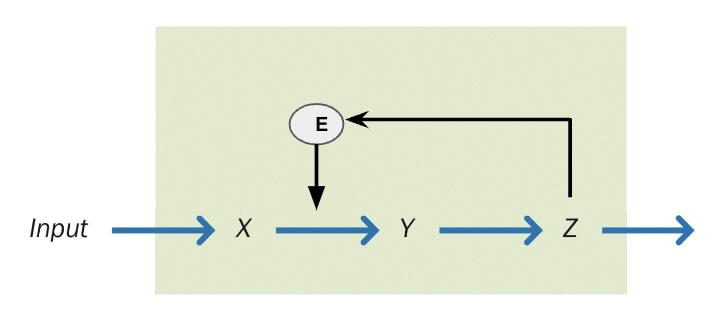
**SysBio Questionnaire for AS#5**

Name: Yunyoung Choi ID: 22100748

**NOTE: Answers for questions shown below should be written in English only. Points may be deducted if the sentence is incomplete, grammar is seriously wrong, or it is difficult to understand due to incorrect use of words and terminology.**

*Suppose you are a systems biologist studying 1) a chemical circuit that controls the concentrations of x, y, and z molecules in a particular human cell, and 2) the enzyme involved in the chemical circuit. Previous research has revealed that the structure of this chemical circuit is as shown in the figure below. This chemical circuit has positive feedback, but we still do not know much about the key enzyme (E) involved in this feedback loop. What is currently known is that 1) this enzyme has one binding site for one Z molecule and 2) under certain conditions, the concentrations of x, y, and z oscillate. Based on these facts, you hypothesize that* ***this enzyme, present in a constant concentration in the cell, is activated by the binding of ligand Z and forms a multimer with positive cooperativity.***

*In order to check if your hypothesis is correct and to analyze the mechanism of action of the enzyme in detail, you created a mathematical model based on your hypothesis (see the equations below). To optimize the parameters of your mathematical model, you measured the concentration of [X] every hour (see the data below) under a specific initial condition:* ***[Input] = 0.75mM, [X], [Y], [Z] = 1mM, [E total] = 1mM****.*

|  |  |
| --- | --- |

1. When we search the model parameters, we know that we need to reasonably limit the size of the parameter space (otherwise it would take a lot of time, or we might not be able to find the optimal parameter set within the limited time). **Suppose you don't know anything about the parameters: *a, b, c, d, n*.** What we know is 1) the facts described above, 2) the initial values of *X, Y, Z*, and *E total*, and 3) the measured concentrations of *X* over a specific time period. In this situation, how did you set the range for each of your model parameters? Is there any logical basis for such a setting? Or did you set the range arbitrarily? Write the range of each parameter (e.g. a = [0,10], interval = 0.01) used for parameter estimation and answer the questions above.

| We really don't know how to determine the range parameters a,b,c and d, so we actually arbitrarily range them based on the parameters we used in the last model.  a: [1.6, 2.5]  b: [0.9, 1.0]  c: [0.9, 1.0]  d: [0.5, 1.2]  n: [2, 4]  When determining the range of n, It was decided following the Hill function. We have to consider the enzyme cooperativity, so ‘n’ must be greater than 1 to generate cooperativity. n=2 is a really ideal condition because it needs accurate concurrency. so we set the range from 2 to 4. |
| --- |

1. Did the Pure Monte Carlo algorithm (which you used in the previous assignment) do a good job of optimizing the parameters for this model? How many trials (N) were conducted in the parameter estimation? How long did the parameter searching take? What are the best parameter values you found, and what is the loss (RMSE)? If your model doesn't fit the data well, can you explain why? Note: When you measure the parameter searching (computation) time, you can use the package ‘pytictoc’ for python and the package ‘tictoc’ for R.

| 1. The Pure Monte Carlo algorithm did a good job of optimizing the parameters for this model?   : The model did a really good job!    The blue line model predicted was drawn close enough to the actual observation value.   1. How many trials (N) were conducted in the parameter estimation?   I set N=10000 to estimate parameters. I tried multiple values of N (100, 1000, … , 1000000) but it took a really long time.   1. How long did the parameter searching take?     Using the pytictoc class, it took 346.68 seconds iterating 10000 times.   1. What are the best parameter values you found, and what is the loss (RMSE)?     In the RMSE graph, a parameter with the minimum RMSE value was set as the best parameter [a=2.2, b=0.94, c=0.98, d=1.14, n=2.91], and the RMSE value was 0.07.   1. If your model doesn't fit the data well, can you explain why?   I think my model fits the data well. I think the probability of getting a low RMSE value is higher because I randomly picked a different value 10,000 times(I think 10,000 is quite a big number). |
| --- |

1. Did the **simulated annealing (SA)** or genetic algorithm (GA) do a good job of optimizing the parameters for this model? Which hyperparameter values (e.g. initial temperature for SA, population number, crossover rate, etc for GA) gave the best results? How long did the parameter searching take? What are the best parameter values you found, and what is the loss (MSE or RMSE)? If your model doesn't fit the data well, can you explain why?

