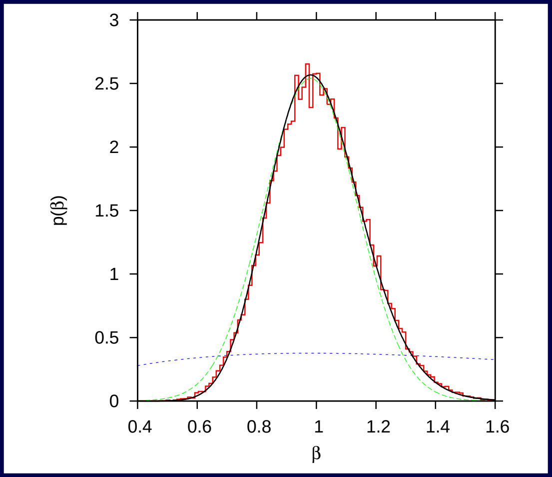
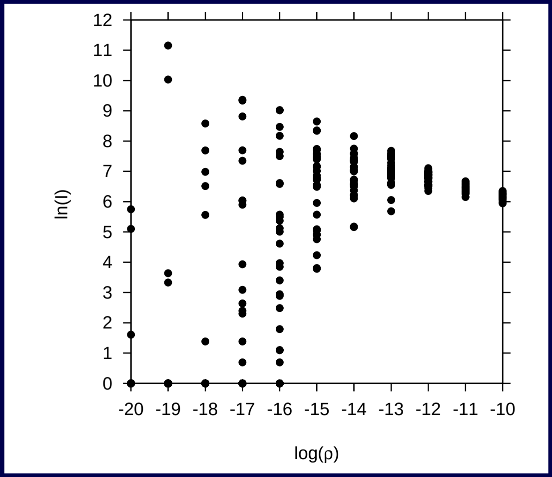
Modelling and model evaluation on empirical data in epidemiology: dynamic noise, chaos and predictability





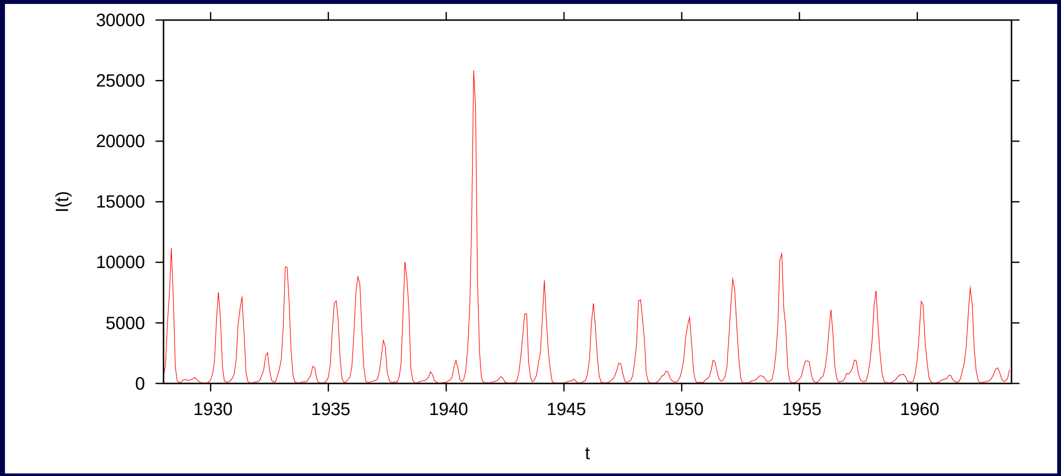
Nico Stollenwerk

Mathematical Biology Group

Centro de Matemática e Aplicações Fundamentais (CMAF)
Univ. Lisboa

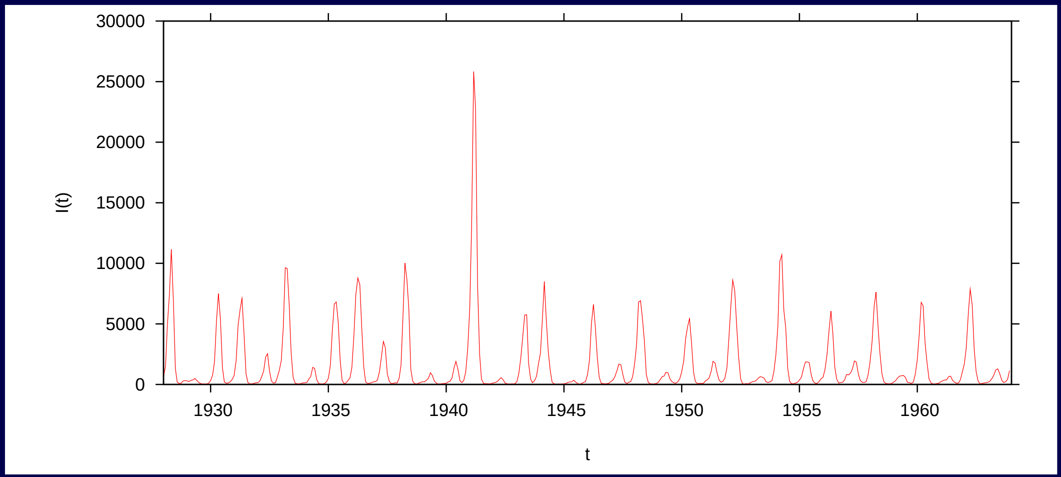
Epidemiological systems

with various qualtitative features

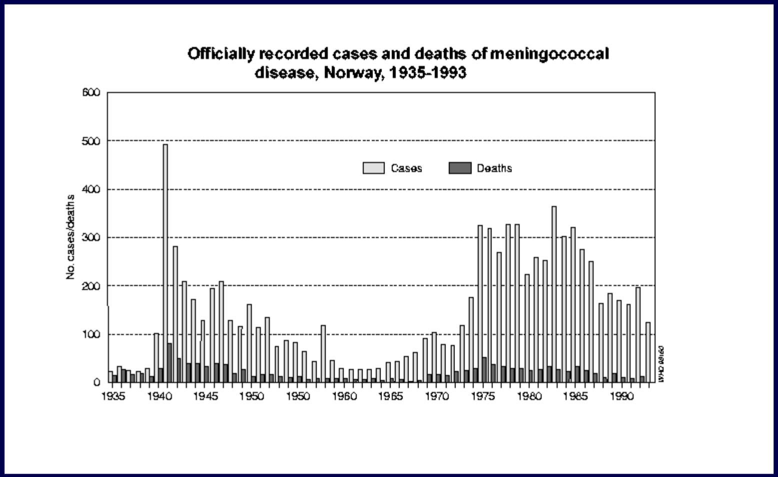


measles in New York City

Epidemiological systems with various qualtitative features



measles in New York City



meningococcal meningitis in Norway

European Union project
DENFREE: ”Dengue reasearch Framework for Resisting Epidemics in Europe”

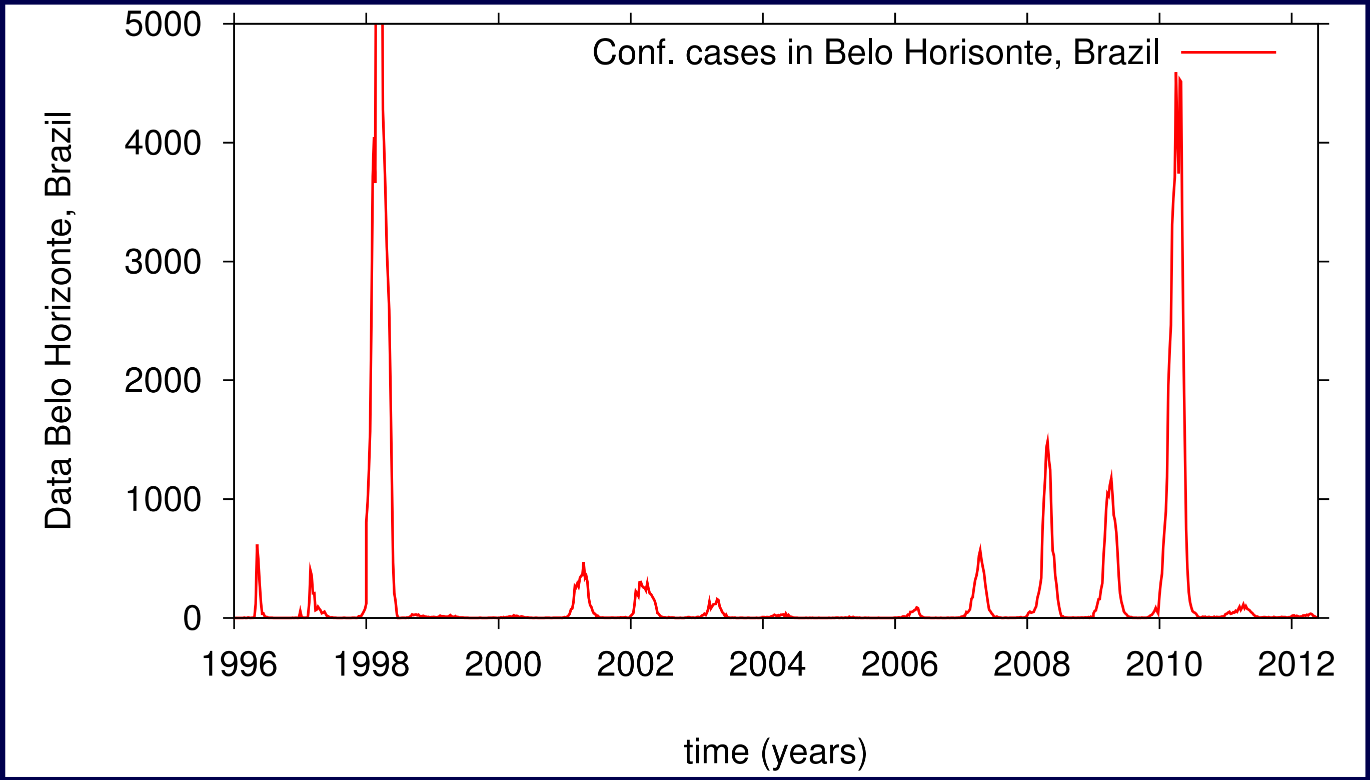
DENFREE

Dengue Research Framework For
Resisting Epidemics In Europe
Since-2012

5 years project, start January 2012

together with 2 more EU project
”the largest financial effort on dengue research world wide”

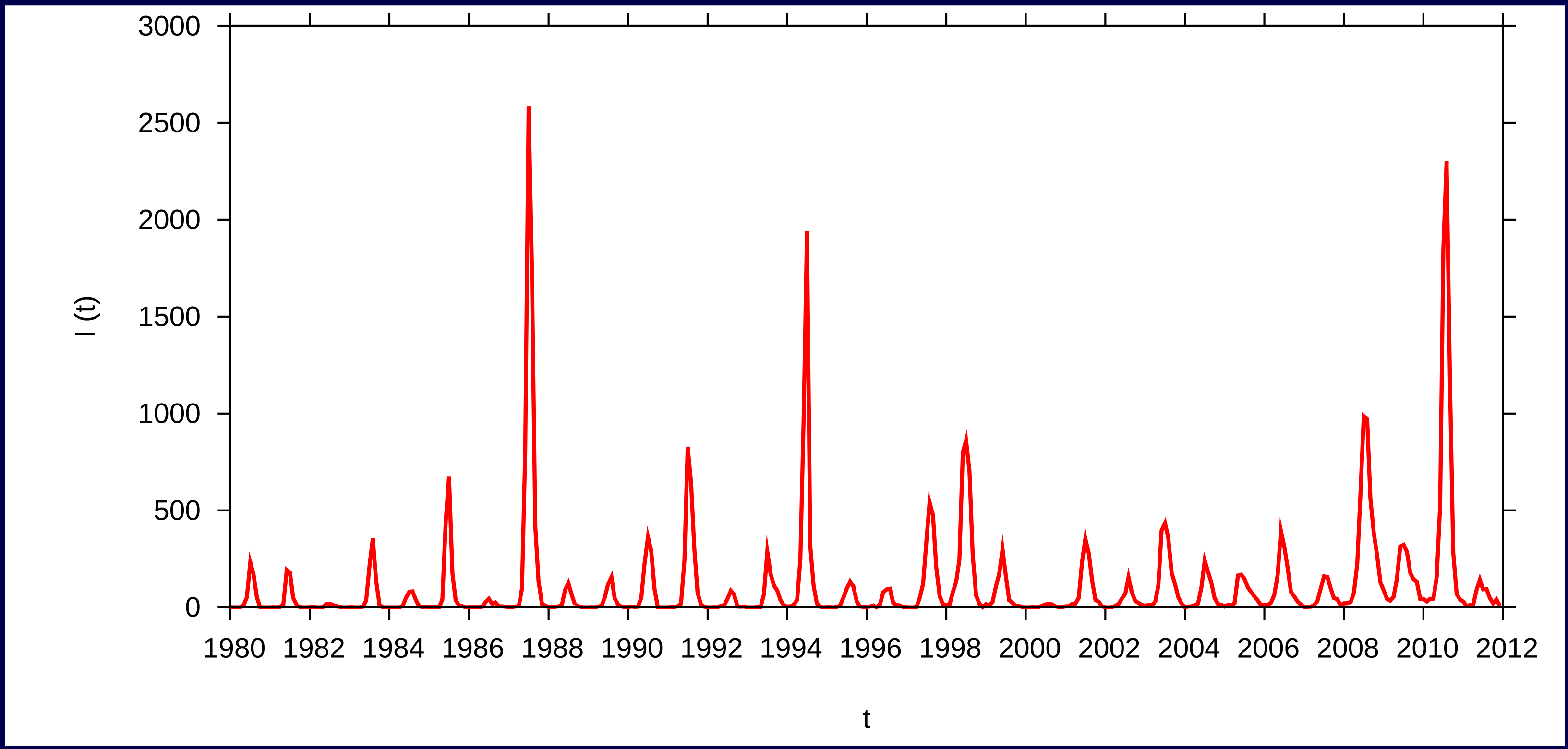
Dengue data from Brazil:
16 years of weekly notified dengue cases



city of Belo Horizonte in the state of Minas Gerais

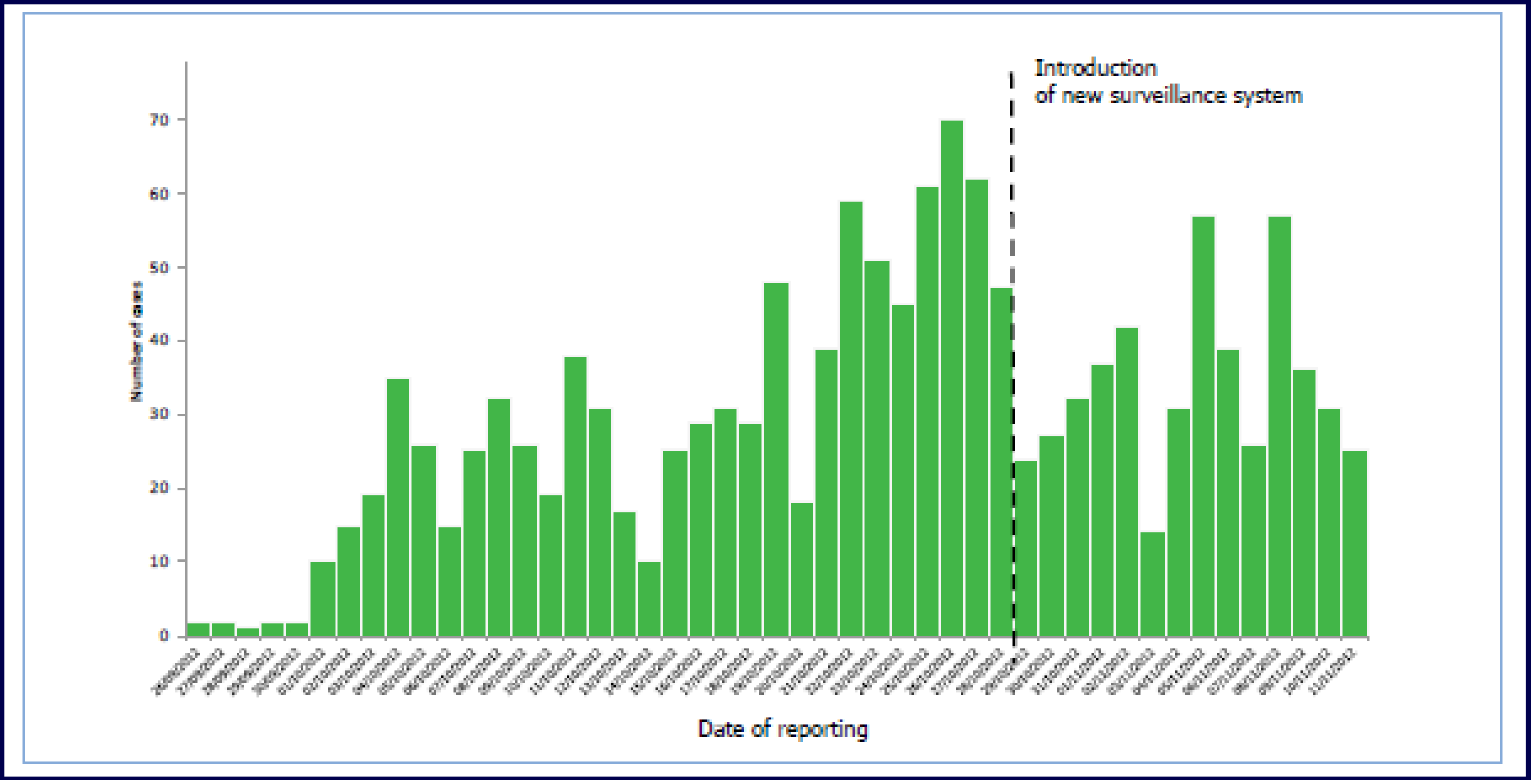
Dengue data from Thailand

32 years of symptomatic dengue cases for all 77 provinces



monthly symptomatic dengue cases
in Chiang Mai 1980-2011

Dengue fever outbreak on Madeira, Portugal, 2012
more than 2000 autochtonous cases detected



European Center for Disease Control (ECDC):

"The largest dengue outbreak in Europe since the 1920th"

Basic probability theory

joint probability

$$p(x, y)$$

marginal distribution

$$p(x) = \int p(x, y)\, dy$$

Bayes' rule

$$p(x, y) = p(x|y) \cdot p(y)$$

distribution that an event x_0 is given with certainty is $p(x) = \delta(x - x_0)$ with Dirac's delta-function

$$\int_a^b f(x) \cdot \delta(x - x_0)\, dx = f(x_0)$$

for x_0 between a and b

Application to epidemic processes

joint probability to find I_{n+1} infected at time $t+\Delta t$ and I_n at t

$$p(I_{n+1}, I_n)$$

marginal distribution to find only one of the variables no matter what the other variable does

$$p(I_{n+1}) = \sum_{I_n=0}^{N} p(I_{n+1}, I_n)$$

Bayes' rule gives conditional probability $p(I_{n+1}|I_n)$ for I_{n+1} knowing for sure I_n times $p(I_n)$

$$p(I_{n+1}, I_n) = p(I_{n+1}|I_n) \cdot p(I_n)$$

giving a dynamic evolution equation for probabilities of infected $p(I_n)$ at time t into $p(I_{n+1})$ at time $t+\Delta t$

$$p_{t+\Delta t}(I_{n+1}) = \sum_{I_n=0}^{N} p(I_{n+1}|I_n) \cdot p_t(I_n)$$

Application to epidemic processes

equation

$$p_{t+\Delta t}(I_{n+1}) = \sum_{I_n=0}^{N} p(I_{n+1}|I_n) \cdot p_t(I_n)$$

is a Perron-Frobenius type equation, and defines a time discrete Markov process

Application to epidemic processes

differential quotient gives time continous Markov process

$$\frac{p_{t+\Delta t}(I) - p_t(I)}{\Delta t} \approx \frac{d}{dt}\, p(I)$$

hence inserting time discrete version with $I := I_{n+1}$ and $\tilde{I} := I_n$

$$\frac{p_{t+\Delta t}(I) - p_t(I)}{\Delta t} = \sum_{\tilde{I}=0}^{N} \left(\frac{1}{\Delta t}\, p(I|\tilde{I}) \right) p_t(\tilde{I}) - \frac{1}{\Delta t}\, p_t(I)$$

and inserting normalization of conditioned probability $\sum_{\tilde{I}=0}^{N} p(\tilde{I}|I) = 1$ into the last term gives

$$\frac{d}{dt}\, p(I) = \sum_{\tilde{I}=0}^{N} w_{I|\tilde{I}}\, p_t(\tilde{I}) - \sum_{\tilde{I}=0}^{N} w_{\tilde{I}|I}\, p_t(I)$$

with transition rates $w_{I|\tilde{I}} := \left(\frac{1}{\Delta t}\, p(I|\tilde{I}) \right)$

Application to epidemic processes

equation

$$\frac{d}{dt}\, p(I) = \sum_{\tilde{I}=0,\tilde{I}\neq I}^{N} w_{I|\tilde{I}}\, p_t(\tilde{I}) - \sum_{\tilde{I}=0,\tilde{I}\neq I}^{N} w_{\tilde{I}|I}\, p_t(I)$$

is also called master equation and defines a time continuous state discrete Markov process

SIS epidemic

stochastic process

$$S + I \xrightarrow{\beta} I + I$$
$$I \xrightarrow{\alpha} S$$

for variable I and $S = N - I$ => probab. $p(I,t)$

$$\frac{d}{dt}\, p(I,t) = \frac{\beta}{N}(I-1)(N-(I-1))\, p(I-1,t) + \alpha(I+1)\, p(I+1,t) - \left(\frac{\beta}{N} I(N-I) + \alpha I\right)\, p(I,t)$$

mean $\langle I \rangle := \sum_{I=0}^{N} I \cdot p(I,t)$

$$\frac{d}{dt}\, \langle I \rangle = (\beta - \alpha)\langle I \rangle - \frac{\beta}{N}\langle I^2 \rangle$$

and only in mean field approx. $var := \langle I^2 \rangle - \langle I \rangle^2 \approx 0$

$$\frac{d}{dt}\, \langle I \rangle = \frac{\beta}{N}\langle I \rangle (N - \langle I \rangle) - \alpha \langle I \rangle$$

we obtain closed ODE

SIR epidemic

stochastic process

$$S + I \xrightarrow{\beta} I + I$$
$$I \xrightarrow{\gamma} R$$
$$R \xrightarrow{\alpha} S$$

for variables S, I and $R = N - S - I$ => probab. $p(S, I, t)$

$$\frac{d}{dt}\, p(S,I,t) = \frac{\beta}{N}(I-1)(S+1)\, p(S+1, I-1, t)$$
$$+\gamma(I+1)\, p(S, I+1, t)$$
$$+\alpha(N-(S+1)-I)\, p(S+1, I, t)$$
$$-\left(\frac{\beta}{N} S I + \gamma I + \alpha(N-S-I)\right)\, p(S,I,t)$$

Linear infection model

SIS model

$$S + I \xrightarrow{\beta} I + I$$
$$I \xrightarrow{\alpha} S$$

with dynamics for the probab. $p(I,t)$

$$\frac{d}{dt}p(I,t) = \frac{\beta}{N}(I-1)(N-(I-1))p(I-1,t) + \alpha(I+1)p(I+1,t) - \left(\frac{\beta}{N}I(N-I) + \alpha I\right)p(I,t)$$

simplified to susceptibles infected only outside the considered population of size N, by meeting a constant number of external infected (from much larger system) I^*, and no recovery (or cumulative cases in SIR)

$$S + I^* \xrightarrow{\beta} I + I^*$$

Linear infection model

$$S + I^* \xrightarrow{\beta} I + I^*$$

for variable I and $S = N - I$ $\quad => \quad$ probab. $p(I,t)$

$$\frac{d}{dt}p(I,t) = \frac{\beta}{N}I^* \cdot (N - (I-1))p(I-1,t) - \frac{\beta}{N}I^* \cdot (N - I)p(I,t)$$

hence constant force of infection $\beta^* := \frac{\beta}{N}I^*$

linear infection model easily solvable

Characteristic function

like ordinary mean now mean of a function

$$\langle e^{i\kappa I} \rangle := \sum_{I=0}^{N} e^{i\kappa I} \cdot p(I,t) =: g(\kappa,t)$$

