Moran model

The Moran model describes the evolution of a collection of cells that is maintained at a constant population size of \( N \). Thinking of the population genetics situation for which it was developed, we will often call the cells individuals. In comparing the results here with those given in the literature one must take into account that in genetics most organisms are diploid (have two copies of their DNA), so the population in the Moran model commonly consists of \( 2N \) copies of a locus.

1 Neutral case

In the simplest version of the model (with no selection or mutation), the dynamics of the Moran model, which occur in continuous time, can be described as follows:

- Each individual is replaced at rate 1. That is, individual \( x \) lives for an exponentially distributed amount with mean 1 and then is “replaced.”
- To replace individual \( x \), we choose an individual at random from the population (including \( x \) itself) to be the parent of the new individual.

Suppose now that each individual has one of two alleles \( A \) and \( a \), and let \( X_t \) be the number of copies of \( A \). The transition rates for \( X_t \) are

\[
\begin{align*}
    i &\rightarrow i + 1 \quad \text{at rate} \quad b_i = (N - i) \cdot \frac{i}{N} \\
    i &\rightarrow i - 1 \quad \text{at rate} \quad d_i = i \cdot \frac{N - i}{N}
\end{align*}
\]

where \( b \) is for birth and \( d \) is for death. In words, \( a \)'s are selected for possible replacement at total rate \( N - i \). The number of \( A \)'s will increase if an \( A \) is chosen to be the parent of the new individual, an event of probability \( i/N \). Similarly, \( A \)'s are selected for possible replacement at total rate \( i \). The number of \( A \)'s will decrease if an \( a \) is chosen to be the parent of the new individual, an event of probability \( (N - i)/N \). Note that \( b_i = d_i \).

Let \( \tau = \min \{ t : X_t = 0 \text{ or } X_t = N \} \) be the fixation time, i.e., the first time at which all individuals have the same type. Since it is possible to reach the absorbing states 0 and \( N \) starting from any interior state \( 0 < i < N \) we have \( P_i(\tau < \infty) = 1 \)

Theorem 1. In the Moran model, the probability that \( A \) becomes fixed when there are initially \( i \) copies is \( i/N \).

Proof. The rates for up and down jumps are the same, so \( (d/dt)E_i X_t = 0 \), and hence \( E_i X_t \) is constant, i.e., \( X_t \) is a martingale. Intuitively this implies that

\[
i = E_i X_\tau = NP_i(X_\tau = N)
\]

To prove this we note that

\[
i = E_i X_t = NP_i(X_\tau = N, \tau \leq t) + E_i(X_t; \tau > t)
\]

Letting \( t \to \infty \) and noting \( P_i(\tau > t) \to 0 \), \( |X_t| \leq N \) the desired result follows. \qed
Let $T_k = \min\{t : X_t = k\}$ be the hitting time of $k$. Writing $\bar{E}_i\tau = E_i(\tau|T_N < T_0)$, we can state

**Theorem 2.** Let $p = i/N$. In the Moran model when $N$ is large

$$\bar{E}_i\tau \approx -\frac{N(1-p)}{p} \log(1-p) \tag{3}$$

As $p \to 0$, $-\log(1-p)/p \to 1$, so

$$\bar{E}_i\tau \sim N \tag{4}$$

**Proof.** Let $S_j$ be the amount of time spent at $j$ before time $\tau$ and note that

$$E_i\tau = \sum_{j=1}^{N-1} E_i S_j$$

Let $N_j$ be the number of visits to $j$. Let $q(j) = 2j(N-j)/N$ be the rate at which the chain leaves $j$. Since each visit to $j$ lasts for an exponential amount of time with mean $1/q(j)$, we have

$$E_i S_j = \frac{1}{q(j)} E_i N_j$$

To compute $E_i N_j$, we begin by noting that

$$P_i(N_j \geq 1) = P_i(T_j < \infty)$$

Letting $T_j^+ = \min\{t : X_t = j \text{ and } X_s \neq j \text{ for some } s < t\}$ be the time of the first return to $j$, we have for $n \geq 1$

$$P_i(N_j \geq n + 1|N_j \geq n) = P_j(T_j^+ < \infty)$$

The last formula shows that, conditional on $N_j \geq 1$, $N_j$ has a geometric distribution with success probability $P_j(T_j^+ = \infty)$. Combining this with our formula for $P_i(N_j \geq 1)$, we have

