MTH 428/528 - Lectures 8-11

Yevgeniy Kovchegov
Oregon State University
Topics:

- Markov chains
- Wright-Fisher Model
- Moran Model
- Fixation times for Moran model
Markov chains.

Consider a sequence of random variables $X_0, X_1, X_2, \ldots$ with values in the discrete state space $S$.

The sequence $\{X_t\}_{t=0,1,\ldots}$ is said to be a discrete time Markov chain if it satisfies the following property, known as Markov property:

$$P(X_{t+1} = j \mid X_t = i, X_{t-1} = i_{t-1}, \ldots, X_1 = i_1, X_0 = i_0) = P(X_{t+1} = j \mid X_t = i)$$

A Markov chain $\{X_t\}_{t=0,1,\ldots}$ is said to be time homogeneous if

$$P(X_{t+1} = j \mid X_t = i) = p(i, j) \quad \text{for all } t = 0, 1, 2, \ldots$$
Wright-Fisher Model.

Model: Let $X_t$ denote the number of alleles $A$ in $t$-th generation. Thus, there are $2N - X_t$ of alleles $a$.

For each of the $2N$ copies of the locus in $(t+1)$-st generation, one of the $2N$ alleles in $t$-th generation is selected uniformly at random. Thus, in these $2N$ Bernoulli trials, allele $A$ comes up with probability

$$p = \frac{X_t}{2N}$$

and allele $a$ comes up with probability $1 - p$.

Hence, given that we know $X_t$, variable $X_{t+1}$ representing the number of alleles $A$ in $(t+1)$-st generation is a **Binomial random variable** with $2N$ trials and $p = \frac{X_t}{2N}$, i.e.

$$P(X_{t+1} = k \mid X_t = m) = \binom{2N}{k} p^k (1 - p)^{2N-k}, \quad \text{where} \quad p = \frac{m}{2N}.$$
Wright-Fisher Model.

Hence, given that we know $X_t$, variable $X_{t+1}$ representing the number of alleles $A$ in $(t+1)$-st generation is a Binomial random variable with $2N$ trials and $p = \frac{X_t}{2N}$, i.e.

$$P(X_{t+1} = k \mid X_t = m) = \binom{2N}{k} p^k (1 - p)^{2N-k}, \quad \text{where } p = \frac{m}{2N}.$$
Birth-and-death chain.

\[ P = \begin{pmatrix} r_0 & p_0 & 0 & 0 & \cdots \\ q_1 & r_1 & p_1 & 0 & \cdots \\ 0 & q_2 & r_2 & p_2 & \cdots \\ 0 & 0 & q_3 & r_3 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \]

Nearest neighbor transitions:

\[ q_j + r_j + p_j = 1. \]
Moran Model.

Consider $N$ individuals, $n = 2N$ copies of the locus.

- Wright-Fisher Model: non-overlapping generations (e.g. annual plants).
- Moran Model (1958): one individual changes at a time (we have $n = 2N$ ‘individuals’).

There, $X_t$ denote the number of alleles $A$ at time $t$.

$$P(X_{t+1} = i + 1 \mid X_t = i) = \frac{i}{n} \cdot \frac{n-i}{n}$$

$$P(X_{t+1} = i - 1 \mid X_t = i) = \frac{n-i}{n} \cdot \frac{i}{n}$$

State space: $S = \{0, 1, 2, \ldots, n\}$. 
Fixation times for Moran model.

Consider a birth-and-death chain $X_t$ on $S = \{0, 1, \ldots, n\}$ with forward probabilities and backward probabilities denoted respectively $p_j$ and $q_j$. Our goal is to find fixation time $T_f$ which is defined as the following first hitting time:

$$T_f = \min \{ t \geq 0 : X_t = 0 \text{ or } n \}.$$ 

We let

$$\varphi(j) = E[T_f \mid X_0 = j],$$

and write the following recurrence equation:

\[
\begin{align*}
\varphi(j) &= 1 + q_j \varphi(j - 1) + r_j \varphi(j) + p_j \varphi(j + 1) \quad \text{for } j = 1, 2, \ldots, n - 1 \\
\varphi(0) &= 0, \\
\varphi(n) &= 0.
\end{align*}
\]
Fixation times for Moran model.

\[
\begin{cases}
\varphi(j) = 1 + q_j \varphi(j - 1) + r_j \varphi(j) + p_j \varphi(j + 1) & \text{for } j = 1, 2, \ldots, n - 1 \\
\varphi(0) = \varphi(n) = 0.
\end{cases}
\]

We let \( \Delta \varphi(j) = \varphi(j + 1) - \varphi(j) \). Then for \( 0 < j < n \),

\[
\Delta \varphi(j) = \frac{q_j}{p_j} \Delta \varphi(j - 1) - \frac{1}{p_j}.
\]

Hence,

\[
\Delta \varphi(j) = \Delta \varphi(j - 1) - \frac{1}{p_j}
\]

for the Moran process since

\[
p_j = q_j = \frac{j(n-j)}{n^2}.
\]

Therefore,

\[
\Delta \varphi(j) = \Delta \varphi(0) - \sum_{k=1}^{j} \frac{1}{p_k} = \varphi(1) - \sum_{k=1}^{j} \frac{1}{p_k}
\]
Fixation times for Moran model.

