

CHALMERS, GÖTEBORGS UNIVERSITET

EXAM for COMPUTATIONAL BIOLOGY

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

Time:	March 16, 2023, at 14 ⁰⁰ – 18 ⁰⁰
Place:	Johanneberg
Teachers:	Bernhard Mehlig, 073-420 0988 (mobile) Ludvig Storm, 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.

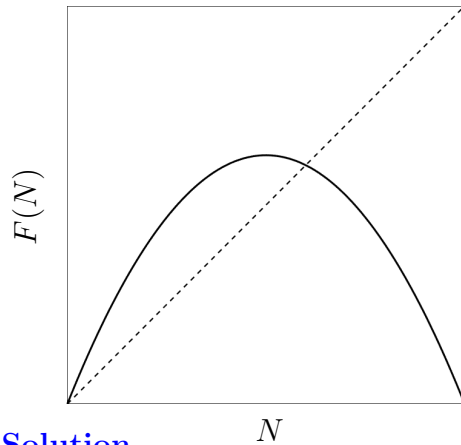
1. Short questions [12 points] For each of the following questions give a concise answer within a few lines per question.

- a) Explain why we can use a continuous description of population growth, even though the population size takes integer values.

Solution

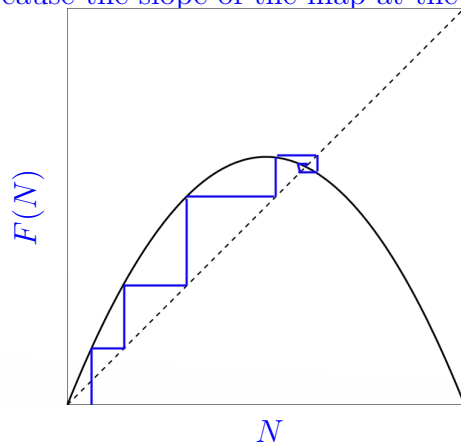
In the continuous models, it is assumed that the population size is very large compared to unity, meaning that the exact number of individuals (fractional or not) does not matter for most practical purposes. Moreover, in most cases we only know the current population size approximately, meaning that we anyway need to round the final result according to the accuracy of the initial condition.

- b) The plot below shows the map $F(N) = 2.5N(1 - N)$ together with a dashed line of unit slope. Copy the plot to your answer sheet and make a cobweb plot, starting from a small value of N .



Solution

The solution should approach the fixed point and have oscillations because the slope of the map at the fixed point lies between -1 and 0 :



- c) Explain the difference between a quasi-steady state and a regular steady state. Give an example of a system with a quasi-steady state.

Solution

A quasi-steady state is a state for which the dynamics approximately does not change for a long time, but where, in the long run, the dynamics will move to a different state. For example in a stochastic model, the distribution of individuals is approximately constant for a long time, but since there is a finite probability that the population goes extinct, the distribution must drift towards a Dirac delta function at zero population as time goes to infinity.

- d) Consider a diffusing concentration $n(x, t)$ in one spatial dimension satisfying the continuity equation

$$0 = \frac{\partial}{\partial t}n(x, t) + \frac{\partial}{\partial x}j(x, t) - f(x, t).$$

Use advection and Fick's law, $j_{\text{diffusion}} = -D\frac{\partial n}{\partial x}$, to write down the corresponding reaction-advection-diffusion equation.

Solution

Using advection and Fick's law, the total flux becomes $j(x, t) = v(x, t)n(x, t) -$

$D \frac{\partial n}{\partial x}$, giving the reaction-advection-diffusion equation

$$\partial_t n(x, t) = f(x, t) - \partial_x(v(x, t)n(x, t)) + D \frac{\partial^2 n}{\partial x^2}$$

- e) Again, consider the continuity and reaction-advection-diffusion equation in the previous subtask. Assume that n satisfies the dynamical system $\dot{n}(x, t) = g(n(x, t))$ in absence of advection and diffusion. Explain what the role of g is in the reaction-advection-diffusion equation.

