Time series analysis via mechanistic models

Interdisciplinary Statistics Working Group

June 26, 2008

Ed Ionides

The University of Michigan, Department of Statistics

Collaborators

Carles Breto (Universidad Carlos III de Madrid)

Aaron King (UM Ecology & Evolutionary Biology)

Mercedes Pascual (UM Ecology & Evolutionary Biology)

Daihai He (UM Ecology & Evolutionary Biology)

Menno Bouma (London School of Hygiene and Tropical Medicine).

What is a "mechanistic" approach to time series analysis?

- Write down equations, based on scientific understanding of a dynamic system, which describe how it evolves with time.
- Further equations describe the relationship of the state of the system to available observations on it.
- Mechanistic time series analysis concerns drawing inferences from the available data about the hypothesized equations.
- Questions of general interest: Are the data consistent with a particular model? If so, for what range of values of model parameters? Does one mechanistic model describe the data better than another?
- A defining principle: the model structure should be chosen based on scientific considerations, rather than statistical convenience.

Why quantify biological population dynamics?

- **Conservation**. Mankind is increasingly responsible for managing ecosystems. This requires a quantitative understanding of population behavior.
- **Public health**. Pathogens are also biological populations. Despite successes of vaccination and medical treatment, new diseases are emerging (SARS, HIV/AIDS) and old ones re-emerging due to drug resistant strains (malaria, tuberculosis). Treating the pathogen as part of an ecosystem is one approach to understanding and controlling emergent and re-emergent diseases.
- **Basic scientific interest**.

Time series data of sufficient quantity and quality to justify mechanistic modeling are increasingly available:

Hot off the Press

- **King, Ionides, Pascual and Bouma**. Inapparent infections and cholera dynamics. *To appear in Nature*.
- **Cauchemez, Valeron, Boëlle, Flahault and Ferguson**. Estimating the impact of school closure on influenza transmission from Sentinel data. *Nature*, 10 April 2008.

Inference for nonlinear mechanistic models in ecology

Six problems of Bjornstad and Grenfell (Science, 2001), identifying obstacles for ecological modeling and inference:

1. Combining measurement noise and process noise.
2. Including covariates in mechanistically plausible ways.
3. Continuous time models.
4. Modeling and estimating interactions in coupled systems.
5. Dealing with unobserved variables.
6. Modeling spatial-temporal dynamics.

Wanted:

A framework for modeling and inference allowing consideration of arbitrary nonlinear, partially observed, vector-valued, time series models.

