

Mathematics 495
Statistical, Dynamical, and Computational Modeling
Test 2

Bardsley, Graham, and Johnson

November 18, 2008

Answer the questions and submit your solutions in the specified manner. The examination is “open book”, but you may not work with other students in the class. All solutions must be submitted by 25 November at 2:10 (in class). The number of points each portion is worth is indicated in the box to the left of the question.

- 25 1. The Gompertz growth function has the form

$$X(t) = Me^{-\ln(M/X_0)e^{-\beta t}}$$

- (a) Show that X is a solution of the differential equation

$$\frac{dX}{dt} = \beta X \ln(M/X), \quad X(0) = X_0.$$

- (b) Plotted is tumor growth data (in number of cancerous cells) from a mouse. It is contained on John Bardsley’s web site. Download it (`tumorgrowth.mat`) and then fit the Gompertz function/ODE to the data. Plot the data and model fit together.

Important Notes: (i) The ODE and analytic solution above assume that the initial time point is $t = 0$. However, the data begins at $t = 2$ days. Take care to incorporate this into your fitting problem. (ii) The optimizers (`nlinfit` and `fminsearch`), as well as the MCMC sampling you’ll do in part (c), work best if the parameters are of roughly the same order of magnitude. To accomplish this, I recommend rescaling M and X_0 as follows: $\tilde{M} = M/10^8$ and $\tilde{X}_0 = X_0/10^8$. Note that the analytic solution can be then be rewritten

$$X(t) = (10^8 \tilde{M})e^{-\ln(\tilde{M}/\tilde{X}_0)e^{-\beta t}}.$$

The optimization and sampling can then be done with respect to the parameters \tilde{M} and \tilde{X}_0 .

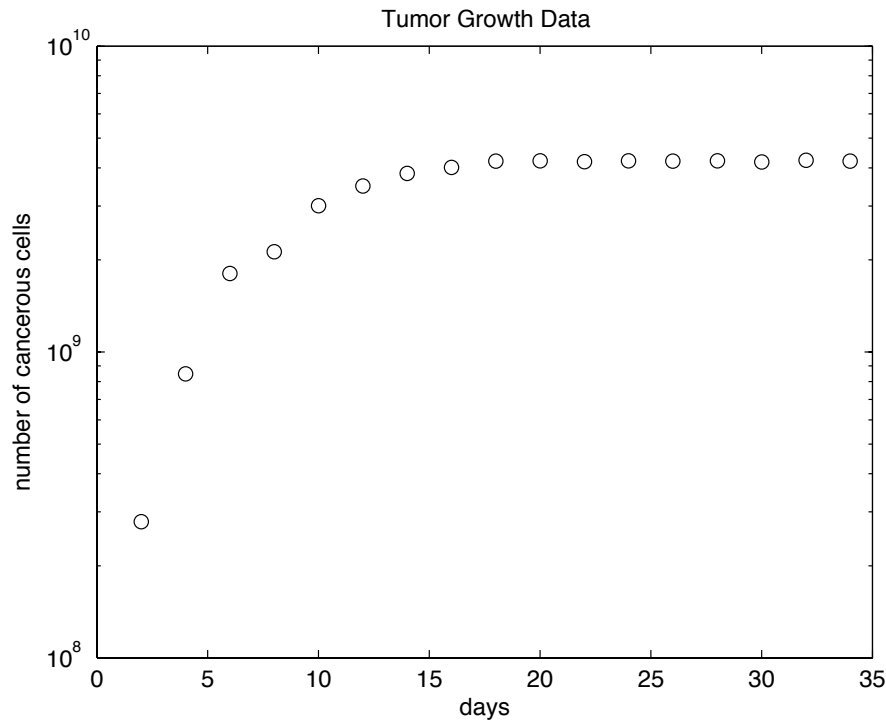


Figure 1: Tumor growth data (in number of cancerous cells) from a mouse. It is contained on John Bardsley's web site. Download it (`tumorgrowth.mat`).

- (c) Using Professor Haario's MCMC sampling codes, obtain 95% confidence intervals for your parameter estimates (M , X_0 , and β). Report these, as well as the mean-chain values and the optimal values from part (b). Provide a diagnostic that suggests your MCMC chain has converged. Plot the pairwise samples of the parameters. And finally, plot the corresponding 95% model prediction bands out to 40 days. Note that all that I have requested can be obtained using Professor Haario's codes (e.g., see `aidsrun.m` from John Bardsley's web site in the directory `MATLABcode/AidsExample/`).

Email all of the MATLAB m-files needed for running your codes in parts (a) and (b) (with the exception of Haario's codes) to John Bardsley. With your written work, hand in the plots requested in part (c) along with relevant commentary.