Characteristic function

like ordinary mean now mean of a function

$$\langle e^{i\kappa I}\rangle := \sum_{I=0}^{N} e^{i\kappa I} \cdot p(I,t) =: g(\kappa,t)$$

generates moments

$$(-i)^n \; \frac{\partial^n}{\partial \kappa^n} g(\kappa,t)\Big|_{\kappa=0} = \langle I^n\rangle$$

Characteristic function

like ordinary mean now mean of a function

$$\langle e^{i\kappa I}\rangle := \sum_{I=0}^{N} e^{i\kappa I}\cdot p(I,t) =: g(\kappa,t)$$

generates moments

$$(-i)^n \; \frac{\partial^n}{\partial\kappa^n} g(\kappa,t)\Big|_{\kappa=0} = \langle I^n\rangle$$

and can be inverted (Fourier transform) with $\kappa =: \frac{2\pi}{N+1}\cdot k$

$$g(\kappa,t) = \sum_{I=0}^{N} e^{i\frac{2\pi}{N+1}k\cdot I}\cdot p(I,t) = \hat{g}(k,t)$$

Characteristic function

like ordinary mean now mean of a function

$$\langle e^{i\kappa I}\rangle := \sum_{I=0}^{N} e^{i\kappa I} \cdot p(I,t) =: g(\kappa,t)$$

generates moments

$$(-i)^n \; \frac{\partial^n}{\partial \kappa^n} g(\kappa,t)\Big|_{\kappa=0} = \langle I^n \rangle$$

and can be inverted (Fourier transform) with $\kappa =: \frac{2\pi}{N+1} \cdot k$

$$g(\kappa,t) = \sum_{I=0}^{N} e^{i\frac{2\pi}{N+1}k\cdot I} \cdot p(I,t) = \hat{g}(k,t)$$

then probability p as function of g

$$p(I,t) = \frac{1}{N+1} \sum_{k=0}^{N} e^{-i\frac{2\pi}{N+1}k\cdot I} \cdot \hat{g}(k,t)$$

Dynamics for $g(\kappa, t)$

use master equation of SIS stochastic system

$$\frac{\partial}{\partial t} g(\kappa, t) = \sum_{I=0}^{N} e^{i\kappa I} \cdot \frac{d}{dt} p(I, t)$$

Dynamics for $g(\kappa, t)$

use master equation of SIS stochastic system

$$\frac{\partial}{\partial t} g(\kappa, t) = \sum_{I=0}^{N} e^{i\kappa I} \cdot \frac{d}{dt} p(I, t)$$

and after some calculation

$$\frac{\partial}{\partial t} g(\kappa, t) = \beta^* N \left((e^{i\kappa} - 1) \right) \cdot g(\kappa, t) + i\beta^* (e^{i\kappa} - 1) \cdot \frac{\partial g}{\partial \kappa}$$

Solution by separation ansatz

solve partial differential equation

$$\frac{\partial}{\partial t} g(\kappa, t) = \beta^* N \left((e^{i\kappa} - 1) \right) \cdot g(\kappa, t) + i\beta^* (e^{i\kappa} - 1) \cdot \frac{\partial g}{\partial \kappa}$$

by separation ansatz first with

$$g(\kappa, t) := h(\kappa) \cdot \ell(\kappa, t)$$

giving another simpler PDE for $\ell(\kappa, t)$, and an easily solvable ODE for $h(\kappa)$

$$\frac{\partial \ell}{\partial t} = i\beta^* \left(e^{i\kappa} - 1) \right) \frac{\partial \ell}{\partial \kappa}$$

$$\frac{dh}{d\kappa} = iN \cdot h(\kappa)$$

last one with special solution $h(\kappa) = e^{iN\kappa}$

Solution by separation ansatz

solve the PDE for $\ell(\kappa, t)$

$$\frac{\partial \ell}{\partial t} = i\beta^* \left(e^{i\kappa} - 1\right)\right) \frac{\partial \ell}{\partial \kappa}$$

by another separation ansatz with

$$\ell(\kappa, t) := m(\kappa) \cdot n(t)$$

giving two separate ODEs for $n(t)$ and $m(\kappa)$ with special solutions

$$\frac{dn}{dt} = i\beta^* \cdot n(t) \qquad \Rightarrow \qquad n(t) = e^{i\beta^* t}$$

and

$$\frac{dm}{d\kappa} = \frac{1}{e^{i\kappa} - 1} \cdot m(\kappa) \qquad \Rightarrow \qquad m(\kappa) = e^{-\kappa} \cdot \left(e^{i\kappa} - 1\right)^{-i}$$

Including initial conditions

for transition probabilities take initially exactly I_0 infected at time t_0, hence

$$p(I,t_0) = \delta_{I,I_0}$$

and hence for the characteristic function

$$g(\kappa,t_0) = \sum_{I=0}^{N} e^{i\kappa I} \cdot p(I,t_0) = e^{i\kappa I_0}$$

and include initial conditions into the separation ansatz via another function $\Phi(z)$ with $z(\kappa,t) = m(\kappa) \cdot n(t)$

$$g(\kappa,t) = h(\kappa) \cdot \Phi(z) = h(\kappa) \cdot \Phi(\ell(\kappa,t))$$

and initial condition equation gives functional form of $\Phi(z)$ by inverting $z(\kappa,t_0)$ to $\kappa(z,t_0)$

$$g(\kappa,t_0) = h(\kappa) \cdot \Phi(z(\kappa,t_0)) = e^{i\kappa I_0}$$

Including initial conditions

$$g(\kappa, t_0) = h(\kappa) \cdot \Phi(z(\kappa, t_0)) = e^{i\kappa I_0}$$

resulting in $e^{-i\kappa} = e^{-i\kappa}(z, t_0)$ as function of z and t_0 as

$$e^{-i\kappa} = 1 - z^i e^{\beta^* t_0}$$

and

$$\Phi(z) = \left(1 - z^i e^{\beta^* t_0}\right)^{N - I_0}$$

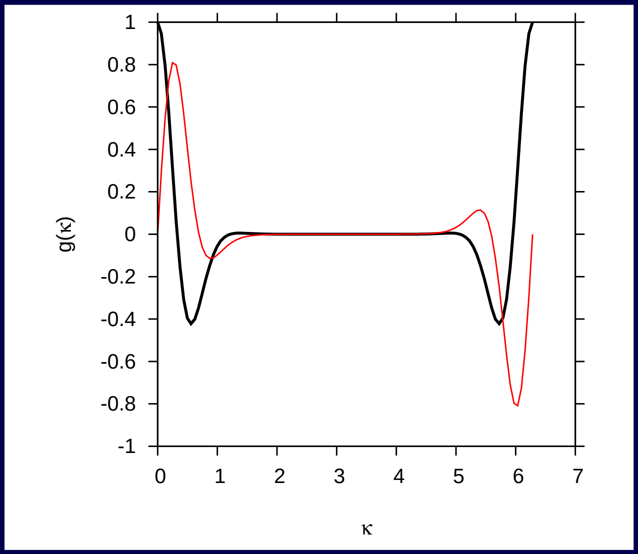
Solution of characteristic function

the solution for all times, including the initial conditions, is now given by

$$g(\kappa, t) = h(\kappa) \cdot \Phi(z(\kappa, t))$$

resulting in

$$g(\kappa, t) = e^{i\kappa N} \cdot \left(e^{-i\kappa} e^{-\beta^*(t-t_0)} + (1 - e^{-\beta^*(t-t_0)}) \right)^{N-I_0}$$



real and imaginary part of $g(\kappa)$ for fixed t

Solution of characteristic function

the solution for all times, including the initial conditions, is now given by

$$g(\kappa, t) = h(\kappa) \cdot \Phi(z(\kappa, t))$$

resulting in

$$g(\kappa, t) = e^{i\kappa N} \cdot \left(e^{-i\kappa} e^{-\beta^*(t-t_0)} + (1 - e^{-\beta^*(t-t_0)}) \right)^{N-I_0}$$

and with $p(I, t) = \frac{1}{N+1} \sum_{k=0}^{N} e^{-i\frac{2\pi}{N+1}k \cdot I} \cdot g(\kappa(k), t)$
(Fourier back-transformation)

$$p(I, t) = \binom{N - I_0}{I - I_0} \left(e^{-\beta^*(t-t_0)} \right)^{N-I} \left(1 - e^{-\beta^*(t-t_0)} \right)^{I-I_0}$$

this is also the transition probability $p(I, t | I_0, t_0)$
needed for the likelihood function

Stochastic simulation

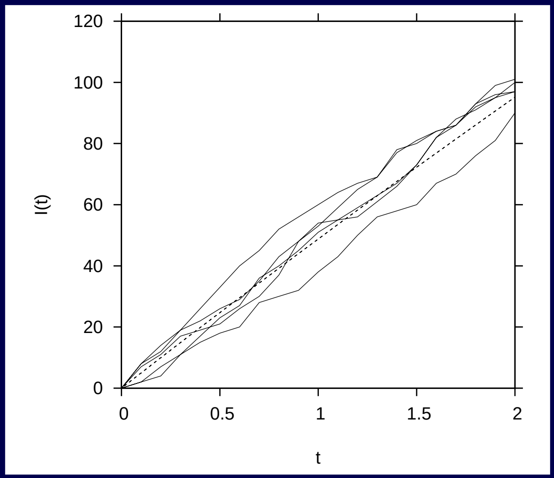
linear infection model as stochastic process

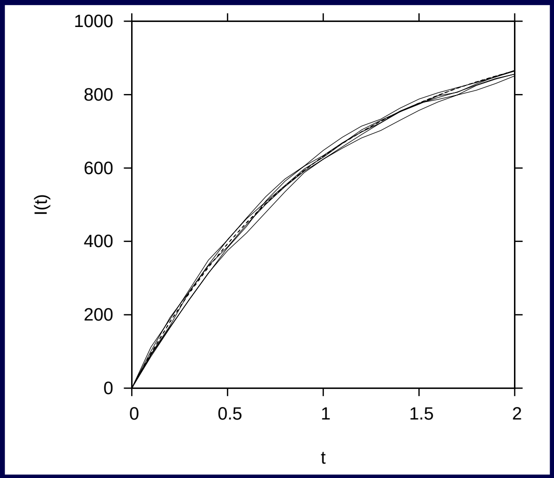
$$S + I^* \xrightarrow{\beta} I + I^*$$

for variable I and $S = N - I$ => probab. $p(I, t)$

$$\frac{d}{dt}p(I,t) = \beta^*(N-(I-1))p(I-1,t) - \beta^*(N-I)p(I,t)$$

simulated by e.g. Gillespie algorithm





$\beta^* = 0.05$ $\beta^* = 1.0$

Stochastic simulation

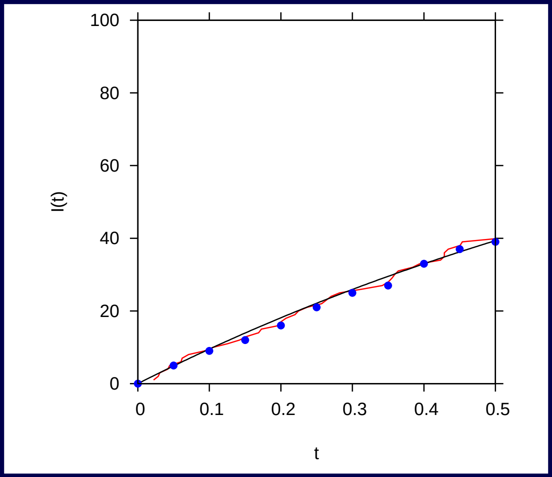
linear infection model as stochastic process

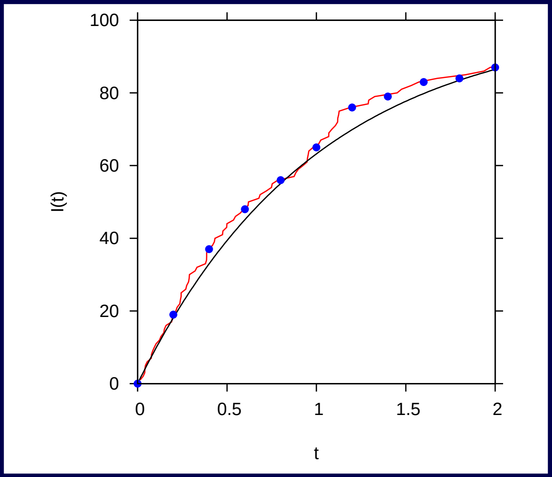
$$S + I^* \xrightarrow{\beta} I + I^*$$

for variable I and $S = N - I$ $\quad => \quad$ probab. $p(I,t)$

$$\frac{d}{dt}p(I,t) = \beta^*(N-(I-1))p(I-1,t) - \beta^*(N-I)p(I,t)$$

simulated by e.g. Gillespie algorithm





take data points for parameter estimation

Likelihood function from data $(I_0, I_1, ... I_n)$

joint probability of data points

$$p(I_n, t_n, I_{n-1}, t_{n-1}, ..., I_1, t_1, I_0, t_0) = \prod_{\nu=0}^{n-1} p(I_{\nu+1}, t_{\nu+1} | I_\nu, t_\nu) \cdot p(I_0, t_0)$$

Likelihood function from data $(I_0, I_1, ...I_n)$

joint probability of data points

$$p(I_n, t_n, I_{n-1}, t_{n-1}, ..., I_1, t_1, I_0, t_0) = \prod_{\nu=0}^{n-1} p(I_{\nu+1}, t_{\nu+1}|I_\nu, t_\nu) \cdot p(I_0, t_0)$$

inserting solution of stochastic process

$$p(I, t|I_0, t_0) = \binom{N - I_0}{I - I_0} \left(e^{-\beta(t-t_0)}\right)^{N-I} \left(1 - e^{-\beta(t-t_0)}\right)^{I-I_0}$$

Likelihood function from data $(I_0, I_1, ... I_n)$

joint probability of data points

$$p(I_n, t_n, I_{n-1}, t_{n-1}, ..., I_1, t_1, I_0, t_0) = \prod_{\nu=0}^{n-1} p(I_{\nu+1}, t_{\nu+1} | I_\nu, t_\nu) \cdot p(I_0, t_0)$$

inserting solution of stochastic process

$$p(I, t | I_0, t_0) = \binom{N - I_0}{I - I_0} \left(e^{-\beta(t-t_0)} \right)^{N-I} \left(1 - e^{-\beta(t-t_0)} \right)^{I-I_0}$$

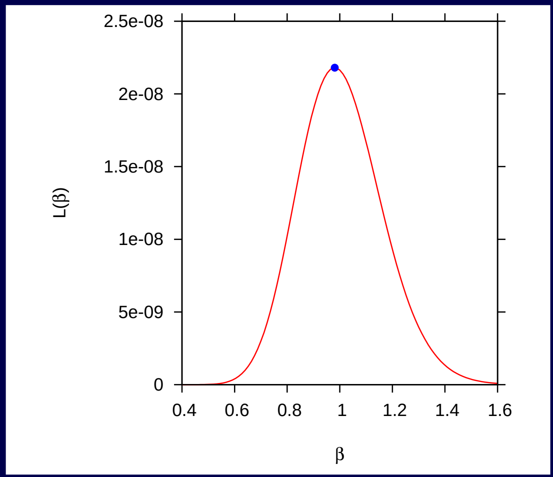
gives likelihood function

$$L(\beta) = \prod_{\nu=0}^{n-1} \binom{N - I_\nu}{I_{\nu+1} - I_\nu} \left(e^{-\beta(\Delta t)} \right)^{N-I_{\nu+1}} \left(1 - e^{-\beta(\Delta t)} \right)^{I_{\nu+1}-I_\nu}$$

Likelihood function from data $(I_0, I_1, \ldots I_n)$

likelihood function

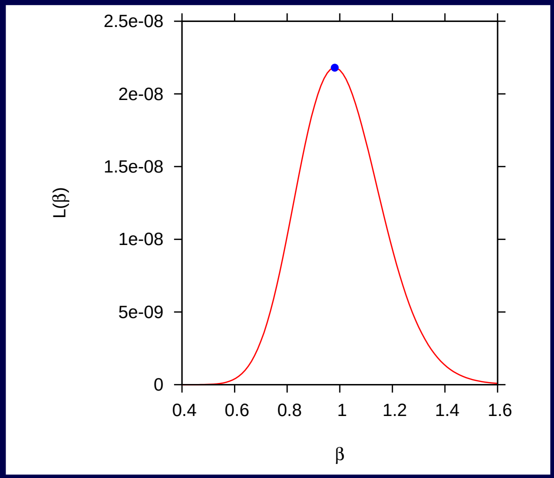
$$L(\beta) = \prod_{\nu=0}^{n-1} \binom{N - I_\nu}{I_{\nu+1} - I_\nu} \left(e^{-\beta(\Delta t)}\right)^{N - I_{\nu+1}} \left(1 - e^{-\beta(\Delta t)}\right)^{I_{\nu+1} - I_\nu}$$



Likelihood function from data $(I_0, I_1, ... I_n)$

likelihood function

$$L(\beta) = \prod_{\nu=0}^{n-1} \binom{N - I_\nu}{I_{\nu+1} - I_\nu} \left(e^{-\beta(\Delta t)}\right)^{N - I_{\nu+1}} \left(1 - e^{-\beta(\Delta t)}\right)^{I_{\nu+1} - I_\nu}$$



maximizing the likelihood $\frac{\partial L}{\partial \beta} = 0$ gives best estimator

$$\hat{\beta} = \frac{1}{\Delta t} \cdot ln \left(\frac{N - \frac{1}{n} \sum_{\nu=0}^{n-1} I_\nu}{N - \frac{1}{n} \sum_{\nu=0}^{n-1} I_{\nu+1}} \right)$$

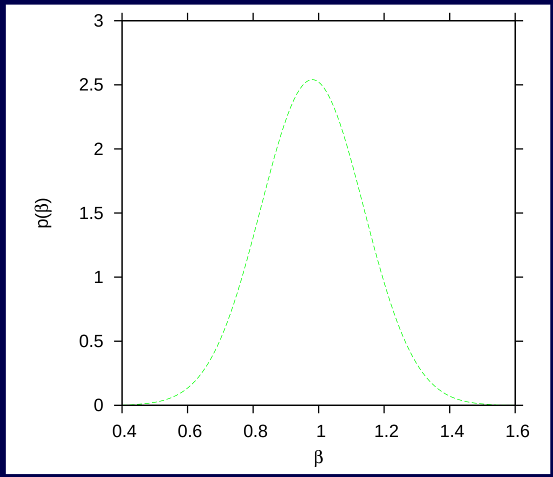
Confidence intervals via Fisher information

assume Gaussianity around the maximum of likelihood

$$p(\beta) := \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{(\beta-\hat{\beta})^2}{2\sigma^2}}$$

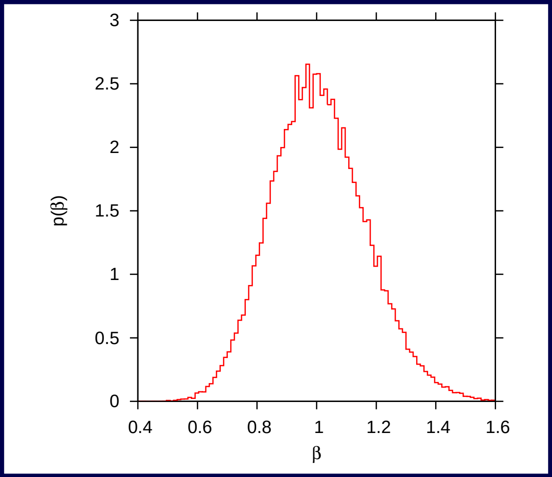
second derivative around maximum gives σ

$$\left.\frac{\partial^2 p(\beta)}{\partial \beta^2}\right|_{\beta=\hat{\beta}} = -\frac{1}{\sigma} \quad , \quad \sigma = -\frac{1}{\left.\frac{\partial^2 L(\beta)}{\partial \beta^2}\right|_{\beta=\hat{\beta}}}$$



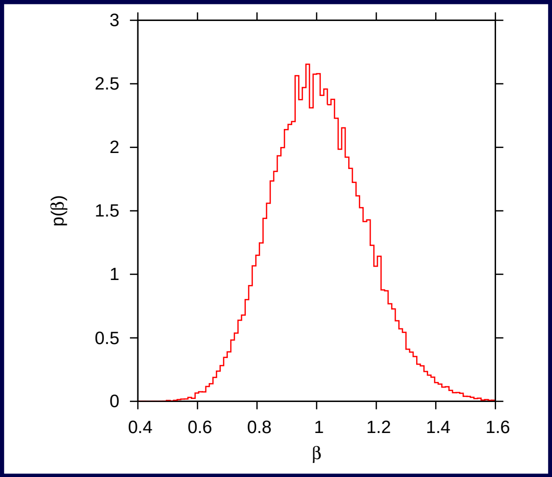
Experiment: many realizations

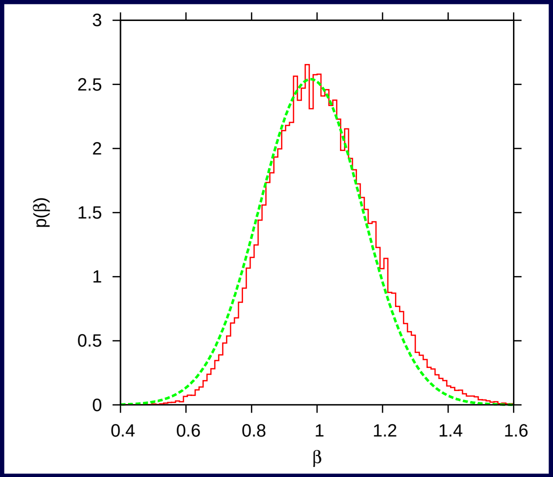
simulate many realizations of stochastic process
take histogram of best estimates



Experiment: many realizations

simulate many realizations of stochastic process
take histogram of best estimates





Gaussian approximation compares relatively well

Likelihood function for multiple paramters

joint probability of data points

$$p(I_n, t_n, I_{n-1}, t_{n-1}, \ldots, I_1, t_1, I_0, t_0) = \prod_{\nu=0}^{n-1} p(I_{\nu+1}, t_{\nu+1} | I_\nu, t_\nu) \cdot p(I_0, t_0)$$

