A VARIANT OF THE LOOK-DOWN MODEL WITH SELECTION

Boubacar Bah\(^1\)

Abstract. The goal of this paper is to study a new version of the look-down model with selection, where the population size \(N\) is finite and fixed. As in [1], we show (see Theorem 1.2) convergence in probability, locally uniformly in \(t\), as the population size \(N\) tends to infinity, towards the Wright-Fisher diffusion with selection.

Résumé. Dans ce papier on présente une variante du modèle étudié dans l'article [1]. On étudie le modèle du look-down avec sélection dans le cas d'une population de taille finie et fixée. Nous montrons que la proportion de l'un des types converge en probabilité, quand la taille \(N\) de la population tend vers l'infini, vers la diffusion de Wright-Fisher avec sélection.

INTRODUCTION

The look-down model and the modified look-down model have been introduced by Donnelly and Kurtz (see [3] and [4]) to give the genealogical process associated to a diffusion model of population evolution. The idea is to distribute the population on sites indexed by \(i \geq 1\), with exactly one individual per site. This powerful representation is now currently used.

We briefly recall the definition of the modified look-down model, without taking into account any spatial motion for the individuals. Consider an infinite size population evolving forward in time. For any \(1 \leq i < j\), at rate one, the individual sitting on site \(i\) gives birth to an individual sitting on site \(j\), and all individuals sitting on a site greater than or equal to \(j\) are shifted to the right (see Figure 1), that is to say each of those individual will move to the site which is at his right. These reproduction events involving levels \(i\) and \(j\) are called look-down events.

The two main differences between this model and the Moran model (see Figure 2), are that first, the arrows representing births are always pointing to the right, that is to say an individual sitting on site \(i\) can only give birth to an individual on a site \(j\) with \(j > i\) (The asymmetry which results from this choice is compensated by exchangeability, which is an important property of the look-down model). This ensures that the infinite model is well defined. Indeed, if we restrict ourselves to the first \(N\) individuals, the evolution is determined by finitely many arrows. This would not be the case with the standard Moran model, which could not be described in the case \(N = \infty\). In the Moran model with infinitely many individuals, there would be infinitely many arrows towards any individual \(i\), in any time interval of positive length.

The second difference is that the individual who was sitting on the site where the offspring took place does not disappear, but instead is moved to the right, just as all the individuals which are on a site to his right.

\(^*\) The author wish to thank an anonymous Referee, whose excellent and very detailed report permitted me to correct some imprecisions in an earlier version of this paper.

\(^1\) CMI, LATP-UMR 6632, Université de Provence, 39 rue F. Joliot Curie, Marseille cedex 13, FRANCE; bbah12@yahoo.fr

\(\copyright\) EDP Sciences, SMAI 2014
In [5], Donnelly and Kurtz added selection for a finite number of type individuals to their model, which involved additional births or possible deaths.

In our model, when a death occurs, the individual who dies is removed from the population, and each of the individuals sitting on a site to the right of his site are shifted to the left.

The aim of this paper is to present a variant of the model studied in [1]. We consider a population of fixed size $N$. We assume that two types of individuals coexist in the population: individuals with the wild-type allele $b$ and the individuals with the advantageous allele $B$. This selective advantageous is modeled by a death rate $\alpha$ for the type $b$ individuals, while the type $B$ individuals are not subject to that specific death mechanism.

We first recall the model from [1] with $N = \infty$ called $(L^\infty)$, and then we will describe the variant which will be the subject of the present paper.

**Look-down model with selection** (called $(L^\infty)$) : We consider a population of infinite size. We will consider the proportion of $b$ individuals. Hence type $b$ individuals are coded by $1$, and $B$ by $0$. We assume that individuals are placed at time $0$ on levels $1, 2, ..., $ each one being, independently from the others, $1$ with probability $x$, $0$ with probability $1 - x$, for some $0 < x < 1$. For any $t \geq 0$, $i \geq 1$, let $\eta_t(i)$ denote the type of the individual sitting on site $i$ at time $t$. Clearly $\eta_t(i) \in \{0, 1\}$. The evolution of the population is governed by the two following mechanisms.

