

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



MÔ HÌNH HOÁ 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 Lâm Khánh Quỳnh – 1813776
Nguyễn Văn Quý – 1813760
Nguyễn Lương Hoài Sơn – 1813854
Nguyễn Duy Thanh – 1813967
Nguyễn Đình Thanh – 1813968

Tp. Hồ Chí Minh, Tháng 7/2020



Mục lục

1 Kiến thức và kết quả chuẩn bị	2
1.1 Tổng quan và bối cảnh đại dịch COVID-19	2
1.2 Các mô hình giả lập dịch bệnh	2
1.3 Hệ động lực và Phương trình vi phân thường	3
1.4 Phương pháp suy luận Bayesian	3
2 Các bài toán	4
2.1 Bài toán 1	4
2.1.1 Các tỉ lệ chuyển đổi	4
2.1.2 Mô hình SIR rời rạc	4
2.1.3 Mô hình SIR liên tục	5
2.1.4 Hệ số lây nhiễm cơ bản R_0	5
2.1.5 Mô hình SIS	6
2.1.6 Mô hình SIRD	7
2.2 Bài toán 2	8
2.2.1 Thuật toán Euler	8
2.3 Bài toán 3	12
2.3.1 Thuật toán Metropolis-Hastings	12
2.4 Bài toán 4	16
2.4.1 Ước lượng hệ số lây nhiễm cơ bản R_0 tại Việt Nam	17
3 Kết luận	19
Tài liệu	19

1 Kiến thức và kết quả chuẩn bị

1.1 Tổng quan và bối cảnh đại dịch COVID-19

Đại dịch COVID-19 là một đại dịch bệnh truyền nhiễm do chủng mới của virus Corona (SARS-CoV-2) gây ra, khởi nguồn vào cuối năm 2019 với tâm dịch đầu tiên được ghi nhận tại thành phố Vũ Hán (Wuhan) thuộc miền trung Trung Quốc. Đến thời điểm hiện tại, đại dịch COVID-19 đã lây lan trên 188 quốc gia và vùng lãnh thổ trên khắp thế giới, với hơn 12 triệu ca nhiễm đã được ghi nhận và khiến hơn 600 nghìn người tử vong.

Virus SARS-CoV-2 chủ yếu lây lan qua các giọt bắn trong không khí khi một cá nhân bị nhiễm bệnh ho hoặc hắt hơi trong phạm vi 2 mét. SARS-CoV-2 có khả năng gây nên các tổn thương nghiêm trọng ở hệ hô hấp của người bị lây nhiễm. Các triệu chứng được báo cáo bao gồm sốt, mệt mỏi, ho khan. Trên các ca nhiễm nặng còn ghi nhận thêm các triệu chứng như khó thở, suy hô hấp. Tuy nhiên, điều đáng lưu ý là 25% số người bệnh có thể không có triệu chứng gì hoặc các triệu chứng xuất hiện không rõ ràng.

Ca nhiễm SARS-CoV-2 đầu tiên tại Việt Nam được xác nhận vào ngày 23 tháng 1 năm 2020. Tính đến 18h00 ngày 17 tháng 7, Việt Nam ghi nhận tổng cộng 382 ca nhiễm, trong đó có 357 bệnh nhân đã được xuất viện và chưa ghi nhận trường hợp tử vong nào do nhiễm SARS-CoV-2.

Với sự vào cuộc quyết liệt của cả hệ thống chính trị cũng như toàn thể người dân, nhiều biện pháp triệt để nhằm phòng chống sự lây lan của đại dịch COVID-19 đã được triển khai trên phạm vi cả nước. Việt Nam đã thực hiện các biện pháp cách ly, theo dõi và hạn chế người đến từ vùng có dịch, đóng cửa biên giới và triển khai việc thực hiện khai báo y tế toàn dân. Nhiều hoạt động tập trung đông người tại các địa phương bị hạn chế, trường học tạm thời đóng cửa nhằm giảm thiểu rủi ro lây nhiễm SARS-CoV-2 trong cộng đồng.

Với những biện pháp quyết liệt và cấp bách kể trên, Việt Nam đã cơ bản kiểm soát tốt dịch COVID-19, trở thành một trong những quốc gia hiếm hoi trên thế giới cơ bản ngăn chặn thành công đà lây nhiễm của SARS-CoV-2 trong cộng đồng. Việt Nam đã vượt qua giai đoạn khó khăn, tiến hành khôi phục các hoạt động sản xuất, kinh doanh, triển khai các kế hoạch phát triển kinh tế - xã hội trong tình hình mới.

Dù vậy, tình hình dịch COVID-19 trên thế giới vẫn đang diễn biến hết sức phức tạp. Trong bối cảnh đó, các mô hình giả lập và dự báo dịch bệnh đóng vai trò quan trọng trong việc đánh giá tình hình dịch bệnh và triển khai các biện pháp phòng chống sự lây lan của dịch bệnh trong cộng đồng.

1.2 Các mô hình giả lập dịch bệnh

Các mô hình toán học có thể được sử dụng để mô phỏng cách một dịch bệnh lây lan trong cộng đồng, qua đó đưa ra các dự báo về tình hình dịch bệnh trong tương lai. Các mô hình sử dụng dữ liệu đã thu thập được về dịch bệnh cùng một vài giả thiết ban đầu, kết hợp với các yếu tố toán học nhằm đưa ra các thông số về dịch bệnh trong tương lai, đồng thời xác định mức độ hiệu quả của các biện pháp phòng chống dịch bệnh.

Một trong những mô hình dự báo dịch bệnh cơ bản nhất là mô hình SIR cùng những

mở rộng của nó. Mô hình này được W.O.Kermack và A.G.McKendrick công bố lần đầu tiên vào năm 1927 trong bài báo *A contribution to the mathematical theory of epidemics*. Mô hình này dựa trên giả thiết dân số trong khoảng thời gian diễn ra dịch bệnh là không đổi, và được chia thành 3 nhóm: Có nguy cơ mắc bệnh (Susceptible); Bị nhiễm bệnh (Infectious); Đã hồi phục (Recovered). Mô hình SIR cũng giả định rằng một người ở nhóm đã phục hồi sẽ có miễn dịch với dịch bệnh và không thể bị tái lây nhiễm. Chi tiết về mô hình này sẽ được giới thiệu kĩ càng hơn ở phần sau của báo cáo.

1.3 Hệ động lực và Phương trình vi phân thường

Trong toán học, một Hệ động lực (Dynamical system) là một hệ các phương trình mô tả sự phụ thuộc của một điểm trong không gian theo thời gian. Tại một thời điểm nhất định, một hệ động lực sẽ tồn tại một trạng thái được mô tả bởi một tập các số thực (một vector) có thể được biểu diễn bằng một điểm trong không gian k -chiều, với k là số lượng phần tử trong tập đó. Hệ động lực được ứng dụng trong rất nhiều lĩnh vực như toán học, hoá học, vật lí, sinh học,...

Một trong những loại phương trình thường được sử dụng để miêu tả một Hệ động lực là phương trình vi phân thường (Ordinary Differential Equation - ODE) - phương trình biểu diễn mối quan hệ giữa một hàm một biến chưa biết với đạo hàm của nó. Trong rất nhiều lĩnh vực, chuyển động của một hệ được mô hình hóa bởi các phương trình vi phân thường. Một trong những mô hình đơn giản nhất là định luật II Newton trong cơ học cổ điển

$$m \frac{d^2 x(t)}{dt^2} = F(x(t))$$

Việc giải phương trình vi phân thường có ý nghĩa quan trọng trong việc xác định tính chất của các hệ động lực. Dù vậy, nhiều phương trình vi phân không tồn tại nghiệm chính xác, hoặc việc giải nghiệm chính xác phức tạp và tốn nhiều thời gian. Vì vậy, đã có nhiều phương pháp được sử dụng để xấp xỉ nghiệm của các phương trình vi phân thường. Trong đó, phương pháp Euler là một phương pháp đơn giản và thường xuyên được sử dụng. Chi tiết về phương pháp này sẽ được trình bày ở phần sau của báo cáo.

