11 Dynamic models and their simulation by Euler's method

Edward Ionides

3/25/2018

Contents

Produced with R version 3.4.3 and **pomp** version 1.16.

Objectives

This chapter develops a general class of dynamic models with particular relevance for biological systems. We have the following goals:

1. Dynamic systems can often be represented in terms of *flows* between *compartments*. We will develop the concept of a *compartment model* for which we specify *rates* for the flows between compartments.

2. We develop deterministic and stochastic interpretations of a compartment model.

3. We introduce Euler's method to simulate from dynamic models, and we apply it to both deterministic and stochastic compartment models.

Compartment models

- A **compartment model** is a model associated with a **flow diagram** specifying how objects move between categories, called **compartments**.

- It is often useful to represent systems by **flow diagrams**. We will need equations to specify formally what the flow diagram means.

- One major applications of compartment models is pharmacokinetics, the study of how pharmacological drugs enter the body, move between organs, metabolize, and leave. The compartments may be the organs; the flow is movement of the drug and its metabolites between organs.

- Another major application of compartment models is epidemiology. Compartments are groups of people; the flow is the movement of an infectious disease through the population.

Compartment models in epidemiology: the SIR model and its generalizations

We will develop deterministic and stochastic representations of a susceptible-infected-recovered (SIR) system, a fundamental class of models for disease transmission dynamics. We will do this using notation which generalizes to more complex systems (Bretó et al. 2009).

$S =$ susceptible
$I =$ infected and infectious
$R =$ recovered and/or removed

- We suppose that each arrow has an associated rate, so here there is a rate $\mu_{SI}(t)$ at which individuals in S transition to I, and μ_{IR} at which individuals in I transition to R.

- To account for demography (births/deaths/immigration/emmigration) we allow the possibility of a source and sink compartment, which is not usually represented on the flow diagram. We write $\mu_{\bullet S}$ for a rate of births into S, and denote mortality rates by $\mu_{S\bullet}$, $\mu_{I\bullet}$, $\mu_{R\bullet}$.

- The rates may be either constant or varying. In particular, for a simple SIR model, the recovery rate μ_{IR} is a constant but the infection rate has the time-varying form

$$\mu_{SI}(t) = \beta\, I(t),$$

  with β being the *contact rate*. Now, for the simplest SIR model, ignoring demography, we set

$$\mu_{\bullet S} = \mu_{S\bullet} = \mu_{I\bullet} = \mu_{R\bullet} = 0.$$

- To develop a systemtic notation, it turns out to be convenient to keep track of the flows between compartments as well as the number of individuals in each compartment. Let

$$N_{SI}(t)$$

  count the number of individuals who have transitioned from S to I by time t. We say that $N_{SI}(t)$ is a *counting process*. A similarly constructed process

$$N_{IR}(t)$$

  counts individuals transitioning from I to R. To include demography, we could keep track of birth and death events by the counting processes $N_{\bullet S}(t)$, $N_{S\bullet}(t)$, $N_{I\bullet}(t)$, $N_{R\bullet}(t)$.

- For discrete population compartment models, the flow counting processes are non-decreasing and integer valued.

- For continuous population compartment models, the flow counting processes are non-decreasing and real valued.

- The numbers of people in each compartment can be computed via these counting processes. Ignoring demography, we have:

$$\begin{array}{rcl} S(t) & = & S(0) - N_{SI}(t) \\ I(t) & = & I(0) + N_{SI}(t) - N_{IR}(t) \\ R(t) & = & R(0) + N_{IR}(t) \end{array}$$

- These equations represent something like *conservation of mass*, or *what goes in must come out*.

2

The *ordinary differential equation* (ODE) interpretation of the SIR model

Together with initial conditions specifying $S(0)$, $I(0)$ and $R(0)$, we just need to write down ODEs for the flow counting processes. These are,

$$dN_{SI}/dt = \mu_{SI}(t)\, S(t),$$

$$dN_{IR}/dt = \mu_{IR}\, I(t).$$

The simple continuous-time Markov chain interpretation of the SIR model

- Continuous-time Markov chains are the basic tool for building discrete population epidemic models.

