CELLULAR AND SYSTEMS MODELING - HW4

Student: Shu-Ting Cho (shc167@pitt.edu)

1. Chemoreceptor clustering (PBOC2, problem 13.8). (10 points)

There is strong evidence that chemoreceptors in *E. coli* tend to cluster near one pole (see Kentner and Sourjik (2006) and Figure 13.23). One hypothesis about the role of such clustering is that it might increase the ability of a bacterium to better detect molecules in its environment. Determine if this is the most efficient strategy for counting (absorbing) molecules of chemoattractant. Approximate *E. coli* as a sphere $a = 1\,\mu\text{m}$ in radius and neglect its motion. Then compare the diffusive current to $N = 1000$ receptors (absorbing patches of radius $s = 1\,\text{nm}$) scattered over the surface of the cell with the diffusive current to the same receptors incorporated into a single patch with the same total area. Make use of the result that the diffusive current onto a sphere of radius a with N absorbing patches of radius s spread uniformly over its surface is

$$I = \frac{4\pi D a c_\infty}{1 + \pi a/Ns}, \tag{13.188}$$

where D is the diffusion constant of the molecules, while c_∞ is their concentration far from the cell. (Adapted from a problem courtesy of H. C. Berg.)

(Hint: consider the case of N=1 to determine the cluster current.)

diffusive current: $I = \frac{4\pi D a c_\infty}{1+\pi a/Ns}$

Receptors scattered over the suface: N=1000

$$I_s = \frac{4\pi D(1\mu m)c_\infty}{1+\pi(1\mu m)/(1000)(1nm)}$$

Receptors incorporated into a single patch: N=1

$$I_c = \frac{4\pi D(1\mu m)c_\infty}{1+\pi(1\mu m)/(1)(1nm)}$$

$$(1000)(1nm) > (1)(1nm) \Rightarrow I_s > I_c$$

The cluster current I_c is less than the scatter current I_s, so the clustering chemoreceptors could decrease the diffusion current and have a better detection efficiency.

CELLULAR AND SYSTEMS MODELING - HW4

Student: Shu-Ting Cho (shc167@pitt.edu)

2. Credible interval (PMLS2, problem 7.5). (10 points)
Ignore the hint in part (a) and use the regularized incomplete beta function instead to evaluate the necessary integrals.

7.5 *Credible interval*

a. Six hundred flips of a coin yielded 301 heads. Given that information, compute the 90% credible interval for the coin fairness parameter ξ by using the method outlined in the text. Assume a Uniform prior for ξ. [*Hint:* Your computer math package may balk at integrating the likelihood function. To help it out, first find the location ξ_* of the peak analytically. Next consider the function $f(\xi) = (\xi/\xi_*)^{\ell}((1-\xi)/(1-\xi_*))^{M-\ell}$. This is the likelihood, divided by its peak value. It still is not normalized, but at least it never exceeds 1.]

$P(model_{\xi}|l = 301, M = 600)$

$\xi_* = 301/600$

Find Δ so that the area under the posterior distribution $P(model_{\xi}|l = 301, M = 600)$ between $\xi_* - \Delta$ *and* $\xi_* + \Delta$ is 90%

Using the regularized incomplete beta function we found that $\Delta \simeq 0.0335$
The 90% credible interval of ξ is between 0.4682 and 0.5352.

b. Six out of 25 animals fed a suspected carcinogen developed a particular cancer. Find the 90% credible interval for the probability ξ to develop the disease, and comment on whether this group is significantly different from a much larger control group, in which ξ_{control} was 17%. Again assume a Uniform prior.

$P(model_{\xi}|l = 6, M = 25)$

$\xi_* = 6/25 = 24\%$

The 90% credible interval of ξ is between 0.1338 and 0.4054.

The $\xi_{control} = 17\%$ is located within the 90% credible interval of ξ, thus thiis experimental group is not significantly different from the control group.

c. This time, suppose that both the control and the experimental groups were of finite size, say, 25 individuals each. Describe a procedure for assessing your confidence that their distributions are different (or not).

For control group, if 4 out of 25 develop the cancer, $\xi_{control} = 4/25 = 16\%$:

$P(model_{\xi}|l = 4, M = 25)$

CELLULAR AND SYSTEMS MODELING - HW4

Student: Shu-Ting Cho (shc167@pitt.edu)

$\xi_* = 4/25 = 16\%$

90% credible interval of $\xi_{control}$ is between 0.0790 and 0.3182
90% credible interval of $\xi_{experiment}$ is between 0.1338 and 0.4054

Since the 90% credible interval for the two datasets overlap, we can say the probability that these two observed results are from the same $P(model_\xi)$ distribution (the null hypothesis) is > 0.05, which means these two distributions are not significantly different.

