LECTURE NOTES
OF
LIFE DATA EPIDEMIOLOGY

Collection of the lectures notes of professors Chiara Polletto and Sandro Meloni.

Edited By

ALICE PAGANO

The University of Padua

Contents

Part I

Meloni's Lectures

1

Basics Definitions and Compartmental Models

> *Some models are wrong, but most of them are useful.*
>
> – Unknown author

Models in science have two different roles: **understanding** what happens and **prediction**. There are two types of models: one more simple and one more complex. In the simplest one you just consider the minimal number of parameters and events involved: this allow to understand what are the main mechanism of a phenomena.

In this course, we gonna start with very simple models in which we firstly assume that there are no structures in the population. This is not accurate, but allows us to understand at a first glance some underlying mechanism. Then, we will consider social structures and contact network models. We will also take into account the interaction of different populations and we analyze how data move from one population to the other. At last, there is another class of models called "Agent Based" models, but we will only brief introduce it.

1.1 Comportamental models

We introduce the **comportamental models** in which most based epidemiologist theories are based on. In reality, there are different levels of understanding how the diseases diffuse, as for instance at a biological level or a more simplest one. It is impossible to insert all the details of a process in a models. We need to summarize all the biological process in few parameters which are the average of what you see inside the population. This is the same principle behind the statistical mechanics in which we concentrate for large scale effects.

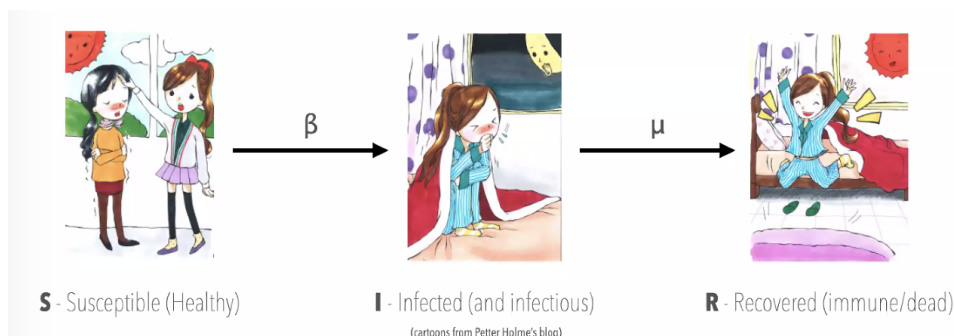Let us consider a population of any individuals and try to characterize them. For

S - Susceptible (Healthy)   **I** - Infected (and infectious)   **R** - Recovered (immune/dead)

(cartoons from Petter Holme's blog)

Figure 1.1: Classification of infected population in three different stage of the disease.
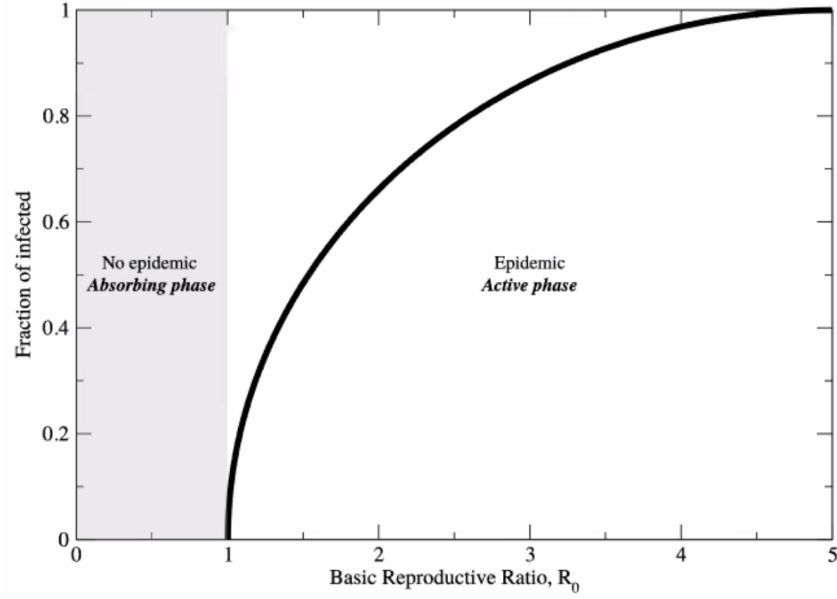
Figure 1.2: Fraction of infected vs basic Reproductive Ratio, R_0.

instance we have three compartments and we want to classify people according to the state of their diseases, as in Fig. 1.1. We have also transition from one state to the other with rates as β and μ which describes the dynamic of the diseases The only problems it is that this approximation is quite strong, because by fixing the rates we are assuming that the process underlying is a Markovian one. In reality, you do not have a exponential distribution but a decay. This will be seen in the course when we will talk about "non-Markovian" epidemics. For instance the β can be the "per contact" infectous rate, so you are just counting the number of contacts. As examples we can consider different models which depend on the type of disease as SI, SIS, SIR and so on and so forth.

