

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC BÁCH KHOA
KHOA KHOA HỌC - KỸ THUẬT MÁY TÍNH



MÔ HÌNH HÓA TOÁN HỌC (CO2011)

Bài tập lớn

“Mô hình SIR trong dự báo COVID-19”

GVHD: Nguyễn An Khương

Nguyễn Tiến Thịnh

SV thực hiện: Nguyễn Huỳnh Hữu Khiêm – 1810998

Trần Văn B – 88471475

Lê Thị C – 36811334

Phạm Ngọc D – 97501334

Kiều Thị E – 12341334

Tp. Hồ Chí Minh, Tháng 7/2020



Mục lục

1 Bài toán 1	2
1.1 Mô hình SIR rời rạc	2
1.2 Mô hình SIR liên tục	5
1.2.1 Phương pháp xấp xỉ Euler	6
1.3 Những vấn đề liên quan đến mô hình SIR	6
1.4 Mở rộng của mô hình SIR - mô hình SIRD	7
Tài liệu	8

1 Bài toán 1

Trình bày lại chi tiết cách xây dựng mô hình SIR (cả trường hợp rời rạc lẫn liên tục) hoặc mở rộng của nó và những vấn đề liên quan.

1.1 Mô hình SIR rời rạc

Hãy xem xét một căn bệnh đang lan rộng khắp nước Mỹ, chẳng hạn như bệnh cúm mới. Các Trung tâm kiểm soát và phòng ngừa dịch bệnh quan tâm đến việc tìm hiểu và thử nghiệm với một mô hình cho căn bệnh mới này trước khi nó trở thành một dịch bệnh thực sự. Ta chia dân số thành ba loại: miễn cảm, nhiễm bệnh và đã loại bỏ. Chúng ta làm các giả định sau đây cho mô hình:

- Không ai vào hoặc ra khỏi cộng đồng, và không có liên hệ bên ngoài cộng đồng.
- Mỗi người đều dễ mắc bệnh S (có thể mắc bệnh cúm mới này); bị nhiễm bệnh I (hiện có cảm cúm và có thể lây cúm); hoặc loại bỏ R (đã hết nhiễm và sẽ không tái phát, trong đó bao gồm trường hợp tử vong).
- Ban đầu, mỗi người trong cộng đồng chỉ có thể là S hoặc I. Một khi ai đó đã bị cúm, người đó không thể mắc lại.
- Giả sử độ dài trung bình của bệnh là $5/3$ tuần, I là người bị coi là bị nhiễm bệnh và có thể truyền bệnh.
- Khoảng thời gian của chúng ta xem xét cho mô hình là mỗi tuần.

Mô hình mà chúng ta sẽ xem xét liên quan đến những người dễ mắc bệnh, những người bị nhiễm bệnh và những người đã hồi phục, thường được gọi là mô hình SIR. Ta định nghĩa cho các biến như sau:

$S(t)$ = số người trong dân số dễ mắc bệnh sau thời điểm t
 $I(t)$ = số người trong dân số đã mắc bệnh và có khả năng lây bệnh tại thời điểm t
 $R(t)$ = số người trong dân số đã không còn mắc bệnh (bao gồm cả trường hợp tử vong)

Hãy bắt đầu mô hình của chúng ta bằng việc xử lý $R(t)$. Giả định của chúng ta có thời gian người trong giai đoạn cúm là $5/3$ (1 và $2/3$) tuần. Như vậy, $3/5$ số người nhiễm bệnh sẽ được loại bỏ mỗi tuần:

$$R(n+1) = R(n) + 0.6I(n)$$

Giá trị 0,6 được gọi là tỷ lệ loại bỏ mỗi tuần. Tỷ lệ loại bỏ đại diện cho tỷ lệ của những người bị nhiễm được loại bỏ khỏi nhiễm bệnh mỗi tuần. Nếu dữ liệu thực có sẵn ta sẽ lấy nó, nếu không sau đó chúng ta có thể thực hiện "phân tích dữ liệu" để lấy tỷ lệ loại bỏ này.