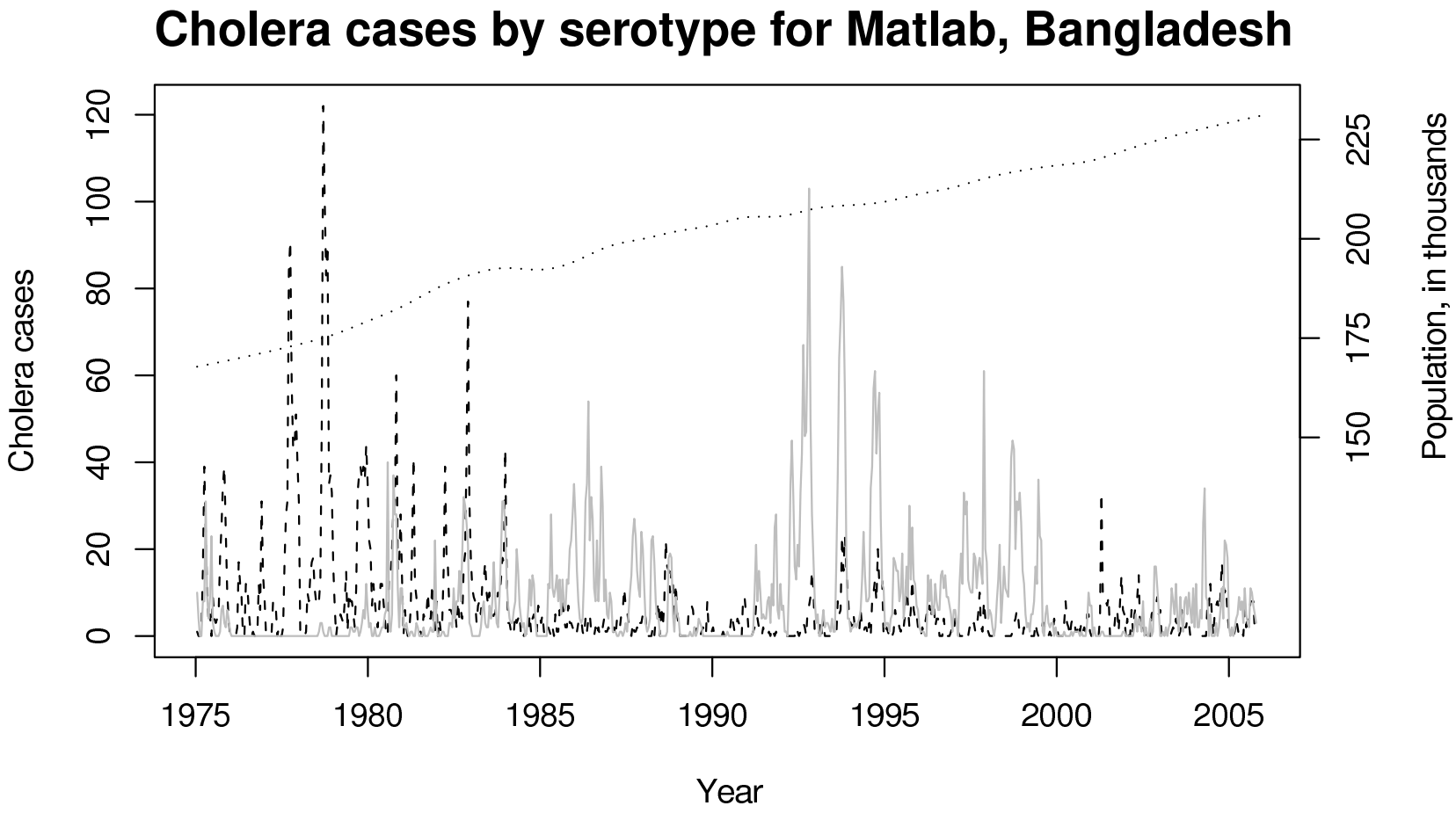
State space models

A state space model is a **partially observed Markov process**. It consists of an unobserved state process x_t and an observation process y_t which is conditionally independent of the past given x_t.

- x_t models a system (discrete or continuous time, usually with some unkown parameters).
- y_t models the available observations (discrete time).

Does this meet the requirements for the modeling framework?

Example: cholera (a diarrheal disease caused by the bacterium *Vibrio cholerae*)



- Biweekly hospital cases from International Center for Diarrheal Disease Research in Matlab, Bangladesh.
- Two serotypes: Inaba (dashed) and Ogawa (solid grey).

A specific question

- Do the data support a hypothesis that cross-immunity between strains explains the apparent strain cycling?

Observational data vs lab experiments for understanding disease dynamics

- To learn how to maintain and control wild lion communities, do you study lions in a zoo? No, you study them in the African bush (maybe this will help construct viable zoo communities).
- To learn about the effects of pH on *V. cholerae*, do you do a lab experiment?

Other open questions for cholera epidemiology

- Seasonality.
- Relation to climate drivers, e.g. El Niño.
- A role for bacteriophage?
- forecasting
  - why do some communities have cholera, others not?
- control measures
  - why are tube wells not more effective?

A general mechanistic model

- A compartment model groups a population into c compartments.
- $X(t) = (X_1(t), \ldots, X_c(t))$ can be written in terms of the flows $N_{ij}(t)$ from i to j, via a **conservation of mass** identity:

$$X_i(t) = X_i(0) + \sum_{j \neq i} N_{ji}(t) - \sum_{j \neq i} N_{ij}(t).$$

- Each **flow** N_{ij} is associated with a **rate** function $\mu_{ij} = \mu_{ij}(t, X(t))$.
- Here, $X_i(t)$ is non-negative integer valued. $X(t)$ models a population divided into c groups; μ_{ij} is the rate at which each individual in compartment i moves to j.
- This makes the compartment model closed. Immigration, birth and death can be included via source and sink compartments.

