Stochastic dynamic models for low count observations

(and forecasting from them)

Dr. Jarad Niemi

Iowa State University

April 19, 2016

Overview

1. Poisson-binomial state-space model
   - Simulations
2. Inference and forecasting
3. Forecasting with noisy observations
4. Forecasting with inference on parameters

Poisson-binomial state-space model

Let X_{tk} be the count of the number of individuals in state k and time t. We model state transitions using Poisson distributions, i.e.

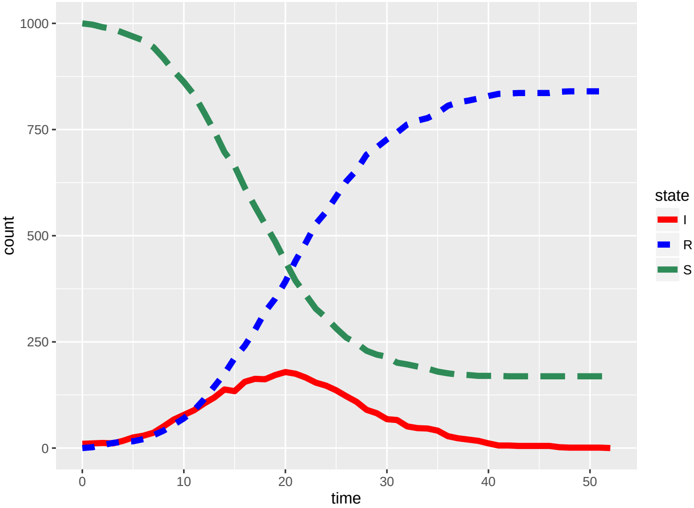
$$\begin{aligned} \Delta X_{tk} &\stackrel{ind}{\sim} Po(\lambda_k f_k(X_{t-1})), \\ X_{tm} &= X_{t-1,m} + \textstyle\sum_{k=1}^{K} v_{mk} \Delta X_{tk}, \quad m = 1, \ldots, M \end{aligned}$$

where $f_k(\cdot)$ are known functions and v_{mk} is a known *stoichiometry*.

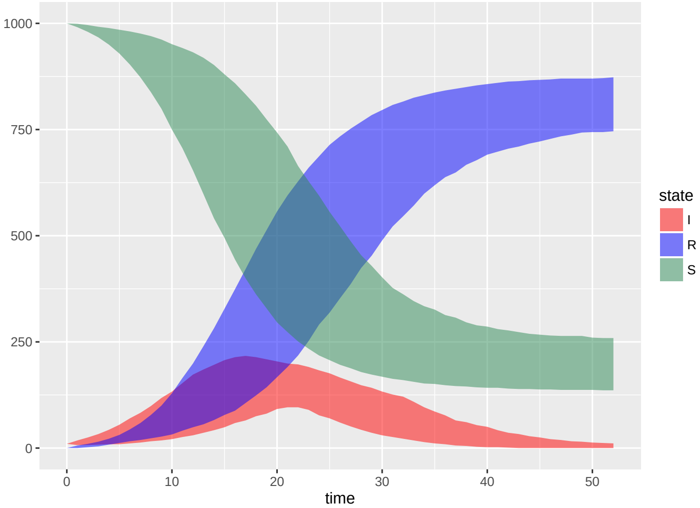
We assume the transitions ΔX_{tk} are partially observed through a binomial distribution, i.e.

