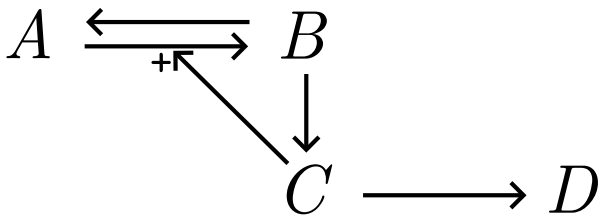
Dugga 2024-02-21

TBMT19 / TBMT37

Write your Dugga-ID on all pages and your answers in Swedish or English. There are in total 5 questions, each worth 3 points. You need at least 2 points per question or 12 points in total to pass. Good luck! /William

1 Model formulation and model parts

Consider the following interaction graph.



(a) Formulate an ODE model based on the interaction graph. Assume that the reaction $C \rightarrow D$ is saturated with respect to C. Also assume that we can measure something proportional to the concentration of B. Make any other necessary assumptions and include these assumptions in the answer. Introduce parameters and initial values with values of your choice. Make sure your suggested model is complete. (3 points)

2 Model simulation

(a) Explain the principles of numerical simulation, and describe the principles of the Euler forward method (2 points)

(b) Why do we need to numerically simulate systems biology models? (1 point)

3 Parameter estimation

(a) You have a model simulation ($\hat{y}$) of four different time-points for a given parameter set (θ):

$$\hat{y}(\theta) = [3,4,6,3]$$

You also have corresponding experimental data (y) with measurement uncertainties (SEM):

$$y = [1,1,2,3], \qquad SEM = [2,1,2,4]$$

Explain the cost function: what are the parts and input/output of the function? What is the cost for the parameter set θ? (2 points)

(b) Are global optimization algorithms guaranteed to find the optimal parameter values? (1 point)

4 Statistical tests

Assume that we again have the same model simulation and data as in question 3 above.

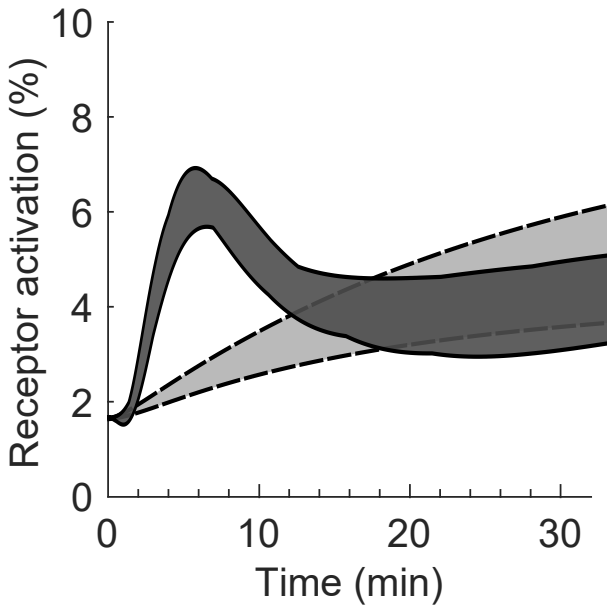
(a) If the χ^2-limit is 9.48, should you reject the model simulation given the parameter set θ? What is the conclusion of the test?
If you did not calculate the cost, assume that the cost is 12.345. (1 point)

(b) Would you reject the model simulation with a whiteness test? Why/why not? (1 point)

(c) If we assume that the parameter set θ is the optimal parameter set, what additional conclusions can you draw? (1 point)

5 Predictions and experimental design

Assume that you have the following model predictions from two different models. The first model is represented by the light gray area with striped edges, and the second model is represented by the dark gray area with solid edges. The predictions are for the activation of a receptor over time.



