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
2. Inference and forecasting
3. Forecasting with noisy observations
4. Forecasting with inference on parameters

S→I→R stochastic compartment model

Focus on Susceptible (S) - Infectious (I) - Recovered (R) model with Poisson transitions, i.e.

$$\overrightarrow{SI} \sim Po(\lambda_{\mathsf{S}\to\mathsf{I}}\mathsf{SI}/\mathsf{N}) \qquad \overrightarrow{IR} \sim Po(\lambda_{\mathsf{I}\to\mathsf{R}}\mathsf{I})$$

S→I→R stochastic compartment model

Focus on Susceptible (S) - Infectious (I) - Recovered (R) model with Poisson transitions, i.e.

$$\overrightarrow{SI} \sim Po(\lambda_{\mathsf{S}\to\mathsf{I}}\mathsf{SI}/\mathsf{N}) \qquad \overrightarrow{IR} \sim Po(\lambda_{\mathsf{I}\to\mathsf{R}}\mathsf{I})$$

updating the states, i.e.

$$S_{t+1} = S_t - \overrightarrow{SI}, \quad I_{t+1} = I_t + \overrightarrow{SI} - \overrightarrow{IR}, \quad R_{t+1} = R_t + \overrightarrow{IR}$$

S→I→R stochastic compartment model

Focus on Susceptible (S) - Infectious (I) - Recovered (R) model with Poisson transitions, i.e.

$$\overrightarrow{SI} \sim Po(\lambda_{\mathsf{S}\to\mathsf{I}}\mathsf{SI}/\mathsf{N}) \qquad \overrightarrow{IR} \sim Po(\lambda_{\mathsf{I}\to\mathsf{R}}\mathsf{I})$$

updating the states, i.e.

$$S_{t+1} = S_t - \overrightarrow{SI}, \quad I_{t+1} = I_t + \overrightarrow{SI} - \overrightarrow{IR}, \quad R_{t+1} = R_t + \overrightarrow{IR}$$

and binomial observations

$$Y_{\mathsf{S}\to\mathsf{I}} \sim Bin(\overrightarrow{SI}, \theta_{\mathsf{S}\to\mathsf{I}}) \qquad Y_{\mathsf{I}\to\mathsf{R}} \sim Bin(\overrightarrow{IR}, \theta_{\mathsf{I}\to\mathsf{R}})$$

S→I→R stochastic compartment model

Focus on Susceptible (S) - Infectious (I) - Recovered (R) model with Poisson transitions, i.e.

$$\overrightarrow{SI} \sim Po(\lambda_{\mathsf{S}\to\mathsf{I}}\mathsf{SI}/\mathsf{N}) \qquad \overrightarrow{IR} \sim Po(\lambda_{\mathsf{I}\to\mathsf{R}}\mathsf{I})$$

updating the states, i.e.

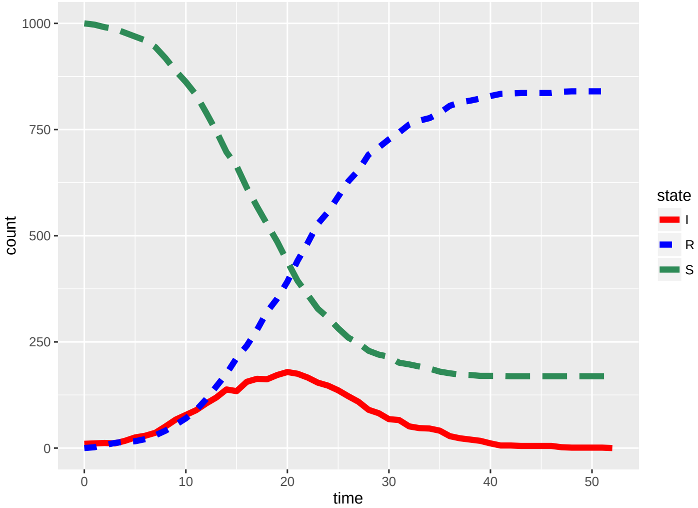
$$S_{t+1} = S_t - \overrightarrow{SI}, \quad I_{t+1} = I_t + \overrightarrow{SI} - \overrightarrow{IR}, \quad R_{t+1} = R_t + \overrightarrow{IR}$$