$$Y_{tk} \stackrel{ind}{\sim} Bin(\Delta X_{tk}, \theta_k).$$

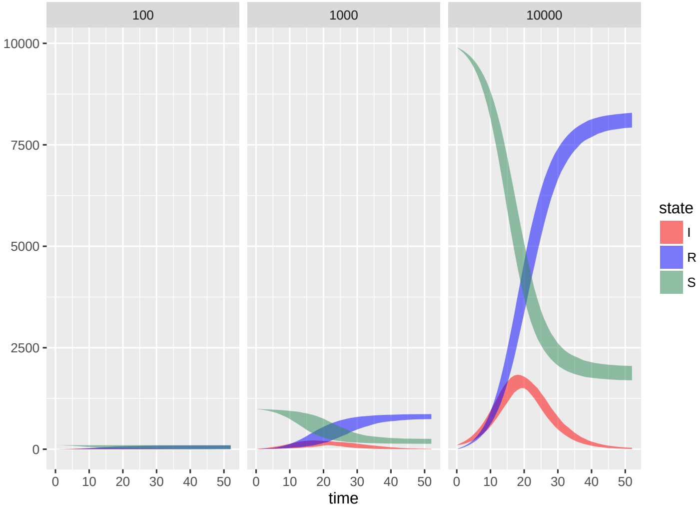
SIR modeling simulations



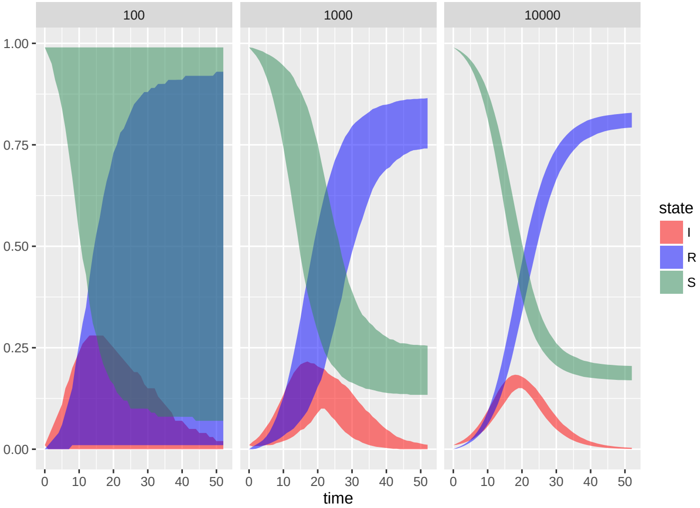
SIR modeling simulations



Variability as a function of population size



Forecasts for proportion of population

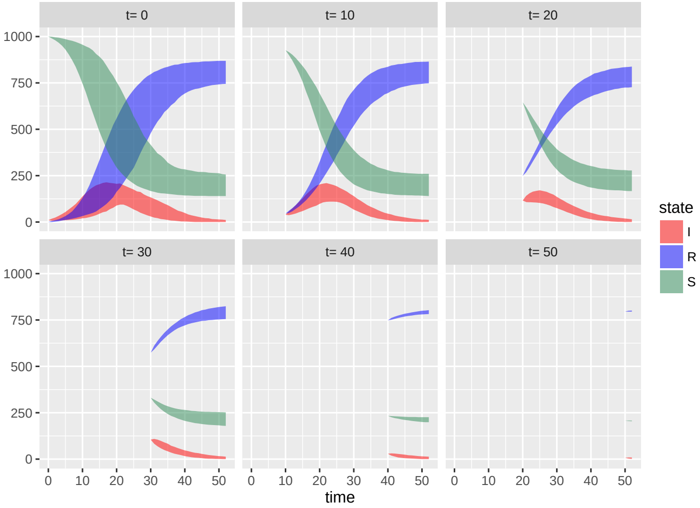


Forecasting with perfect information

Suppose you know transition rates λ, observation probabilities θ, and the states $X_{0:t}$ and your only goal is to forecast the future states $X_{t+1:T}$, i.e.

$$p(X_{t+1:T}|\theta, \lambda, X_{0:t}) = p(X_{t+1:T}|\theta, \lambda, X_t)$$

this distribution is estimated via Monte Carlo simulation.


t= 0
t= 10
t= 20
t= 30
t= 40
t= 50
1000
750
500
250
0
0 10 20 30 40 50
time
state
I
R
S

Delay in data analysis

Suppose you have a one or two week delay in collecting, processing, and analyzing so that when trying to forecast future states you are using "old" data, i.e.
Real-time:

