

CHALMERS, GÖTEBORGS UNIVERSITET

EXAM for COMPUTATIONAL BIOLOGY

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

Time:	March 17, 2022, at 14 ⁰⁰ – 18 ⁰⁰
Place:	Johanneberg
Teachers:	Kristian Gustafsson, 070-050 2211 (mobile), visits once around 15 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) Give two examples of biological systems, one where a time delay model is a suitable model and one where a discrete growth model is suitable.

Solution

Delay example: house flies (delay due to time spent as egg). Discrete example: Synchronised growth of cells.

- b) In the Lotka-Volterra model for prey and predator populations N and P

$$\dot{N} = N(a - bP)$$

$$\dot{P} = P(cN - d)$$

predators have limitless appetite. Explain what this means and how one can model a limited appetite.

Solution

With limitless appetite interactions (of rate $\propto NP$) have the same outcome independent of the population sizes and result in terms proportional to NP in the dynamics. Limited appetite can be modelled by multiplying the encounter rate by a function that decays with N .

- c) Write down a growth model for the concentrations in the spontaneous reaction from substrate S to product P with rate constant k , $S \xrightarrow{k} P$.

Solution

Let $s = [S]$ and $p = [P]$. The growth model becomes

$$\begin{aligned}\dot{s} &= -ks \\ \dot{p} &= ks\end{aligned}$$

- d) Give examples of three distinct stochastic mechanisms or effects that we have described using stochastic models in the course.

Solution

For example: fluctuations of birth rate in growth equations, random spatial motion, and stochastic gene histories.

- e) Consider a striped snake skin pattern created by diffusion driven instability of morphogen concentrations. Do you expect the stripes to run from head to tail or like rings around the body? Motivate your answer.

Solution

The stripes should be rings around the body. In a domain with very small aspect ratio, many wave lengths are allowed along the long axis but fewer wave lengths along the shorter axis. This is not consistent with stripes from head to tail, which correspond to a single long wave length along the long axis (from head to tail) and a much smaller wave length around the body (separating the different stripes).

- f) Explain the mechanism for wave patterns in unstirred concentrations undergoing the Belousov-Zhabotinsky reaction in a vertical tube.

Solution

See Lecture notes 13.1

- g) Explain what is meant by the phase of an oscillator. What does it mean that several oscillators are phase locked in the Kuramoto model?

Solution

The phase of an oscillator is 2π times the fraction of the period time of the oscillator. In Kuramoto's model phase locked oscillators rotate rigidly with a shared constant frequency ('phase locked' means that the relative phases between oscillators are constant).

- h) **This problem is not part of the course material this year.**

The autocorrelation function of a time series $x_0, x_1 \dots x_{N-1}$ is defined by

$$C(k) = \frac{\sum_{n=0}^{N-k-1} (x_n - \langle x \rangle)(x_{n+k} - \langle x \rangle)}{\sum_{n=0}^{N-1} (x_n - \langle x \rangle)^2}.$$

Explain how $C(k)$ can be used to detect periodic oscillations in data.

Solution

If the data has oscillations with period T , then x_n and x_{n+T} are correlated and the autocorrelation function has peaks at intervals of approximate length T (approximate because T is not an integer in general).

2. Continuous VS discrete fishery management models [8 points]

Consider the following continuous growth model for how a fish population of $N(t)$ individuals at time t is affected by fishing

$$\dot{N} = rN \left(1 - \frac{N}{K} \right) - EN. \quad (1)$$

Here E is a fishing rate with $0 \leq E < r$. r and K are positive parameters denoting the growth rate and carrying capacity in absence of fishing.

- a) By rearranging the terms, show that Eq. (1) can be rewritten as a continuous logistic growth model, $\dot{N} = r_E N (1 - N/K_E)$, with effective growth rate r_E and carrying capacity K_E . Express r_E and K_E in terms of the original parameters r , K and E . Verify that $r_E > 0$.

Solution

Rewrite Eq. (1)

$$\dot{N} = (r - E)N - r \frac{N^2}{K} = \underbrace{(r - E)}_{r_E} N \left(1 - \underbrace{\frac{r}{(r - E)K}}_{1/K_E} N \right)$$

where $r_E = r - E$ and $K_E = K(r - E)/r$. Since $r > E$, we have $r_E > 0$.

- b) Using the result in subtask a), write down the stable steady state N^* of Eq. (1). Define the yield as the removal rate of the population due to fishing. Which fishing rate E gives the maximal yield in the steady state of the model (1)? What is the corresponding maximal yield?