and binomial observations

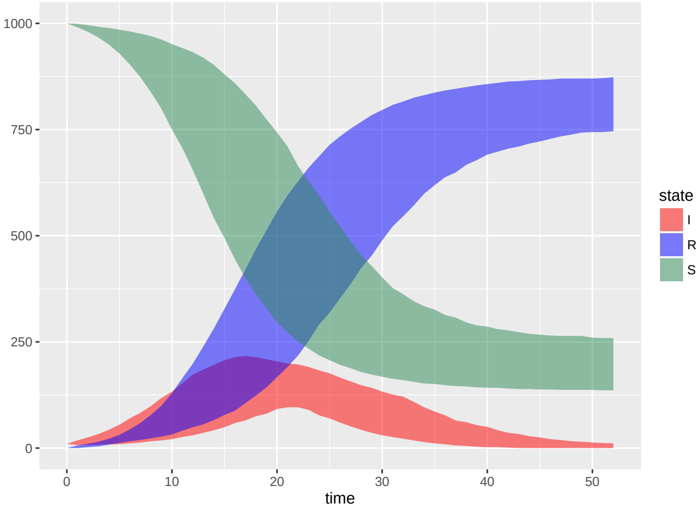
$$Y_{\mathsf{S}\to\mathsf{I}} \sim Bin(\overrightarrow{SI}, \theta_{\mathsf{S}\to\mathsf{I}}) \qquad Y_{\mathsf{I}\to\mathsf{R}} \sim Bin(\overrightarrow{IR}, \theta_{\mathsf{I}\to\mathsf{R}})$$

A more general stochastic dynamic modeling structure can be used to extended to geographical regions, subpopulations, etc.

SIR modeling simulations



SIR modeling simulations



Forecasting with perfect information

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

Forecasting with perfect information

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

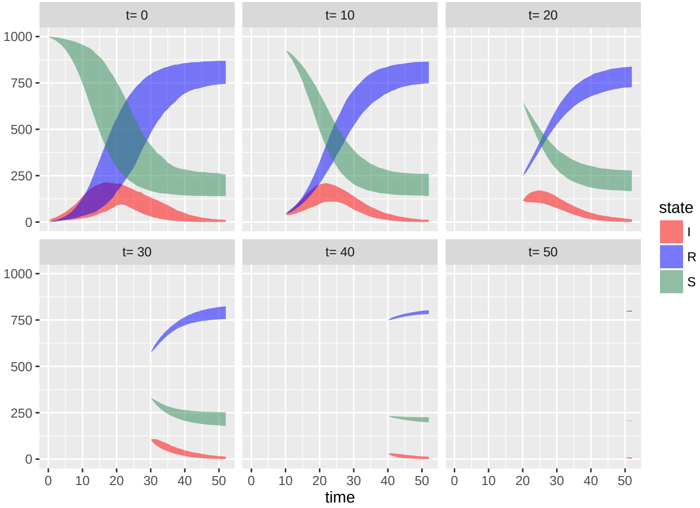
$$p(X_{t+1:T}|\theta, \lambda, X_{0:t}) = p(X_{t+1:T}|\theta, \lambda, X_t)$$

Forecasting with perfect information

Suppose you know transition rates λ, observation probabilities $\theta = 1$, 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,

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.

