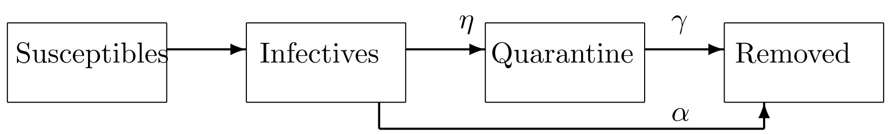
Project: An SIQR model for COVID-19.

Introduction The project "COVID-19" presented an initial attempt to model the spread of COVID-19 with a SEIR (susceptibles-exposed-infectives-removed) compartment model. However, after that project was written it became more and more apparent that a significant number of people could have an *asymptomatic* case of COVID-19, and that these people could be infectious and circulating in the population without knowing they were sick. This was especially the case since testing was often limited to symptomatic individuals who also met other criteria like travel to areas where COVID-19 was known to exist or known exposure to a documented case. The inability to identify infectious people accurately makes it hard to use an SEIR model and reconcile it with the available data. Because of this, here we will look at a different kind of compartment model, denoted SIQR, where the compartments are susceptibles-infectives-quarantined-removed.

The SIQR model. As in an SIR or SEIR model, the S compartment consists of the susceptible individuals. From this compartment, an individual can move into the infectives compartment I, and this compartment includes asymptomatic, presymptomatic (infectious, but not yet showing symptoms, though they will eventually do so), and those with symptoms who have not yet been isolated so they are still able to transmit the disease to others. However, once an infectious person is identified as COVID-19 positive we assume that person is then isolated (either at home or in a hospital, voluntarily or involuntarily) and thus moves to the quarantined compartment Q. Such an individual can then no longer transmit the disease to anyone else. These four compartments are pictured below; the removed compartment R (recovered or dead individuals) is discussed in detail below (where it will be refined).

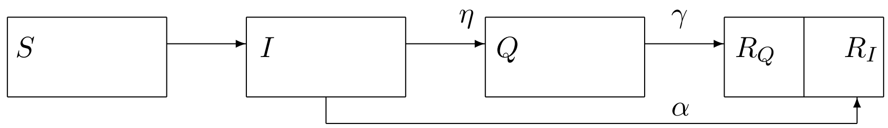


We let $S(t)$, $I(t)$, $Q(t)$, and $R(t)$ denote the number of people in each compartment at time t. Ignoring births or deaths due to non-COVID causes, we will assume that we have a community of total size N, which is constant. As a first model we will use the equations

$$\begin{aligned}
S'(t) &= -\frac{\beta}{N}SI \\
I'(t) &= \frac{\beta}{N}SI - (\alpha+\eta)I \\
Q'(t) &= \eta I - \gamma Q \\
R'(t) &= \gamma Q + \alpha I
\end{aligned}$$

Note that some infected persons are never identified as such and thus never move to the quarantined compartment.

Comparing the model to available data. We are going to use our model with data from Italy in the period from February 21, 2020 to April 22, 2020. Italy was the first European country to be severely affected by COVID-19, and containment/mitigation measures were instituted there in several steps. When an individual was identified as COVID-19 positive in Italy, the person was immediately isolated (at home or in a hospital), thus moving them from the I compartment to the Q compartment. Available data[1] gives the cumulative number of *identified* COVID-19 positive cases, and at time t would be given by the sum of $Q(t)$ and the number of individuals who have moved from compartment Q to compartment R by time t. Asymptomatic cases which appear in compartment I but never move to Q (i.e. are never identified) before eventually moving to R are "invisible". Thus to compare our model with available data we need to separate the R compartment into two sub-compartments, R_Q and R_I, of removed individuals from compartment Q and compartment I, respectively. In doing this, $Q(t) + R_Q(t)$ measures the cumulative number of observed cases at time t. Thus we modify our compartment diagram as follows:



The corresponding system of differential equations is now:

$$
\begin{aligned}
S'(t) &= -\frac{\beta}{N} SI \\
I'(t) &= \frac{\beta}{N} SI - (\alpha + \eta) I \\
Q'(t) &= \eta I - \gamma Q \\
R_Q'(t) &= \gamma Q \\
R_I'(t) &= \alpha I
\end{aligned}
$$

Parameter values. We will be modeling the spread of coronavirus in Italy from February 21 ($t = 0$) to April 22 ($t = 61$). Italy instituted an initial lockdown of the country (starting with the northern part) over a period of a few days in early March (March 8-10), and more stringent measures including a broad lockdown of non-essential workplaces took place on March 20. To reflect these changes in public health policy, we will consider our differential equation system over three time periods: February 21 to March 10 ($t = 0$ to $t = 18$), March 10 to March 20 ($t = 18$ to $t = 28$), and March 20 to April 22 ($t = 28$ to $t = 61$).