CELLULAR AND SYSTEMS MODELING - HW4

Student: Shu-Ting Cho (shc167@pitt.edu)

3. Likelihood analysis of a Poisson process (PMLS2, problem 9.7). (10 points)

9.7 *Likelihood analysis of a Poisson process*

Suppose that you measure a lot of waiting times from some random process, such as the stepping of a molecular motor. You believe that these times are draws from an Exponential distribution: $\wp(t) = Ae^{-\beta t}$, where A and β are constants. But you don't know the values of those constants. Moreover, you only had time to measure six steps, or five waiting times $t_{w,1}, \ldots, t_{w,5}$, before the experiment ended.[27]

a. A and β are not independent quantities: Express A in terms of β. State appropriate units for A and for β.

$A = \beta$

b. Write a symbolic expression for the likelihood of any particular value of β, in terms of the measured data $t_{w,1}, \ldots, t_{w,5}$.

original function: $p(t) = Ae^{-\beta t}$

The likelihood function

$$L(\beta) = \prod_{i=1}^{5} \beta \times e^{-\beta t_{w,i}}$$

The log-likelihood function

$$logL(\beta) = 5log\beta + \sum_{i=1}^{5} (-\beta \times t_{w,i})$$

c. Find the maximum-likelihood estimate of the parameter β; give a short derivation of your formula.

Derivation: compute gradient

$$\frac{\partial logL(\beta)}{\partial \beta} = 5 \times \frac{1}{\beta} - \sum_{i=1}^{5} (t_{w,i})$$

The equation gradient = 0

$$5 \times \frac{1}{\beta} - \sum_{i=1}^{5} (t_{w,i}) = 0$$

$$\Rightarrow 5 \times \frac{1}{\beta} = \sum_{i=1}^{5} (t_{w,i})$$

$$\Rightarrow \beta_{MLE} = \frac{5}{\sum_{i=1}^{5} (t_{w,i})}$$

CELLULAR AND SYSTEMS MODELING - HW4

Student: Shu-Ting Cho (shc167@pitt.edu)

The next three problems makes use of the following model, also known as the "Standard model of viral kinetics," which describes the dynamics of a viral infection within a host. The basic model consists of the following four reactions:

$$S + V \xrightarrow{b} I$$
$$I \xrightarrow{d} \emptyset$$
$$I \xrightarrow{p} I + V$$
$$V \xrightarrow{g} \emptyset$$

where *S* represents a cells susceptible to infection, *V* represents viral particles, and *I* represents infected cells. The rate of each reaction is governed by the kinetic Law of Mass Action, i.e., its rate is governed by the rate constant times the product of reactant concentrations. In this problem you will take the initial size of the susceptible cell population, S_0, to be 10^7, the initial number of infected cells, I_0, to be 1, and the initial number of viral particles, V_0, to be pI_0/g. In addition we will hold the rate constant g fixed at 15/day.

4. Constructing and analyzing an ODE model (10 points)

(a) What are the differential equations that define the dynamics of the model?

Using formula: $\frac{dy}{dt} = input\ rate - output\ rate$

S: No reaction for input. 1st reaction for output

$$\frac{dS}{dt} = -\ b[S][V]$$

V: 3rd reaction for input. 1st & 4th reactions for output

$$\frac{dV}{dt} = p[I] - b[S][V] - g[V]$$

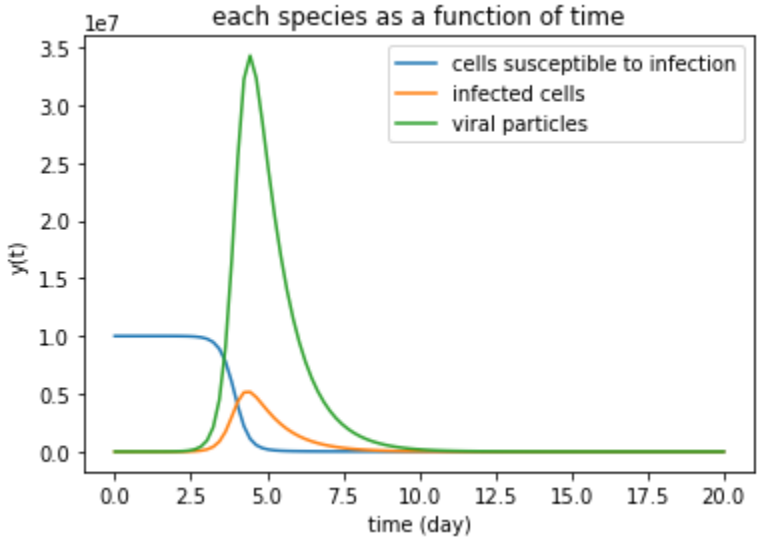
I: 1st & 3rd reactions for input. 2nd & 3rd reactions for output