1. **Births**. For any $1 \leq i < j < \infty$, arrows are placed from $i$ to $j$ according to a rate one Poisson process, independently of the other pairs $i' < j'$. Suppose there is an arrow from $i$ to $j$ at time $t$. Then a descendent (of the same type) of the individual sitting on level $i$ at time $t^-$ occupies the level $j$ at time $t$, while for any $k \geq j$, the individual occupying the level $k$ at time $t^-$ is shifted to level $k + 1$ at time $t$. In other words, $\eta_t(k) = \eta_{t^-}(k)$ for $k < j$, $\eta_t(j) = \eta_{t^-}(i)$, $\eta_t(k) = \eta_{t^-}(k - 1)$ for $k > j$. 

\[\text{Figure 1. The graphical representation of a modified look-down construction. At times } s \text{ the individual sitting on level 2 gives birth to an individual sitting on set 5. For all } k \geq 5, \text{ the individual at level } k \text{ is instantaneously shifted to level } k + 1.\]
Figure 2. The graphical representation of the Moran model of size $N = 9$. Arrows between lines indicate resampling event. By resampling the genealogical relationships between individuals change. At time $s$ the individual at the level 5 gives birth to an individual who replaces the individual sitting on level 1.

(2) Deaths. Any type 1 individual dies at rate $\alpha$, his vacant level being occupied by his right neighbor, who himself is replaced by his right neighbor, etc. In other words, independently of the above arrows, crosses are placed on each level according to a rate $\alpha$ Poisson process, independently of the other levels. Suppose there is a cross at level $i$ at time $t$. If $\eta_t-(i) = 0$, nothing happens. If $\eta_t-(i) = 1$, then $\eta_t(k) = \eta_t-(k)$ for $k < i$, and $\eta_t(k) = \eta_t-(k + 1)$ for $k \geq i$.

This model has been formulated by Anton Wakolbinger in an oral presentation [6]. Note that with those deaths, the infinite model is no longer immediately well defined, since for each $N \geq 1$, the evolution of the individuals sitting on the first $N$ sites depend, in case of death, on the individual sitting on the following sites. Contrary to models defined in [3], [4], [5], the process $Y_t^N = N^{-1}(\eta_t(1), \ldots, \eta_t(N))$ is not a Markov process but it is approximately Markovian (see [1] for more details). We can modify the model ($L^\infty$) as follows.

$N$-Look-down model with selection (called ($N-L^\infty$)) : For each $N$, consider the process $\{\eta_t^N(i), i \geq 1, t \geq 0\}$, obtained by applying only the arrows between 1 \leq i \leq j \leq N, and the crosses on levels 1 to $N$. In other words, all arrows pointing to levels above $N$, and all crosses on levels above $N$ have been erased. We then have a finite number of arrows and crosses on any finite time interval, and $\{\eta_t^N(i), i \geq 1, t \geq 0\}$ is constructed in an obvious way, by implementing the effect of the arrows and crosses, in the order in which they are met.

We are going to present now the variant of the look-down construction with selection called ($L^N$), where the size $N \in \mathbb{N} = \{1, 2, \ldots, \}$ of the population is finite and fixed. As in the model ($L^\infty$), we denote by $\zeta_t^N(i) \in \{0, 1\}$ the type of the individual sitting on site $i$ at time $t$. The following events occurs:
• **Births**: For each $1 \leq i < j \leq N$, we put arrows from $i$ to $j$ at rate 1. At such an arrow, the individual at level $i$ put a child at level $j$. Individuals previously at levels $j, \ldots, N - 1$ are shifted one level up; individuals at site $N$ dies.

• **Deaths**: Any type 1 individual dies at rate $\alpha$, his vacant level being occupied by his right neighbor, who himself is replaced by his right neighbor, etc. We complete the population by an individual of type 1 at level $N$ with probability $a_N(t)$, type 0 with probability $1 - a_N(t)$, where $a_N(t)$ is the proportion of type $b$ individuals before the death event $t$, see below.

We can represent this evolution graphically, drawing the set $[1, N] \times \mathbb{R}$. A point $(i, t) \in [1, N] \times \mathbb{R}$ stands for the individual that occupies level $i$ at time $t$. Times goes from the top to the bottom. The jump times of the Poisson process associated to the pair $(i, j)$, $1 \leq i < j \leq N$ are represented by arrows from $i$ to $j$. We refer the reader to Figure 3 for a pictural presentation of our model.

