DM&P 2011: Problem Set 3 Solutions

Daniel Weissman

These solutions explain the answers in more detail than you should use in your homeworks.

Problem 1

Suppose that we have a branching process in which the number of offspring follows a Poisson distribution, with mean $\lambda = 1 + s$: the probability that an individual has $j$ offspring is $\psi_j = e^{-\lambda} \frac{\lambda^j}{j!}$.

a) Write down the probability of extinction at time $t + 1$, $Q_{t+1}$, in terms of $Q_t$. (You will need to use the fact that $\sum_{j=0}^{\infty} \frac{\lambda^j}{j!} = e^{\lambda}$.)

A: We learned in lecture that the probability distribution of a branching process evolves according to

$$Q_{t+1} = \sum_{j=0}^{\infty} \psi_j Q_t^j.$$
For a Poisson offspring distribution, this gives

\[ Q_{t+1} = \sum_{j=0}^{\infty} e^{-\lambda} \frac{\lambda^j}{j!} Q_t^j \]

\[ = e^{-\lambda} \sum_{j=0}^{\infty} \frac{(\lambda Q_t)^j}{j!} \]

\[ = e^{-\lambda} e^{\lambda Q_t} \]

\[ = e^{\lambda(Q_t-1)}, \]

where in the third line we have used the identity in the hint.

b) What is the ultimate probability of extinction, for \( s = 0.1 \)?

A: The ultimate probability of extinction, \( Q_\infty \), is the solution to

\[ Q_\infty = e^{(1+s)(Q_\infty-1)}. \]

For \( s = 0.1 \), we can solve this numerically to find that \( Q_\infty \approx 0.8 \).

c) How accurate is the approximation \( P_\infty = \frac{2s}{\text{var}(J)} \), where \( J \) is the (random) number of offspring?

A: \( P_\infty \), the probability of ultimate non-extinction, is \( P_\infty = 1 - Q_\infty \approx 0.2 \).

The variance of a Poisson distribution is the same as its mean, so the approximate formula is \( P_\infty \approx \frac{2s}{1+s} \approx 0.2 \), the same as we got from the numerical solution to the exact expression.

**Problem 2**

There are four possible values for a single DNA nucleotide: A, C, G, and T. The most common mutations in DNA are **transitions**, between A\(\leftrightarrow\)G and C\(\leftrightarrow\)T. (Other kinds of mutations, e.g., A\(\leftrightarrow\)C are called **transversions**, and are rarer;
we will ignore them here.) Let the rate of each of the four possible transition mutations be $\mu$.

a) What are the stationary states?

A: Since we are ignoring transversions, the two pairs \{A, G\} (“purines”) and \{C, T\} (“pyrimidines”) are completely out of contact with each other. Since the two mutation rates within a pair are equal, the probabilities of having either member of a pair must be equal in a stationary states. Thus, the stationary states are all states of the form $\pi_A = \pi_G = a$, $\pi_C = \pi_T = 1/2 - a$ (where $\pi_X$ is the stationary probability of having nucleotide $X$, and $0 \leq a \leq 1/2$).

b) How fast does an evolving genome approach the stationary state?

A: Let’s say that we’re nearly at the stationary state with $\pi_A = \pi_G = 1/2$, but we’re a little bit likelier to have A: $p_A = 1/2 + \epsilon$, $p_G = 1/2 - \epsilon$. In the next generation, we’ll have $p'_A = 1/2 + (1 - 2\mu)\epsilon$, $p'_G = 1/2 - (1 - 2\mu)\epsilon$. The distance from the stationary state decreased by a factor of $1 - 2\mu$, so the rate of approach to the stationary state is $-\log(|1 - 2\mu|) \approx 2\mu$ (assuming $\mu \ll 1$).

c) Write down a formula for the proportions of \{A, G, C, T\} in matrix form, i.e., write a matrix equation showing how to get the probability distribution at time $t$ from the initial distribution.

A: Let $p = \{p_A, p_G, p_C, p_T\}$. Then the transition matrix is

$$M = \begin{pmatrix} 1 - \mu & \mu & 0 & 0 \\ \mu & 1 - \mu & 0 & 0 \\ 0 & 0 & 1 - \mu & \mu \\ 0 & 0 & \mu & 1 - \mu \end{pmatrix}.$$
The probability distribution at time $t$ is

$$p(t) = M^t p(0).$$

**d) Show that the eigenvalues and eigenvectors correspond to your answers to (a), (b).**

A: Since $M$ is block-diagonal, we can analyze the upper left and bottom right blocks separately. The eigenvalues of $m = \begin{pmatrix} 1 - \mu & \mu \\ \mu & 1 - \mu \end{pmatrix}$ are $\lambda_1 = 1$ and $\lambda_2 = 1 - 2\mu$, with corresponding eigenvectors $e_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $e_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$. So any vector $v = \{a, a, b, b\}$ is an eigenvector of $M$ with eigenvalue $\lambda_1 = 1$, and the stationary states we found in (a) are all such vectors that are also valid probability distributions (i.e., that have all entries non-negative and summing to 1). Any vector $v = \{a, -a, b, -b\}$ is an eigenvector of $M$ with eigenvalue $\lambda_2 = 1 - 2\mu$. The rate of approach to the stationary states is given by the log of the reciprocal of the absolute value of this second-leading eigenvector, which is $-\log |\lambda_2| = -\log |1 - 2\mu|$, just like we found in (b).