1.4 Phương pháp suy luận Bayesian

Phương pháp suy luận Bayesian dựa trên định lý Bayesian về xác suất có điều kiện

$$P(H_0|E) = \frac{P(E|H_0)P(H_0)}{P(E)}$$

Trong đó

- H_0 là giả thuyết cần xác minh.
- E là tập các dữ liệu quan sát đã cho trước.
- $P(H_0|E)$ là xác suất hậu nghiệm của H_0 nếu biết E .
- $P(E|H_0)$ là hàm khả dĩ (likelihood), biểu diễn mức độ thích hợp của giả thuyết H_0 đối với tập dữ liệu cho trước E .
- $P(H_0)$ là xác suất tiên nghiệm của H_0 .
- $P(E)$ là xác suất biên của E .

Trong báo cáo này sử dụng phương pháp suy luận Bayesian, kết hợp với việc chọn mẫu dựa trên thuật toán Random Walk Metropolis-Hastings.

2 Các bài toán

2.1 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.

Mô hình SIR là một trong những mô hình ngăn (compartmental model) đơn giản nhất để mô hình hoá dịch bệnh. Mô hình này chia tổng dân số trong một khu vực thành ba phần riêng biệt:

- **S**: Tổng số người có khả năng nhiễm bệnh (Susceptible). Khi một người có nguy cơ nhiễm bệnh tiếp xúc với một người đang nhiễm bệnh, người đó sẽ có khả năng bị lây nhiễm.
- **I**: Tổng số người bị nhiễm bệnh (Infectious). Những người này đã được xác nhận nhiễm bệnh và có khả năng lây nhiễm cho những người có khả năng nhiễm bệnh (S).
- **R**: Tổng số người đã hồi phục (Recovered). Đây là nhóm những người đã bị nhiễm bệnh và hồi phục sau khi được chăm sóc y tế. Mô hình SIR giả định rằng những người đã hồi phục sẽ có miễn dịch với bệnh và không thể tái lây nhiễm cho người khác hoặc bị tái nhiễm.

Ba thành phần kể trên (S, I, R) thể hiện số người ở mỗi trạng thái trong một thời điểm nhất định. Để thể hiện số người có khả năng nhiễm bệnh, số người nhiễm bệnh, và số người đã hồi phục ở một thời điểm t nhất định (kể cả khi đã giả định dân số trong toàn bộ khoảng thời gian được xét là không đổi), ta định nghĩa ba hàm số đối với thời gian t : $S(t)$, $I(t)$, $R(t)$. Với một dịch bệnh lây lan trong một khu vực dân số nhất định, ba hàm số trên sẽ được sử dụng để dự đoán tình hình lây lan của dịch bệnh trong tương lai.

Trong mô hình SIR, mỗi người sẽ có thể trải qua ba trạng thái: Có khả năng nhiễm bệnh (**S**), nhiễm bệnh (**I**), hồi phục (**R**) như hình vẽ bên dưới



Hình 1: Mô hình SIR

2.1.1 Các tỉ lệ chuyển đổi

Giữa hai thành phần S và I, tỉ lệ chuyển đổi theo mô hình SIR là $\frac{d(S/N)}{dt} = -\beta \frac{SI}{N^2}$, trong đó N là tổng dân số (cố định theo giả định của mô hình), β là tỉ lệ tiếp xúc của mỗi người trong nhóm S với người trong nhóm I, và $\frac{SI}{N^2}$ là tỉ lệ người trong nhóm S bị lây nhiễm sau khi tiếp xúc với người trong nhóm I. Vì N cố định, phương trình trên có thể được rút gọn về dạng $\frac{dS}{dt} = -\beta \frac{SI}{N}$.

Giữa hai thành phần I và R, tỉ lệ chuyển đổi được xác định bằng số bệnh nhân ở nhóm I đã hồi phục γI , với γ là tỉ lệ hồi phục của các bệnh nhân nhiễm bệnh. Nếu một bệnh nhân hồi phục sau khoảng thời gian D , thì $\gamma = \frac{1}{D}$.

2.1.2 Mô hình SIR rời rạc

Với h là khoảng thời gian giữa hai lần quan sát các giá trị S, I, R; xét $t = t_0, t_1, \dots, t_n$, với $t_k = t_{k-1} + h$ cùng các giá trị ban đầu tại thời điểm t_0 là $S(0)$, $I(0)$, $R(0)$, ta có mô hình SIR

rời rạc như sau, với $k = 1, \dots, n$

$$S(t_k) = S(t_{k-1}) + (-\beta \frac{S_{k-1}I_{k-1}}{N})h \quad (1)$$

$$I(t_k) = I(t_{k-1}) + (\beta \frac{S_{k-1}I_{k-1}}{N} - \gamma I_{k-1})h \quad (2)$$

$$R(t_k) = R(t_{k-1}) + \gamma I(t_{k-1})h \quad (3)$$

2.1.3 Mô hình SIR liên tục

Tốc độ lây lan của dịch bệnh trong thực tế thường lớn hơn rất nhiều lần so với tỉ lệ sinh và tử vong trong dân số. Vì vậy, mô hình SIR giả định rằng dân số trong khoảng thời gian diễn ra dịch bệnh là không đổi. Mô hình SIR không có các yếu tố nhân khẩu học như trên được biểu diễn bằng một hệ động lực với ba phương trình vi phân thường sau:

$$\frac{dS}{dt} = -\beta \frac{SI}{N} \quad (4)$$

$$\frac{dI}{dt} = \beta \frac{SI}{N} - \gamma I \quad (5)$$

$$\frac{dR}{dt} = \gamma I \quad (6)$$

Với $S(t)$, $I(t)$, $R(t)$ là các hàm số đã được đề cập ở trên. N là tổng dân số và không đổi theo thời gian: $N = S(t) + I(t) + R(t)$.

2.1.4 Hệ số lây nhiễm cơ bản R_0

Trong dịch tễ học, hệ số lây nhiễm cơ bản (kí hiệu R_0) là số trường hợp lây nhiễm mới được tạo ra bởi một trường hợp lây nhiễm ban đầu. Nói một cách đơn giản, $R_0 = 2$ có nghĩa rằng một người đang nhiễm bệnh sẽ có khả năng lây nhiễm cho trung bình 2 người khác.

Một trong những ứng dụng quan trọng của việc ước tính hệ số R_0 là xác định xem một bệnh truyền nhiễm có khả năng lây lan trong cộng đồng và bùng phát thành đại dịch hay không. Trong các mô hình lây nhiễm thường được sử dụng, khi $R_0 \geq 1$, sự lây nhiễm sẽ có thể bắt đầu lây lan nhanh trong cộng đồng. Và nếu $R_0 \leq 1$, sự bùng phát dịch bệnh sẽ khó xảy ra.

