Bayesian Inference and Model Selection for Low-Count Time Series Models with Intractable Likelihoods

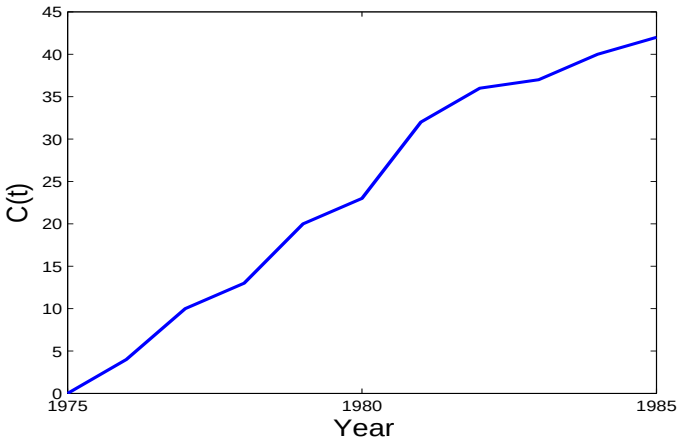
Dr Chris Drovandi
Queensland University of Technology
c.drovandi@qut.edu.au

Collaborators: Tony Pettitt, Roy McCutchan
Acknowledgements: Australian Research Council Discovery Grant and organisers of ARS

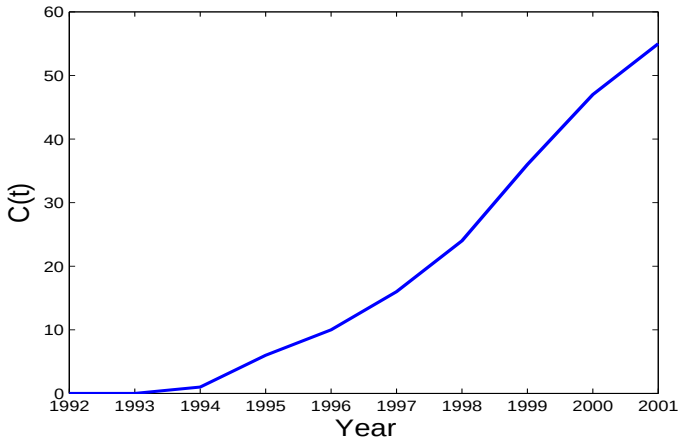
December 17, 2015

Motivating Example - Chronic Wasting Disease

- Dataset: Cumulative number of death of deer due to chronic wasting disease (two separate epidemics)
- Objective: Determine best transmission model from data and estimate parameters



(a) epidemic 1



(b) epidemic 2

Modelling

- Three populations: Susceptible $S(t)$, Infected $I(t)$ and Cumulative death $C(t)$ (Sun et al 2014)

$$\begin{aligned} P(S = i+1, I = j, C = k) &= a\Delta_t + o(\Delta_t), \\ P(S = i-1, I = j, C = k) &= mi\Delta_t + o(\Delta_t), \\ P(S = i-1, I = j+1, C = k) &= \beta ij\Delta_t + o(\Delta_t), \\ P(S = i, I = j-1, C = k) &= mj\Delta_t + o(\Delta_t), \\ P(S = i, I = j-1, C = k+1) &= \mu j\Delta_t + o(\Delta_t), \end{aligned} \quad (1)$$

- β, μ unknown and $I(0)$, $S(0)$ unknown for two epidemics (6 parameters)
- Compare with 'direct' model with 'latent' model. Additional population $L(t)$ and parameter $0 < \alpha < 1$ (proportion of the clinical course spent in latency)

Inference

- Likelihood for such models requires matrix exponential e^{tG} $\rightarrow$ computationally intractable
- But simulation is straightforward $\rightarrow$ Gillespie Algorithm
- Objective: Performing Bayesian estimation and model selection for low count time series models:
  - With intractable likelihoods (but simulation is feasible)
  - Partially observed

Notation

- Time series data $\boldsymbol{y}_{1:T}$, with $\boldsymbol{y}_t$ tth observation
- Prior distribution $p(\boldsymbol{\theta})$, likelihood function $p(\boldsymbol{y}_{1:T}|\boldsymbol{\theta})$
- Require estimate of posterior $p(\boldsymbol{\theta}|\boldsymbol{y}_{1:T})$ and 'evidence' $p(\boldsymbol{y}_{1:T}) = \int_{\boldsymbol{\theta}} p(\boldsymbol{y}_{1:T}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}$ (difficult).
- Model with largest evidence is preferred

