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, λ_1 and the rate of recovery, λ_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

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

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 following

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

(c) Performance, over/underfitting.

(d) Metropolis

(e) Sampling vs solving

3

Given two points, the eqn of a line is derived as follows:

(Points given as t2 x2 t1 x1)

t = (x2 - x2)/(t2 - t1) x + b

t1 - (x2 - x2)/(t2 - t1) x1 = b

or

t2 - (x2 - x2)/(t2 - t1) x2 = b

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

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) TODO interp formula

(c) TODO interp formula + some deriv stuff.

4

(a)

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

$$Pr(b_i = 0|f) = (1 - f)^{s_i}$$

$$Pr(b_i = 1|f) = 1 - Pr(b_i = 0|f) = 1 - (1 - f)^{s_i}$$

$$Pr(b|f) = \prod_{i=1}^{n} Pr(b_i = b_i)$$

(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