

---

**Recursive Schemes**

1. *Fibonacci's Rabbits*: Consider a generalization of Fibonacci's model for rabbit populations, in which some of the rabbits (whether young or adult) die in each generation, and not all adults reproduce. (i) Assume the same mortality rate for young and adult rabbits, such that a fraction  $p$  of young rabbits matures to adulthood, and the same fraction  $p$  of adult rabbits survives to the next generation. (ii) Assume that a fraction  $f$  of adults reproduce the next generation of young rabbits.

(a) Write the  $2 \times 2$  (transfer) matrix that relates the populations of young and adult rabbits from one generation to the next, i.e. find the elements of the matrix  $T$ , such that

$$\begin{pmatrix} Y_{N+1} \\ A_{N+1} \end{pmatrix} = T \begin{pmatrix} Y_N \\ A_N \end{pmatrix},$$

where  $Y_N$  and  $A_N$  are the (average) numbers of young and adult rabbits in generation  $N$ .

(b) Show that for large  $N$  the population grows (or decays) exponentially, and find the ratio  $\lambda(p, f)$  between numbers of successive generations. What is the condition that separates growing and decaying populations?

(c) Find the asymptotic ratio of young to adult rabbits, and note its connection to an eigenvector of the transfer matrix.

\*\*\*\*\*

2. *Point mutations in DNA*: Since the four nucleotides in DNA have different chemical compositions and energetics, they could mutate at different rates. We shall explore whether, without natural selection at work, such preferential mutation may lead to different compositions of nucleotides.

(a) Consider a simple model in which all *transversions* (i.e. mutations between purines A and G, or between pyrimidines T and C) occur with probability  $q$ , while *transitions* (i.e. any mutation from a purine to a pyrimidine or vice versa) occur with probability  $p$ , in each generation. Write down the  $4 \times 4$  (Markov) transition matrix,  $\Pi_1$ , that relates the frequencies of nucleotides ( $p_A, p_G, p_T, p_C$ ) from one generation to the next. (Make sure that the normalization condition  $p_A + p_G + p_T + p_C = 1$  is preserved.)

(b) Find the eigenvalues of the transition matrix  $\Pi_1$ . (**Hint:** You should be able to simply guess the eigenvectors by considering the symmetries of the matrix.)

(c) Find the matrix  $\Pi_t = \Pi_1^t$ , describing the evolution of probabilities after  $t$  generations.

(d) Show that in steady state (after many duplications), all nucleotides occur with the same frequency. Estimate the number of generations (as a function of  $p$  and  $q$ ) needed to reach such a steady state.

(e) You should be able to convince yourself that for any model in which mutation rates between pairs of bases are the same in the forward and backward directions, all nucleotides are equally likely in the steady state. However, in the human genome the nucleotides C and G occur less often than A and T. This is partly due to methylation of successive CG pairs which makes them more susceptible to mutations. To mimic this asymmetry, consider an unrealistic model in which transitions from A to C and T to G occur with probability  $p_+$ , while the reverse transitions (from C to A or G to T) occur at a lower probability of  $p_-$ . (The other transitions occur at rate  $p$ , and transversions at rate  $q$  as before.) Write the modified transfer matrix corresponding to this model, and obtain the resulting frequencies of nucleotides in steady state.

\*\*\*\*\*

(Optional) **3. Correlations in the *E. coli* genome:** In the models examined in the previous problem, point mutations at each position on the DNA occur at rates independent of other locations. Consequently, they predict  $p_{XY} = p_X p_Y$ , where  $p_{XY}$  is the *joint probability* of finding nucleotides X and Y, at different locations. Test this hypothesis on the genome of *E. coli* (available on the course web-page) as follows:

(a) Calculate the frequencies of the four nucleotides in the genome.

(b) Write a program to count all 16 possible pairs of neighboring bases (e.g. AT); hence obtain the joint probabilities  $p_{XY}$ , and construct the  $4 \times 4$  matrix of correlations  $c_{XY}^{(1)} = p_{XY}/(p_X p_Y)$ .

(c) Repeat the above calculation for nucleotides that are further neighbors, and find the corresponding matrices  $c_{XY}^{(n)}$  (e.g. consider next nearest neighbor locations  $j$  and  $j + 2$  to calculate  $c_{XY}^{(2)}$ ). How do correlations decay as a function of the separation  $n$ ?

\*\*\*\*\*

**4. Bound Directed Polymer:** Consider a variant of the directed polymer model introduced in lecture 6, which is limited to the half plane  $x \geq 0$ , and attracted to the boundary at  $x = 0$ . The weight  $W(x, t)$  of paths terminating at the point  $(x, t)$  satisfies the recursion relations

$$\begin{cases} W(x, t + 1) = W(x, t - 1) + \gamma W(x - 1, t) + \gamma W(x + 1, t), & \text{for } x \geq 1, \text{ and} \\ W(0, t + 1) = \mu W(0, t - 1) + \gamma W(1, t). \end{cases}$$

Here  $\gamma = e^{-\beta g} < 1$  is the Boltzmann weight of a diagonal segment (of unfavorable energy  $g$ ), while  $\mu = e^{\beta b} > 1$  is the weight of attaching to the boundary (favored by an energy  $b$ ).

(a) Show that configurations of the polymer bound to the interface can be described by a steady state weight  $W(x, t) \propto z^t e^{-\lambda x}$ .

(b) Calculate the partition function  $Z(t)$ , after a large number of steps  $t$ .

(c) Show that for the bound state to exist,  $\mu$  must exceed a critical value  $\mu_c(\gamma)$ . What happens for  $\mu < \mu_c$ ?

(d) Plot the ‘localization length’  $\langle \ell \rangle = 1/\lambda$  as a function of  $\mu$ .

\*\*\*\*\*

**5. DNA alignment:** Consider a scheme for aligning DNA sequences in which a score  $s = 1$  is assigned to a match, while  $s = 0$  for a transversion ( $A \leftrightarrow G$  or  $T \leftrightarrow C$ ) and  $s = -\mu$  for a transition (e.g.  $A \leftrightarrow C$ ). (Assume all four nucleotides occur with equal frequency.)

(a) Find the parameter  $\lambda(\mu)$  which governs the statistics of such random gapless alignments.

(b) Show that alignments are possible only for  $\mu > \mu_c$ , and plot  $\lambda$  as a function of  $\mu$ .

\*\*\*\*\*