Why do we want a Markov chain model?

- Occasional low counts / fadeouts occur. A continuous approximation to the population (e.g. stochastic differential equations) may be inappropriate.

Why do we want to add noise to the rates?

- Without noise, infinitesimal mean equals infinitesimal variance.
- Noise is a way to give **over-dispersion**, often critical to fit data (think of generalized linear models).
- For each rate, we specify an **integrated noise process** $\Gamma_{ij}(t)$ giving rise to a **noise process** $\xi_{ij}(t) = \frac{d}{dt}\Gamma_{ij}(t)$.

Properties of the integrated noise processes $\Gamma_{ij}(t)$

P1 Independent increments.

P2 Stationary increments.

P3 Non-negative increments. Therefore, $\xi_{ij}(t) = \frac{d}{dt}\Gamma_{ij}(t)$ is **non-negative white noise**.

P4 Unbiased multiplicative noise: $E[\Gamma_{ij}(t)] = t$.

P5 Partial independence: Γ_{ij} independent of Γ_{ik} for $j \neq k$.

P6 Full independence: all noise processes independent.

P7 **Gamma noise**: marginally, $\Gamma_{ij}(t)$ is a gamma process.

Implicitly defined models

We use an Euler approximation to define a process, rather than vice versa:

1. Divide the interval $[0, T]$ into N intervals of width $\delta = T/N$
2. Set initial value $X(0)$
3. FOR $n = 0$ to $N - 1$
4. Generate noise increments $\{\Delta\Gamma_{ij} = \Gamma_{ij}(n\delta + \delta) - \Gamma_{ij}(n\delta)\}$
5. Generate process increments $(\Delta N_{i1}, \ldots, \Delta N_{i,i-1}, \Delta N_{i,i+1}, \Delta N_{ic}, R_i)$
   $\sim$ Multinomial$(X_i(n\delta), p_{i1}, \ldots, p_{i,i-1}, p_{i,i+1}, \ldots, p_{ic}, 1-\sum_{k\neq i} p_{ik})$
   where $p_{ij} = p_{ij}(\{\mu_{ij}(n\delta, X(n\delta))\}, \{\Delta\Gamma_{ij}\})$ is given in (1)
6. Set $X_i(n\delta + \delta) = R_i + \sum_{j\neq i} \Delta N_{ji}$
7. END FOR

The limiting Markov chain is specified follows:

$$P[\Delta N_{ij} = n_{ij}, \text{ for all } 1 \le i \le c, 1 \le j \le c, i \ne j \mid X(t) = (x_1, \ldots, x_c)]$$

$$= E\left[\prod_{i=1}^{c}\left\{\binom{x_i}{n_{i1}\ \ldots\ n_{ii-1}\ n_{ii+1}\ \ldots\ n_{ic}\ \ r_i}(1-\textstyle\sum_{k\ne i}p_{ik})^{r_i}\displaystyle\prod_{j\ne i}p_{ij}^{n_{ij}}\right\}\right] + o(\delta)$$

where $r_i = x_i - \sum_{k\ne i} n_{ik}$, $\binom{n}{n_1\ \ldots\ n_c}$ is a multinomial coefficient and

$$\begin{aligned} p_{ij} &= p_{ij}(\{\mu_{ij}(t,x)\}, \{\Delta\Gamma_{ij}(t)\}) \\ &= (1-\exp\{-\sum_k \mu_{ik}\Delta\Gamma_{ik}\})\mu_{ij}\Delta\Gamma_{ij}\Big/\sum_k \mu_{ik}\Delta\Gamma_{ik}, \end{aligned}$$

with $\mu_{ij} = \mu_{ij}(t,x)$.

Theorem 1 (Breto, He, Ionides & King: in review). Supposing assumptions (P1–P5) about the noise process, this limit does indeed specify a Markov chain.

Proof. An explicit construction involving exponential transition clocks for each individual, based on the method of Sellke (1983). Such methods are standard for networks of interacting Poisson processes (i.e., our compartment model with no noise). Care is required here due to the introduction of noise.

