Mathematical Statistic Project

Romà Domènech Masana A20450272

April 2020

Contents

1 Intro

Using the SIR model: Stands for Suscepible, Infected, Recovered. I got the information from (Gonçalves 2020) .

This model compartiments the population in three groups: *Susceptible, Infected and Recovered.* The transitions are:

$$\begin{gathered} S + I \xrightarrow{\beta} I \\ I \xrightarrow{\mu} R \end{gathered} \tag{1}$$

The first line represents the rate of infections, and the second line represents the rate of recovery. More precisely we could try to write down the fluctuations like this: For a given interval of time $[0, t]$, we could model the fluctuations like:

$$\begin{aligned} S_t = &\quad S_0 - \text{ new inf. during } [0,t] \\ I_t = &\quad I_0 + \text{ new inf. during } [0,t] - \text{ recovered during } [0,t] \\ R_t = &\quad R_0 + \text{ recovered during } [0,t]. \end{aligned} \tag{2}$$

And now comes the interesting part. How do we model the **new infections during** the interal of time $[0, t]$?

1. Well, we could say that the longer the interval, the more infections we will probably get. So we can call this amount $NI_{[0,t]}$ and make it linearly dependent on time like this

$$NI_{[0,t]} \propto t.$$

2. But then, observe that it also makes sense to think that the new infections will be proportional to the current amount of people that can be infected. That's to say S_0. So so far we have

$$NI_{[0,t]} \propto tS_0$$

3. Now, finally observe that the people get infected from other infected people. So I_0 should play a role here. So actually, the bigger the density of infected people among the population, the bigger the chance of getting infected. So we will make $NI_{[0,t]}$ also proportional to I_0/N, where N is the total amount of alive individuals. This makes our formula

$$NI_{[0,t]} \propto tS_0 \frac{I_0}{N}.$$

4. Finally just multiply the obtained formula by a constant. Let's call this constant β and we get that

$$NI_{[0,t]} = tS_0 \tfrac{I_0}{N}\beta. \tag{3}$$

Similarily we could model the new recovered people during this interval of time $[0,t]$.

1. As time passes more infected people recover from the disease unavoidingly. So this quantity that we can call $NR_{[0,t]}$ must be

   $$NR_{[0,t]} \propto t$$

   as well.

2. Now we don't see that the current or past amount of recovered people makes infected people recover in any way. So S_t or S_0 cannot play a role here.

3. Finally, just observe that the amount of newly recovered people will be a fraction or portion from the currently infected people. So we can write

   $$NR_{[0,t]} \propto tI_t$$

   or, in other words,

   $$NR_{[0,t]} = tI_t\mu, \tag{4}$$

   for some constant μ.

Now, this gives us the change in an interval of time $[0,t]$. If we want to compute the infinitessimal variation or derivative, with respect to time, we could just take $\frac{\partial}{\partial t}S_t$, $\frac{\partial}{\partial t}I_t$, $\frac{\partial}{\partial t}R_t$ on Equation 2 and we would obtain:

$$\begin{aligned} \tfrac{\partial}{\partial t}S_t =& \quad -S_0\tfrac{I_0}{N}\beta \\ \tfrac{\partial}{\partial t}I_t =& \quad S_0\tfrac{I_0}{N}\beta - I_0\mu \\ \tfrac{\partial}{\partial t}R_t =& \quad I_0\mu. \end{aligned} \tag{5}$$

Now, if we are taking the derivative we are actually assuming that the points 0 and t are infinitely close, so that *we can actually assume they are the same*[1], so the derivatives become:

$$\begin{aligned} \tfrac{\partial}{\partial t}S_t =& \quad -S_t\tfrac{I_t}{N}\beta \\ \tfrac{\partial}{\partial t}I_t =& \quad S_t\tfrac{I_t}{N}\beta - I_t\mu \\ \tfrac{\partial}{\partial t}R_t =& \quad I_t\mu. \end{aligned} \tag{6}$$

1.1 Density of infected population stays close to 0

In other words, if we assume that the amount of people to be infected is almost the 100% of the population, or

$$\frac{S_t}{N} \simeq 1.$$

This allows us to solve the Infected population term I_t in Equation 6, and we obtain, with this assumption:

$$\tfrac{\partial}{\partial t}I_t = I_t(\beta - \mu), \tag{7}$$

which has solution:

$$I_t = I_0 e^{(\beta-\mu)t} = I_0 e^{\mu(R_\theta - 1)t} \tag{8}$$

where

1. β is as we mentioned before the constant indicating the proportion of new infected individuals at every instant of time.