$$E_i N_j = \frac{P_i(T_j < \infty)}{P_j(T_j^+ = \infty)}$$

Since the average value of $X_t$ is constant in time, the martingale argument in (2) shows that for $0 \leq i \leq j$

$$i = jP_i(T_j < T_0) + 0 \cdot [1 - P_i(T_j < T_0)]$$

and solving gives

$$P_i(T_j < T_0) = \frac{i}{j} \quad P_i(T_j > T_0) = \frac{j-i}{j}$$

Similar reasoning shows that for $j \leq i \leq N$,

$$i = jP_i(T_j < T_N) + N[1 - P_i(T_j < T_N)]$$

and solving gives

$$P_i(T_j < T_N) = \frac{N-i}{N-j} \quad P_i(T_j > T_N) = \frac{i-j}{N-j}$$
When the process leaves $j$, it goes to $j-1$ or $j+1$ with equal probability, so
\[ P_j(T^+_j = \infty) = \frac{1}{2} \cdot P_{j+1}(T_j > T_N) + \frac{1}{2} \cdot P_{j-1}(T_j > T_0) \]
\[ = \frac{1}{2} \cdot \frac{1}{N-j} + \frac{1}{2} \cdot \frac{1}{j} = \frac{N}{2j(N-j)} \]

Putting our results into (7) gives
\[ E_i N_j = \begin{cases} \frac{i}{j} \cdot \frac{2j(N-j)}{N} & i \leq j \\ \frac{N-i}{N-j} \cdot \frac{2j(N-j)}{N} & j \leq i \end{cases} \]

Since $q(j) = 2j(N-j)/N$, (6) gives
\[ E_i S_j = \begin{cases} \frac{i}{j} & i \leq j \\ \frac{N-i}{N-j} & j \leq i \end{cases} \]

If we let $h(i) = P_i(T_N < T_0)$ and let $p_t(i,j)$ be the transition probability for the Moran model, then it follows from the definition of conditional probability and the Markov property that
\[ \bar{p}_t(i,j) = \frac{P_i(X_t = j, T_N < T_0)}{P_i(T_N < T_0)} = p_t(i,j) \cdot \frac{h(j)}{h(i)} \]

Integrating from $t = 0$ to $\infty$, we see that the conditioned chain has
\[ \bar{E}_i S_j = \int_0^\infty \bar{p}_t(i,j) \, dt = \frac{h(j)}{h(i)} E_i S_j \]

By the reasoning that led to (5),
\[ \bar{E}_i \tau = \sum_{j=1}^{N-1} \bar{E}_i S_j = \sum_{j=1}^{N-1} \frac{1}{j} \cdot \sum_{j=1}^{i-1} \frac{j}{N-j} \]

The first sum is $N - i$. For the second we note that
\[ \sum_{j=1}^{i-1} \frac{j}{N-j} = N \sum_{j=1}^{i-1} \frac{j}{1-j/N} \cdot \frac{1}{N} \approx N \int_0^p \frac{u}{1-u} \, du \]

where $p = i/N$. To evaluate the integral, we note that it is
\[ = \int_0^p -1 + \frac{1}{1-u} \, du = -p - \log(1-p) \]
Combining the last three formulas gives

\[ \bar{E}_i \tau \approx N(1 - p) + \frac{N(1 - p)}{p}(-p - \log(1 - p)) \]

\[ = -\frac{N(1 - p)}{p} \log(1 - p) \]

which gives (3). \qed
2 Directional selection

In this section, we will introduce selection letting 1 and 1\(-s\) be the relative fitnesses of the two alleles, \(A\) and \(a\). Let \(X_t\) be the number of \(A\)'s at time \(t\). Thinking of the fitnesses as the probability that an offspring of that type is viable, we can formulate the transition rates of the Moran model with selection as

\[
i \rightarrow i + 1 \quad \text{at rate} \quad b_i = (N - i) \cdot \frac{i}{N}
\]

\[
i \rightarrow i - 1 \quad \text{at rate} \quad d_i = i \cdot \frac{N - i}{N} \cdot (1 - s)
\]

(13) \quad sMoranrate

In words, \(a\)'s are selected for possible replacement at total rate \(N - i\). The number of \(A\)'s will increase if an \(A\) is chosen to be the parent of the new individual, an event of probability \(i/N\). The reasoning is similar for the second rate, but in this case the replacement only occurs with probability \(1 - s\).

