution with parameter $\binom{k}{2}=k(k-1)/2$. If we denote by $\{i_0,j_0\}$ the pair of indices for which the minimum is achieved, then the two blocks B_{i_0} and B_{j_0} are merged at the instant $\mathbf{e}$, that is, they are replaced by $B_{i_0}\cup B_{j_0}$. The process eventually reaches the trivial partition with a single block, which is the absorbing state. The time ζ_N to absorption is the age of the most recent common ancestor to all individuals in the present population. Note that it can be expressed as

$$\zeta_N=\sum_{k=2}^{N}\frac{2}{k(k-1)}\epsilon_k$$

where the ϵ_k are independent standard exponential variables, since then $\epsilon_k/\binom{k}{2}$ has the exponential law with parameter $\binom{k}{2}$. In particular there is the bound

$$\mathbb{E}(\zeta_N)\leq\sum_{k=2}^{\infty}\frac{2}{k(k-1)}=2.$$

Kingman pointed at the remarkable property of sampling consistency of N-coalescents. For every integer $N'<N$, if we write $\Pi'_N(t)$ for the restriction of the random partition $\Pi_N(t)$ to $\{1,\ldots,N'\}$, then the process $(\Pi'_N(t))_{t\geq 0}$ is an N'-coalescent. By taking projective limits, this enables the construction of a version for infinite populations. Namely, there is a process with values in the space of partitions of $\mathbb{N}=\{1,2,\ldots\}$, which we denote by $(\Pi(t))_{t\geq 0}$, such that for every integer N, the restriction of $\Pi(t)$ to $\{1,\ldots,N\}$ is an N-coalescent. One calls $(\Pi(t))_{t\geq 0}$ *Kingman's coalescent*.

The calculations for the expected absorption time in the paragraph above show, that even though $\Pi(0)$ is the partition into singletons and has thus infinitely many blocks, for every $t>0$, $\Pi(t)$ has almost surely finitely many blocks. One says that the coalescent comes down from infinity. This property has notably a key role in the problem of estimating the age of "mitochondrial Eve", i.e. the most-recent common female ancestor of all present-date humans; see Chang [7] and the discussion thereafter. This is an interesting instance of a concrete problem where a mathematical model is crucially needed in order to infer estimations of quantities to which it impossible to have a direct access.

During the last 15 years or so, various extensions of Kingman's coalescent have been considered. The so-called Λ-coalescents, which were introduced independently by Pitman and Sagitov, cover situations where multiple mergers may occur, whereas in Kingman's coalescent, each coalescent event involves exactly 2 blocks. In particular, this may be relevant to describe genealogies for species with extreme reproductive behavior (occasionally, a single parent may have a huge offspring of the same order as the whole population), such as certain marine species. In a different direction, the analysis of the genealogy of spatially structured populations motivated the introduction of spatial versions of coalescent processes. Overall, coalescent theory has been mainly developed in the neutral case (absence of selection), with the notable exception of recent developments initiated by Brunet and Derrida. These authors considered population of branching type in which only the best fitted children are selected for the next generation. This changes considerably the genealogy. Roughly speaking it is no longer described by Kingman's coalescent as one might have expected, but rather by the Bolthausen-Sznitman coalescent, an important special case of a Λ-coalescent which has first appeared in connection with spin-glass models in Statistical Physics. See Berestycki, Berestycki and Schweinsberg [3] and references therein for much more on this topic.

Kingman's analysis of genealogies relies crucially on the hypothesis that populations have a constant size, which is of course not very realistic when applied to real life models (think for instant of the growth of the human population in history), and a difficult question is to develop useful models which cover the case when the size of the population is time-varying. Further Kingman's coalescent only applies to haploid populations, and genealogies for in the diploid situation is far more complex. In particular, key biological phenomena such as recombination have to be taken into account, see notably the work of Baake [2] and collaborators in this area.