Theorem 1, formal statement. Suppose (P1–P5) and that $\mu_{ij}(t,x)$ is uniformly continuous as a function of t. Let $C(\zeta,0)$ be the compartment containing individual ζ at time $t=0$. Set $\tau_{\zeta,0}=0$, and generate independent $\mathrm{Exponential}(1)$ random variables $M_{\zeta,0,j}$ for each ζ and $j \neq C(\zeta,0)$. For $m \geq 1$, recursively set

$$\tau_{\zeta,m,j} = \inf\left\{t : \int_{\tau_{\zeta,m-1}}^{t} \mu_{C(\zeta,m-1),j}(s,X(s))\, d\Gamma_{C(\zeta,m-1),j}(s) > M_{\zeta,m-1,j}\right\}.$$

At time $\tau_{\zeta,m} = \min_j \tau_{\zeta,m,j}$, set $C(\zeta,m) = \arg\min_j \tau_{\zeta,m,j}$ and for each $j \neq C(\zeta,m)$ generate an independent $\mathrm{Exponential}(1)$ random variable $M_{\zeta,m,j}$. The increments

$dN_{ij}(t) = \sum_{\zeta,m} \mathbb{I}\{C(\zeta,m-1)=i, C(\zeta,m)=j, \tau_{\zeta,m}=t\}$

specify a Markov chain $X(t)$ whose infinitesimal transition probabilities are given by the limit of the numerical algorithm as $\delta \to 0$.

Theorem 2 (Breto, He, Ionides & King: in review). For the case of independent gamma noise, an analytic formula is available for the infinitesimal probabilities and infinitesimal moments of this chain.

Theorem 2, formal statement. Supposing (P1–P7), the infinitesimal transition probabilities are

$$P[\Delta N_{ij} = n_{ij}, \text{ for all } i \neq j \mid X(t) = (x_1, \ldots, x_c)]$$
$$= \prod_i \prod_{j \neq i} \pi(n_{ij}, x_i, \mu_{ij}, \sigma_{ij}) + o(\delta)$$

where

$$\pi(n, x, \mu, \sigma) = 1_{\{n=0\}} + \delta \binom{x}{n} \sum_{k=0}^{n} \binom{n}{k} (-1)^{n-k+1} \sigma^{-2} \ln\left(1 + \sigma^2 \mu (x-k)\right)$$

Inference for partially observed Markov processes (pomps)

- Ionides, Bretó and King (Proc. Natl. Acad. Sci., 2006) developed an "iterated filtering" method for likelihood based inference for general pomp models.
- This is implemented in the R package pomp, available from CRAN.
- Iterated filtering, implemented by sequential Monte Carlo (a "particle filter") has a **plug-and-play property**: only an algorithm for simulation of sample paths need be supplied to the inference methodology.
- Iterated filtering thus fits in nicely with our notion of implicitly defined mechanistic models.

Key idea of iterated filtering

- Bayesian inference for time-varying parameters becomes a solveable filtering problem. Set $\theta = \theta_t$ to be a random walk with

$$E[\theta_t|\theta_{t-1}] = \theta_{t-1} \qquad \mathrm{Var}(\theta_t|\theta_{t-1}) = \sigma^2$$

- The limit $\sigma \to 0$ can be used to maximize the likelihood for fixed parameters.

Two analogies

- Like the **EM algorithm**, iterated filtering is an optimization trick that takes advantage of a special model structure (partially observed Markov processes).
- Like **simulated annealing**, iterated filtering introduces stochasticity, resulting in “thermal fluctuations” which “cool” toward a “freezing point” at a likelihood maximum.

Theorem 1. (Ionides, Bretó & King, 2006)

Suppose $\hat{\theta}_0$, C *and* $y_{1:T}$ *are fixed and define*

$$\begin{aligned} \hat{\theta}_t &= \hat{\theta}_t(\sigma) = E[\theta_t | y_{1:t}] \\ V_t &= V_t(\sigma) = \mathrm{Var}(\theta_t | y_{1:t-1}) \end{aligned}$$

Assuming sufficient regularity conditions for a Taylor series expansion,

$$\lim_{\sigma \to 0} \sum_{t=1}^{T} V_t^{-1} (\hat{\theta}_t - \hat{\theta}_{t-1}) = (\partial/\partial\theta) \log f(y_{1:T} | \theta, \sigma{=}0) \Big|_{\theta=\hat{\theta}_0}$$

The limit of an appropriately weighted average of local filtered parameter estimates is the derivative of the log likelihood.