There are difference between medical status and infection status. We do not care about medical status but only about the disease and how the immuning system is reacting. We have four main stages: starting from a helthy state, the individual contract the disease and then it become infectious until it recover. The most important things is that the compartments are not the same of the medical status.

Now, let us introduce the **Basic Reproductive Number** R_0 (pronunced R naught) which is a measure of the infection of the population. We put one guy inside a group for instance for three days and at the end of these days we count the number of secondary cases that we have. This is the main idea of R_0. This number determines wheter a disease will spread or not:

$$\begin{cases} R_0 < 1 \\ R_0 = 1 \\ R_0 > 1 \end{cases} \tag{1.1}$$

Let us consider the plot of Fig. 1.2, we have a sort of second order phase transition at the point $R_0 = 1$. Note that the R_0 number for the SARS is higher thant the one of COVID-19, but we did not experienced an outbreak of this disease. For computing R_0 we are assuming that the population is totally susceptible, but this assumption is valid only at the very early stages. Then, we care ombining also epidemiological and demographical features. The conclusion is that this number could vary from one population to another. Since we are doing a coarse-graining of the dynamics, the number R_0 represents the mean of possibly different distributions. It can give the

wrong idea of similar R_0, means similar outbreaks. The distribution of infections can be quite heterogeneous, hence the mean could be quite representative in homogeneous populations. For instance, let us consider the plot in Fig. 1.3. We see that SARS is etherogeneous, while Spanish Flu was a homogeneous one. The COVID-19 is most likely somewhere in the middle.
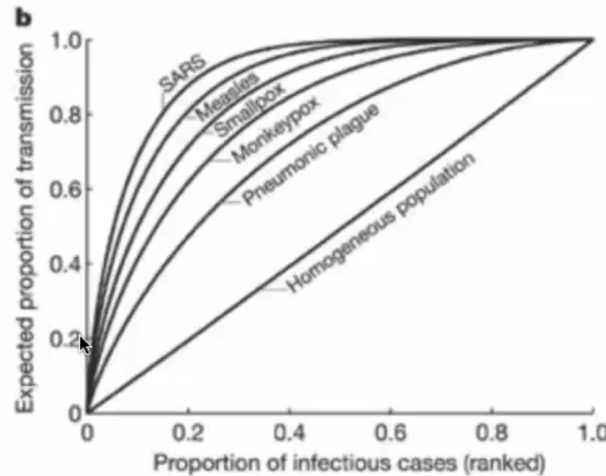

Figure 1.3: Figure from: Lloyd-Smith et al. Nature 438, 355–359 (2005).

Now, we introduce the **Effective Reproductive Number** $R(t)$, which is the same of R_0 but it varies in time. Hence, it is the average number of secondary cases a case produces in a population at time t.

Another important quantity is the **Infectious period**:

$$\tau = \frac{1}{\mu}, \quad \tau = \frac{1}{(\alpha + \mu)} \tag{1.2}$$

Other relevant quantities are the **Incubation period** (from infection to symptoms), the **Generation time** (from infection to infection), the **Serial interval** (from symptoms to symptoms) or the **TOST** (measures from symptoms to infection). The problems is that TOST in many cases can be negative (see Fig. 1.4 for more details).
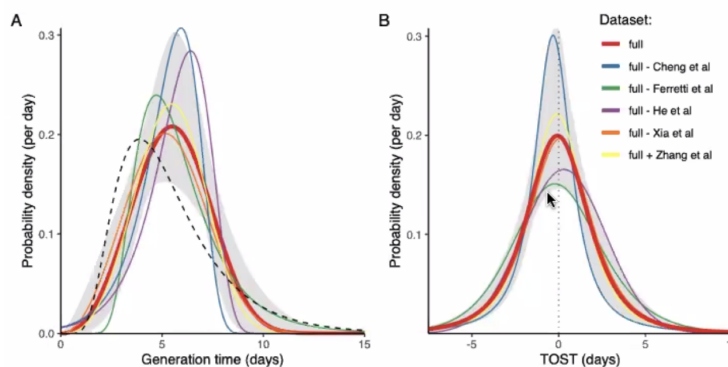

Figure 1.4: Figure from: Ferretti et al. https://www.medrxiv.org/content/10.1101/2020.09.04.20188516v1

1.2 Basic models

We are gonna introduce some of the basic models we will deal for the entire course. We are assuming that we are in **well-mixed populations**, or homogeneous mixing.

