



*The Abdus Salam
International Centre for Theoretical Physics*



1863-21

**Advanced School and Conference on Statistics and Applied
Probability in Life Sciences**

24 September - 12 October, 2007

A Galton-Watson model for evolutionary path to escape

Serik Sagitov

*Chalmers University of Technology and Gothenburg University
SE-412 96 Gothenburg, Sweden*

Serik Sagitov (Trieste 2007) <http://www.math.chalmers.se/~serik/>

A Galton-Watson model for evolutionary path to escape

(based on a forthcoming paper with M.C.Serra)

1. Evolutionary dynamics of escape: a motivating example
2. Multitype Galton-Watson model
3. The two type case
4. A forward mutation model
5. Asymptotic distribution for the time to escape
6. References

1 Evolutionary dynamics of escape: a motivating example

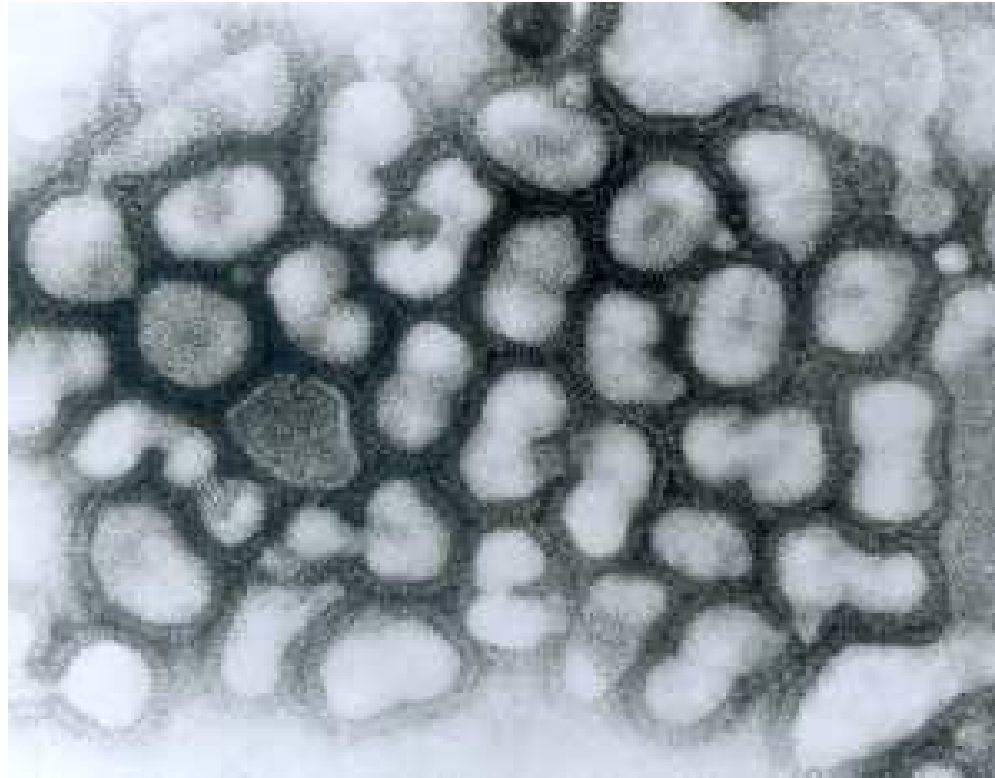
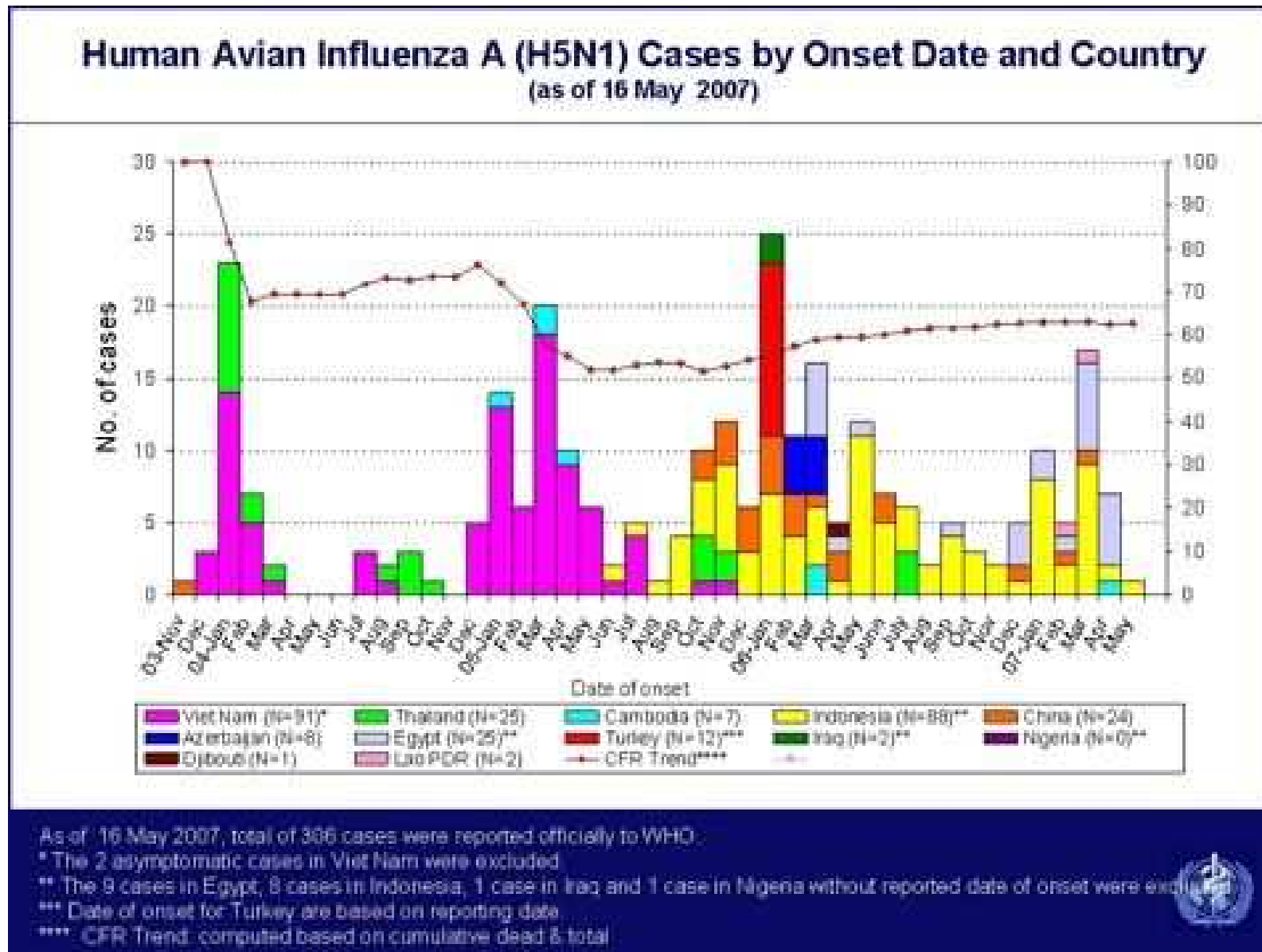