- Recall that a *Markov chain* is a discrete-valued stochastic process with the *Markov property*: the future evolution of the process depends only on the current state.

- Surprisingly many models have this Markov property. If all important variables are included in the state of the system, then the Markov property appears automatically.

- The Markov property lets us specify a model by the transition probabilities on small intervals (together with the initial conditions). For the SIR model, we have

$$
\begin{array}{rcl}
\mathbb{P}\big[N_{SI}(t+\delta){=}N_{SI}(t)+1\big] &=& \mu_{SI}(t)\, S(t)\delta + o(\delta) \\
\mathbb{P}\big[N_{SI}(t+\delta){=}N_{SI}(t)\big] &=& 1 - \mu_{SI}(t)\, S(t)\delta + o(\delta) \\
\mathbb{P}\big[N_{IR}(t+\delta){=}N_{IR}(t)+1\big] &=& \mu_{IR}\, I(t)\delta + o(\delta) \\
\mathbb{P}\big[N_{IR}(t+\delta){=}N_{IR}(t)\big] &=& 1 - \mu_{IR}(t)\, I(t)\delta + o(\delta)
\end{array}
$$

- Here, we are using little o notation. We write

$$h(\delta) = o(\delta)$$

  to mean

$$\lim_{\delta \to 0} \frac{h(\delta)}{\delta} = 0.$$

Question: What is the link between little o notation and the derivative?

- Explain why

$$f(x + \delta) = f(x) + \delta g(x) + o(\delta)$$

 is the same statement as

$$\frac{df}{dx} = g(x).$$

- What considerations might help you choose which of these notations to use?

Exercises

From Markov chain to ODE

Find the expected value of $N_{SI}(t + \delta) - N_{SI}(t)$ and $N_{IR}(t + \delta) - N_{IR}(t)$ given the current state, $S(t)$, $I(t)$ and $R(t)$. Take the limit as $\delta \to 0$ and show that this gives the ODE model.

- A *simple* counting process is one which cannot count more than one event at a time (Wikipedia: Point_process). Thus, in a technical sense, the SIR Markov chain model we have written is simple. One may want to model the extra randomness resulting from multiple simultaneous events: someone sneezing in a bus; large gatherings at football matches; etc. This extra randomness may even be critical to match the variability in data.

- Later in the course, we may see situations where this extra randomness plays an important role. Setting up the model using counting processes, as we have done here, turns out to be useful for this.
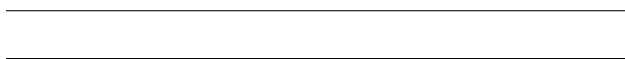
Euler's method for ordinary differential equations (ODEs)

- Euler took the following approach to numeric solution of an ODE:

  - He wanted investigate an ODE
  $$dx/dt = h(x)$$
  with an initial condition $x(0)$. He supposed this ODE has some true solution $x(t)$ which could not be worked out analytically. He therefore wished to approximate $x(t)$ numerically.
  - He initialized the numerical solution at the known starting value,
  $$\tilde{x}(0) = x(0).$$
  Then, for $k = 1, 2, \ldots$, he supposed that the gradient dx/dt is approximately constant over the small time interval $k\delta \leq t \leq (k+1)\delta$. Therefore, he defined
  $$\tilde{x}\big((k+1)\delta\big) = \tilde{x}(k\delta) + \delta\, h\big(\tilde{x}(k\delta)\big).$$

- This only defines $\tilde{x}(t)$ when t is a multiple of δ, but let's suppose $\tilde{x}(t)$ is constant between these discrete times.

- We now have a numerical scheme, stepping forwards in time increments of size δ, that can be readily evaluated by computer (or by hand, if you are Euler).

- Mathematical analysis of Euler's method says that, as long as the function $h(x)$ is not too exotic, then $x(t)$ is well approximated by $\tilde{x}(t)$ when the discretization time-step, δ, is sufficiently small.