State Space Representation

- Hidden states: $\boldsymbol{s}_t$ is a simulated version of $\boldsymbol{y}_t$. $\boldsymbol{x}_t$ allows sequential simulation of model (more later)
- Transition density: $p(\boldsymbol{x}_t, \boldsymbol{s}_t | \boldsymbol{x}_{t-1}, \boldsymbol{s}_{t-1}, \boldsymbol{\theta})$ can be simulated.
- Observation density: $\mathbb{1}(\rho(\boldsymbol{s}_t, \boldsymbol{y}_t) \leq \epsilon_t)$
- ρ is a discrepancy function to compare $\boldsymbol{s}_t$ and $\boldsymbol{y}_t$.
- ϵ_t allows for non-exact matching (in some cases can set $\epsilon_t = 0$)

Bootstrap Particle Filter

- Can use particle filter to obtain unbiased estimate $\hat{p}(\boldsymbol{y}_{1:t}|\boldsymbol{\theta}, \epsilon_{1:t})$ (Del Moral 2004).
- Bootstrap particle filter of Gordon et al (1993) is applicable. Requires simulation from transition density only.
- Consists of a series of re-sampling, propagation and weighting steps
- Weight equation proportional to observation density. In our context the weights may be proportional to 1 or 0.
- May end up with all particles having 0 weight $\rightarrow$ degeneracy.

Alive Particle Filter

- Could think about iteration t of bootstrap particle filter as binomial sampling $\rightarrow$ may end up with 0 successes.
- Alive particle filter (Del Moral et al 2015, Drovandi et al 2015): Performs negative binomial sampling until N successes obtained
- But... Still require unbiased estimate of $\hat{p}(\boldsymbol{y}_{1:t}|\boldsymbol{\theta}, \epsilon_{1:t})$.
- Obtain an extra 'success'. For ith observation let n_i be number of simulations required to produce $N+1$ 'successes'. Unbiased estimate of success probability is then $N/(n_i - 1)$

Alive Particle Filter (Cont...)

Algorithm 1 The alive particle filter with auxiliary variables for data $\boldsymbol{y}_{1:t}$.

1: For notational simplicity set $\log \hat{p}(\boldsymbol{y}_{1:0}|\boldsymbol{\theta}, \epsilon_{1:0}) = 0$
2: Obtain initial simulated data, $\boldsymbol{s}_0^i$, and auxiliary variable information, $\boldsymbol{x}_0^j$, for $j = 1, \ldots, N_x$ if necessary
3: **for** $i = 1$ **to** t **do**
4: Set $n_i = 0$
5: **for** $k = 1$ **to** $N_x + 1$ **do**
6: matched = 'no'
7: **while** matched == 'no' **do**
8: Resample an index r from the set $\{1, \ldots, N_x\}$ with equal weights
9: Generate $\boldsymbol{s}_i^*$ and $\boldsymbol{x}_i^*$ from $p(\boldsymbol{s}_i, \boldsymbol{x}_i | \boldsymbol{s}_{i-1}^r, \boldsymbol{x}_{i-1}^r, \boldsymbol{\theta})$
10: Set $n_i = n_i + 1$
11: **if** $n_i == K$ **then**
12: Set $\hat{p}(\boldsymbol{y}_{1:t}|\boldsymbol{\theta}, \epsilon_{1:t}) = 0$ and return
13: **end if**
14: **if** $\rho(\boldsymbol{s}_i^*, \boldsymbol{y}_i) \leqslant \epsilon_i$ **then**
15: Set $\boldsymbol{s}_i^k = \boldsymbol{s}_i^*$, $\boldsymbol{x}_i^k = \boldsymbol{x}_i^*$ and matched = 'yes'
16: **end if**
17: **end while**
18: **end for**
19: Set $\log \hat{p}(\boldsymbol{y}_i|\boldsymbol{y}_{1:i-1}, \boldsymbol{\theta}, \epsilon_{1:i}) = \log(\frac{N_x}{n_i - 1})$
20: Set $\log \hat{p}(\boldsymbol{y}_{1:i}|\boldsymbol{\theta}, \epsilon_{1:i}) = \log \hat{p}(\boldsymbol{y}_{1:i-1}|\boldsymbol{\theta}, \epsilon_{1:i-1}) + \log \hat{p}(\boldsymbol{y}_i|\boldsymbol{y}_{1:i-1}, \boldsymbol{\theta}, \epsilon_{1:i})$
21: **end for**

Alive Particle Filter (Cont...)

