Particle learning for low counts in disease outbreaks

Jarad Niemi
with Mike Ludkovski and Nicholas Michaud

Iowa State University

April 18, 2016

Outline

- Measles outbreak in Zimbabwe
- Model for low counts in disease outbreaks
- Particle learning
- Simulation study
- Application to outbreak in Zimbabwe

Making decisions based on surveillance data

The primary purpose of this work is to

> *use surveillance data to help inform public health officials on control measures.*

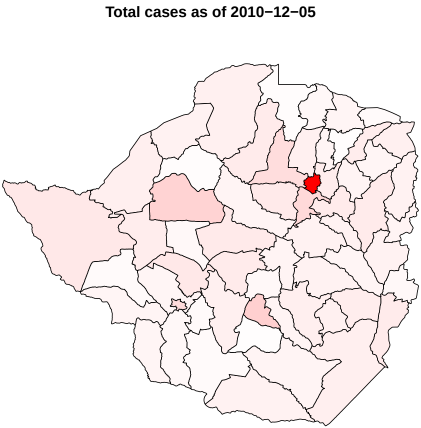
Measles outbreak in Zimbabwe (2009-2010):

- Late summer of 2009, measles detected in Zimbabwe
- Reporting of measles added to regular cholera reporting

    > *Lab confirmed: Suspected case of measles with positive serum IgM antibody, with no history of measles vaccination in the past 4 weeks.*

- Fall of 2009, localized vaccination campaign
- Measles spread across the country
- Summer 2010, mass vaccination campaign
- Fall 2010, no additional measles cases reported

Measles outbreak in Zimbabwe (2009-2010)

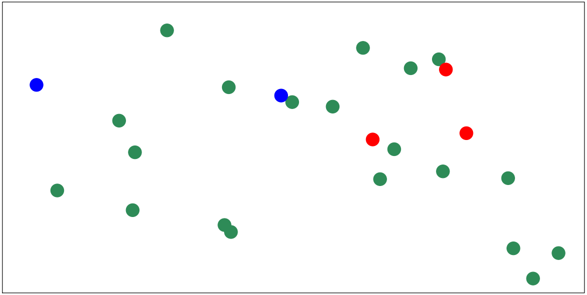


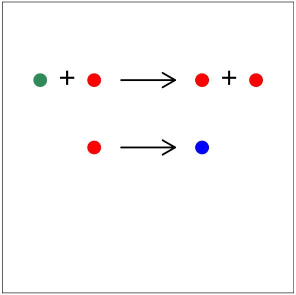
Outline

- Measles outbreak in Zimbabwe
- Model for low counts in disease outbreaks
- Particle learning
- Simulation study
- Application to outbreak in Zimbabwe

Imagine a *well-mixed* space in *thermal equilibrium* with

- M states: $S_1, \ldots, S_M$ with
- number of individuals $X_1, \ldots, X_M$ with elements $X_m \in \mathbb{Z}^+$
- which change according to K transitions: $R_1, \ldots, R_K$ with
- propensities $a_1(x), \ldots, a_K(x)$.
- The propensities are given by $a_k(x) = \lambda_k f_k(x)$
- where $f_k(x)$ is a known function of the system state.
- If transition k occurs, the state is updated by the stoichiometry v_k with
- elements $v_{ij} \in \{-2, -1, 0, 1, 2\}$.





τ-leaping

- If transition $k \in \{1, \ldots, K\}$ has the following probability

$$\lim_{\tau \to 0} \frac{P(\text{transition } k \text{ within the interval } (t, t+\tau) | X_t)}{\tau} = \lambda_k f_k(X_t),$$

then this defines a continuous-time Markov jump process.

This model can be discretized using the τ-leaping approximation:

$$\Delta X_{tk} \stackrel{ind}{\sim} Po(\lambda_k f_k(X_t)\tau)$$

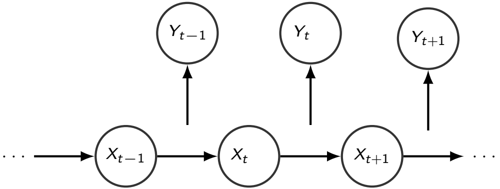
and updating

$$X_{t+\tau,m} = X_{tm} + \sum_{k=1}^{K} v_{mk} \Delta X_{tk}$$

For simplicity, we'll set $\tau = 1$, the observation interval.