Trong mô hình SIR, ta có thời gian tiếp xúc trung bình giữa những người có khả năng nhiễm bệnh (nhóm S) và những người đã nhiễm bệnh (nhóm I) là $T_c = \beta^{-1}$, và thời gian hồi phục trung bình của một bệnh nhân đã nhiễm bệnh là $T_r = \gamma^{-1}$. Từ đó, ta có thể suy ra rằng tỉ lệ lây lan của một người đã nhiễm bệnh cho những người chưa nhiễm bệnh là $\frac{T_c}{T_r}$, hay nói cách khác:

$$R_0 = \frac{\beta}{\gamma}$$

Tầm quan trọng của hệ số lây nhiễm cơ bản R_0 có thể được giải thích bằng mô hình SIR liên tục được giới thiệu ở phần trên. Ta nhắc lại hai phương trình vi phân thường được sử dụng để mô tả tốc độ thay đổi của hai giá trị S và I trong mô hình SIR liên tục

$$\frac{dS}{dt} = -\beta \frac{SI}{N} \quad (7)$$

$$\frac{dI}{dt} = (\beta \frac{S}{N} - \gamma)I \quad (8)$$

Để ý rằng mô hình SIR chỉ có ý nghĩa khi các giá trị S, I, R là không âm. Vì vậy, một khi giá trị $S(t)$ và $I(t)$ có giá trị bằng 0, ta coi mô hình SIR từ thời điểm t trở đi là không có ý nghĩa. Từ hai phương trình vi phân trên, ta thấy $S' < 0$ với mọi t và $I' > 0$ khi và chỉ khi $\frac{S}{N} > \frac{\gamma}{\beta}$. Từ đó, trong giai đoạn dịch bệnh xuất hiện trong cộng đồng, số người bị nhiễm bệnh I sẽ tăng khi $\frac{S}{N} > \frac{\gamma}{\beta}$. Nhưng vì $S' < 0$ với mọi t đồng nghĩa với việc hàm S giảm nghiêm ngặt, đến một thời điểm t_k nào đó, giá trị $\frac{S}{N}$ sẽ bé hơn $\frac{\gamma}{\beta}$, dẫn đến giá trị I sẽ giảm. Điều này cũng giải thích cho việc trong mô hình SIR, giá trị I ban đầu sẽ tăng, sau đó đạt mức đạt cực đại khi $\frac{S}{N} = \frac{\gamma}{\beta}$, và cuối cùng sẽ giảm về 0.

Vì $R_0 = \frac{\beta}{\gamma}$, ta có thể viết lại các công thức đã nêu ở trên theo R_0 . Nếu $\frac{SR_0}{N} > 1$, dịch bệnh có nguy cơ lây lan rộng và bùng phát trong cộng đồng. Ngược lại, nếu $\frac{SR_0}{N} < 1$, dịch bệnh sẽ dần được dập tắt.

Giá trị R_0 có thể được ước lượng khi có đầy đủ dữ liệu về một dịch bệnh trong quá khứ. Từ hai phương trình (7) và (8), ta có

$$\frac{dI}{dS} = \frac{(\beta S - \gamma)I}{-\beta SI} = -1 + \frac{\gamma}{\beta S}$$

Giả sử I là hàm phụ thuộc vào biến S , giải phương trình vi phân trên ta được

$$I = -S + \frac{\gamma}{\beta} \ln S + c$$

Với c là hằng số. Một cách khác để biểu diễn mối quan hệ giữa S và I trong mô hình SIR là định nghĩa hàm hai biến $V(S, I)$

$$V(S, I) = S + I - \frac{\gamma}{\beta} \ln S$$

Đồ thị biểu diễn mối quan hệ giữa S và I sẽ là đồ thị của phương trình $V(S, I) = c$, với c là hằng số được xác định từ các giá trị đầu $S(0)$ và $I(0)$. Giả sử $S(0) \approx N$, với N là tổng dân số, $I(0) \approx 0$ và $R_0 = \frac{\beta}{\gamma}$, đồng thời định nghĩa $S_\infty = \lim_{t \rightarrow \infty} S(t)$. Với việc $\lim_{t \rightarrow \infty} I(t) = 0$, ta có mối quan hệ $V(S_0, I_0) = V(S_\infty, 0)$, hay nói cách khác

$$N - \frac{\gamma}{\beta} \ln S_0 = S_\infty - \frac{\gamma}{\beta} \ln S_\infty$$

Bằng biến đổi đại số, ta có

$$R_0 = \frac{(\ln S_0 - \ln S_\infty)}{N - S_\infty}$$

2.1.5 Mô hình SIS

Với một vài bệnh truyền nhiễm như cảm lạnh hoặc cúm, một người sau khi mắc bệnh sẽ không tạo miễn dịch đặc hiệu đối với bệnh đó. Nói cách khác, một người sau khi mắc bệnh và được chữa khỏi hoàn toàn có thể bị tái nhiễm nhiều lần. Mô hình SIS được sử dụng để mô tả những loại dịch bệnh đó.

Trong mô hình SIS, dân số được chia làm hai loại: Có nguy cơ nhiễm bệnh (**S**) và nhiễm bệnh (**I**). Một người ở nhóm **S** sau khi nhiễm bệnh sẽ được chuyển sang nhóm **I**. Ngược lại, một người

ở nhóm I sau khi được chữa khỏi sẽ được chuyển sang nhóm S. Hệ động lực biểu diễn mô hình SIS bao gồm hai phương trình vi phân như sau

$$\frac{dS}{dt} = -\beta \frac{SI}{N} + \gamma I \quad (9)$$

$$\frac{dI}{dt} = \beta \frac{SI}{N} - \gamma I \quad (10)$$

2.1.6 Mô hình SIRD

Trong mô hình SIR, một người được coi là ở nhóm R khi: (1) được chữa khỏi bệnh và có miễn dịch đặc hiệu; và (2) tử vong. Mô hình SIRD là một mở rộng của mô hình SIR, với ba nhóm S, I, R và các hệ số β và γ như mô hình SIR, ngoài ra có thêm nhóm D dùng để phân loại những người mắc bệnh đã tử vong và hệ số tử vong μ .

Hệ động lực biểu diễn mô hình SIRD bao gồm bốn phương trình vi phân như sau

$$\frac{dS}{dt} = -\beta \frac{SI}{N} \quad (11)$$

$$\frac{dI}{dt} = \beta \frac{SI}{N} - \gamma I - \mu I \quad (12)$$

$$\frac{dR}{dt} = \gamma I \quad (13)$$

$$\frac{dD}{dt} = \mu I \quad (14)$$

2.2 Bài toán 2

Viết chương trình sử dụng thuật toán Euler tìm nghiệm của hệ SIR hoặc mở rộng của nó với tham số đầu vào gồm biến thời gian t , các hệ số tiếp xúc β , hệ số phục hồi γ và điều kiện đầu là số ca mắc bệnh $I(t_0)$ và số ca phục hồi $R(t_0)$ của mô hình tính tại thời điểm đầu tiên ghi nhận được các ca nhiễm bệnh. Giá trị trả về là mảng chứa số người nhiễm bệnh $I(t)$ và số người đã hồi phục $R(t)$ tính tại thời điểm $t \geq t_0$. Cho một số ví dụ về điều kiện đầu và các hệ số trong mô hình và dùng chương trình đã viết để tìm nghiệm xấp xỉ. Biểu diễn nghiệm xấp xỉ bằng cách vẽ đồ thị. Trường hợp là hệ SIR mở rộng như SIRD thì cần trả về $I(t)$, $R(t)$ và $D(t)$ là số ca tử vong tại thời điểm $t \geq t_0$. Trình bày chi tiết kết quả trong báo cáo.

Với β và γ không đổi, nghiệm chính xác của mô hình SIR liên tục đã đề cập ở phần trên là:

$$S(t) = S_0(1+k)^{\frac{\beta}{\beta-\gamma}}(1+ke^{(\beta-\gamma)(t-t_0)})^{-\frac{\beta}{\beta-\gamma}} \quad (15)$$

$$I(t) = I_0(1+k)^{\frac{\beta}{\beta-\gamma}}(1+ke^{(\beta-\gamma)(t-t_0)})^{-\frac{\beta}{\beta-\gamma}}e^{(\beta-\gamma)(t-t_0)} \quad (16)$$

$$R(t) = N - S(t) - I(t) \quad (17)$$

Trong đó $k = \frac{I_0}{S_0}$.

Việc giải nghiệm chính xác của mô hình SIR thường khá phức tạp. Vì vậy, một số thuật toán tìm nghiệm xấp xỉ của hệ phương trình vi phân bậc nhất được sử dụng để xấp xỉ nghiệm của mô hình SIR, trong đó thuật toán đơn giản và phổ biến nhất là thuật toán Euler.