- Euler's method is not the only numerical scheme to solve ODEs. More advanced schemes have better convergence properties, meaning that the numerical approximation is closer to $x(t)$. However, there are 3 reasons we choose to lean heavily on Euler's method:

1. Euler's method is the simplest (following the KISS principle).

2. Euler's method extends naturally to stochastic models, both continuous-time Markov chains models and stochastic differential equation (SDE) models.

3. Close approximation of the numerical solutions to a continuous-time model is less important than it may at first appear, a topic worth further discussion. . .

Some comments on using continuous-time models and discretized approximations

- In some physical and engineering situations, a system follows an ODE model closely. For example, Newton's laws provide a very good approximation to the motions of celestial bodies.

- In many biological situations, ODE models only become close mathematical approximations to reality at reasonably large scale. On small temporal scales, models cannot usually capture the full scope of biological variation and biological complexity.

- If we are going to expect substantial error in using $x(t)$ to model a biological system, maybe the numerical solution $\tilde{x}(t)$ represents the system being modeled as well as $x(t)$ does.

- If our model fitting, model investigation, and final conclusions are all based on our numerical solution $\tilde{x}(t)$ (i.e., we are sticking entirely to simulation-based methods) then we are most immediately concerned with how well $\tilde{x}(t)$ describes the system of interest. $\tilde{x}(t)$ becomes more important than the original model, $x(t)$.

- When following this perspective, it is important that the scientists fully describe the numerical model $\tilde{x}(t)$. The main advantage of the continuous-time model $x(t)$ is then that it gives a succinct way to describe how $\tilde{x}(t)$ was constructed.

- All numerical methods are, ultimately, discretizations. Epidemiologically, setting δ to be a day, or an hour, can be quite different from setting δ to be two weeks or a month. For continuous-time modeling, we still require that δ is small compared to the timescale of the process being modeled, and the choice of δ does not play an explicit role in the interpretation of the model.

- Putting more emphasis on the scientific role of the numerical solution itself reminds you that the numerical solution has to do more than approximate a target model in some asymptotic sense: the numerical solution should be a sensible model in its own right.

Euler's method for a discrete SIR model

- Recall the simple continuous-time Markov chain interpretation of the SIR model without demography:

$$
\begin{array}{rcl}
\mathbb{P}\big[N_{SI}(t+\delta)\!=\!N_{SI}(t)+1\big] & = & \mu_{SI}(t)\,S(t)\delta + o(\delta), \\
\mathbb{P}\big[N_{IR}(t+\delta)\!=\!N_{IR}(t)+1\big] & = & \mu_{IR}\,I(t)\delta + o(\delta).
\end{array}
$$

- We look for a numerical solution with state variables $\tilde{S}(k\delta)$, $\tilde{I}(k\delta)$, $\tilde{R}(k\delta)$.

- The counting processes for the flows between compartments are $\tilde{N}_{SI}(t)$ and $\tilde{N}_{IR}(t)$. The counting processes are related to the numbers of individuals in the compartments by the same flow equations we had before:

$$
\begin{array}{rcl}
\tilde{S}(k\delta) & = & S(0) - \tilde{N}_{SI}(k\delta) \\
\tilde{I}(k\delta) & = & I(0) + \tilde{N}_{SI}(k\delta) - \tilde{N}_{IR}(k\delta) \\
\tilde{R}(k\delta) & = & R(0) + \tilde{N}_{IR}(k\delta)
\end{array}
$$

- Let's focus on a numerical solution to $N_{SI}(t)$, since the same methods can also be applied to $N_{IR}(t)$.

- Here are three possibilities, defining a process at times $t = k\delta$ for all integer k and some fixed $\delta > 0$.

1. A Poisson approximation.

$$
\tilde{N}_{SI}(t+\delta) = \tilde{N}_{SI}(t) + \text{Poisson}\big[\mu_{SI}\big(\tilde{I}(t)\big)\,\tilde{S}(t)\,\delta\big],
$$

where Poisson(μ) is a Poisson random variable with mean μ and

$$\mu_{SI}\big(\tilde{I}(t)\big) = \beta\,\tilde{I}(t).$$

2. A binomial approximation with transition probabilities approximated by rate times time.

$$\tilde{N}_{SI}(t+\delta) = \tilde{N}_{SI}(t) + \text{Binomial}\big[\tilde{S}(t), \mu_{SI}\big(\tilde{I}(t)\big)\,\delta\big],$$

where Binomial(n, p) is a binomial random variable with mean np and variance $np(1-p)$.