Assume that x satisfies the dynamical system $\dot{x} = h(x, t)$ in absence of diffusion. Explain what the role of h is in the reaction-advection-diffusion equation.

Solution

In absence of advection and diffusion ($j = 0$), both equations read $\frac{\partial n}{\partial t} = f(n(x, t))$ implying $\frac{dn}{dt} = f(n(x, t))$ (since $\dot{x} = 0$ when $j = 0$). We therefore identify $g(n(x, t))$ with the source term $f(n(x, t))$ in the reaction-advection-diffusion equation.

In absence of diffusion, the flux is given by the advective contribution, $j = v(x, t)n(x, t)$, originating from particles that follows a fluid flow, $\dot{x} = v(x, t)$. Thus, $h(x, t)$ exactly corresponds to the flow $v(x, t)$ in the reaction-advection-diffusion equation.

- f) What is meant by the ‘order parameters’ in the Kuramoto model? What do they quantify?

Solution

Bernhard’s lecture notes 2.4.

- g) Oscillations are often modeled by second-order time derivatives. Explain why only first-order derivatives appear in the Kuramoto model.

Solution

The oscillators in the Kuramoto model can be understood as driven and overdamped. A driven damped oscillator will oscillate (second-order differential equation as for the harmonic oscillator) but eventually damping will stabilize the dynamics to a steady state given by the driving. If the time scale of the damping is short compared to the other time scales in the system, we can neglect this transient behavior and only consider the stabilized behavior. In this limit it is enough to consider a first-order equation to describe a system with a stable steady state.

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

Explain what the difference is between measurement noise and dynamical noise in a time series obtained from experimental measurements of a biological process.

Solution

The measurement noise corresponds to fluctuations in an observable

due to inaccuracies in a measurement device. The dynamical noise on the other hand corresponds to inherent fluctuations in the evolution of the underlying dynamics, for example due to imperfections of the model or due to stochastic fluctuations.

2. Von Foerster's doomsday model [10 points] Based on a fit to data on worldwide human population from the last two millenia, von Foerster suggested in 1962 the following growth model for the human population $N(t)$

$$\dot{N} = \frac{r}{B}N^2, \quad (1)$$

where r and B are positive parameters, and $N(0)$ is smaller than B .

a) It is possible to rewrite Eq. (1) on the form

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

with a function K that depends on N . What is the form of $K(N)$?

Solution

Equating the two expressions for \dot{N} gives

$$\frac{r}{B}N^2 = rN \left(1 - \frac{N}{K(N)} \right) \Rightarrow \frac{N}{K(N)} = 1 - \frac{N}{B} \Rightarrow K(N) = \frac{N}{1 - N/B}$$

b) Explain why the form of $K(N)$ in subtask a) with $N < B$ can be argued to be a good model for growth of the human population, but perhaps not for other species. Does something strange happen when N becomes larger than B ?

Solution

$K(N)$ can be interpreted as a population-dependent carrying capacity. When $N < B$, this carrying capacity increases as the human population increases. This can be argued to be a good model for humans that have been (and possibly are) able to increase their carrying capacity through inventions. Animals are not expected to be able to increase their carrying capacity in this way.

When N approaches B , the carrying capacity goes to infinity, to become negative for $N > B$. This is something strange indeed, and the interpretation of K as a carrying capacity is not longer justified.

c) Solve the dynamics for N in Eq. (1) as a function of time, given the initial size N_0 at $t = 0$.

Solution

Solving Eq. (1) by separation of variables

$$dt = \frac{B}{r} \frac{dN}{N^2} \Rightarrow t = -\frac{B}{r} \left[\frac{1}{N} - \frac{1}{N_0} \right]$$

Solve for N to obtain

$$N(t) = \frac{1}{\frac{1}{N_0} - \frac{r}{B}t},$$

where N_0 is the population at $t = 0$.

