Chapter 5

Time dependent probabilities

5.1 Transition rates

1. *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 *transitions* (i.e. mutations between purines A and G, or between pyramidines T and C) occur with probability q, while *transversions* (i.e. any mutation from a purine to a pyrimidine or vice versa) occur with probability p, in each generation. Write down the 4×4 (Markov) transition matrix, Π_1, that relates the frequencies of nucleotides ($p_{\rm A}$, $p_{\rm G}$, $p_{\rm T}$, $p_{\rm C}$) from one generation to the next. (Make sure that the normalization condition $p_{\rm A}+p_{\rm G}+p_{\rm T}+p_{\rm C}=1$ is preserved.)

   (b) Find the eigenvalues of the transition matrix Π_1. (**Hint:** You should be able to simply guess the eigenvectors by considering the symmetries of the matrix.)
   item Find the matrix $\Pi_t = \Pi_1^t$, describing the evolution of probabilities after t generations.

   (c) 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.

   (d) 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 transversions from A to C and T to G occur with probability p_+, while the reverse transversions (from C to A or G to T) occur at a lower probability of p_-. (The

other transversions occur at rate p, and transitions at rate q as before.) Write the modified transfer matrix corresponding to this model, and obtain the resulting frequencies of nucleotides in steady state.

*****

2. *Activation/deactivation reaction:* Many molecules in biology can be made active or inactive through the addition of a phosphate group. The enzyme that adds the phosphate group is usually termed a kinase, while a phosphatase removes this group. Let us consider a case where a finite number N of such molecules within a cell can be exchanged between the two forms at rates a and b, i.e.

$$\mathrm{A} \rightleftharpoons_b^a \mathrm{B} \quad ,$$

where we have folded the probabilities to encounter the enzymes in the reaction rates.

   (a) Write down the Master equation that governs the evolution of the probabilities $p(N_A = n, N_B = N - n, t)$.

   (b) Assuming that initially all molecules are in state A, i.e. $p(n, t = 0) = \delta_{n,N}$, find $p(n, t)$ at all times. You may find it easier to guess the solution, but should then check that it satisfies the equations obtained before.

*****

5.2 Continuum limit

1. *The Moran process,* named after Patrick Moran, is a simple method for modeling a population of constant size. At each step one individual from the population is randomly selected for duplication/reproduction, and another for elimination/death, thus maintaining a fixed size.

   (a) For a haploid population of size N, with one locus and two alleles A_1 and A_2, compute the changes $\langle \Delta N_1 \rangle$ and $\langle \Delta N_1^2 \rangle$ in number of individuals with allele N_1 after one step.

   (b) Construct the drift-diffusion equation for this model, assuming that $N/2$ steps of the Moran process correspond to one generation time.

   (c) How would you modify the process to implement differing fitness values for the two alleles?

*****

2. *Treadmilling Actin:* Actin filaments are long, asymmetric, polymers involved in a variety of cellular functions. In some cases the filaments are in a dynamic state in which monomers are removed from one end and added to the other. (The two ends are called minus and plus respectively, and this process is known as treadmilling.)

(a) Assume that monomers are added to the plus-end at rate a, and removed from the minus end at rate b. Write down the equations governing the rate of change of the probabilities $\{p(\ell, t)\}$, for finding a filament of length ℓ at time t. Note that $\ell = 1,\ 2,\ , 3, \cdots$, and that the equation of $p(1, t)$ is different from the rest.

(b) It is possible to have a dynamic steady state with probabilities $p^*(\ell)$ that do not change with time. Find the (properly normalized) distribution $p^*(\ell)$ in such a case.

(c) What is the condition for the existence of a time independent steady state, and the mean length of the filament in such a case?

(d) For $a > b$, what is the average length of a filament at time t, starting from individual monomers at time $t = 0$? Calculate the fluctuations (variance) in length, and write down an approximate probability distribution $p(\ell, t)$ with the correct first and second moment.

*****

3. *Activation/deactivation reaction:* Many molecules in biology can be made active or inactive through the addition of a phosphate group. The enzyme that adds the phosphate group is usually termed a kinase, while a phosphatase removes this group. Let us consider a case where a finite number N of such molecules within a cell can be exchanged between the two forms at rates a and b, i.e.

$$\mathrm{A} \rightleftharpoons_b^a \mathrm{B} \quad ,$$

where we have folded the probabilities to encounter the enzymes in the reaction rates.

(a) Write down the Master equation that governs the evolution of the probabilities $p(N_A = n, N_B = N - n, t)$.

(b) Assuming that initially all molecules are in state A, i.e. $p(n, t = 0) = \delta_{n,N}$, find $p(n, t)$ at all times. You may find it easier to guess the solution, but should then check that it satisfies the equations obtained before.

*****

5.3 Brownian motion

1. *Foraging:* Typical foraging behavior consists of a random search for food, followed by a quick return to the nest. For this problem, assume that the nest is at the origin, and the search consists of a random walk *in two dimensions* around the nest.

(a) Modeling the search as a random walk with diffusion constant D, what is the probability density for the searcher to be a distance r from the nest, at a time t after leaving the nest?

(b) Assume that durations of search segments are exponentially distributed, i.e. with probability $p(t) \propto e^{-t/\tau}$. Further assume that the times spent in returning to the nest, and stay at nest between searches, are negligible compared to search times. After times much longer than τ, what is the probability to find the searcher at a distance r from the nest. Use saddle-point integration to find the asymptotic probability for large r.

*****

2. *Chemotaxis:* The motion of *E. Coli* in a solution of nutrients consists of an *alternating* sequence of *runs* and *tumbles*. During a run the bacterium proceeds along a straight line for a time t_r with a velocity v. It then tumbles for a time t_t, after which it randomly chooses a new direction $\hat{n}$ to run along. Let us assume that the times t_r and t_t are independently selected from probability distributions

$$p_r(t_r) = \frac{4t_r}{\tau_r^2} \exp\left(-\frac{2t_r}{\tau_r}\right) \quad , \quad \text{and} \quad p_t(t_t) = \frac{4t_t}{\tau_t^2} \exp\left(-\frac{2t_t}{\tau_t}\right) \quad .$$

(a) Assuming values of $\tau_r \approx 2$s, $\tau_t \approx 0.2$s, and $v \approx 30\mu\text{ms}^{-1}$, calculate the diffusion coefficient D for the bacterium at long times.

(b) In the presence of a chemical gradient the run times become orientation dependent, and are longer when moving in a favorable direction. For preferred motion up the z axis, let us assume that the average run time depends on its orientation $\hat{n}$ according to $\tau_r(\hat{n}) = \tau_0 + g\hat{n} \cdot \hat{z}$. Calculate the average drift velocity at long times.

*****