- Add in stopping factor K. Guards against poor proposals where it takes too long to get N 'matches'

Metropolis-Hastings + Alive Particle Filter

Drovandi et al 2015 (Bayesian Analysis)

- Feed unbiased estimate from alive particle filter into MH algorithm (ala Andrieu and Roberts 2009, Andrieu et al 2010)
- Reversible jump to move between models

Drawbacks:

- Relies on reversible jump moves
- Must start MCMC at a 'good' value
- Other drawbacks associated with MCMC

Sequential Monte Carlo + Alive Particle Filter

- Chopin et al 2013 consider using bootstrap particle filter within SMC algorithm $\rightarrow$ SMC^2
- We adapt this to use alive particle filter instead $\rightarrow$ alive SMC^2

SMC with likelihood (Chopin et al 2002)

- Sequence of targets from prior to posterior (e.g. data annealing)
- Traverse a set of N 'particles' through sequence of distributions using re-weighting, re-sampling and perturbation steps.
- Re-weighting Step:

$$w_t = W_{t-1} p(\boldsymbol{y}_t | \boldsymbol{y}_{1:t-1}, \boldsymbol{\theta}), \tag{2}$$

- Resampling: When $ESS = 1/\sum (W_t^i)^2 < N/2$ resample particle set proportional to weights
- Perturbation: Move particles with MCMC kernel of invariant distribution $p(\boldsymbol{\theta}|\boldsymbol{y}_{1:t})$. Repeat R times.
- Estimate of Evidence: Product of sum of unnormalised weights $\widehat{Z_T} = \prod_{t=1}^{T} \sum_{i=1}^{N} w_t^i$

SMC with alive particle filter

- Standard SMC requires likelihood evaluations for re-weighting and perturbation steps.
- Re-weight: Estimate conditional likelihood $p(\boldsymbol{y}_t|\boldsymbol{y}_{1:t-1}, \boldsymbol{\theta}, \epsilon_{1:t})$:
  - Run iteration t of alive particle filter
  - Requires the use of auxiliary variables $\{\boldsymbol{x}^j_{t-1}, \boldsymbol{s}^j_{t-1}\}_{j=1}^{N_x}$ saved from iteration t of alive particle filter.
- Perturbation step:
  - For each $\boldsymbol{\theta}^*$ run alive particle filter based on data $\boldsymbol{y}_{1:t}$
  - Will produce a new set of auxiliary information at time t $\{\boldsymbol{x}^j_{t-1}, \boldsymbol{s}^j_{t-1}\}_{j=1}^{N_x}$ to use for time $t+1$
- Evidence: estimate has the same form as before

Handling highly informative first observations

- If a 'vague' prior is used then first observation may result in large reduction of ESS
- Our Approach:
  - Repeated simulations from prior until N_θ 'matches' with first observation $\boldsymbol{y}_1$
  - For each $\boldsymbol{\theta}$ kept estimate likelihood based on $\boldsymbol{y}_1$
  - Each $\boldsymbol{\theta}$ kept has equal weight

Alive SMC^2

Benefits

- Reversible jump avoided
- Overcomes 'starting value' problem of MCMC approach
- Adaptation easy within SMC
- Easily parallelisable

Drawbacks

- Any observation may be highly informative (large reduction in ESS)
- Does not scale well with number of observations (neither does MCMC approach)

Integer Autoregressive Moving Average Models

INARMA(p,q)

$$Y_t = \sum_{i=1}^{p} \alpha_i \circ Y_{t-i} + u_t + \sum_{j=1}^{q} \beta_j \circ u_{t-j},$$

where $u_t \overset{iid}{\sim} Discrete$. Here we use $u_t \overset{\text{iid}}{\sim} \text{ZIP}(\lambda, \rho)$. $\circ$ is the binomial thinning operator.

- INAR(1) likelihood is a convolution of binomial and ZIP distributions (tractable)
- However $p \geq 2$ becomes more and more intractable
- For $q \geq 1$ have non-Markovian model

Jazi et al 2012 consider only INAR(1) type models

INARMA Example

- We consider 4 models: INAR(1), INAR(2), INMA(1), INARMA(1,1)
- Consider exact matching $\rightarrow \epsilon_{1:T} = 0$.
- When MA(1) component considered we require auxiliary variable $\boldsymbol{x_t} = u_t$.

Simulation study

- Considered 3 datasets from each model (12 datasets in total)
- Alive SMC^2 recovered true model for 11 out of 12 datasets. ZIP INMA(1) sometimes confused with the ZIP INARMA(1,1) model