Binomial-Poisson discrete-time state-space model

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



S→I→R stochastic compartment model

An SIR compartment model tracks the number of

- Susceptibles (S)
- Infectious (I)
- Recovered (R)

usually with the stipulation that $N = S + I + R$ is constant.

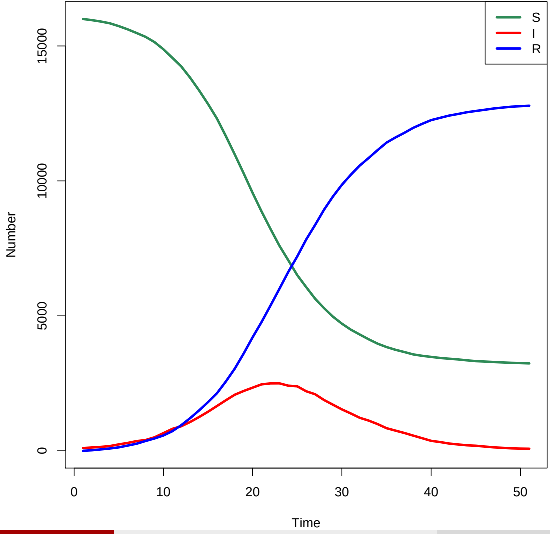
A stochastic SIR model has $M = 3$ (states) and $K = 2$ (transitions) with $X_t = (S_t, I_t, R_t)$,

$$v = \begin{array}{cc} & \begin{array}{cc} S \to I & I \to R \end{array} \\ \begin{array}{c} S \\ I \\ R \end{array} & \left(\begin{array}{cc} -1 & 0 \\ 1 & -1 \\ 0 & 1 \end{array} \right) \end{array},$$

$f_1(X_t) = S_t I_t / N$, and $f_2(X_t) = I_t$.

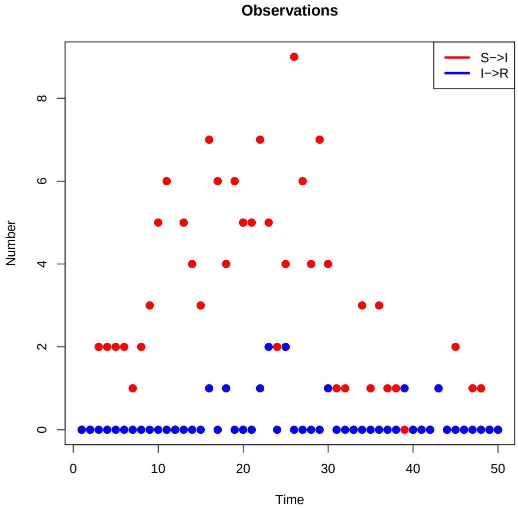
S→I→R stochastic compartment model

$$\Delta X_{S\to I} \sim Po(\lambda_{S\to I} SI/N) \qquad \Delta X_{I\to R} \sim Po(\lambda_{I\to R} I)$$



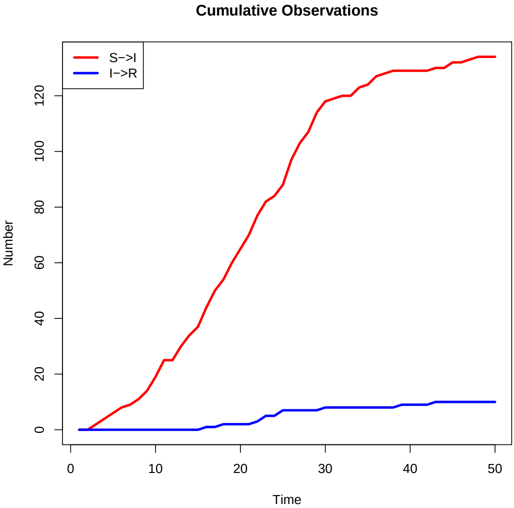
Binomial sampling of transitions

$$Y_{S \to I} \sim Bin(\Delta X_{S \to I}, \theta_{S \to I}) \qquad Y_{I \to R} \sim Bin(\Delta X_{I \to R}, \theta_{I \to R})$$



