CHALMERS, GÖTEBORGS UNIVERSITET

EXAM for
COMPUTATIONAL BIOLOGY

COURSE CODES: **FFR 110, FIM740GU, PhD**

Time:	June 10, 2022, at $08^{30} - 12^{30}$
Place:	Johanneberg
Teachers:	Kristian Gustafsson, 070-050 2211 (mobile), visits once around 10^{00}
Allowed material:	Mathematics Handbook for Science and Engineering
Not allowed:	any other written material, calculator

Maximum score on this exam: 50 points (need 20 points to pass).
Maximum score for homework problems: 50 points (need 20 points to pass).
CTH ≥40 grade 3; ≥60 grade 4; ≥80 grade 5,
GU ≥40 grade G; ≥70 grade VG.

---

1. Short questions [12 points] For each of the following questions give a concise answer within a few lines per question.

a) For what population sizes do you expect the Malthus growth model to be most suitable? Why is the form of the Malthus growth model often valid for these population sizes, also when additional factors (crowding, competition for resources, etc) are taken into account?

Solution
The Malthus model is suitable for population sizes much smaller than all other population scales in the system, but it should not be too small, such that it matters if one individual is added|removed (then a stochastic model is needed).

Additional factors such as crowding, competition for resources, etc due to interaction are non-linear. For population sizes much smaller than the populations scales associated with the additional factors, the linear model dominates.

b) Explain how to draw a cobweb plot for a map $N_{\tau+1} = F(N_\tau)$ and explain why this is the right procedure to obtain the solution orbit N_1, N_2, . . . , starting from N_0.

Solution
Lecture notes 2.3

c) Explain the principle of competitive exclusion and give an example of a possible application of it.

Solution

If two species are competing for the same resource, the disadvantageous species must either go extinct or adjust its behavior to a different niche. Possible applications are pest control or to treat infections by introducing a competing species to the pest or infectious bacteria.

d) Assume that ammonia (NH_3) is spilled on the floor two meters from where you stand. Estimate the time scale until a significant amount of NH_3 molecules reach your nose, if they are solely transported by three-dimensional molecular diffusion through air with diffusion coefficient $2 \cdot 10^{-5}\,\mathrm{m}^2/\mathrm{s}$. Comment on the result, is it realistic?

Solution

The displacement due to molecular diffusion is $\langle \Delta \boldsymbol{x}^2 \rangle \sim 6Dt$ in 3D. The distance from the source to your head is approximately $|\Delta \boldsymbol{x}| \approx \sqrt{8}$ m, leading to the time scale $t \sim |\Delta \boldsymbol{x}|^2/(6D) \approx \frac{2}{3} \cdot 10^5\,\mathrm{s} \approx 20\,\mathrm{h}$ (any estimate $t \sim 10^5\,\mathrm{s}$ was accepted).

This time scale is not realistic because the dynamics of the molecules are dominated by turbulent transport/diffusion. Any small wind puff will take then to you in order of seconds.

e) Fisher's equation in one spatial dimension is given by

$$\frac{\partial}{\partial t} n(x,t) = r n(x,t) \left(1 - \frac{n(x,t)}{K}\right) + D \frac{\partial^2}{\partial x^2} n(x,t) \,.$$

Explain how a travelling wave solution to Fisher's equation can travel much faster into an initially empty region than the solution to the diffusion equation (obtained by letting $r = 0$ in Fisher's equation).

Solution

Diffusion is efficient to smear out sharp gradients, but does not provide efficient transport into the empty region, the displacement in time scale t is of the order $\sqrt{2Dt}$ due to the law of diffusion. The form of travelling wave solutions to Fisher;s equation gives a displacement linear in time, much faster than the diffusive transport. The reason is that at a given position x, the per capita growth rate due to production, $\dot{n}/n \sim r(1 - n/K)$, is largest for small population sizes, creating large gradients to closeby empty regions. In combination with diffusion which effectively smear these gradients out, the population moves to initially empty regions, resulting in a stable wave front, a travelling wave.

f) The Kuramoto model consists of N oscillators with phases θ_i and natural angular frequencies ω_i

$$\dot{\theta}_i = \omega_i + \frac{K}{N} \sum_{j=1}^{N} \sin(\theta_j - \theta_i) \,.$$

Explain how a fraction of these oscillators can be synchronized despite having different angular frequencies.

Solution
Bernhard's lecture notes 7.6

g) Consider the stochastic SIS model for disease spreading. Let λ_n denote the rate of infection and μ_n the rate of recovery for n infectives. Write down a Master equation for the distribution $\rho_n(t)$ for observing n infectives at time t. Assume that each infection event results in at most one individual becoming infected.