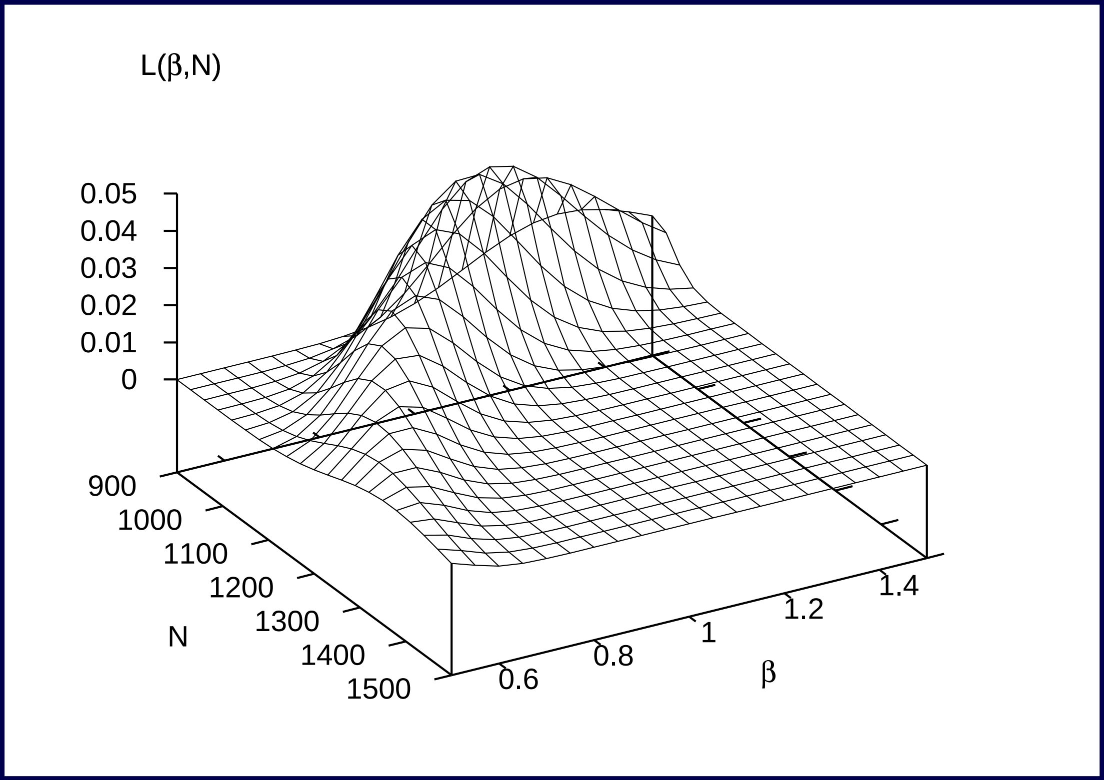
inserting solution of stochastic process

$$p(I, t | I_0, t_0) = \binom{N - I_0}{I - I_0} \left(e^{-\beta(t-t_0)}\right)^{N-I} \left(1 - e^{-\beta(t-t_0)}\right)^{I-I_0}$$

gives likelihood function

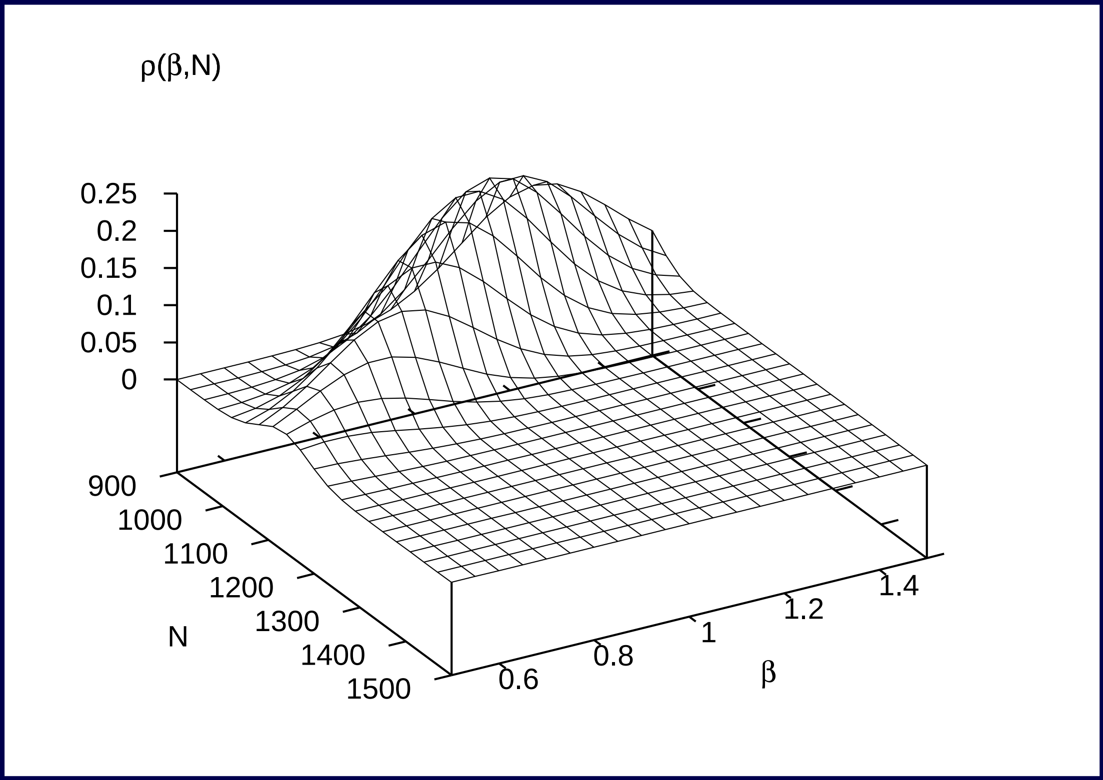
$$L(\beta, N) = \prod_{\nu=0}^{n-1} \binom{N - I_\nu}{I_{\nu+1} - I_\nu} \left(e^{-\beta(\Delta t)}\right)^{N-I_{\nu+1}} \left(1 - e^{-\beta(\Delta t)}\right)^{I_{\nu+1}-I_\nu}$$

Likelihood function



Likelihood per data point

Likelihood function



Gaussian approximation

Generalization to further models: Euler-multinomial approximation

approximation for small time steps $\Delta t = t - t_0$

$$S + I_0 \xrightarrow{\beta} I + I_0$$

gives stochastic process for decay of suseptibles S

$$\frac{d}{dt} p(S,t) \ = \ \frac{\beta}{N} I_0 (S+1) p(S+1,t) - \frac{\beta}{N} I_0 S p(S,t)$$

giving

$$p(S,t|S_0,t_0) = \begin{pmatrix} S_0 \\ S \end{pmatrix} \left(e^{-\frac{\beta}{N} I_0 (t-t_0)} \right)^S \left(1 - e^{-\frac{\beta}{N} I_0 (t-t_0)} \right)^{S_0 - S}$$

updating at time t_1 to $S_1 = S$ and $I_1 = I_0 + (S_0 - S_1)$ giving

$$p(S_1, t_0 + \Delta t|S_0,t_0) = \begin{pmatrix} S_0 \\ S_1 \end{pmatrix} \left(e^{-\frac{\beta}{N} I_0 \Delta t} \right)^{S_1} \left(1 - e^{-\frac{\beta}{N} I_0 \Delta t} \right)^{S_0 - S_1}$$

Generalization to further models: Euler-multinomial approximation

approximation for small time steps $\Delta t = t - t_0$

$$S + I_0 \xrightarrow{\beta} I + I_0$$

gives stochastic process for decay of suseptibles S

$$\frac{d}{dt} p(S,t) = \frac{\beta}{N} I_0 (S+1) p(S+1,t) - \frac{\beta}{N} I_0 S p(S,t)$$

giving

$$p(S,t|S_0,t_0) = \begin{pmatrix} S_0 \\ S \end{pmatrix} \left(e^{-\frac{\beta}{N} I_0 (t-t_0)} \right)^S \left(1 - e^{-\frac{\beta}{N} I_0 (t-t_0)} \right)^{S_0 - S}$$

updating at time t_1 to $S_1 = S$ and $I_1 = I_0 + (S_0 - S_1)$ giving

$$p(I_1, t_0 + \Delta t | I_0, t_0) = \begin{pmatrix} N - I_0 \\ N - I_1 \end{pmatrix} \left(e^{-\frac{\beta}{N} I_0 \Delta t} \right)^{N - I_1} \left(1 - e^{-\frac{\beta}{N} I_0 \Delta t} \right)^{I_1 - I_0}$$

Generalization to further models: Euler-multinomial approximation

in the same way "decay of infected"

$$I \xrightarrow{\alpha} S$$

gives stochastic process for decay of infected I

$$\frac{d}{dt} p(I,t) \, = \, \alpha(I+1)p(I+1,t) - \alpha I p(S,t)$$

updating at time $t_0 + \Delta t$ to I_2 and $S_2 = S_0 + (I_0 - I_1)$ giving

$$p(I_2, t_0 + \Delta t | I_0, t_0) = \begin{pmatrix} I_0 \\ I_2 \end{pmatrix} \, \left(e^{-\alpha \Delta t} \right)^{I_2} \left(1 - e^{-\alpha \Delta t} \right)^{I_0 - I_2}$$

Generalization to further models: Euler-multinomial approximation

and putting everything together to the final update for the full SIS model

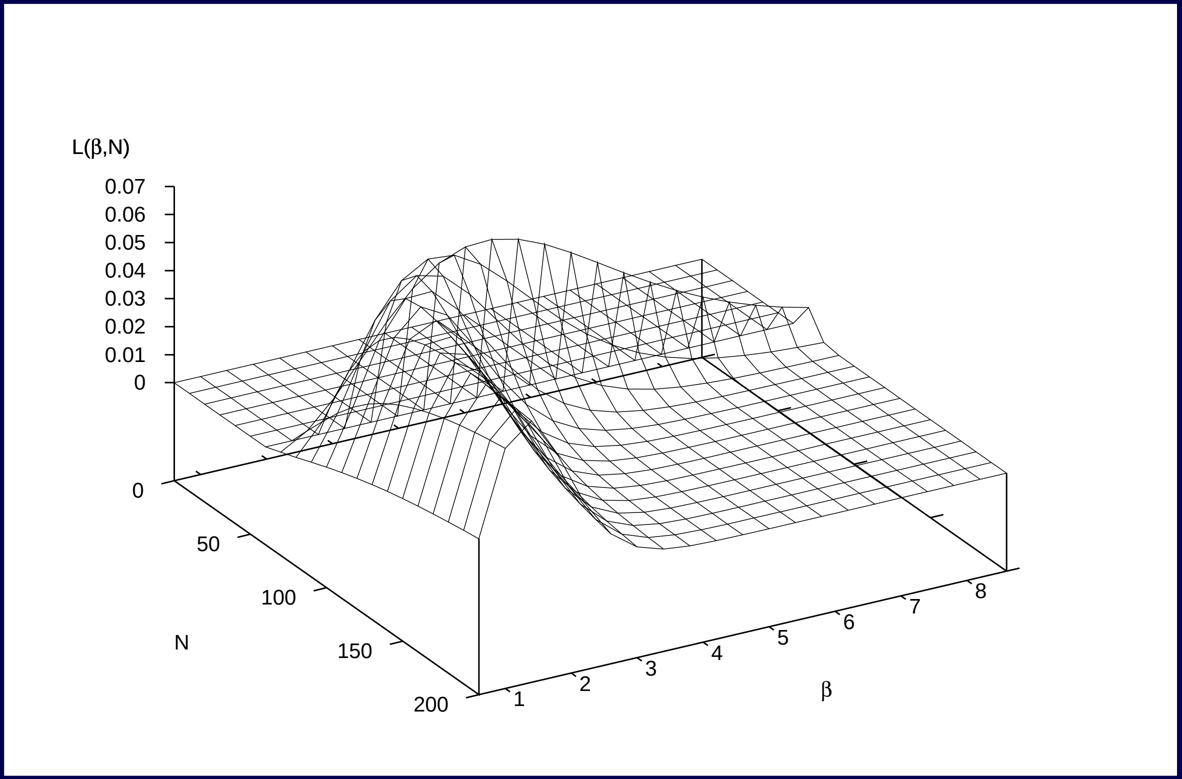
$$S + I_0 \xrightarrow{\beta} I + I_0$$
$$I \xrightarrow{\alpha} S$$

gives with update rules $I_t = I_0 + I_1 - (N - I_2)$ and its stochastic version $p(I_t|I_1, I_2) = \delta_{I_2, N-I_0+I_t-I_1}$

$$p(I_t, t|I_0, t_0) = \sum_{I_1=0}^{N-I_0} \sum_{I_2=0}^{I_0} p(I_t|I_1, I_2) \cdot p(I_2, t_0 + \Delta t|I_0, t_0) \cdot p(I_1, t_0 + \Delta t|I_0, t_0)$$

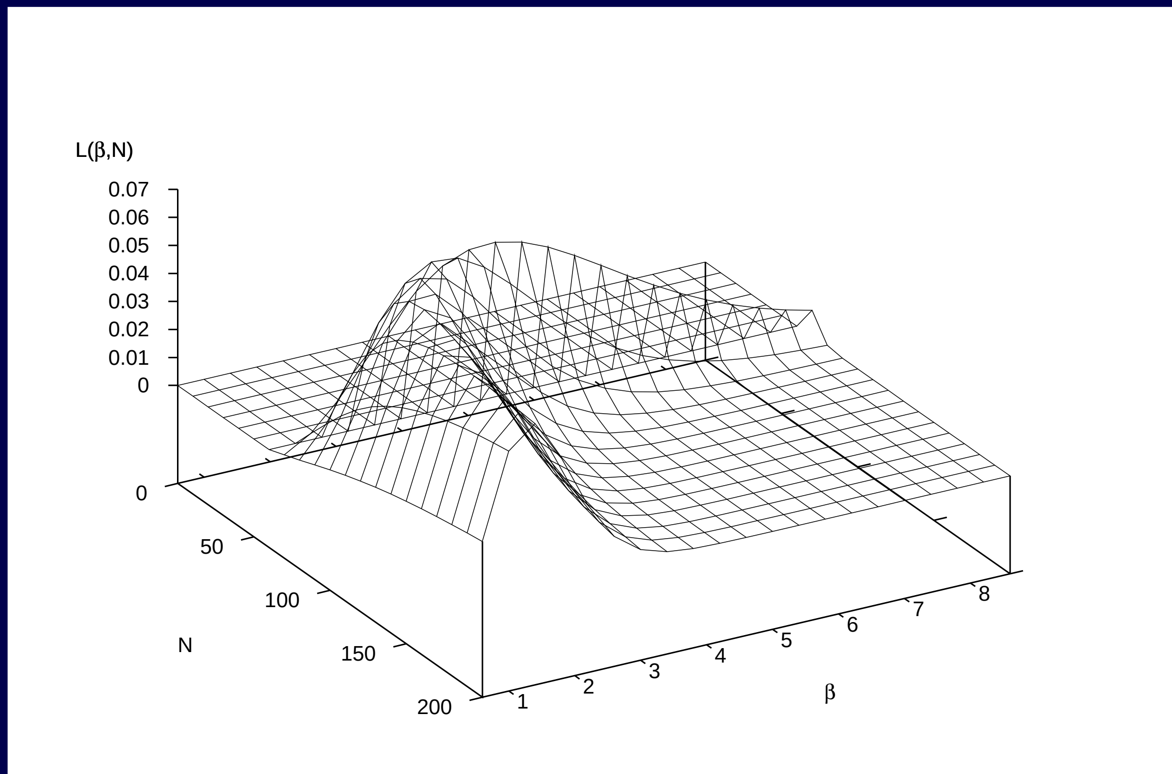
and from this again the likelihood, but sticking with eventually large summations in it

Likelihood function: Euler-multinomial approximation



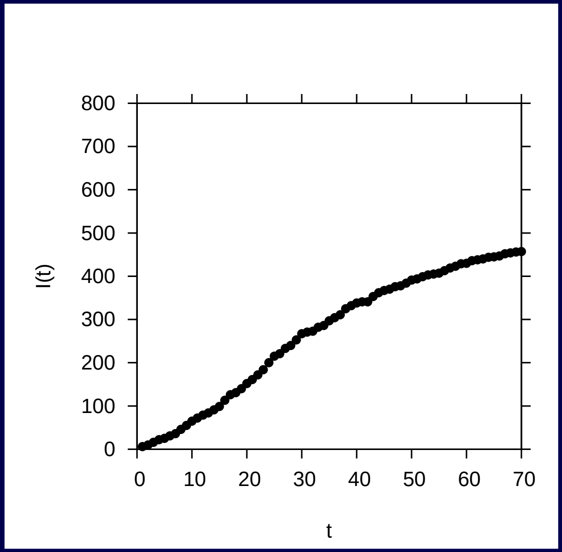
Likelihood per data point

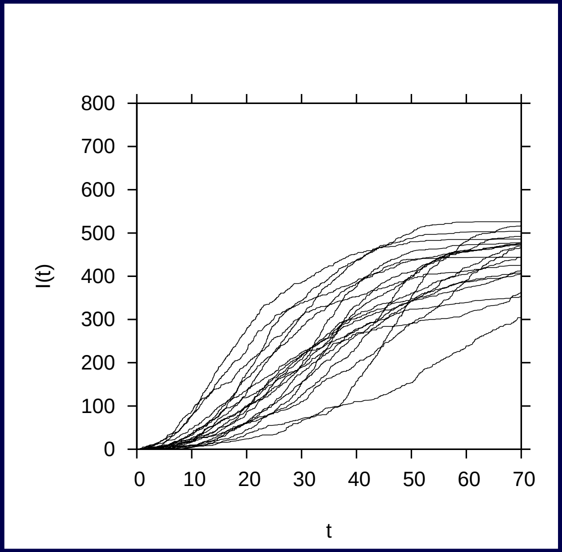
Likelihood function: Euler-multinomial approximation



likelihood profile versus likelihood section

Comparison of data with simulations





flu data cumulative

simulations of SIR-system

number of simulations in η-ball vicinity to data set gives likelihood of data under this model parameter set

=> estimate of likelihood function (Stollenwerk, Briggs 2000)

Comparison of data with simulations

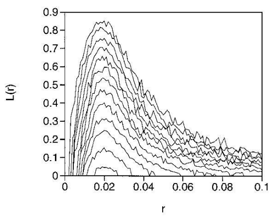


Fig. 4. Empirical likelihood curves for the one parameter r for various values of η-neighbourhoods. The maximum does not change much with varying η, showing that the estimate for the parameter is rather robust.

mates of the parameters used for our likelihood sections are obtained with this method.

From the Padé approximation in Section 6 instead of likelihood sections we could also easily generate likelihood profiles as described for example in Ref. [12]. These profiles are calculated by varying one parameter and maximizing the likelihood in respect to all other model parameters, which is rather cumbersome for the Empirical Likelihood Method due to the fluctuations around the empirical likelihood maximum (see Fig. 4). In biological systems one often has information about some of the model parameters from other experiments and searches for an otherwise difficult obtainable parameter like the contact rate, which is r in our case. In such situations the Empirical Likelihood Method is easiest and best applicable. However, we have also investigated empirical likelihoods with variation of two parameters [11].

8. Summary and prospect

We have solved the Master equation for a plant disease model analytically and also obtained numerically stable solutions over the whole range of states, which was previously not possible using the matrix exponential.

The solution is used for constructing likelihood sections from empirical microcosm data. The Master equation approach can be easily generalized to more complex models, allowing for likelihood estimations on the basis of simulated trajectories. Further research on this Empirical Likelihood Method is in progress.

The form of the Master equation we use here gives exponential waiting times between events and in the Gillespie algorithm this property is used explicitly to construct stochastic realizations of the process. However, the exponential waiting time is not a principal restriction, but arbitrary waiting time distributions can be included in a Master equation with time-convolution [13,14]. It would be an interesting extension of the present work to combine numerically this time-convolved Master equation with our Empirical Likelihood Method.

Also the Master equation approach opens naturally the way to a Fock space formulation of stochastic processes [15] which is easily generalizable to the field theoretic treatment of spatial epidemic systems see Ref. [16], and related Refs. [17–21]). Such a field theory is needed to describe the underlying experimental system more appropriately, as first experiments by Bailey et al. indicate [22]). The time decaying susceptibility drives the system through a threshold region between a simple spreading regime and a non-spreading regime.

Acknowledgements

We gratefully acknowledge discussions and provision of experimental data by Gavin Gibson, Adam Kleczkowski and Doug Bailey, and financial support of the BBSRC obtained by Chris Gilligan. We also thank the referees for some helpful references.

References

[1] N.G. van Kampen, Stochastic Processes in Physics and Chemistry, North-Holland, Amsterdam, 1992.
[2] N.S. Goel, N. Richter-Dyn, Stochastic models in Biology, Academic Press, New York, 1974.
[3] A. Kleczkowski, D.J. Bailey, C.A. Gilligan, Proc. R. Soc. (London) B 263 (1996) 777.
[4] C.W. Gardiner, Handbook of Stochastic Methods, Springer, Berlin, 1985;

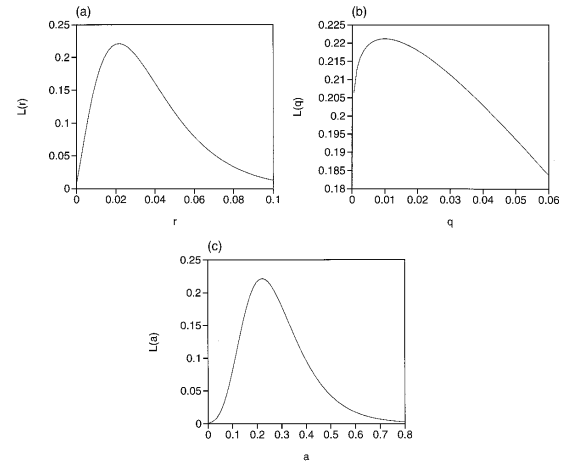


Fig. 3. Likelihood sections for all three parameters, i.e. variation of one parameter, keeping the others fixed at their maximal values, as obtained from the likelihood maximization. The estimates are: $r = 0.022$, $q = 0.0099$ and $a = 0.22$.

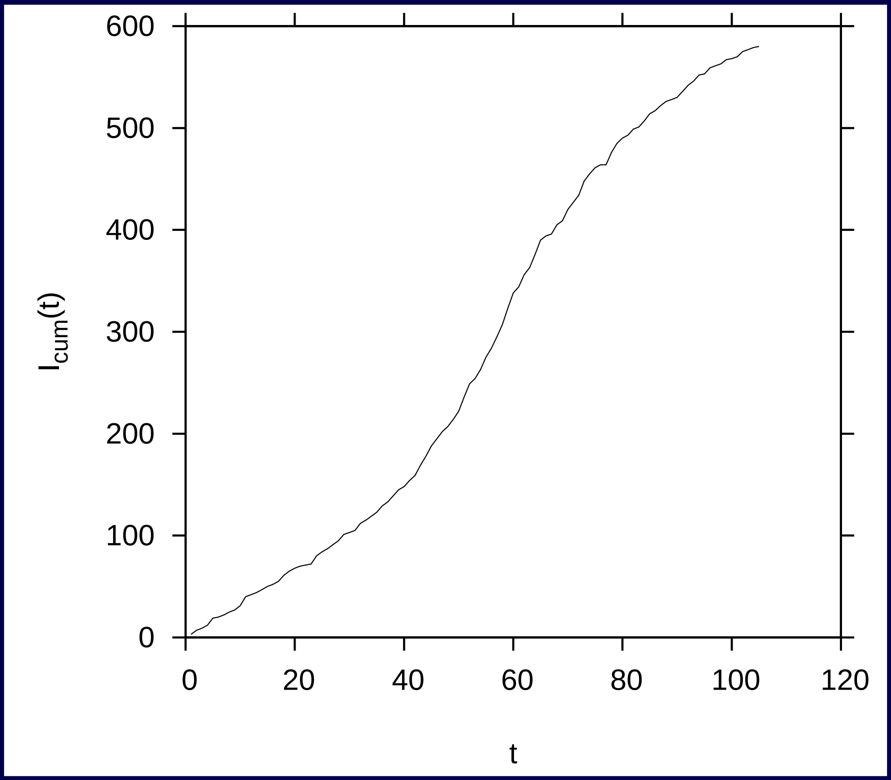
using the β-recursion, i.e. using Eq. (14). We obtained in this way the same value for L from both methods. Only the machine precision prevented using the β-recursion for higher values of k_i.