Binomial sampling of transitions

$$Y_{S\to I} \sim Bin(\Delta X_{S\to I}, \theta_{S\to I}) \qquad Y_{I\to R} \sim Bin(\Delta X_{I\to R}, \theta_{I\to R})$$



Outline

- Measles outbreak in Zimbabwe
- Model for low counts in disease outbreaks
- Particle learning
- Simulation study
- Application to outbreak in Zimbabwe

Bayesian inference

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

$$\begin{aligned}
\theta_k &\overset{ind}{\sim} Be(a_{0k}, b_{0k}), \\
\lambda_k &\overset{ind}{\sim} Ga(c_{0k}, d_{0k}), \\
X_0 &\sim Mult(N; \chi_1, \ldots, \chi_M)
\end{aligned}$$

Filtered distribution:

$$p(X_t, \lambda, \theta | y_{1:t})$$

Forecast distribution

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

Particle learning

Approximating a filtered distribution:

$$p(X_t, \lambda, \theta | y_{1:t}) \approx J^{-1} \sum_{j=1}^{J} \delta_{(X_t, \psi)^{(j)}} p(\lambda | \psi^{(j)}) p(\theta | \psi^{(j)})$$

where

- $\delta_{(X_t, \psi)^{(j)}}$ indicates a *particle* location
- ψ are particle sufficient statistics
- $p(\lambda | \psi^{(j)})$ is a joint distribution for all rate parameters
- $p(\theta | \psi^{(j)})$ is a joint distribution for all sampling parameters

Intuition:

- each particle represents a current belief about the world
- lots of particles provide uncertainty about this belief

Particle learning: going from t to $t+1$

Start with $p(X_t, \lambda, \theta | y_{1:t}) \approx J^{-1} \sum_{j=1}^{J} \delta_{(X_t, \psi_t)^{(j)}} p(\lambda | \psi_t^{(j)}) p(\theta | \psi_t^{(j)})$

1. For all particles,
    a. Sample $\theta^{(j)} \sim p(\theta | \psi^{(j)})$.
    b. Calculate $w_j \propto p(y_{t+1} | X_t^{(j)}, \theta^{(j)}, \psi^{(j)})$.
2. For $j = 1, \ldots, J$
    a. Sample j^* with probability w_{j^*}.
    b. Sample $\lambda^{(j)} \sim p(\lambda | \psi^{(j^*)})$
    c. Sample $\Delta X_{t+1}^{(j)} \sim p(\Delta X | \lambda^{(j)}, \theta^{(j^*)}, X_t^{(j^*)}, y_{t+1})$.
    d. Update $X_{t+1}^{(j)}$ based on $X_t^{(j^*)}$ and $\Delta X_{t+1}^{(j)}$.
    e. Update $\psi_{t+1}^{(j)} = \mathcal{S}(\psi_t^{(j^*)}, y_{t+1}, \Delta X_{t+1}^{(j)})$.

End with $p(X_{t+1}, \lambda, \theta | y_{1:t+1}) \approx J^{-1} \sum_{j=1}^{J} \delta_{(X_{t+1}, \psi_{t+1})^{(j)}} p(\lambda | \psi_{t+1}^{(j)}) p(\theta | \psi_{t+1}^{(j)})$

Particle sufficient statistics

(k subscript is implicit on the next 3 slides)

Recall the model

$$\begin{aligned} Y_{t+1} &\sim Bin(\Delta X_{t+1}, \theta) & \Delta X_{t+1} &\sim Po(\lambda_t f(X_t)), \\ \theta | y_{1:t} &\sim Be(a_t, b_t), & \lambda | y_{1:t} &\sim Ga(c_t, d_t) \end{aligned}$$

Set $\psi_t = (a_t, b_t, c_t, d_t)$, then

$$\begin{aligned} a_{t+1} &= a_t + y_{t+1}, \\ b_{t+1} &= b_t + \Delta X_{t+1} - y_{t+1}, \\ c_{t+1} &= c_t + \Delta X_{t+1}, \\ d_{t+1} &= d_t + f(X_t). \end{aligned}$$

This defines $\psi_{t+1} = \mathcal{S}(\psi_t, y_{t+1}, \Delta X_{t+1})$.