Mathematically, it is what is called mean field approximation. In the well-mixed population assumptions, we are assuming that:

- All individuals are equivalent, hence every one has the same probability of getting the infection;

- every individual has the same number of contacts $N - 1$, or on average $\langle k \rangle$;

- another important assumption is that we are in a closed population. Hence, the sum of the density distribution of the individuals is 1 and we have no deaths or births. In practice, we are assuming that our time scale is so little that we can consider the population constant.

1.2.1 SI model

This simple model is the **SI** (susceptible infected). You can get the infection and once you get it you cannot recover (you stay infected forever). The transition is:

$$S + I \xrightarrow{\beta} I + I$$
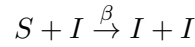
The β is the "per contact" infection rate and dictates the speed of the spreading. We can write down the equation and solve it deterministically:

$$\frac{\mathrm{d}s}{\mathrm{d}t} = -\beta \langle k \rangle \, si, \qquad \frac{\mathrm{d}i}{\mathrm{d}t} = \beta \langle k \rangle \, si$$

where $\langle k \rangle$ represents the contacts, the i means the fraction of infected in the population ($i = I/N$), while s the fraction of susceptible in the population ($s = S/N$). The product si is the probability of contacts, and βsi is the proabability of having one more infected.

To solve it analitically, we should remember that our population is closed hence $s + i = 1$. See the calculus in Fig. 1.5. The result is:

Since $s + i = 1$ we only have one equation with $s = 1 - i$

$$\frac{di}{dt} = \beta i(1 - i) \longrightarrow \frac{1}{\beta i(1 - i)} di = dt \longrightarrow \frac{1}{\beta(1 - i)} di + \frac{1}{\beta i} di = dt$$

Integrating both sides

$$-\log|1 - i| + \log|i| = \beta(t + C) \longrightarrow \frac{i}{1 - i} = e^{\beta(t+C)} = Ae^{\beta t} \quad \text{With} \quad A = i_0/(1 - i_0)$$

Thus

$$i(t) = \frac{i_0 e^{\beta t}}{1 - i_0 + i_0 e^{\beta t}}$$

Figure 1.5: Calculation of the analitic solution for the SI model.

$$i(t) = \frac{i_0 e^{\beta t}}{1 - i_0 + i_0 e^{\beta t}}$$

which is a sigmoid function (Fig. 1.6) which always saturates at 1. We have the first part where we have the exponential growth (which is the one we have seen in the media for covid-19), then at a certain point you are slowing down. The reason of slowing down is because of the term si, the probability of funding new supsceptible is going down. Finally, you reach 1 after a very long time. The value of β is the one which drives the spreading. Increasing β we have a faster exponential growth. This was the simplest model.
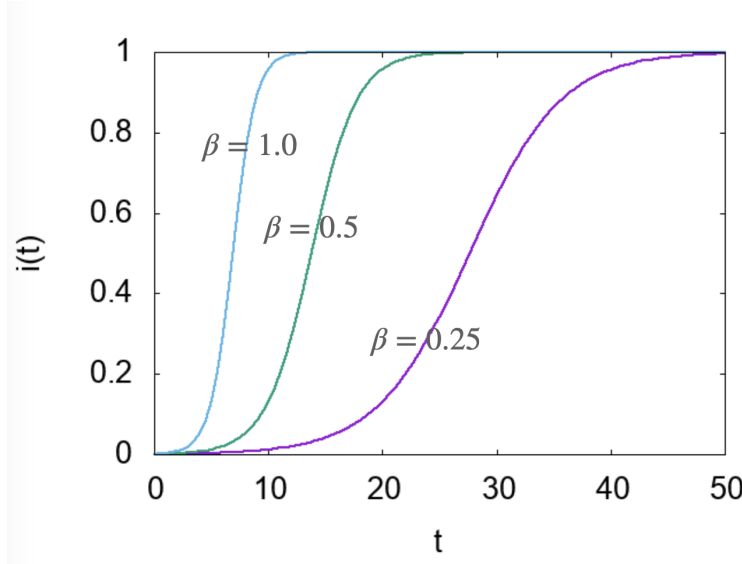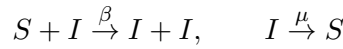
Figure 1.6: Plot of the solution of the SI model for different β.