3. A binomial approximation with exponential transition probabilities.

$$\tilde{N}_{SI}(t+\delta) = \tilde{N}_{SI}(t) + \text{Binomial}\big[\tilde{S}(t), 1 - \exp\big\{-\mu_{SI}\big(\tilde{I}(t)\big)\delta\big\}\big].$$

- What are the advantages of these different schemes? Conceptually, it is simplest to think of (1.) or (2.). Numerically, it is usually preferable to implement (3.).

Exercises

Theoretical exercise: Compartment models via stochastic differential equations (SDEs)

The Euler method extends naturally to stochastic differential equations. A natural way to add stochastic variation to an ODE $dx/dt = h(x)$ is

$$dX/dt = h(X) + \sigma\,dB/dt$$

where $\{B(t)\}$ is Brownian motion and so dB/dt is Brownian noise. Then, an Euler approximation $\tilde{X}(t)$ is generated by

$$\tilde{X}\big((k+1)\delta\big) = \tilde{X}(k\delta) + \delta\,h\big(\tilde{X}(k\delta)\big) + \sigma\sqrt{\delta}\,Z_k$$

where $Z_1, Z_2, \ldots$ is a sequence of independent standard normal random variables, i.e., $Z_k \sim N[0,1]$. Although SDEs are often considered an advanced topic in probability, the Euler approximation doesn't demand much more than familiarity with the normal distribution.

Write down a stochastic Euler method for an SDE representation of the SIR model. Consider some difficulties that might arise with non-negativity constraints, and propose some practical way one might deal with that issue.

- A useful method to deal with positivity constraints is to use Gamma noise rather than Brownian noise (Bhadra et al. 2011, Laneri et al. (2010)). SDEs driven by Gamma noise can be investigated by Euler solutions simply by replacing the Gaussian noise by an appropriate Gamma distribution.

Conceptual exercise: Euler's method vs Gillspie's algorithm

- A widely used, exact simulation method for continuous time Markov chains is Gillspie's algorithm. We do not put much emphasis on Gillespie's algorithm here. Why? When would you prefer an implementation of Gillespie's algorithm to an Euler solution?

- Numerically, Gillespie's algorithm is often approximated using so-called tau-leaping methods. These are closely related to Euler's approach. Is it reasonable to call a suitable Euler approach a tau-leaping method?

Compartmental models in pomp.

As an example that we can probe in some depth, let's look at an isolated outbreak of influenza that occurred in a boarding school for boys in England (Anonymous 1978). Let's examine the data:

```
bsflu <- read.table("bsflu_data.txt")
head(bsflu)
```

```
##               B  C day
## 1978-01-22    1  0   1
## 1978-01-23    6  0   2
## 1978-01-24   26  0   3
## 1978-01-25   73  1   4
## 1978-01-26  222  8   5
## 1978-01-27  293 16   6
```