Conditional forward propagation

Recall the model

$$
\begin{aligned}
Y_{t+1} &\sim Bin(\Delta X_{t+1}, \theta) \\
\Delta X_{t+1} &\sim Po(\lambda f(X_t))
\end{aligned}
$$

Then $p(\Delta X_{t+1}|\lambda, \theta, X_t, y_{t+1})$ is

$$
\begin{aligned}
\Delta X_{t+1} &= y_{t+1} + Z_{t+1} \\
Z_{t+1} &\sim Po([1-\theta]\lambda f(X_t))
\end{aligned}
$$

by an appeal to Bayes' Rule, a change of variables, and the marginal distribution for Y_t:

$$
Y_{t+1} \sim Po(\theta \lambda f(X_t)).
$$

One-step ahead predictive distribution

From the previous slide and model construction:

$$\begin{aligned} Y_{t+1} &\sim Po(\theta\lambda f(X_t)) \\ \lambda|y_{1:t} &\sim Ga(c_t, d_t) \end{aligned}$$

Then

$$Y_{t+1}|\theta, \psi_t, X_t \sim NegBin(c_t, e_t)$$

where

$$e_t = \frac{\theta f(X_t)}{d_t + \theta f(X_t)}.$$

and

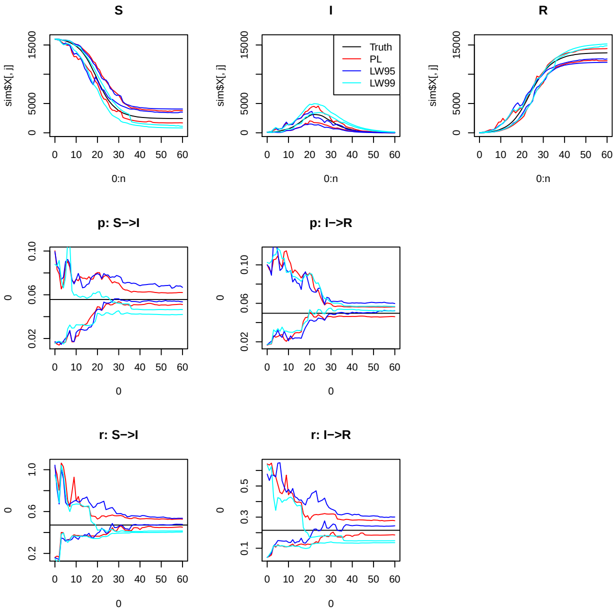
- c_t is the number of failures,
- Y_{t+1} is the number of successes, and
- e_t is the success probability.

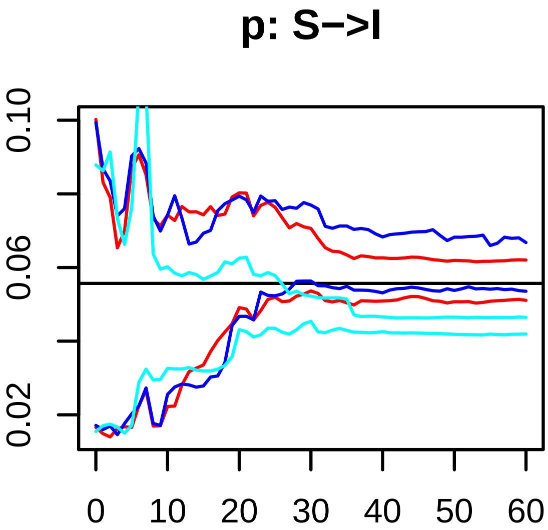
Outline

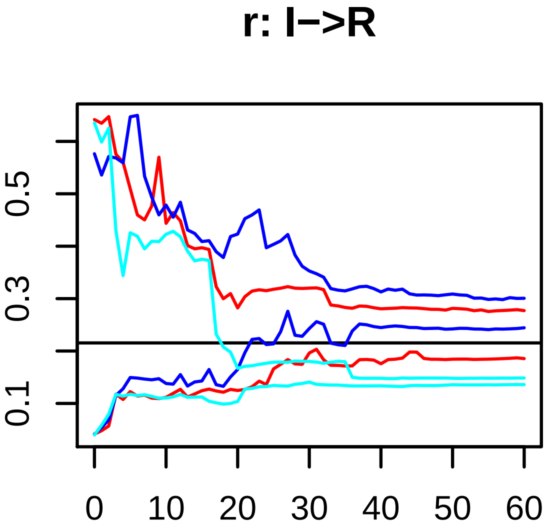
- Measles outbreak in Zimbabwe
- Model for low counts in disease outbreaks
- Particle learning
- Simulation study
- Application to outbreak in Zimbabwe