Figure 1: *Electron micrograph of avian flu viruses.*

Placed in to a new host, a virus can adjust to a hostile environment via a sequence of mutations on say L particular nucleotide sites



Escape events: via a sequence of mutations the avian flu virus escapes extinction causing an onset of Human Avian Influenza



Questions:

- introduce a simple stochastic model of virus reproduction and mutation
- assuming mutations are rare find the asymptotics of the escape probability for a population of viruses stemming from a single wild-type virus
- describe a stochastic process of virus reproduction and mutation given an escape event has occurred

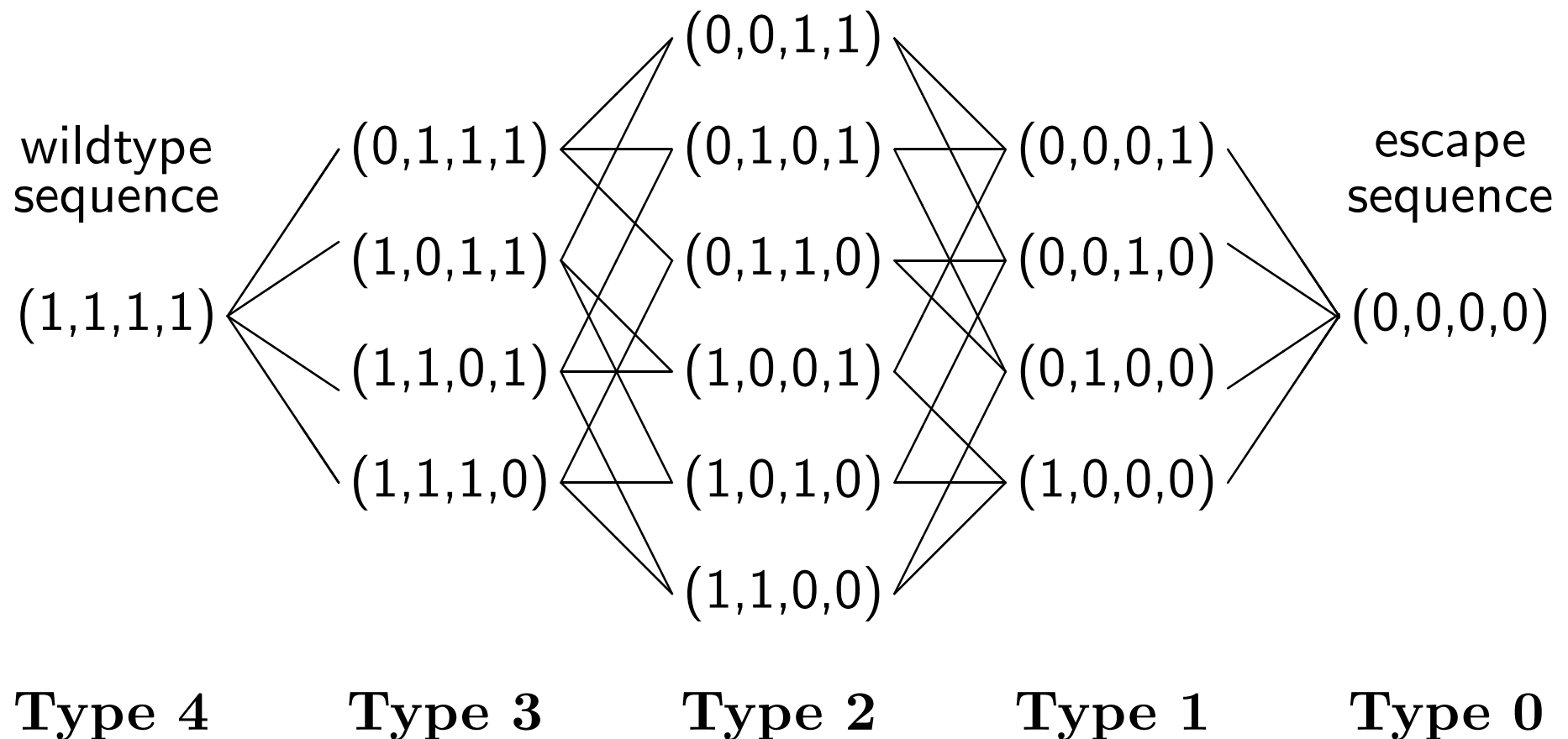
Iwasa et al suggested using a branching process model:

particles reproduce asexually and independently

small population sizes, competition among particles can be ignored

2 Multitype Galton-Watson model

The network of 0-1 sequences of length $L = 4$, where the edges represent single point mutations of probability $\mu \ll 1$



Let all the sequences with i ones have the same reproduction law
probability of having k offspring $p_i(k)$

fitness as the mean offspring number $m_i = \sum_{k=1}^{\infty} kp_i(k)$

Subcritical reproduction for all sequences except the escape one

$$0 < m_1, \dots, m_L < 1 < m_0 < \infty$$

A multitype Galton-Watson model with $L + 1$ types

time is measured in generations

particle's type i = the number of ones in the sequence

offspring's type can differ due to mutation at one or several sites

The reproduction law with mutation

$$p_i^{(\mu)}(k_0, \dots, k_L) = p_i(k) \binom{k}{k_0, \dots, k_L} \left(q_{i0}^{(\mu)}\right)^{k_0} \cdots \left(q_{iL}^{(\mu)}\right)^{k_L}$$

where $k = k_0 + \dots + k_L$ and

$$q_{ij}^{(\mu)} = P_i(\text{the given offspring becomes of type } j).$$

Clearly,

$$q_{ii}^{(\mu)} \rightarrow 1, \quad \mu \rightarrow 0$$

and for $j < i$

$$q_{ij}^{(\mu)} \sim \binom{i}{j} \mu^{i-j}. \quad (1)$$

3 The two type case

Consider a GW process with two types of particles labelled by 0 and 1

- type 1 is the subcritical wild-type with $m_1 < 1$
- type 0 is the escape type with $m_0 > 1$.

Both backward and forward mutations are allowed, but our asymptotical analysis confirms, that given mutations are rare, we can neglect the backward mutations.