Remark. In the course we are gonna use capital letter for integer numbers and small lecter for densities.

1.2.2 SIS model

Now, let us go to the **SIS** model. This model starts to be more complicated. We have two different transitions:

$$S + I \xrightarrow{\beta} I + I, \qquad I \xrightarrow{\mu} S$$

whose second one is spontaneous. This model is used for diseases that do not confer immunity. **Endemic state** means that the disease circulates in the population for very large times. The important things is that it is the simplest models in which a dynamical equilibrium can be reached. An individual could recover after the disease but it do not get immunity, indeed there are always people infected that can propagate the disease. The μ is the recovery rate which determines the time-scale of the infection. Dividing β by μ you can rescale all the dynamics. The equations are exactly the same of before except for a term:

$$\frac{\mathrm{d}s}{\mathrm{d}t} = -\beta \langle k \rangle \, si + \mu i\,, \qquad \frac{\mathrm{d}i}{\mathrm{d}t} = \beta \langle k \rangle \, si - \mu i$$

and you can solve them in the way of before. Also the shape of the solution is exactly the same:

$$i(t) = i_0 \frac{(\beta - \mu)e^{(\beta-\mu)t}}{\beta - \mu + \beta i_0 e^{(\beta-\mu)t}}$$

If we plot it we have the same form but with the difference that we are not reaching one, but $\frac{\beta-\mu}{\beta}$. Hence, as said, we have some sort of dynamical equilibrium: the new infected are the same of the new recovery that you are getting. The population will fluctuate around this value $\frac{\beta-\mu}{\beta}$ and enlarging μ will give to larging fluctuations (Fig. 1.7).

It could be more instructive to study what happens at the beginning for this model. At the beginning I can assume that almost my population is composed by my susceptible ($s \sim 1$) and the number of infected is very little ($i \ll 1$). Hence, we can rewrite the differential equations as:

$$\frac{\mathrm{d}i}{\mathrm{d}t} = \beta \langle k \rangle \, si - \mu i \sim \beta \langle k \rangle \, i - \mu i \rightarrow i(t) \sim i_0 e^{(\beta \langle k \rangle - \mu)t}$$

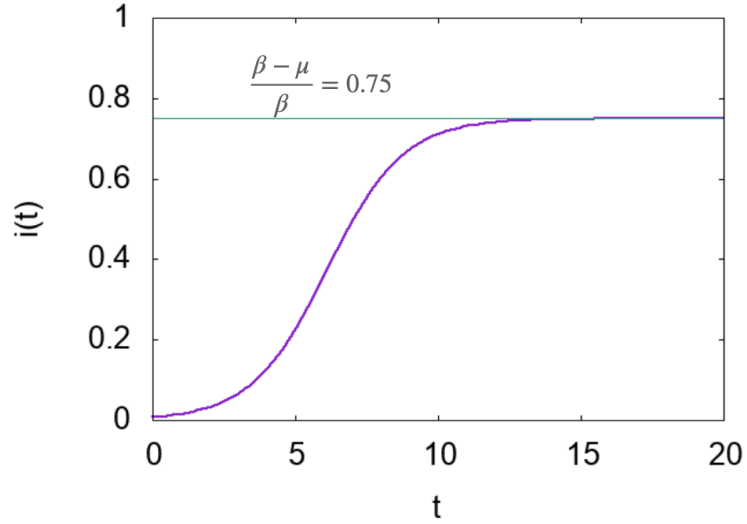**Figure 1.7:** Plot of the solution of the SIS model.

We have that if $\beta \langle k \rangle < \mu$ I have not spreading at this point, while if $\beta \langle k \rangle > \mu$ the exponential becomes positive and I have the exponential growing at the beginning (Fig. 1.8).
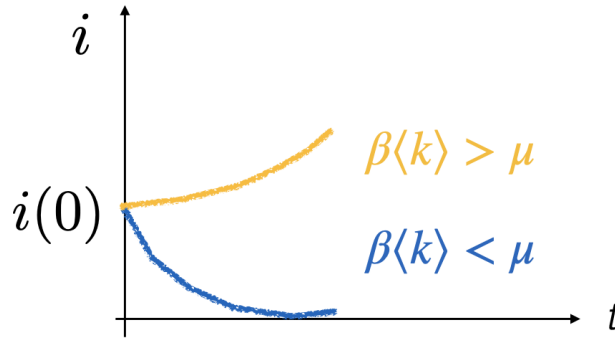

Figure 1.8: Initial transient for the SIS model.