$I(t)$ sẽ có khả năng tăng và giảm số lượng của chính nó theo thời gian. $I(t)$ bị giảm bởi số lượng bị loại bỏ mỗi tuần: $-0.6 * I(t)$. Và $I(t)$ được tăng lên bởi số người nhạy cảm tiếp xúc với một người bị nhiễm bệnh và mắc bệnh, $a * S * I(t)$. Chúng ta xác định tỷ lệ a là tỷ lệ mà bệnh lây lan, hoặc hệ số truyền. Chúng ta nhận ra đây là một xác suất hệ số. Ban đầu, chúng ta sẽ cho rằng tỷ lệ này là một giá trị không đổi có thể được tìm thấy từ điều kiện ban đầu. Hãy giả sử minh họa sau: Giả sử chúng ta có 1000 sinh viên cư trú các ký túc xá. Y tá đã tìm thấy 5 học sinh báo cáo cho bệnh xá ban đầu: $I(0)=5$ và $S(0)=995$. Sau một tuần, tổng số người bị cúm là 9. Chúng ta tính a như sau:

$$I(0) = 5, I(1) = I(0) - 0.6 * I(0) + aI(0) * S(0)$$

$$I(1) = 9 = 5 - 3 + a * 5 * 995$$

$$7 = a(4957)$$

$$a = 0,001407$$

Hãy xem xét $S(n)$. Con số này chỉ giảm bởi số lượng bị nhiễm bệnh. Chúng ta có thể sử dụng tỷ lệ a như trước đây để có được mô hình:

$$S(n+1) = S(n) - aS(n)I(n)$$

Khi đó, chúng ta thu được mô hình kết hợp:

$$R(n+1) = R(n) + 0.6I(n)$$

$$I(n+1) = I(n) - 0.6I(n) + 0.001407I(n)S(n)$$

$$S(n+1) = S(n) - 0.001407S(n)I(n)$$

$$I(0) = 5, S(0) = 995, R(0) = 0$$

Phương trình mô hình SIR trên, có thể được giải quyết lặp lại và xem xét bằng đồ thị. Ta lặp lại giải pháp và thu được biểu đồ để quan sát hành vi để có được một số thông tin cần biết.

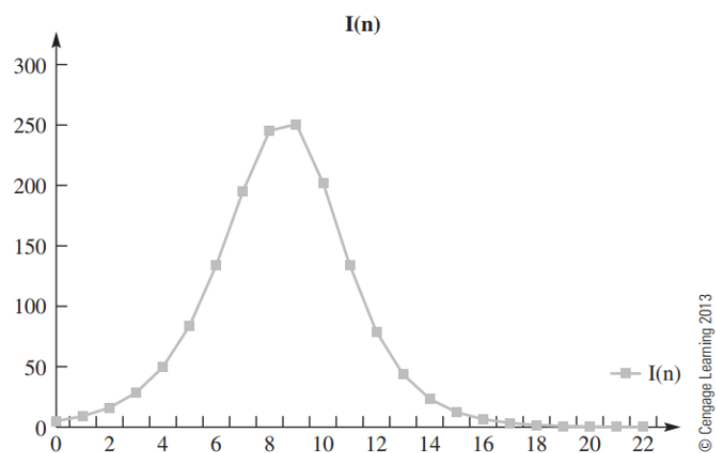
Hình 1: Biểu đồ phân bố số liệu SIR rời rạc

Week	$R(n)$	$I(n)$	$S(n)$
0	0	5	995
1	3	8.999825	988.0002
2	8.399895	16.11073	975.4894
3	18.06633	28.55649	953.3772
4	35.20023	49.72832	915.0714
5	65.03722	83.91682	851.046
6	115.3873	134.0505	750.5621
7	195.8176	195.1831	608.9993
8	312.9275	245.3182	441.7543
9	460.1184	250.6044	289.2772
10	610.481	202.241	187.2779
11	731.8256	134.1869	133.9874
12	812.3378	78.97173	108.6905
13	859.7208	43.66564	96.61352
14	885.9202	23.40195	90.67782
15	899.9614	12.34649	87.69211
16	907.3693	6.46194	86.16877
17	911.2465	3.368218	85.38533
18	913.2674	1.751936	84.98068
19	914.3185	0.910249	84.7712
20	914.8647	0.472668	84.66264
21	915.1483	0.245372	84.60633
22	915.2955	0.127358	84.57712

Giải thích: Chúng ta phân tích cả bảng số và Hình 1.31 Lờil.34 để kiểm tra những gì đang xảy ra trong dịch. Đỉnh điểm của dịch cúm xảy ra vào khoảng tuần 9, trong đó đồ thị của những người bị nhiễm đạt giá trị tối đa. Số lượng tối đa là lớn hơn 250 một chút; nhìn từ bảng ta thấy là 250.6044. Sau khoảng 17 tuần, ít nhất 85 người vẫn chưa bị cúm.