Solution

Since $r_E > 0$, the logistic growth model has a stable steady state at $N^* = K_E$. The yield at this steady state is $Y(E) = EN^* = EK_E = KE - KE^2/r$. We obtain the maximal yield when $Y'(E) = 0$, i.e. when $E = r/2$. The corresponding value is $Y_{\max} = Y(E = r/2) = Kr/4$.

If generations of the fish population do not overlap, it may be better to consider the following discrete growth model for fishing

$$N_{\tau+1} = N_{\tau} + rN_{\tau} \left(1 - \frac{N_{\tau}}{K} \right) - EN_{\tau}. \quad (2)$$

Assume that the time units are scaled so that r and E takes the same values in Eqs. (1) and (2).

- c) Verify that the steady state N^* in subtask b) is a steady state also of Eq. (2). Contrast the range of stability of N^* in the two models.
Hint: To simplify the calculation, you can rewrite Eq. (2) in terms of the parameters r_E and K_E of subtask a).

Solution

In terms of the parameters in subtask a), we have (by simply adding N_τ to the rearrangement in subtask a))

$$N_{\tau+1} = (1 + r_E)N_\tau - \frac{r_E}{K_E}N_\tau^2.$$

Inserting $N^* = K_E$ gives

$$\text{LHS: } N_{\tau+1} = K_E$$

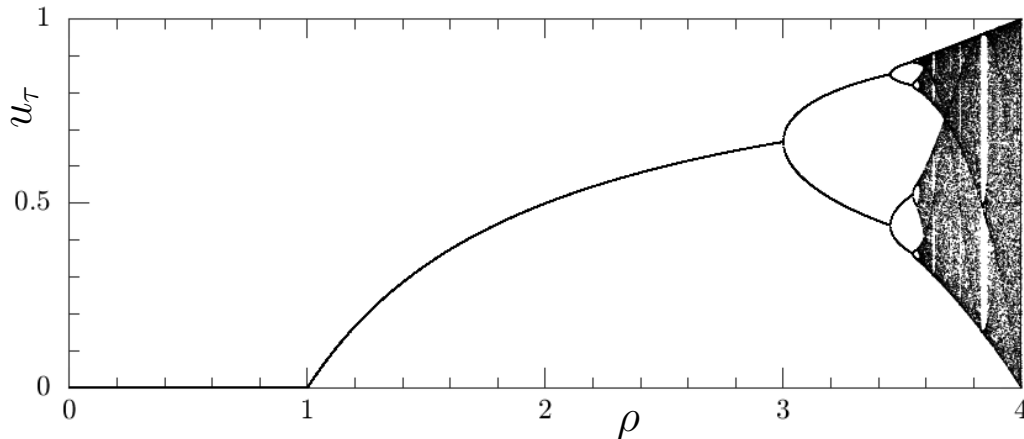
$$\text{RHS: } N_{\tau+1} = (1 + r_E)K_E - r_E K_E = K_E.$$

i.e. N^* is still a steady state. Determine its stability

$$\Lambda = \left. \frac{dN_{\tau+1}}{dN_\tau} \right|_{N_\tau=K_E} = 1 + r_E - 2\frac{r_E}{K_E}K_E = 1 - r_E,$$

i.e. stable for $0 < r_E < 2$. In contrast, the steady state of the continuous model is stable for all $r_E > 0$.

- d) The plot below shows the bifurcation diagram for the dimensionless logistic map $u_{\tau+1} = \rho u_\tau(1 - u_\tau)$ with population size u and parameter ρ .



Verify that the range of stability in subtask c) is consistent with this diagram. What happens outside the range of stability of the steady state N^* ? Discuss whether there are situations when fishing can have a beneficial effect on the population dynamics according to the model (2).

Solution

The parameter ρ in the logistic model corresponds to $1 + r_E$ in our model (the coefficient of the linear term). In the range $1 < \rho < 3$ (equivalent to $0 < r_E < 2$) the bifurcation diagram shows a stable steady state, consistent with the result in subtask b). For $r_E < 0$ the diagram shows that there is a transcritical bifurcation to a steady state

at zero population. For $r_E > 2$ the diagram shows that there is a period-doubling bifurcation (Λ passes through -1).

According to the model (and observations), due to their large reproductive rate fish may show large periodic or chaotic fluctuations between generations. Since fishing is essentially a lowering of the reproductive rate, fishing may have a stabilising effect, giving periodic solutions or a stable steady state in a system that would otherwise be chaotic.

3. Travelling waves [10 points] The following is a model for spread of a population with concentration $n(x, t)$ in a one-dimensional habitat:

$$\frac{\partial n}{\partial t} = rn \left(1 - \frac{n}{K}\right) \left(\frac{n}{A} - 1\right) + D \frac{\partial^2 n}{\partial x^2}. \quad (3)$$

Here, r , K , A , and D are positive constant parameters. Assume that $A < K$.