3. When should one expect deterministic averages ?

Roughly speaking, the pertinence of deterministic models is often justified by the assertion that, due to the Law of Large Numbers, a quantity evaluated for each individual and averaged over a large population shall be nearly deterministic. A possible objection when applying bluntly this simple rule of thumb, is that the implicit hypothesis that the population can be modeled as a family of independent individuals with the same distribution may be unrealistic in practice, putting in doubt the legitimacy of the conclusions. Actually, much less restrictive requirements than independence are needed for the validity of the conclusions of Law of Large Numbers; and the issue of whether an average over a large population is essentially deterministic or random, can be clarified by a simple covariance analysis. Recall that, roughly speaking, a square integrable random variable is close to a constant (which then coincides with its mathematical expectation) if and only if its variance is small. We shall now present this covariance analysis tailored for our purposes.

Consider for every integer $n \geq 1$, a random population of size $P(n) \geq 1$, say $\mathcal{P}_n = \{x_i : i = 1, \ldots, P(n)\}$, where the labelling of the individuals is irrelevant for our purposes, and let $f_n : \mathcal{P}_n \to \mathbb{R}$ denote some real-valued function evaluated for each individual of the population. We may think of $f_n(x)$ as some real trait, that is, a real number measuring some characteristic of the individual x; for instance, $f_n(x)$ may denote the adult size of x, or the weight of x at birth, etc. We are interested in the average of f_n over the population,

$$\bar{f}_n = \frac{1}{P(n)} \sum_{i=1}^{P(n)} f_n(x_i),$$

which is a random quantity since the population is random. Roughly speaking, the next result provides an elementary answer – actually, almost a tautology – to the question of whether $\bar{f}_n$ is nearly deterministic when n is large. Recall that when ξ and ξ' are two (real) random variables, their covariance is denoted by

$$\mathrm{Cov}(\xi, \xi') = \mathbb{E}(\xi\xi') - \mathbb{E}(\xi)\mathbb{E}(\xi'),$$

whenever this quantity is well-defined.

Lemma 3.1. *For each $n \geq 1$, sample two individuals X_n and X'_n in the population $\mathcal{P}_n$ uniformly at random with replacement, and set $\xi_n = f_n(X_n)$ and $\xi'_n = f_n(X'_n)$. Assume further that*

$$\sup_{n \geq 1} \mathbb{E}(\xi_n^2) < \infty.$$

Then for every $\ell \in \mathbb{R}$, the following two assertions are equivalent:

(i) $\lim_{n\to\infty} \bar{f}_n = \ell$ *in* $L^2(\mathbb{P})$.

(ii) $\lim_{n\to\infty} \mathbb{E}(\xi_n) = \ell$ *and* $\lim_{n\to\infty} \mathrm{Cov}(\xi_n, \xi'_n) = 0$.

Proof. Because ξ_n and ξ'_n are two traits sampled uniformly at random with replacement, we have

$$\mathbb{E}\left(g(\xi_n, \xi'_n)\right) = \mathbb{E}\left(\frac{1}{P(n)^2} \sum_{i=1}^{P(n)} \sum_{j=1}^{P(n)} g(f_n(x_i), f_n(x_j))\right)$$

for every measurable function $g : \mathbb{R}^2 \to \mathbb{R}_+$. We deduce that

$$\mathbb{E}(\bar{f}_n) = \mathbb{E}(\xi_n) = \mathbb{E}(\xi'_n) \quad \text{and} \quad \mathbb{E}(\bar{f}_n^2) = \mathbb{E}(\xi_n \xi'_n)\,.$$

As a consequence, the variance of $\bar{f}_n$ is simply given by

$$\mathrm{Var}(\bar{f}_n) = \mathrm{Cov}(\xi_n, \xi'_n),$$

and our claim then follows easily. □

Remark 3.2. This elementary second moment analysis lies at the heart of the notion of propagation of chaos, developed by Marc Kac [16], which is also relevant for asymptotic study of large populations.

Roughly speaking, Lemma 3.1 shows that for the empirical average of a real trait over a large random population to be close to a constant, the correlation between the traits of two randomly sampled individuals has to be small. Let us now conclude this section by illustrating Lemma 3.1 with the following simple example. Consider a population model with non-overlapping generations, and some real trait t for that population. Imagine now that traits are transmitted from parents to children up-to an independent perturbation η which has a fixed distribution. That is, if y is a child of x, then $t(y) = t(x) + \eta_y$, where η_y is a random variable distributed as η. Assume also that the η_y are further independent for different individuals, and that η is centered [$\mathbb{E}(\eta) = 0$], and has finite variance $\sigma^2 = \mathbb{E}(\eta^2) < \infty$.