Theorem 2. (Ionides, Bretó & King, 2006)

Set $\hat{\theta}^{(n+1)} = \hat{\theta}^{(n)} + \sigma_n^2 M(\nabla \ell(\hat{\theta}^{(n)}) + \eta_n)$, where M is a positive definite symmetric matrix. Suppose the following:

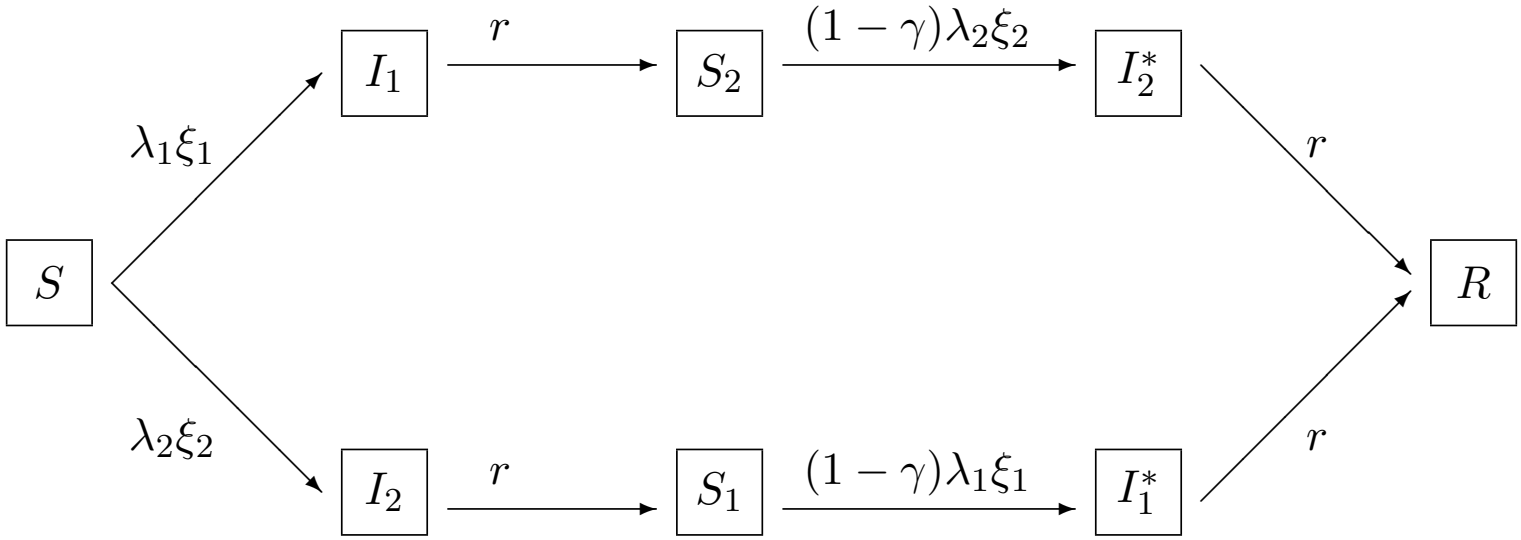
1. $\ell(\theta)$ is twice continuously differentiable and uniformly convex.
2. $\lim_n \sigma_n^2 n^{1-\alpha} > 0$ for some $\alpha \in (0,1)$.
3. $\{\eta_n\}$ has $E[\eta_n] = o(1)$, $\mathrm{Var}(\sigma_n^2 \eta_n) = o(1)$, $\mathrm{Cov}(\eta_m, \eta_n) = 0$ for $m \neq n$.

If there is a $\hat{\theta}$ with $\nabla \ell(\hat{\theta}) = 0$ then $\hat{\theta}^{(n)}$ converges in probability to $\hat{\theta}$.

With appropriate assumptions, iterated filtering does converge to a local maximum if "cooled" sufficiently slowly.