Solution
The possible transitions to a population of size n are $n-1 \to n$ due to infection or $n+1 \to n$ due to recovery. The transitions from a population of size n are $n \to n+1$ due to infection or $n \to n-1$ due to recovery. The change in probability in a small time step δt is therefore

$$\rho_n(t) = \rho_n(t-\delta t) + \delta t[\lambda_{n-1}\rho_{n-1}(t-\delta t) + \mu_{n+1}\rho_{n+1}(t-\delta t) - (\lambda_n + \mu_n)\rho_n(t-\delta t)]\,.$$

The corresponding Master equation is obtained by rearranging the terms and taking the limit $\delta t \to 0$

$$\frac{\mathrm{d}\rho_n}{\mathrm{d}t} = \lambda_{n-1}\rho_{n-1} + \mu_{n+1}\rho_{n+1} - (\lambda_n + \mu_n)\rho_n\,.$$

h) This problem is not part of the course material this year.
Discuss and contrast the effects of measurement noise and dynamical noise on a time series which is generated by linear (Malthus) decay (negative growth rate).

Solution
Lecture notes 14.2

2. Delay model of houseflies [12 points] Consider the following model for the growth of a population of house flies of size N

$$\dot{N} = -dN(t) + bN(t-T)(k - bsN(t-T)) \,. \tag{1}$$

Here d is the per capita death rate and b is the per capita rate of laying eggs. Furthermore, k, s and T are positive parameters.

a) Give a plausible explanation for the form of the system (1). What is the significance of the parameters k, s and T?

Solution
The first term is a linear Malthus death term proportional to the population size, with per capita death rate d. The second term is a birth term that takes into account the time T spent as an egg, i.e. at time t the $bN(t-T)$ eggs laid at time $t-T$ hatches to contribute to the population. $k - bsN(t-T)$ is egg-to-adult survival ratio: k is maximal fraction that survives and s is a reduction in survival due to each additional produced egg.

b) Find all steady states ($N(t) = \text{const.}$) of the Eq. (1). Find a condition for one positive steady state to exist and use your explanation of the parameters in subtask a) to argue why the form of the condition is reasonable.

Solution
There are two steady states when $N^* = 0$ and when $N^* = (bk-d)/(b^2 s)$. The second steady state is positive if $bk > d$. This makes sense because then the egg-laying rate times the fraction of eggs surviving to adulthood is larger than the death rate (the second-order term $-b^2 sN(t-T)^2$ does not matter close to the transition of N from negative to positive).

Let $b = 4$ and $k = d = s = 1$ in the subtasks below.

c) Show that close to the positive steady state, the dynamics of a small perturbation η can be approximated by the form

$$\frac{\mathrm{d}\eta}{\mathrm{d}t} \approx C_1\eta(t) + C_2\eta(t-T) \,. \tag{2}$$

What are the expressions for the coefficients C_1 and C_2?

Solution
The positive steady state is $N^* = (bk-d)/(b^2 s)$. Write $N = (bk - d)/(b^2 s) + \eta$ and expand the dynamics (1) to first order in η:

$$\begin{aligned}
\frac{\mathrm{d}\eta}{\mathrm{d}t} &= -d\left[\frac{bk-d}{b^2 s} + \eta(t)\right] + b\left[\frac{bk-d}{b^2 s} + \eta(t-T)\right]\left(k - bs\left[\frac{bk-d}{b^2 s} + \eta(t-T)\right]\right) \\
&= -d\left[\frac{bk-d}{b^2 s} + \eta(t)\right] + \frac{bk-d}{bs}\left(k - bs\left[\frac{bk-d}{b^2 s} + \eta(t-T)\right]\right) + b\eta(t-T)\left(k - \frac{bk-d}{b}\right) \\
&= (bk-d)\frac{1}{b^2 s}[-d - (bk-d) + bk] - d\eta(t) + [-(bk-d) + bk - (bk-d)]\eta(t-T) \\
&= -d\eta(t) + [2d - bk]\eta(t-T) \,,
\end{aligned}$$

i.e. on the form (2) with $C_1 = -d$ and $C_2 = 2d - bk$. Using the specified parameter values, we have $C_1 = -1$ and $C_2 = -2$.

d) Use the ansatz $\eta(t) = Ae^{\lambda t}$ in Eq. (2) to derive an equation for λ.

