Assignment: Make your own model

Lokesh Venkatesh
DS 3154: Applied Math Methods

August 21, 2024

Describe how you would build and evaluate two models of the same physical phenomenon but have different purposes. Describe in detail:

1. Inputs, Outputs and the Generative Process
2. What parameters it may have (try to be as detailed as possible, but don't describe 1000 parameters!)
3. Evaluation strategy you would employ, based on the structure of the model and the purpose of the model

I will describe two popular compartment models in Epidemiology: the classic SIR model and a Diffusion-based SIR Model. In both models, it is assumed that there are no deaths due to the spreading infection and any fluctuations in the size of the total population due to birth or death are neglected.

1 The Classic SIR Model

1.1 Inputs:

The inputs to this model are time series collected for each of the three compartments: Susceptible $S(t)$, Infected $I(t)$ and Recovered $R(t)$, and this is done for a population as a whole without looking at any variation within. Data for these inputs may be sampled across the whole population at periodic intervals, using which a time series may be constructed.

1.2 Outputs:

Since this is a *predictive* model, the Output variables are almost the same Input variables, just for different time points than in the input time series. The difference is that even though the three input time series are in discrete integers, the resulting output of the system of differential equations will be a continuous real variable, which will have to be converted into discrete integers by using, say the greatest integer function.

1.3 Generative Process:

This is a Dynamical model based on a set of three Linear Ordinary Differential Equations in time, representing the three compartments:

$$\frac{dS}{dt} = -\frac{\beta IS}{N} \tag{1}$$

$$\frac{dI}{dt} = \frac{\beta IS}{N} + \gamma I \tag{2}$$

$$\frac{dR}{dt} = \gamma I \tag{3}$$

1.4 Parameters:

This model has exactly two parameters.

1. β represents the effective transmission rate. This encapsulates factors like the movement of individuals, how virulent the infection is, any other epidemic-control measures imposed on the population by a government, etcetera.

2. γ represents the recovery rate, which also assimilates factors like accessibility to healthcare, innate immunity, etcetera.

Both these parameters may be estimated by fitting the model to the three input-time series, through some statistical tool like a Regression-Analysis.

1.5 Purpose of the model:

This is an **extrapolative** model, to predict how the numbers of the three compartments will evolve. The aim is to get a broad understanding of the trajectory of the epidemic, trying to estimate, for example, the 'peak' of the infection curve, and thus whether healthcare resources will be overloaded at the height of the epidemic, or say the values of the two parameters, whose ration R_0 indicates whether the government should declare an outbreak and even care to impose appropriate counter-measures such as social distancing norms widespread testing.

1.6 (*Some*) ways the model is *wrong*

These are a few ways in which the model is inaccurate when thinking of the real-life phenomenon of epidemics but is still useful to model in such a way.

1. The model returns the Outputs as continuous variables when in fact the inputs are three time series of discrete numbers
2. The model does not take into account any fluctuations that arise due to intrinsic birth and death rates in the population
3. The model treats the population as one entity and does not account for any spatial dynamics, which means that anybody at any location can get infected by anybody in another location.
4. The model assumes that the mortality rate due to the disease is zero.

2 A diffusion-based SIR Model

2.1 Inputs:

In this model, the inputs are time-series of
The inputs to this model are essentially collected as a time series per person, but while also following each person's motion in the population space over time. When the time series is collected per person, what part of the time series corresponds to one of three compartments is noted, and their motion across the landscape while the individual is in that compartment is also analysed.

2.2 Outputs:

When this system is modelled as a bunch of particles floating in space, passing each other and 'infecting' the 'susceptible' molecules, the output is a dynamical map of particles in the population space, and depending on what compartment these particles belong to, they can represent one of three colours. These particles are constantly moving in random motion and go through the stages of infection over time.

2.3 Generative Process:

The partial differential equations describing this system over time are as follows:

$$\frac{\partial S}{\partial t} = D_S \nabla^2 S - \frac{\beta I S}{N} \tag{4}$$

$$\frac{\partial I}{\partial t} = D_I \nabla^2 I + \frac{\beta I S}{N} + \gamma I \tag{5}$$

$$\frac{\partial R}{\partial t} = D_R \nabla^2 R + \gamma I \tag{6}$$

The variables are essentially functions of space and time, i.e. something of the form $f(x, y; t)$.

2.4 Parameters:

1. β represents the effective transmission rate. This encapsulates factors like the movement of individuals, how virulent the infection is, any other epidemic-control measures imposed on the population by a government, etcetera.

2. γ represents the recovery rate, which also assimilates factors like accessibility to healthcare, innate immunity, etcetera.

3. D_S is the diffusion coefficient for susceptible individuals in the population.

4. D_I is the diffusion coefficient for infected individuals, which is intuitively expected to be lesser than that for susceptible or recovered individuals since an infected individual is more likely to be hospitalised or expected to rest.

5. D_R is the diffusion coefficient for recovered individuals, which is expected to be more or less similar to that for susceptible individuals unless there is a system of social distancing in place for susceptible individuals, assuming that the government has declared an epidemic.

While β and γ may be estimated similarly to that of the classic SIR model, to have an 'effective' number for both, the three Diffusion coefficients will have to be estimated by mapping the motion of individuals in the three compartments over time and coming up with some number after some statistical analysis. These parameters can be tuned depending on the effect of quarantine for infected individuals and social-distancing measures for susceptible individuals.

2.5 Purpose of the model:

Like the classic SIR model, this is also an extrapolative model intended to predict how the infection evolves, but also **where** the infection spreads faster when compared to other regions. The aim is to identify 'hot spots', and depending on the quarantine measures imposed by the government, i.e tuning the diffusion coefficient for infected individuals, one can, for example, estimate what the threshold of cases should be for imposing a quarantine zone, or how much should the government limit the motion of susceptible individuals to prevent hot-spots from growing too rapidly. The 'location' component of the model gives it an advantage over the classic SIR model so that governments can be given recommendations to focus on targeted measures rather than imposing nationwide lockdowns, or how widespread and aggressively the government should focus on testing for infected individuals.

2.6 (*Some*) ways the model is *wrong*

1. Just like the classic SIR model, the diffusion-based model also does not take into account any fluctuations that arise due to intrinsic birth and death rates in the population

2. Just like before, this model too assumes that the mortality rate due to the disease is zero.

3. The model assumes that there are no 'hot-spot' centres, i.e. that individuals just randomly move across the landscape and do not spend prolonged periods at specific locations, which would be shopping centres or parks in the real world.

4. This specific model assumes that the diffusion coefficients for each of the three compartments are spatially constant, which is likely untrue.