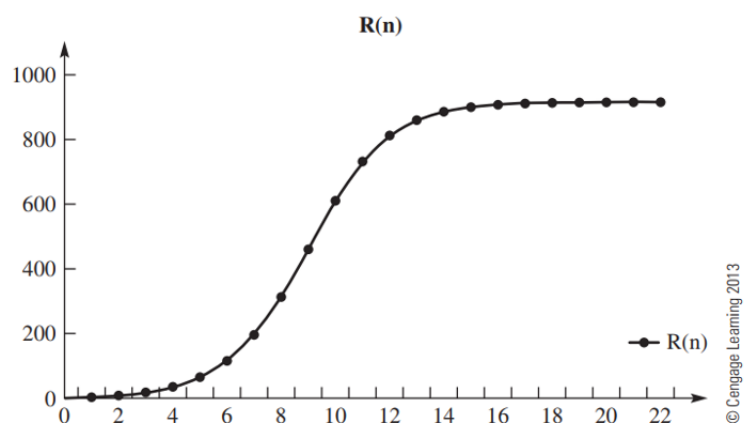
■ **Figure 1.31**

Plot of $I(n)$ versus time



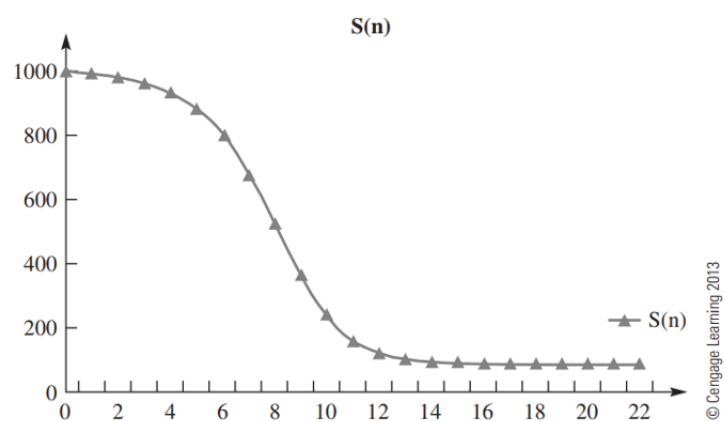
■ **Figure 1.32**

Plot of $R(n)$ over time



■ **Figure 1.33**

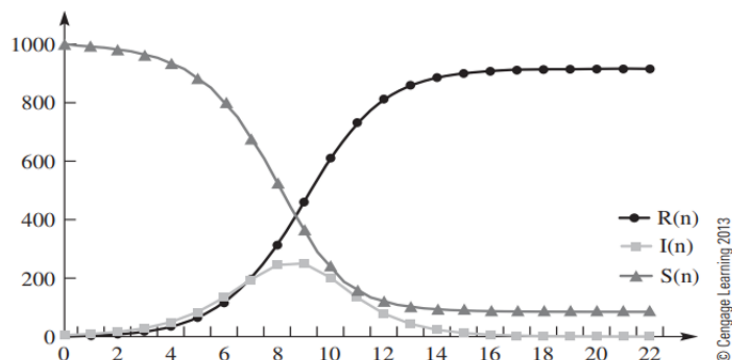
Plot of $S(n)$ over time



Hình 2

Figure 1.34

Plot of $I(n)$, $S(n)$, and $R(n)$ overlaid together



1.2 Mô hình SIR liên tục

Mô hình SIR (Susceptible - Infectious - Recovered) là một trong các mô hình cách ly cơ bản được sử dụng nhiều nhất trong quá khứ và hiện tại để mô tả dịch bệnh. Mô hình được lần đầu tiên phát biểu vào thế kỷ 20. Mô hình thể hiện ba trạng thái (Có nguy cơ mắc bệnh - Mắc bệnh - Hồi phục) cho một nhóm người được cách ly với giả thiết rằng sẽ miễn dịch với bệnh nếu đã phục hồi. Mô hình SIR là một hệ động lực gồm ba phương trình vi phân sau:

$$\begin{aligned}\frac{dS}{dt} &= -\frac{\beta}{N}IS \\ \frac{dI}{dt} &= \frac{\beta}{N}IS - \gamma I \\ \frac{dR}{dt} &= \gamma I\end{aligned}$$

trong đó tại mỗi thời điểm $t \geq t_0 \geq 0$ với t_0 là thời điểm đầu ghi nhận,