| 1. Did the simulated annealing (SA) or genetic algorithm (GA) do a good job of optimizing the parameters for this model?   : Yes, the model did a great job.    Although it is a little different from the simple Monte Carlo model, it can be seen that the model fits well with the data with a small RMSE value.   1. Which hyperparameter values (e.g. initial temperature for SA, population number, crossover rate, etc for GA) gave the best results?   When using the SA model, I set the initial temperature to 10 and the cooling rate to 0.95. I've increased and decreased these two hyperparameters several times, but I think those two values produce the best parameters empirically.   1. How long did the parameter searching take?     Using the pytictoc class, it took 40.74 seconds to search parameters.   1. What are the best parameter values you found, and what is the loss (MSE or RMSE)?     In the RMSE graph, a parameter with the minimum RMSE value was set as the best parameter [a=2.19, b=0.98, c=0.92, d=1.07, n=2.69], and the RMSE value was 0.10.   1. If your model doesn't fit the data well, can you explain why?   I think the model fits quite well with the data. By the way, when using a simple Monte Carlo algorithm, RMSE was 0.07, but when using SA algorithm, RMSE was calculated as 0.10. I thought it might be because the ‘number of iterations’ value was different, so I matched the numbers with a simple monte carlo algorithm. But in the SA algorithm, the parameters were exactly the same when 1000 and 10000 times were given. I thought that the SA algorithm would be better than randomly assigning process, but I'm not sure why this is happening. |
| --- |

1. In order to further narrow the parameter space, you searched for more papers and found that the binding affinity (Kd) value of the enzyme monomer for the *Z* molecule is 1 mM in a similar physiological condition. Using this value, you can infer 1) the value of *d*, and 2) that the range of the value of *a* is about [1.5, 2.5]. What do you think the value of ***d*** is? And how can the range of ***a*** be inferred?

| 1. The value of d       > d would be 1.   1. The range of the value of *a* is about [1.5, 2.5]     I know from the above inequality that *a* should be greater than 1.5, but I can't infer the upper bound why *a* should be less than 2.5. |
| --- |

1. Re-run the pure 1) Monte Carlo and 2) SA or GA (that you used in Question 3) with the values from the answer to question 4. How long did the parameter searching take? Compared to the result of Question 3, have the computation time and the result of the model improved? If so, how much? If not improved, can you explain why?

| 1. How long did the parameter searching take?     Using the pytictoc class, it took 33.06 seconds.   1. Compared to question 3, the computation time was reduced from 40 seconds to 33 seconds. And the RMSE was also reduced from 0.10 to 0.08. |
| --- |

1. Suppose that the value of hill coefficient ***n*** (a model parameter) you estimated is very close to the truth. What does the value (inferred through a computer modeling) mean biologically? Is the value consistent with your hypothesis and the facts mentioned above?

| According to the parameter *n* found in the SA model, n was 2.87(best). Since *n* is a natural number, rounding is n=3. The hypothesis for this problem was that the enzyme forms a multimer and causes positive cooperativity when Z is combined into ligand. Since n is greater than 1, it can be interpreted as causing positive cooperativity is correct. |
| --- |

1. Given only a handful of facts about the chemical circuit and the values of [*X*] measured over a short period of time, you have estimated the key parameters for *X*, *Y, Z*, and the enzyme *E*. If you successfully find the model parameters, they all have biological meaning. What does it mean to you that you can discover important information about *X*, *Y,* *Z* and *E* through mathematical modeling without actually experimenting with them?

| Mathematical modeling without experiments gives us some advantages. First of all, it can predict more various situations. In real-world experiments, even if only a short time can be measured, the computer can see how it will change after a longer time. We can also see how it will change when more diverse variables are added. Secondly, it can provide simple insights from complex data, and it can also integrate various domains. Finally, It helps our understanding. Sometimes even things that are too complicated to interpret can be easily interpreted in a mathematical way. Modeling can handle calculations that are difficult for humans to do. |
| --- |

1. Please write down what you learned from this assignment or what was most memorable to you (Any answer will earn points).

| I actually got help from ChatGPT when I did this assignment. While I was having so much trouble writing the SA code, I thought I should use it, and after a lot of questions, I finally got the good results. The reason why this was memorable was that it was amazing to let me know some information that I couldn't find easily even if I kept googling. Also it helped me solve the problem in a short time. It was both hopeful and shocking that we could now get help from machines, not only from humans. |
| --- |

1. Do you have any questions or things you don't understand on this topic? (Any question you ask will earn points)

| As I said in the previous question, I wonder why the RMSE value of the SA algorithm is higher than that of the simple monte carlo algorithm. It doesn't seem to be just a difference in the number of iterations. |
| --- |

1. If you have your team member or outside member who helped you understand, please write the name and what you are grateful for. If there is a person who contributes and helps your team the most, please list that person's name as well.

| Systematic trio team, and Soonjun Kwon |
| --- |

1. If your team worked together and helped each other to complete this assignment, attach a photo of it here. I'll give you 10 bonus points. Note: Photos should show the names of the people who participated in the meeting.