2. μ is the contant indicating the proportion of new recoveries at every instant of time.

[1] I am actually not that sure of this step

We also wanted to note that in order to do the regression and find out the parameter R_0, we prefer to write the above equation as

$$\log I_t = \log I_0 + \mu(R_\theta - 1)t = b_0 + b_1 t. \tag{9}$$

So, from the above description it is difficult to avoid asking yourself, if

$$R_\theta = \frac{\beta}{\mu},$$

does the infection kind of stop when $R_\theta < 1$ and spread when $R_\theta > 1$? This is the kind of quantity that the governments and health organization want to know and look at to have an idea of the instant danger of the illness on their country.

1.1.1 Solving S_t for this case

If we solve Equation 6 given the condition $S_t/N \simeq 1$, for S_t, we obtain that (using the solution Equation 8)

$$\begin{aligned} S_t = & \quad Ce^{-\frac{\beta}{\mu}\frac{I_0}{N}\frac{1}{R_0-1}e^{\mu(R_0-1)t}} \\ = & \quad S_0 e^{-\frac{\beta}{\mu}\frac{I_0}{N}\frac{1}{R_0-1}\left(e^{\mu(R_0-1)t}-1\right)} \end{aligned} \tag{10}$$

2 Solving the SIR epidemic model analytically

As derived by (Harko, Lobo, and Mak 2014) we can convert the set of differential equations Equation 6 into a Bernoulli equation that we will finally leave as implicit integration. The steps are the following:

1. Take the first equation of the set Equation 6 and write it as

$$-\tfrac{N}{\beta}\tfrac{S'_t}{S_t} = I_t \tag{11}$$

2. Now, if we take the derivative of this with respect to time we obtain

$$-\tfrac{N}{\beta}\left(\tfrac{S''_t}{S_t} - \left(\tfrac{S'_t}{S_t}\right)^2\right) = I'_t. \tag{12}$$

3. If we plug in these to in the second equation of the set of differential equations we have been using so far, we obtain the following

$$\begin{gathered} -\tfrac{N}{\beta}\left(\tfrac{S''_t}{S_t} - \left(\tfrac{S'_t}{S_t}\right)^2\right) = -(\tfrac{S_t}{N}\beta - \mu)\tfrac{N}{\beta}\tfrac{S'_t}{S_t} \\ \downarrow \\ -\tfrac{N}{\beta}\left(\tfrac{S''_t}{S_t} - \left(\tfrac{S'_t}{S_t}\right)^2\right) = -S'_t + \mu\tfrac{N}{\beta}\tfrac{S'_t}{S_t} \end{gathered} \tag{13}$$

4. Now from the third equation of the original set we get that

$$R'_t = -\tfrac{N\mu}{\beta}\tfrac{S'_t}{S_t}. \tag{14}$$

5. This can be integrated and allows us to say that

$$\begin{gathered} R_t = -\tfrac{N\mu}{\beta}\log S_t + C_1 \\ \downarrow \\ S_t = C_1 e^{-\frac{\beta}{N\mu}R_t} \\ \downarrow \\ S_t = S_0 e^{-\frac{\beta}{N\mu}(R_t - R_0)} \end{gathered} \tag{15}$$

6. From this last equation we can differentiate with respect to time and obtain that

$$S_t' = -\tfrac{\beta}{N\mu} S_0 e^{-\frac{\beta}{N\mu}(R_t - R_0)} R_t'. \tag{16}$$

We can also differentiate Equation 14 in order to obtain

$$R_t'' = -\tfrac{N\mu}{\beta}\left(\tfrac{S_t''}{S_t} - \left(\tfrac{S_t'}{S_t}\right)^2\right). \tag{17}$$

7. Now, if we substitute the last two equations that we got for S_t' and R_t'', but also the one that we got for R_t' on Equation 13, we obtain the following equation in terms of R_t, R_t' and R_t'' only:

$$\begin{gathered} \tfrac{1}{\mu} R_t'' = \tfrac{\beta}{N\mu} S_0 e^{-\frac{\beta}{N\mu}(R_t - R_0)} R_t' - R_t' \\ \downarrow \\ R_t'' = R_t' \left(\tfrac{\beta}{N} S_0 e^{-\frac{\beta}{N\mu}(R_t - R_0)} - \mu\right). \end{gathered} \tag{18}$$