- a) Give an explanation of the first term $rn \left(1 - \frac{n}{K}\right) \left(\frac{n}{A} - 1\right)$ in Eq. (3). What does K and A signify?

Solution

Carrying capacity K and threshold population A for Allee effect (negative growth for small population sizes due to difficulty of finding mates, reduced cooperation, etc).

- b) Introduce dimensionless time τ , position ξ and population size u and rewrite the dynamics (3) in terms of these coordinates and a single dimensionless parameter of your choice.

Solution

Let $n = n_0 u$, $t = t_0 \tau$ and $x = x_0 \xi$ to get

$$\frac{\partial u}{\partial \tau} = \frac{t_0}{n_0} \frac{\partial n}{\partial t} = t_0 r u \left(1 - \frac{n_0 u}{K}\right) \left(\frac{n_0 u}{A} - 1\right) + \frac{t_0}{x_0^2} D \frac{\partial^2 u}{\partial \xi^2}$$

One example is to choose $t_0 = 1/r$, $n_0 = K$ and $x_0 = \sqrt{D/r}$, and define $\alpha = A/K$ ($0 < \alpha < 1$) to obtain

$$\frac{\partial u}{\partial \tau} = u(1 - u) \left(\frac{u}{\alpha} - 1\right) + \frac{\partial^2 u}{\partial \xi^2}.$$

- c) Assume that u only depends on ξ and τ through the combination $z = \xi - c\tau$ and convert the partial differential equation in subtask b) to a dynamical system for u and $v = \frac{du}{dz}$.

Hint: If you did not solve subtask b), you can instead use Eq. (3) with $r = K = D = 1$ and parameter A .

Solution

In terms of the wave variable z , the partial differential equation in subtask b) becomes (use $\frac{\partial}{\partial \xi} = \frac{\partial z}{\partial \xi} \frac{d}{dz} = \frac{d}{dz}$ and $\frac{\partial}{\partial \tau} = \frac{\partial z}{\partial \tau} \frac{d}{dz} = -c \frac{d}{dz}$)

$$-c \frac{du}{dz} = u(1 - u) \left(\frac{u}{\alpha} - 1\right) + \frac{d^2 u}{dz^2}.$$

Let $v = \frac{du}{dz}$ and rearrange the terms to obtain the dynamical system

$$\begin{aligned}\frac{du}{dz} &= v \\ \frac{dv}{dz} &= -cv - u(1-u) \left(\frac{u}{\alpha} - 1 \right)\end{aligned}$$

- d) Find all fixed points in the system derived in subtask c) and determine their stability if $c < 0$ (**OBS: Negative sign of c**).

Solution

The system has fixed points at $(u_1^*, v_1^*) = (0, 0)$, $(u_2^*, v_2^*) = (\alpha, 0)$, and $(u_3^*, v_3^*) = (1, 0)$. The stability matrix is

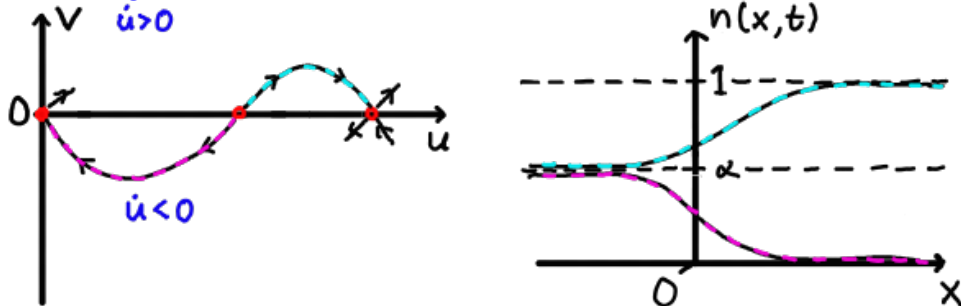
$$\begin{aligned}\mathbb{J} &= \begin{pmatrix} 0 & 1 \\ 1 - 2u + \frac{1}{\alpha}u(3u - 2) & -c \end{pmatrix} \\ \text{tr}\mathbb{J} &= -c > 0 \\ \det\mathbb{J} &= 2u - 1 + \frac{1}{\alpha}u(2 - 3u) \\ \det\mathbb{J}(u_1^*, v_1^*) &= -1 < 0 \\ \det\mathbb{J}(u_2^*, v_2^*) &= 1 - \alpha > 0 \\ \det\mathbb{J}(u_3^*, v_3^*) &= 1 - \frac{1}{\alpha} < 0\end{aligned}$$

(u_1^*, v_1^*) is a saddle point
 (u_2^*, v_2^*) is an unstable node/spiral
 (u_3^*, v_3^*) is a saddle point

- e) Sketch the shapes of possible travelling wave solutions with $c < 0$ (**OBS: Negative sign of c**) in the phase plane for u and v . Sketch the corresponding wave profiles $n(x, t)$ against t . In which direction does the wave travel?