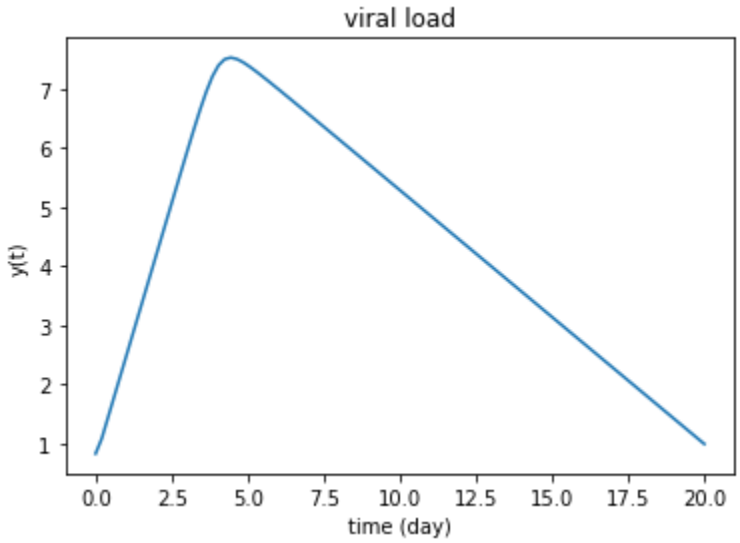
$$\frac{dI}{dt} = b[S][V] + p[I] - d[I] - p[I] = b[S][V] - d[I]$$

CELLULAR AND SYSTEMS MODELING - HW4

Student: Shu-Ting Cho (shc167@pitt.edu)

(b) Write a Python function to simulate the model using numerical integration (scipy's *odeint* or *solve_ivp* are good choices). Plot the results you obtain with the values of the remaining parameters taken as (b, d, p) = (1.0e-7, 1, 100). You should make two plots, one showing each species as a function of time and another showing the viral load, which is defined as the log10 of the number of virus particles. This is the standard way that viral kinetic data is displayed and you should use this form for the remaining plots.