Previous work on likelihood inference for nonlinear partially observed dynamical systems

- In principle, the Bayesian posterior can be found for fixed (non time-varying) parameters. In practice, this is hard to do (Liu and West, 2001).
- Direct calculation of the likelihood surface (Hurzeler and Kunsch, 2001) is not practically feasible on moderate or large dynamical systems.
- Markov Chain Monte Carlo methods such as the Stochastic Expectation-Maximization algorithm (Cappe, 2005) are not readily applicable to continuous time models. In addition, they lack the plug-and-play property.



Cholera model with interacting serotypes.

S, susceptible to both Inaba and Ogawa serotypes;

I_1, infected with Inaba; I_2, infected with Ogawa;

S_1, susceptible to Inaba (immune to Ogawa); S_2, susceptible to Ogawa (immune to Inaba);

I_1^*, infected with Inaba (immune to Ogawa); I_2^*, infected with Ogawa (immune to Inaba);

R, immune to both serotypes.

Births enter S, and all individuals have a mortality rate m.

$$
\begin{array}{rclrcl}
\lambda_1 & = & \beta(t)\frac{(I_1(t)+I_1^*(t))^\alpha}{P(t)} + w & \lambda_2 & = & \beta(t)\frac{(I_2(t)+I_2^*(t))^\alpha}{P(t)} + w \\
\log\beta(t) & = & b_0(t-1990) + \sum_{i=1}^6 b_i s_i(t) & & & \\
\mu_{SI_1} & = & \lambda_1 & \mu_{SI_2} & = & \lambda_2 \\
\mu_{S_1 I_1^*} & = & (1-\gamma)\lambda_1 & \mu_{S_2 I_2^*} & = & (1-\gamma)\lambda_2 \\
\mu_{I_1 S_2} & = & \mu_{I_2 S_1} = r & \mu_{I_2^* R} & = & \mu_{I_1^* R} = r \\
\mu_{X_j D} & = & m \quad \text{for } X_j \in \{S, I_1, I_2, S_1, S_2, I_1^*, I_2^*, R\} & & & \\
\xi_{SI_2} & = & \xi_{S_2 I_2^*} = \xi_2(t) & \xi_{SI_1} & = & \xi_{S_1 I_1^*} = \xi_1(t)
\end{array}
$$