Solution

Travelling waves correspond to trajectories connecting fixed points. The unstable fixed point (u_2^*, v_2^*) can connect to the stable manifolds of either of the both saddle points. Since $\dot{u} = v$, u increases in the upper half plane and decreases in the lower half plane, the heteroclinic trajectories and corresponding travelling wave solutions should look something as follows:



Both waves move to the left when $c < 0$ (the wave front at say $z =$

$\xi - c\tau = 0$ takes the coordinate $\xi = c\tau$, meaning its location on the ξ axis moves to smaller ξ when $c < 0$).

The case where (u_2^*, v_2^*) is an unstable spiral is similar, but the travelling waves will oscillate around $n = \alpha$. There is also the possibility that the two saddle points are connected, forming a travelling wave between $n = 0$ and $n = 1$, but this is more unlikely since the manifolds of the saddle points must match exactly.

4. Disease spreading of two variants of a disease [10 points] The SIRS model describes the spread of a non-lethal disease where immunity only lasts for a finite time:

$$\begin{aligned}\dot{s} &= -\beta si + \gamma r \\ \dot{i} &= \beta si - \alpha i \\ \dot{r} &= \alpha i - \gamma r.\end{aligned}\tag{4}$$

Here s , i and r denote the fractions of the total population that is susceptible (s), infective (i), or recovered (r). The parameters α , β and γ are positive.

- a) Explain the meaning of the time scales associated with the inverse parameters: $1/\alpha$, $1/\beta$ and $1/\gamma$.

Solution

$1/\alpha$ is typical time for an infective to become recovered.
 $1/\beta$ is typical time for a susceptible to be infected by a single infective.
 $1/\gamma$ is typical time for a recovered to lose its immunity and become susceptible.

- b) Verify that $s + i + r$ is conserved in Eq. (4), and use this fact to remove r from the equations, leaving two equations for s and i .

Solution

The dynamics of the total population $s + i + r$ is

$$\dot{s} + \dot{i} + \dot{r} = -\beta si + \gamma r + \beta si - \alpha i + \alpha i - \gamma r = 0.$$

Since $s + i + r$ is constant, we can replace $r = 1 - s - i$ in Eq. (4) to obtain the two equations

$$\begin{aligned}\dot{s} &= -\beta si + \gamma(1 - s - i) \\ \dot{i} &= \beta si - \alpha i\end{aligned}$$

- c) Show that this model has two steady states, one disease free and one endemic (you do not need to evaluate their stability). Give a condition for the endemic steady state to exist.

Solution

The fixed points of the system in subtask b) are $(s_1^*, i_1^*) = (1, 0)$ and $(s_2^*, i_2^*) = (\frac{\alpha}{\beta}, \frac{\gamma}{\alpha + \gamma}(1 - \frac{\alpha}{\beta}))$. (s_1^*, i_1^*) is disease free and (s_2^*, i_2^*) is endemic. The endemic steady state exists if $\beta > \alpha$ (and if $\gamma > 0$).

Assume that the SIRS model (4) successfully describes the spread of a disease, for example the Delta variant of COVID-19. Now assume a new variant of the disease forms, for example Omicron. Assume that the new variant gives similar symptoms (same α and γ), but it is more infectious, and each variant only gives partial immunity to the other variant.

To simplify, assume further:

- One individual can at most be infected by one variant at a time
 - Individuals do not die, when they recover they become immune to the variant they were infected by and partially immune to the other variant.
 - Both immunity and partial immunity is lost at the same rate.
 - If a partially immune individual gets infected, it loses all immunity (i.e. it behaves as an infected susceptible).
- d) Extend and modify the SIRS model (4) to model the spread of two variants of a disease in a population following the assumptions above.