- d) Von Foerster used the solution in subtask c), with fitted $B/r \approx 9/5 \cdot 10^{11}$ years, to predict the future population. Starting from 3 billion people in 1963, what population size did he predict for the current year (2023)? Is the answer reasonable? If not, why does the model fail?

Solution

Inserting $N_0 = 3 \cdot 10^9$, $B/r \approx 9/5 \cdot 10^{11}$ and $t = 60$ years into the solution in subtask c) give

$$N(t) = \frac{1}{\frac{1}{3 \cdot 10^9} - \frac{5}{9 \cdot 10^{11}} 60} = \frac{3 \cdot 10^9}{1 - 1} = \infty$$

I.e. the human population grows to infinity when the denominator becomes zero. This is of course unreasonable, looking at the population size today. There are several arguments of why the model gives unreasonable predictions, for example it is unrealistic with infinite carrying capacity or that the per capita growth rate becomes infinite as $N \rightarrow \infty$. This example shows that a model based on interpolation of data often fails for the purpose of extrapolation.

3. Model for spreading of measles [10 points] The following is a model for spreading of measles in a population of S susceptibles, I infectives, and R recovered/immune

$$\begin{aligned}\dot{S} &= (1-p)b(S+I+R) - \beta S \frac{I}{S+I+R} - dS \\ \dot{I} &= \beta S \frac{I}{S+I+R} - \alpha I - d(1+\epsilon)I \\ \dot{R} &= pb(S+I+R) + \alpha I - dR\end{aligned}\tag{2}$$

A fraction p ($0 \leq p \leq 1$) of newborns are vaccinated. The parameters b , d , β and α are positive and $0 < \epsilon \ll 1$.

- a) Give brief explanations of the different terms in Eq. (2). How is the deadliness of measles modeled?

Solution

A fraction $1-p$ of births go into susceptibles, modeled by $(1-p)b(S+I+R)$, and the rest $(pb(S+I+R))$ go into recovered/immune. Since the total number of births is proportional to the total population size $S+R+I$, all individuals are assumed to give birth at equal rate. It is assumed that vaccination and recovery gives lifelong immunity.

There are three death terms, all proportional to the number of individuals in respective category. The model assumes that the death rate is increased by a factor $1+\epsilon$ for infectives, taking account for the deadliness of the disease.

The terms $\beta SI/(S+I+R)$ is the per susceptible infection term, proportional to the fraction of the population being infective (the normalization $1/(S+I+R)$ is usually included in β , but the population size is not constant here). The terms αI is the removal rate of infectives due to recovery.

- b) Find a condition on the parameters such that the model (2) has at least one disease-free steady state with positive population size. Comment on the form of this steady state when $p=0$ and when $p=1$.

Solution

Disease free states have $I=0$, giving $\dot{I}=0$ in (2), leaving

$$\begin{aligned}\dot{S} &= (1-p)b(S+R) - dS \\ \dot{R} &= pb(S+R) - dR.\end{aligned}$$

For general parameter values, this system only has the fixed-point $R^*=S^*=0$, corresponding to an extinct population. Solving the system with the requirement that either R or S is non-zero gives the condition $b=d$. For this case, there is a line of fixed points satisfying

$$pS^* = (1-p)R^*.$$

When $p = 0$ we have $R^* = 0$, i.e. the population consists of only susceptibles in the steady state (if we wait long enough after the disease dies out also all recovered die). When $p = 1$ we have $S^* = 0$, i.e. the population consists of only immune/recovered in the steady state since all newborns belong to this category after vaccination.

- c) Assume that the condition you found in subtask b) is satisfied and that the initial population is equal to the disease-free steady state. Find a condition on the parameters such that disease introduced by a small number of infectives in the population does not spread.

Solution

Having a single compartment for infectives, I , the next-generation matrix is scalar. The initial change in I is therefore given by

$$\dot{I}|_{t=0} \approx \left(\beta \frac{S^*}{S^* + R^*} - \alpha - d(1 + \epsilon)\right)I_0 = (\beta(1 - p) - \alpha - d(1 + \epsilon))I_0$$

The condition for the disease not to spread is

$$\beta(1 - p) < \alpha + d(1 + \epsilon).$$

This is equivalent to a basic reproduction number $r_0 = \frac{\beta(1-p)}{\alpha+d(1+\epsilon)}$.