2.2.1 Thuật toán Euler

Giả sử ta có phương trình vi phân bậc nhất

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

Ý tưởng của thuật toán Euler là xấp xỉ các điểm trên đồ thị hàm số y bằng các điểm A_1, A_2, \dots, A_n (hay nói cách khác, là xấp xỉ nghiệm của phương trình vi phân trên bằng dãy số y_1, y_2, \dots, y_n).

Với y_0 cho trước, ta có thể tính đạo hàm của hàm y tại điểm y_0 và dựng được đường tiếp tuyến tại điểm A_0 ứng với y_0 . Sau bước nhảy Δ_t , điểm A_1 sẽ được xác định trên đường tiếp tuyến của A_0 . Với Δ_t đủ nhỏ, dy sẽ xấp xỉ Δ_y . Sau khi lặp lại quy trình trên, ta sẽ có tập hợp các điểm A_1, A_2, \dots, A_n ứng với y_1, y_2, \dots, y_n .

Một cách tổng quát, thuật toán Euler được mô tả như sau:

Cho phương trình vi phân bậc nhất $y' = f(t, y(t))$ với $y(t_0) = y_0$ cho trước. Chọn giá trị bước nhảy h đủ nhỏ và $t_n = t_0 + nh$. Khi đó, mỗi bước nhảy của phương pháp Euler từ t_n đến $t_{n+1} = t_n + h$ được xác định bởi công thức

$$y_{n+1} = y_n + hf(t_n, y_n)$$

Giá trị của y_n chính là nghiệm xấp xỉ của phương trình vi phân bậc nhất tại t_n : $y_n \approx y(t_n)$. Ứng dụng phương pháp Euler vào mô hình SIR rời rạc đã được trình bày ở trên, ta có đoạn chương trình được viết bằng ngôn ngữ R để tìm nghiệm xấp xỉ của hệ SIR.

```
# Import necessary libraries
library(ggplot2)
library(tidyr)

# Initialize variables from user input
s0 <- readline("Initial value of S: ")
i0 <- readline("Initial value of I: ")
r0 <- readline("Initial value of R: ")
beta <- readline("Possion rate (beta): ")
gamma <- readline("Recovery rate (gamma): ")
t <- readline("Period of time: ")

s0 <- as.numeric(s0)
i0 <- as.numeric(i0)
r0 <- as.numeric(r0)
beta <- as.numeric(beta)
gamma <- as.numeric(gamma)
t <- as.numeric(t)

#Predict SIR value based on Euler's algorithm
euler <- function(s0, i0, r0, beta, gamma, t) {
  s = i = r = numeric(t + 1)
  N = s0 + i0 + r0
  s[1] = s0; i[1] = i0; r[1] = r0
  j = 1
  while (j <= t) {
    s[j+1] = s[j] + (-beta*i[j]*s[j])
    i[j+1] = i[j] + (beta*i[j]*s[j]) - gamma*i[j]
    r[j+1] = r[j] + (gamma*i[j])
    j = j + 1
  }
  x = 0:t
  return (data.frame(Time = x, Susceptible = s, Infectious = i, Recovered = r))
}

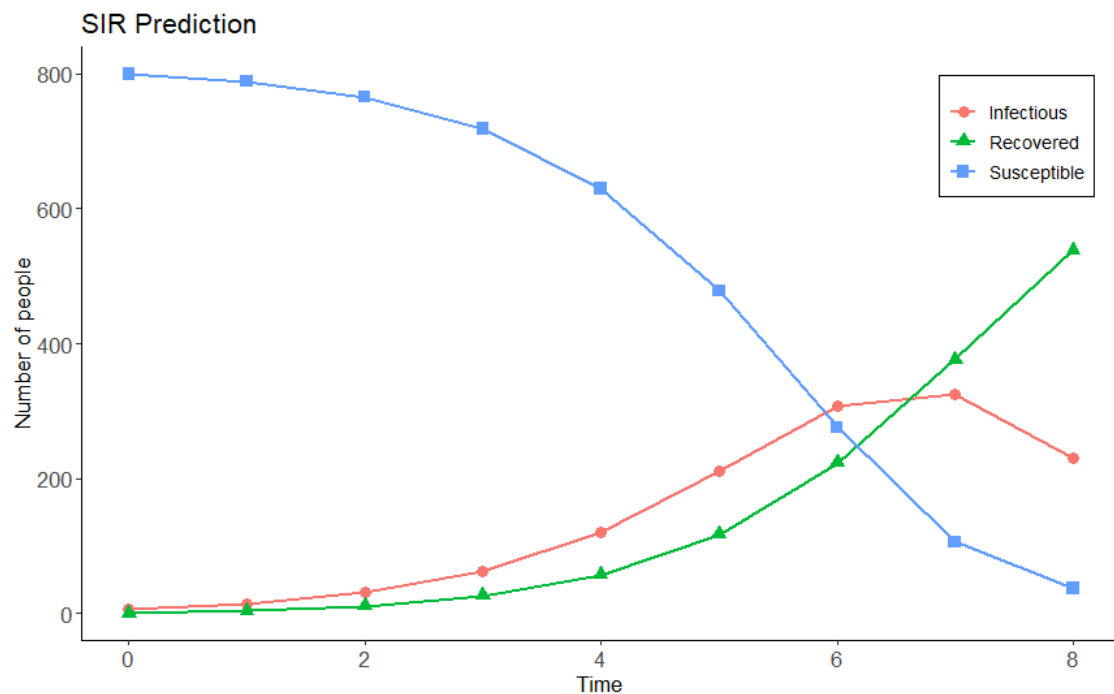
#Get the result and display it
res <- euler(s0, i0, r0, beta, gamma, t)
View(res)

#Create a line plot to visualize results
res %>% gather(key, value, Susceptible, Infectious, Recovered) %>% ggplot(aes(x =
  Time, y = value, colour = key, shape = key)) +
  geom_line(size = 1) +
  geom_point(size = 3) +
  labs(title = "SIR Prediction", x = "Time", y = "Number of people") +
  theme_classic() +
  theme(plot.title = element_text(size = 16), text = element_text(size = 13), axis
    .text = element_text(size = 13)) +
  theme(legend.position = c(0.9, 0.85), legend.title = element_blank(), legend.
    background = element_blank(), legend.box.background = element_rect(color = "
    black"))
```

Ví dụ 1: Sử dụng lại ví dụ trong đề bài, với các tham số đầu vào $S(t_0) = 800, I(t_0) = 7, R(t_0) = 0, \beta = 0.002, \gamma = 0.5$, ta có bảng dữ liệu thể hiện nghiệm của mô hình SIR trong 8 tuần, cùng đồ thị biểu diễn bảng dữ liệu đó.