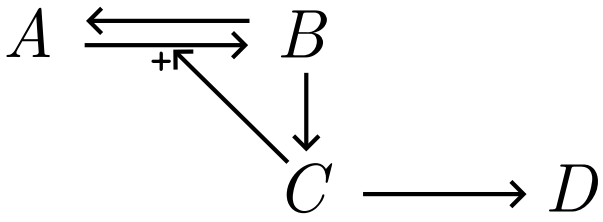
(a) If you could ask an experimentalist to measure the receptor activation at only one time-point, which time-point would you pick? What could be the potential outcomes when you get the data? Motivate your answer! (2 points)

(b) What is a core prediction? (1 point)

Answers: Dugga 2024-02-21

1 Model formulation and model parts

Consider the following interaction graph.



(a) Formulate an ODE model based on the interaction graph. Assume that the reaction $C \to D$ is saturated with respect to C. Also assume that we can measure something proportional to the concentration of B. Make any other necessary assumptions and include these assumptions in the answer. Introduce parameters and initial values with values of your choice. Make sure your suggested model is complete. (3 points)

Answer:
I assume that the states correspond to concentrations.

Start by identifying the states of the model: [A], [B], [C], [D].

Next, identify the reactions and reaction rates:

$$A \to B: \quad v1 = k1 \cdot [A] \cdot [C] \tag{1}$$
$$B \to A: \quad v2 = k2 \cdot [B] \tag{2}$$
$$B \to C: \quad v3 = k3 \cdot [B] \tag{3}$$
$$C \to D: \quad v4 = V_{max} \frac{[C]}{K_m + [C]} \tag{4}$$

I assumed that all reactions except $C \to D$ are with mass action kinetics.

Give values to introduced parameters: [k1, k2, k3, Vmax, Km] = [1, 2, 2, 1, 3].

Construct the ODEs using the reaction rates above:

$$\frac{d[A]}{dt} = -v1 + v2 \tag{5}$$
$$\frac{d[B]}{dt} = v1 - v2 - v3 \tag{6}$$
$$\frac{d[C]}{dt} = v3 - v4 \tag{7}$$
$$\frac{d[D]}{dt} = v4 \tag{8}$$

Introduce initial values: [A](0) = 1, [B](0) = 0, [C](0) = 0, [D](0) = 1.

Define the measurement equation: $\hat{y} = ky \cdot [B]$. Assume ky = 4.

2 Model simulation

(a) Explain the principles of numerical simulation, and describe the principles of the Euler forward method. (2 points)

Answer:
The principles of numerical simulation involves solving ODEs numerically instead of analytically, by taking multiple small steps along the direction of the flow/gradient. The gradient can be calculated using different methods, such as the Euler forward method or the Runge-Kutta method(s).

In the Euler forward method, we start with the initial values of the states and calculate the values of the ODEs at the initial time-point. We then take a small step in the direction of the ODEs, and use the new values to calculate the ODEs at the next time-point. This is repeated until the time-frame asked for by the user has been simulated.

The Euler forward method can be written as:

$$x(t + \Delta t) = x(t) + d/dt(x(t)) \cdot \Delta t \tag{9}$$

(b) Why do we need to numerically simulate systems biology models? (1 point)

Answer:
In systems biology, we often have complex models with many states and reactions, and it is often not possible to solve these models analytically. Therefore, we need to use numerical simulation to solve the models and obtain the model simulations.

3 Parameter estimation

(a) You have a model simulation ($\hat{y}$) of four different time-points for a given parameter set (θ):

$$\hat{y}(\theta) = [3,4,6,3]$$

You also have corresponding experimental data (y) with measurement uncertainties (SEM):

$$y = [1,1,2,3], \qquad SEM = [2,1,2,4]$$

Explain the cost function: what are the parts and input/output of the function? What is the cost for the parameter set θ? (2 points)

Answer:
The cost function is a function that measures the difference between the model simulation (for a given parameter set) and the experimental data. The cost is calculated by calculating the difference between the model simulation and the experimental data (residuals), and then weighing the residuals by the measurement uncertainties, and summing the squared weighted residuals together. The input to the cost function is the parameter values, and the output is the agreement to data.