We want to choose the type of the $N$ individuals at time 0 in an exchangeable way, with the constraint that the proportion of $b$ individuals is given. One possibility is to draw without replacement $N$ balls from an urn where we have put $k$ red balls (which represent the type $b$ individuals) and $N - k$ black balls (which represent the type $B$ individuals). At each draw, each of the balls which remain in the urn has same probability of being chosen.

It follows from the above considerations and Proposition 3 in [1] that at each time $t > 0$, the types of the $N$ individuals are exchangeable. Now, for each $t \geq 0$, we define

$$X_t^N = \frac{1}{N} \sum_{i=1}^{N} \zeta^N_t(i),$$ (0.1)

and if $t$ is a death time

$$a_N(t) = \frac{1}{N} \sum_{i=1}^{N} \zeta^N_t(i) = X_t^N.$$

The random variable $X_t^N \in [0, 1]$ represents the proportion of type $b$ individuals in the population.

In the next section, we prove that $X_t^N \to X_t$ in probability, locally uniformly in $t \geq 0$, where $X_t$ is a $[0, 1]$-valued Markov process, solution to the stochastic differential equation

$$dX_t = -\alpha X_t(1 - X_t)dt + \sqrt{X_t(1 - X_t)}dB_t, \quad t \geq 0,$$

where $B$ is standard Brownian motion. From this, we compare our new model with the model from [1], i.e we compare the model $(L^N)$ with the model $(L^\infty)$.

1. **Convergence to the Wright-Fisher diffusion with selection**

Recall the process $\{\eta_t(i), i \geq 1, t \geq 0\}$ and $\{\zeta^N_t(i), i \geq 1, t \geq 0\}$ defined in the introduction. Let $N > 2$ denote a fixed integer, which will represent the size of the population in the model $(L^N)$. For each $1 \leq M \leq N - 2$, we define

$$X_t^{M,N} = \frac{1}{M} \sum_{i=1}^{M} \zeta^N_t(i),$$

and

$$Y_t^M = \frac{1}{M} \sum_{i=1}^{M} \eta_t(i).$$

$X_t^{M,N}$ represents the proportion of type $b$ individuals at time $t$ in the model $(L^N)$ among the first $M$ individuals, and $Y_t^M$ represents the proportion of type $b$ individuals at time $t$ in the model $(L^\infty)$ among the $M$ first individuals.
Figure 3. The graphical representation of a modified look-down construction with selection of size $N = 17$. The vertical time axis is shown on the left to the right, time flows from top to bottom. Solid lines represent type $b$ individuals, while dotted lines represent type $B$ individuals. At times $s$ the level of the dying individual is 2, for all $2 < k \leq N$, the individual at level $k$ replaces instantaneously the individual at level $k - 1$. We complete the population by an individual of type 1 at level $N$. 
For each \( t \geq 0 \), let
\[
Y_t = \lim_{M \to \infty} Y_t^M.
\]
We already know that \( Y_t \) exists a.s. (see Corollary 1 in [1]).

**Remark 1.1.** We first remark that the infinite model \( (L^\infty) \) comes down from infinity. In other words, there exists a stopping time \( \zeta_\infty \) such that \( Y_{\zeta_\infty} \in \{0, 1\} \). That is to say one of the two types \( b \) or \( B \) fixates in a finite time. To see this, we will look backwards from time \( t \) to time 0.

For each \( t > 0 \), we denote by \( Z_t^N \) the highest level occupied by the ancestors at time \( t \) of the \( N \) first individuals at time 0. For simplifying, we suppose that at time 0 the \( N \) first individuals are of \( b \) type. In this condition, the process \( (Z_t^N)_{t \geq 0} \) is a jump Markov process with state space \( \{1, 2, \ldots, \infty\} \). When in state \( n \), the process jumps to

1. \( n - 1 \) at rate \( \binom{n}{2} \);
2. \( n + 1 \) at rate \( \alpha n \), \( \alpha > 0 \).