We write $\mathcal{P}_n$ for the population at the n-th generation, which has size $\mathrm{Card}(\mathcal{P}_n) = P(n)$. For every $r \in \mathbb{R}$, we are interested in the proportion of individuals at the n-th generation having a trait smaller than $r\sigma^2\sqrt{n}$, that is

$$\bar{f}_n(r) = \frac{\mathrm{Card}\{x \in \mathcal{P}_n : t(x) \leq r\sigma^2\sqrt{n}\}}{\mathrm{Card}(\mathcal{P}_n)}.$$

Let us assume that if X_n and X'_n denote two individuals picked uniformly at random in the population $\mathcal{P}_n$, and if $\gamma_n \leq n$ denotes the generation of the most recent common ancestor of X_n and X'_n, then

$$\lim_{n\to\infty} \mathbb{E}\left(\frac{\gamma_n}{\sqrt{n}}\right) = 0. \tag{6}$$

This is a very mild assumption which is fulfilled by many natural models; note that it also implies that $\lim_{n\to\infty} P(n) = \infty$, as otherwise the probability that $X_n = X'_n$ would not tend to 0 as $n \to \infty$ and (6) would fail.

We assert that then, the repartition of traits in the population is asymptotically Gaussian, viz.

$$\lim_{n\to\infty} \bar{f}_n(r) = \frac{1}{\sqrt{2\pi}} \int_\infty^r \exp(-x^2/2)\mathrm{d}x. \tag{7}$$

Let us now briefly show how this follows from Lemma 3.1. Note that the traits of X_n and X'_n can be expressed in terms of the trait, say τ_n, of their most recent common ancestor, in the form

$$t(X_n) = \tau_n + \beta_n, \quad t(X_n) = \tau_n + \beta'_n$$

with

$$\beta_n = \eta_{\gamma_n+1} + \cdots + \eta_n, \quad \beta'_n = \eta'_{\gamma_n+1} + \cdots + \eta'_n,$$

where γ_n stands for the generation of the most recent common ancestor of the two individuals, and $\eta_1, \ldots, \eta_n$ and $\eta'_1, \ldots, \eta'_n$ are independent copies of η.

On the one hand, the trait of the most recent common ancestor of X_n and X'_n has the same law as $t_0 + \eta_1 + \cdots + \eta_{\gamma_n}$, where t_0 denotes the trait of the ancestor of the entire population and may be assumed deterministic for the sake of simplicity, and as a consequence

$$\mathbb{E}(\tau_n) = t_0 \quad \text{and} \quad \mathrm{Var}(\tau_n) = \mathbb{E}(\eta) \times \mathbb{E}(\gamma_n).$$

We then see from (6) that $\mathbb{E}(\tau_n^2/n)$ converges to 0 as $n \to \infty$.

On the other hand, an easy application of the Central Limit Theorem combined with the assumption (6) shows that $\beta_n/\sigma\sqrt{n}$ and $\beta'_n/\sigma\sqrt{n}$ converge in distribution as n tend to ∞ towards a pair of independent standard Gaussian variables. We conclude from above that the same holds for the rescaled traits $t(X_n)/\sigma\sqrt{n}$ and $t(X'_n)/\sigma\sqrt{n}$. In particular, if f_n denotes the indicator function of the interval $(-\infty, r\sigma^2\sqrt{n}]$, then, as $n \to \infty$

$$\mathbb{E}(f_n(X_n)) = \mathbb{P}(t(X_n) \leq r\sigma^2\sqrt{n}) \longrightarrow \frac{1}{\sqrt{2\pi}} \int_\infty^r \exp(-x^2/2)\mathrm{d}x,$$

and $\mathrm{Cov}(f_n(X_n), f_n(X'_n)) \to 0$. Since $\bar{f}_n(r)$ can be expressed in the form

$$\bar{f}_n(r) = P(n)^{-1} \sum_{x \in \mathcal{P}_n} f_n((x)),$$

the conclusion (7) follows from Lemma 3.1.