Moreover, the **escape event** can (and will) be treated as **birth of at least one virus of the escape type**.

In this case the restriction $m_0 > 1$ can be dropped.

We will allow for the mutation probability per birth

$$q_{10}^{(\mu)} = \mu a_{10}^{(\mu)}(k)$$

to depend on the offspring number k . We will assume uniform convergence to a bounded function

$$a_{10}^{(\mu)}(k) \rightarrow a_{10}(k), \quad \mu \rightarrow 0.$$

Assertion 1. If the constant

$$b_{10} = \sum_{k=1}^{\infty} k p_1(k) a_{10}(k)$$

is strictly positive, then the probability of escape has asymptotics

$$Q_{10}^{(\mu)} \sim \mu b_{10} (1 - m_1)^{-1}.$$

Assertion 2. Conditioned on escape, the GW process with types 0 and 1 stemming from a single particle of type 1 is asymptotically described by a decomposable GW process with types 10 and 11 stemming from a single 10 particle.

Type 10 particles form the stem lineage (stopping on the mutation event) and type 11 particles form the side (no mutation) lineages

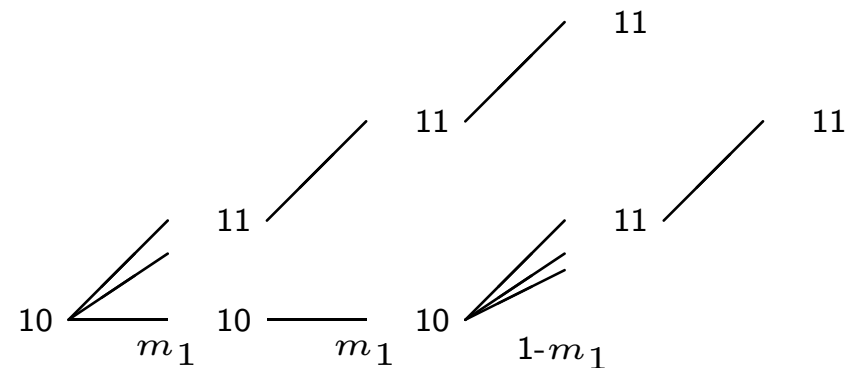
$$P_{10}(\nu_{10} = 1) = m_1$$

$$P_{10}(\nu_{10} = 0) = 1 - m_1$$

$$P_{10}(\nu_{11} = k | \nu_{10} = 1) = \frac{kp_1(k)}{m_1}$$

$$P_{10}(\nu_{11} = k | \nu_{10} = 0) = \frac{kp_1(k)a_{10}(k)}{b_{10}}$$

$$P_{11}(\nu_{11} = k) = p_1(k)$$



Remark 1. In the network model case $q_{10}^{(\mu)} \sim \mu$, so that $Q_{10}^{(\mu)} \sim \frac{\mu m_1}{1-m_1}$ and $P_{10}(\nu_{11} = k) = \frac{k p_1(k)}{m_1}$.

Remark 2. The *size-biased* reproduction law $\hat{p}(k) = \frac{k p(k)}{m}$ characterizes the stem lineage in a subcritical reproduction conditioned on non-extinction.

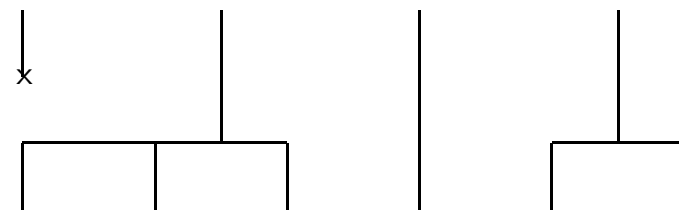
Two sampling designs to find the family size distribution

prospective $p(k)$: pick a mother and count her children

retrospective $\hat{p}(k)$: pick a daughter and count her and her sisters

$$p(0) = p(1) = p(2) = p(3) = \frac{1}{4}, m = \frac{3}{2}$$

$$\hat{p}(0) = 0, \hat{p}(1) = \frac{1}{6}, \hat{p}(2) = \frac{2}{6}, \hat{p}(3) = \frac{3}{6}$$



4 A forward mutation model

Suppose we can distinguish between $L + 1$ types of particles, labelled $0, \dots, L$ with mean offspring numbers

$$0 < m_1, \dots, m_L < 1, \quad 0 < m_0 < \infty.$$

Type i particles can only produce particles of the types $0, \dots, i$, whatever is $i \in [0, L]$.

Notice that this forward mutation model prohibits the reverse mutations for the sake of simplicity.

