

## 02-512 Assignment 06

Karan Sikka

ksikka@cmu.edu

November 26, 2014

---

1

---

(a) The variables we are trying to find are the rate of infection,  $\lambda_1$  and the rate of recovery,  $\lambda_2$ .

We can run the CTMM as a simulation and compare it with real data. The output of the simulation will be sequence of states over time where each state will be  $(S_t, I_t, R_t)$ .

Let  $(Sr_t, Ir_t, Rr_t)$  be the real data. One possible objective function to minimize is

$$L(\lambda_1, \lambda_2) = \sum_{\text{realdatapoints}} (S_t - Sr_t)^2 + (I_t - Ir_t)^2 + (R_t - Rr_t)^2$$

You can use steepest/gradient descent, Newton-Raphson's method, or a similar algorithm to find parameters yielding a local minimum. Rather than analytically computing the gradient, you'd have to approximate it using finite difference methods. Performance may be a concern, depending on how long the simulation has to run for.

(b) Let  $G$  be the growth rate,  $x_1$  be the conc of nutrient 1  $x_2$  be the conc of nutrient 2.

$$G = \theta_1 x_1^2 + \theta_2 x_1 + \theta_3 x_2^2 + \theta_4 x_2 + \theta_5$$

Where  $\vec{\theta}$  are the parameters we're trying to estimate.

Let  $Gr(x_1, x_2)$  be the experimenally determined growth rate.

One possible objective function to minimize is

$$L(x_1, x_2) = \sum (Gr(x_1, x_2) - G(x_1, x_2))^2$$

You can use steepest descent once again, like in part a.

(c) Call parameters we are estimating,  $f_B$  and  $f_b$ , which are the frequencies of the B allele and b allele respectively.

Let  $B$  be the number of people observed with brown eyes, and  $b$  be the number of people observed with blue eyes.

The likelihood function is as follows:

$$Pr(B = B, b = b | p = p) =$$

---

2

---

(a) Say there are  $m$  biomarkers. Let  $\vec{\theta}$  be an  $m + 1$  dimensional vector of parameters.

Let  $\mu = \theta_{m+1} + \sum_{i=1}^m \theta_i x_i$  in the following

$$L(\mu, \sigma^2; \theta) = \frac{1}{2\pi\sigma^2} \left( \frac{1}{e} \right)^{\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2}}$$

(b) Say there are  $m$  biomarkers. Let  $\vec{\theta}$  be an  $2m + 1$  dimensional vector of parameters.

Let  $\mu = \theta_{2m+1} + \sum_{i=1}^m \theta_i x_i^2 + \sum_{i=m+1}^{2m} \theta_i x_i$  in the same likelihood function from part a.

(c) Performance, over/underfitting.

(d) Metropolis

(e) Sampling vs solving

### 3

Given two points  $(t1, x1), (t2, x2)$ , the eqn of a line for general  $t, x$  is as follows:

$$t - t2 = (x2 - x1)/(t2 - t1) * (x - x2)$$

The the following is the piecewise linear function interpolation, found by plugging in values from the table into the above equation:

If  $0 \leq t < 2$ ,  $t - 2 = (5/2)(x - 5)$

If  $2 \leq t < 5$ ,  $t - 5 = ((6 - 5)/(5 - 2))(x - 6)$

If  $5 \leq t < 8$ ,  $t - 8 = ((10 - 6)/(8 - 5))(x - 10)$

If  $8 \leq t \leq 10$ ,  $t - 10 = ((20 - 10)/(8 - 10))(x - 20)$

(b) Let  $(t_1, x_1) = (0, 0), (t_2, x_2) = (2, 5), \dots, (t_5, x_5) = (10, 20)$

Let  $[5]$  be short notation for  $1, 2, 3, 4, 5$ .

Then:

$$x = \sum_{i=1}^5 \frac{\prod_{j \in [5]: j \neq i} (t - t_j)}{\prod_{j \in [5]: j \neq i} (t_i - t_j)} x_i$$

(c) We have 4 quadratic equations of the form

$$S_{i,i+1}(t) = c_{i,0} + c_{i,1}t + c_{i,2}t^2$$

for  $i = 1$  to  $i = 4$ . The derivative of each equation is of the form

$$\frac{dS_{i,i+1}}{dt} = c_{i,1} + 2c_{i,2}t$$

.

Given the constraints of the problem, we can solve for all  $3 * 4 = 12$  parameters by expressing the constraints as equations:

$$S_{1,2}(0) = 0$$

$$S_{1,2}(2) = 5$$

$$S_{2,3}(2) = 5$$

$$S_{2,3}(5) = 6$$

$$S_{3,4}(5) = 6$$

$$S_{3,4}(8) = 10$$

$$S_{4,5}(8) = 10$$

$$S_{4,5}(10) = 20$$

Constraints for the derivative continuity:

$$\frac{dS_{1,2}}{dt}(2) = \frac{dS_{2,3}}{dt}(2)$$

$$\frac{dS_{2,3}}{dt}(5) = \frac{dS_{3,4}}{dt}(8)$$

$$\frac{dS_{3,4}}{dt}(8) = \frac{dS_{4,5}}{dt}(10)$$

Additional constraint:

$$\frac{dS_{1,2}}{dt}(0) = 0$$

Now we have 12 parameters and 12 constraints, we represent it as a linear system and solve using a gaussian elimination.

$$\begin{bmatrix} 1 & t_1 & t_1^2 & & & & & & & & & \\ 1 & t_2 & t_2^2 & & & & & & & & & \\ 0 & 1 & 2t_2 & 0 & -1 & -2t_2 & & & & & & \\ & & & 1 & t_2 & t_2^2 & & & & & & \\ & & & 1 & t_3 & t_3^2 & & & & & & \\ & & & 0 & 1 & 2t_3 & 0 & -1 & -2t_3 & & & \\ & & & & & & 1 & t_3 & t_3^2 & & & \\ & & & & & & 1 & t_4 & t_4^2 & & & \\ & & & & & & 0 & 1 & 2t_4 & 0 & -1 & -2t_4 \\ & & & & & & & & & 1 & t_4 & t_4^2 \\ & & & & & & & & & 1 & t_5 & t_5^2 \\ 0 & 1 & 2t_1 & & & & & & & & & \end{bmatrix} \begin{bmatrix} c_{1,0} \\ c_{1,1} \\ c_{1,2} \\ c_{2,0} \\ c_{2,1} \\ c_{2,2} \\ c_{3,0} \\ c_{3,1} \\ c_{3,2} \\ c_{4,0} \\ c_{4,1} \\ c_{4,2} \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ 0 \\ x_2 \\ x_3 \\ 0 \\ x_3 \\ x_4 \\ 0 \\ x_4 \\ x_5 \\ 0 \end{bmatrix}$$

---

4

---

(a)

If  $b_i$  is 0, there are no boojum on island  $i$ . Based on this observation, we note the following:

$$\begin{aligned} Pr(b_i = 0|f) &= (1 - f)^{s_i} \\ Pr(b_i = 1|f) &= 1 - Pr(b_i = 0|f) = 1 - (1 - f)^{s_i} \\ Pr(b|f) &= \prod_{i=1}^n Pr(b_i = b_i) \end{aligned}$$

(b)

$$\hat{f} = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n s_i}$$

(c)

$$E[y_i] = b_i * \hat{f} * s_i$$

(d)

Submitted online

(e)

TODO