Remark 3.3. The example presented above can be considerably generalized in the setting of branching random walk, where (7) can be viewed as an elementary version of much deeper limit theorems due mainly to Biggins; see e.g. [4].

4. Sources of random averages in large populations

We shall now discuss situations where the empirical average of a trait remains intrinsically random, even for large populations. Keeping in mind the elementary Lemma 3.1, we shall describe circumstances where, even though the size of the population tends to infinity, the traits of two randomly sampled individuals remain correlated. We stress that these simple examples aim at illustrating typical phenomena rather than at describing realistic population models.

4.1. The effect of small ancestral populations. The first example illustrates the classical effect of small ancestral populations; the informal idea is that random events should have in general a larger stochastic impact on small populations than on large ones, due to averaging effects for large populations, and that this randomness can then propagates to the next generations.

We consider a population modeled by a simple *Pólya urn*. That is, imagine that at initial time, we have two individuals, one carrying allele A and the other one carrying allele a, which are viewed as two balls, one labelled A and one labelled

a, placed in a urn with infinite capacity. At each step, we pick a ball in the urn uniformly at random, note its label, and then replace it into the urn together with an additional ball having the same label. The interpretation in terms of population dynamics is that at each step, an individual is picked uniformly at random in the current population, and gives birth to a clone, that is that types are transmitted without change from parents to children.

At first sight, it may seem awkward to model a population as an urn[2], however, this is precisely what happens in the following situation. A *Yule process* $(Y(t))_{t\geq 0}$ is one of the simplest random population process in continuous time. It is a pure birth process, which describes the evolution of the size of a population in which each individual gives birth to a clone at rate 1, independently of the other individuals. Imagine now that a two-type Yule process starts from an individual with allele A and an individual with allele a (alleles are implicitly assumed to be neutral for the reproduction). When the population reaches size n, it remains unchanged for an exponentially distributed time with parameter n, and then an individual is selected uniformly at random in the current population and duplicated. So if we observe a two-type Yule process at the sequence of times when individuals duplicate, we get precisely the dynamics of a Pólya urn.

We are interested in the repartition of alleles in the population, and write $R(n)$ for the proportion of individuals carrying allele A when the total population reaches size n, that is after $n-2$ steps. It is well-known and easy to prove that as $n \to \infty$, $R(n)$ converges to a random variable $R(\infty)$ which has the uniform distribution on $[0,1]$. For instance, one can use the easy fact that a Yule process started from a single individual grows exponentially fast as a function of time; more precisely $\lim_{t\to\infty} \mathrm{e}^{-t} Y(t) = W$ where W is a random variable with the standard exponential distribution, viz. $\mathbb{P}(W > x) = \mathrm{e}^{-x}$ for all $x \geq 0$. In our situation, we have two independent Yule processes, say $(Y_A(t))_{t\geq 0}$ and $(Y_a(t))_{t\geq 0}$ in the obvious notation, and the proportion of individuals carrying allele A thus fulfills

$$\begin{aligned}\lim_{t\to\infty} \frac{Y_A(t)}{Y_A(t)+Y_a(t)} &= \lim_{t\to\infty} \frac{\mathrm{e}^{-t}Y_A(t)}{\mathrm{e}^{-t}Y_A(t)+\mathrm{e}^{-t}Y_a(t)} \\ &= \frac{W_A}{W_A+W_a}\end{aligned}$$

where W_A and W_a are two independent standard exponential random variables, so that the ratio $W_A/(W_A+W_a)$ has the uniform distribution on $[0,1]$.

So even though the size of the population tends to infinity, the repartition of alleles remains intrinsically random. Actually, the randomness of the repartition is essentially built at the early stages of the process when the population is still

[2]Actually, urn models can be quite useful for population modeling. A further important example is Hoppe's urn which can be depicted as follows. We start with an urn with two balls, one with a label, say A, and one unlabeled ball and proceed as in Pólya's model, except that when the unlabelled ball is picked, then it is replaced in the urn together with a ball having a new label which was not present in the urn before. This can be used as a simple model for neutral mutations, and explain the celebrated Ewens sampling formula (which is further discussed in the next section) and its connections to Poisson-Dirichlet random partition for the repartition of alleles.