Solution

One example model is to consider the following compartments:

s susceptible to both variants

i_1 infected by variant 1

i_2 infected by variant 2

r_1 immune to variant 1, susceptible to variant 2

r_2 immune to variant 2, susceptible to variant 1

One model for the dynamics satisfying the assumptions is

$$\dot{s} = -\beta_1 s i_1 - \beta_2 s i_2 + \gamma(r_1 + r_2)$$

This is the same as the dynamics for s in the SIRS model, but we have two infection terms with different infection rates, $\beta_2 > \beta_1$ since the second variant is more infective.

$$\dot{r}_1 = -p_2 \beta_2 r_1 i_2 + \alpha i_1 - \gamma r_1$$

$$\dot{r}_2 = -p_1 \beta_1 r_2 i_1 + \alpha i_2 - \gamma r_2$$

Same as the dynamics for recovered in the SIRS model, with contribution from partial immunity by multiplying the infection rate between recovered and infectious of opposite type by $0 < p_1 < 1$ and $0 < p_2 < 1$. Using continuity, the dynamics for i_1 and i_2 is obtained (infected minus recovered)

$$\dot{i}_1 = \beta_1 i_1 s + p_1 \beta_1 r_2 i_1 - \alpha i_1$$

$$\dot{i}_2 = \beta_2 i_2 s + p_2 \beta_2 r_1 i_2 - \alpha i_2$$

- e) Does your model in subtask d) have a steady state where one of the variants disappears due to the competition (you do not need to evaluate the stability of this steady state)? What is the form of the dynamics for the remaining disease?

Solution

Assume Omicron makes Delta go extinct, i.e. search for solutions where $i_1 = 0$ and $\dot{i}_1 = 0$:

$$\begin{aligned}\dot{s} &= -\beta_2 s i_2 + \gamma(r_1 + r_2) \\ \dot{r}_1 &= -p_2 \beta_2 r_1 i_2 - \gamma r_1 \\ \dot{r}_2 &= \alpha i_2 - \gamma r_2 \\ \dot{i}_1 &= 0 \\ \dot{i}_2 &= \beta_2 i_2 s + p_2 \beta_2 r_1 i_2 - \alpha i_2\end{aligned}$$

When $i_1 = 0$, then r_1 decreases, meaning it will approach zero as $t \rightarrow \infty$, the fixed points of the \dot{r}_1 equation. The remaining dynamics becomes

$$\begin{aligned}\dot{s} &= -\beta_2 s i_2 + \gamma r_2 \\ \dot{r}_2 &= \alpha i_2 - \gamma r_2 \\ \dot{i}_2 &= \beta_2 i_2 s - \alpha i_2\end{aligned}$$

i.e. an SIRS model for the remaining variant (the steady state values s^* and i_2^* are given in subtask b) and $r_2^* = 1 - s^* - i_2^*$).

5. Coalescent process [10 points]

- a) [5 points] 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:

$$P(T_j) = \lambda_j \exp(-\lambda_j T_j) \quad \text{with} \quad \lambda_j = \frac{1}{N} \binom{j}{2}. \quad (5)$$

Solution

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

- b) [3 points] 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

$$\langle S_n \rangle = \theta \sum_{j=1}^{n-1} \frac{1}{j} \quad \text{and} \quad \left\langle \frac{1}{\binom{n}{2}} \sum_{i < j} \Delta_{ij} \right\rangle = \theta. \quad (6)$$

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. (6) 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

Averaging over Poisson process gives $\langle S_n \rangle = \mu T_{\text{tot}}^{(n)}$. Now average over genealogies using $\langle T_{\text{tot}}^{(n)} \rangle = \sum_{j=2}^n j \langle T_j \rangle$ (need to draw a genealogy to check that this is correct). Averaging using Eq. (5) gives $2N \sum_{j=1}^{n-1} \frac{1}{j}$. Need to remember that $\binom{j}{2} = j(j-1)/2$. The j cancels, redefine the summation variable to get the result.

For the second relation, use that $\langle \Delta_{ij} \rangle = \langle S_2 \rangle$. Averaging over the Poisson process and genealogies gives $\langle \Delta_{ij} \rangle = 2\mu \langle T_2 \rangle = \theta$.

c) [2 points] The two estimators

$$\hat{\theta}_1 = \left(\sum_{j=1}^{n-1} \frac{1}{j} \right)^{-1} S_n \quad \text{and} \quad \hat{\theta}_2 = \frac{1}{\binom{n}{2}} \sum_{i < j} \Delta_{ij}$$

have the same average for neutral genealogies. Explain qualitatively in a couple of sentences why the estimators tend to have different averages when selection is important.

Hint: Remember how selection tends to change genealogies.

Solution

Selection makes the genealogies more tree-like. This makes it more likely to observe singletons in the sample, where an SNP occurs only in one individual, but not in the others. In this case one expects $\langle \hat{\theta}_2 \rangle < \langle \hat{\theta}_1 \rangle$. For a neutral genealogy, the relations (6) show that $\langle \hat{\theta}_2 \rangle = \langle \hat{\theta}_1 \rangle$.