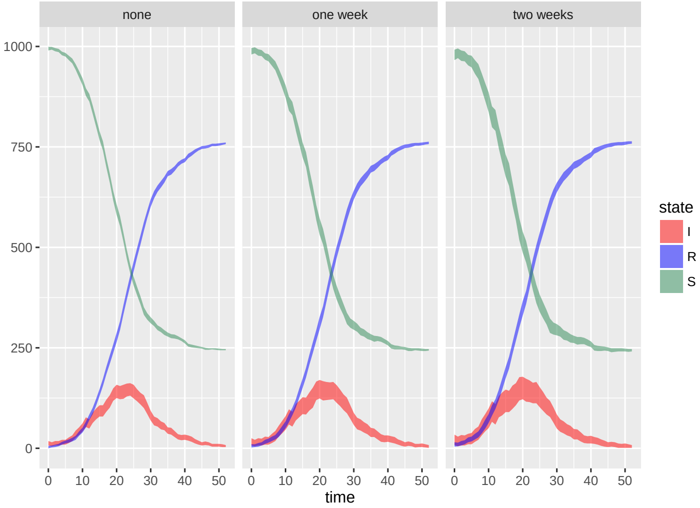
$$p(X_{t+1:T}|\theta, \lambda, X_t)$$

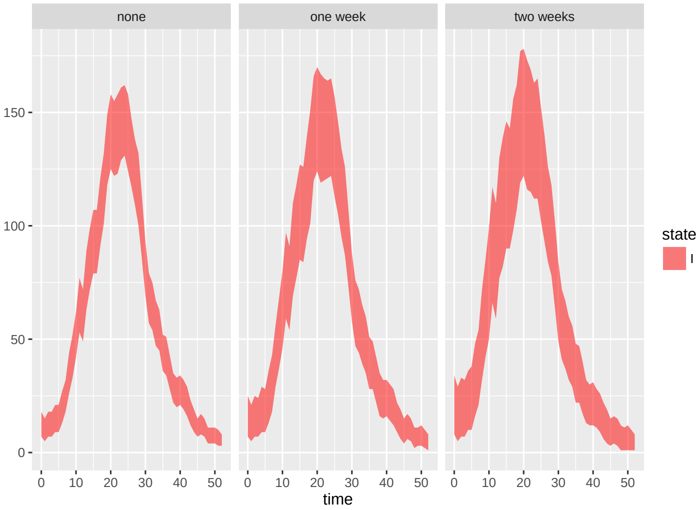
One-week delay:

$$p(X_{t+1:T}|\theta, \lambda, X_{t-1})$$

Two-week delay:

$$p(X_{t+1:T}|\theta, \lambda, X_{t-2})$$


none
one week
two weeks
1000
750
500
250
0
0
10
20
30
40
50
time
state
I
R
S


none
one week
two weeks
150
100
50
0
0
10
20
30
40
50
time
state
I

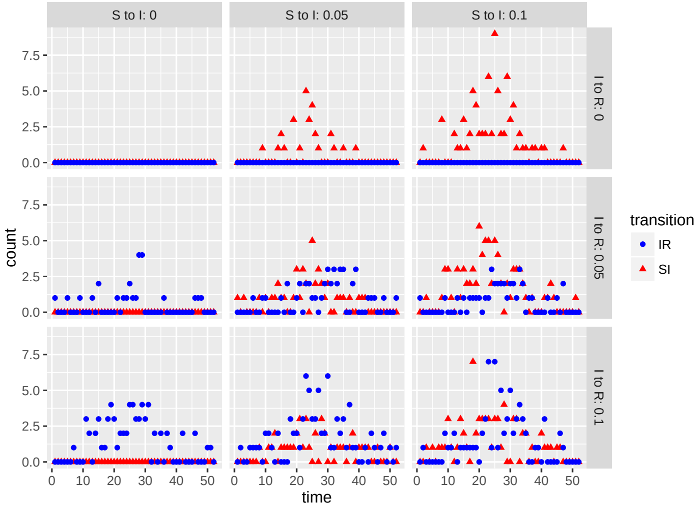
Forecasting with noisy observations

Suppose, we know the transition rates (λ) and the observation probabilities (θ), but we only observe a noisy version of the state transitions, i.e. .

$$Y_{tk} \stackrel{ind}{\sim} Bin(\Delta X_{tk}, \theta_k).$$

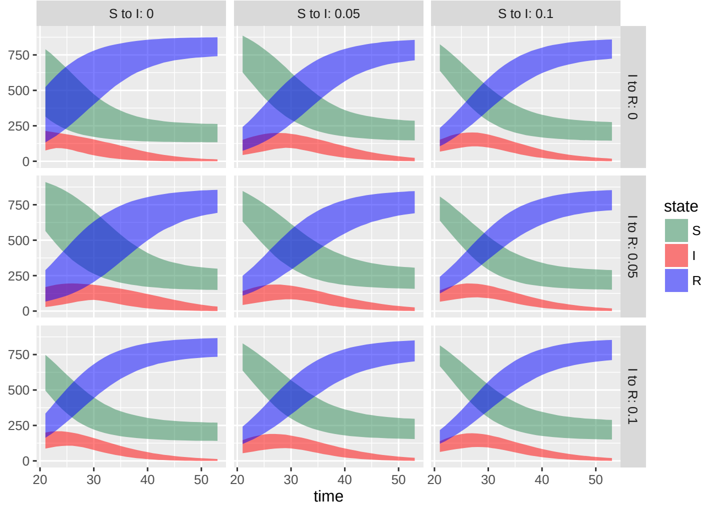
Now the forecast distribution we need is

$$p(X_{t+1:T}|\lambda, \theta, y_{0:t}) = \int p(X_{t+1:T}, \lambda, \theta|X_t)p(X_t|\lambda, \theta, y_{0:t})dX_t.$$