Solution
Inserting the ansatz gives

$$\lambda A e^{\lambda t} = C_1 A e^{\lambda t} + C_2 A e^{\lambda(t-T)} .$$
$$\Rightarrow \lambda = C_1 + C_2 e^{-\lambda T} = 1 - 2e^{-\lambda T} .$$

e) Find a condition on T for which the steady state population of house-flies is unstable. What kind of long-term behavior do you expect?

Solution
Write $\lambda = \lambda' + \mathrm{i}\lambda''$. The system undergoes a bifurcation to unstable when λ with maximal real part λ' passes zero. Expressing the equations in subtask d) with $\lambda = \mathrm{i}\lambda''$ (so that $\lambda' = 0$) gives

$$\mathrm{i}\lambda'' = -1 - 2e^{-\mathrm{i}\lambda'' T} = -1 - 2[\cos(-\lambda'' T) + \mathrm{i}\sin(-\lambda'' T)]$$
$$\Rightarrow \begin{cases} 0 = -1 - 2\cos(\lambda'' T) \\ \lambda'' = 2\sin(\lambda'' T) \end{cases}$$

The first equation gives

$$\lambda'' T = \mathrm{acos}\left(-\tfrac{1}{2}\right) + 2\pi n = \tfrac{\pi}{3} + 2\pi n ,$$

where n is a positive integer. Inserting $\lambda'' T$ into the second equation gives

$$\lambda'' = \sqrt{3}$$

Thus, the eigenvalue with maximal real part passes zero when T becomes larger than $T_{\mathrm{c}} = \frac{1}{\sqrt{3}}\frac{\pi}{3}$ (using $n = 0$ in the condition for $\lambda'' T$ above).

Since λ'' is non-zero, we expect growing oscillations for solutions starting close to the unstable steady state. Since the system does not have stable attractors we expect the dynamics to approach a limit cycle (could in principle also be a chaotic dynamics).

3. Effect of spruce budworms on a forest [10 points] In the lectures the spruce budworm model was introduced for the fast growth of budworms in a forest under predation. To instead make a model for the effect of budworms feeding on the foliage (i.e. leaves and branches) of the forest, assume a constant budworm population of size B at one of its stable steady states: either outbreak (large B) or refuge (small B). Let $S(t)$ denote the average surface area of the branches of a tree and let $H(t)$ denote a general measure of the forest's health. A growth model for S and H is

$$\begin{aligned}\dot{S} &= r_S S\left(1 - \frac{S}{K_S}\frac{K_H}{H}\right)\\ \dot{H} &= r_H H\left(1 - \frac{H}{K_H}\right) - P\frac{B}{S}\end{aligned} \tag{3}$$

where r_S, r_H, K_S, K_H and P are positive parameters.

a) Interpret the terms in Eq. (3) biologically.

Solution
S is governed by a logistic growth because the maximal surface area is bounded, with upper limit K_S corresponding to the natural size in absence of budworms. The factor K_H/K models decline in growth due to stress since the surface area may decline through death of branches or whole trees.

Also H is governed by a logistic growth with a maximal value K_H in absence of budworms. The second term in $\dot{H}$ models the stress on the trees from budworms eating the foliage. It is proportional to B/S because a large budworm population consumes more, but bigger trees are less affected.

b) Introduce dimensionless units to write the system (3) in as few dimensionless parameters as possible.

Solution
Let $S = s_0 s$, $H = h_0 h$ and $t = t_0\tau$ to get

$$\begin{aligned}\frac{\mathrm{d}s}{\mathrm{d}\tau} &= \frac{t_0}{s_0} r_S s_0 s\left(1 - \frac{s_0 s}{K_S}\frac{K_H}{h_0 h}\right)\\ \frac{\mathrm{d}h}{\mathrm{d}\tau} &= \frac{t_0}{h_0}\left[r_H h_0 h\left(1 - \frac{h_0 h}{K_H}\right) - P\frac{B}{s_0 s}\right]\end{aligned}$$

Choose for example $h_0 = K_H$, $s_0 = K_S$, $t_0 = 1/r_S$ to get

$$\begin{aligned}\frac{\mathrm{d}s}{\mathrm{d}\tau} &= s\left(1 - \frac{s}{h}\right)\\ \frac{\mathrm{d}h}{\mathrm{d}\tau} &= \rho h\,(1-h) - \alpha\frac{1}{s}\end{aligned}$$

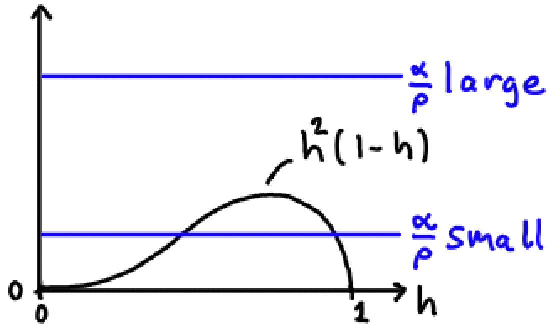
with $\rho = r_H/r_S$ and $\alpha = PB/(r_S K_S K_H)$.

c) Show, for example geometrically, that the system has two positive fixed points if B is small and no positive ones if B is large.