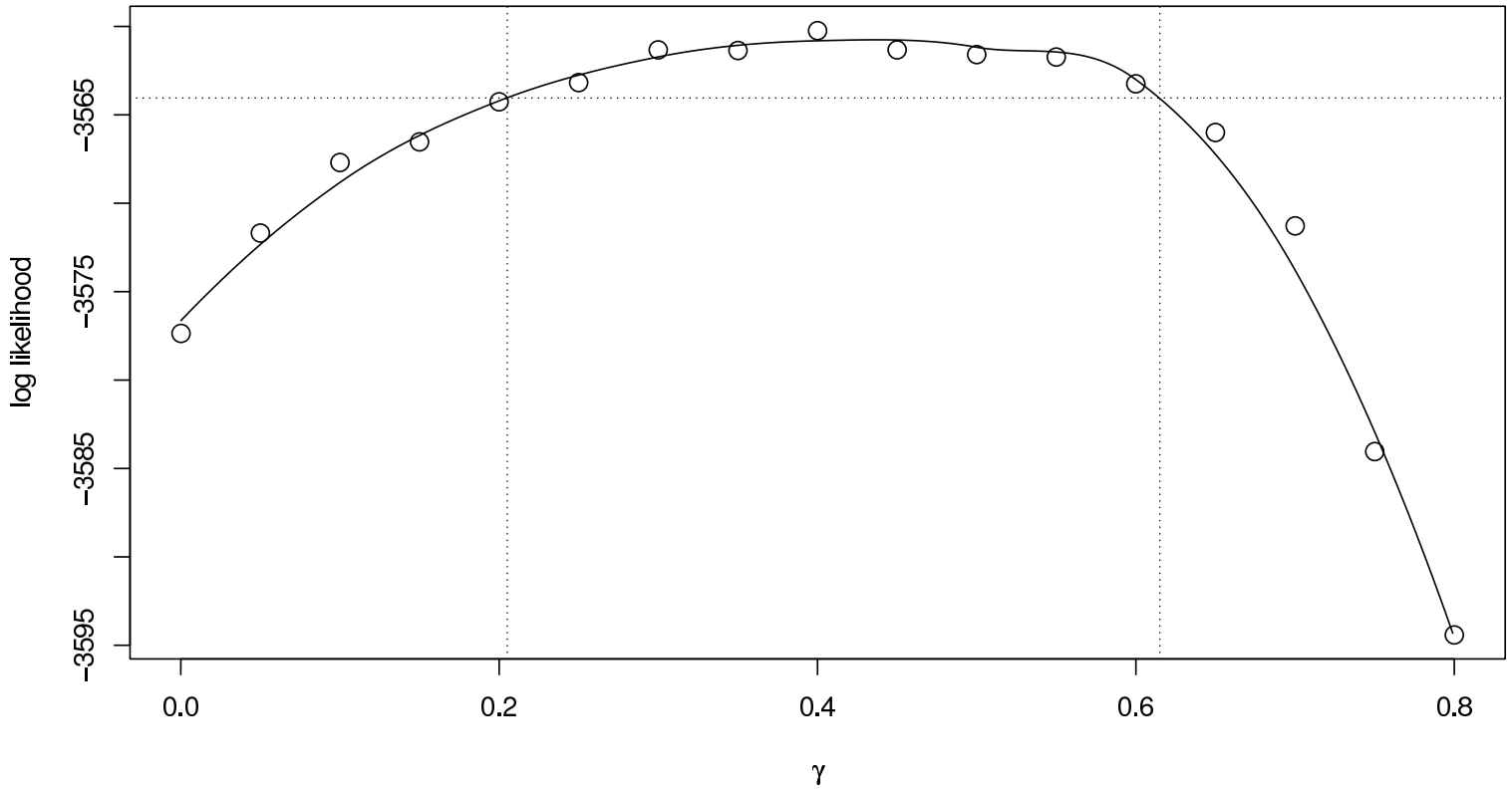
Interpretation of diagram as a multinomial process with random rates.

$\xi_2(t)$ and $\xi_1(t)$ are independent gamma noise processes, each with infinitesimal variance parameter σ^2.

Compartments B and D model demographic events (births, deaths).

Seasonality is modeled via a periodic cubic B-spline basis $\{s_i(t), i = 1, \ldots, 6\}$.

	$\hat{\theta}_A$	SE_A	$\hat{\theta}_B$	SE_B
r	38.42	–	36.91	3.88
ρ	0.067	–	0.653	0.069
γ	0.400	0.087	1.00	0.41
σ	0.1057	0.0076	0.0592	0.0075
ϕ	0.014	0.30	0.0004	0.024
$w \times 10^3$	0.099	0.022	0.0762	0.0072
α	0.860	0.015	0.864	0.017
b_0	−0.0275	0.0017	−0.0209	0.0015
b_1	4.608	0.098	3.507	0.083
b_2	5.342	0.074	3.733	0.091
b_3	5.723	0.075	4.448	0.055
b_4	5.022	0.076	3.534	0.065
b_5	5.508	0.064	4.339	0.053
b_6	5.804	0.059	4.274	0.039
ℓ	−3560.23		−3539.11	



Cross-immunity profile likelihood for regime A, yielding a 99% confidence interval for γ of (0.20, 0.61).

Interpretation of results

- Cross-immunity is weakly identified in both regimes A and B (standard errors are large).
- Regime B, which fits the data considerably better, essentially ignores asymptomatic cholera (estimated reporting rate is close to 1).
- Since Regime A has been the orthodoxy, this is a controversial finding. However, neither A nor B is entirely satisfactory. Improvements in the modeling of short-term disease dynamics (King et al, 2008) may further clarify the role of cross-immunity.

Conclusions

- Plug-and-play statistical methodology permits likelihood-based analysis of flexible classes of stochastic dynamic models.
- This has led to a need for economically-parameterized models for interacting populations (compartment models).
- **It is increasingly possible to carry out data analysis via nonlinear mechanistic stochastic dynamic models**. This should help to build a link between the mathematical modeling community (for whom models are typically conceptual and qualitative) and quantitative applications (testing hypotheses about mechanisms, forecasting, evaluating the consequences of interventions).