- 25 2. Perform an MCMC analysis of the problem from Jon Graham's handout on 11/17. The data is given in `blood.mat`, which you can get from either John or Jon's web sites, and the ODE model has the form

$$\frac{dy}{dt} = \theta_1 \exp(-\theta_1 t) - \theta_2 y, \quad y(0) = y_0.$$

Plot the data and model fit. Then, from the MCMC chain that you compute with Professor Haario's codes, obtain 95% confidence intervals for your parameter estimates (θ_1, θ_2 , and y_0). Report these, as well as the mean-chain values and the values from the optimal fit. Provide a diagnostic that suggests your MCMC chain has converged. Plot the pairwise samples of the parameters. And finally, plot the corresponding 95% model prediction bands out to 15 days.

Email all of the MATLAB m-files needed for running your codes (with the exception of Haario's codes) to John Bardsley. Also, hand in the requested plots along with relevant commentary.

- 25 3. Begin with the SIR equations describing a population held constant by a balance of births and deaths. Birth rate equals death rate equals μ .

$$\frac{dS}{dt} = \mu N - \beta SI - \mu S \quad (1)$$

$$\frac{dI}{dt} = \beta SI - (\gamma + \mu)I \quad (2)$$

$$\frac{dR}{dt} = \gamma I - \mu R \quad (3)$$

Do the following:

- Add a latent phase (E) of individuals that have been infected but are not yet infectious. That is, they carry the disease but do not transmit it to others. Write out the resulting system of ordinary differential equations. This is called the SEIR model.
- Find an expression for R_o for this model.
- Write a MATLAB program to solve this model numerically, for a disease with $R_o = 15$ (as estimated for measles) in a population of 3,000,000 and with the latent and infectious stages each lasting one week. Do numerical experiments on this model, with seasonal variation in contact rate, to test the claim that the latent proportion $E(t)/(E(t) + I(t))$ remains roughly constant.
- Seasonal variation in contact rate is often modeled with

$$\beta(t) = \hat{\beta}(1 + \phi \cos(2\pi t)) \quad (4)$$

with time measured in years. Values of ϕ for measles range between 0.1 and 0.3, depending on who you ask. Large values of ϕ stress the numerical algorithm, especially if $S(0) \sim N$. Explore a range of values of ϕ . Include a few well-chosen and well-designed graphs to support your claims.

Describe your

- methods*: describe the simulations that you conducted, provide source code;
- results*: give a verbal summary of the results;
- conclusions*: was the claim valid?

- 25 4. Reconsider the data on the cumulative number of polio cases over 12 months during the 1949 polio epidemic as given in the table below.

Month	0	1	2	3	4	5	6	7	8	9	10	11
Cases	494	759	1016	1215	1619	2964	8489	22377	32618	38153	41462	42375

In problem 1 on homework #8, you were asked to fit these data with the standard SIR model where R corresponded to the cumulative number of polio cases at time t . Consider again the SIR model shown to the right with a slight modification where the parameters (a, b) are each divided by 1000 to reduce the scale differences between the initial conditions (S_0, I_0, R_0) and these parameters. This will greatly improve our ability to numerically estimate the Jacobian matrix.

$$\begin{aligned}\frac{dS}{dt} &= -(a/1000)IS, & S(0) &= S_0, \\ \frac{dI}{dt} &= (a/1000)IS - (b/1000)I, & I(0) &= I_0, \\ \frac{dR}{dt} &= (b/1000)I, & R(0) &= R_0.\end{aligned}$$

- (a) Modify the bootstrap code provided by Jon Graham in the lab on Monday, November 17 (code can be found on the course webpage), to fit these data with the modified SIR model shown to the right and bootstrap the estimates of the five parameters (a, b, S_0, I_0, R_0) . Report the model estimates, their bootstrap standard errors, estimates of bias of the parameter estimates, and 95% bootstrap confidence intervals for the parameters. Use at least 200 bootstrap samples. Reasonable starting values for the five parameters are: $(45000, 0.09, 494, 0.07, 975)$. These took a little trial and error, so don't modify these much if at all.
- (b) Write a paragraph summarizing this output and comparing the results to those obtained using MCMC in problem 1(c) of homework #8.

In modifying the bootstrap code, you will want to be aware of the following items:

- Remember to remove the places where a "0" is appended to the times and to change the vectors which run from "2:nout" to "1:nout" since these data begin with month 0. You will need to do this in **bootoptimize** and **bootNLM**.
- In creating your SIR function, remember to divide the parameters (a, b) by 1000.
- Because this model is more complex than those we bootstrapped in class, add the line **opts = optimset('MaxFunEvals',5000)** before the bootstrap loop in **bootNLM.m**. You will then need to add **opts** as the last argument in the **fminsearch** call within the bootstrap loop. This is necessary to allow the optimization routine a sufficient number of iterations to converge.

- Finally, in **bootoptimize.m**, you will need to change **yout** to **yout(:,3)** in the **sse** line to compute the SSE using the *recovered* values, since R is the cumulative number of polio cases. In **bootNLM.m**, the “y” returned by the ODE solver (called twice in this function) will have three columns for the three initial conditions. Again, you will only want to use the 3rd column **y(:,3)** in the lines that follow the solver.

This quiz has 4 questions worth a total of 100 points.
--