In other words, the infinitesimal generator of \( \{Z_t^N, t \geq 0\} \) is given by:
\[
Q^n f(n) = \binom{n}{2} [f(n - 1) - f(n)] + \alpha n [f(n + 1) - f(n)].
\]

Now, let
\[
\zeta_N = \inf\{t \geq 0 : Z_t^N = 1\}.
\]
We first note that \( Z_0^N = N \) and \( Z_\infty^N = 1 \). Clearly \( (\zeta_N)_{N \geq 1} \) is increasing and \( \lim_{N \to \infty} \zeta_N := \zeta_\infty \) is the time of fixation. One says that the model \( (L^\infty) \) comes down from infinity when the fixation time \( \zeta_\infty \) is finite.

We are going to show that \( \zeta_\infty < \infty \) a.s. Let
\[
f(n) = \sum_{k=n+1}^{\infty} \frac{1}{\binom{k}{2} - \alpha k} = \sum_{k=n+1}^{\infty} \frac{2}{k(k-1-2\alpha)}.
\]

For each \( n > 2\alpha \), \( f(n) \) is well defined. In the next, we suppose that \( n > 2 \). We have
\[
f(n - 1) - f(n) = \frac{1}{\binom{n}{2} - \alpha n} \quad \text{and} \quad f(n + 1) - f(n) = \frac{-1}{\binom{n+1}{2} - \alpha(n + 1)}
\]
Since \( n \to 1/\binom{n}{2} - \alpha n \) is decreasing, we obtain
\[
Q^n f(n) = \frac{\binom{n}{2}}{\binom{n}{2} - \alpha n} - \frac{\alpha n}{\binom{n+1}{2} - \alpha(n + 1)} \\
\geq \frac{\binom{n}{2}}{\binom{n}{2} - \alpha n} - \frac{\alpha n}{\binom{n}{2} - \alpha(n)} = 1.
\]

The process \( f(Z_t^N) - f(N) - \int_0^t Q^n f(Z_s^N) ds \) is a martingale. The quantity \( \zeta_N \) is a finite stopping time for each \( N \geq 1 \). Let \( k \geq 1 \), applying the optional sampling theorem to the bounded stopping time \( \zeta_N \land k \), we get:
\[
\mathbb{E}(f(Z_{\zeta_N \land k}^N) - f(N) - \int_0^{\zeta_N \land k} Q^n f(Z_s^N) ds) = f(N).
\]

With the inequality \( Q^N f(N) \geq 1 \) for some \( N > 2\alpha \), we deduce that
\[
\mathbb{E}(\zeta_N \land k) \leq \mathbb{E}(f(Z_{\zeta_N \land k}^N)) - f(N).
\]
By monotone convergence and Lebesgue’s theorem, we have \( \mathbf{E}(\zeta_N) \leq f(1) - f(N) \). Passing to the limit in \( N \), we have \( \zeta_N \uparrow \zeta_\infty \) and \( f(N) \to \infty \), thus
\[
\mathbf{E}(\zeta_\infty) \leq f(1) < \infty,
\]
and then \( \zeta_\infty < \infty \) a.s.

In the next of this section, we assume that at time \( t = 0, \zeta_0^N(i) = \eta_0(i) \), for any \( 1 \leq i \leq N \). We have the following theorem

**Theorem 1.2.** Suppose that \( X_0^N \to x \) a.s., as \( N \to \infty \), where \( 0 < x < 1 \). Then for all \( t > 0 \) \( X_t^N \to X_t \) a.s., where \( \{X_t, t \geq 0\} \) is a weak solution to the following SDE:

\[
\begin{aligned}
  &dx_t = -\alpha x_t(1 - x_t)dt + \sqrt{x_t(1 - x_t)}dB_t, \quad t \geq 0, \\
  &x_0 = x, \quad 0 < x < 1,
\end{aligned}
\]

(1.1)

\( B \) is a standard Brownian motion. This diffusion is called the Fisher-Wright diffusion with selection.

**Proof:** We first compare \( X_t^{M,N} \) and \( Y_t^M \). We have:

**Proposition 1.3.** For any \( 1 \leq M \leq N - 2 \) such that \( M^2 \geq 16\alpha(N + 1) \),

\[
\mathbf{P}\left( \exists t > 0 : X_t^{M,N} \neq Y_t^M \right) \leq (N + 1) \left( \frac{4\alpha(N + 1)}{M^2} \right)^{N-M},
\]

in particular as \( N \to \infty \), this probability tends to 0 if \( \frac{N}{M} \to c \), where \( c \) is a constant.