**Theorem 3.** In the Moran model with selection \(s > 0\)

\[
P_i(T_N < T_0) = \frac{1 - (1 - s)^i}{1 - (1 - s)^N}
\]

(14) \quad fixselect

When \(i = 1\), the numerator is just \(s\). If selection is strong, i.e., \(Ns\) is large, then \((1 - s)^N \approx 0\) and the probability of fixation of a new mutant is just \(s\). When \(Ns = O(1)\), \((1 - s) \approx e^{-s}\), so (14) can be written as

\[
P_i(T_N < T_0) \approx \frac{1 - e^{-is}}{1 - e^{-Ns}}
\]

(15) \quad fixselecta

This case in which \(s = O(1/N)\) is called the weak selection regime.

**Proof.** Let \(h(i) = P_i(T_N < T_0)\). Births happen at rate \(b_i\) and deaths at rate \(d_i\), so the probability a birth occurs before a death is \(b_i/(b_i + d_i)\) and we have

\[
h(i + 1) - h(i) = \frac{d_i}{b_i} (h(i) - h(i - 1)) = (1 - s)(h(i) - h(i - 1))
\]

Now \(h(0) = 0\), so if we let \(c = h(1)\) and iterate, it follows that

\[
(\star) \quad h(i + 1) - h(i) = c(1 - s)^i
\]

Summing we have

\[
h(j) = \sum_{i=0}^{j-1} c(1 - s)^i = c \frac{1 - (1 - s)^j}{s}
\]

We must have \(h(N) = 1\) so \(c = s/(1 - (1 - s)^N)\) and the desired result follows. \(\square\)
We can also prove the result using a more intuitive martingale argument.

Another derivation of (14). To motivate the computation, we begin by recalling the martingale proof of Theorem 1. Let $\tau = T_0 \wedge T_N$. When $s = 0$, $E X_t$ is constant in time, so we have

$$i = N \cdot P_i(X_\tau = N) + 0 \cdot P_i(X_\tau = 0)$$

Solving, we have $P_i(X_\tau = N) = i/N$.

When $s > 0$, $b_i/(b_i + d_i) = 1/(2 - s)$. A little calculation shows that

$$i \mapsto i + 1 \quad \text{at rate} \quad \frac{1}{2 - s}$$

$$i \mapsto i - 1 \quad \text{at rate} \quad \frac{1}{2 - s} = \frac{1}{1 - s}$$

so, in this case, the value of $E(1 - s)^{X_t}$ stays constant in time. Reasoning as before,

$$(1 - s)^{i} = (1 - s)^{N} P_i(X_\tau = N) + 1 \cdot [1 - P_i(X_\tau = N)]$$

Solving we have

$$P_i(X_\tau = N) = \frac{1 - (1 - s)^i}{1 - (1 - s)^N}$$

in agreement with (14).

One can generalize Theorem 2 to compute the expected time to fixation in the model with selection. However to obtain more insight into what is happening during the fixation of a favorable allele, we will take a different approach.

\begin{theorem}[Theorem 4] In the Moran model with selection as $N \to \infty$

$$E_1 \tau \sim \frac{2}{s} \log N$$

\end{theorem}

\begin{proof}

The key is to establish that there are three phases in the fixation process.

1. While the advantageous $A$ allele is rare, the number of $A$’s can be approximated by a supercritical branching process.

2. While the frequency of $A$’s, $u_A \in [\epsilon, 1 - \epsilon]$ there is very little randomness and $u_A$ follows the solution of the logistic differential equation: $du_A/dt = su_A(1 - u_A)$.

3. While the disadvantageous $a$ allele is rare, the number of $a$’s can be approximated by a subcritical branching process.