Solution
Solving $\dot{s} = 0$ gives the positive solution $s = h$ if $h > 0$, which gives the following condition in $\dot{h} = 0$

$$0 = \rho h\,(1-h) - \alpha\frac{1}{h} \;\Rightarrow\; h^2\,(1-h) = \frac{\alpha}{\rho}$$

Sketching these curves for small α/ρ (small B) and large α/ρ (large B) show the asked number of positive fixed points:



d) Discuss how the spruce budworm population affects the forest according to the model. Discuss whether the results are realistic.

Solution

According to the model, for the case of small B, there are two steady states, out of which one is stable (the one that approaches $h = 1$ as $\rho \to 0$). The effect of the budworm is therefore to reduce the surface area s and health h of the forest. For the case of large B, the forest does not have any stable fixed points, meaning that it will decrease in size and eventually pass $s = 0$ to die out.

However, as the environment changes, the spruce budworm equilibrium population will slowly change. In particular we expect its carrying capacity to become smaller as they reduce the health of the forest. This may cause a jump to a lower population (refuge equilibrium, small B), where the forest can recover and approach its stable fixed point. The prediction that the forest dies out therefore seems unrealistic.

4. Diffusion driven instability [8 points] Consider the following reaction-diffusion equation system in one spatial dimension for a version of the SI model of susceptibles $S(x,t)$ and infectives $I(x,t)$

$$\begin{aligned}\frac{\partial S}{\partial t} &= b - d_S S - \beta SI + D_S \frac{\partial^2 S}{\partial x^2} \\ \frac{\partial I}{\partial t} &= \beta SI - d_I I + D_I \frac{\partial^2 I}{\partial x^2}\end{aligned}, \tag{4}$$

where b, d_S, d_I, β, D_S and D_I are positive parameters. For simplicity you can let $d_S = d_I = \beta = D_S = 1$ throughout this problem (corresponding to a suitable dedimensionalisation that you do not need to derive).

a) Show that if $b > d_I d_S/\beta = 1$, the system (4) has a positive homogeneous steady state which is stable.

Solution

The positive homogeneous steady state of Eq. (4) is obtained from the equation system

$$0 = b - d_S S^* - \beta S^* I^* \tag{5}$$

$$0 = \beta S^* I^* - d_I I^* \tag{6}$$

with positive solution

$$(S^*, I^*) = \left(\frac{d_I}{\beta}, \frac{b}{d_I} - \frac{d_S}{\beta}\right).$$

Evaluate $\mathbb{J}$ at the steady state

$$\mathbb{J}(S^*, I^*) = \begin{pmatrix} -d_S - \beta I^* & -\beta S^* \\ \beta I^* & \beta S^* - d_I \end{pmatrix} = \begin{pmatrix} -\beta \frac{b}{d_I} & -d_I \\ \beta \frac{b}{d_I} - d_S & 0 \end{pmatrix}$$

The fixed point is stable if the trace is negative and the determinant is positive. For the current system, the trace is always negative and the determinant becomes

$$\det(\mathbb{J}(S^*, I^*)) = -(-d_I)(\beta \tfrac{b}{d_I} - d_S) = \beta b - d_I d_S,$$

which is positive when the fixed point is positive, i.e. when $b > d_I d_S/\beta$

In the lectures, we showed that small perturbations $\delta u(x,t) \equiv u(x,t) - u^*$ and $\delta v(x,t) \equiv v(x,t) - v^*$ from a stable homogeneous steady state (u^*, v^*) for the concentrations u and v following a generic reaction-diffusion system, can be decomposed using the ansatz

$$\begin{pmatrix}\delta u \\ \delta v\end{pmatrix} = e^{\lambda t + ikx} \begin{pmatrix}\delta u_0 \\ \delta v_0\end{pmatrix}$$

yielding the following equation:

$$0 = [\lambda - \mathbb{K}] \begin{pmatrix}\delta u_0 \\ \delta v_0\end{pmatrix}, \text{ where } \mathbb{K} = \mathbb{J}(u^*, v^*) - k^2 \begin{pmatrix} D_u & 0 \\ 0 & D_v \end{pmatrix}.$$

Here D_u and D_v are the diffusion coefficients of u and v respectively.

b) Does the system (4) have a diffusion-driven instability for any parameter values $b_c(k)$ for which space-dependent perturbations with a single wave number k becomes unstable?