The variable B refers to boys confined to bed and C to boys in convalescence. Let's restrict our attention for the moment to the B variable.

```
bsflu <- subset(bsflu,select=c(day,B))
ggplot(data=bsflu,aes(x=day,y=B))+geom_line()+geom_point()
```
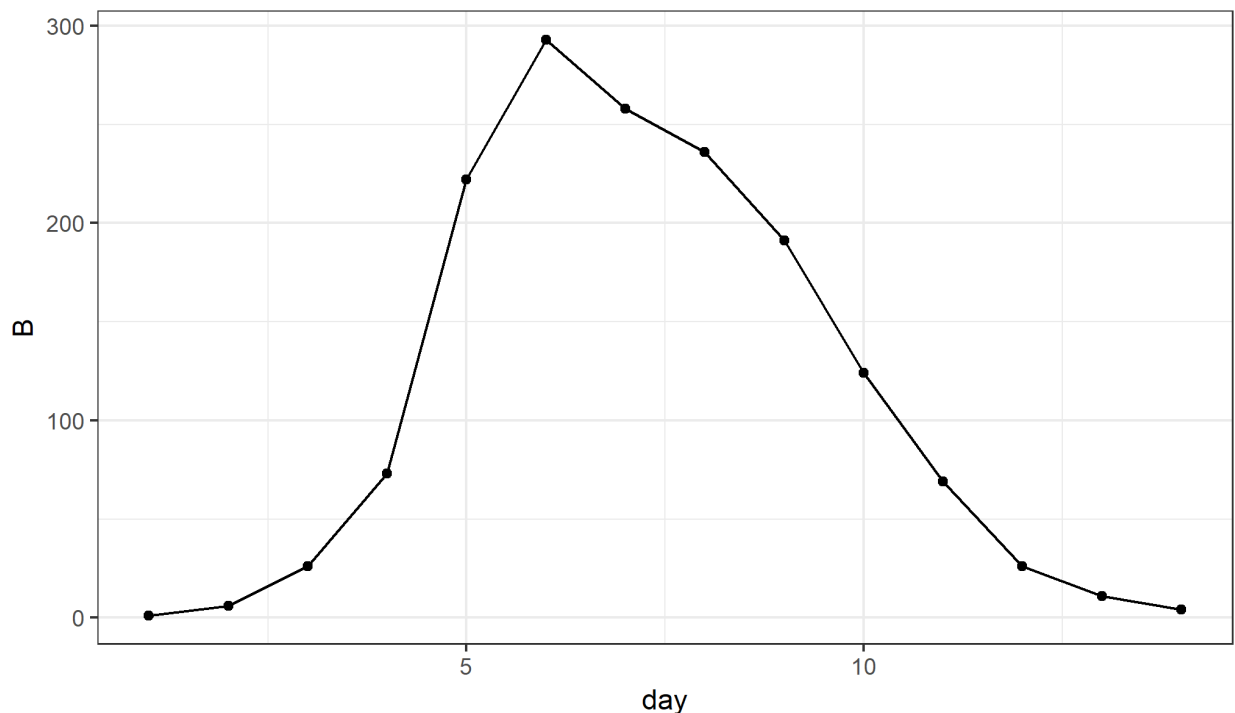

Let's assume that B indicates the number of boys confined to bed the preceding day and that the disease follows the simple SIR model. Our tasks will be, first, to estimate the parameters of the SIR and, second, to decide whether or not the SIR model is an adequate description of these data.

Below is a diagram of the SIR model. The host population is divided into three classes according to their infection status: S, susceptible hosts; I, infected (and infectious) hosts; R, recovered and immune hosts. The

rate at which individuals move from S to I is the force of infection, $\lambda = \beta I/N$, while that at which individuals move into the R class is γ.

Let's look at how we can view the SIR as a POMP model. The unobserved state variables, in this case, are the numbers of individuals, S, I, R in the S, I, and R compartments, respectively. It's reasonable in this case to view the population size $N = S + I + R$, as fixed. The numbers that actually move from one compartment to another over any particular time interval are modeled as stochastic processes. In this case, we'll assume that the stochasticity is purely demographic, i.e., that each individual in a compartment at any given time faces the same risk of exiting the compartment.

To implement the model in **pomp**, the first thing we need is a stochastic simulator for the unobserved state process. We've seen that there are several ways of approximating the process just described for numerical purposes. An attractive option here is to model the number moving from one compartment to the next over a very short time interval as a binomial random variable. In particular, we model the number, ΔN_{SI}, moving from S to I over interval Δt as

$$\Delta N_{SI} \sim \text{Binomial}\left(S, 1 - e^{-\lambda \Delta t}\right),$$

and the number moving from I to R as

$$\Delta N_{IR} \sim \text{Binomial}\left(I, 1 - e^{-\gamma \Delta t}\right).$$

A `Csnippet` that encodes such a simulator is as follows:

```
sir_step <- Csnippet("
  double dN_SI = rbinom(S,1-exp(-Beta*I/N*dt));
  double dN_IR = rbinom(I,1-exp(-gamma*dt));
  S -= dN_SI;
  I += dN_SI - dN_IR;
  R += dN_IR;
")
```