	Time	Susceptible	Infectious	Recovered
1	0	800.00000	7.00000	0.00000
2	1	788.80000	14.70000	3.50000
3	2	765.60928	30.54072	10.85000
4	3	718.84476	62.03488	26.12036
5	4	629.65787	120.20433	57.13780
6	5	478.28266	211.47737	117.23996
7	6	275.99074	308.03061	222.97865
8	7	105.96355	324.04250	376.99396
9	8	37.29016	230.69463	539.01520

Hình 2: Nghiệm của mô hình SIR trong ví dụ 1



Hình 3: Đồ thị nghiệm của mô hình SIR trong ví dụ 1

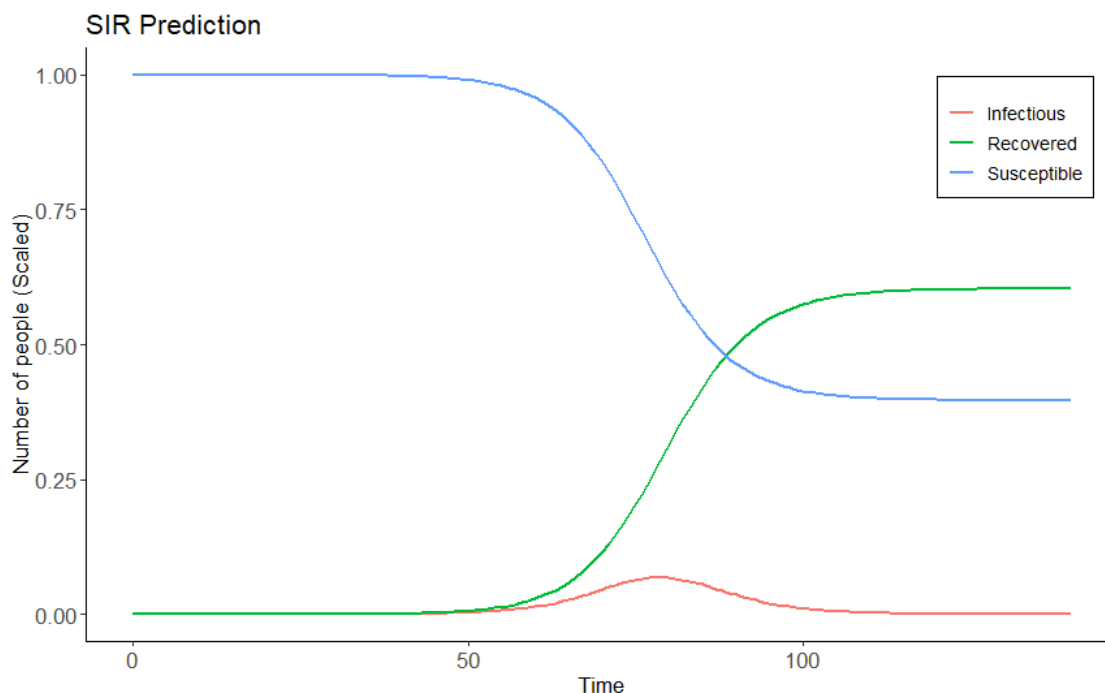
Ví dụ 2: Dịch cúm Hong Kong bùng phát tại thành phố New York, Hoa Kỳ vào cuối những năm 60 của thế kỷ trước. Ở thời điểm đó, rất ít người dân New York miễn dịch với bệnh cúm này, vì vậy hầu như tất cả người dân đều ở trong trạng thái có thể bị nhiễm bệnh. Giả sử ở thời điểm bắt đầu ghi nhận dịch có 7,900,000 người có nguy cơ nhiễm bệnh, đã 10 ca nhiễm bệnh được xác nhận, và chưa có người nào tử vong vì dịch. Từ đó, ta có các điều kiện đầu như sau:

- $S_0 = 7,900,000$
- $I_0 = 10$
- $R_0 = 0$

Để thuận tiện cho việc tính toán, ta sẽ chuyển các điều kiện trên theo tỉ lệ $s_0 = 1$:

- $s_0 = 1$
- $i_0 = 1.27 * 10^{-6}$
- $r_0 = 0$

Giả sử thời gian hồi phục sau khi mắc bệnh là 3 ngày, khi đó $\gamma = \frac{1}{3}$. Đồng thời, giả sử một người trong nhóm nhiễm bệnh (I) cứ hai ngày lại tiếp xúc với một người ở nhóm có khả năng nhiễm bệnh (S), khi đó $\beta = \frac{1}{2}$. Đồ thị bên dưới thể hiện tình hình dịch bệnh trong vòng 140 ngày kể từ khi ghi nhận dịch bệnh.



Hình 4: Đồ thị nghiệm của mô hình SIR trong ví dụ 2

2.3 Bài toán 3

Viết chương trình theo ngôn ngữ tự chọn để lấy mẫu sử dụng thuật toán Metropolis–Hastings với tham số đầu vào là phân bố xác suất tiên nghiệm $\pi(\beta, \gamma)$ cho trước. Giá trị trả về là một mẫu gồm các cặp β và γ có phân bố xác suất $\pi(\beta, \gamma)$. Vẽ biểu đồ thể hiện quá trình chọn mẫu. Trình bày chi tiết kết quả trong báo cáo.

2.3.1 Thuật toán Metropolis-Hastings

Gọi p là phân bố cần lấy mẫu với hàm mật độ xác suất $p(x)$. Ý tưởng cơ bản của thuật toán Metropolis-Hastings là ở mỗi bước, thuật toán sẽ đưa ra một mẫu x' dự định, sau đó cân nhắc xem có nên chuyển từ trạng thái hiện tại x sang trạng thái dự định x' hay không, bằng cách tính toán xác suất $q(x'|x)$, với q là phân bố xác suất dự định.

Phân bố dự định thường được sử dụng là phân bố chuẩn nhiều chiều $q(x'|x) \sim \mathcal{N}(x, \Sigma)$, tức là phân bố chuẩn với giá trị trung bình là trạng thái hiện tại x và ma trận hiệp phương sai $\Sigma = \sigma^2 I_n$. Lý do chọn phân bố chuẩn nhiều chiều làm phân bố dự định vì nó là phân bố có tính đối xứng tại giá trị trung bình x , qua đó rút gọn việc tính toán xác suất chấp nhận ở bước tiếp theo.

Nếu dự định x' được chấp nhận, ta sẽ thêm x' vào tập mẫu và chuyển trạng thái hiện tại từ x sang x' , ngược lại ta sẽ giữ nguyên trạng thái hiện tại x và thêm nó vào tập mẫu. Việc chấp nhận hay bác bỏ trạng thái dự định x' được xác định dựa trên xác suất được tính bởi công thức sau:

$$r = \min\left\{1, \frac{p(x')}{p(x)}\right\}$$

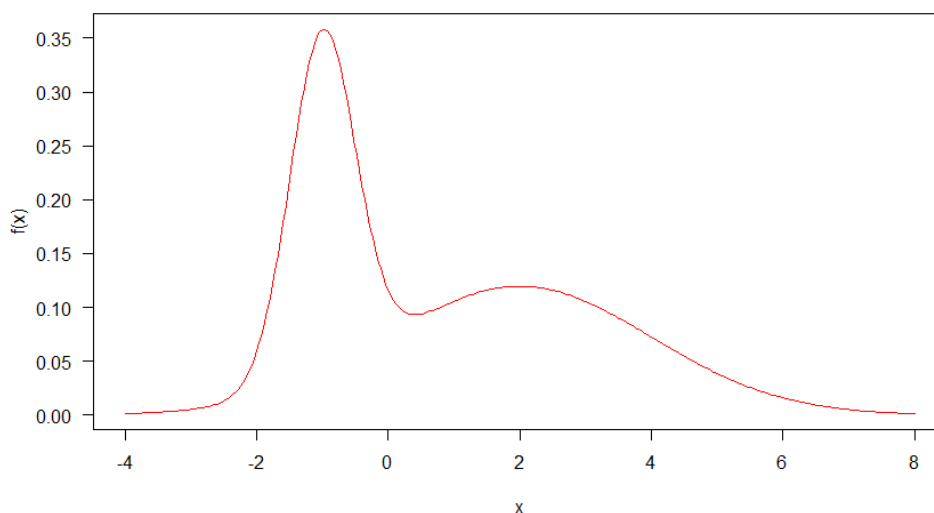
Nếu xác suất chấp nhận $r \leq u$, với $u \sim U(0, 1)$, ta chấp nhận trạng thái dự định x' , ngược lại ta bác bỏ trạng thái này.

```
#Import necessary libraries
library(MASS)

5 #Generate samples based on MCMC algorithm

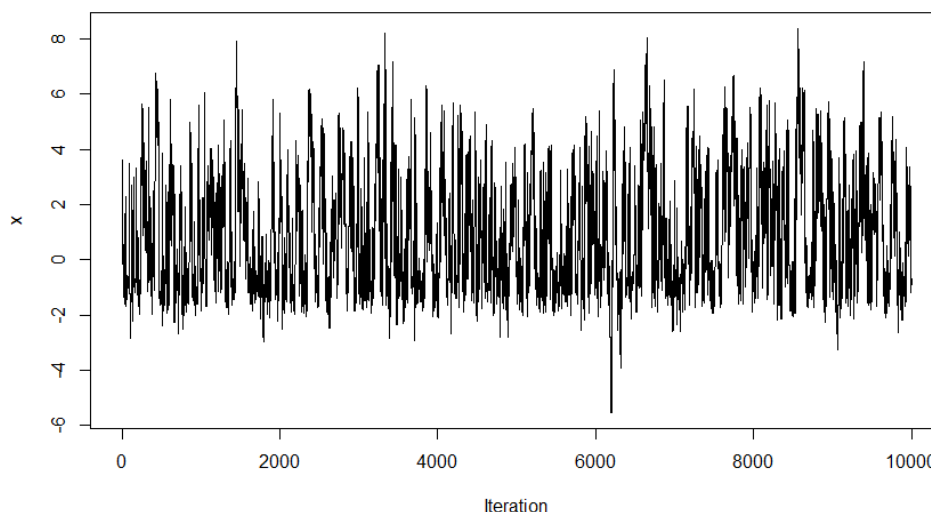
metro <- function(target, N, x, burnin = 0, sd) {
  samples <- x
  for (i in 2:(burnin + N)) {
    proposal <- mvrnorm(n = 1, x, sd)
    10 if (runif(1) < min(1, target(proposal)/target(x))) {
      x <- proposal
    }
    samples <- rbind(samples, x)
  }
  15 samples[(burnin + 1):(N + burnin),]
}
```