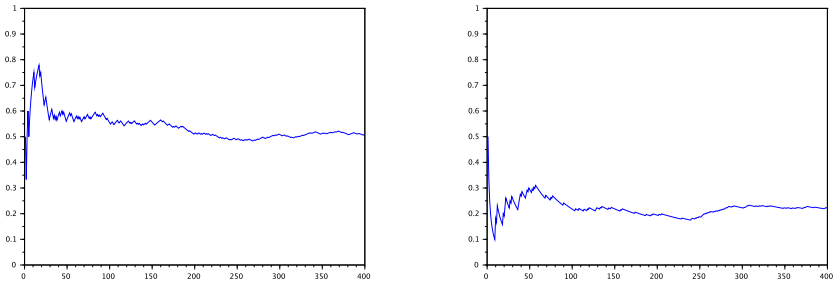

Figure 2 :
Two simulations of a Polya urn illustrating convergence to a random limit.

small. This phenomenon is of course much more general than discussed in this elementary example.

Last, it may be also interesting to compare with Lemma 3.1. What goes wrong when we try to apply Lemma 3.1, is that when we sample at random two individuals in the population when it reaches size n, and for $i = 1, 2$, write $\xi_i(n) = 1$ if the i-th sampled individual carries allele A and $\xi_i(n) = 0$ otherwise, then when n is large

$$\mathbb{P}(\xi_1(n) = \xi_2(n) = 1) = \mathbb{P}(\xi_1(n) = \xi_2(n) = 0) \sim 1/3$$

and

$$\mathbb{P}(\xi_1(n) = 1, \xi_2(n) = 0) = \mathbb{P}(\xi_1(n) = 0, \xi_2(n) = 1) \sim 1/6.$$

That is, the correlation between two randomly sampled individuals persists even when the population is large.

4.2. Role of the regimes. Lemma 3.1 shows that non-deterministic averages over large populations should be expected whenever two randomly picked individuals remain asymptotically correlated when the size of the population tends to infinity. Intuitively, the correlation between individuals stems from their common history, whereas the genetic material transmitted from an ancestor tends to fade away generation after generation, for instance due to new mutations. Roughly speaking, we may expect that the older the common ancestors are and the faster traits evolve, the smaller the correlation between individuals is.

We shall illustrate this idea with two fairly different examples which are both based on Moran's model for a population with large size. In the first example, neutral mutations are superposed to the model, whereas in the second example, we shall be interested in the invasion of an advantageous allele.

4.2.1. Rare mutations. So let us consider Moran's population model, for a population with a large size N. Imagine that some gene has different alleles, which are all neutral for the reproduction. Let $\mathcal{A} = \{a_1, \ldots, a_j\}$ denote the set of alleles, and for the purpose of modeling, simply assume that a mutation process

is superposed (independently) to the evolution of the Moran process. Specifically, consider a transition kernel q on $\mathcal{A}$, that is, for each $a \in \mathcal{A}$, $q(a, \cdot)$ is a probability measure on $\mathcal{A}$ with $q(a, a) = 0$ for all $a \in \mathcal{A}$. So for two different alleles, a and a', $q(a, a')$ is the probability that, provided that a mutation occurs while an individual carrying allele a duplicates, this produces an individual with allele a'. In other words, during a reproduction event when an individual, say x, is replaced by a copy y' of an individual, say y, if y carries allele $a \in \mathcal{A}$, then y' also carries allele a with probability $1 - p(N)$, and carries a different allele $a' \in \mathcal{A}$ with probability $p(N)q(a, a')$, where $p(N)$ is a small parameter which depends only on N and represents the rate of mutations.