The very important thing is that if I consider what it is happening I have two choices for the steady state:

$$\frac{\mathrm{d}i}{\mathrm{d}t} = 0 \rightarrow \begin{cases} i = 0 & \beta \langle k \rangle < \mu \\ i > 0 & \beta \langle k \rangle > \mu \end{cases}$$

and we have that:

$$i > 0 \iff \beta > \beta_c = \frac{\mu}{\langle k \rangle}$$

which is the **epidemic threshold**. This is telling you if the disease is gonne spread. The epidemic threshold is the minimum value of the infection probability for which the disease survives. This is what in physics is called a second order phase transition (Fig. 1.9). In this case the critical exponents are the same of the Ising model (they are in the same class of universality). This is one of the most important quantities we are gonna study.

What is the relation between R_0 and the epidemic threshold? Obviously, they are strongly correlated. We are saying that we have a critical value and below it we have no spreading, while above we have a fraction of infected people. The epidemic threshold is given you the condition under which you have the spreading. Mathematically,
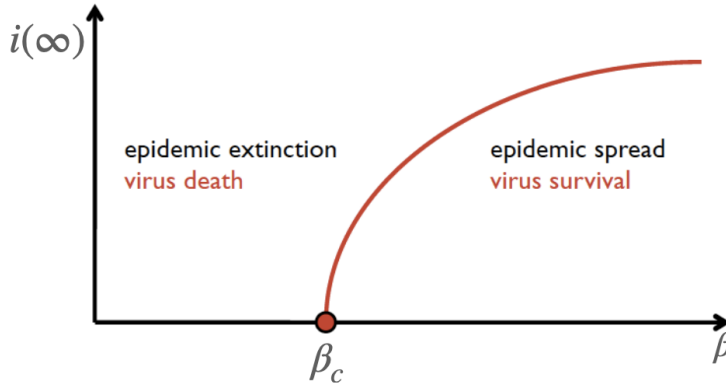
Figure 1.9: Epidemic diagram.

giving a speficic model, its critical version is giving the value for which $R_0 = 1$, which means that if you are above the threshold you need a minimum of infected people which is 1. In the case of the SIS model:

$$R_0 = \frac{\beta \langle k \rangle}{\mu} = 1$$

1.2.3 SIR model

The idea is the same of the SIS, but we are adding a new state which accounts for long lasting immunity. Hence, once you got the disease you can have a long immunity. However, the density of the population is still fixed to 1. The transitions of this model are:

$$S + I \xrightarrow{\beta} I + I, \qquad I \xrightarrow{\mu} R$$
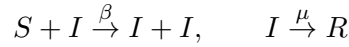
and we have no endemic state. The differential equations are:

$$\frac{\mathrm{d}s}{\mathrm{d}t} = -\beta \langle k \rangle si, \qquad \frac{\mathrm{d}i}{\mathrm{d}t} = \overbrace{\beta \langle k \rangle si}^{\text{New infections}} - \overbrace{\mu i}^{\text{Recovery}}, \qquad \frac{\mathrm{d}r}{\mathrm{d}t} = \mu i$$

It is a good point to introduce the different regimes that you have during a spreading and which are represented in Fig. 1.10 for the SIS and SIR models. Initially, at the beginning of each spreading, you have the **noisy phases** where the numbers are too small to cause a spreading, hence you have a sort of stocastic fluctuations. In most of the cases, you can end up without infected. If you stop a guy in this noisy phase, you are able to stop the disease (if it is hetherogeneous). Then, we have the **exponential growth**. Then, the disease is slowing down. Finally, you reach the steady state for the SIS (endemic state), while for the SIR the disease disappear (absorbing state).

