

Modified SIR Model and Nondimensionalization

Brian Hunt

University of Maryland

AMSC/MATH 420, Spring 2015

Modified SIR Model

- A standard extension to the SIR model adds terms representing births and deaths that are proportional to the overall population.
- If we're modeling an adult subpopulation that is either infected or “at risk”, it may be more appropriate to add a net influx that is independent of the current susceptible/infectious/removed populations:

$$dS/dt = q - pSI$$

$$dI/dt = pSI - rI$$

$$dR/dt = rI$$

- The cumulative number of people infected is $I + R$, and the rate of new infections is pSI .

Units of Variables and Parameters

- The variables S , I and R have units of “population”. A unit of population could be one person, but sometimes other units are used; e.g., census data is often tabulated in units of thousands of people.
- The derivatives dS/dt , dI/dt , dR/dt have units of population/time. Therefore:
 - p has units of $1/(\text{population} \cdot \text{time})$.
 - q has units of population/time.
 - r has units of $1/\text{time}$.

Change of Parameters

- In the SI model, we developed new parameters N , λ , and δ that had units of population, 1/time, and time, respectively.
- This separated a parameter N that controlled the size of the outbreak from parameters λ and δ that controlled the speed and timing of the outbreak.
- Let's do something similar for our modified SIR model, which we'll call the **SIRg model** (SIR with linear population growth).
- If we let $N = S + I + R$, then N changes over time (if $q \neq 0$). To make N a constant, let's set $N = S(0) + I(0) + R(0)$.

Normalized/Nondimensional Parameters

- Let's make N be the only parameter whose units involve population.
- As before, let $\lambda = pN$; then λ has units of 1/time.
- Let's make the other parameters be “dimensionless” – i.e., be independent of the units of time and population.
- Let $\mu = q/(\lambda N)$; then μ is dimensionless.
- Let $\nu = r/\lambda$; then ν is dimensionless. The quantity $1/\nu$ is, for this model, what epidemiologists call the **basic reproduction number** of an outbreak; it is the average number of other people infected by each infectious person early in the epidemic.

Model with New Parameters

- The modified SIR model with the new parameters is

$$dS/dt = \lambda(\mu N - SI/N)$$

$$dI/dt = \lambda(SI/N - \nu I)$$

$$dR/dt = \lambda\nu I$$

- This is a bit messier than before, but it allows us to isolate better the effects of changing a parameter.
- If we multiply N by a constant a , then multiplying S , I , R by the same constant yields a solution to the model with the same values of λ, μ, ν .
- If we change λ , this only changes the rate of the outbreak (not the size).

Interpretation of New Parameters

- N controls the size of the outbreak. However, it does not represent the total number who will eventually be infected.
- λ controls the rate of the outbreak.
- μ and ν are rates of the renewal (new susceptibles) and recovery/removal processes **relative** to the infection process.
- There are two more parameters that determine the relative sizes of $S(0)$, $I(0)$, and $R(0)$, whose sum we have called N . For example, we could use dimensionless parameters $\alpha = I(0)/N$ and $\beta = R(0)/N$. If time 0 is early in the outbreak, it may be reasonable to assume that β is essentially 0 .

Fitting Data with More Parameters

- When trying to optimize (in our setting, minimize the sum of the squares of the residuals) over more than two or three parameters, it becomes difficult to use contour plots productively.
- The “Etch-a-Sketch” method we discussed previously works with any number of parameters, but is not efficient.
- There are general-purpose algorithms, such as MATLAB’s `fminsearch`, that take a function and an initial guess for the location of its minimum, and attempt to home in on a local minimum.
- These methods are not guaranteed to work, and there is no easy way in general to determine whether a function has multiple minima.

Questions to Consider

- What happens to $S(t)$, $I(t)$, $R(t)$, and to the rate of new infections as $t \rightarrow \infty$? The most relevant parameters are μ and ν , which determine the “shape” of the solutions, as opposed to N and λ , which determine their size and speed. The relative sizes of the initial conditions may also be relevant.
- How much do the new processes (renewal, governed by q and μ , and recovery/removal, governed by r and ν) affect the model’s ability to fit a data set? (It may help to consider separately the cases $q = \mu = 0$ and $r = \nu = 0$.)

Fits to San Francisco Monthly Diagnoses

Model	SI	SIR	SIg	SIRg
N	29857	30227	25901	28488
λ	0.031275	0.036046	0.031163	0.0039613
μ	—	—	0.022166	0.40842
ν	—	0.10569	—	-5.5787
α	0.024120	0.019253	0.026024	0.37272
E_y	122980	121850	118940	105350
rms	22.636	22.532	22.261	20.951

- $N, \lambda, \mu, \nu, \alpha$ are model parameters (β was set to 0 for these fits)
- E_y is the sum of the squares of all 240 residuals (1982–2001)
- $\text{rms} = \sqrt{E_y/252}$ is the root-mean-square error

Observations

- For the San Francisco data set, the SIR and SIg models fit slightly better than the SI model.
- The SIRg model shows a more significant reduction in the error, but the parameter values seem questionable; the value of λ is an order of magnitude smaller than for the other models, μ and α are an order of magnitude larger, and ν is negative.
- This is a mild form of overfitting; the model has few enough parameters that it will interpolate the data reasonably, but since the parameter values are unrealistic, it may not extrapolate well.