(a) Explain why the ratio $(R_Q(n) - R_Q(n-1))/Q(n-1)$ gives a rough approximation to the value of the parameter γ (for any integer value of $n \geq 1$). Also, explain why we have data available to compute these ratios, assuming that the Health Department in Italy provides daily figures for the number of active COVID-19 cases, the number of deaths, and the number of recovered individuals. The authors of the paper[2] compute the average of these ratios over the three time periods $0 \leq t \leq T_1$, $T_1 \leq t \leq T_2$, and $T_2 \leq t \leq T_3$ where T_1 is March 10, T_2 is March 20, and T_3 corresponds to April 22, and used these averages to give $\gamma_1 = 0.041$, $\gamma_2 = 0.038$, and $\gamma_3 = 0.023$ for the three time periods, respectively. These are the values we will use in our model. What are the units on γ?

(b) To understand the parameters α and η, first recall that moving from compartment I to compartment Q happens when an individual tests positive for COVID-19 or is presumed positive based on their symptoms. We estimate this happens for about 1/3 of the people in compartment I; evidence for this figure of 1/3 can be found in several situations where all individuals were tested and data were obtained

[1] From the Italian Department of Health at http://www.salute.gov.it/nuovocoronavirus

[2] Morten Gram Pedersen and Matteo Meneghini, Quantifying undetected COVID-19 cases and effects of containment measures in Italy: Predicting phase 2 dynamics, preprint March 2020, available at researchgate.net/publication/339915690.

on how many who tested positive were asymptomatic (e.g. passengers on the Diamond Princess cruise ship, or the residents of the Italian village Vo'Euganeo). Correspondingly, about 2/3 of all people in compartment I will move into compartment R_I. Regardless of whether a person moves from I to Q or from I to R_I, the average time spent in compartment I is 10 days. Thus, for the parameters η we obtain $\eta = \frac{1}{3}\frac{1}{10} \approx 0.0333$. What is the corresponding estimate for the parameter α? What are the units on η and α?

(c) The number $d = \frac{1}{(\alpha+\eta)}$ can be interpreted as the average number of days that an individual spends in the I compartment. Another important number is the **reproduction number** $R_0 = \frac{\beta}{(\alpha+\eta)}$. Its interpretation is that, on average, each infected individual will directly infect R_0 additional individuals over the course of their infection. At the early stages of the epidemic (this includes all times under consideration in this project), $S \approx N$ (almost everyone in the community is susceptible). If we accordingly make the approximation $S = N$ in our differential equation for I, we obtain the *single linear* equation

$$I'(t) = \beta I - (\alpha + \eta)I = \rho I \tag{1}$$

where $\rho = \beta - (\alpha + \eta)$. Your task here is to use this approximating linear equation to show that over the average time period that an individual remains in the I compartment, the infected population will grow (or shrink) approximately by a factor of $e^{(R_0-1)}$, that is, if t_0 is any time in the early stages of the epidemic, then show that

$$I(t_0 + d) \approx e^{(R_0-1)} I(t_0).$$

(d) Returning to the data available from the Italian Department of Health, we will determine the value of β in each of the three time periods $0 \leq t \leq T_1$, $T_1 \leq t \leq T_2$, and $T_2 \leq t \leq T_3$ (where $T_1 = 18$, $T_2 = 28$, and $T_3 = 61$) that best fits the available data. Parts (e)-(g) below outline how this is done. There you will obtain the values $\beta_1 = 0.2898$ for $0 \leq t \leq T_1$, $\beta_2 = 0.2018$ for $T_1 \leq t \leq T_2$, and $\beta_3 = 0.0818$ for $T_2 \leq t \leq T_3$. Compute the reproduction number $R_0 = \frac{\beta}{(\alpha+\eta)}$ for each of the three time periods.

(e) **Determining β_1.** Recall the approximation $I'(t) = \beta I - (\alpha + \eta)I = \rho I$ (Equation (1)), where $\rho = \beta - (\alpha + \eta)$. From (b) we have $\alpha + \eta = 0.1$. We are going to be considering three different values of β, and hence three different values of ρ, one for each time period $0 \leq t \leq T_1$, $T_1 \leq t \leq T_2$, and $T_2 \leq t \leq T_3$. We will denote these ρ_1 and β_1, ρ_2 and β_2, and ρ_3 and β_3, respectively.

Show that for $0 \leq t \leq T_1$ Equation (1) with initial condition $I = I(0)$ at $t = 0$ has solution

$$I(t) = I(0)e^{\rho_1 t}. \tag{2}$$

Since

$$Q'(t) = \eta I - \gamma Q$$

and

$$(R_Q)'(t) = \gamma Q$$

we have $(Q + R_Q)'(t) = \eta I = \eta I(0)e^{\rho_1 t}$ where we have used our approximation (2) for I. Show that with $Q(0) = 0$ and $R_Q(0) = 0$, this gives

$$(Q + R_Q)(t) = \frac{\eta I(0)}{\rho_1}(e^{\rho_1 t} - 1) \tag{3}$$

for $0 \leq t \leq T_1$, where $\rho_1 = \beta_1 - (\alpha + \eta) = \beta_1 - 0.1$. Using *Mathematica* you will determine values of ρ_1 (and hence β_1) and $I(0)$ to best fit the data from the Italian health department for $Q + R_Q$,