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

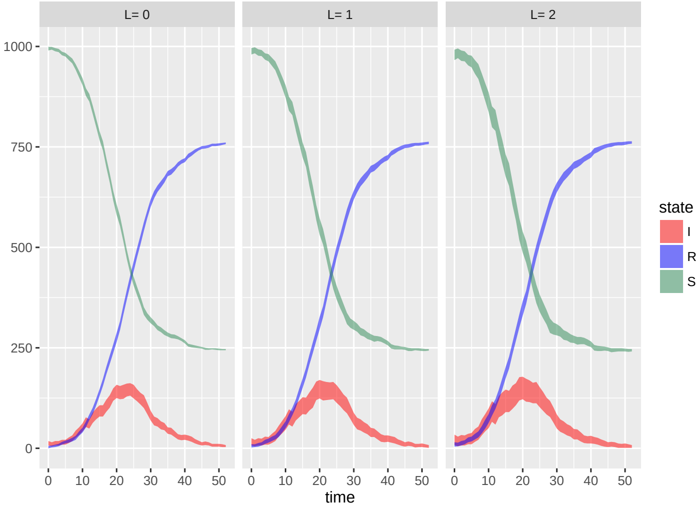
Delay in data analysis

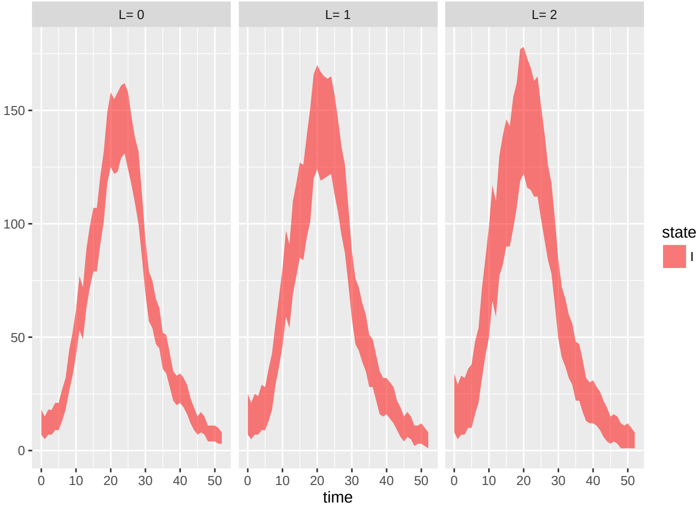
Suppose you have a one or two week delay in collecting, processing, and analyzing so that when trying to forecast future states you are using "old" data, i.e.

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

where

- $L = 0$ indicates up-to-date data
- $L = 1$ indicates one-week old data
- $L = 2$ indicates two-week old data


L= 0
L= 1
L= 2
1000
750
500
250
0
0 10 20 30 40 50
time
state
I
R
S


L= 0
L= 1
L= 2
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. .

$$Y_{\mathsf{S}\to\mathsf{I}} \sim Bin(\overrightarrow{SI}, \theta_{\mathsf{S}\to\mathsf{I}}) \qquad Y_{\mathsf{I}\to\mathsf{R}} \sim Bin(\overrightarrow{IR}, \theta_{\mathsf{I}\to\mathsf{R}})$$

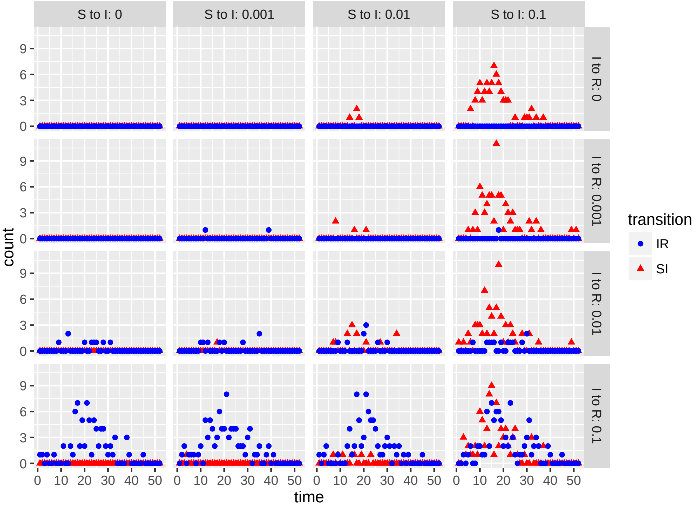
Forecasting with noisy observations

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. .

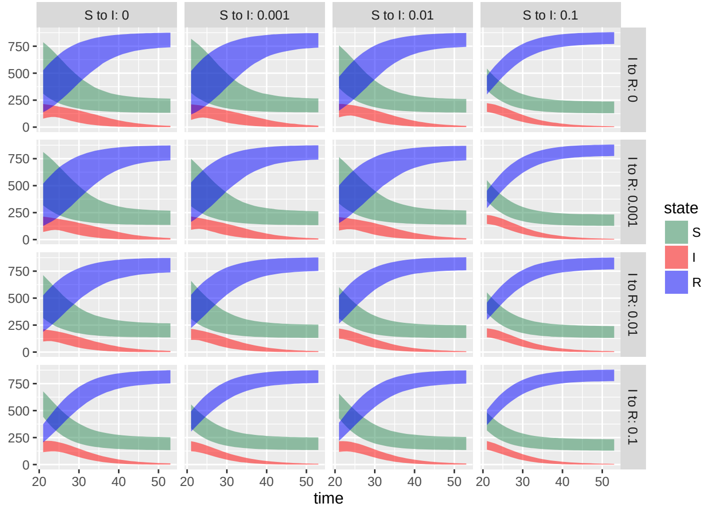
$$Y_{\mathsf{S}\to\mathsf{I}} \sim Bin(\overrightarrow{SI}, \theta_{\mathsf{S}\to\mathsf{I}}) \qquad Y_{\mathsf{I}\to\mathsf{R}} \sim Bin(\overrightarrow{IR}, \theta_{\mathsf{I}\to\mathsf{R}})$$