We let $\Delta \varphi(j) = \varphi(j + 1) - \varphi(j)$. Then for $0 < j < n$,

$$
\Delta \varphi(j) = \Delta \varphi(0) - \sum_{k=1}^{j} \frac{1}{p_k} = \varphi(1) - \sum_{k=1}^{j} \frac{1}{p_k}
$$

Hence, for any $j = 1, 2, \ldots, n$,

$$
\varphi(j) = \Delta \varphi(j-1) + \Delta \varphi(j-2) + \ldots + \Delta \varphi(0) = j \varphi(1) - \sum_{k:1 \leq k \leq j-1} \frac{j - k}{p_k}.
$$

Thus we can find the $\varphi(1)$ as for $j = n$,

$$
0 = \varphi(n) = n \varphi(1) - \sum_{k=1}^{n-1} \frac{j - k}{p_k} = n \left( \varphi(1) - n \sum_{k=1}^{n-1} \frac{1}{k} \right)
$$

and therefore $\varphi(1) = n \sum_{k=1}^{n-1} \frac{1}{k}$. 
Fixation times for Moran model.

For any $j = 1, \ldots, n$, we derived

$$\phi(j) = j\phi(1) - \sum_{k:1 \leq k \leq j-1} \frac{j-k}{p_k}.$$ 

Observe that

$$\frac{j-k}{p_k} = n^2 \left( \frac{j}{n} \cdot \frac{1}{k} - \left(1 - \frac{j}{n}\right) \cdot \frac{1}{n-k} \right).$$

We let $\ell(1) = 0$ and for $j \geq 2$,

$$\ell(j) := \sum_{k:1 \leq k \leq j-1} \frac{1}{k}.$$ 

Then $\phi(1) = n\ell(n)$ and

$$\phi(j) = jn\ell(n) - n^2 \cdot \frac{j}{n} \sum_{k:1 \leq k \leq j-1} \frac{1}{k} + n^2 \cdot \left(1 - \frac{j}{n}\right) \sum_{k:1 \leq k \leq j-1} \frac{1}{n-k}.$$
Fixation times for Moran model.

We let $\ell(1) = 0$ and for $j \geq 2$, $\ell(j) := \sum_{k:1 \leq k \leq j-1} \frac{1}{k}$.

Then $\varphi(1) = n\ell(n)$ and

$$
\varphi(j) = jn\ell(n) - n^2 \cdot \frac{j}{n} \sum_{k:1 \leq k \leq j-1} \frac{1}{k} + n^2 \cdot \left(1 - \frac{j}{n}\right) \sum_{k:1 \leq k \leq j-1} \frac{1}{n-k}
$$

rewrites as

$$
\varphi(j) = jn\ell(n) - n^2 \cdot \frac{j}{n} \ell(j) + n^2 \cdot \left(1 - \frac{j}{n}\right) (\ell(n) - \ell(n-j+1)).
$$

Thus,

$$
\varphi(j) = n^2 \left[ \frac{j}{n} (\ell(n) - \ell(j)) + \left(1 - \frac{j}{n}\right) (\ell(n) - \ell(n-j+1)) \right].
$$
Fixation times for Moran model: discussion.

We let $\ell(1) = 0$ and for $j \geq 2$, $\ell(j) := \sum_{k:1 \leq k \leq j-1} \frac{1}{k}$, and obtained

$$\varphi(j) = n^2 \left[ \frac{j}{n} (\ell(n) - \ell(j)) + \left( 1 - \frac{j}{n} \right) (\ell(n) - \ell(n - j + 1)) \right].$$

Observe that

$$\ell(j) := \sum_{k:1 \leq k \leq j-1} \frac{1}{k} = \int_{1}^{j} \frac{1}{x} dx + C_j = \ln(j) + C_j,$$

where $0 < C_j < \gamma = 0.57721 \ldots$ (the Euler constant).
Fixation times for Moran model: discussion.

\[ \ell(j) := \sum_{k:1 \leq k \leq j-1} \frac{1}{k} = \int_1^j \frac{1}{x} \, dx + C_j = \ln(j) + C_j, \]

where \(0 < C_j < \gamma = 0.57721\ldots\) (the Euler constant).

Source: Wikipedia.org
Fixation times for Moran model: discussion.

We let $\ell(1) = 0$ and for $j \geq 2$, $\ell(j) := \sum_{k:1 \leq k \leq j-1} \frac{1}{k}$, and obtained

$$\varphi(j) = n^2 \left[ \frac{j}{n} (\ell(n) - \ell(j)) + \left(1 - \frac{j}{n}\right) (\ell(n) - \ell(n - j + 1)) \right].$$

Observe that

$$\ell(j) := \sum_{k:1 \leq k \leq j-1} \frac{1}{k} = \int_1^j \frac{1}{x} dx + C_j = \ln(j) + C_j,$$

where $0 < C_j < \gamma = 0.57721\ldots$ (the Euler constant). Hence,

$$\varphi(j) = -n^2 \left[ \frac{j}{n} \ln \frac{j}{n} + \left(1 - \frac{j}{n}\right) \ln \left(1 - \frac{j-1}{n}\right) + \text{Const} \right].$$
Fixation times for Moran model: discussion.

Observe that

\[ \ell(j) := \sum_{k:1 \leq k \leq j-1} \frac{1}{k} = \int_{1}^{j} \frac{1}{x} dx + C_j = \ln(j) + C_j, \]

where \(0 < C_j < \gamma = 0.57721 \ldots\) (the Euler constant). Hence,

\[ \varphi(j) = -n^2 \left[ \frac{j}{n} \ln \left( \frac{j}{n} \right) + \left( 1 - \frac{j}{n} \right) \ln \left( 1 - \frac{j-1}{n} \right) + \text{Const} \right]. \]

If we let \(x = \frac{j}{n}\) be the fraction of \(A\) alleles, then for \(x \in [\epsilon, 1-\epsilon]\), the constant term goes to zero as \(n \to \infty\), and

\[ E[T_f | X_0/n = x] = -n^2 \left[ x \ln(x) + (1-x) \ln(1-x) + o(1) \right]. \]