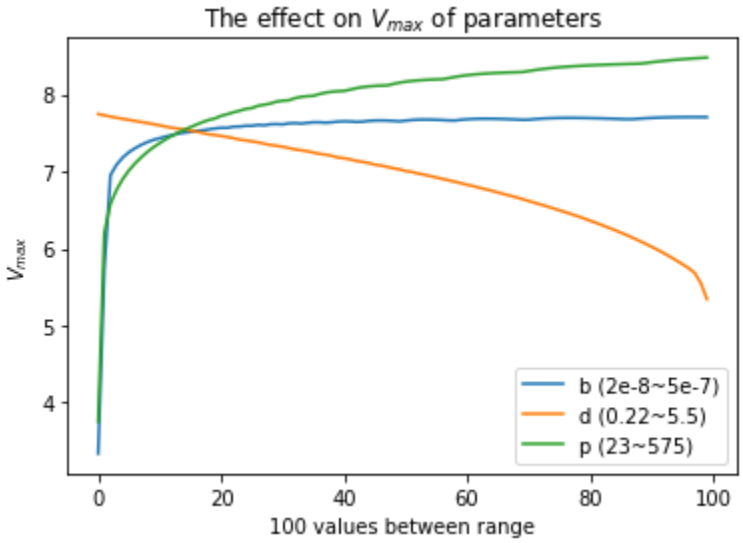


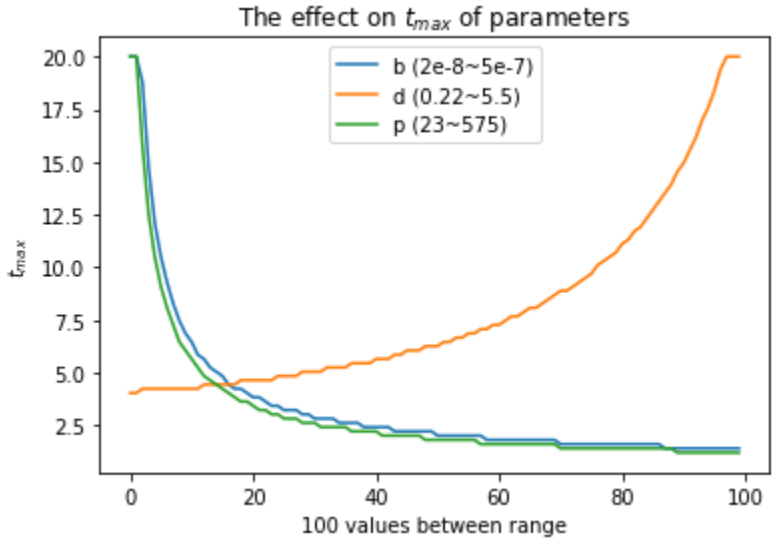


CELLULAR AND SYSTEMS MODELING - HW4

Student: Shu-Ting Cho (shc167@pitt.edu)

(c) Define functions to compute the peak viral load (V_{max}) and the time post infection at which peak load occurs (t_{max}). Make plots showing the effect on both V_{max} and t_{max} of varying each of the three parameters (b, d, p) about a five-fold range up and down varying one parameter at a time. For example vary *b* over the range 2e-8 to 5e-7. Which parameters have the strongest effect on each quantity? [*Hint: It may help to use interpolation to determine the value of t_{max} from the timecourses you compute so that it varies smoothly as each parameter is changed.*]





While setting the varying range for each of the three parameters to get around the same range of Vmax (4~8) and tmax (1~20), the range for b is the smallest (2e-8~5e-7). Thus, b has the strongest effect on both Vmax and tmax, changing b will change Vmax and tmax the most.

CELLULAR AND SYSTEMS MODELING - HW4

Student: Shu-Ting Cho (shc167@pitt.edu)

(d) The *sensitivity coefficients* are often used to characterize the sensitivities of a model's properties to changes in parameters. These are scaled derivatives that are defined as $S_p^Y = \frac{\partial \ln Y}{\partial \ln p} = \frac{p}{Y}\frac{\partial Y}{\partial p}$, where Y is the property of interest (here, V_{max} or t_{max}) and the is the parameter of interest. These are measures of local sensitivity that are computed at a single point the parameter space. Report the sensitivity coefficients for V_{max} and t_{max} for the model (b, d, p) at (1.0e-7, 1, 100).

Report the sensitivity coefficients:

$$S_b^{V_{max}} = \frac{b}{V_{max}}\frac{\partial V_{max}}{\partial b} = 0.022164091199234456$$

$$S_d^{V_{max}} = \frac{d}{V_{max}}\frac{\partial V_{max}}{\partial d} = -0.03276733843137881$$

$$S_p^{V_{max}} = \frac{p}{V_{max}}\frac{\partial V_{max}}{\partial p} = 0.08207005860466135$$

$$S_b^{t_{max}} = \frac{b}{t_{max}}\frac{\partial t_{max}}{\partial b} = -0.9378177809679603$$

$$S_d^{t_{max}} = \frac{d}{t_{max}}\frac{\partial t_{max}}{\partial d} = -0.011536799876461877$$