S to I: 0
S to I: 0.05
S to I: 0.1
I to R: 0
I to R: 0.05
I to R: 0.1
count
7.5
5.0
2.5
0.0
0
10
20
30
40
50
time
transition
IR
SI

Noisily observed state



Forecasting with inference on parameters

In reality, we don't know the transition rates (λ) and the observation probabilities (θ), and we only observe a noisy version of the state transitions.
Now the forecast distribution we need is

$$p(X_{t+1:T}|y_{0:t}) = \int\int\int p(X_{t+1:T}, \lambda, \theta|X_t)p(X_t, \lambda, \theta|y_{0:t})d\lambda d\theta dX_t.$$

Prior distributions

In order to calculate (or approximate) the integral

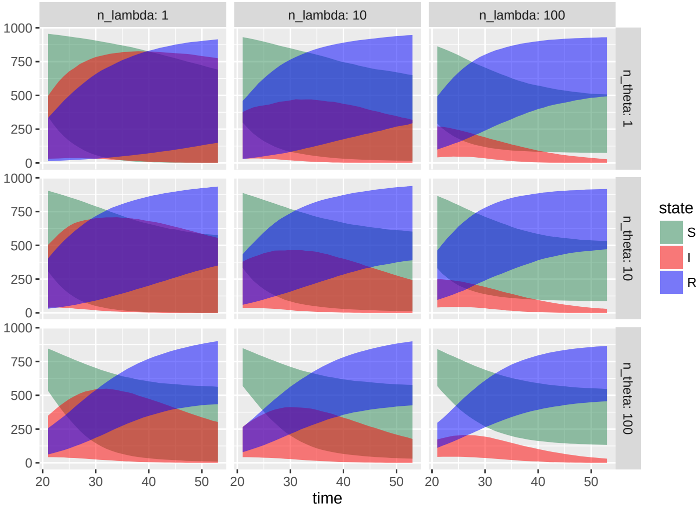
$$\int\int\int p(X_{t+1:T}, \lambda, \theta | X_t) p(X_t, \lambda, \theta | y_{0:t}) d\lambda d\theta dX_t$$

we need to assign priors to λ, θ, and X_0. Suppose, we assume

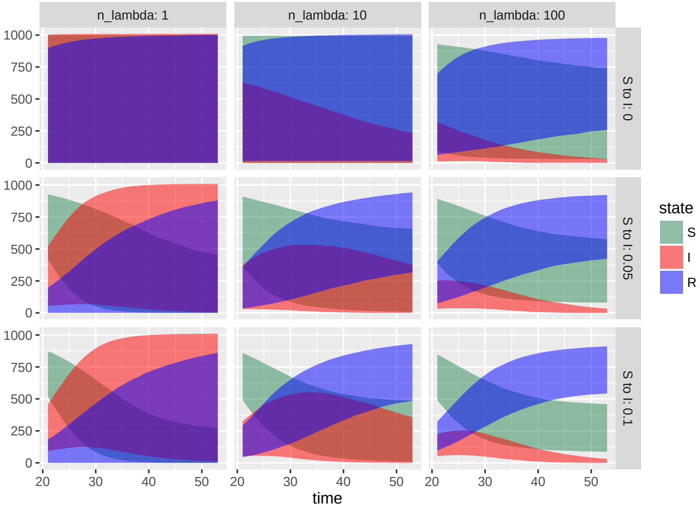
$$\begin{array}{ll} \theta_k & \stackrel{ind}{\sim} Be(n_\theta p_k, n_\theta[1-p_k]) \\ \lambda_k & \stackrel{ind}{\sim} Ga(n_\lambda c_k, n_\lambda) \\ X_0 & \sim Mult(N; z_1, \ldots, z_S) \end{array}$$

We can control how informative the priors are with n_θ and n_λ.

Informative priors



Balance priors and data



We need information

Information can come from

- Data
- Priors

We can quantitatively assess the impact of better information, i.e.

- Increasing prior information, e.g. $\lambda \sim Ga(n_\lambda c_k, n_\lambda)$ by increasing n_λ.
- Increasing surveillance, e.g. $Y_t \sim Bin(S \to I, \theta)$ by increasing θ.
- Increasing timeliness, e.g. $p(X_{t+1:T}|y_{1:t-L})$ by decreasing L.

Then, we can discuss how to assign resources depending on the costs associated with each impact above.