To calculate the epidemic threshold in the case of the SIR the calculations are the same of before. In particular, we assume that $r \ll 1$ so that:

$$\frac{\mathrm{d}i}{\mathrm{d}t} = \beta \langle k \rangle si - \mu i \sim \beta \langle k \rangle i - \mu i \to i(t) \sim i_0 e^{(\beta \langle k \rangle - \mu)t}$$

and the result is again:

$$\beta > \beta_c = \frac{\mu}{\langle k \rangle}$$

Since we can get an analitic expression for S and I in this SIR model, we want to study what is the behavior at the end for $t = \infty$. We get that:

$$\frac{\mathrm{d}s}{\mathrm{d}r} = \frac{-\beta \langle k \rangle s}{\mu}$$
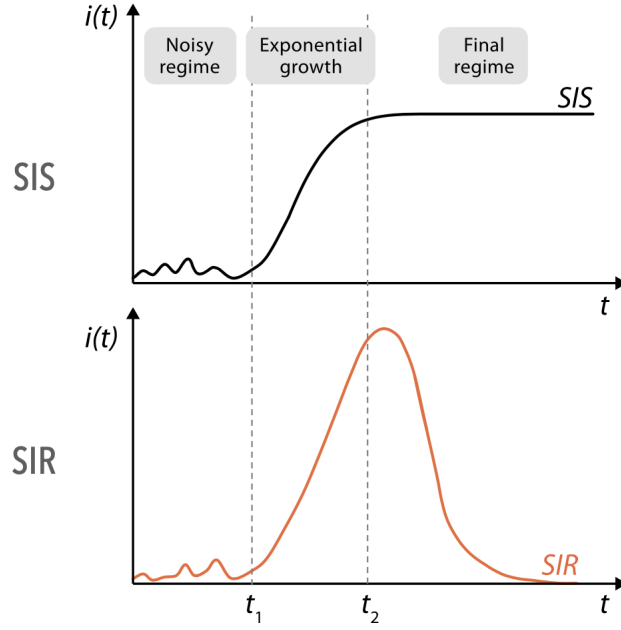
Figure 1.10:   Epidemic regimes.

If we assume $r_0 = 0$ and we integrate the above expression with respect to r, we obtain:

$$s(t) = s_0 e^{-r(t)\frac{\beta\langle k\rangle}{\mu}}$$

As said, we cannot solve this equation directly, but we can study the behavior in the long term. At $t = \infty$, we have that $i(\infty) = 0$ and thus $s(\infty) = 1 - r(\infty)$:

$$1 - r(\infty) - s_0 e^{-r(\infty)\overbrace{\frac{\beta\langle k\rangle}{\mu}}^{R_0}} = 0$$

This is a transcendental equation which cannot be solved analytically but it gives important hints on the behavior of the disease. Note that $R_0 = \beta\langle k\rangle/\mu$ and this explains why R_0 drives the exponential growth of the disease. Moreover, we note that the initial fraction s_0 of susceptible plays a role in shaping the final fraction of recovered. In particular, if $s_0 \ll 1$ the disease cannot spread.

1.3   Extensions of the SIR model

We want to modify the SIR to include something that we want in our model. We were assuming that the population was totally closed and so it always sum up to 1. This is one thing that we want to remove because it is unrealistic. We will assume that there could be births and deaths.

If we consider the demography, we see that every year there are new child that are infected by disease as Measles and Chickenpox. We expect that usually they die out over weeks.

We are assuming that α is the death rate in all the classes (deaths are not due to the disease).

We are assuming also constant population:

$$\frac{\mathrm{d}s}{\mathrm{d}t} + \frac{\mathrm{d}i}{\mathrm{d}t} + \frac{\mathrm{d}r}{\mathrm{d}t} = 0$$

We change our equation as:

$$\frac{\mathrm{d}s}{\mathrm{d}t} = \alpha - \beta si - \alpha s, \qquad \frac{\mathrm{d}i}{\mathrm{d}t} = \beta si - \mu i - \alpha i, \qquad \frac{\mathrm{d}r}{\mathrm{d}t} = \mu i - \alpha r$$

This complicates the study of the dynamic and for solving it you are assuming that the population is constant, hence at the equilibrium state:

$$\frac{\mathrm{d}s}{\mathrm{d}t} = \frac{\mathrm{d}i}{\mathrm{d}t} = \frac{\mathrm{d}r}{\mathrm{d}t} = 0$$

We obtain the equation:

$$i^*[\beta s^* - (\mu + \alpha)] = 0$$

which is not differential anymore. The two solutions are $i^* = 0$ (**disease free state**) or $s^* = \frac{\alpha + \mu}{\beta} = \frac{1}{R_0}$ which is the **endemic state**.

We can also calculate of the three values of the fraction of infected in the endemic state:

$$(s^*, i^*, r^*) = \left(\frac{1}{R_0}, \frac{\alpha}{\beta}(R_0 - 1), 1 - \frac{1}{R_0} - \frac{\alpha}{\beta}(R_0 - 1) \right)$$

This exists only if $R_0 > 1$.

Part II

Polletto's Lectures

Bibliography