We are interested in the repartition of alleles in the current population, and more precisely, in the proportion $\bar{f}_N$ of individuals carrying a given allele a. That is, for every individual x, we write $f_N(x) = 1$ if x carries allele a and $f_N(x) = 0$ otherwise, and $\bar{f}_N = \frac{1}{N} \sum f_N(x)$, where the sum is taken over the N individuals of the present population. For this, we pick two individuals X_N and X'_N uniformly at random in the present population, and write $\xi_N = f_N(X_N)$ and $\xi'_N = f_N(X'_N)$. With no loss of generality, we assume that the lifetime of each individual in the Moran model has the exponential distribution with parameter $r = (N - 1)/2$, so that the genealogy is described by an N-coalescent; see Section 2.4. In particular, the age of the most recent common ancestor, say Y_N, of X_N and X'_N has a standard exponential distribution, say $\mathbf{e}$, and therefore there are about $\mathbf{e} \times N/2$ individuals along the ancestral lineage from X_N (respectively, X'_N) to Y_N.

We now see that the asymptotic correlation between ξ_N and ξ'_N depends on whether the rate of mutations $p(N)$ is much smaller than $1/N$, or much larger that $1/N$, or of order $1/N$. Specifically:

- if $p(N) \ll 1/N$, then the probability that a mutation occurred on the ancestral lineages from X_N and from X'_N to their common ancestor is small when N is large, and therefore $\xi_N = \xi'_N$ with high probability. In this situation, the population at the present time is essentially monomorphic, that is nearly all individuals carry the same allele (which is the allele carried by their most recent common ancestor Y_N). The precise value of that allele is random, its distribution being given by the invariant law π of the Markov chain on $\mathcal{A}$ with transition kernel q. Then $\bar{f}_N$ is statistically close to a Bernoulli random variable β, with $\mathbb{P}(\beta = 1) = \pi(a)$ and $\mathbb{P}(\beta = 0) = 1 - \pi(a)$.

- if $p(N) \gg 1/N$, then with high probability, a large number of random mutations have occurred on the ancestral lineages from X_N and from X'_N to their common ancestor, and ξ_N and ξ'_N are asymptotically uncorrelated. The proportion of individuals carrying allele a is nearly deterministic, with $\bar{f}_N \sim \pi(a)$.

- In the critical case when $p(N) \sim \theta/N$ for some constant $\theta > 0$, then one can prove that the pair (ξ_N, ξ'_N) converges in distribution to a pair of random variables which are neither identical, nor independent. It follows that $\bar{f}_N$ converges in distribution to a random variable, which has not a Bernoulli law.

We further mention that in the same circle of ideas, but now in the setting of the infinite allele model of Kimura and Crow, one of the most remarkable applications of Kingman's coalescent is an illuminating explanation of the celebrated sampling formula due to Warren Ewens. Roughly speaking, imagine that each individual in Moran's model may mutate at (critical) rate θ/N, and then bears a new (neutral) allele which was never observed before. Each allelic population eventually becomes extinct, and one can check that the partition of the population at time t into subfamilies sharing the same allele converges in distribution to a certain statistical equilibrium as t goes to infinity. Ewens obtained an explicit formula for the equilibrium distribution, which arises not only in the context of population models, but much more generally in a variety of combinatorial structures. Kingman has shown that Ewens sampling formula can be recovered by superposing random marks on the branches of the genealogical tree describing an N-coalescent, and then analyzing the clusters of individuals which are connected by branches having no mark.

4.2.2. Strong or weak selection. Here, just as in Section 2.4, we work with Moran's model for a population of size $N \gg 1$, and suppose for simplicity that individuals carry either allele a or allele A. Whereas we assumed in Section 2.4 that these two alleles are neutral for reproduction, we suppose here that allele a is advantageous with selection coefficient $s \in (0, 1)$, which may depend on the size of the population. This means that for each reproduction event, when a pair of individuals of different types (a, A) is picked, then it is replaced by a pair (a, a) with probability $(1 + s)/2$, and by a pair (A, A) with probability $(1 - s)/2$ (so the neutral case would correspond to setting the selection coefficient $s = 0$). Writing $P_N^s(t)$ for the number of individuals carrying allele a at time t, just as in the case without selection discussed in Section 2.4, one easily checks that the process $(P_N^s(t) : t \geq 0)$ is Markovian, now with infinitesimal generator

$$\mathcal{A}_N^s f(j) = r\left((1+s)f(j+1) + (1-s)f(j-1) - 2f(j)\right)\frac{j(N-j)}{N-1}$$

Note that this can also be expressed in the form

$$\mathcal{A}_N^s f(j) = \mathcal{A}_N f(j) + rs\left(f(j+1) - f(j-1)\right)\frac{j(N-j)}{N-1},$$

where $\mathcal{A}_N f(j) = \mathcal{A}_N^0 f(j)$ is given in (5).