7. Empirical likelihood

The above mentioned solution cannot be carried through to more general Master equations, which have different time-dependent transition rates for different transitions as likely in multicompartmental models, for example models with an additional exposed class (SEI models) between susceptible and infected classes (as in the SI models we consider here). Still, the single trajectory simulation method holds for time-dependent multicompartment models and even can be used for constructing empirically obtained likelihoods. We experiment with such a method by estimating the joint probability of the data, that is, Eq. (12), directly from simulated stochastic trajectories. In the space of dimensionality of the number of data points the estimate is given by using balls around the measured data with radius η (η-balls) and counting the number of simulated trajectories inside these neighborhoods (for details see a forthcoming article by Stollenwerk [11]). The esti-

estimate of likelihood function (Stollenwerk, Briggs 2000)

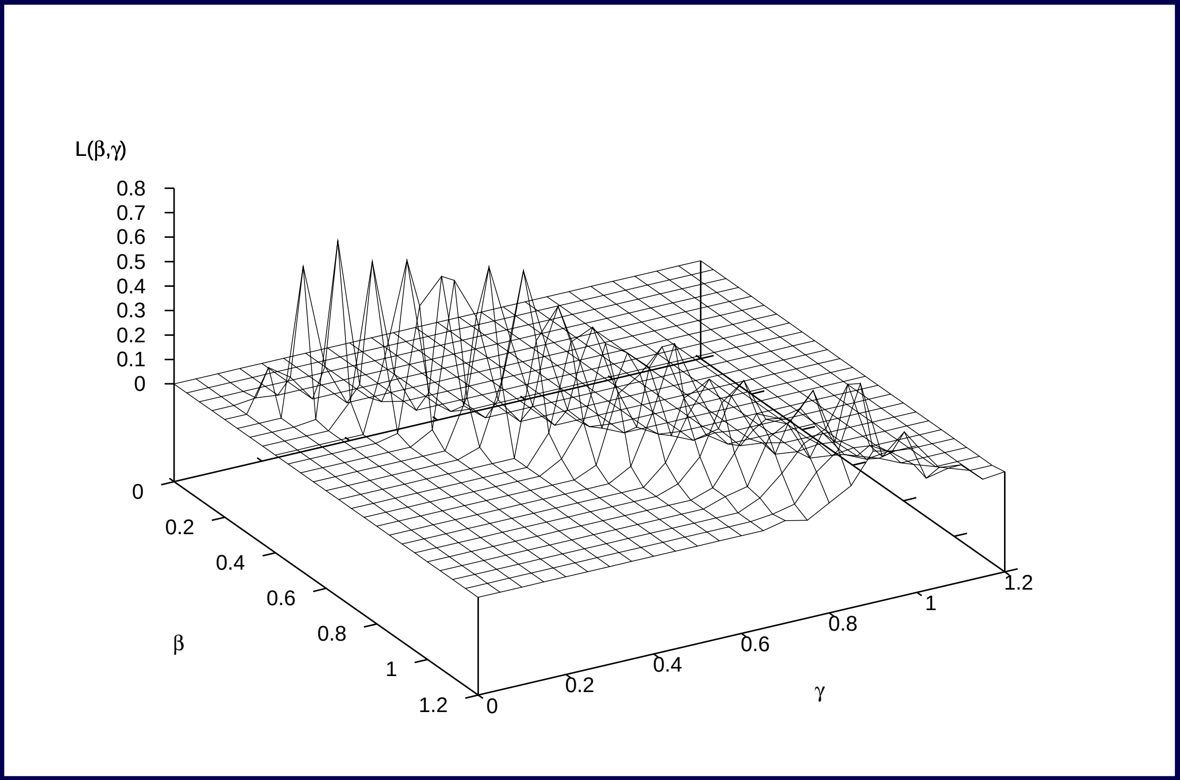
η-ball method for Dutch influenza data



daily influenza data between 1st of January and 15th of April 2007 for the Netherlands (from InfluenzaNet, EPIWORK project)

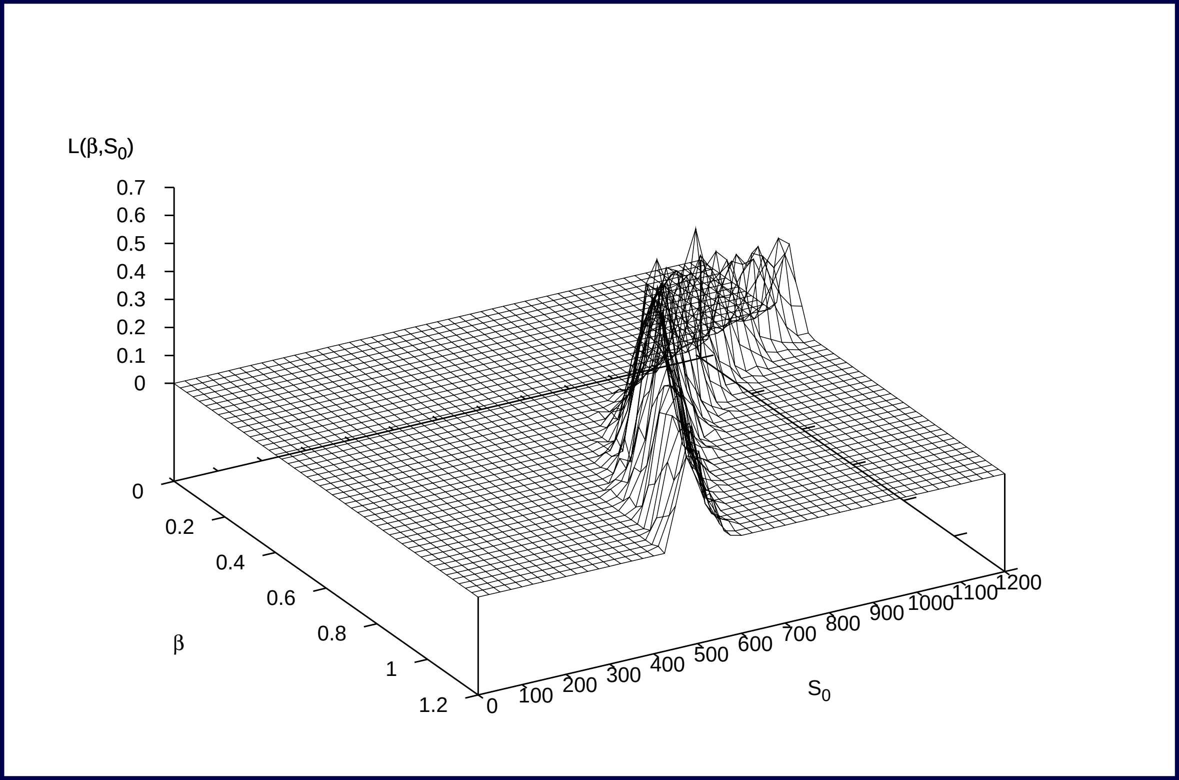
... to be compared with SIR stochastic simulations for various parameter values

Estimated likelihood function



Likelihood per data point

Estimated likelihood function



Likelihood per data point

Stochastic simulation

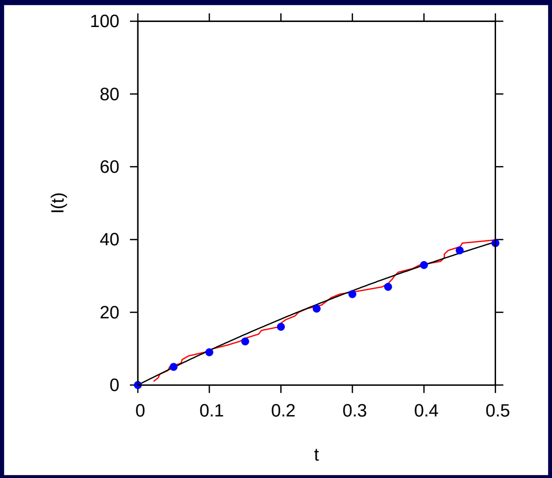
linear infection model as stochastic process

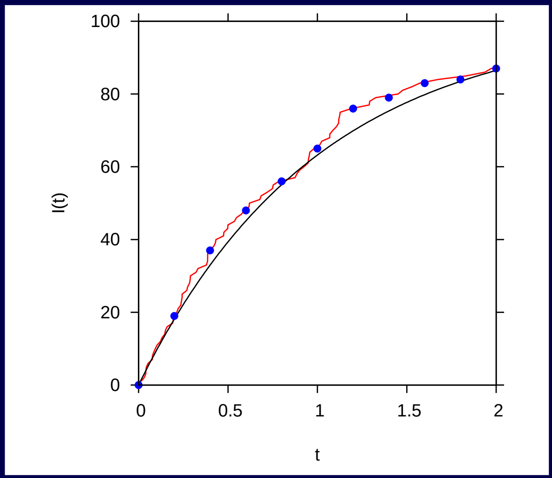
$$S + I^* \xrightarrow{\beta} I + I^*$$

for variable I and $S = N - I$ => probab. $p(I, t)$

$$\frac{d}{dt}p(I,t) = \beta^*(N-(I-1))p(I-1,t) - \beta^*(N-I)p(I,t)$$

simulated by e.g. Gillespie algorithm

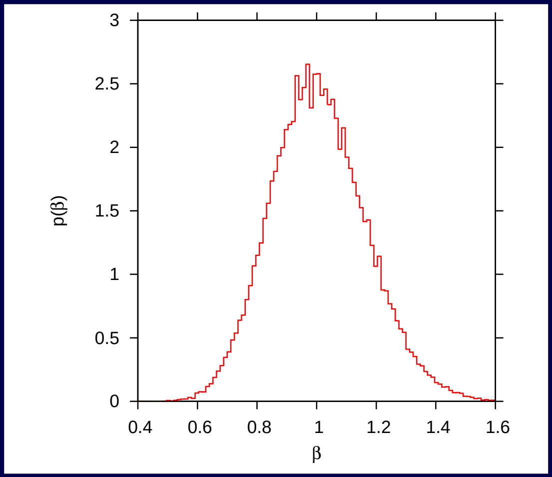


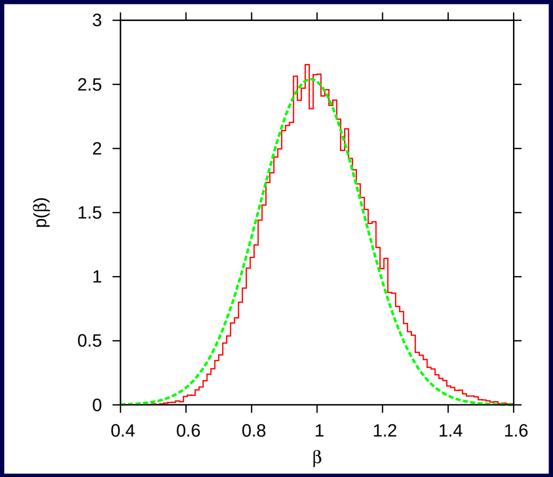


take data points for parameter estimation
via likelihood maximization

Experiment: many realizations

simulate many realizations of stochastic process
take histogram of best estimates

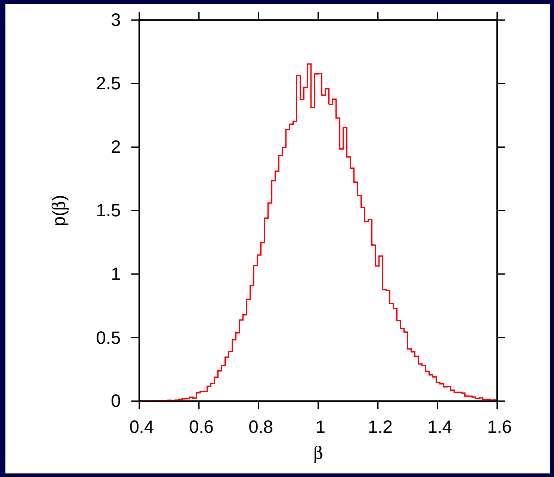


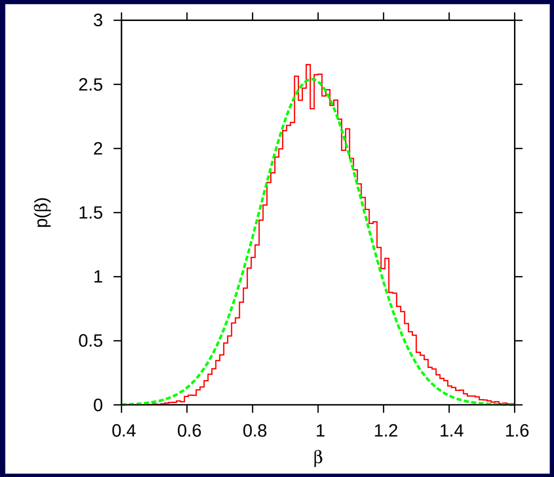


Gaussian approximation compares relatively well

Experiment: many realizations

simulate many realizations of stochastic process
take histogram of best estimates





Gaussian approximation compares relatively well
but can be improved :-)

Bayesian approach to improve confidence intervals

as before data vector $\underline{I} = (I_0, I_1, ...I_n)$ consider joint probability of data and parameter

$$p(\beta, \underline{I}) = p(\underline{I}, \beta)$$

gives via conditional probabilities $p(\beta|\underline{I}) \cdot p(\underline{I}) = p(\underline{I}|\beta) \cdot p(\beta)$ the probability of the parameter given the data $p(\beta|\underline{I})$, the Bayesian posterior

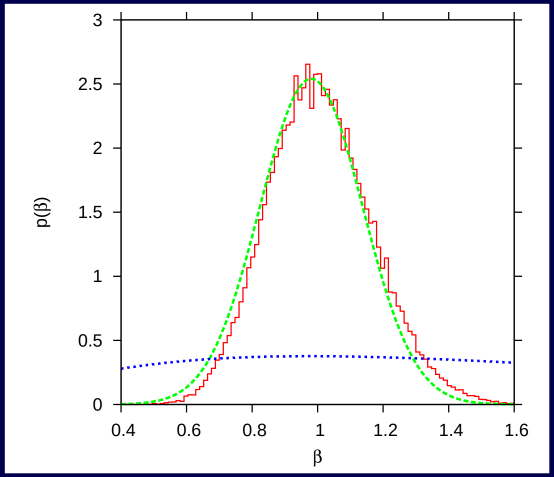
$$p(\beta|\underline{I}) = \frac{p(\underline{I}|\beta)}{p(\underline{I})}\, p(\beta)$$

again with previously used likelihood function $p(\underline{I}|\beta)$

Bayesian approach to improve confidence intervals

$$p(\beta|\underline{I}) = \frac{p(\underline{I}|\beta)}{p(\underline{I})}\, p(\beta)$$

with previously used likelihood function $p(\underline{I}|\beta)$



conjugate prior is a beta-distribution with parameters a and b

Bayesian approach to improve confidence intervals

$$p(\beta|\underline{I}) = \frac{p(\underline{I}|\beta)}{p(\underline{I})}\, p(\beta)$$

with previously used likelihood function $p(\underline{I}|\beta) = L(\beta)$

$$p(\underline{I}|\beta) = \prod_{\nu=0}^{n-1} \binom{N - I_\nu}{I_{\nu+1} - I_\nu} \left(e^{-\beta\cdot\Delta t}\right)^{N - I_{\nu+1}} \left(1 - e^{-\beta\cdot\Delta t}\right)^{I_{\nu+1} - I_\nu}$$

or with abreviation $\theta := 1 - e^{-\beta\cdot\Delta t}$

$$p(\underline{I}|\theta) = \left(\prod_{\nu=0}^{n-1} \binom{N - I_\nu}{I_{\nu+1} - I_\nu}\right) (1-\theta)^{\sum_{\nu=0}^{n-1}(N - I_{\nu+1})} \theta^{\sum_{\nu=0}^{n-1}(I_{\nu+1} - I_\nu)}$$

has the functional form

$$p(\underline{I}|\theta) = k_1\, \theta^{k_2}\, (1-\theta)^{k_2}$$

Bayesian approach to improve confidence intervals

$$p(\beta|\underline{I}) = \frac{p(\underline{I}|\beta)}{p(\underline{I})}\, p(\beta)$$

with previously used likelihood function $p(\underline{I}|\beta) = L(\beta)$

$$p(\underline{I}|\beta) = \prod_{\nu=0}^{n-1} \begin{pmatrix} N - I_\nu \\ I_{\nu+1} - I_\nu \end{pmatrix} \left(e^{-\beta\cdot\Delta t}\right)^{N-I_{\nu+1}} \left(1 - e^{-\beta\cdot\Delta t}\right)^{I_{\nu+1}-I_\nu}$$

or with abreviation $\theta := 1 - e^{-\beta\cdot\Delta t}$ has the functional form

$$p(\underline{I}|\theta) = k_1\, \theta^{k_2}\, (1-\theta)^{k_2}$$

and with beta-function $B(a,b) := \int_0^1 x^{a-1}(1-x)^{b-1}\, dx$ and $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ the conjugate prior is

$$p(\theta) = \theta^{a-1}\,(1-\theta)^{b-1}\,/B(a,b)$$

Bayesian approach to improve confidence intervals

$$p(\theta|\underline{I}) = \frac{p(\underline{I}|\theta)}{p(\underline{I})}\, p(\theta)$$

with above given likelihood function $p(\underline{I}|\theta)$ and prior $p(\theta)$ we only need still the normalization constant

$$p(\underline{I}) = \int_0^1 p(\underline{I}|\theta)p(\theta)\, d\theta$$

and the transformation to the original variable β

$$p(\beta|\underline{I}) = p(\theta|\underline{I})\frac{d\theta}{d\beta}$$

to calculate the desired posterior $p(\beta|\underline{I})$, i.e. the probability for the parameter given the data

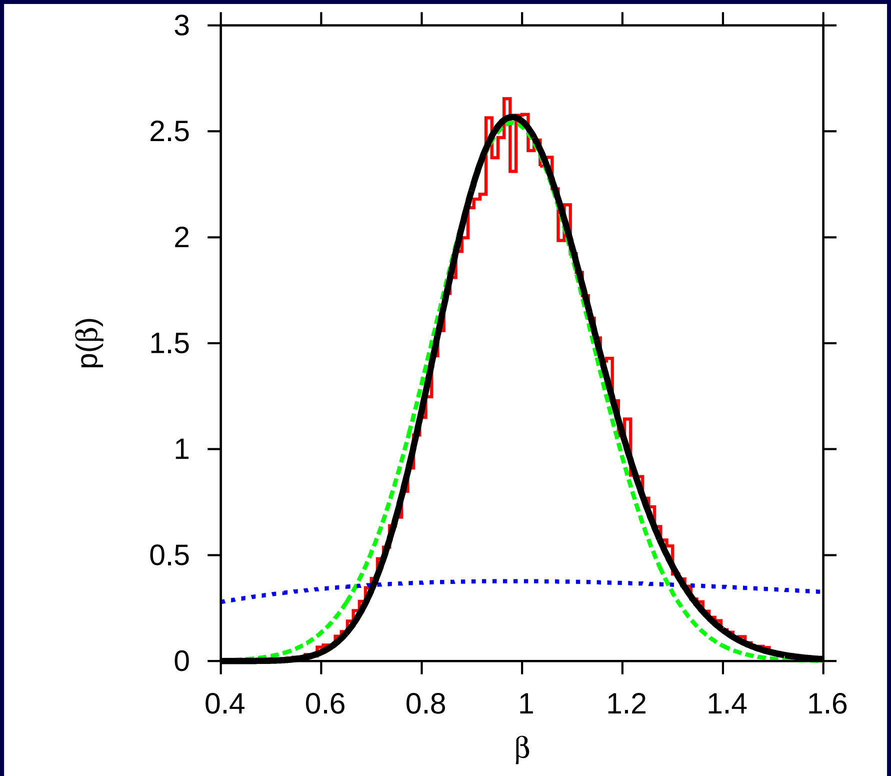
Bayesian posterior

after some calculation the posterior is

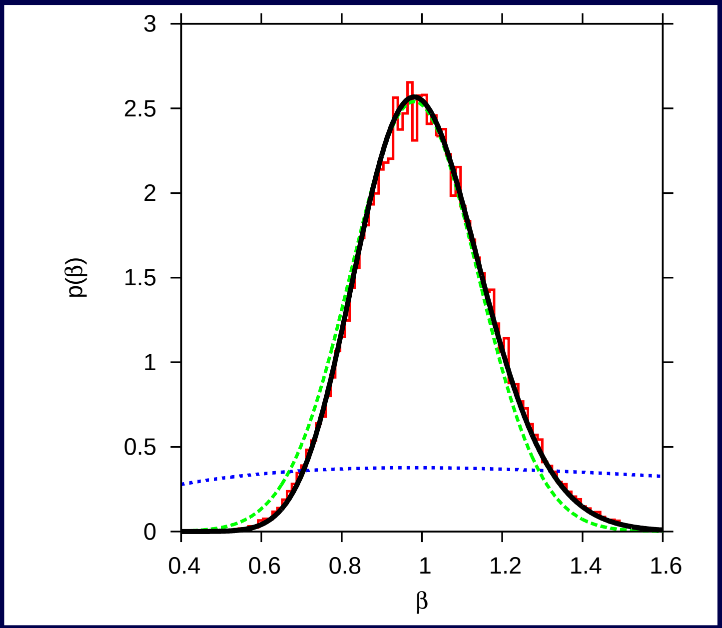
$$p(\beta|\underline{I}) = \frac{\Gamma(a+b+\sum_{\nu=0}^{n-1}(N-I_\nu))}{\Gamma(a+\sum_{\nu=0}^{n-1}(I_{\nu+1}-I_\nu))\,\Gamma(b+\sum_{\nu=0}^{n-1}(N-I_{\nu+1}))}$$

$$\cdot\left(1-e^{-\beta\Delta t}\right)^{a+\sum_{\nu=0}^{n-1}(I_{\nu+1}-I_\nu)-1}\left(e^{-\beta\Delta t}\right)^{b+\sum_{\nu=0}^{n-1}(N-I_{\nu+1})-1}$$

$$\cdot e^{-\beta\Delta t}\cdot\Delta t$$

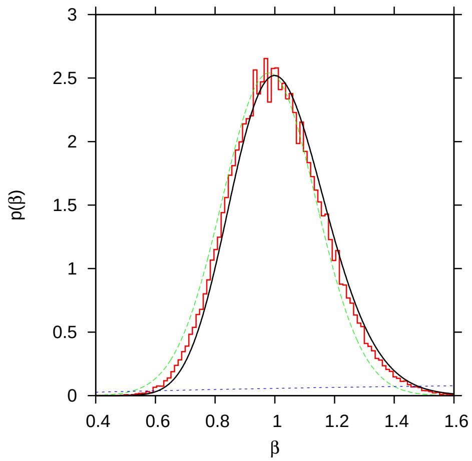


Bayesian posterior

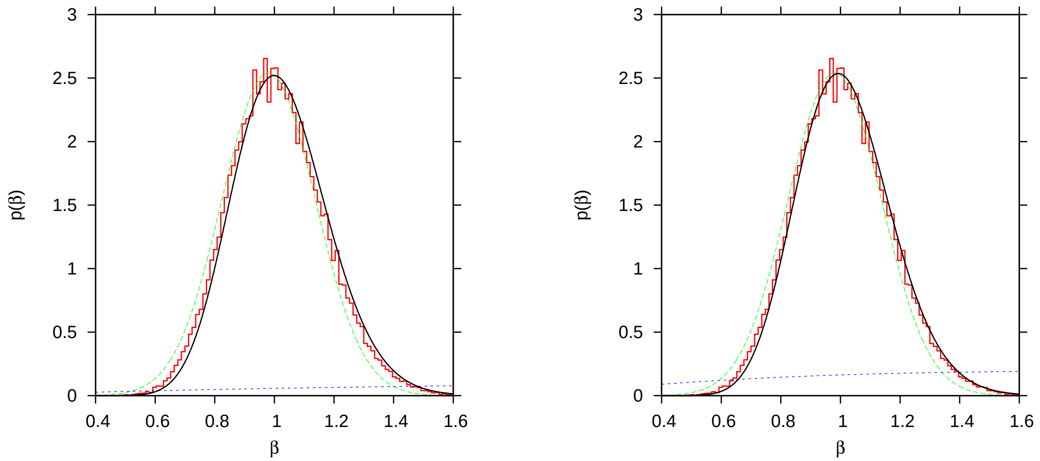


when using soft prior and with good data, the likelihood function carries most of the information