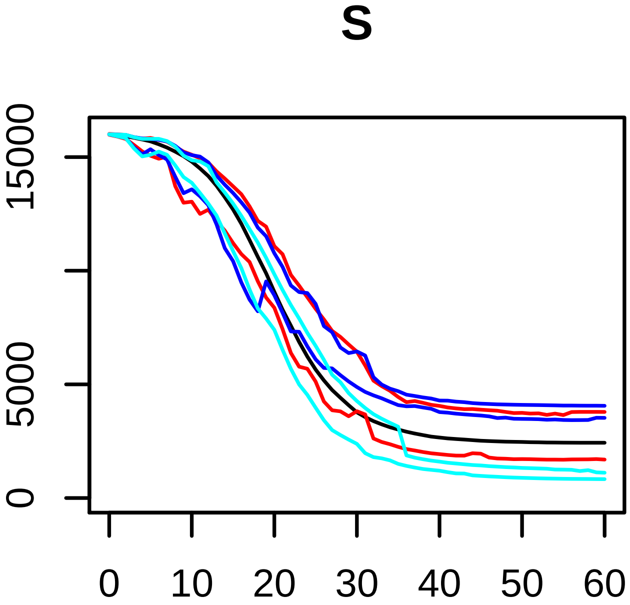
Simulation study

- 100 simulations from
  - $X_0 \sim Mult(16100; (.994, .006, 0))$ corresponds to $E[I_0] = 100$
  - $\theta_{S\to I}, \theta_{I\to R} \sim Be(50, 950)$ for all k
  - $\lambda_{S\to I} \sim Ga(50, 100)$
  - $\lambda_{I\to R} \sim Ga(25, 100)$
  - Ensured simulations had at least one $S \to I$ observation in first 5 time points
- Settings
  - 500 particles
  - multinomial resampling