The upshot of computing explicitly infinitesimal generators is that this enables the use of powerful techniques known as *diffusion-approximation* to establish limit theorems (in a strong or in a weak sense) for Markov processes from the asymptotic behaviour of their infinitesimal generators. We refer to Ethier and Kurtz [12] or Etheridge [11] for a detailed account of this concept, which is especially useful for the study of large population models. In the present setting, this enables to prove that as the size N of the total population goes to infinity, in the regime of called *strong selection* when the parameter s does not depend on N, the ratio process

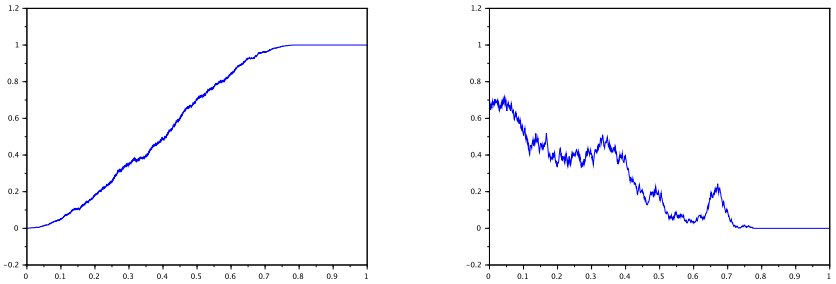

Figure 3 : Ratio process for a Moran process with selection:
strong selection (left) and weak selection (right)

$R_N^s(t) = P_N^s(t)/N$ converges, $\lim_{N\to\infty} R_N^s(t) = R^s(t)$, where $R^s : [0,\infty) \to [0,1]$ is a deterministic function which solves the logistic growth equation

$$\frac{\mathrm{d}R^s(t)}{\mathrm{d}t} = rsR^s(t)(1 - R^s(t)).$$

On the other hand, when the selection parameter s depends on N such that $s \sim c/N$ with $c > 0$ constant, which is known as a regime of *weak selection*, the ratio process properly time-rescaled, $R_N^s(tN) = P_N^s(tN)/N$, converges, now merely in distribution, to a Wright-Fisher diffusion process with selection; see for instance Etheridge [11]. Figure 3 above illustrates these two cases; note that in the strong selection case, the ratio process is close to the curve of a deterministic logistic growth function, and that in the weak selection case, the favorable allele may nonetheless disappear (which would be much more unlikely in the strong selection regime).

5. Conclusions

Modeling biological phenomena has been an important source of studies in applied mathematics for many years. The main issues are to come up with models which are both realistic enough, and thus capture the essence of the phenomena of interest, and nonetheless simple enough, and thus remain tractable for statistical analysis. Finding the right trade-off between realism and simplicity is often a difficult problem, and depends o the nature of the questions to be answered. The same applies when deciding whether to opt for a deterministic or a stochastic model.

In this text, we have merely discussed a few important population models, even though of course, mathematical modeling in Biology, and more generally in Life Sciences, concerns a great variety of aspects aside from population dynamics. Further, many more population models can be found in the literature, e.g. for propagation of infectious diseases, evolutionary invasion or adaptive dynamics, parasites, etc. For individual-based models, traits of individuals are correlated

through common ancestors, and in general cannot be viewed as independent variables. This may play an important role for determining whether average features over large populations are nearly deterministic or intrinsically random quantities. We have illustrated the possible impact of the small size of ancestral populations (which is often referred to as bottleneck), of the rates of mutations, and of the strength of selection. Other phenomena may of course also have a crucial role, for instance rare extreme events (think of the impact of the collision of a large asteroid with Earth).