At day zero, we'll assume that $I = 1$ and $R = 0$, but we don't know how big the school is, so we treat N as a parameter to be estimated and let $S(0) = N - 1$. Thus an initializer `Csnippet` is

```
sir_init <- Csnippet("
  S = N-1;
  I = 1;
  R = 0;
")
```

We fold these `Csnippet`s, with the data, into a `pomp` object thus:

```
pomp(bsflu,time="day",t0=0,rprocess=euler.sim(sir_step,delta.t=1/6),
     initializer=sir_init,paramnames=c("N","Beta","gamma"),
     statenames=c("S","I","R")) -> sir
```

Now let's assume that the case reports, B, result from a process by which new infections result in confinement with probability ρ, which we can think of as the probability that an infection is severe enough to be noticed by the school authorities. Since confined cases have, presumably, a much lower transmission rate, let's treat B as being a count of the number of boys who have moved from I to R over the course of the past day. We need a variable to track this. Let's modify our `Csnippet` above, adding a variable H to track the incidence. We'll then replace the `rprocess` with the new one.

```
sir_step <- Csnippet("
  double dN_SI = rbinom(S,1-exp(-Beta*I/N*dt));
  double dN_IR = rbinom(I,1-exp(-gamma*dt));
  S -= dN_SI;
  I += dN_SI - dN_IR;
```

```
  R += dN_IR;
  H += dN_IR;
")

sir_init <- Csnippet("
  S = N-1;
  I = 1;
  R = 0;
  H = 0;
")

pomp(sir,rprocess=euler.sim(sir_step,delta.t=1/6),initializer=sir_init,
     paramnames=c("Beta","gamma","N"),statenames=c("S","I","R","H")) -> sir
```

Now, we'll model the data, B, as a binomial process,

$$B_t \sim \text{Binomial}\left(H(t) - H(t-1), \rho\right).$$

But we have a problem, since at time t, the variable H we've defined will contain $H(t)$, not $H(t) - H(t-1)$. We can overcome this by telling `pomp` that we want H to be set to zero immediately following each observation. We do this by setting the `zeronames` argument to `pomp`:

```
pomp(sir,zeronames="H") -> sir
```

Now, to include the observations in the model, we must write both a `dmeasure` and an `rmeasure` component:

```
dmeas <- Csnippet("lik = dbinom(B,H,rho,give_log);")
rmeas <- Csnippet("B = rbinom(H,rho);")
```

and put these into our `pomp` object:

```
sir <- pomp(sir,rmeasure=rmeas,dmeasure=dmeas,statenames="H",paramnames="rho")
```