Ví dụ 1: Phân bố xác suất 1 biến. Giả sử ta cần lấy một mẫu có phân bố xác suất $p_1(x) \sim 0.4\mathcal{N}(-1, 0.5) + 0.6\mathcal{N}(2, 2)$. Hàm mật độ xác suất của phân phối p_2 có đồ thị như bên dưới



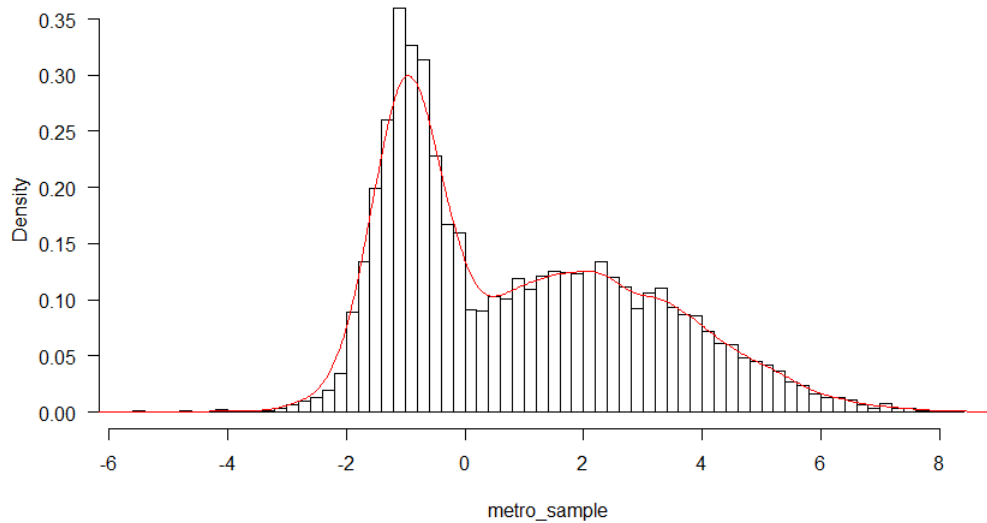
Hình 5: Mật độ xác suất của phân phối p_1

Chạy chương trình trên với giá trị đầu $x_0 = 0$, phương sai của phân phối dự định $\Sigma = (1)$ để tạo mẫu gồm 10,000 giá trị, trong đó loại bỏ (burn-in) 1,000 giá trị đầu tiên, ta có biểu đồ thể hiện quá trình chọn mẫu như sau.



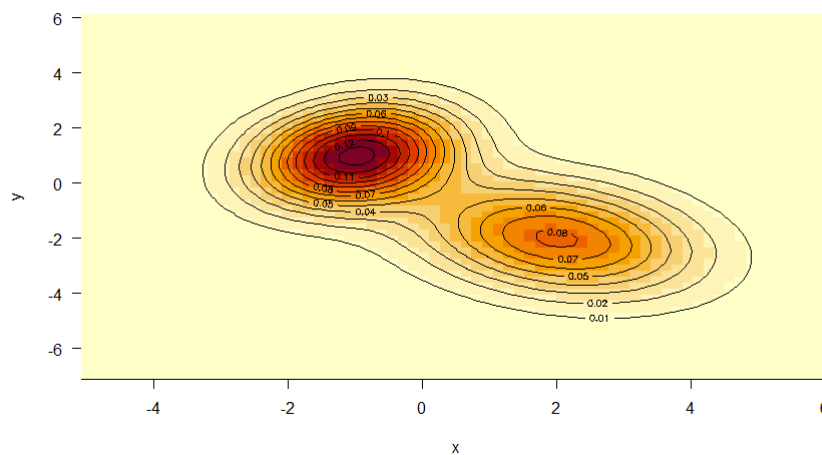
Hình 6: Quá trình chọn mẫu p_1

Đồ thị Histogram và mật độ xác suất của mẫu vừa tạo



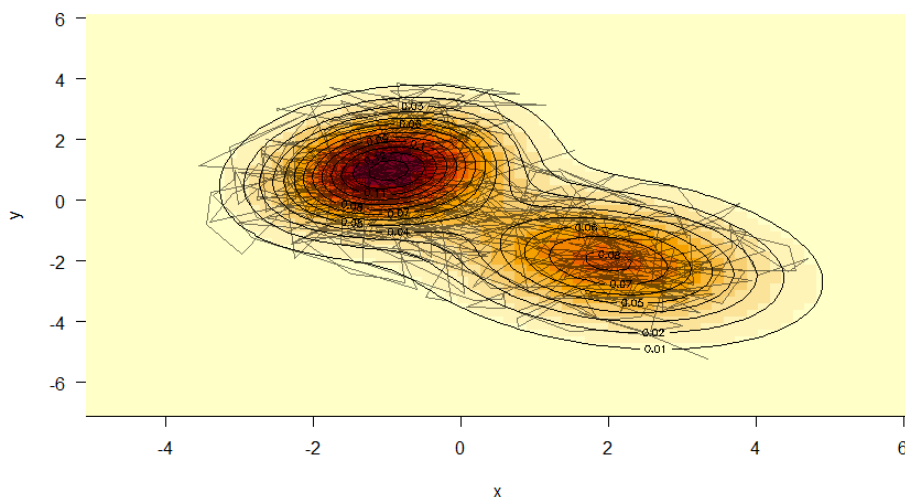
Hình 7: Histogram và mật độ xác suất của mẫu

Ví dụ 2: Phân bố xác suất 2 biến. Ta sử dụng một phân phối xác suất p_2 là tổng của hai phân phối chuẩn 2 chiều $\mathcal{N}_1(\mu_1, \Sigma_1)$ và $\mathcal{N}_2(\mu_2, \Sigma_2)$, trong đó $\mu_1 = \begin{pmatrix} 1 & 1 \end{pmatrix}$, $\Sigma_1 = \begin{pmatrix} 1 & 0.25 \\ 0.25 & 1.0 \end{pmatrix}$, $\mu_2 = \begin{pmatrix} 2 & -2 \end{pmatrix}$, $\Sigma_2 = \begin{pmatrix} 2 & -0.5 \\ -0.5 & 2 \end{pmatrix}$. Hàm mật độ xác suất của phân phối p_2 có đồ thị như bên dưới