As the asymptotic analysis of the two type case shows, a more general model with reversed mutations should lead to the same asymptotic behavior.

As in the two type case the forward mutation probabilities ($j < i$)

$$q_{ij}^{(\mu)} = \mu^{i-j} a_{ij}^{(\mu)}(k)$$

may depend on the offspring number k . We will assume uniform convergences to bounded functions

$$a_{ij}^{(\mu)}(k) \rightarrow a_{ij}(k), \mu \rightarrow 0.$$

Let the constants

$$b_{ij} = \sum_{k=1}^{\infty} k p_i(k) a_{ij}(k)$$

be positive for all $j < i$.

Define a matrix $\mathbb{A} = [A_{ij}]_{i,j=0}^L$ by

$$\mathbb{A} = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ \frac{b_{10}}{1-m_1} & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{b_{i0}}{1-m_i} & \frac{b_{i1}}{1-m_i} & \cdots & \frac{b_{i,i-1}}{1-m_i} & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{b_{L0}}{1-m_L} & \frac{b_{L1}}{1-m_L} & \cdots & \frac{b_{L,i-1}}{1-m_L} & \frac{b_{L,i}}{1-m_L} & \cdots & \frac{b_{L,L-1}}{1-m_L} & 0 \end{bmatrix}$$

In the network case due to (1) on page 8

$$a_{ij}(k) = \binom{i}{j}, \quad b_{ij} = \binom{i}{j} m_i$$

Assertion 3. The probability of escape $Q_{i0}^{(\mu)} \sim \mu^i \chi_i$, where

$$\chi_1 = A_{10}$$

$$\chi_2 = A_{20} + A_{21}\chi_1$$

$$\chi_3 = A_{30} + A_{31}\chi_1 + A_{32}\chi_2$$

...

In terms of the matrix powers $\mathbb{A}^n = [A_{ij}^{(n)}]_{i,j=0}^L$ we can write

$$\begin{aligned} \chi_i &= A_{i0}^{(i)} \\ &= \sum_{k=0}^{i-1} \sum_{0=j_0 < j_1 < j_2 < \dots < j_k < i} A_{ij_k} \dots A_{j_1 0}. \end{aligned}$$

Put $\chi_0 = 1$ and define a probability distribution on $j = 0, \dots, i - 1$ by

$$B_{ij} = \frac{\chi_j}{\chi_i} A_{ij}$$

Assertion 4. Conditioned on escape, the limit process is described by a decomposable $2L$ -type GW process

$$P_{i0}(\nu_{i0} = 1) = m_i$$

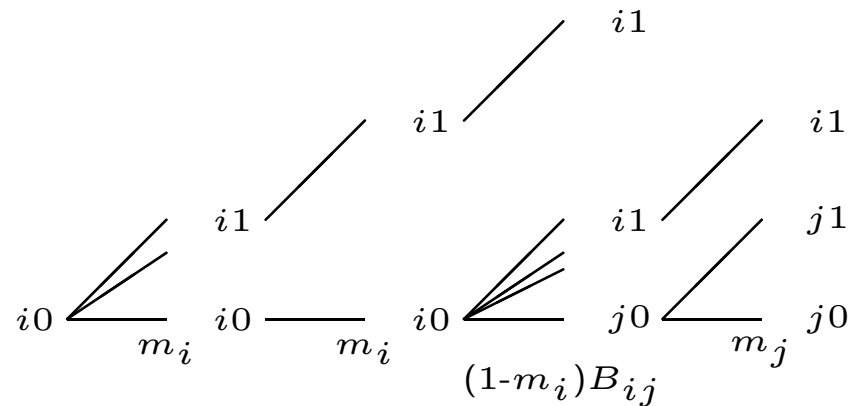
$$P_{i0}(\nu_{i0} = 0) = 1 - m_i$$

$$P_{i0}(\nu_{j0} = 1 | \nu_{i0} = 0) = B_{ij}$$

$$P_{i0}(\nu_{i1} = k | \nu_{i0} = 0) = \frac{kp_i(k)}{m_i}$$

$$P_{i0}(\nu_{i1} = k | \nu_{j0} = 1) = \frac{kp_i(k)a_{ij}(k)}{b_{ij}}$$

$$P_{i1}(\nu_{i1} = k) = p_i(k)$$



5 The time to escape

The asymptotically viable path of mutations is described by a Markov chain $\{Y(n)\}_{n \geq 0}$ with the transition matrix

$$\mathbb{D} = [D_{ij}]_{i,j=0}^L, \quad D_{ij} = (1 - m_i)B_{ij} + m_i 1_{\{i=j\}}.$$

For application purposes, it is important to study the waiting time W_i to produce the escape type 0 starting from type i , that is the waiting time until absorption at state 0

$$P(W_i \leq n) = P(Y(n) = 0 | Y(0) = i).$$

W_i is a sum of a random number of indep geometric random variables.