- d) Assuming that the model does not allow limit cycle or chaotic solutions, what is the long-term behavior of S , I , and R according to the model without vaccination (measles, being a contagious disease, spreads in this case)? If you find that the result is unrealistic, explain what property of the model underlies your finding and discuss how the model could be improved to become more realistic.

Hint: It may be helpful to look at the dynamics of the total population.

Solution

Without vaccination, the positive disease-free steady state is unstable, meaning that the population will end up at the fixed point $S^* = I^* = R^* = 0$, i.e. the population goes extinct. The reason for this is the mortality ϵ , slowly reducing the total population size $N = S + I + R$, not allowing a constant population

$$\dot{N} = \dot{S} + \dot{I} + \dot{R} = (b - d)(S + I + R) - d\epsilon I = -d\epsilon I$$

As long as $I > 0$, the total population will decline until it reaches $N = 0$ (all other steady states with $I = 0$ are unstable according to the previous subtasks). Alternatively, one can evaluate the linear stability for $S^* = I^* = R^* = 0$, to find that it is stable for all parameters.

The problem with the model is related to the fact that without infections, the growth is linear, requiring a sensitive balance between births and deaths. We can resolve this problem by introducing a carrying capacity, defining a stable steady state in the absence of the disease.

4. Diffusion-driven instability in a periodic domain [10 points]

Consider a reaction-diffusion equation for two concentrations $u(x, t)$ and $v(x, t)$ in one spatial dimension

$$\begin{aligned}\frac{\partial u}{\partial t} &= f_u(u, v) + D_u \frac{\partial^2 u}{\partial x^2} \\ \frac{\partial v}{\partial t} &= f_v(u, v) + D_v \frac{\partial^2 v}{\partial x^2}\end{aligned}$$

in a system of length L . Compactly, we write

$$\frac{\partial \mathbf{n}}{\partial t} = \mathbf{f}(\mathbf{n}) + \mathbb{D} \frac{\partial^2 \mathbf{n}}{\partial x^2}, \quad \text{with } \mathbf{n} = \begin{pmatrix} u \\ v \end{pmatrix}, \quad \mathbf{f} = \begin{pmatrix} f_u \\ f_v \end{pmatrix} \quad \text{and } \mathbb{D} = \begin{pmatrix} D_u & 0 \\ 0 & D_v \end{pmatrix}. \quad (3)$$

- a) What are the dimensions of u , f_u and D_u ?

Solution

Being a concentrations, the dimension of u is either size over length, N/L , or simply one over length, $1/L$ if the size is considered dimensionless (both are accepted). The dimension of f_u must have an additional time factor compared to n , $[f_u] = N/(LT)$, due to the partial derivative on the left-hand side. The dimension of the diffusion coefficient is $[D_u] = L^2/T$ to match up with the other terms.

- b) Assume that \mathbf{n}^* is a spatially homogeneous steady state to the system (3). Make a small spatio-temporal perturbation to this steady state and show, by separations of variables, that the spatial part of the perturbation, $X(x)$, satisfies Helmholtz equation

$$\frac{d^2}{dx^2} X(x) = -k^2 X(x).$$

Here k is a constant coming from the separation of variables.