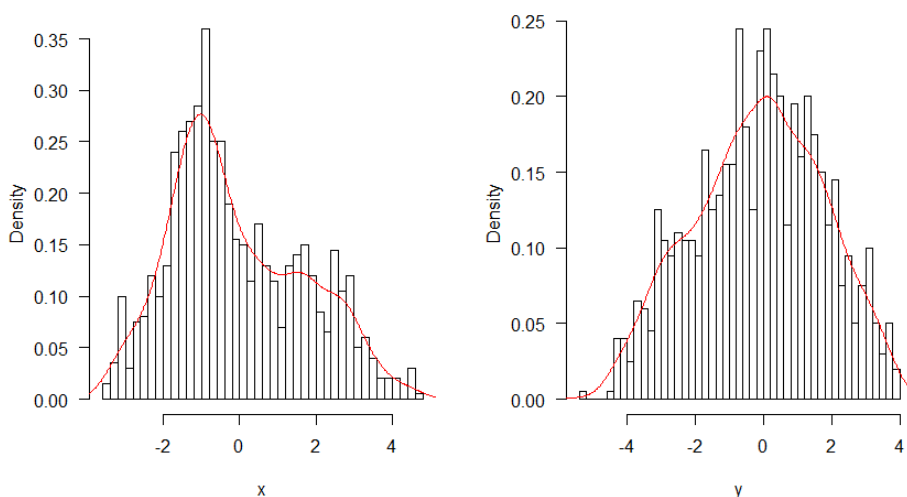
Hình 8: Mật độ xác suất của phân phối p_2

Giả sử rằng ta không biết trước rằng phân bố xác suất cần lấy mẫu là tổng của hai phân bố chuẩn nhiều chiều (rõ ràng là trong thực tế chúng ta sẽ không biết điều này). Chạy chương trình trên với giá trị đầu $x_0 = (-4 \ -4)$, $\Sigma = I_2$ để tạo mẫu gồm 1,000 giá trị, sau khi lược bỏ (burn-in) 100 giá trị đầu tiên, ta có biểu đồ thể hiện quá trình chọn mẫu như sau:



Hình 9: Quá trình chọn mẫu p_2

Đồ thị Histogram và mật độ xác suất của các mẫu x và y vừa tạo



Hình 10: Histogram và mật độ xác suất của mẫu

2.4 Bài toán 4

Mỗi nhóm hãy tự chọn lấy một khu vực gồm một số quốc gia, dùng các chương trình ở Bài tập 3 để ước lượng giá trị trung bình hệ số R_0 ở khu vực này. Phân tích rõ chính sách hạn chế đi lại và cách ly đã ảnh hưởng đến hệ số R_0 ở khu vực này như thế nào. Nêu rõ dẫn chứng. Trình bày chi tiết kết quả trong báo cáo.

Việc ước lượng hệ số lây nhiễm cơ bản R_0 có ý nghĩa quan trọng trong việc theo dõi tình hình lây lan của dịch bệnh ở thời điểm hiện tại và trong tương lai; cũng như đánh giá mức độ ảnh hưởng của các biện pháp phòng chống dịch bệnh đã được triển khai. Ở phần dưới, nhóm sử dụng dữ liệu thời gian về tình hình dịch COVID-19 ở hai quốc gia nổi bật là Việt Nam và Hoa Kỳ để phân tích chiến lược tiếp cận và phòng chống dịch ở nước ta.

Trong đoạn chương trình bên dưới, nhóm sử dụng dữ liệu thời gian về dịch COVID-19 được cung cấp từ thư viện Tidycovid19 của tác giả Joachim Gassen, đồng thời sử dụng thư viện R_0 của hai tác giả Pierre-Yves Boelle và Thomas Obadia để đưa ra ước lượng về hệ số R_0 theo thời gian.

```
#Import necessary libraries

library(devtools)
#remotes::install_github("joachim-gassen/tidycovid19")

5 library(tidycovid19)
  library(dplyr)
  library(tidyr)
  library(ggplot2)
10 library(R0)
   library(MASS)

#Get the merged data from tidycovid19 library.
#In case of other countries, replace the 'iso3c' value
15 #The latest date of received data is 20-07-2020

df <- download_merged_data(cached = TRUE, silent = TRUE)
vietnamdata <- df %>% filter(iso3c == "VNM", date != "2020-07-21") %>% dplyr::
  select(date, confirmed, recovered)

20 #The 'id' column contains the number of infected people everyday, which equals the
   number of confirmed case minus recovered ones.

vietnamdata <- vietnamdata %>% mutate(Id = confirmed - recovered)

#Normalize the data
25 for (i in 1:(nrow(vietnamdata))) {
  if (is.na(vietnamdata$Id[i])) vietnamdata$Id[i] <- 1
  else if (vietnamdata$Id[i] == 0) vietnamdata$Id[i] <- 1
}

30 #Fit the data to the log-normal distribution

param <- as.vector(fitdistr(vietnamdata$Id, "lognormal")$estimate)

35 #Generate the generation time distribution

mGT <- generation.time("lognormal", param)
```

```

40 #Estimate  $R_0(t)$  with time-dependant algorithm

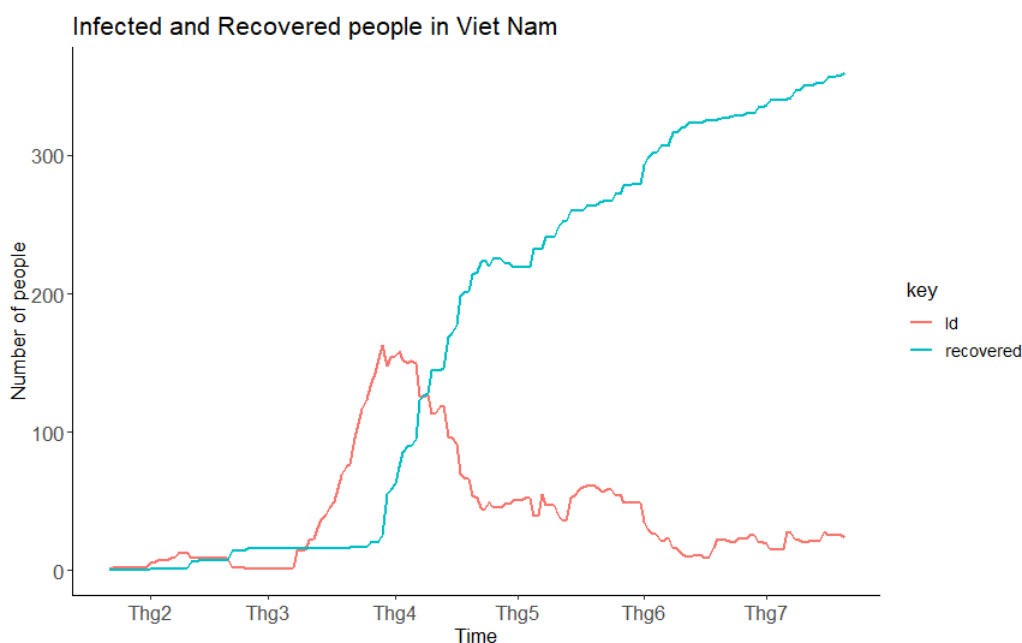
est <- est.R0.TD(vietnamdata$Id, mGT, begin = 0, end = 180, pop.size = 90000000,
  nsim = 500)

#Plot  $R_0(t)$  and  $E(R)$ 
45 plot <- ggplot(data = data.frame(x = vietnamdata$date[1:179], y = est$R[1:179]),
  aes(x = x, y = y)) + geom_line() +
  geom_hline(yintercept = mean(est$R[1:179]), color="blue") +
  labs(title = "R0 prediction in Vietnam", x = "Time", y = "R0") +
  theme_classic() +
  theme(plot.title = element_text(size = 16), text = element_text(size = 13), axis.
    text = element_text(size = 13))