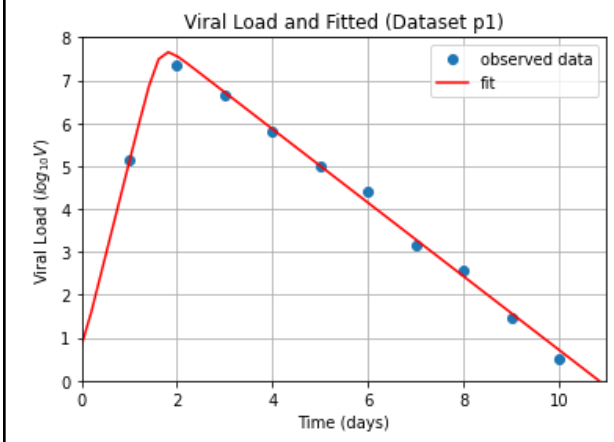
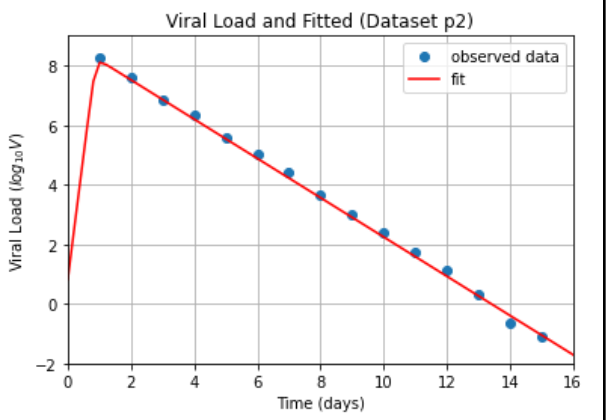
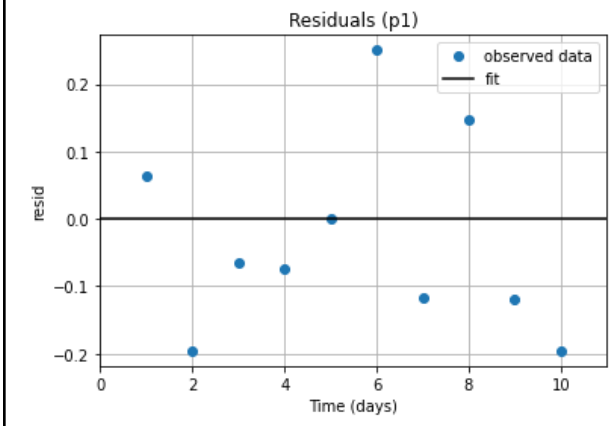
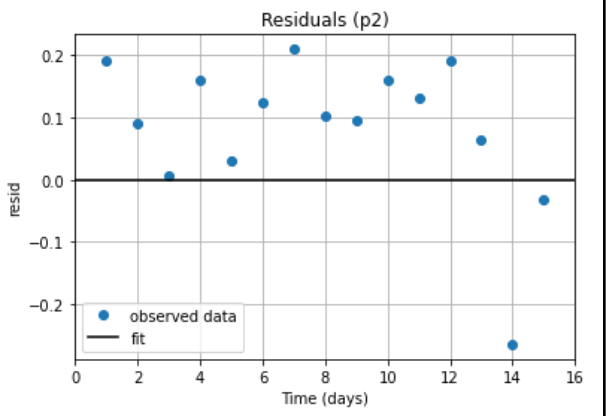
$$S_p^{t_{max}} = \frac{p}{t_{max}}\frac{\partial t_{max}}{\partial p} = -0.8544100492746549$$

CELLULAR AND SYSTEMS MODELING - HW4

Student: Shu-Ting Cho (shc167@pitt.edu)

5. Parameter estimation (10 points)

You will use data in the files p1.csv and p2.csv to perform separate fits of the model parameters *b*, *d*, and *p*. I suggest using the scipy function *least_squares*. It's not as convenient as the *curve_fit* routine shown in the class example because it requires you to define the residual function that is used to calculate χ^2, but you will need to write a function to compute χ^2 for this problem anyway. Perform fits of the data in both datasets and report the optimized values you find for the parameters and χ^2. Make separate plots of the residuals in each case and explain why you think the fit in each case is good or not.

# dataset p1:	# dataset p2:
log10(b): -6.54214800 d: 1.97396965 p: 118.096156 χ^2: 0.00775741 cost: 97.82284036629906	log10(b): -6.53481719 d: 1.51678374 p: 262.995232 χ^2: 0.01143879 cost: 151.01164161943368
	
	

The fits in both cases are good because residuals and χ^2 are small. The sum of residuals is 1.2285 for p1 dataset and 1.847 for p2 dataset.

CELLULAR AND SYSTEMS MODELING - HW4

Student: Shu-Ting Cho (shc167@pitt.edu)

6. Confidence intervals (10 points)

Use the rule-of-thumb given in class to estimate a single variance parameter for each dataset based on your best fit. Use these values to determine the 95% confidence intervals for each model parameter for each data set (χ^2=4) using both the covariance matrix returned by your least squares fitting routine and using the profile likelihood method. What aspects of the data are leading to some practical nonidentifiability for some parameters?

95% confidence interval borders

	p1 left	p1 right	p2 left	p2 right
log10_b	-6.67898	-6.36475	-inf	-6.44684
d	1.90367	2.03981	1.50039	1.53986
p	82.44572	163.51963	236.13662	+inf

For dataset p2, when log10(b) < -6.639634621145124, the probability doesn't change anymore, so the left border can go to -inf. Similarly, the probability stays the same for p > 273.88516906931864, so the right border can go to +inf.