**Proof:** For each \( i \geq 1, t > 0 \), let \( \xi_i^{1,N} \) denote the level on which the individual who was sitting on level \( i \) at time 0 sits at time \( t \), where the evolution corresponds to the model \((N-L^\infty)\) defined in the Introduction (all arrows pointing to levels above \( N \), and all crosses on levels above \( N \) have been erased). The process \( \xi_i^{1,N} \) takes its values in \( \{1, 2, \ldots, N\} \). Each time there is a birth on a level smaller than or equal to \( \xi_i^{1,N}, \xi_i^{1,N} \) has a jump of size 1. Each time there is a death on a level smaller than or equal to \( \xi_i^{1,N}, \xi_i^{1,N} \) has a jump of size -1. In other words, \( \xi_i^{1,N} \) follows the position of the individual who was sitting on level \( i \) at time \( t = 0 \) until his possible death, then follows the position of his left neighbor if \( \xi_i^{1,N} \geq 2 \), etc. We insist upon the rule that when this individual is killed, he is replaced by his immediate left neighbor. We have

\[
\{ \exists t > 0 : X_t^{M,N} \neq Y_t^M \} = \{ \exists 1 \leq i \leq M, t > 0 \text{ such that } \xi_s^{i,N} \neq \eta(t) \} \\
= \{ \exists 1 \leq i \leq N + 1, 0 \leq s < t \text{ such that } \xi_s^{i,N} > N, \xi_t^{i,N} = M \}. 
\]

In other words, in order to have \( X_t^{M,N} \neq Y_t^M \), for some \( t > 0 \), we need that at least one individual following the look-down model with selection (the model \((L^\infty)\)) visits the level \( M \), after having visited the level \( N + 1 \), and the identity follows from the following monotonicity property : \( i < j \Rightarrow \xi_i^{1,N} \leq \xi_j^{1,N} \) a.s. for all \( t > 0 \). Consequently

\[
\mathbf{P}\left( \exists t > 0 : X_t^{M,N} \neq Y_t^M \right) \leq \sum_{i=1}^{N+1} \mathbf{P}\left( \exists 0 \leq s < t \text{ such that } \xi_s^{i,N} > N, \xi_t^{i,N} = M \right).
\]

We first show that for each \( M^2 \geq 16\alpha(N + 1) \),

\[
\mathbf{P}\left( \exists t > 0 \text{ such that } \xi_t^{N+1,N} = M \right) \leq \left( \frac{4\alpha(N + 1)}{M^2} \right)^{N-M}.
\]

(1.2)
Any individual evolving to the model \((N-L^\infty)\) (we insist that all arrows pointing to levels above \(N\), and all crosses on levels above \(N\) have been erased) sitting on a level higher than or equal to \(M+1\) is shifted to the right at rate at least \(M(M+1)/2\), while any individual is shifted to the left at a rate which is bounded above by \(\alpha(N+1)\). Hence, we can couple the process \(\xi_t^{N+1,N}\) with a birth and death process \(\rho_t^{N,M}\), where \(\rho_0^{N,M} = N+1\), with birth rate \(M(M+1)/2\) and death rate \(\alpha(N+1)\). We have

\[
\rho_t^{N,M} \leq \xi_t^{N+1,N}, \quad 0 \leq t \leq \tau_M^N,
\]

where

\[
\tau_M^N = \inf\{t > 0, \rho_t^{N,M} = M\}.
\]

Indeed, an individual starting from \(N+1\) at time 0 will always be at level higher than or equal to \(\rho_t^{N,M}\) at time \(t\). Clearly

\[
P(\exists t > 0 \text{ such that } \xi_t^{N+1,N} = M) \leq P(\tau_M^N < \infty),
\]

hence (1.2) follows from

**Lemma 1.4.**

\[
P(\tau_M^N < \infty) \leq \left(\frac{4\alpha(N+1)}{M^2}\right)^{N-M}.
\]