Let's perform some simulations to verify that things are working. To do so, we'll need some parameters. A little thought will get us some ballpark estimates. In the data, it looks like there were a total of 1540 infections, so the population size, N, must be somewhat in excess of this number. In fact, we can use the final-size equation
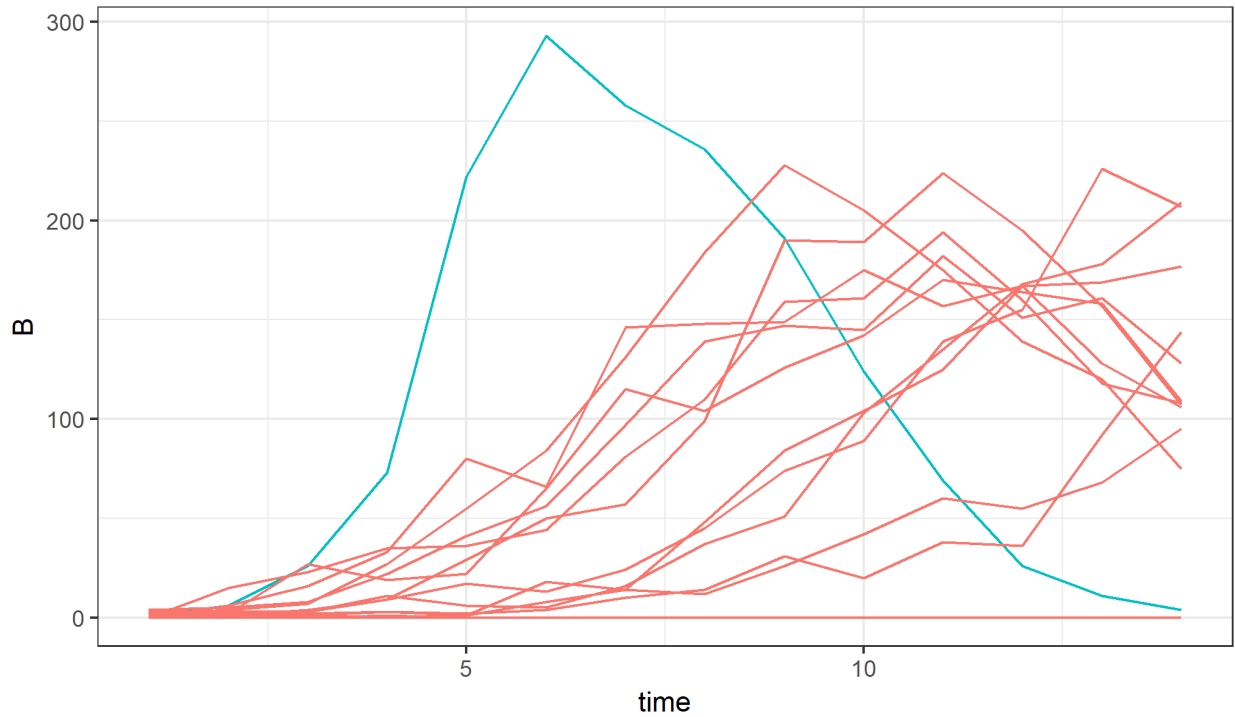
$$R_0 = -\frac{\log\left(1 - f\right)}{f},$$

where $f = R(\infty)/N$ is the final size of the epidemic, together with the idea that R_0 must be, say, around 1.5, to estimate that $f \approx 0.6$, whence $N \approx 2600$. If the infectious period is roughly 1 da, then $1/\gamma \approx 1$ da and $\beta = \gamma R_0 \approx 1.5 \text{ da}^{-1}$.

```
sims <- simulate(sir,params=c(Beta=1.5,gamma=1,rho=0.9,N=2600),
                 nsim=20,as=TRUE,include=TRUE)

ggplot(sims,mapping=aes(x=time,y=B,group=sim,color=sim=="data"))+
  geom_line()+guides(color=FALSE)
```

Exercises

Explore the SIR model

Fiddle with the parameters to see if you can't find parameters for which the data are a more plausible realization.

The SEIR model

Below is a diagram of the so-called SEIR model. This differs from the SIR model in that infected individuals must pass a period of latency before becoming infectious.

Modify the codes above to construct a `pomp` object containing the flu data and an SEIR model. Perform simulations as above and adjust parameters to get a sense of whether improvement is possible by including a latent period.