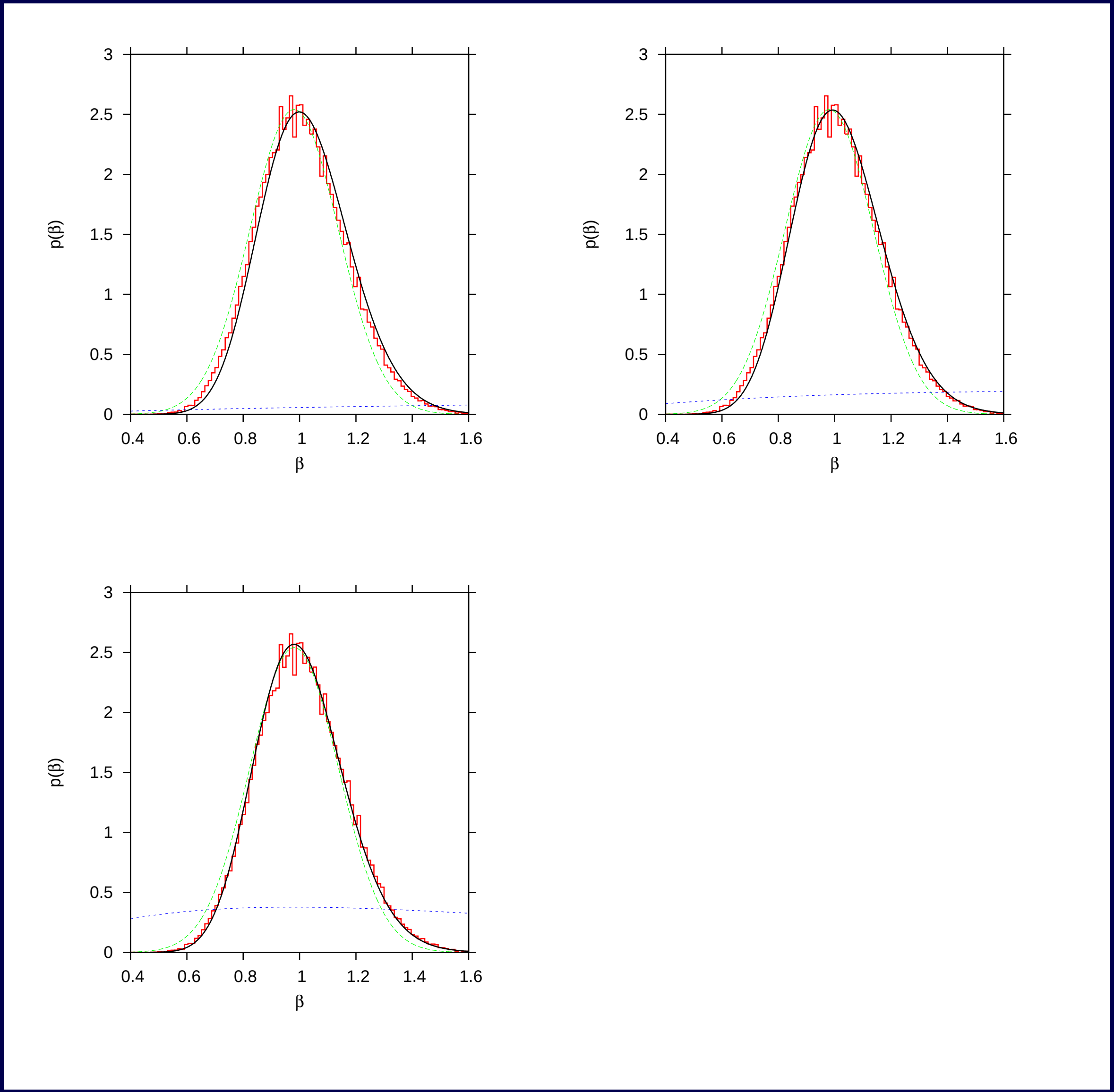
Changing Bayesian prior



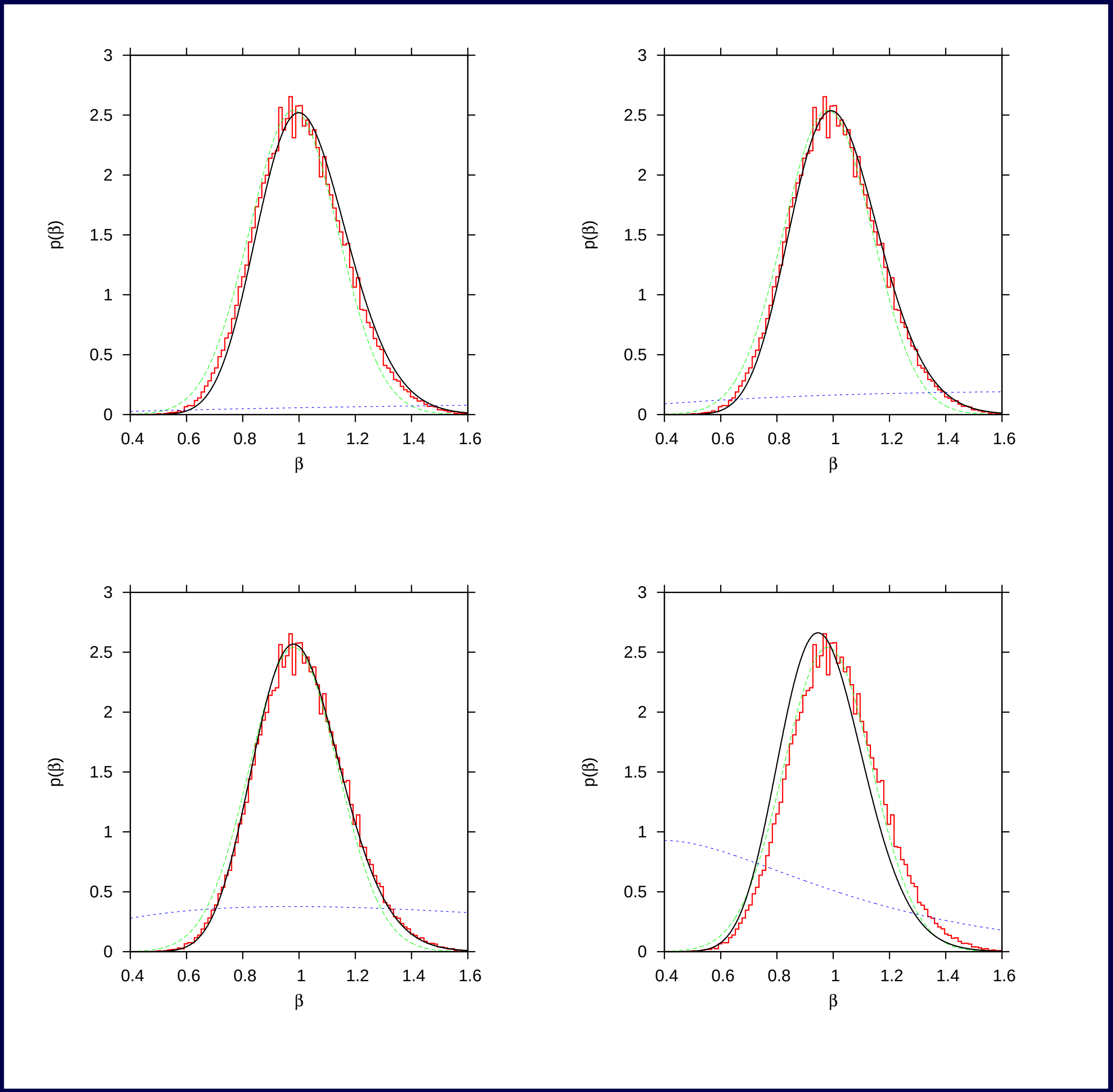
Changing Bayesian prior



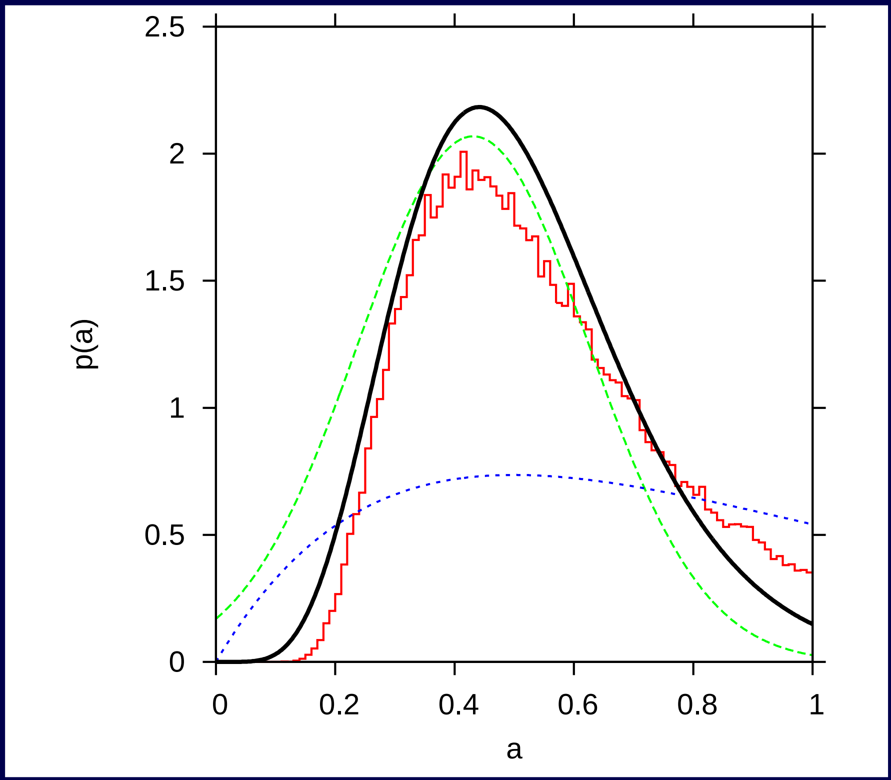
Changing Bayesian prior



Changing Bayesian prior



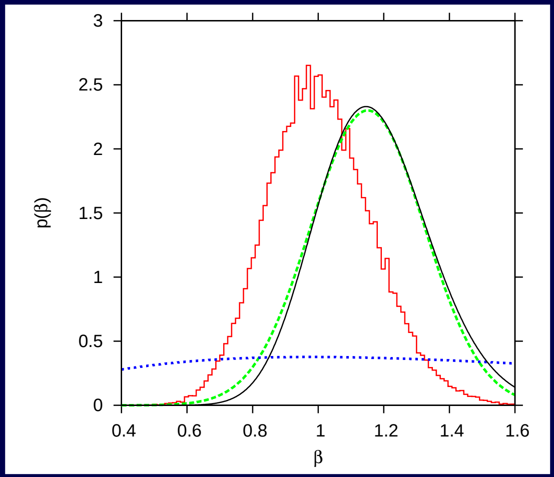
Another example: estimating exponential distribution



the effects are even more pronounced

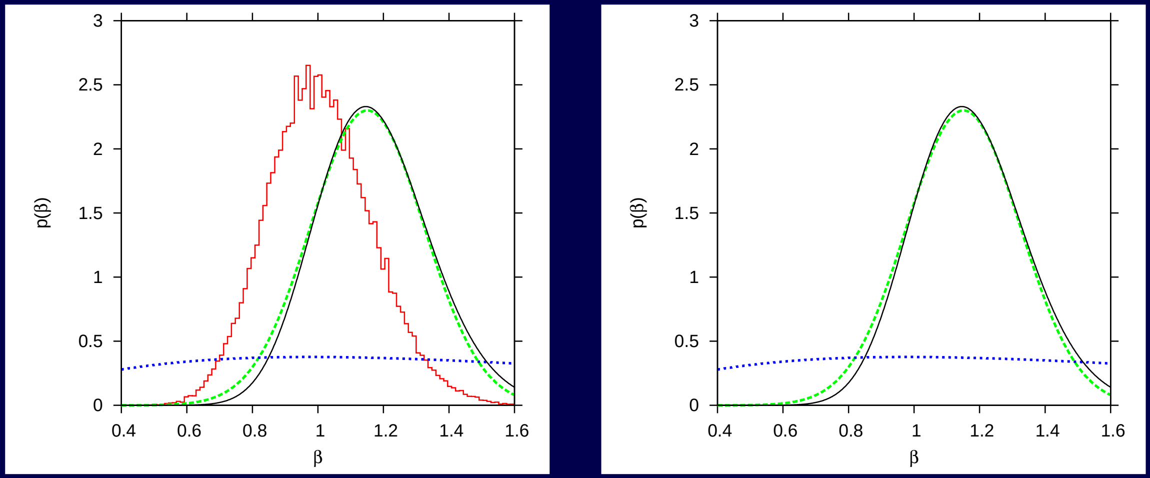
Empirical situation

observed realisation might be "atypical"



Empirical situation

observed realisation might be "atypical"



and we might never know how atypical our data are

Model comparison: Linear Inf. model versus Poisson model

Linear Infection model

$$S + I^* \xrightarrow{\beta} I + I^*$$

with dynamics for the probab. $p(I,t)$

$$\frac{d}{dt}p(I,t) = \beta^*(N-(I-1))p(I-1,t) - \beta^*(N-I)p(I,t)$$

can be further simplified. For large N

$$\beta^*(N-(I-1)) \approx \beta^*(N-I) \approx \beta^* N =: \lambda$$

Master equation reduced to

$$\frac{d}{dt}p(I,t) = \lambda \cdot p(I-1,t) - \lambda \cdot p(I,t)$$

Solution (with initial conditions $p(I,t_0) = \delta_{I,I_0}$**)**

$$p(I,t) = \frac{(\lambda \Delta t)^{I-I_0}}{(I-I_0)!}\, e^{-\lambda \Delta t}$$

Likelihood function from data $(I_0, I_1, \ldots I_n)$

Joint probability of data points for Markovian processes

$$p(I_n, t_n, I_{n-1}, t_{n-1}, \ldots, I_1, t_1, I_0, t_0) = \prod_{\nu=0}^{n-1} p(I_{\nu+1}, t_{\nu+1} | I_\nu, t_\nu) \cdot p(I_0, t_0)$$

Insert transition probabilities

$$p(I, t | I_0, t_0) = \frac{(\lambda \Delta t)^{I - I_0}}{(I - I_0)!} e^{-\lambda \Delta t}$$

Likelihood function

$$L(\lambda) = \prod_{\nu=0}^{n-1} \frac{(\lambda \Delta t)^{I_{\nu+1} - I_\nu}}{(I_{\nu+1} - I_\nu)!} e^{-\lambda \Delta t}$$

Poisson model, Bayesian

Likelihood Function

$$L(\lambda) = \prod_{\nu=0}^{n-1} \frac{(\lambda \Delta t)^{I_{\nu+1}-I_\nu}}{(I_{\nu+1} - I_\nu)!}\, e^{-\lambda \Delta t}$$

with new parameter

$$\theta := \lambda \Delta t$$

equal to

$$L(\theta) = \left(\prod_{\nu=0}^{n-1} \frac{1}{(I_{\nu+1} - I_\nu)!} \right) \theta^{\sum_{\nu=0}^{n-1}(I_{\nu+1}-I_\nu)} e^{-\theta n}$$

Poisson model, Bayesian

With constants

$$k_2 := \sum_{\nu=0}^{n-1} (I_{\nu+1} - I_\nu)$$

and

$$k_5 := \prod_{\nu=0}^{n-1} \frac{1}{(I_{\nu+1} - I_\nu)!}$$

we rewrite L

$$L(\theta) = k_5 \theta^{k_2} e^{-\theta n} =: p(\underline{I}|\theta)$$

conjugate prior

$$p_{a_2,b_2}(\theta) = \frac{b_2^{a_2}}{\Gamma(a_2)}\, \theta^{a_2-1} e^{-b_2\theta} := p(\theta)$$

Poisson model, Bayesian

In our case

$$p(\underline{I}|\theta) \cdot p(\theta) = k_5 \, \frac{b_2^{a_2}}{\Gamma(a_2)} \, \theta^{a_2+k_2-1} e^{-(b_2+n)\theta} = k_6 \, \theta^{a_2+k_2-1} e^{-(b_2+n)\theta}$$

Normalizing constant

$$p(\underline{I}) = \int_0^\infty p(\underline{I}|\theta) p(\theta) d\theta = k_5 \, \frac{b_2^{a_2}}{\Gamma(a_2)} \cdot \frac{\Gamma(a_2+k_2)}{(b_2+n)^{a_2+k_2}}$$

$$p(\theta|\underline{I}) = \frac{p(\underline{I}|\theta) p(\theta)}{p(\underline{I})} = \frac{(b_2+n)^{a_2+k_2}}{\Gamma(a_2+k_2)} \, \theta^{(a_2+k_2-1)} e^{-(b_2+n)\theta}$$

Finally we get

$$p(\lambda|\underline{I}) = p(\theta|\underline{I}) \frac{d\theta}{d\lambda} = \frac{(b_2+n)^{a_2+k_2}}{\Gamma(a_2+k_2)} \, (\lambda \Delta t)^{(a_2+k_2-1)} e^{-\lambda \Delta t (b_2+n)} \Delta t$$

Model Comparison

Consider, for a given data set $\underline{I}$, two models: M_1 with parameter β and M_2 with parameter λ

$$\frac{p(M_1|\underline{I})}{p(M_2|\underline{I})} = \frac{\frac{p(\underline{I}|M_1)}{p(\underline{I})} \cdot p(M_1)}{\frac{p(\underline{I}|M_2)}{p(\underline{I})} \cdot p(M_2)} = \frac{p(\underline{I}|M_1)}{p(\underline{I}|M_2)} \cdot \frac{p(M_1)}{p(M_2)}$$

Model Comparison

Consider, for a given data set $\underline{I}$, two models: M_1 with parameter β and M_2 with parameter λ

$$\frac{p(M_1|\underline{I})}{p(M_2|\underline{I})} = \frac{\frac{p(\underline{I}|M_1)}{p(\underline{I})} \cdot p(M_1)}{\frac{p(\underline{I}|M_2)}{p(\underline{I})} \cdot p(M_2)} = \frac{p(\underline{I}|M_1)}{p(\underline{I}|M_2)} \cdot \frac{p(M_1)}{p(M_2)}$$

Assuming $p(M_1) = p(M_2) = \frac{1}{2}$ we obtain the Bayes factor k via

$$\frac{p(M_1|\underline{I})}{p(M_2|\underline{I})} = \frac{p(\underline{I}|M_1)}{p(\underline{I}|M_2)} := k$$

Model Comparison

Consider, for a given data set $\underline{I}$, two models: M_1 with parameter β and M_2 with parameter λ

$$\frac{p(M_1|\underline{I})}{p(M_2|\underline{I})} = \frac{\frac{p(\underline{I}|M_1)}{p(\underline{I})} \cdot p(M_1)}{\frac{p(\underline{I}|M_2)}{p(\underline{I})} \cdot p(M_2)} = \frac{p(\underline{I}|M_1)}{p(\underline{I}|M_2)} \cdot \frac{p(M_1)}{p(M_2)}$$

Assuming $p(M_1) = p(M_2) = \frac{1}{2}$ we obtain the Bayes factor k via

$$\frac{p(M_1|\underline{I})}{p(M_2|\underline{I})} = \frac{p(\underline{I}|M_1)}{p(\underline{I}|M_2)} := k$$

and with $p(\underline{I}|M_1) := \int p(\underline{I}|\beta, M_1)p(\beta, M_1)\ d\beta$

$$p(\underline{I}|M_1) = k_1 \cdot \frac{\Gamma(a_1 + b_1)}{\Gamma(a_1)\Gamma(a_2)} \cdot \frac{\Gamma(a_1 + k_2)\Gamma(b_1 + k_3)}{\Gamma(a_1 + k_2 + b_1 + k_3)}$$

and $p(\underline{I}|M_2) := \int p(\underline{I}|\lambda, M_2)p(\lambda, M_2)\ d\lambda$

$$p(\underline{I}|M_2) = k_5 \cdot \frac{b_2^{a_2}}{\Gamma(a_2)} \cdot \frac{\Gamma(a_2 + k_2)}{(b_2 + n)^{a_2 + k_2}}$$

Model Comparison

we get for Bayes factor k

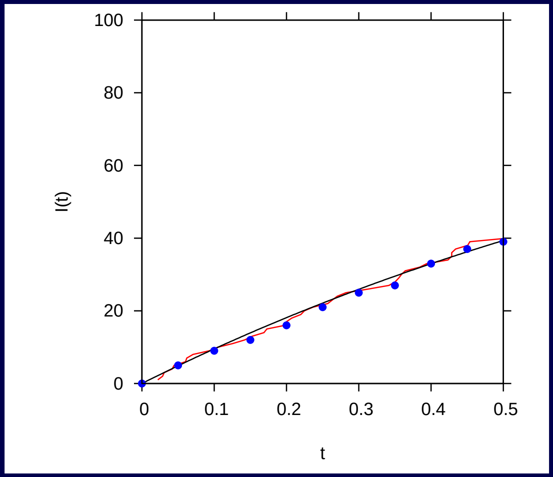
$$k = \frac{k_1 \cdot \Gamma(a_1 + b_1) \cdot \Gamma(a_1 + k_2) \cdot \Gamma(b_1 + k_3) \cdot \Gamma(a_2) \cdot (b_2 + n)^{a_2 + k_2}}{k_5 \cdot \Gamma(a_1) \cdot \Gamma(b_1) \cdot \Gamma(a_1 + k_2 + b_1 + k_3) \cdot \Gamma(a_2 + k_2) \cdot b_2^{a_2}}$$

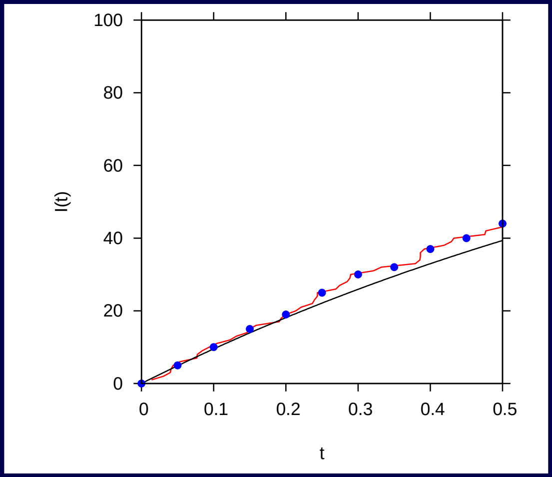
where

$$k_1 := \left(\prod_{\nu=0}^{n-1} \binom{N - I_\nu}{I_{\nu+1} - I_\nu} \right) \qquad k_2 := \sum_{\nu=0}^{n-1} (I_{\nu+1} - I_\nu)$$

$$k_3 := \sum_{\nu=0}^{n-1} (N - I_{\nu+1}) \qquad k_5 := \prod_{\nu=0}^{n-1} \frac{1}{(I_{\nu+1} - I_\nu)!}$$