```

2.4.1 Ước lượng hệ số lây nhiễm cơ bản R_0 tại Việt Nam

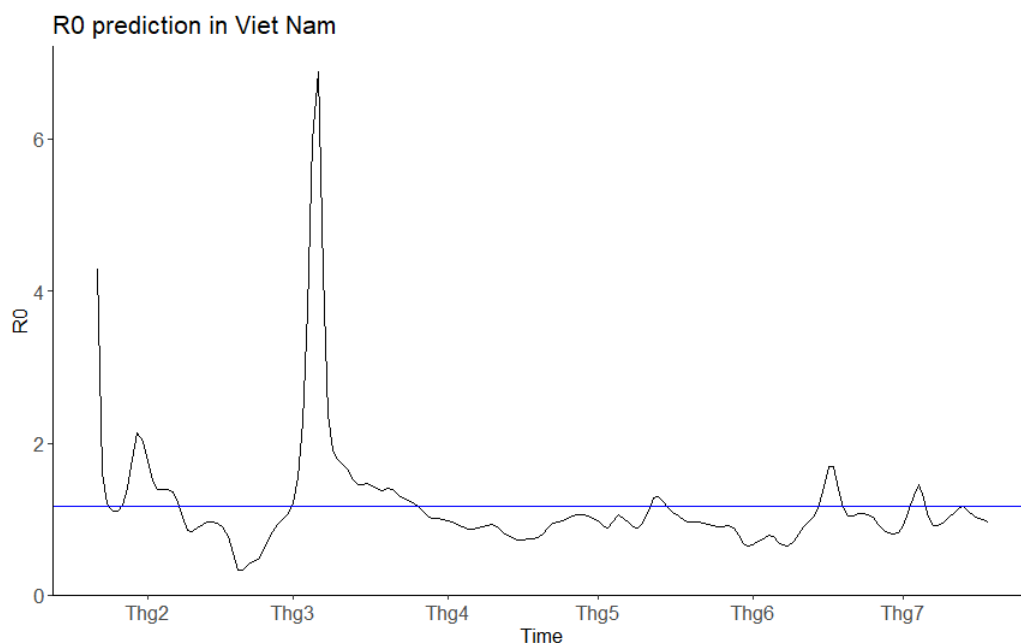
Biểu đồ bên dưới thể hiện số ca nhiễm bệnh và số ca đã hồi phục tại Việt Nam từ khi bùng phát dịch cho đến thời điểm hiện tại (lưu ý rằng số ca nhiễm bệnh là số ca vẫn còn được xác định là dương tính ở thời điểm ghi nhận, không phải là tổng số ca nhiễm bệnh lũy kế từ khi bùng phát dịch).



Hình 11: Tổng số ca nhiễm bệnh và số ca hồi phục tại Việt Nam

Có thể thấy giai đoạn từ đầu tháng 3 đến đầu tháng 4 là giai đoạn có nhiều ca dương tính nhất, vì vậy ta có thể dự đoán rằng hệ số lây nhiễm R_0 trong giai đoạn này sẽ ở mức khá cao.

Chạy chương trình trên để ước lượng giá trị R_0 theo thời gian tại Việt Nam, ta có biểu đồ thể hiện giá trị R_0 như bên dưới.



Hình 12: Ước lượng giá trị R_0 theo thời gian tại Việt Nam

Có thể thấy mô hình đã thể hiện đúng những gì ta đã dự đoán ở trên: khoảng thời gian từ đầu tháng 3 đến đầu tháng 4 ghi nhận hệ số R_0 ở mức cao đột biến. Đây cũng chính là thời điểm diễn ra giai đoạn 2 của dịch COVID-19 tại Việt Nam, khi các ca lây nhiễm cộng đồng liên tục được ghi nhận. Ngày 06/03/2020, Hà Nội ghi nhận ca dương tính SARS-CoV-2 đầu tiên, và cũng là ca bệnh thứ 17 Việt Nam. Liên tục những ngày sau đó, nhiều ổ dịch liên tục xuất hiện tại Hà Nội, Vĩnh Phúc, Bình Thuận, Thành phố Hồ Chí Minh,... Số ca dương tính trong cộng đồng liên tục tăng.

Trước tình hình đó, Chính Phủ đã triển khai nhiều biện pháp chặt chẽ để ngăn chặn đà lây lan của dịch COVID-19, bao gồm việc tạm dừng nhập cảnh đối với tất cả du khách nước ngoài, đồng thời thực hiện khai báo y tế bắt buộc toàn dân. Ngày 31/03/2020, thủ tướng Chính Phủ ban hành Chỉ thị số 16 về việc cách ly toàn xã hội trên phạm vi toàn quốc kể từ 0h ngày 01/04. Tất cả các hoạt động tụ tập đông người ở nơi công cộng đều bị hạn chế ở mức tối đa.

Với những biện pháp trên, Việt Nam đã cơ bản ngăn chặn thành công sự lây lan của COVID-19 trong cộng đồng. Kể từ giữa tháng 4 đến nay, Việt Nam không ghi nhận thêm ca mắc SARS-CoV-2 trong cộng đồng. Tất cả các ca nhiễm SARS-CoV-2 có yếu tố nước ngoài đều đã được kiểm soát và cách ly ngay từ khi nhập cảnh. Hệ số R_0 từ đó đã giảm về mức ổn định ≈ 1 từ đó đến nay. Theo mô hình trên, hệ số R_0 trung bình tại Việt Nam kể từ khi bùng phát dịch đến nay ở mức 1.1713.

3 Kết luận

Đại dịch COVID-19 vẫn đang diễn biến hết sức phức tạp trên thế giới và để lại hậu quả lớn lao cả về tính mạng lẫn kinh tế cho hầu hết các quốc gia. Bằng các biện pháp kịp thời và chặt chẽ, Việt Nam đã phần nào ngăn chặn thành công đại dịch này trên phạm vi quốc gia, đồng thời quay trở lại phát triển kinh tế - xã hội trong tình hình mới.

Thông qua báo cáo này, nhóm đã trình bày tóm lược về mô hình SIR trong dự báo đại dịch, với trường hợp cụ thể là đại dịch COVID-19, đồng thời cũng đã viết chương trình bằng ngôn ngữ R để giải nghiệm của mô hình SIR. Ngoài ra, nhóm cũng đã trình bày về thuật toán Metropolis-Hastings sử dụng Monte Carlo Markov Chain để tạo mẫu cho phân bố xác suất chưa biết, đồng thời sử dụng thuật toán này cùng các thư viện hỗ trợ để ước lượng hệ số lây nhiễm cơ bản R_0 tại Việt Nam kể từ khi đại dịch COVID-19 xuất hiện cho đến nay.

Việc sử dụng các mô hình toán học để giả lập và dự báo dịch bệnh có ý nghĩa quan trọng trong việc đưa ra các biện pháp nhằm ngăn chặn sự lây lan của dịch bệnh. Hy vọng qua báo cáo này, độc giả sẽ có được cái nhìn tổng quan về mô hình SIR, một trong những mô hình được ứng dụng rộng rãi, đồng thời hiểu thêm về các thuật toán được ứng dụng để ước lượng các chỉ số quan trọng trong đại dịch, cũng như nắm được những điểm khác biệt trong chiến lược phòng chống COVID-19 của Việt Nam.

Tài liệu

Kermack, W. O., & McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, 115(772), 700-721.

Bohner, M., Streipert, S., & Torres, D. F. (2019). Exact solution to a dynamic SIR model. *Nonlinear Analysis: Hybrid Systems*, 32, 228-238.

Brauer, F. (2008). Compartmental models in epidemiology. In *Mathematical epidemiology* (pp. 19-79). Springer, Berlin, Heidelberg.