Assignment 3

BT5240: Computational and Systems Biology

Parameter Estimation

Shreya Nema (BS17B033)

(04-04-2020)

Model parameter estimation is a key challenge in biological systems. Most of the time in an experiment only a fraction of parameter is estimated while the rest are often fitted and are important for model prediction.

The given model of the MAPK cascade is defined by kinetic equations with known and unknown parameters. The known parameters are given in 2 sets while the unknown parameters are required to estimate to predict the model.

The following assumptions are made already-

Model/mechanism of reaction:

$$\frac{d[MKKK]}{dt} = r_2 - r_1$$
$$\frac{d[MKKK-P]}{dt} = r_1 - r_2$$
$$\frac{d[MKK]}{dt} = r_6 - r_3$$
$$\frac{d[MKK-P]}{dt} = r_3 + r_5 - r_4 - r_6$$
$$\frac{d[MKK-PP]}{dt} = r_4 - r_5$$
$$\frac{d[MAPK]}{dt} = r_{10} - r_7$$
$$\frac{d[MAPK-P]}{dt} = r_7 + r_9 - r_8 - r_{10}$$
$$\frac{d[MAPK-PP]}{dt} = r_8 - r_9$$

Reactions:

$$r_1 = V_1 * \frac{[MKKK]}{((1 + (\frac{[MAPK-PP]}{KI})^n) * (K_1 + [MKKK]))}$$

$$r_2 = p_3 * \frac{[MKKK - P]}{(K_2 + [MKKK - P])}$$

$$r_3 = p_1 * [MKKK - P] * \frac{[MKK]}{(K_3 + [MKK])}; \quad r_4 = p_1 * [MKKK - P] * \frac{[MKK - P]}{(K_4 + [MKK - P])}$$

$$r_5 = p_2 * \frac{[MKK - PP]}{(K_5 + [MKK - PP])}; \quad r_6 = p_2 * \frac{[MKK - P]}{(K_6 + [MKK - P])}$$

$$r_7 = p_1 * [MKK - PP] * \frac{[MAPK]}{(K_7 + [MAPK])}; \quad r_8 = p_1 * [MKK - PP] * \frac{[MAPK - P]}{(K_8 + [MAPK - P])}$$

$$r_9 = V_9 * \frac{[MAPK - PP]}{(K_9 + [MAPK - PP])}; \quad r_{10} = V_{10} * \frac{[MAPK - P]}{(K_{10} + [MAPK - P])}$$

Moiety conservation equations:

$$[MKKK]_{total} = [MKKK] + [MKKK - P] = 100$$
$$[MKK]_{total} = [MKK] + [MKK - P] + [MKK - PP] = 300$$
$$[MAPK]_{total} = [MAPK] + [MAPK - P] + [MAPK - PP] = 300$$

Initial known parameters:

Set 1: $K_I = 9, n = 1, V_1 = 2.5, V_9 = V_{10} = 0.5, K_1 = 10, K_2 = 8, K_3 = K_4 = K_5 = K_6 = K_7 = K_8 = K_9 = K_{10} = 15.$
Set 2: $K_I = 18, n = 2, V_1 = 2.5, V_9 = V_{10} = 1.25, K_1 = 50, K_2 = 40, K_3 = K_4 = K_5 = K_6 = K_7 = K_8 = K_9 = K_{10} = 100.$

Experimental data is of sufficient quality and quantity.

The model is structurally identifiable.

The experimental measurements of concentration tend to be very noisy and taken at a limited number of time points. To tackle the issue of experimental data containing NAN values at the position where data could not be recorded, we neglect all the NAN values.

First step involves setting up optimization problem. The problem can be nonlinear least-squares-based(minimization of sum of least square) and population based (metaheuristic, multi agent approaches etc). Different optimization algorithms can be used for example standard minimization algorithm like fminsearch or fmincon in MATLAB or Simulated annealing or Evolutionary algorithms.

I have used constrained nonlinear optimization or say nonlinear programming method (fmincon) that finds a constrained minimum of a scalar function of several variables starting at an initial estimate with upper and lower bounds. But disadvantage of fmincon is, it may stop at local minima and is really sensitive to initial parameter values. To overcome this problem several random input estimates are given and the function is optimized at each initial parameter values. The least error value of all estimates is chosen as global minima.

Objective function:

$\sum_{i=1}^{m}(x_{m,i} - x_{p,i})$^2 - (1)

Or

$$\sum \left(\frac{x_{m,i} - x_{p,i}}{\partial_i x_{m,i}} \right)^2 \qquad \text{-(2)}$$

In (2) ∂ is the allowed error and the error is normalised by measured values.

Equation (1) is used as objective function in this case because by normalisation correct estimate of error can not be obtained owing to machine precision rounding off errors. Normalization is very useful in multi – objective optimization problems to ensure consistency of optimal solution. As here I used combined sum of squares optimisation for D1 ∪ D2, normalisation is not required.

The differential equations given in the model is solved using MATLAB ode23s function. As the model contains multiple time scales that is the model is stiff, ode23s with sufficient number of iterations and ability of handle stiff systems is reasonable to use in this case.

As moiety conservation relations are satisfies by the initial concentration there is no need to provide constraint. Ode function do not take into account variable value constrains, but these constraints can be incorporated by writing an event function that will check if the estimated values are within the bound. If not, the function will ignore the current value and will estimate the value again. This function can be provided under option parameter 'OutputFcn' and will be called after each iteration.

Code:

Run hw3_sdata.m for dataset1 set1, dataset2 set1; dataset1 set2; dataset2 set2.

And run hw3_cdata.m for dataset 1+2 set1 and set2 both.

RESULTS:

Dataset1, set1: 0.0249, 0.9629, 0.2456 error: 2.2846e + 05

Dataset1, set2: 0.0301, 1.1194, 0.2085 error: 8.8342e + 05

Dataset2, set1: 0.0111, 0.2635, 0.7298 error: 7.5314e + 05

Dataset2, set2: 0.0256, 0.9056, 0.2321 error: 1.7855e + 05

Dataset1+2, set1: 0.0209, 0.5241, 0.3710 error: 1.8831e + 06

Dataset1+2, set2: 0.0290, 1.1076, 0.2046 error: 1.4283e + 06

Squared sum error is maximum for combined dataset 1 and 2.

Minimum for Dataset2 set 2

While for single dataset it is maximum for dataset 2 set2