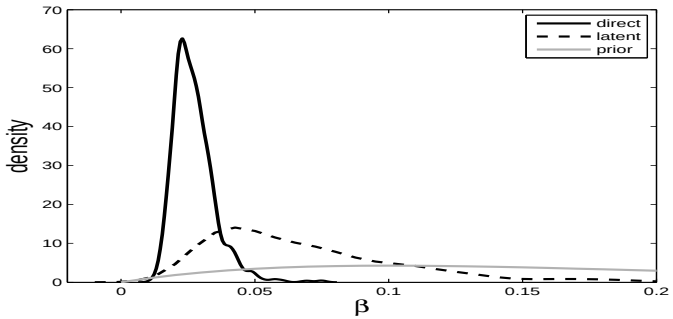
INARMA Model - Bovine datasets

- Data is the monthly number of submissions of bovine displaying certain symptoms to animal health laboratories
- Symptoms abortion, anorexia, illthrift, skin lesions and sudden death (datasets also considered in Jazi et al 2012)
- Fit the four models to each dataset

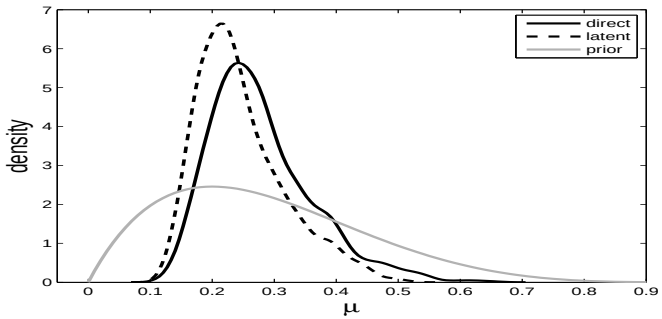
Bovine datasets - Results

data	method	INAR(1)	INAR(2)	INMA(1)	INARMA(1,1)
abort	SMC^2	0.35 (0.06)	**0.45** (0.08)	0.014 (0.003)	0.18 (0.05)
abort	IS	0.34 (0.01)	**0.51** (0.02)	0.013 (0.002)	0.14 (0.01)
anor	SMC^2	0.29 (0.05)	**0.43** (0.07)	0.016 (0.002)	0.27 (0.05)
anor	IS	0.31 (0.02)	**0.47** (0.02)	0.014 (0.001)	0.20 (0.01)
illthrift	SMC^2	0.23 (0.04)	**0.52** (0.08)	0.00 (0.00)	0.25 (0.06)
skin	SMC^2	0.17 (0.03)	**0.68** (0.04)	0.09 (0.02)	0.05 (0.006)
sudden	SMC^2	0.21 (0.05)	0.09 (0.02)	0.01 (0.002)	**0.69** (0.06)

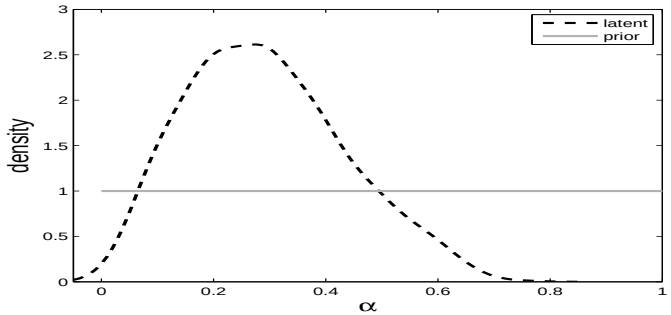
Chronic Wasting Disease - Results (Exact Matching)



(c) β



(d) μ



(e) α

The estimated posterior model probability of the latent model is 0.74 (0.03).

Chronic Wasting Disease - Results ($\epsilon_t = 1$)

- Similar parameter posterior distributions
- The estimated posterior model probability of the latent model is 0.77 (0.02).
- Roughly halved the computing time

Final Comments

Comments

- Avoids the need for summary statistics (excessive loss of information). Further, model choice is delicate with summary statistics (Robert, Marin and colleagues (2014,2015))
- EP ABC (Barthelme and Chopin) is applicable but has limitations

Future Work

- At observation t go through a sequence of tolerances $\epsilon_t^1 > \epsilon_t^2 > \cdots > \epsilon_t^L = \epsilon_t$. SMC^3 algorithm???
- Internal estimate of variability of evidence estimate?
- Incorporate alive particle filter within lazy ABC framework (Prangle 2015) to remove effect of fudge parameter K

References

- Drovand and McCutchan (2015). Alive SMC^2: Bayesian Model Selection for Low-Count Time Series Models with Intractable Likelihoods. Biometrics.
- Drovandi, Pettitt and McCutchan (2015). Exact and approximate Bayesian inference for low count time series models with intractable likelihoods. Bayesian Analysis.
- Andrieu and Roberts (2009). The pseudo-marginal approach for efficient Monte Carlo computations. The Annals of Statistics.
- Andrieu, Doucet and Holenstein (2010). Particle Markov chain Monte Carlo methods. JRSS B.
- Chopin, Jacob and Papaspiliopoulos (2013). SMC^2: an efficient algorithm for sequential analysis of state space models. JRSS B.