Numerical examples





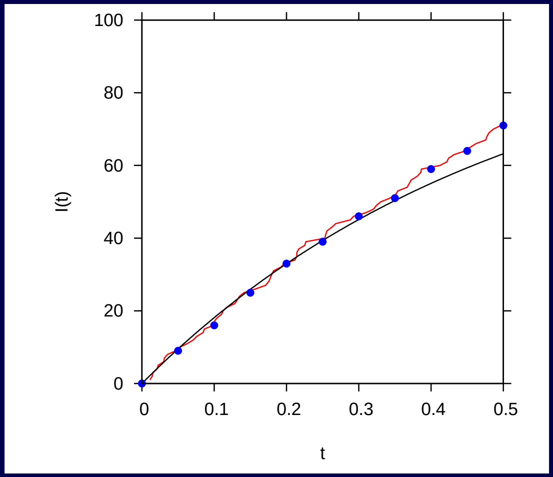
$\beta = 1.0$

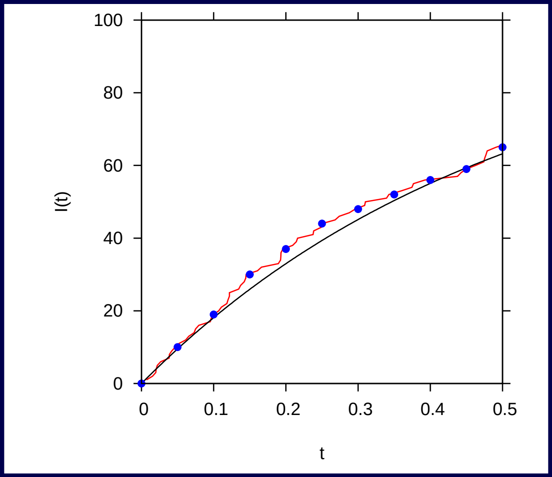
$k = 0.957514$

$\beta = 1.0$

$k = 1.394059$

Numerical examples





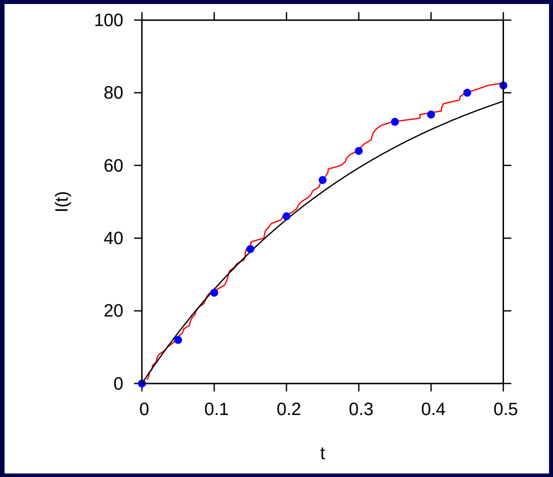
$\beta = 2.0$

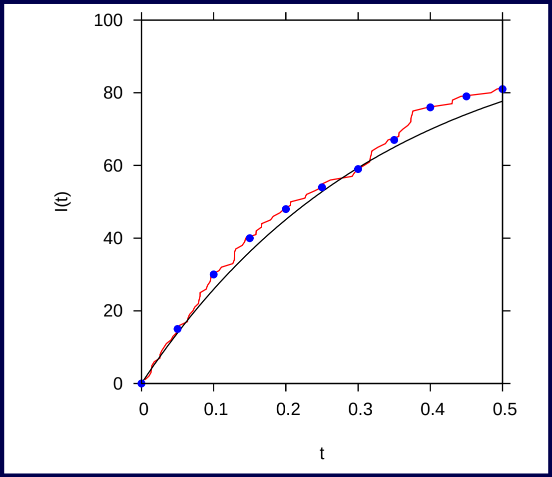
$k = 0.650476$

$\beta = 2.0$

$k = 83.234998$

Numerical examples





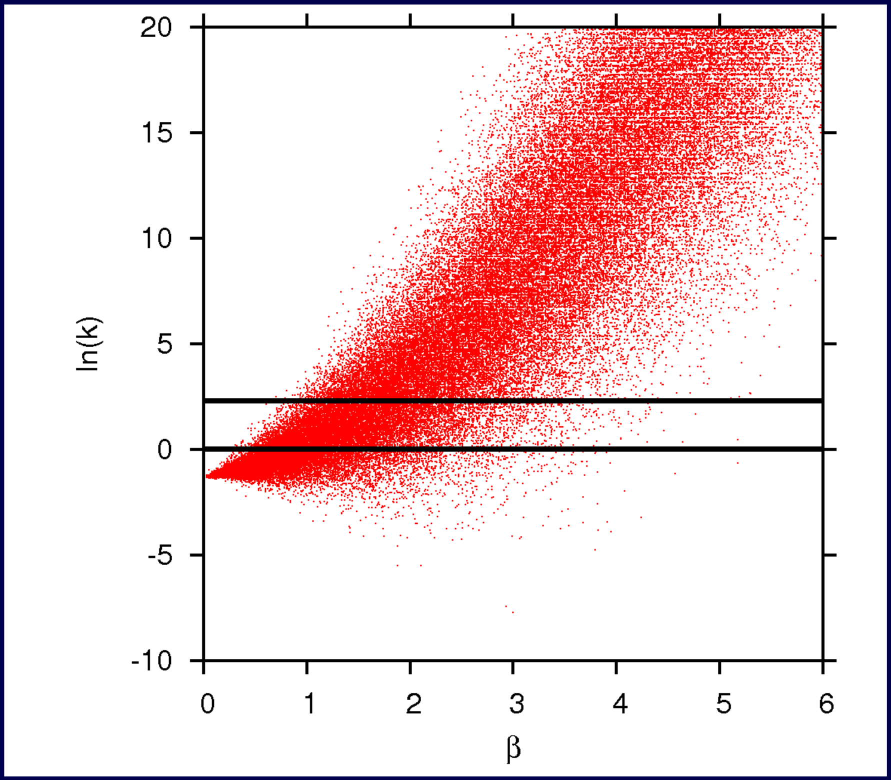
$\beta = 3.0$

$k = 3343.696$

$\beta = 3.0$

$k = 10941.218$

Bayes factor for many realizations over changing parameter



many realizations show more evidence for simplistic model than for the underlying model

(lines for ln(1), no evidence, and ln(10), "strong evidence" for more complex model)

Prediction into future based on data $(I_0, I_1, ...I_n)$

Prediction into future based on data $(I_0, I_1, ...I_n)$

joint probability of data points gives likelihood e.g. for the linear infection model $L(\beta)$

$$p(I_n, t_n, I_{n-1}, t_{n-1}, ..., I_1, t_1, I_0, t_0|\beta) = \prod_{\nu=0}^{n-1} p(I_{\nu+1}, t_{\nu+1}|I_\nu, t_\nu, \beta) \cdot p(I_0, t_0) = L(\beta)$$

and transition probability now into the future $t > t_n = t_{max}$ knowing I_n at t_n was already calculated previously :-)

$$p(I, t|I_n, t_n, \beta) = \binom{N - I_n}{I - I_n} \left(e^{-\beta(t-t_n)}\right)^{N-I} \left(1 - e^{-\beta(t-t_n)}\right)^{I-I_n}$$

is a function of the estimated model parameter β

$$p(I, t|I_n, t_n, \beta) = p(I, t|I_n, t_n, \hat{\beta})$$

with maximum likelihood estimate $\hat{\beta}$ or any best value from the Bayesian posterior $p(\beta|\underline{I})$, maximum, median etc., inserted

Prediction into future based on data $(I_0, I_1, ...I_n)$

then best prediction $\hat{I}_{n+1}$ for next time step t_{n+1} given by maximum of $p(I_{n+1}, t_{n+1}|I_n, t_n, \hat{\beta})$

$$\frac{\partial}{\partial I_{n+1}} ln\, p(I_{n+1}, t_{n+1}|I_n, t_n, \hat{\beta})\Big|_{\hat{I}_{n+1}} = 0$$

using $x! = \Gamma(x+1)$ or for large values Stirling's formula $x! \approx e^{x\, ln(x)}$ and for quantifying the insecurity of this prediction use

$$p(I_{n+1}, t_{n+1}|I_n, t_n, \hat{\beta})$$

but:

Where is the insecurity

of the underlying previous data $(I_0, I_1, ...I_n)$???

Prediction into future based on data $(I_0, I_1, ...I_n)$

from the prediction probability $p(I_{n+1}, t_{n+1}|I_n, t_n, \hat{\beta})$ and the Bayesian posterior $p(\beta|\underline{I})$

$$p(\beta|\underline{I}) = p(\beta|I_1, I_2, ..., I_n)$$

we can construct a joint probability as the product

$$p(I_{n+1}, t_{n+1}|I_n, t_n, \beta) \cdot p(\beta|\underline{I}) = p(I_{n+1}, t_{n+1}, \beta|\underline{I})$$

and integrate over the model parameter β to obtain the prediction based on the underlying data only (and including the parameter insecurity naturally)

$$p(I_{n+1}, t_{n+1}|\underline{I}) = \int_0^\infty p(I_{n+1}, t_{n+1}|I_n, t_n, \beta) \cdot p(\beta|\underline{I})\, d\beta$$

and only in the limiting case of exactly known parameter $p(\beta|\underline{I}) := \delta(\beta - \hat{\beta})$ we obtain the previous result $p(I_{n+1}, t_{n+1}|\underline{I}, \hat{\beta})$.

explicit calculation as homework :-)

Prediction into future based on data $(I_0, I_1, ...I_n)$

prediction probability $p(I_{n+1}, t_{n+1}|\underline{I})$ for the linear infection model (including parameter insecurity)

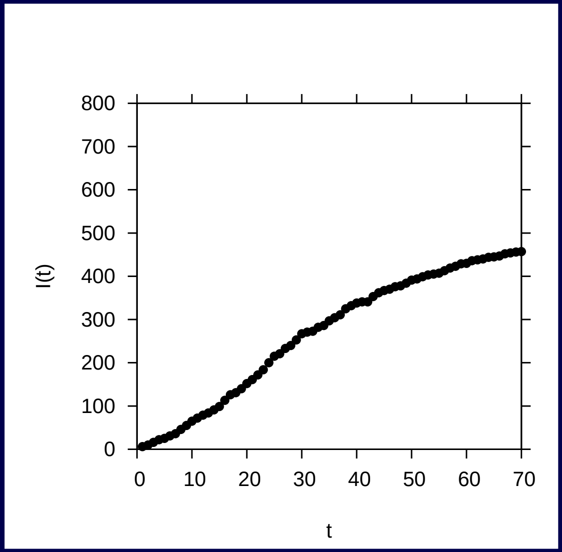
$$p(I_{n+1}, |\underline{I}) = \int_0^\infty p(I_{n+1}, t_{n+1}|I_n, t_n, \beta) \cdot p(\beta|\underline{I})\, d\beta$$

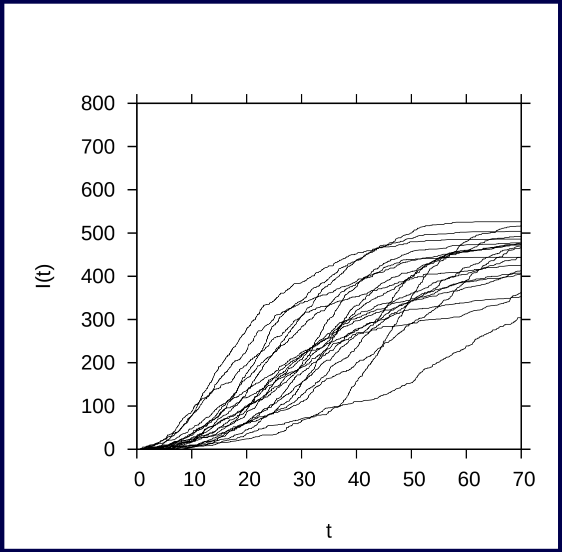
$$= \binom{N - I_n}{I_{n+1} - I_n} \frac{B(a + I_{n+1} - I_n + k_2, b + N - I_{n+1} + k_3)}{B(a + k_2, b + k_3)}$$

again in terms of the beta-function, still depending on prior parameters but not explicitly on model parameter β, with $k_2 := \sum_{\nu=0}^{n-1}(I_{\nu+1} - I_\nu)$ and $k_3 := \sum_{\nu=0}^{n-1}(N - I_{\nu+1})$ only being data dependent

expected to have wide distribution in case of few data

Application to more complex systems: Comparison of data with simulations





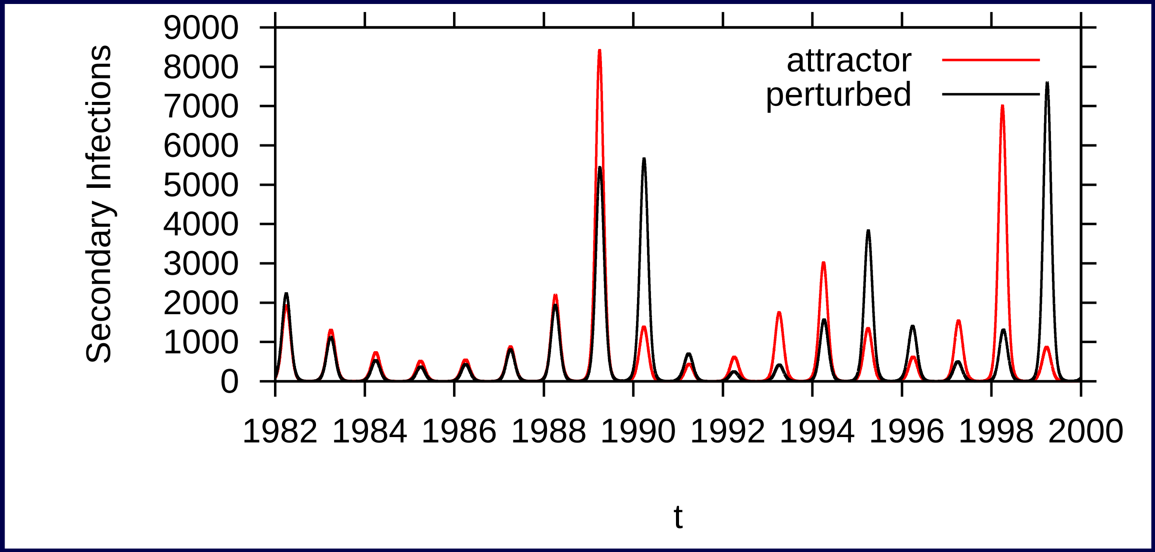
flu data cumulative

simulations of SIR-system

number of simulations in η-ball vicinity to data set gives likelihood of data under this model parameter set

=> estimate of likelihood function (Stollenwerk, Briggs 2000)

Short term predictability, long term unpredictability



simulations with different initial conditions

implications for data analysis: Maximum Likelihood Iterated Filtering (MIF) is choice for such systems (Ionides et al 2006/ Bretó et al. 2009)

Iterated Filtering

algorithmic descritption after Bretó et al. 2009:

MODEL INPUT: $f(\cdot)$, $g(\cdot|\cdot)$, $y_1, ..., y_N$, $t_0, ..., t_N$

ALGORITHMIC PARAMETERS: integers J, L, M; scalars $0 < a < 1$, $b > 0$; vectors $X_I^{(1)}$, $\theta^{(1)}$; positive definite symmetric matrices Σ_I ,Σ_θ.

FOR $m = 1$ to M

  $X_I(t_0, j) \sim N[X_I^{(m)}, a^{m-1}\Sigma_I], \quad j = 1, ..., J$

  $X_F(t_0, j) = X_I(t_0, j)$

  $\theta(t_0, j) \sim N[\theta^{(m)}, ba^{m-1}\Sigma_\theta]$

  $\bar{\theta}(t_0) = \theta^{(m)}$

  FOR $n = 1$ to N

    $X_P(t_n, j) = f(X_F(t_{n-1}, j), t_{n-1}, t_n, \theta(t_{n-1}, j), W)$

    $w(n, j) = g(y_n | X_P(t_n, j), t_n, \theta(t_{n-1}, j))$

    draw $k_1, ..., k_J$ such that $\text{Prob}(k_j = i) = w(n, i) / \sum_\ell w(n, \ell)$

    $X_F(t_n, j) = X_P(t_n, k_j)$

    $X_I(t_n, j) = X_I(t_{n-1}, k_j)$

    $\theta(t_n, j) \sim N[\theta(t_{n-1}, k_j), a^{m-1}(t_n - t_{n-1})\Sigma_\theta]$

    Set $\bar{\theta}_i(t_n)$ to be the sample mean of $\{\theta_i(t_{n-1}, k_j), \ \ j = 1, ..., J\}$

    Set $V_i(t_n)$ to be the sample variance of $\{\theta_i(t_n, j), \ \ j = 1, ..., J\}$

  END FOR

  $\theta_i^{(m+1)} = \theta_i^{(m)} + V_i(t_1) \sum_{n=1}^{N} V_i^{-1}(t_n)(\bar{\theta}_i(t_n) - \bar{\theta}_i(t_{n-1}))$

  Set $X_I^{(m+1)}$ to be the sample mean of $\{X_I(t_L, j), \ \ j = 1, ..., J\}$

END FOR

RETURN

maximum likelihood estimate for parameters, $\hat{\theta} = \theta^{(M+1)}$

maximum likelihood estimate for initial values, $\hat{X}(t_0) = X_I^{(M+1)}$

maximized conditional log likelihood estimates, $\ell_n(\hat{\theta}) = log(\sum_j w(n, j)/J)$

maximized log likelihood estimate, $\ell(\hat{\theta}) = \sum_n \ell_n(\hat{\theta})$

A fresh look at Iterated Filtering
to include dynamic noise appropriately

algorithmic descritption after Bretó et al. 2009:

MODEL INPUT: $f(\cdot)$, $g(\cdot|\cdot)$, $y_1, ..., y_N$, $t_0, ..., t_N$

ALGORITHMIC PARAMETERS: integers J, L, M; scalars $0 < a < 1$, $b > 0$; vectors $X_I^{(1)}$, $\theta^{(1)}$; positive definite symmetric matrices Σ_I ,Σ_θ.

```
FOR m = 1 to M
   X_I(t_0, j) ~ N[X_I^(m), a^(m-1) Σ_I],   j = 1, ..., J
   X_F(t_0, j) = X_I(t_0, j)
   θ(t_0, j) ~ N[θ^(m), b a^(m-1) Σ_θ]
   θ̄(t_0) = θ^(m)
   FOR n = 1 to N
      X_P(t_n, j) = f(X_F(t_{n-1}, j), t_{n-1}, t_n, θ(t_{n-1}, j), W)
      w(n, j) = g(y_n | X_P(t_n, j), t_n, θ(t_{n-1}, j))
      draw k_1, ..., k_J such that Prob(k_j = i) = w(n, i) / Σ_ℓ w(n, ℓ)
      X_F(t_n, j) = X_P(t_n, k_j)
      X_I(t_n, j) = X_I(t_{n-1}, k_j)
      θ(t_n, j) ~ N[θ(t_{n-1}, k_j), a^(m-1) (t_n - t_{n-1}) Σ_θ]
      Set θ̄_i(t_n) to be the sample mean of {θ_i(t_{n-1}, k_j),  j = 1, ..., J}
      Set V_i(t_n) to be the sample variance of {θ_i(t_n, j),  j = 1, ..., J}
   END FOR
   θ_i^(m+1) = θ_i^(m) + V_i(t_1) Σ_{n=1}^{N} V_i^{-1}(t_n) (θ̄_i(t_n) - θ̄_i(t_{n-1}))
   Set X_I^(m+1) to be the sample mean of {X_I(t_L, j),  j = 1, ..., J}
END FOR
```

RETURN

maximum likelihood estimate for parameters, $\hat{\theta} = \theta^{(M+1)}$

maximum likelihood estimate for initial values, $\hat{X}(t_0) = X_I^{(M+1)}$

maximized conditional log likelihood estimates, $\ell_n(\hat{\theta}) = log(\sum_j w(n,j)/J)$

maximized log likelihood estimate, $\ell(\hat{\theta}) = \sum_n \ell_n(\hat{\theta})$

A fresh look at Iterated Filtering
to include dynamic noise appropriately

algorithmic descritption after Bretó et al. 2009:

MODEL INPUT: $f(\cdot)$, $g(\cdot|\cdot)$, $y_1, ..., y_N$, $t_0, ..., t_N$

ALGORITHMIC PARAMETERS: integers J, L, M; scalars $0 < a < 1$, $b > 0$; vectors $X_I^{(1)}$, $\theta^{(1)}$; positive definite symmetric matrices Σ_I ,Σ_θ.

FOR $m = 1$ to M
    $X_I(t_0, j) \sim N[X_I^{(m)}, a^{m-1}\Sigma_I], \quad j = 1, ..., J$
    $X_F(t_0, j) = X_I(t_0, j)$
    $\theta(t_0, j) \sim N[\theta^{(m)}, ba^{m-1}\Sigma_\theta]$
    $\bar{\theta}(t_0) = \theta^{(m)}$
    FOR $n = 1$ to N
        $X_P(t_n, j) = f(X_F(t_{n-1}, j), t_{n-1}, t_n, \theta(t_{n-1}, j), W)$
        $w(n, j) = g(y_n | X_P(t_n, j), t_n, \theta(t_{n-1}, j))$
        draw $k_1, ..., k_J$ such that $\text{Prob}(k_j = i) = w(n, i) / \sum_\ell w(n, \ell)$
        $X_F(t_n, j) = X_P(t_n, k_j)$
        $X_I(t_n, j) = X_I(t_{n-1}, k_j)$
        $\theta(t_n, j) \sim N[\theta(t_{n-1}, k_j), a^{m-1}(t_n - t_{n-1})\Sigma_\theta]$
        Set $\bar{\theta}_i(t_n)$ to be the sample mean of $\{\theta_i(t_{n-1}, k_j), \;\; j = 1, ..., J\}$
        Set $V_i(t_n)$ to be the sample variance of $\{\theta_i(t_n, j), \;\; j = 1, ..., J\}$
    END FOR
    $\theta_i^{(m+1)} = \theta_i^{(m)} + V_i(t_1) \sum_{n=1}^{N} V_i^{-1}(t_n)(\bar{\theta}_i(t_n) - \bar{\theta}_i(t_{n-1}))$
    Set $X_I^{(m+1)}$ to be the sample mean of $\{X_I(t_L, j), \;\; j = 1, ..., J\}$
END FOR

RETURN
maximum likelihood estimate for parameters, $\hat{\theta} = \theta^{(M+1)}$
maximum likelihood estimate for initial values, $\hat{X}(t_0) = X_I^{(M+1)}$
maximized conditional log likelihood estimates, $\ell_n(\hat{\theta}) = log(\sum_j w(n, j)/J)$
maximized log likelihood estimate, $\ell(\hat{\theta}) = \sum_n \ell_n(\hat{\theta})$

A fresh look at Iterated Filtering
to include dynamic noise appropriately

algorithmic descritption after Bretó et al. 2009:

MODEL INPUT: $f(\cdot)$, $g(\cdot|\cdot)$, $y_1, ..., y_N$, $t_0, ..., t_N$

ALGORITHMIC PARAMETERS: integers J, L, M; scalars $0 < a < 1$, $b > 0$; vectors $X_I^{(1)}$, $\theta^{(1)}$; positive definite symmetric matrices Σ_I ,Σ_θ.

1. FOR $m = 1$ to M
2. $X_I(t_0, j) \sim N[X_I^{(m)}, a^{m-1}\Sigma_I], \quad j = 1, ..., J$
3. $X_F(t_0, j) = X_I(t_0, j)$
4. $\theta(t_0, j) \sim N[\theta^{(m)}, ba^{m-1}\Sigma_\theta]$
5. $\bar{\theta}(t_0) = \theta^{(m)}$
6. FOR $n = 1$ to N
7. $X_P(t_n, j) = f(X_F(t_{n-1}, j), t_{n-1}, t_n, \theta(t_{n-1}, j), W)$
8. $w(n, j) = g(y_n|X_P(t_n, j), t_n, \theta(t_{n-1}, j))$
9. draw $k_1, ..., k_J$ such that Prob$(k_j = i) = w(n, i)/\sum_\ell w(n, \ell)$

...