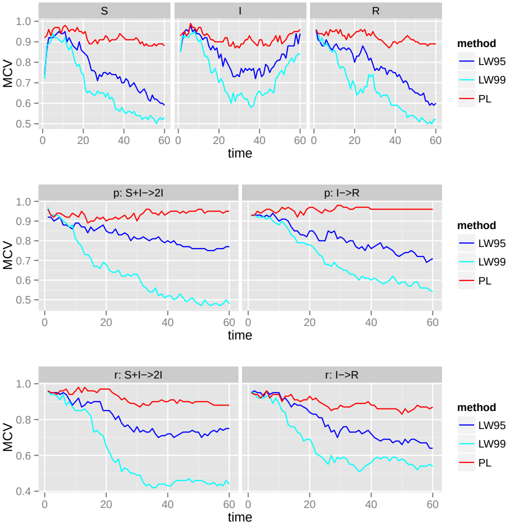

S
I
R
p: S−>I
p: I−>R
r: S−>I
r: I−>R
sim$X[, j]
0:n
Truth
PL
LW95
LW99


p: S−>I
0.10
0.06
0.02
0
10
20
30
40
50
60

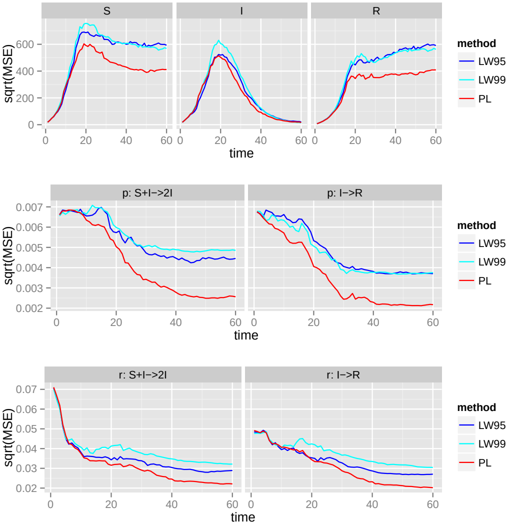

r: I−>R
0.1
0.3
0.5
0
10
20
30
40
50
60


S
0
5000
15000
0
10
20
30
40
50
60

Coverage


S
I
R
MCV
1.0
0.9
0.8
0.7
0.6
0.5
0
20
40
60
time
method
LW95
LW99
PL
p: S+I−>2I
p: I−>R
r: S+I−>2I
r: I−>R
0.4

RMSE



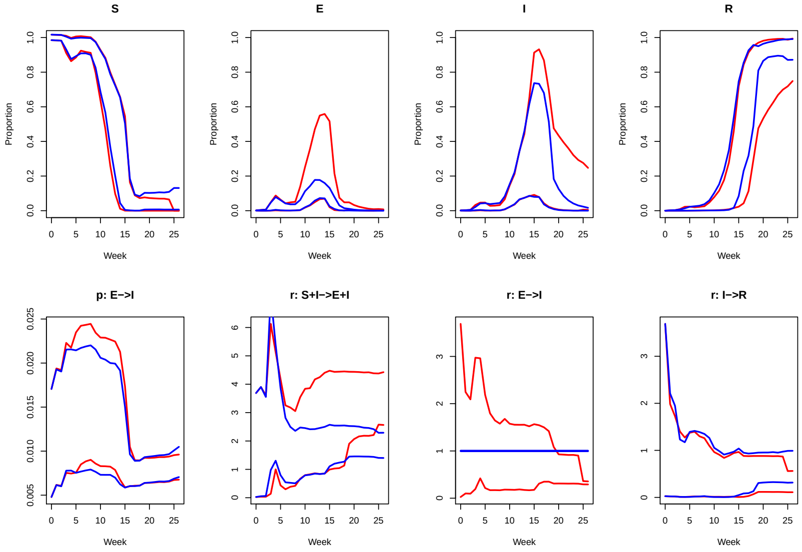
Outline

- Measles outbreak in Zimbabwe
- Model for low counts in disease outbreaks
- Particle learning
- Simulation study
- Application to outbreak in Harare

Harare measles outbreak

- Model
  - Known incubation period: $S \to E \to I \to R$
  - Only observe weekly $E \to I$ transitions
- Priors
  - $N \sim Bin(1.5M, 0.01)$
  - $X \sim Mult(N, (.998, .001, .001, 0))$
  - $\theta_{S\to E} = \theta_{I\to R} = 0$
  - $\theta_{E\to I} \sim Be(10, 990)$
  - $\lambda_{S\to E} \sim Ga(1, 1)$
  - $\lambda_{E\to I} = 1$ and $\lambda_{E\to I} \sim Ga(1, 1)$
  - $\lambda_{I\to R} \sim Ga(1, 1)$
- Settings
  - 10,000 particles
  - stratified resampling

Harare measles outbreak



Summary

- discrete-time binomial-Poisson state-space model
- Particle learning (with integration of some parameters)
- Computationally efficient
- Data - timely, accurate, disaggregated, usable, e.g. `https://github.com/rambaut/MERS-Cases/blob/gh-pages/data/cases.csv`
- Slides: `https://github.com/jarad/IDM2016`
- `tlpl` R package: `https://github.com/jarad/tlpl`

Thank you!

Theoretical results

Specifically, from Section 3.5.1 of Del Moral 2004, for bounded functions f_t and any $p > 1$, the following result holds

$$E_{e_0}^{J}\left[\left|e_t^J(f_t) - e_t(f_t)\right|^p\right]^{1/p} \leq \frac{a(p)b(t)||f||}{\sqrt{J}}$$

where

- $e_t(f_t)$ is the expectation of f_t under the true filtered distribution at time t,
- $e_t^J(f_t)$ is the expectation of f_t under the particle approximation at time t using J particles,
- $a(p)$ is a function of p,
- $b(t)$ is an increasing function of t that depends on which algorithm is used, and
- $||\cdot||$ is the supremum norm.