The cost for the parameter set θ is calculated as follows:

$$\begin{aligned}
\text{residuals} &= y - \hat{y} = [1,1,2,3] - [3,4,6,3] = [-2,-3,-4,0] \\
v(\theta) &= \sum_{t=1}^{N} \frac{(y(t) - \hat{y}(t,\theta))^2}{SEM(t)^2} = \frac{(-2)^2}{2^2} + \frac{(-3)^2}{1^2} + \frac{(-4)^2}{2^2} + \frac{0^2}{4^2} \\
&= \frac{4}{4} + \frac{9}{1} + \frac{16}{4} + \frac{0}{16} = 1+9+4+0 = 14
\end{aligned}$$

(b) Are global optimization algorithms guaranteed to find the optimal parameter values? (1 point)

Answer:
No, global optimization algorithms are not guaranteed to find the optimal parameter values.

4 Statistical tests

Assume that we again have the same model simulation and data as in question 3 above.

(a) If the χ^2-limit is 9.48, should you reject the model simulation given the parameter set θ?

What is the conclusion of the test?
If you did not calculate the cost, assume that the cost is 12.345. (1 point)

Answer:
The χ^2-limit is 9.48, and the cost is 14 (or 12.345). Since the cost is larger than the χ^2-limit, we should reject the model simulation given the parameter set θ. The conclusion of the test is that the model simulation does not agree with the experimental data (the residuals are too big).

(b) Would you reject the model simulation with a whiteness test? Why/why not? (1 point)

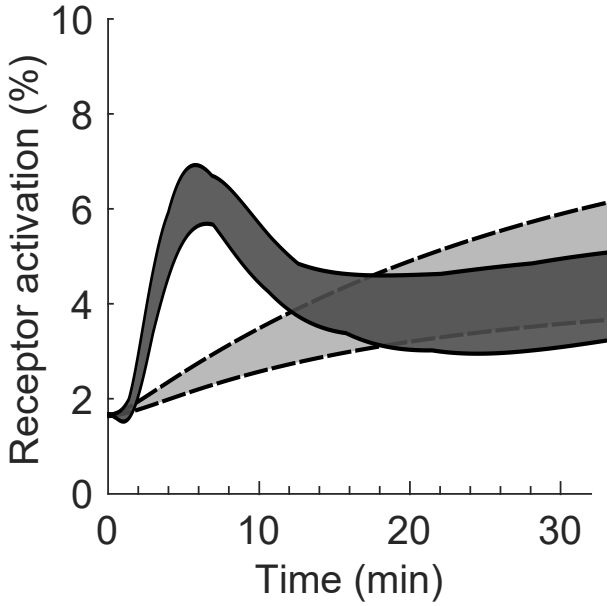
Answer:
The residuals are correlated (since the model simulation is greater or equal to the data at all time-points), and therefore we would reject the model simulation with a whiteness test (the residuals are too correlated).

(c) If we assume that the parameter set θ is the optimal parameter set, what additional conclusions can you draw? (1 point)

Answer:
If the parameter set is the optimal parameter set, we can conclude that the model (structure and/or hypothesis) is not able to describe the experimental data and must be rejected.

5 Predictions and experimental design

Assume that you have the following model predictions from two different models. The first model is represented by the light gray area with striped edges, and the second model is represented by the dark gray area with solid edges. The predictions are for the activation of a receptor over time.



(a) If you could ask an experimentalist to measure the receptor activation at only one time-point, which time-point would you pick? What could be the potential outcomes when you get the data? Motivate your answer! (2 points)

Answer:
I would pick the time-point where both of the two model predictions have a small uncertainty and also differ the most from each other. Somewhere between approximately 3 and 7 minutes. If the new experimental data does not agree with either of the model predictions, we can conclude that the model(s) not in agreement with the data must be rejected. If the experimental data has a reasonable uncertainty, we will be able to reject at least one model.

(b) What is a core prediction? (1 point)

Answer:
A core prediction is a well-determined prediction that can be measured experimentally.