Solution

Calculate $\det \mathbb{K}$ of the matrix $\mathbb{K}$ for the system (4)

$$\mathbb{K} = \mathbb{J}(u^*, v^*) - k^2 \begin{pmatrix} D_S & 0 \\ 0 & D_I \end{pmatrix} = \begin{pmatrix} -\beta \frac{b}{d_I} - k^2 D_S & -d_I \\ \beta \frac{b}{d_I} - d_S & -k^2 D_I \end{pmatrix}$$

$$\det \mathbb{K} = k^4 D_I D_S + \beta \frac{b}{d_I} k^2 D_I + \beta b - d_S d_I$$

Since $\mathrm{tr}\mathbb{K} < 0$ (we add something negative to $\mathrm{tr}\mathbb{J}$ which is negative), the real part of the maximal eigenvalue $\lambda(k)$ passes zero when $\det \mathbb{K}$ goes from positive to negative. This happens when $\det \mathbb{K} = 0$, i.e. when

$$b = b_c(k) = \frac{d_S d_I - k^4 D_I D_S}{\beta[1 + \frac{1}{d_I} k^2 D_I]} = \frac{1 - k^4 D_I}{1 + k^2 D_I}\,.$$

But since $b_c(k) \leq 1$ in this expression, the system is unstable for all non-zero wave numbers. The system therefore does not have a diffusion driven instability.

c) Assume a system with a diffusion-driven instability that is unstable to perturbations with wave numbers in a range $k_- \leq k \leq k_+$. Initially, the concentration is obtained by a small random perturbation from the homogeneous stable steady state. The domain is one-dimensional with length L and no-flux boundary conditions. What is the minimal length L required to have a diffusion-driven instability in this system?

Solution

Superposition of the spatial part of the solution e^{ikx} gives trigonometric spatial solutions

$$R_k(x) = A_k \cos(kx) + B_k \sin(kx)\,.$$

Setting the coordinate system such that the domain is $0 \leq x \leq L$ and using the no flux boundary conditions $\frac{\partial}{\partial x} R_k(0) = \frac{\partial}{\partial x} R_k(L) = 0$ gives the solution $R_k(x) = A_k \cos(kx)$ with $k = 0, \pi/L, 2\pi/L, \ldots$.

To have a diffusion-driven instability, any of the non-zero modes must lie in the interval $k_- \leq k \leq k_+$. The smallest value of L for which this can happen is $\pi/L = k_+$, i.e. $L = \pi/k_+$.

5. Coalescent process [8 points]

a) The coalescent process is a model for neutral sample genealogies, consistent with the Fisher-Wright model. Describe the coalescent process in its simplest form, for a sample of size n from a large population, $N \gg n$, and derive the following distribution of the time T_j to the next coalescent event, given that there are j ancestral lines:

$$P(T_j) = \lambda_j \exp(-\lambda_j T_j) \quad \text{with} \quad \lambda_j = \frac{1}{N}\binom{j}{2}\,. \tag{7}$$

Solution

See pages 14-16 in Bernhard's lecture notes.

b) Tajima suggested a test for selection by comparing whether a genetic mosaic is compatible with a neutral sample genealogy, or not. The test is based upon two different estimators for the mutation parameter $\theta = 2N\mu$ that are derived from the following equations

$$\langle S_n \rangle = \theta \sum_{j=1}^{n-1} \frac{1}{j} \quad \text{and} \quad \left\langle \frac{1}{\binom{n}{2}} \sum_{i<j} \Delta_{ij} \right\rangle = \theta\,. \tag{8}$$

Here $\langle S_n \rangle$ is the average number of single-nucleotide polymorphisms (SNPs) in the sample of size n, and Δ_{ij} is the number of SNPs between two individuals in the sample, i and j. Derive the two relations in Eq. (8) using the coalescent process.

Hint: For the first relation, use that the number S_n of SNPs in a given genealogy for n individuals is Poisson distributed,

$$P(S_n = j) = \frac{(\mu T_{\text{tot}}^{(n)})^j}{j!} \exp(-\mu T_{\text{tot}}^{(n)})\,,$$

where $T_{\text{tot}}^{(n)}$ is the total branch length of the genealogy. Compute the expected number of SNPs, and then average over genealogies. For the second relation, compute $\langle \Delta_{ij} \rangle$ by considering $n = 2$.

Solution

Given in Exam20220317