$0 \le t \le T_1$, with a function of the form

$$\frac{\eta I(0)}{\rho_1}(e^{\rho_1 t} - 1);$$

the *Mathematica* starter outlines the relevant commands. For use in part (f) we also compute $Q + R_Q$ at the end of the first time period as

$$(Q + R_Q)(T_1) = \frac{\eta I(0)}{\rho_1}\left(e^{\rho_1 T_1} - 1\right). \tag{4}$$

(f) **Determining** $\boldsymbol{\beta_2}$**.** Next you will determine the "best fit" value of β_2 for the time period $T_1 \le t \le T_2$. First we solve $I'(t) = \rho_2 I$ with $I(T_1) = I(0)e^{\rho_1 T_1}$ from Equation (2) to obtain

$$I(t) = I(0)e^{\rho_1 T_1}e^{\rho_2 (t - T_1)}$$

for $T_1 \le t \le T_2$. Note that

$$I(T_2) = I(0)e^{\rho_1 T_1}e^{\rho_2 (T_2 - T_1)}; \tag{5}$$

you will need this in (g).

Show that the solution to

$$(Q + R_Q)'(t) = \eta I(t) = \eta I(0)e^{\rho_1 T_1}e^{\rho_2 (t - T_1)}$$

having value

$$(Q + R_Q)(T_1) = \frac{\eta I(0)}{\rho_1}\left(e^{\rho_1 T_1} - 1\right)$$

(from Equation (4)) can be written as

$$(Q + R_Q)(t) = \eta I(0)\frac{e^{\rho_1 T_1}}{\rho_2}\left(e^{\rho_2 (t - T_1)} - 1\right) + \frac{\eta I(0)}{\rho_1}\left(e^{\rho_1 T_1} - 1\right). \tag{6}$$

Using the outline in the *Mathematica* starter, determine ρ_2 and hence also β_2, to best fit the data for the time period $T_1 \le t \le T_2$. In doing this, you will be using the values of η, $I(0)$, and ρ_1 already determined.

Note that from Equation (6) we can compute

$$(Q + R_Q)(T_2) = \eta I(0)\frac{e^{\rho_1 T_1}}{\rho_2}\left(e^{\rho_2 (T_2 - T_1)} - 1\right) + \frac{\eta I(0)}{\rho_1}\left(e^{\rho_1 T_1} - 1\right).$$

(g) **Determining** $\boldsymbol{\beta_3}$**.** Finally we determine our best fit value of ρ_3 and thus β_3 for the time period $T_2 \le t \le T_3$. To do this, first solve $I'(t) = \rho_3 I$ with

$$I(T_2) = I(0)e^{\rho_1 T_1}e^{\rho_2 (T_2 - T_1)}$$

to obtain

$$I(t) = I(0)e^{\rho_1 T_1}e^{\rho_2 (T_2 - T_1)}e^{\rho_3 (t - T_2)}.$$

Using this, solve

$$(Q + R_Q)'(t) = \eta I(0)e^{\rho_1 T_1}e^{\rho_2 (T_2 - T_1)}e^{\rho_3 (t - T_2)}$$

with initial condition

$$(Q + R_Q)(T_2) = \eta I(0)\frac{e^{\rho_1 T_1}}{\rho_2}\left(e^{\rho_2 (T_2 - T_1)} - 1\right) + \frac{\eta I(0)}{\rho_1}\left(e^{\rho_1 T_1} - 1\right)$$

(from (f)) and show that the result can be written as

$$(Q + R_Q)(t) = \eta I(0) \frac{e^{\rho_1 T_1}}{\rho_2} \left(e^{\rho_2 (T_2 - T_1)} - 1\right) + \frac{\eta I(0)}{\rho_1} \left(e^{\rho_1 T_1} - 1\right) + \frac{\eta I(0)}{\rho_3} e^{\rho_1 T_1} e^{\rho_2 (T_2 - T_1)} \left(e^{\rho_3 (t - T_2)} - 1\right).$$

All constants on the right-hand side are known except for ρ_3. Using the commands outlined in the *Mathematica* starter notebook to determine ρ_3, and hence β_3, to best fit our data for $T_2 \leq t \leq T_3$.

(h) Summarizing our parameter values we have

$$\alpha = 0.06667, \ \eta = 0.03333,$$

$$\gamma_1 = 0.041, \ \gamma_2 = 0.038, \ \gamma_3 = 0.023,$$

and

$$\beta_1 = 0.2898, \ \beta_2 = 0.2018, \ \beta_3 = 0.0818.$$

Using these values in our nonlinear SIQR system and the numerical solution feature in *Mathematica*, solve for $Q(t)$ and $R_Q(t)$ in each of the three time periods, and graphically compare your solution for $Q + R_Q$ to the provided data in each time period. The *Mathematica* starter indicates how to do this. Assume $Q(0) = 0$, $R_Q(0) = 0$, and as determined in (e), $I(0) = 2048$. Give a final graph that shows the solution for $Q + R_Q$ for $0 \leq t \leq 61$ together with the data points for actual cumulative cases (as given in the *Mathematica* starter).