Without any doubt, Life Sciences will continue to motivate frontline researches in mathematical modeling for many more years. Let us merely point at one challenging problem amongst others in this area. Most models of evolution focus on the haploid case and on already-existing genes, and describe how natural selection affects their frequencies depending on their relative fitnesses. The problem of modeling sexual reproduction, including genetic recombination and generation of new alleles that may appear through mutations, and of describing their dynamics and their genealogies, has only been partly addressed so far and should be the subject of deeper investigations in the future.

6. References

[1] Allman, Elizabeth S. and Rhodes, John A. *Mathematical models in biology: an introduction*, Cambridge University Press, Cambridge, 2004.

[2] Baake, Ellen. Deterministic and stochastic aspects of single-crossover recombination. In: *Proceedings of the International Congress of Mathematicians* **IV**, pp. 3037–3053, Hindustan Book Agency, New Delhi, 2010.

[3] Julien Berestycki and Nathanaël Berestycki and Jason Schweinsberg. The genealogy of branching Brownian motion with absorption. Ann. Probab. **41** (2013), 527–618.

[4] Biggins, J. D. Uniform convergence of martingales in the branching random walk, *Ann. Probab.* **20** (1992), 137–151.

[5] Bürger, R. *The mathematical theory of selection, recombination, and mutation*, Wiley Series in Mathematical and Computational Biology, John Wiley & Sons, Ltd., Chichester, 2000.

[6] Cannings, C. The latent roots of certain Markov chains arising in genetics: a new approach. I. Haploid models. *Advances in Appl. Probability* **6** (1974), 260–290.

[7] Chang, Joseph T. Recent common ancestors of all present-day individuals (with discussion and reply by the author), *Adv. in Appl. Probab.* **31** (1999), 1002–1038.

[8] Dawson, Donald A. *Introductory lectures on stochastic population systems.* (2010) Available at : `http://www.researchgate.net/profile/Donald_Dawson`

[9] Durrett, Rick. *Probability models for DNA sequence evolution*, Probability and its Applications, Springer-Verlag, New York, 2002.

[10] Edelstein-Keshet, Leah. *Mathematical models in biology*, SIAM Classics in Applied Mathematics, 2005.

[11] Etheridge, Alison. Some mathematical models from population genetics. In *École d'été de Probabilités de Saint-Flour*, Lecture Notes in Mathematics vol. 2012. Springer, 2011.

[12] Ethier, Stewart N. and Kurtz, Thomas G. *Markov processes: characterization and convergence*. Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, Inc., New York, 1986.

[13] Haccou, Patsy and Jagers, Peter and Vatutin, Vladimir A. *Branching processes: variation, growth, and extinction of populations*. Cambridge Studies in Adaptive Dynamics, Cambridge University Press, Cambridge, 2007.

[14] Hein, Jotun and Schierup, Mikkel H. and Wiuf, Carsten, *Gene genealogies, variation and evolution, a primer in coalescent theory*. Oxford University Press, Oxford, 2005.

[15] Hofbauer, Josef and Sigmund, Karl. *The theory of evolution and dynamical systems*. London Mathematical Society Student Texts, Cambridge University Press, Cambridge, 1988.

[16] Kac, M. Foundations of kinetic theory, in: *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. III*, pp. 171–197, University of California Press, Berkeley and Los Angeles, 1956.

[17] Kingman, J. F. C. On the genealogy of large populations, *J. Appl. Probab.* **19A** (1982), 27–43.

[18] Kingman, J. F. C. The coalescent, *Stochastic Process. Appl.* **13** (1982), 235–248.

[19] Lambert, Amaury. The branching process with logistic growth, *Ann. Appl. Probab.* **15** (2005), 1506–1535.

[20] Turchin, Peter. *Complex population dynamics: a theoretical/empirical synthesis*, Monographs in Population Biology **35**, Princeton University Press, Princeton, NJ, 2003.

Jean Bertoin, Institut für Mathematik, Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

E-mail: jean.bertoin@math.uzh.ch