Rethinking the boarding-school flu data

In the preceding, we've been assuming that B_t represents the number of boys *sent* to bed on day t. Actually, this isn't correct at all. As described in the report (Anonymous 1978), B_t represents the total number of boys *in* bed on day t. Since boys were potentially confined for more than one day, the data count each infection multiple times. On the other hand, we have information about the total number of boys at risk and the total number who were infected. In fact, we know from Anonymous (1978) that $N = 763$ boys were at risk and 512 boys in total spent between 3 and 7 days away from class (either in bed or convalescent). Moreover, there is information in the data on the number of boys, C_t, convalescent at day t. Since 1540 boy-da/512 boy $\approx$ 3 da, we know that the average duration spent in bed was 3 da and, since $\sum_t C_t = 0$, we can infer that the average time spent convalescing was 0 boy-da/512 boy $\approx$ 0 da.

```
require(reshape2)
ggplot(data=melt(bsflu,id="day"),mapping=aes(x=day,y=value,color=variable))+
  geom_line()+geom_point()
```
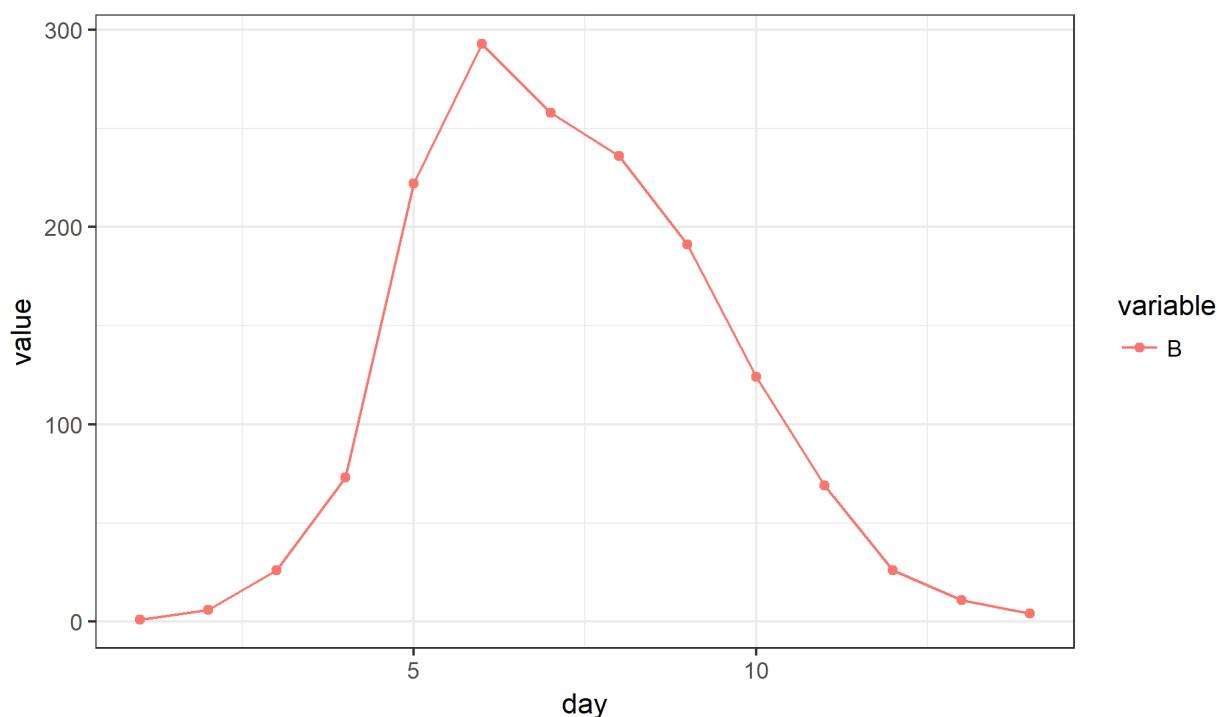

Formulate a model with both confinement and convalescent stages. Implement it in **pomp** using a compartmental model like that diagrammed below.

You will have to give some thought to just how to model the relationship between the data (B and C) and the state variables. How many parameters can reasonably be fixed? How many must be estimated? Obtain some ballpark estimates of the parameters and simulate to see if you can plausibly explain the data as a realization of this model.

Acknowledgment

References

Anonymous. 1978. Influenza in a boarding school. British Medical Journal 1:587.

Bhadra, A., E. L. Ionides, K. Laneri, M. Pascual, M. Bouma, and R. Dhiman. 2011. Malaria in Northwest India: Data analysis via partially observed stochastic differential equation models driven by Lévy noise. Journal of the Americal Statistical Association 106:440–451.

Bretó, C., D. He, E. L. Ionides, and A. A. King. 2009. Time series analysis via mechanistic models. Annals of Applied Statistics 3:319–348.

Laneri, K., A. Bhadra, E. L. Ionides, M. Bouma, R. Yadav, R. Dhiman, and M. Pascual. 2010. Forcing versus feedback: Epidemic malaria and monsoon rains in NW India. PLoS Comput. Biol. 6:e1000898.