use e.g. η-balls to construct likelihood

Example study for particle filter: SIRS with seasonality and import

stochastic process

$$S + I \xrightarrow{\beta(t)} I + I$$
$$I \xrightarrow{\gamma} R$$
$$R \xrightarrow{\alpha} S$$
$$S \xrightarrow{\varrho} I$$

with seasonal forcing given by

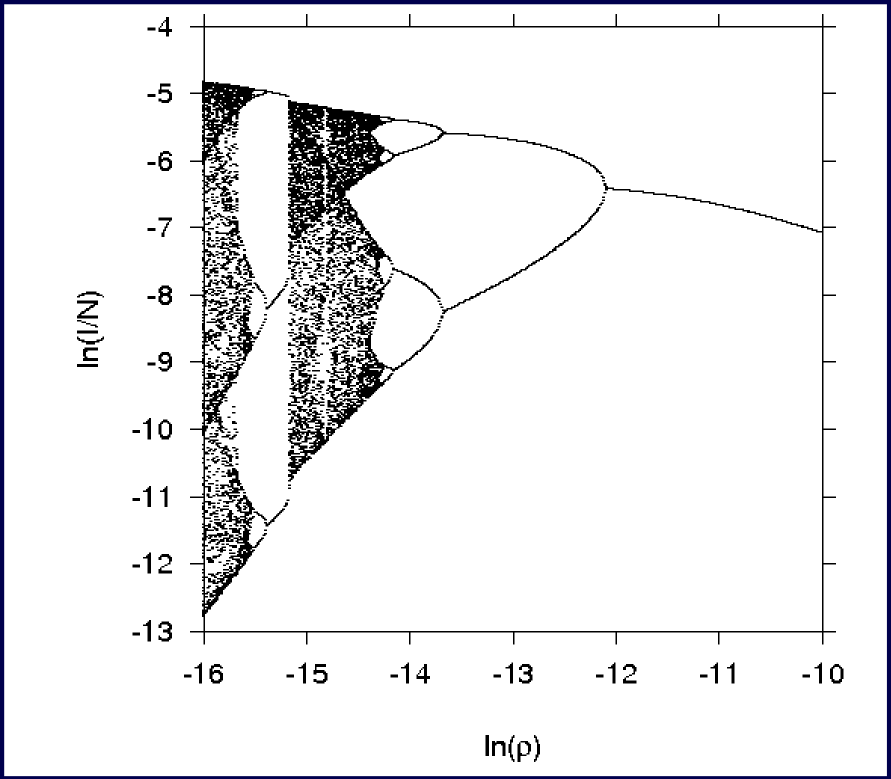
$$\beta(t) = \beta_0 \cdot (1 + \theta \cdot \cos(\omega t))$$

and parameters in the UPCA region, relevant for influenza, $\alpha = \frac{1}{6\text{y}}$, $\gamma = \frac{1}{3\text{d}} = \frac{365}{3}\text{y}^{-1}$, $\beta_0 = 1.5 \cdot \gamma$, **and**

$$\theta = 0.12$$

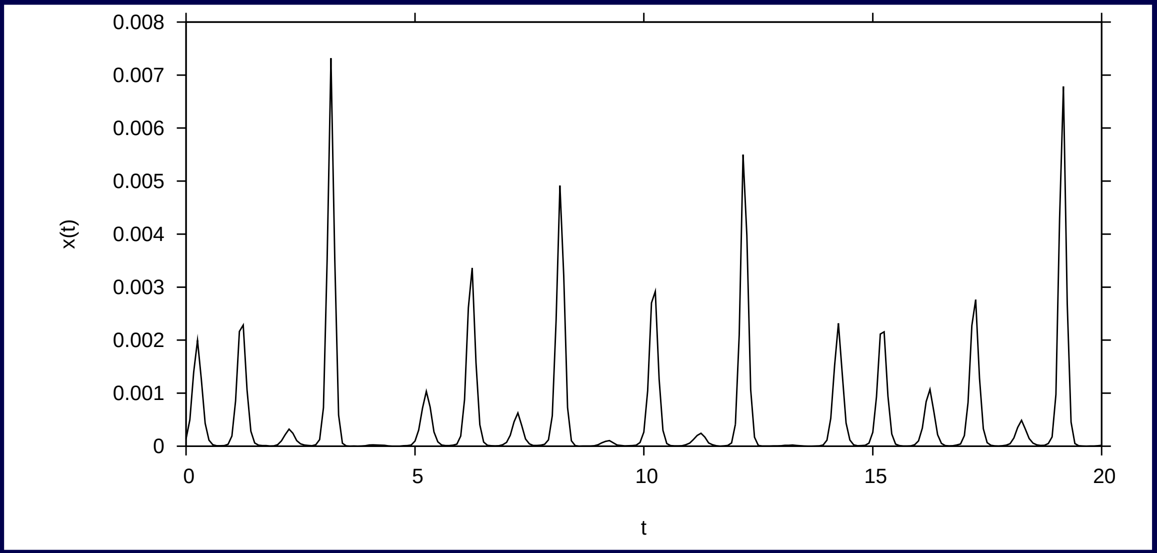
$$\ln(\varrho) = -15$$

Example study for particle filter: SIRS with seasonality and import



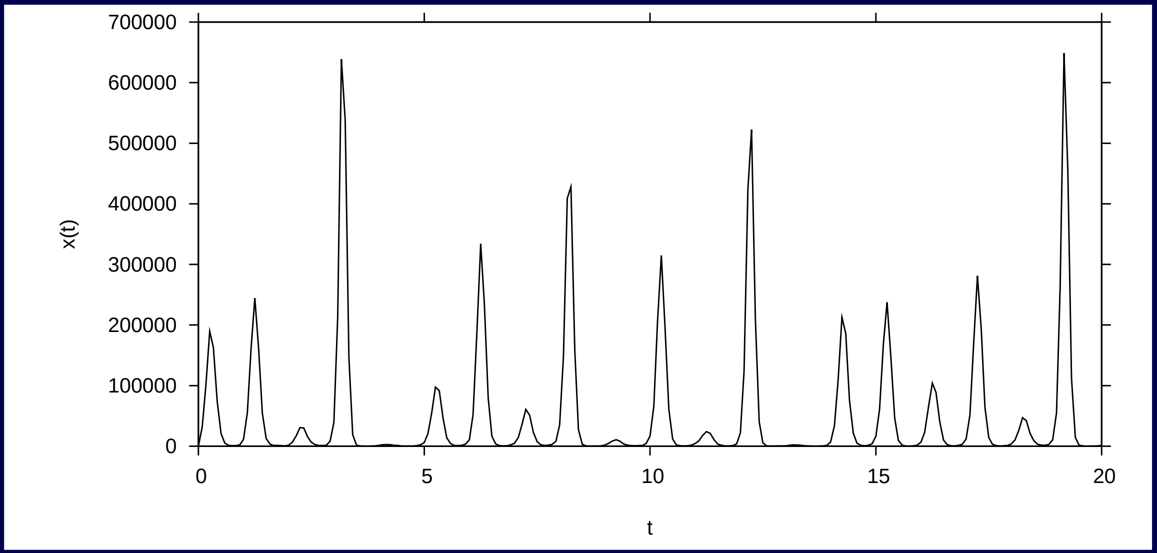
Bifurcation diagram for import $\ln(\varrho)$

Time series generated via Gillespie algorithm



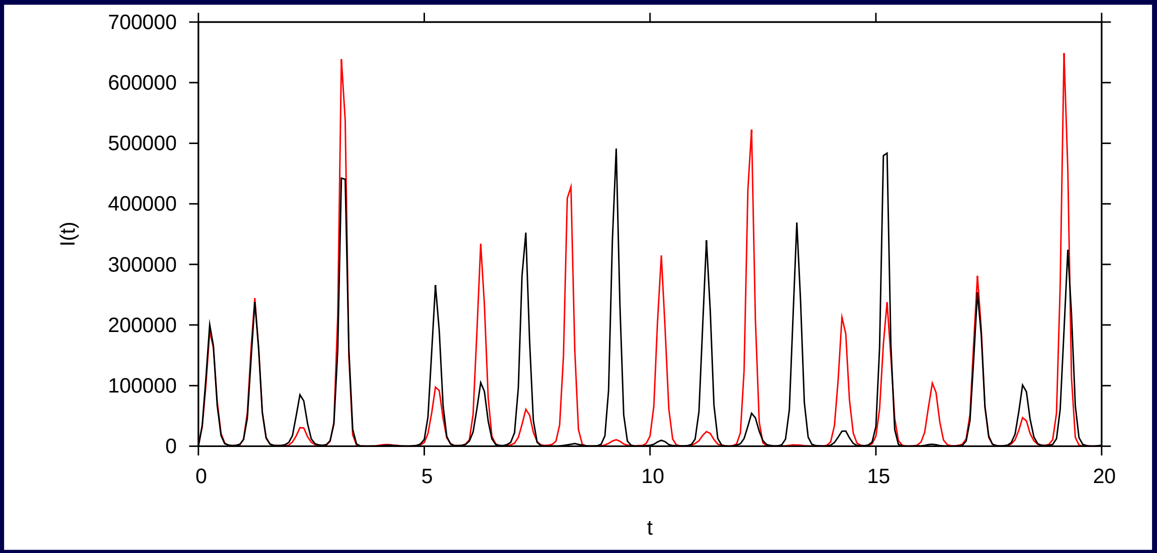
stochastic simulation with exact method, typical run

Time series generated via Gillespie algorithm



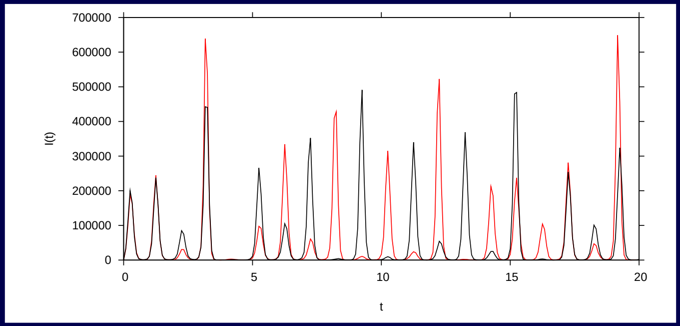
total number of incidences, serves as toy data set
monthly sampled over 20 years

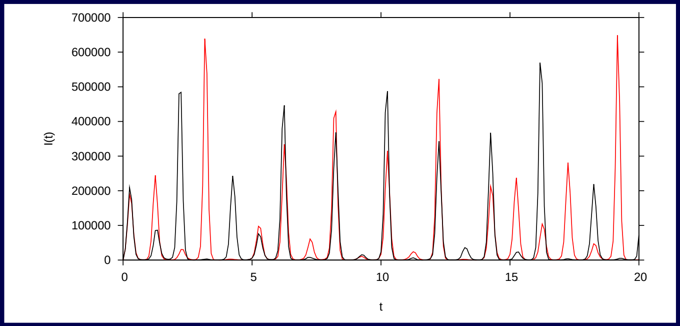
Comparison with Euler-multinomial approximation



Euler multinomial approximation in black, $\Delta t = 0.001\text{d} = (0.001/365)\text{y}$

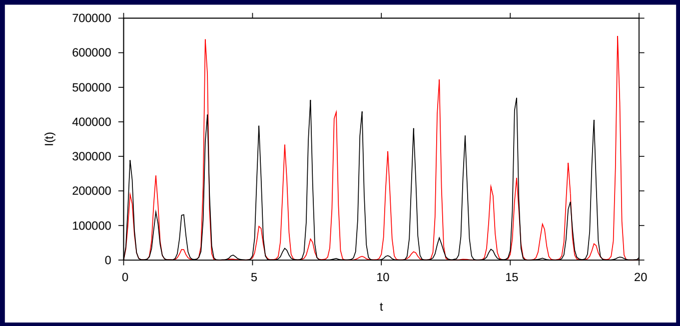
Euler-multinomial approximation: changing Δt

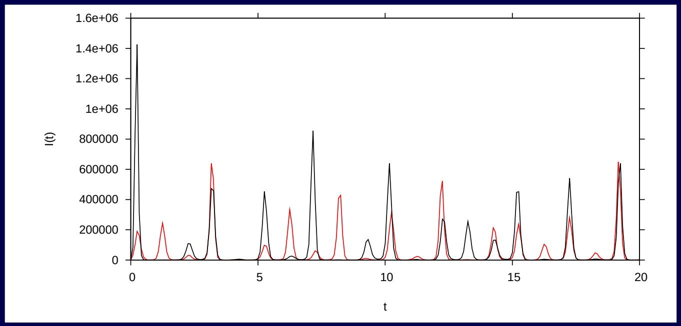




$$\Delta t = 0.001\text{d}$$

$$\Delta t = 0.01\text{d}$$

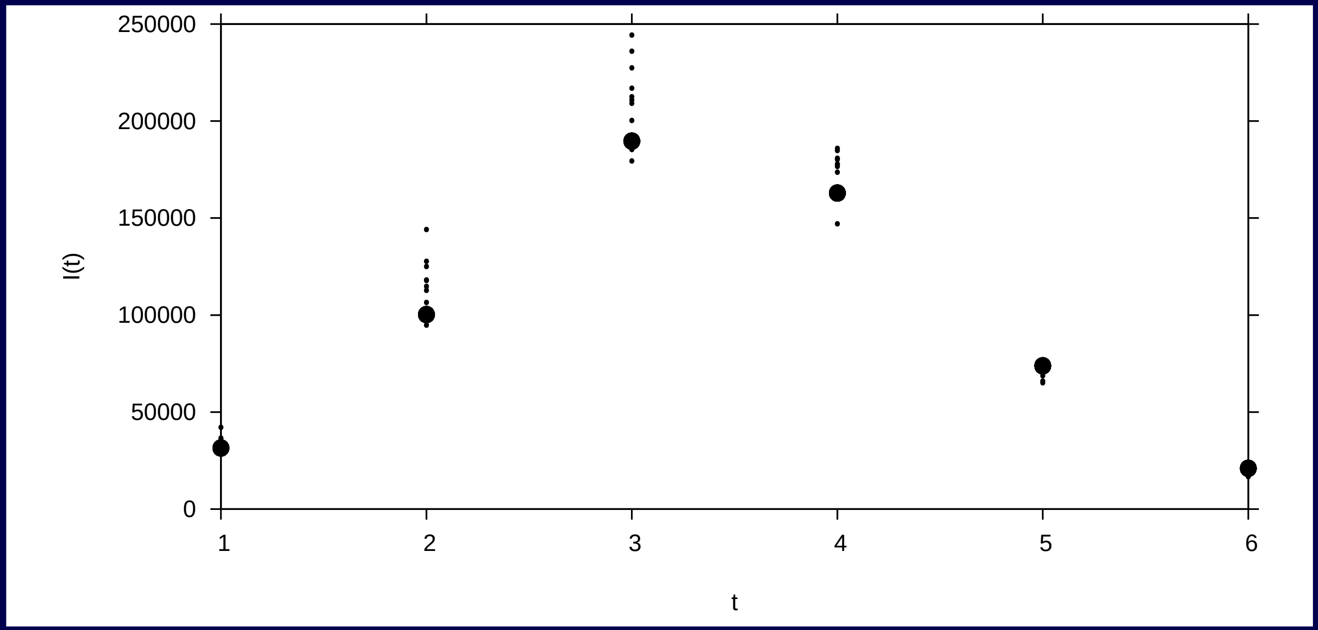




$$\Delta t = 0.1\text{d}$$

$$\Delta t = 1\text{d}$$

Constructing particle filter: particle weights from dynamic noise



cloud of simulations around the first 6 months of data

Euler-multinomial with $\Delta t = 0.01$d

Constructing particle filter: particle weights from dynamic noise

compare the data set $\underline{I}_E = (I_1, I_2, ..., I_E)$, with dimension E (here $E = 6$ months) with K Euler-multinomial simulations $\underline{I}_k(\underline{\theta}_j)$ performed with parameter set $\underline{\theta}_j$ ("particles")

$$\hat{p}(\underline{I}_E|\underline{\theta}_j) = \frac{1}{K}\sum_{k=1}^{K} H\left(\eta - ||\underline{I}_E - \underline{I}_k(\underline{\theta}_j)||_E\right)$$

simulations in η-ball around the data, with $H(x)$ being the Heaviside step function, give estimate of the time-local likelihood function $p(\underline{I}_E|\underline{\theta}_j)$, hence for $K \longrightarrow \infty$ and $\eta \longrightarrow 0$

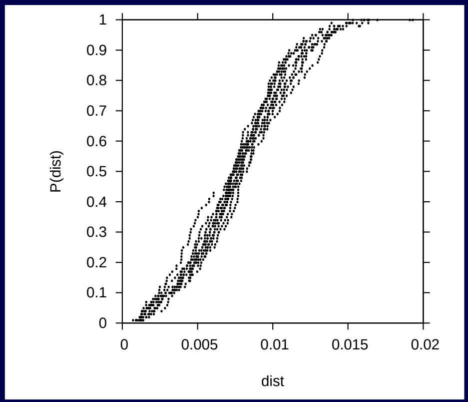
$$w_j := \hat{p}(\underline{I}_E|\underline{\theta}_j) \rightarrow p(\underline{I}_E|\underline{\theta}_j)$$

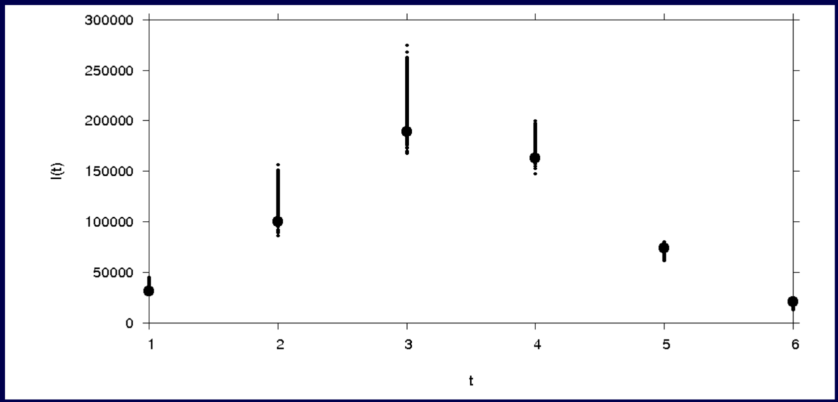
giving the weights of particles w_j for the particle filter

Constructing particle filter: distribution of distances

compare the data set $\underline{I}_E = (I_1, I_2, ..., I_E)$, with dimension E (here $E = 6$ months) with K Euler-multinomial simulations $\underline{I}_k(\underline{\theta}_j)$ performed with parameter set $\underline{\theta}_j$ ("particles")

$$\hat{p}(\underline{I}_E|\underline{\theta}_j) = \frac{1}{K}\sum_{k=1}^{K} H\left(\eta - ||\underline{I}_E - \underline{I}_k(\underline{\theta}_j)||_E\right)$$





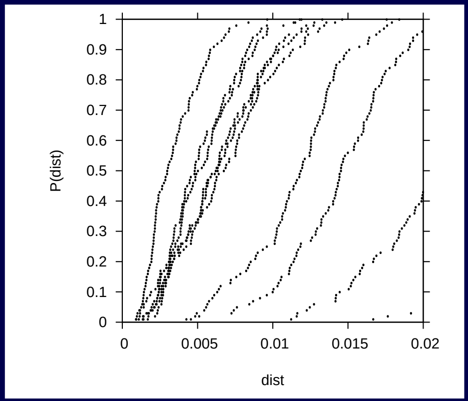
$J = 10$ particles, original parameter set $\underline{\theta}_j$, with $K = 100$ simulations each, distances $dist := ||\underline{I}_E - \underline{I}_k(\underline{\theta}_j)||$

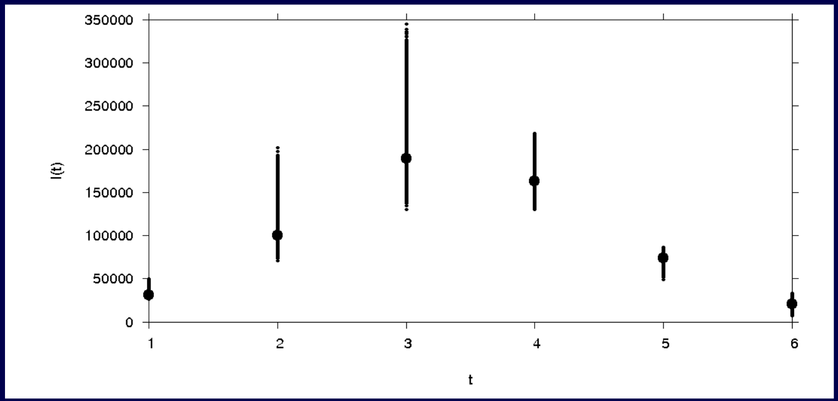
Constructing particle filter: variation of parameters

vary e.g. seasonality θ by 10% with a Gaussian distribution

$$p(\theta) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(\theta-\mu)^2}{2\cdot\sigma^2}}$$

with $\mu = \theta_{orig} = 0.12$, the original value, and $\sigma = \mu/10$ (acts like a Gaussian prior)



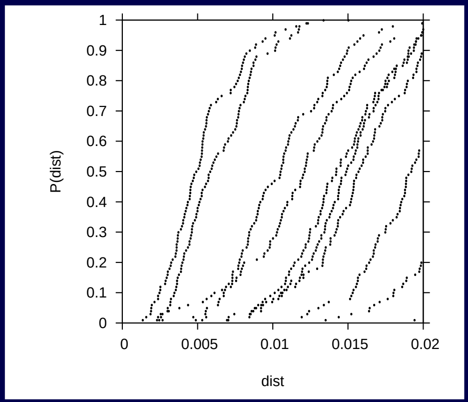


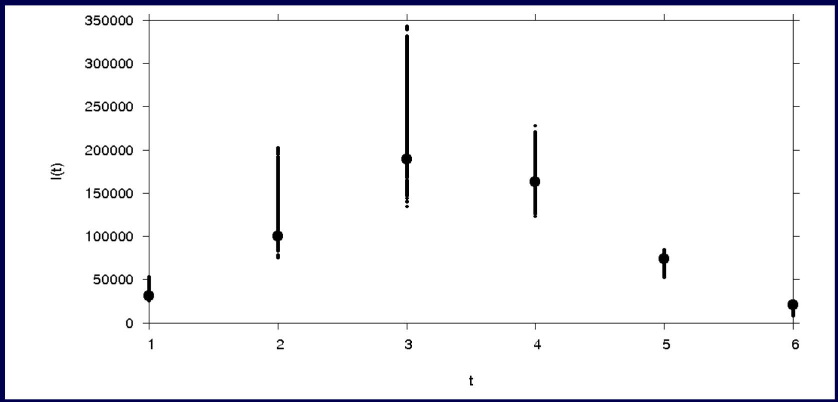
$J = 10$ particles, with $K = 100$ simulations each, most distances are larger, but some even smaller now :-)

Constructing particle filter: variation of several param. and initial cond.

vary seasonality θ, import $ln(\varrho)$ and intital conditions I_0 and R_0, all Gaussian, same order of magnitude

$$\underline{\theta} = (\theta, \varrho, I_0, R_0)$$



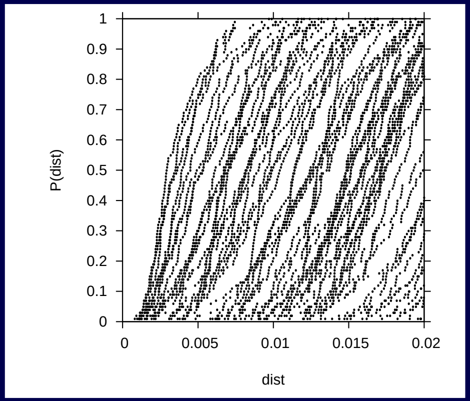


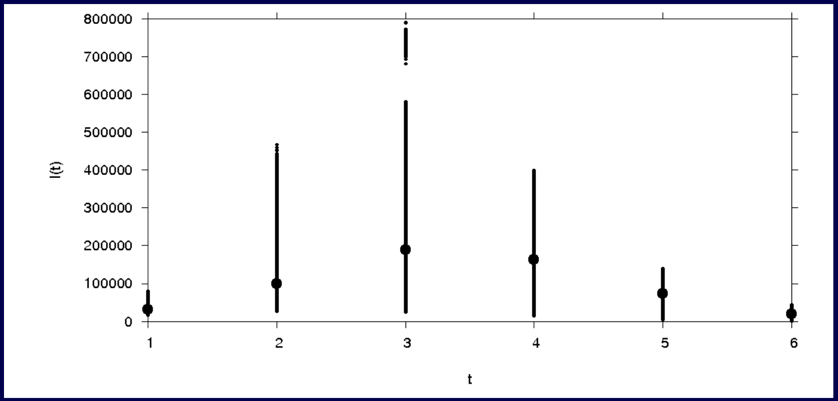
$J = 10$ particles, with $K = 100$ simulations each

Constructing particle filter: calculation of weights of each particle

weight w_j of particle $\underline{\theta}_j$ from estimating time-local likelihood function for dynamic noise

$$w_j := \hat{p}(\underline{I}_E|\underline{\theta}_j) = \frac{1}{K}\sum_{k=1}^{K} H\left(\eta - ||\underline{I}_E - \underline{I}_k(\underline{\theta}_j)||_E\right)$$



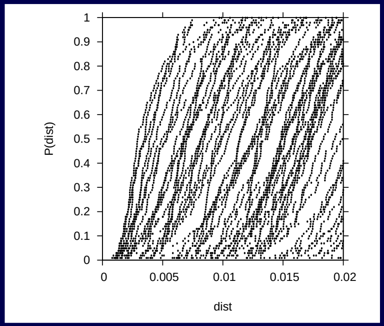


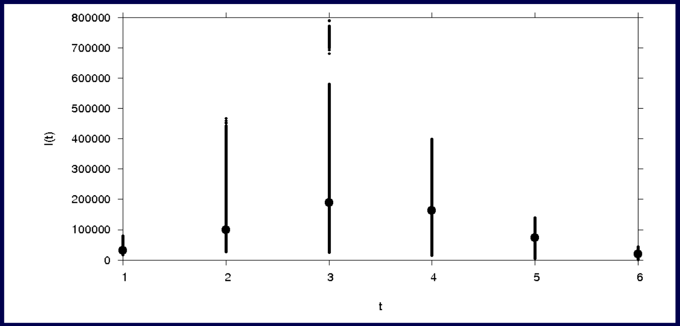
$J = 100$ particles, with $K = 100$ simulations each

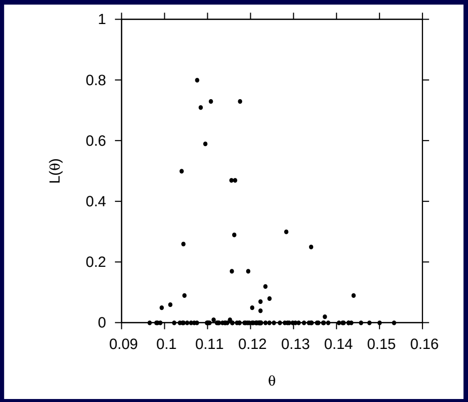
Constructing particle filter: calculation of weights of each particle