8. Then this equation is the one I said that can be transformed into a Bernoulli one with $\alpha = 2$. To see that, we will take two variable transformations. The first one consists on taking

$$\begin{gathered} u(t) = e^{-\frac{\beta}{N\mu} R_t} \\ \downarrow \\ R_t = -\tfrac{N\mu}{\beta}\log(u_t) \\ \downarrow \\ R_t' = -\tfrac{N\mu}{\beta}\tfrac{u_t'}{u_t} \\ \downarrow \\ R_t'' = -\tfrac{N\mu}{\beta}\left(\tfrac{u_t''}{u_t} - \left(\tfrac{u_t'}{u_t}\right)^2\right) \end{gathered} \tag{19}$$

9. But we just obtained to different expressions for R_t''! If we equal them together, and then plug in the value we derived of R_t' in terms of u_t and u_t', we get that

$$\begin{gathered} -\tfrac{N\mu}{\beta}\left(\tfrac{u_t''}{u_t} - \left(\tfrac{u_t'}{u_t}\right)^2\right) = -\tfrac{N\mu}{\beta}\tfrac{u_t'}{u_t}\left(\tfrac{\beta}{N} S_0 e^{\frac{\beta R_0}{N\mu}} u_t - \mu\right) \\ \downarrow \\ \tfrac{u_t''}{u_t} - \left(\tfrac{u_t'}{u_t}\right)^2 = \tfrac{u_t'}{u_t}\left(\tfrac{\beta}{N} S_0 e^{\frac{\beta R_0}{N\mu}} u_t - \mu\right) \\ \downarrow \\ u_t u_t'' - (u_t')^2 = u_t' u_t \left(\tfrac{\beta}{N} S_0 e^{\frac{\beta R_0}{N\mu}} u_t - \mu\right) \end{gathered} \tag{20}$$

10. We will do the last change of variables. Take

$$\begin{gathered} \phi := \tfrac{dt}{du} = \tfrac{1}{u_t'} \\ \downarrow \\ \tfrac{d\phi}{du} = \phi' \tfrac{dt}{du} = \phi' \phi. \end{gathered} \tag{21}$$

In particular this implies that

$$\begin{gathered} u_t' = 1/\phi \\ \downarrow \\ u_t'' = -\tfrac{1}{\phi^2}\phi' = -\tfrac{1}{\phi^3}\phi'\phi = -\tfrac{1}{\phi^3}\tfrac{d\phi}{du} \end{gathered} \tag{22}$$

11. Now, if we substitute these last expressions in the last equation we got relating u_t'', u_t' and u_t we obtain:

$$\begin{gathered} -\tfrac{u_t}{\phi^3}\tfrac{d\phi}{du} - \tfrac{1}{\phi^2} = \tfrac{u_t}{\phi}\left(\tfrac{\beta}{N} S_0 e^{\frac{\beta R_0}{N\mu}} u_t - \mu\right) \\ \downarrow \\ -u_t \tfrac{d\phi}{du} - \phi = u_t \phi^2 \left(\tfrac{\beta}{N} S_0 e^{\frac{\beta R_0}{N\mu}} u_t - \mu\right) \\ \downarrow \\ \tfrac{d\phi}{du} + \tfrac{\phi}{u_t} = \phi^2\left(\mu - \tfrac{\beta}{N} S_0 e^{\frac{\beta R_0}{N\mu}} u_t\right). \end{gathered} \tag{23}$$

12. A Bernoulli differential equation of $b(a)$ has the following form:
$$\tfrac{db}{da} + P(a)b = b^\alpha Q(a). \tag{24}$$
Therefore we obtained a Bernoulli equation with $b = \phi, a = u_t$ and $\alpha = 2, P(a) = 1/a, Q(a) = \frac{\beta}{N} S_0 e^{\frac{\beta R_0}{N\mu}} a - \mu$. This type of equations can be linearized, and then solved by using an integrating factor. The solution for this one (I didn't really compute it, I took this part from the paper (Harko, Lobo, and Mak 2014)) is
$$\begin{array}{c} \frac{dt}{du} = \phi = \frac{1}{u_t\left(C_2 - \mu \log u_t + \frac{\beta}{N} S_0 e^{\frac{\beta R_0}{N\mu}} u_t\right)} \\ \downarrow \\ t = \int^u \frac{1}{s\left(C_2 - \mu \log s + \frac{\beta}{N} S_0 e^{\frac{\beta R_0}{N\mu}} s\right)} ds + C_3 \end{array} \tag{25}$$