$S(t)$: Số người có nguy cơ mắc bệnh

$I(t)$: Số người nhiễm bệnh

$R(t)$: Số người phục hồi sau bệnh

$\beta(t)$: Tỷ lệ tiếp xúc của mỗi người trong nhóm $S(t)$ với người trong nhóm $I(t)$

$\gamma(t)$: Tỷ lệ hồi phục khi mắc bệnh

$N(t)$: Tổng số người trong cộng đồng bị cách ly được tính bằng: $N(t) := S(t) + I(t) + R(t)$

Hệ phương trình vi phân trên có thể được hiểu như sau:

- Phương trình (1) thể hiện sự suy giảm số người có nguy cơ mắc bệnh tại thời điểm $t \geq t_0$. Sự suy giảm được tính theo xác suất lây bệnh khi có tiếp xúc giữa nhóm $S(t)$ và nhóm $I(t)$;
- Phương trình (2) thể hiện độ biến thiên số người mắc bệnh tại thời điểm $t \geq t_0$. Sự biến thiên này được tính bằng cách lấy số người ở nhóm $S(t)$ đã bị lây nhiễm sau khi tiếp xúc với người bệnh nhóm $I(t)$ và trừ đi số người ở nhóm $I(t)$ đã phục hồi với tỷ lệ $\gamma I(t)$;
- Phương trình (3) thể hiện số người đã hồi phục từ nhóm $I(t)$ theo tỷ lệ hồi phục là γ .

Lấy minh họa với các số liệu tương tự như mô hình SIR rời rạc, ta có hệ các phương trình kết hợp sau:

$$\begin{aligned}\frac{dR}{dt} &= 0.6I(t) \\ \frac{dI}{dt} &= -0.6I(t) + 0.001407I(t)S(t) \\ \frac{dS}{dt} &= 0.001407I(t)S(t)\end{aligned}$$

$$I(0)=5, S(0)=995, R(0)=0$$

Chúng ta sẽ giải mô hình SIR liên tục này bằng phương pháp xấp xỉ Euler.

1.2.1 Phương pháp xấp xỉ Euler

Phương pháp Euler là một phương pháp bậc một thường được sử dụng trong việc giải các phương trình vi phân thường. Phương pháp được đặt tên theo Leonhard Euler, người đã giới thiệu phương pháp trong quyển sách Institutionum Calculi Integralis cùng tên xuất bản trong khoảng thời gian 1768 đến 1770. Giả sử ta có phương trình vi phân bậc nhất:

$$y' = f(t, y(t))$$

Khi đó ý tưởng của phương pháp Euler là xấp xỉ nghiệm y bằng dãy y_n sao cho

$$y_{n+1} := y_n + f(t_n, y_n) \Delta t$$

với Δt là xấp xỉ đủ nhỏ và $f(t_n, y_n)$ là độ dốc của đường cong y tính tại thời điểm t

Ở dạng tổng quát, một hệ phương trình vi phân được viết dưới dạng

$$y'_1 = f_1(t, y_1, \dots, y_N)$$

\vdots

$$y'_N = f_N(t, y_1, \dots, y_N)$$

Khi đó, áp dụng xấp xỉ Euler vào mô hình SIR liên tục ta được:

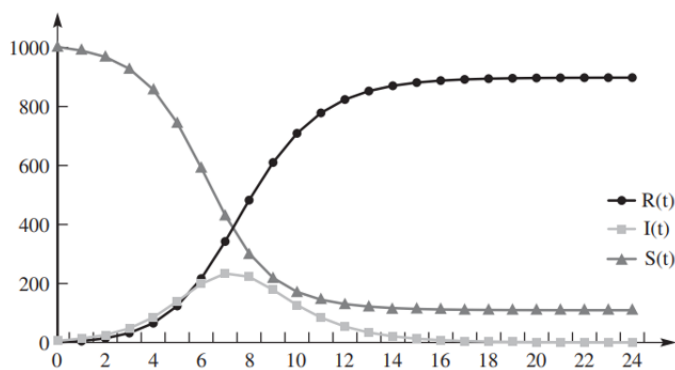
$$S_n = S_{n-1} - b S_{n-1} I_{n-1} \Delta t$$

$$I_n = I_{n-1} + (b S_{n-1} I_{n-1} - k I_{n-1}) \Delta t$$

$$R_n = R_{n-1} + k I_{n-1} \Delta t$$

Sau khi giải bằng phương pháp Euler với $\Delta t = 0,5$ ta được kết quả sau:

