Computational Biology (FFR110/FIM740)

Problem set 3
for March 11, 23.59, 2022

Problem 1, Stochastic dynamics in large but finite populations [9p] In this task you will contrast the deterministic disease dynamics (valid in the limit of infinite populations) to the corresponding disease dynamics in large but finite populations. The analysis will be made for the SIS model introduced in the lecture notes. In what follows, S denotes the population size of susceptibles, and I denotes the population size of infectives. The deterministic dynamics for the SIS model are given by

$$\frac{\mathrm{d}I}{\mathrm{d}t} = \frac{\alpha}{S+I} SI - \beta I, \tag{1a}$$

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -\frac{\alpha}{S+I} SI + \beta I. \tag{1b}$$

Here, the parameters α and β are positive constants.

a) **Deterministic model** [**1.5p**] For which parameter values will the infection sustain ad infinitum in the model (1)? Support your result by finding the steady states of the system (1) and performing linear stability analysis.

b) **Stochastic model for finite population size** [**1.5p**] Following the steps in the lecture notes, contrast the deterministic model (1) to the corresponding stochastic model. Assume that the total population consists of N individuals (where N is constant in time). An individual in the population is either infected or susceptible. Assume that in a short time interval, the number of infected individuals, n, changes by $+1$ or -1 because of

$$\text{new infection}: \; n \to n+1 \text{ at rate } b_n = \alpha\, n \left(1 - \frac{n}{N}\right),$$
$$\text{recovery}: \; n \to n-1 \text{ at rate } d_n = \beta\, n\,.$$

Write down the Master equation for the probability $\rho_n(t)$ to observe n infected individuals at time t in a finite population consisting of N individuals. Derive Eq. (1a) in the limit of $N \to \infty$. Assume that the parameter values are such that the infection under the deterministic model persists ad infinitum. Does this result apply to the stochastic model as well? Discuss in which way the stochastic model differs from the deterministic one. Relate your discussion to the *quasi-steady state*.

The stochastic model described above can in principle be simulated by performing a sequence of time steps of fixed small size $\mathrm{d}t$. At each time step, the number of infected n either increases by 1 with probability $b_n \mathrm{d}t$, decreases by 1 with probability $d_n \mathrm{d}t$, or remains the same with probability $1 - (b_n + d_n)\mathrm{d}t$. This is however not efficient because most time steps the population does not change.

c) **Towards efficient simulation of the stochastic model** [**3p**] Instead of taking many small time steps one can, in order to speed up the simulations, use the probability distributions $P(t_b)$ and $P(t_d)$ of the times until the next infection (t_b) or recovery (t_d). To find these probability distributions, start from n infectives and assume that the total infection and recovery rates b_n and d_n take three sets of values: $\{b_n = 0.1, d_n = 0.2\}$, $\{b_n = 1, d_n = 2\}$, and $\{b_n = 10, d_n = 5\}$. For each set $\{b_n, d_n\}$ do the following

(i) Take small time steps dt as described above and record the first time t_b or t_d where an infection or recovery event takes place.

(ii) Reset the population back to n and repeat step (i) many times in order to get good statistics for t_b and t_d for the rates $\{b_n, d_n\}$.

(iii) Plot the logarithm of the resulting distributions of times between two subsequent infections $\log P(t_b)$ against t_b, and between two subsequent recoveries $\log P(t_d)$ against t_d.

(iv) Verify that the distributions decay exponentially with exponent $-\lambda t$, where $\lambda = b_n$ for infections and $\lambda = d_n$ for recoveries.

Using the obtained distributions, what is the average time until an infection or recovery? Explain why.

Gillespie algorithm: The above procedure gives an efficient algorithm for simulating the model: starting from n infectives, draw two independent random numbers using the distributions in subtask c). The minimum of the two random numbers gives the time and type of the next event.

d) **Population distribution at different times in the stochastic model** [**3p**] Use the Gillespie algorithm described above to find how the distribution $P(n_t)$ of infectives in the stochastic model changes as a function of time. Start by choosing the parameters α, β so that the deterministic dynamics has a positive steady state I^*, and initialize the population with $n_0 = I^*$. Find the distribution $P(n_t)$ by averaging over many realisations of the process starting from n_0. As discussed in the lecture notes, the lifetime of the quasi-steady state (T_{ext}) depends sensitively on α, β. Choose three different times $t_1 < T_{\text{ext}}, t_2 \sim T_{\text{ext}}, t_3 > T_{\text{ext}}$. Plot $-\log P(n_t)$ against n_t for $t = t_1, t_2, t_3$. Describe the observations: in particular how does $P(n_{t_1})$ compare to a Gaussian distribution? In the lecture notes an approximate expression for the quasi-steady state distribution is derived from the Master equation. Compare this expression to the results of your simulations. Does the theory correctly predict deviations from a Gaussian? Does it correctly describe the far left tail of the distribution (corresponding to few infected individuals)?

Problem 2, Population genetics [8p] Following the lecture notes, denote by $P(S_n = j)$ the probability that there are j Single Nucleotide Polymorphisms (SNPs) in a sample of size n. Consider the infinite-size model, i.e. a population where the size N approaches infinity and the mutation rate per individual and generation μ approaches zero, such that $2N\mu = \theta = \text{const.}$.

a) [**4p**] Derive an expression for $P(S_n = 0)$, i.e. the probability to not have any SNPs in a sample of size n.

b) [**4p**] Derive the distribution of the number of SNPs in a sample of size $n = 2$. You should find

$$P(S_2 = j) = \frac{1}{1+\theta}\left(\frac{\theta}{1+\theta}\right)^j .$$