13. I want to rewrite now the initial system of differential equations Equation 6 in terms of u_t only, so that if we solve the last equation for u_t, we will have actually solved analytically the simplified SIR model. From the change of variables
$$u(t) = e^{-\frac{\beta}{N\mu} R_t}$$
that we did, we obtain that
$$R_t = -\frac{N\mu}{\beta} \log u_t.$$
Now from the last form of Equation 15 we can see that
$$S_t = S_0 e^{\frac{\beta}{N\mu} R_0} u_t,$$
and from the fact that
$$\begin{aligned} I_t &= -\tfrac{N\mu}{\beta} \tfrac{S'_t}{S_t} \tfrac{1}{\mu}, \\ \tfrac{N\mu}{\beta} \tfrac{S'_t}{S_t} &= R'_t, \end{aligned}$$
we get that
$$I_t = \tfrac{1}{\mu} R'_t.$$
Now using the derived expression for R'_t in terms of u_t, u'_t we obtain that
$$\begin{aligned} I_t = \; & -\tfrac{N}{\beta} \tfrac{u'_t}{u_t} = -\tfrac{N}{\beta} \tfrac{1}{u_t \phi} \\ & = -\tfrac{N}{\beta u_t} u_t \left(C_2 - \mu \log u_t + \tfrac{\beta}{N} S_0 e^{\frac{\beta R_0}{N\mu}} u_t\right) \\ & = -\tfrac{N}{\beta} C_2 + \tfrac{N\mu}{\beta} \log u_t - S_0 e^{\frac{\beta R_0}{N\mu}} u_t \end{aligned}$$

Thus we transformed the original model into the following
$$\begin{aligned} S_t &= S_0 e^{\frac{\beta}{N\mu} R_0} u_t, \\ R_t &= -\tfrac{N\mu}{\beta} \log u_t, \\ I_t &= -\tfrac{N}{\beta} C_2 + \tfrac{N\mu}{\beta} \log u_t - S_0 e^{\frac{\beta R_0}{N\mu}} u_t. \end{aligned} \tag{26}$$
Where u_t is the integration limit of this equation:
$$t = \int^u \frac{1}{s\left(C_2 - \mu \log s + \frac{\beta}{N} S_0 e^{\frac{\beta R_0}{N\mu}} s\right)} ds + C_3, \tag{27}$$
and C_2, C_3 are two constants of integration that we should set by using the initial value problem for I_t and u_t.

2.1 Solving u_t

In order to solve the last equation and get the explicit value of u we should first

1. Integrate numerically the equation in order to obtain
$$t = f(u)$$
for some function f. And

2. Invert numerically the previous equation in order to obtain
$$u_t = f^{-1}(t)$$

Even if we manage to do all this, the problem is: **How do we fit any coefficients then?** Because the results will be numerical and we can't spot any constant in the formula that we could estimate using statistical tools.

3 Appling the SIR model to the data from Spain

In Spain there are 46.94 million people. This would be our N, where

$$S_t + I_t + R_t = N = 46'940'000 \tag{28}$$

Since we don't consider any births, or deaths, we actually consider the deaths from Covid as part of the recovered variable R_t. Thus we will only be focusing on how many people can get infected, how many people are infected, and how many people left this stage (either by death or by recovery) this last quantity is what we called R_t in the text. In other words:

$$\begin{aligned} S_t &= N - \text{ coronavirus cases} \\ R_t &= \text{ recovered + deaths} \\ I_t &= \text{ coronavirus cases } - R_t \end{aligned} \tag{29}$$

In this case

$$\tfrac{S_t}{N} = \tfrac{46.70M}{46.94M} \simeq 0.995, \tag{30}$$

So I think it is reasonable to make the assumption $\frac{S_t}{N} \sim 1$ and use the simplified version of the SIR model of Equation 8.
So if we want to fit and find the value R_θ in the equation

$$\log I_t = \log I_0 + \mu(R_\theta - 1)t = b_0 + b_1 t. \tag{31}$$

we first approximate the value for μ by doing

$$\mu \simeq \text{ median } \tfrac{R_{t+1} - R_t}{I_t} \simeq 0.037, \tag{32}$$

which is motivated by the third of the equations of the system Equation 6.