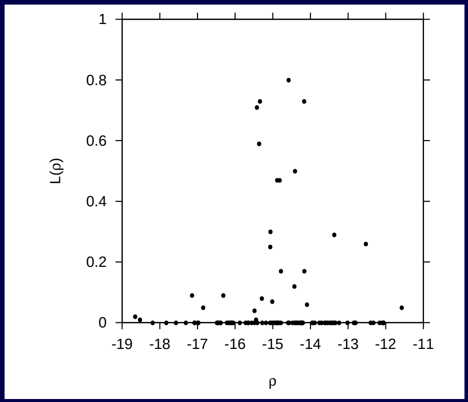
weight w_j of particle $\underline{\theta}_j$ from estimating time-local likelihood function for dynamic noise

$$w_j := \hat{p}(\underline{I}_E|\underline{\theta}_j) = \frac{1}{K}\sum_{k=1}^{K} H\left(\eta - ||\underline{I}_E - \underline{I}_k(\underline{\theta}_j)||_E\right)$$



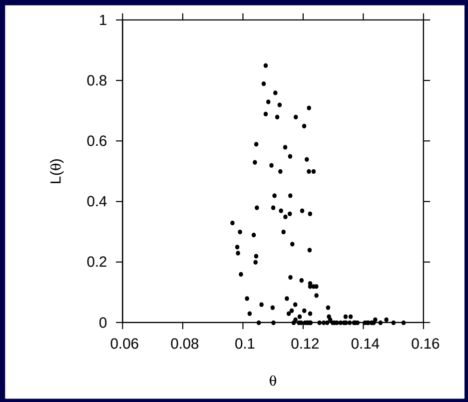


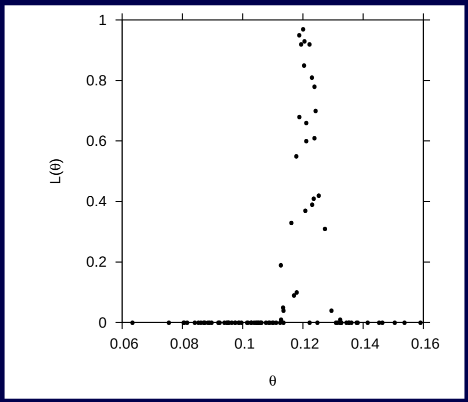




Constructing particle filter: filtering after each 6 months slice

filtering (resample) proportionally to weights w_j of particles $\underline{\theta}_j$ after each 6 months time slice, η-ball size of $\eta = 0.005$





initial distribution final distribution of θ

Particle filter in action

now going $M = 5$ times through the time series with each $\mathcal{L} = 40$ time slices of 6 months,

starting parameter values now not any more $\theta = 0.12$, but $\theta = 0.14$, and not $ln(\varrho) = -15.0$ but $ln(\varrho) = -13.0$

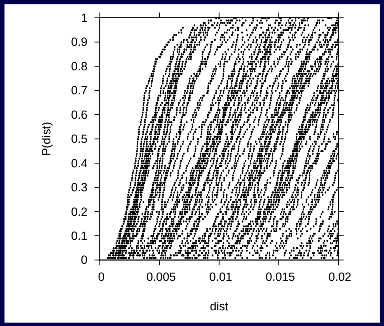
simulated annealing parameters $a = 0.8$ and at each m-tour initial variance factor $b = 2$ (for details see e.g. Bretó et al. 2009), update rule with sample mean over particles $\bar{\theta}_i^{(m)}(\ell)$ at each time slice

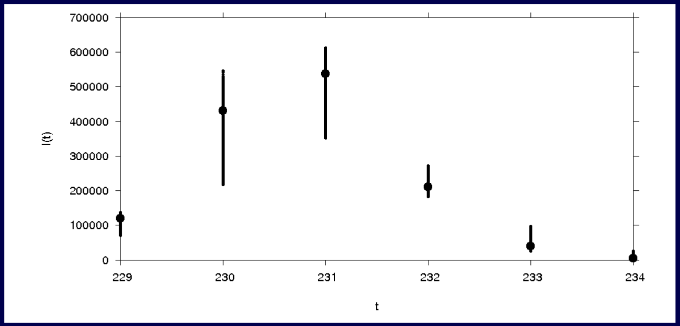
$$\theta_i^{(m+1)} = \sum_{\ell=1}^{\mathcal{L}} \bar{\theta}_i^{(m)}(\ell)$$

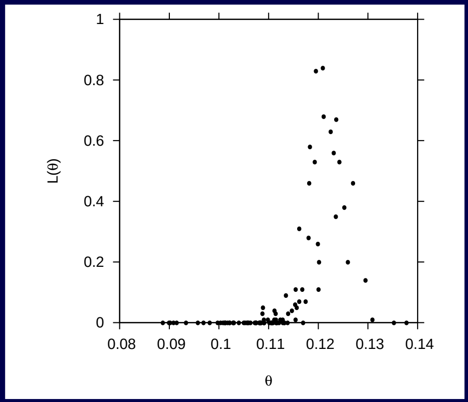
Particle filter in action

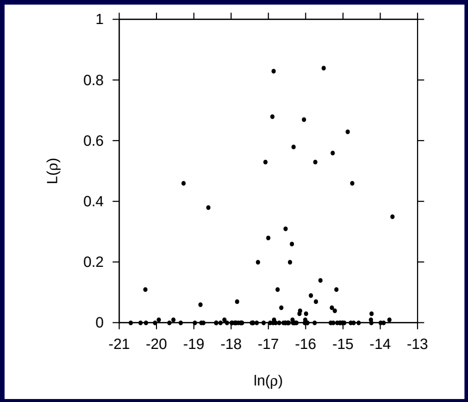
now going $M = 5$ times through the time series with each $\mathcal{L} = 40$ time slices of 6 months,

starting parameter values now not any more $\theta = 0.12$, but $\theta = 0.14$, and not $ln(\varrho) = -15.0$ but $ln(\varrho) = -13.0$





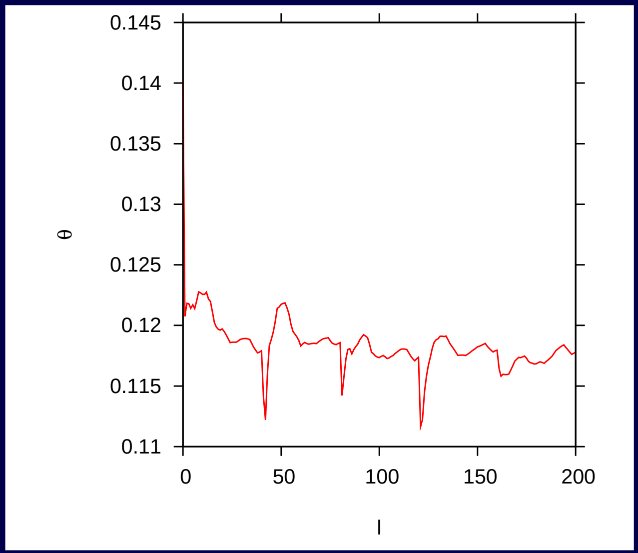


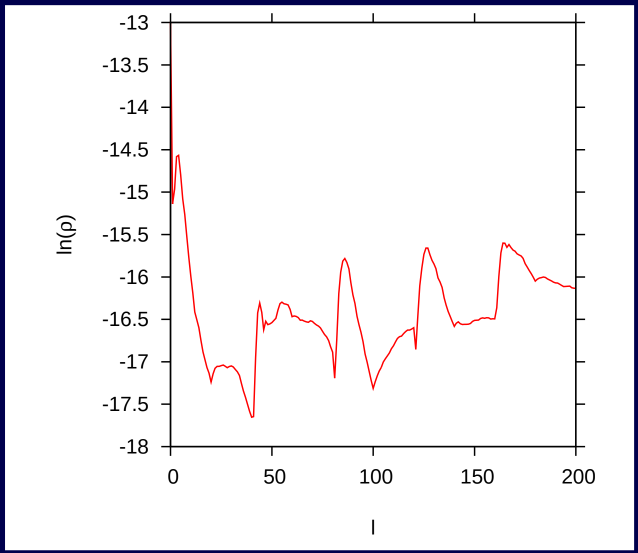


results of final time slice

Particle filter in action: convergence in parameter space

now going $M = 5$ times through the time series with each $\mathcal{L} = 40$ time slices of 6 months,

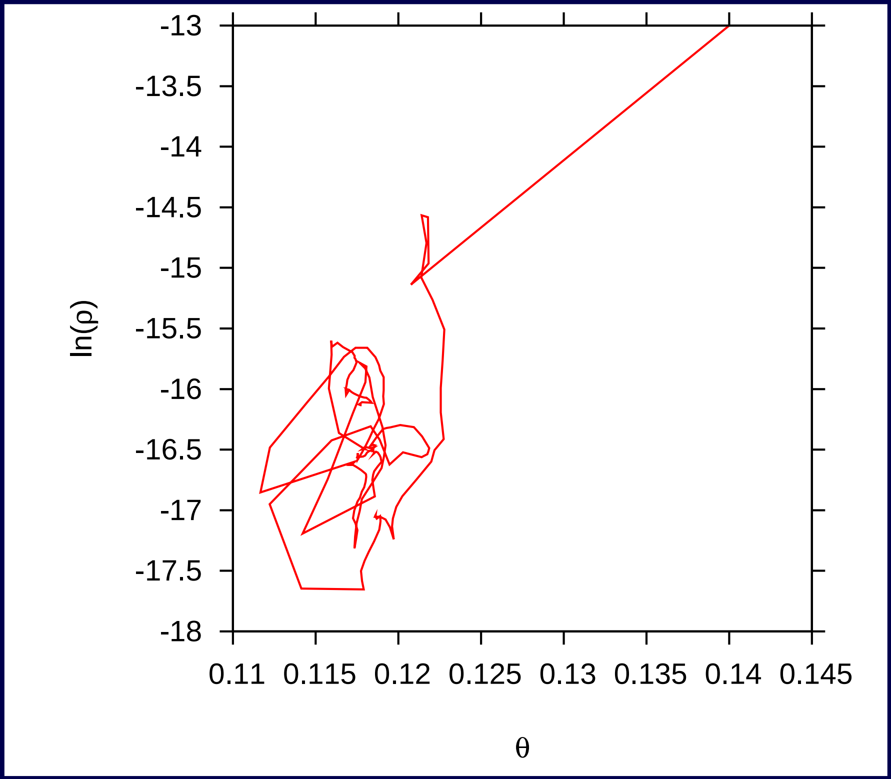




estimates of the parameters along the $M = 5$ runs through the time series with 5×40 time slices covered

Particle filter in action: convergence in parameter space

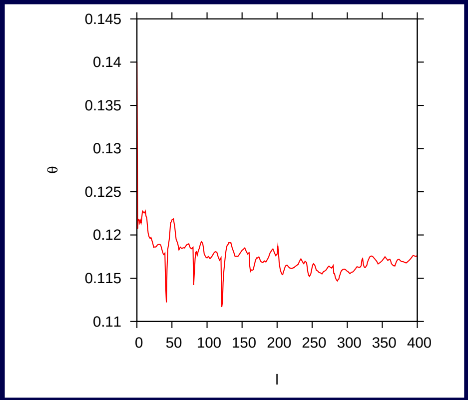
now going $M = 5$ times through the time series with each $\mathcal{L} = 40$ time slices of 6 months,

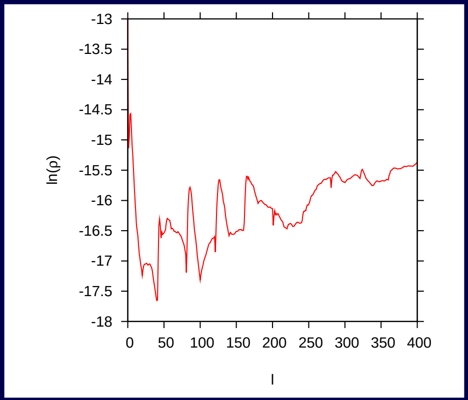


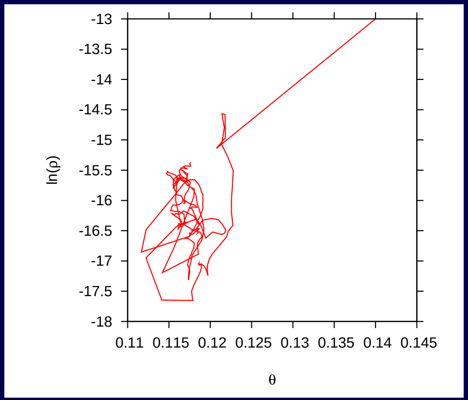
estimates of two parameters jointly

Particle filter in action: convergence in parameter space

now going $M = 10$ times through the time series



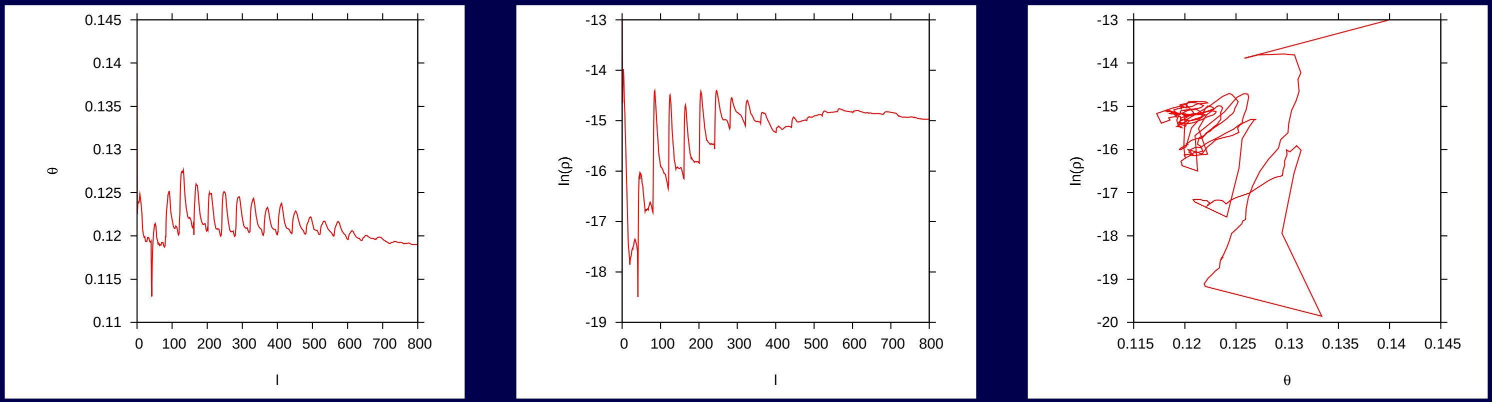




effect of simulated annealing now visible

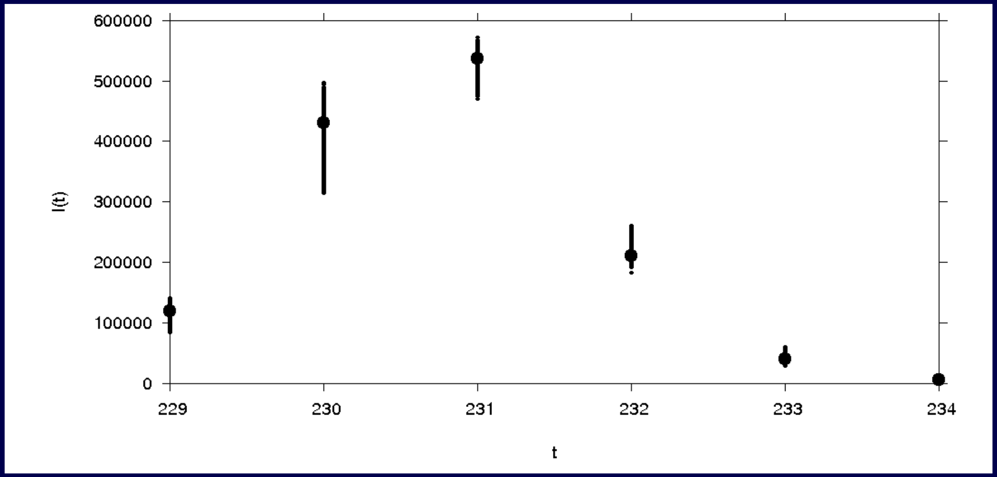
Particle filter in action: convergence in parameter space

now going 20 times through the time series and more particles, better η resolution etc.



completing the iterated filtering for dynamic noise in chaotic population systems

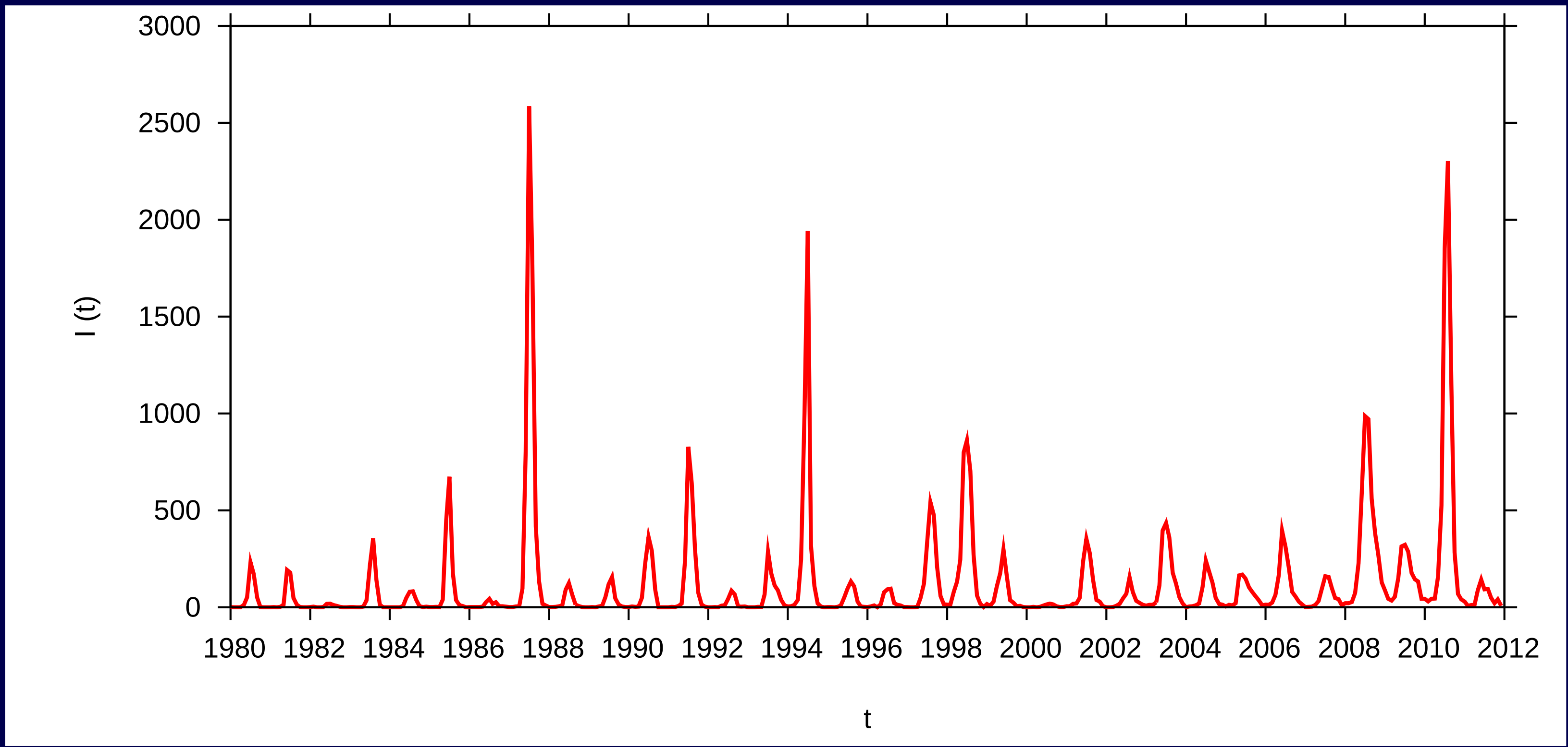
Particle filter in action: good description of the data



cloud of simulations stay close to the data
for the selected parameter sets (particles)

Dengue data from Thailand

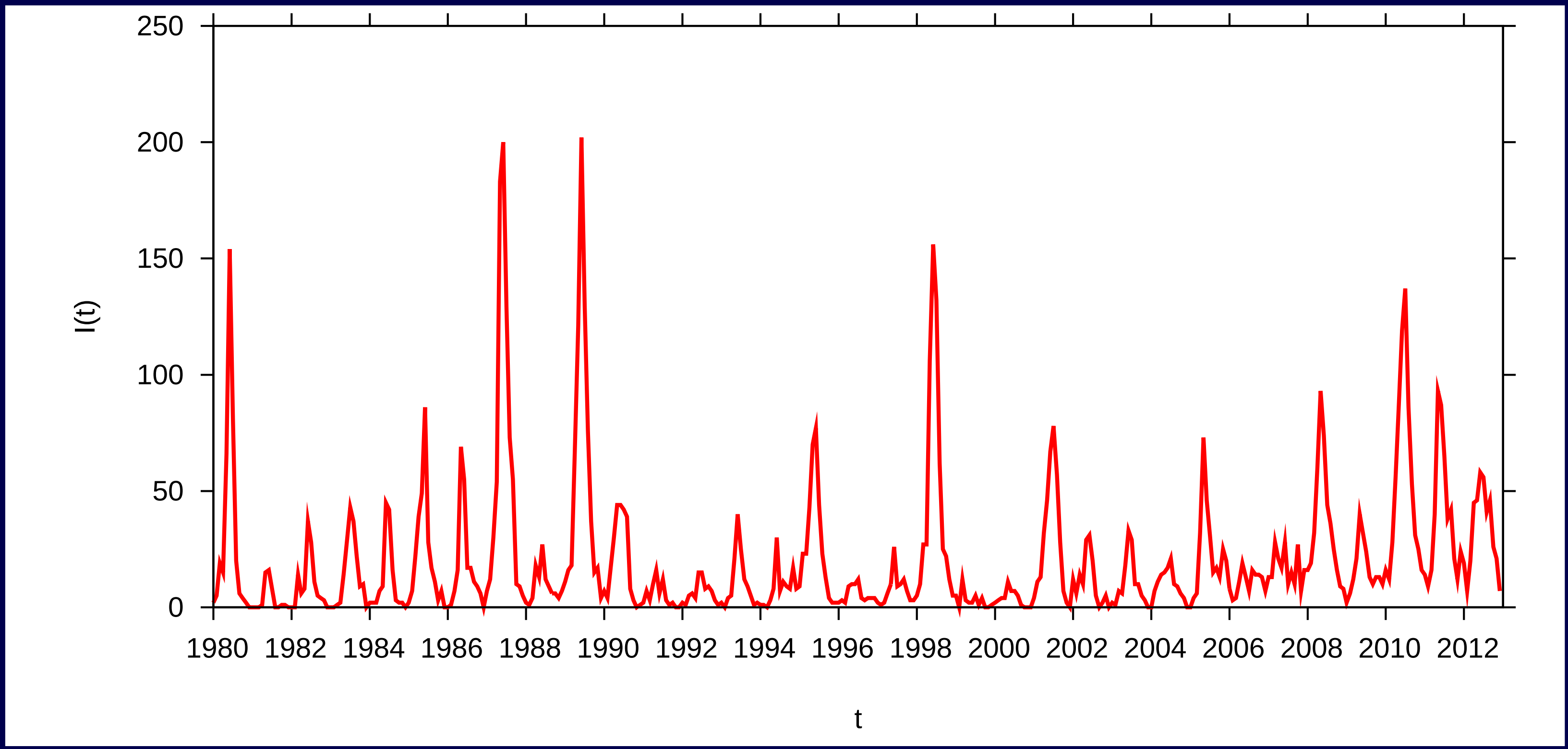
predictability needed to set up iterated filtering



monthly symptomatic dengue cases in Chiang Mai 1980-2011

Dengue data from Thailand

with updated data real time predictability now possible



monthly symptomatic dengue cases
in Trat 1980 to end of 2012