**Proof:** Let \(\{X_n, n \geq 1\}\) and \(\{Y_n, n \geq 1\}\) be two mutually independent sequence of i. i. d. r. v.'s, the \(X_n\)'s being exponential with parameter \(M(M+1)/2\), the \(Y_n\)'s being exponential with parameter \(\alpha(N+1)\). We have

\[
P(\tau_M^N < \infty) \leq \sum_{n=1}^{\infty} P(X_1 + \cdots + X_n > Y_1 + \cdots + Y_{n+N-M}).
\]

Indeed, for the process \(\rho_t^{N,M}\) to reach \(M\), we need an excess of \(N+1-M\) deaths compared to the number of births. Now

\[
P(X_1 + \cdots + X_n > Y_1 + \cdots + Y_{n+N-M}) = P(e^{M(M+1)/4}X_1 \cdots X_n - Y_1 \cdots Y_{n+N-M} > 1)
\]

\[
\leq (E[e^{M(M+1)/4}X_1])^n (E[e^{-M(M+1)/4}Y_1])^{n+N-M}
\]

\[
= 2^n \left(\frac{\alpha(N+1)}{\alpha(N+1) + M(M+1)/4}\right)^{n+N-M}
\]

\[
\leq \left(\frac{8\alpha(N+1)}{M^2}\right)^n \left(\frac{4\alpha(N+1)}{M^2}\right)^{N-M}.
\]

Summing from \(n = 1\) to \(\infty\) yields the result, since

\[
\sum_{n=1}^{\infty} \left(\frac{8\alpha(N+1)}{M^2}\right)^n \leq 1, \quad \text{provided } M^2 \geq 16\alpha(N+1).
\]

We can now conclude the proof of Proposition 1.3. We note that for \(M^2 \geq 16\alpha(N+1)\), any \(1 \leq i \leq N+1\),

\[
P(\exists 0 < s < t \text{ such that } \xi_s^{i,N} = N+1, \xi_t^{i,N} = M) \leq \left(\frac{4\alpha(N+1)}{M^2}\right)^{N-M}.
\]
Indeed, wait until \( \theta_{i,N} = \inf\{t \geq 0, \xi^{i,N}_t = N + 1\} \), which is a stopping time at which the Markov process \( \xi^{i,N}_t \) starts afresh, and then use the same argument as that of Lemma 1.4. Consequently

\[
P(\exists 1 \leq i \leq N + 1, 0 < s < t \text{ such that } \xi^{i,N}_s = N + 1, \xi^{i,N}_t = M) 
\leq (N + 1) \left( \frac{4\alpha(N + 1)}{M^2} \right)^{N-M}.
\]

The result follows.

We can now proceed with the proof of Theorem 1.2. Recall the definition of \( Y^M_t \). We have shown in [1] that

\[ Y^M_t \to X_t \text{ a.s., as } M \to \infty, \quad (1.3) \]

where \( X_t \) is a \([0, 1]\)-valued Markov process which solution of (1.1). Now, we are going to show that \( X^N_t \to X_t \text{ a.s.} \) for all \( T > 0 \).

Let \( M = N - 2 \). We have

\[ |M X^M_{t,N} - N X^N_t| \leq 2. \]

From which, we deduce that

\[ |X^M_{t,N} - X^N_t| \leq \frac{4}{M}. \quad (1.4) \]

The last assertion together with Proposition 1.3 and (1.3), implies that

\[ X^N_t \to X_t \text{ a.s.} \]

The result follows.

We can in fact prove an additional property.

**Corollary 1.5.** For all \( T > 0 \),

\[ \sup_{0 \leq t \leq T} |X^N_t - X_t| \to 0 \text{ in probability, as } N \to \infty. \]

**Proof:** Thanks to (1.4), it suffices to show that \( \sup_{0 \leq t \leq T} |X^{N-2,N}_t - X_t| \to 0 \) in probability, as \( N \to \infty. \)

For any \( N \geq 3 \), we have

\[
P\left( \sup_{0 \leq t \leq T} |X^{N-2,N}_t - X_t| > \varepsilon \right) \leq P\left( \exists 0 < s < t : X^{N-2,N}_s \neq Y^{N-2}_t \right) + P\left( \sup_{0 \leq t \leq T} |Y^{N-2}_t - X_t| > \frac{\varepsilon}{2} \right).
\]

The result follows from the Proposition 1.3 and Corollary 4 in [1].

**References**