Now if we do the fit of Equation 31 for all the data, we obtain a

$$\begin{aligned} b_0 &= 2.067 \\ b_1 &= 0.1622 \\ R_\theta &= \tfrac{b_1}{\mu} + 1 \simeq 5.39 \end{aligned} \tag{33}$$

This means that up to today there is a rate of infection of 500% which is too much I think. Because this means that each person would infect around 5.39 people.

If instead we fit Equation 31 for **the last 15 days**, we obtain a

$$\begin{aligned} b_0 &= 11.98 \\ b_1 &= -0.0083 \\ R_\theta &= \tfrac{b_1}{\mu} + 1 \simeq 0.776 \end{aligned} \tag{34}$$

This means that the most recent data we have is that there is a rate of infection of 78% which means that the disease is not infectious anymore because the average amount of infected people is less than 1.

4 Plots for the exponential regression

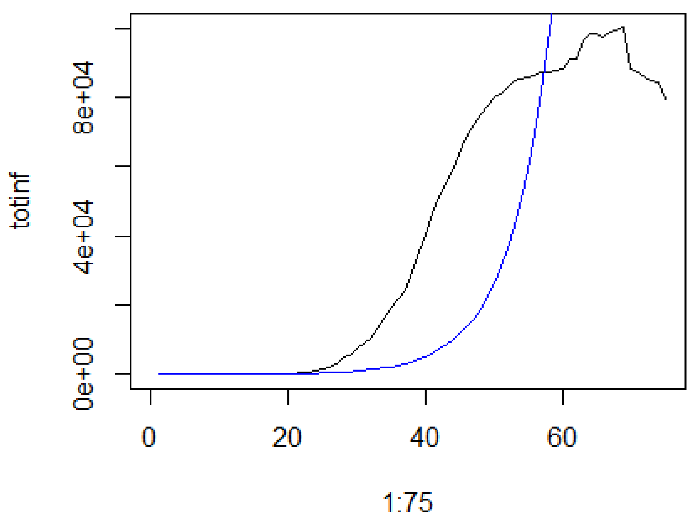

Figure 1: Total series of infected people in Spain over the last 75 days. We include in blue the first fit $I_t = I_0 e^{\mu(R_\theta - 1)t} = \log 2.06 e^{0.037*(5.39-1)}$.

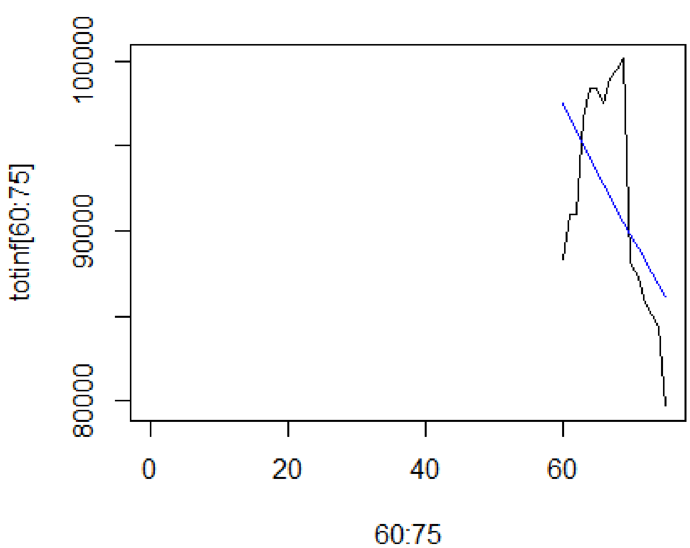

Figure 2: Total series of infected people in Spain over the last 15 days. We include in blue the second fit $I_t = \log 11.98 e^{0.037(0.776-1)t}$.

References

[1] Gonçalves, Bruno. "Epidemic Modeling 101: Or why your CoVID-19 exponential fits are wrong". In: *Medium: Data For Science* (2020).

[2] Harko, T., Lobo, F., and Mak, M. "Exact analytical solutions of the Susceptible-Infected-Recovered (SIR) epidemicmodel and of the SIR model with equal death and birth rates". In: *arxiv.org* (2014).