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: 0.001
S to I: 0.01
S to I: 0.1
I to R: 0
I to R: 0.001
I to R: 0.01
I to R: 0.1
count
time
0
3
6
9
0
10
20
30
40
50
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.

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.

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

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

Prior distributions

In order to calculate (or approximate) the integral

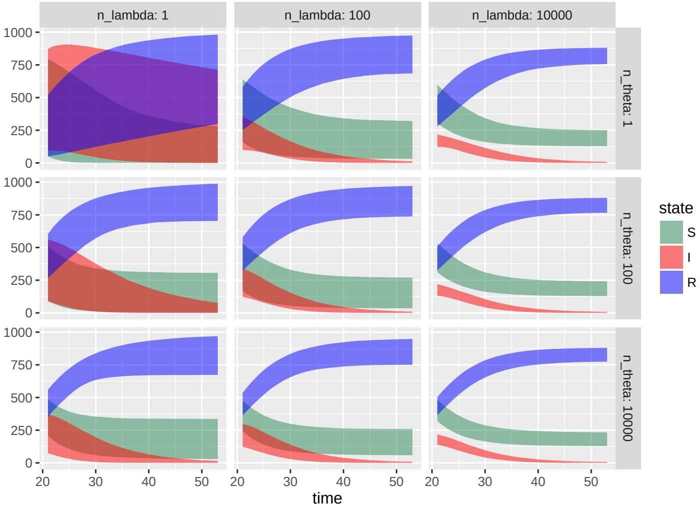
$$\int\int\int p(X_{t+1:T},\lambda,\theta|X_t)p(X_t,\lambda,\theta|y_{0:t})d\lambda d\theta dX_t$$

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

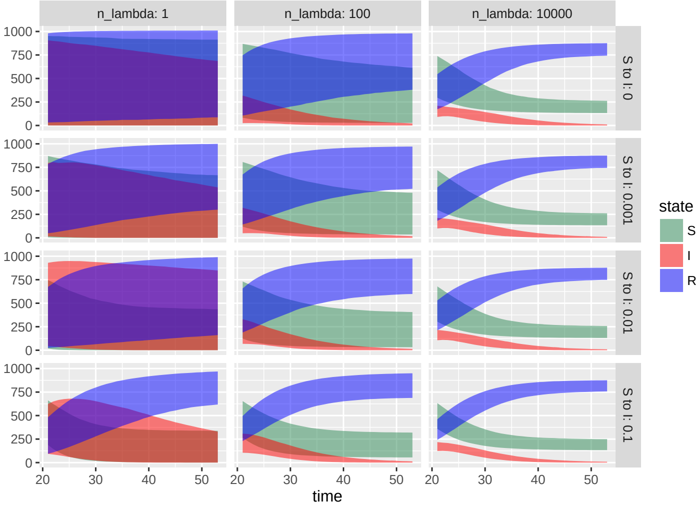
$$\begin{array}{ll} \theta_k & \stackrel{ind}{\sim} Be(n_\theta p_k, n_\theta[1-p_k]) \\ \lambda_k & \stackrel{ind}{\sim} Ga(n_\lambda c_k, n_\lambda) \\ X_0 & \sim Mult(N; z_1, \ldots, z_S) \end{array}$$

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 need information

Information can come from

- Data
- Priors

We can quantitatively assess the impact of better information,

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,

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. $Y \sim Bin(S \to I, \theta)$ by increasing θ, and

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. $Y \sim Bin(S \to I, \theta)$ by increasing θ, and
- increasing timeliness, e.g. $p(X_{t+1:T}|y_{1:t-L})$ by decreasing L.

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. $Y \sim Bin(S \to I, \theta)$ by increasing θ, and
- increasing timeliness, e.g. $p(X_{t+1:T}|y_{1:t-L})$ by decreasing L.

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. $Y \sim Bin(S \to I, \theta)$ by increasing θ, and
- 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.