The Chapman-Kolmogorov equation yields a recursion for the probability $P_i(n) = P(W_i = n)$

$$P_i(n) = m_i P_i(n-1) + (1 - m_i) \sum_{j=1}^{i-1} B_{ij} P_j(n-1).$$

Turning to the expected waiting time $M_i = \sum_{n=1}^{\infty} n P_i(n)$ we derive

$$\begin{aligned} M_i &= \frac{1}{1 - m_i} + \sum_{j=1}^{i-1} \frac{B_{ij} + B_{ij}^{(2)} + \dots + B_{ij}^{(i-j)}}{1 - m_j} \\ &= \frac{1}{1 - m_i} + \sum_{j=1}^{i-1} \frac{\chi_j (A_{ij} + \dots + A_{ij}^{(i-j)})}{\chi_i (1 - m_j)} \end{aligned}$$

since the matrix powers \mathbb{A}^n and \mathbb{B}^n are connected by $B_{ij}^{(n)} = \frac{\chi_j}{\chi_i} A_{ij}^{(n)}$.

Observe that the last formula is a weighted sum of the individual waiting times $E(T_j) = \frac{1}{1-m_j}$. The corresponding weight

$$\begin{aligned} \frac{\chi_j}{\chi_i} (A_{ij} + \dots + A_{ij}^{(i-j)}) &= \frac{A_{ij} A_{j0}^{(j)} + \dots + A_{ij}^{(i-j)} A_{j0}^{(j)}}{A_{i0}^{(i)}} \\ &= P(Y(n) = j \text{ for some } n) \end{aligned}$$

gives the probability that the chain $Y(n)$ visits the state j before it is absorbed at 0.

In the case of "neutral mutation" with $m_j = m$, $j = 1, \dots, L$ we get

$$M_L = \frac{1}{1-m} \left(L - \frac{A_{L0} + \dots + A_{L0}^{(L-1)}}{\chi_L} \right).$$

Finally, we describe a case where there is a simple formula for the coefficients χ_i . Suppose that $a_{ij}(k) \equiv a_i(k)$ is the same for all daughter types j given the mother type i . Then with simplified notation $b_{ij} = c_i$ we obtain

$$\chi_i = \frac{c_i}{1 - m_i} \left(1 + \frac{c_{i-1}}{1 - m_{i-1}} \right) \cdots \left(1 + \frac{c_1}{1 - m_1} \right).$$

In this case we can also compute

$$P(Y(n) = j \text{ for some } n) = \frac{c_j}{1 + c_j - m_j}$$

and the expected total time to escape becomes

$$M_L = \frac{1}{1 - m_L} + \sum_{j=1}^{L-1} \frac{c_j}{(1 + c_j - m_j)(1 - m_j)}.$$

In particular, if $a_{ij}(k) \equiv 1$, then $c_j = m_j$ and

$$M_L = \frac{1}{1 - m_L} + \sum_{j=1}^{L-1} \frac{m_j}{1 - m_j}.$$

If furthermore $m_j \equiv m$, then

$$\chi_j = m(1 - m)^{-j}$$

and

$$P(Y(n) = j \text{ for some } n) = m.$$

In this special case the number of intermediate types has a binomial distribution $\text{Bin}(L - 1, m)$ and

$$M_L = \frac{1 + (L - 1)m}{1 - m}.$$

References

- [1] Haccou P., Jagers P., Vatutin V.A. Branching Processes: Variation, Growth and Extinction of Populations. Cambridge University Press, Cambridge, 2005.
- [2] Iwasa, Y., Michor, F., Nowak, M.A., 2003. Evolutionary dynamics of escape from biomedical intervention. Proc. Roy. Soc. London B 270, 2573-2578.
- [3] Iwasa, Y., Michor, F., Nowak, M.A., 2004. Evolutionary dynamics of invasion and escape. Journal of Theoretical Biology 226, 205-214.
- [4] Sagitov, S. and Serra M.C., 2007. Multitype Galton-Watson processes escaping extinction, 1-25, submitted to Adv. Appl. Probab.

Thank you!