Solution

Following the lecture notes, inserting $\mathbf{n} = \mathbf{n}^* + \delta \mathbf{n}(x, t)$ into Eq. (3) gives to linear order in $\delta \mathbf{n}$

$$\frac{\partial}{\partial t} \delta \mathbf{n} = \mathbb{J}(\mathbf{n}^*) \delta \mathbf{n} + \frac{\partial^2}{\partial x^2} \mathbb{D} \delta \mathbf{n}$$

Separation of variables using the ansatz $\delta \mathbf{n}(x, t) = T(t)X(x)\delta \mathbf{n}_0$ with constant $\delta \mathbf{n}_0$ gives the Helmholtz equation

$$\begin{aligned}\frac{\partial}{\partial t} T(t)X(x)\delta \mathbf{n}_0 &= T(t)X(x)\mathbb{J}(\mathbf{n}^*)\delta \mathbf{n}_0 + T(t)\frac{\partial^2}{\partial x^2} X(x)\mathbb{D}\delta \mathbf{n}_0 \\ \Rightarrow \frac{T'(t)}{T(t)}\delta \mathbf{n}_0 &= \mathbb{J}(\mathbf{n}^*)\delta \mathbf{n}_0 + \frac{X''(x)}{X(x)}\mathbb{D}\delta \mathbf{n}_0.\end{aligned}$$

Only the last term depends on x , meaning that if x is varied, the last term is constant, implying that the prefactor $X''(x)/X(x)$ is constant. Denoting the constant $-k^2$ gives the sought Helmholtz equation.

- c) Solve the Helmholtz equation for no flux boundary conditions (vanishing derivatives at the boundaries). Which wave numbers are allowed? Illustrate the patterns corresponding to the two first positive wave numbers.

Solution

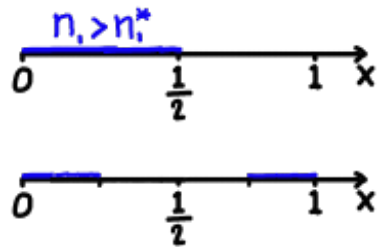
The solution to Helmholtz equation is

$$X(x) = A \cos(kx) + B \sin(kx)$$

Implementing the boundary conditions $X'(0) = X'(L) = 0$, using $X'(x) = -Ak \sin(kx) + Bk \cos(kx)$, gives

$$X(x) = A \cos(kx) \text{ with } k = 0, \frac{\pi}{L}, \frac{2\pi}{L}, \dots$$

Assuming color where the concentration is above the steady state, the pattern for the first two non-zero wave numbers are (the upper bound should be L , not 1):



- d) Now consider instead a circular domain with periodic boundary conditions and circumference L . Which wave numbers are allowed in this case? Describe the patterns corresponding to the two first positive wave numbers.

Solution

As before, the solution to the Helmholtz equation

$$X(x) = A \cos(kx) + B \sin(kx)$$

Use periodic boundary condition: $X(x) = X(x + L)$ and $\frac{dX}{dx}(x) = \frac{dX}{dx}(x + L)$. These conditions are true for any x . Choose $x = 0$ to obtain

$$\begin{aligned} A &= A \cos(kL) + B \sin(kL) \\ Bk &= -Ak \sin(kL) + Bk \cos(kL) \end{aligned}$$

One solution is $k = 0$. If $k \neq 0$, we simplify the second equation by dividing it by k . Multiply the first equation by B and the second equation by A and subtract to obtain

$$0 = (A^2 + B^2) \sin(kL).$$

The only non-trivial solution must have $\sin(kL) = 0$, i.e. k is a multiple of π/L . Inserting $\sin(kL) = 0$ into the original equations gives

$A = A \cos(kL)$ and $B = B \cos(kL)$. We find the additional constraint $\cos(kL) = +1$. In conclusion k must take values $0, 2\pi/L, 4\pi/L, \dots$

The pattern for the first non-zero mode is a colored half-circle, the pattern for the second mode is two equally-separated colored quartercircles. The starting position of the patterns are arbitrary due to circular symmetry.

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:

$$P(T_j) = \lambda_j \exp(-\lambda_j T_j) \quad \text{with} \quad \lambda_j = \frac{1}{N} \binom{j}{2}. \quad (4)$$

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

$$\langle S_n \rangle = \theta \sum_{j=1}^{n-1} \frac{1}{j} \quad \text{and} \quad \left\langle \frac{1}{\binom{n}{2}} \sum_{i < j} \Delta_{ij} \right\rangle = \theta. \quad (5)$$

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. (5) 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