Thank you!

These slides are available at

`www.stat.lsa.umich.edu/~ionides/pubs/mechanistic-talk.pdf`

References

Arulampalam, M. S., Maskell, S., Gordon, N., and Clapp, T. (2002). A tutorial on particle filters for online nonlinear, non-Gaussian Bayesian tracking. *IEEE Trans. Sig. Proc.*, 50:174 – 188.

Cappé, O., Moulines, E., and Rydén, T. (2005). *Inference in Hidden Markov Models*. Springer, New York.

Cauchemez, S., Valleron, A., Boëlle, P., Flahault, A., and Ferguson, N. M. (2008). Estimating the impact of school closure on influenza transmission from sentinel data. *Nature*, 452:750–754.

Doucet, A., de Freitas, N., and Gordon, N. J., editors (2001). *Sequential Monte Carlo Methods in Practice*. Springer, New York.

Hürzeler, M. and Künsch, H. R. (2001). Approximating and maximising the likelihood for a general state-space model. In Doucet, A., de Freitas, N., and

Gordon, N. J., editors, *Sequential Monte Carlo Methods in Practice*, pages 159–175. Springer, New York.

Ionides, E. L., Bretó, C., and King, A. A. (2006). Inference for nonlinear dynamical systems. *Proc. Natl. Acad. Sci. USA*, 103:18438–18443.

King, A. A., Ionides, E. L., Pascual, M., and Bouma, M. J. (2008). Inapparent infections and cholera dynamics. *To appear in Nature*.

Liu, J. and West, M. (2001). Combining parameter and state estimation in simulation-based filtering. In Doucet, A., de Freitas, N., and Gordon, N. J., editors, *Sequential Monte Carlo Methods in Practice*, pages 197–224. Springer, New York.

Maximum Likelihood via Iterated Filtering (MIF)

(Ionides, Bretó & King, PNAS, 2006)

Select inital value $\hat{\theta}^{(1)}$ and algorithmic parameters σ_1, c, α and N.

For n in $1, \ldots, N$

(i) set $\sigma = \sigma_1 \alpha^{n-1}$ and initialize $E[\theta_0^{(n)}] = \hat{\theta}^{(n)}$, $\mathrm{Var}(\theta_0^{(n)}) = c\sigma^2$.

(ii) evaluate the filtering means $\hat{\theta}_t^{(n)} = E[\theta_t^{(n)} | y_{1:t}]$ and the prediction variances $V_{t,n} = \mathrm{Var}(\theta_t^{(n)} | y_{1:t-1})$, for $t = 1, \ldots, T$.

(iii) $\hat{\theta}^{(n+1)} = \hat{\theta}^{(n)} + V_{1,n} \sum_{t=1}^{T} V_{t,n}^{-1} (\hat{\theta}_t^{(n)} - \hat{\theta}_{t-1}^{(n)})$

An average of the filtering means, with weights depending on the filtering variances, converges to the maximum of the likelihood (under regularity conditions)

Implementing MIF using Monte Carlo: A brief tutorial

- Let $\{X^F_{t,j}, j = 1, \ldots, J\}$ solve the filtering problem at time t by having (approximately) marginal density $f(x_t|y_{1:t})$.

- **Move particles according to the state process dynamics:**
  Make $X^P_{t+1,j}$ a draw from $f(x_{t+1}|x_t{=}X^F_{t,j})$. Then $\{X^P_{t+1,j}\}$ is a draw from $f(x_{t+1}|y_{1:t})$, solving the prediction problem at time $t+1$.

- **Prune particles according likelihood given data:**
  Make $X^F_{t+1,j}$ a drawn from $\{X^P_{t+1,j}\}$ with probability proportional to $w_j = f(y_t|x_t{=}X^P_{t,j})$. Then $\{X^F_{t+1,j}\}$ solves the filtering problem at $t+1$.

- $E[x_t|y_{1:t}]$ and $\mathrm{Var}(x_t|y_{1:t-1})$ are calculated as the sample mean and variance of $X^F_{t,k}$ and $X^P_{t,k}$ respectively.