Week	$R(t)$	$I(t)$	$S(t)$
0	0	5	995
1	4.29623	13.0107	982.693
2	13.6301	25.0714	961.299
3	31.503	47.0769	921.42
4	64.6671	84.2088	851.124
5	122.74	138.163	739.097
6	214.755	197.109	588.136
7	339.747	231.908	428.346
8	478.985	221.114	299.902
9	605.748	176.888	217.364
10	704.076	125.809	170.115
11	772.755	83.435	143.81
12	817.839	53.1578	129.003
13	846.397	33.0964	120.506
14	864.118	20.332	115.55
15	874.983	12.3924	112.624
16	881.598	7.51803	110.884
17	885.608	4.54821	109.844
18	888.033	2.74696	109.22
19	889.497	1.6574	108.845
20	890.381	0.99941	108.62
21	890.913	0.60242	108.484
22	891.234	0.36305	108.403
23	891.428	0.21876	108.354
24	891.544	0.13181	108.324



■ Figure 12.31

SIR model

1.3 Những vấn đề liên quan đến mô hình SIR

*Hệ số R_0 :

Một trong những đại lượng quan trọng nhất đối với một mô hình dịch bệnh là hệ số lây nhiễm cơ bản, hay thường gọi là “hệ số R_0 ”. Nếu $R_0 < 1$, dịch sẽ tắt trước khi kịp bùng phát, còn nếu $R_0 > 1$, dịch sẽ bùng phát. Thí dụ, trong mô hình SIR đơn giản ở trên, $R_0 = N/\gamma$, trong đó N là số người khỏe mạnh trung bình mà một người mắc bệnh có thể lây cho trong khoảng thời gian mắc bệnh $1/\gamma$. Một cách trực giác, điều này khá hợp lý: nếu trung bình một người mắc bệnh lây cho nhiều hơn một người thì số người mắc bệnh phải tăng (theo cấp số nhân), còn nếu trung bình một người mắc bệnh lây cho ít hơn một người khác thì số người mắc bệnh phải giảm dần.

Ước lượng này có vai trò rất quan trọng và được thể hiện trên hệ số $R_0 := \frac{\beta}{\gamma}$

1.4 Mở rộng của mô hình SIR - mô hình SIRD

Chúng ta xem xét một cộng đồng N người không đổi và không có sự dịch chuyển giữa bên trong và bên ngoài. Khi đó, mô hình SIRD đặt mỗi người trong cộng đồng vào một trong năm trạng thái sau:

$$S_t + I_t + R_t + D_t + C_t = N_t$$

với:

S_t = Susceptible

I_t = Infectious

R_t = Resolving

D_t = Dead

C_t = ReCovered

Lưu ý: R trong mô hình SIRD là Resolving khác với R trong mô hình SIR là Resolved. Ý nghĩa thực tế thể hiện số người được quan tâm chữa trị.

Sau khi thời kỳ lây nhiễm của người đó kết thúc, người đó ở trong trạng thái đã giải quyết vấn đề, R . Một hằng phân số θ của những người thoát khỏi trạng thái này mỗi kỳ (tức là sau thời gian chữa trị và không còn khả năng lây nhiễm) sẽ rơi vào một trong hai trường hợp (Tử vong – Bình phục):

Death: fraction δ

Recovery: fraction $1-\delta$

Trong các mô hình phiên bản trước đây của dự án này, chúng ta thấy rằng điều quan trọng là phải có một mô hình phân biệt giữa thời kỳ truyền nhiễm và thời gian phục hồi. Sự khác biệt này là chìa khóa để khớp dữ liệu với các giá trị tham số hợp lý về mặt sinh học khi chúng ta đặt các hạn chế trên đường thời gian của β

Ta có hệ mô hình SIRD được thể hiện đầy đủ dưới đây:

$$\Delta S_{t+1} = -\beta_t S_t I_t / N$$

$$\Delta I_{t+1} = \beta_t S_t I_t / N - \gamma I_t$$

$$\Delta R_{t+1} = \gamma I_t - \theta R_t$$

$$\Delta D_{t+1} = \delta \theta R_t$$

$$\Delta C_{t+1} = (1 - \delta) \theta R_t$$



Tài liệu

- [Dal] Dalgaard, P. *Introductory Statistics with R*. Springer 2008.
- [K-Z] Kenett, R. S. and Zacks, S. *Modern Industrial Statistics: with applications in R, MINITAB and JMP*, 2nd ed., John Wiley and Sons, 2014.
- [Ker] Kerns, G. J. *Introduction to Probability and Statistics Using R*, 2nd ed., CRC 2015.