**Phase 1.** Let $i$ be the number of $A$’s. If $i/N$ is small, then the transition rates in (13) simplify:

$$i \rightarrow i + 1 \quad \text{at rate} \quad b_i \approx i$$

$$i \rightarrow i - 1 \quad \text{at rate} \quad d_i \approx (1 - s)i$$

\end{proof}
This is a continuous time branching process in which each of the $i$ individuals gives birth at rate 1 and dies at rate $1 - s$. Letting $Z_t$ be the number of individuals at time $t$, it is easy to see from the description that

$$\frac{d}{dt}EZ_t = sEZ_t$$

so $EZ_t = Z_0e^{st}$. A result from the theory of branching processes, see Athreya and Ney (1972), shows that as $t \to \infty$

$$e^{-st}Z_t \to W$$  \hspace{1cm} (16)

The limit $W$ may be 0, and will be if the branching process dies out, that is, if $Z_t = 0$ for some $t$. However, on the event that the process does not die out $\Omega_\infty = \{Z_t > 0 \text{ for all } t\}$, we have $W > 0$.

Let $T_1$ be the first time that $X_t = M = N/\log N$. Using (16), we see that $(e^{-st}Z_t|\Omega_\infty) \to W = (W|W > 0)$ so if we condition on survival

$$\frac{N}{\log N} \approx \exp(sT_1)W$$

and solving gives

$$T_1 \approx \frac{1}{s} \log \left( \frac{N}{W \log N} \right) \approx \frac{1}{s} \log(N)$$

**Phase 2.** Let $T_2$ be the first time that $X_t = N - M$, where $M = N/\log N$. As we will now show, during the second phase from $T_1$ to $T_2$ the process behaves like the solution of the logistic differential equation. Let $X_t$ be the number of copies of the mutant allele at time $t$, and let $Y_t^N = X_t/N$. $Y_t^N$ makes transitions as follows:

$$\frac{i}{N} \to \frac{i + 1}{N} \quad \text{at rate } b_i = N - i \cdot \frac{i}{N}$$

$$\frac{i}{N} \to \frac{i - 1}{N} \quad \text{at rate } d_i \approx (1 - s)i \cdot \frac{N - i}{N}$$

When $Y_0^N = i/N = y$, the infinitesimal mean

$$\frac{d}{dt}EY_t^N = b_i \cdot \frac{1}{N} + d_i \cdot \left(- \frac{1}{N} \right) = s\frac{N - i}{N} \cdot \frac{i}{N} = sy(1 - y)$$

while the infinitesimal variance

$$\frac{d}{dt}E(Y_t^N - y_0)^2 = (b_i + d_i) \cdot \frac{1}{N^2} = (2 - s)\frac{N - i}{N} \cdot \frac{i}{N} \cdot \frac{1}{N} \to 0$$

In this situation, results in Section 7.4 of Ethier and Kurtz (1986), show that as $N \to \infty$, $Y_t^N$ converges to $Y_t$, the solution of the logistic differential equation

$$dY_t = sY_t(1 - Y_t)$$

It is straightforward to check that the solution of this equation is

$$Y_t = \frac{1}{1 + Ce^{-st}}$$
where \( C = (1 - Y_0)/Y_0 \). In the case of interest, \( Y_0 = 1/\log(N) \), so \( C \approx \log(N) \). Thus \( Y_t = 1 - 1/(\log N) \) when

\[
(\log N)e^{-t} = \frac{\log N}{\log N - 1} - 1 = \frac{1}{\log N - 1} \sim \frac{1}{\log N}
\]

Solving, we find that \( T_2 - T_1 \approx 2\log\log N \).

**Phase 3.** To achieve fixation of the \( A \) allele mutation after time \( T_2 \), the \( M = N/(\log N) \) \( a \) alleles must decrease to 0. The number of \( a \) alleles, \( Z_t \), makes transitions

\[
\begin{align*}
& j \to j + 1 \quad \text{at rate } d_{N-j} \approx (1 - s)j \\
& j \to j - 1 \quad \text{at rate } b_{N-j} \approx j
\end{align*}
\]

That is, \( Z_t \) is a continuous time branching process in which each of the \( j \) individuals gives birth at rate \((1 - s)\) and dies at rate 1. By arguments in phase 1, \( EZ_t = Z_0e^{-st} \) so it takes about \((1/s)\log(2N)\) units of time to reach 0.

The times in the three phases were

- Phase 1: \((1/s)\log(N)\)
- Phase 2: \(\log\log(N)\)
- Phase 3: \((1/s)\log(N)\)

and we have proved Theorem 4.