Real time changepoint detection with applications in dynamic networks

Melissa Turcotte and Nick Heard

Department of Mathematics
Institute for Security Science and Technology
Imperial College London

27 March 2012

1. Dynamic Networks
2. Continuous time changepoint model
3. Sequential Monte Carlo Algorithm
4. Examples

Networks

Networks are representational forms for systems in nature, society and technology, for example

- data traffic flows on a computer network
- communications between individuals in a social network such as facebook or other communities of interest

Dynamic networks are characterised by

- members becoming active/inactive
- links between members weakening/strengthening
- appearance/disappearance of members

Graphs are considered the most appropriate data form to store information about the network.

Networks

Networks are representational forms for systems in nature, society and technology, for example

- data traffic flows on a computer network
- communications between individuals in a social network such as facebook or other communities of interest

Dynamic networks are characterised by

- members becoming active/inactive
- links between members weakening/strengthening
- appearance/disappearance of members

Graphs are considered the most appropriate data form to store information about the network.

Networks

Networks are representational forms for systems in nature, society and technology, for example

- data traffic flows on a computer network
- communications between individuals in a social network such as facebook or other communities of interest

Dynamic networks are characterised by

- members becoming active/inactive
- links between members weakening/strengthening
- appearance/disappearance of members

Graphs are considered the most appropriate data form to store information about the network.

Motivating Problem

Detect anomalies in dynamic networks in real time.

Precise definition of an anomaly depends on application of interest

- sudden changes in connectivity between nodes
- links between unrelated nodes
- changes within substructures

Anomaly detection has important security applications

- virus detection in a computer network
- monitoring telecommunication networks

Motivating Problem

Detect anomalies in dynamic networks in real time.

Precise definition of an anomaly depends on application of interest

- sudden changes in connectivity between nodes
- links between unrelated nodes
- changes within substructures

Anomaly detection has important security applications

- virus detection in a computer network
- monitoring telecommunication networks

Features of the data

- Large number of nodes
- High frequency traffic arriving as a data stream
- Often the data exhibit seasonality; weekends

Need fast, parallelisable methods to do inference in real time.

Methodology

If a network has undergone some structural changes this suggests

- some entities are communicating more or less frequently
- there are new links between entities

Heard *et al* (2010) looked at a two-stage approach

- treat nodes or edges independently and identify potentially anomalous nodes through deviation from their usual connectivity
- create a subgraph around anomalous nodes and use standard graph analysis tools in this reduced subnetwork

Concentrate on stage 1 in a continuous time frame using changepoint models...

Methodology

If a network has undergone some structural changes this suggests

- some entities are communicating more or less frequently
- there are new links between entities

Heard *et al* (2010) looked at a two-stage approach

- treat nodes or edges independently and identify potentially anomalous nodes through deviation from their usual connectivity
- create a subgraph around anomalous nodes and use standard graph analysis tools in this reduced subnetwork

Concentrate on stage 1 in a continuous time frame using changepoint models...

Methodology

If a network has undergone some structural changes this suggests

- some entities are communicating more or less frequently
- there are new links between entities

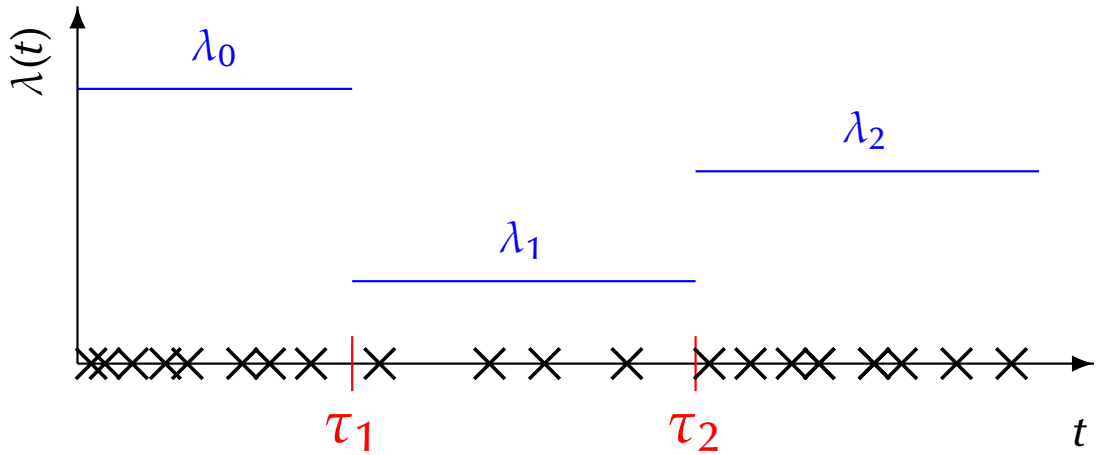
Heard *et al* (2010) looked at a two-stage approach

- treat nodes or edges independently and identify potentially anomalous nodes through deviation from their usual connectivity
- create a subgraph around anomalous nodes and use standard graph analysis tools in this reduced subnetwork

Concentrate on stage 1 in a continuous time frame using changepoint models...

Bayesian Changepoint Model for Communications

- Suppose node or edge communications in the network follow a Poisson process with piecewise constant intensity $\lambda(t)$.
- Jumps or changepoints represent anomalies and split up the data into disjoint segments.
- In the simplest case, the i^{th} segment is modelled by a homogeneous Poisson process with intensity λ_i.



Priors

For a process observed over $[t_0, t_n]$

- k_n changepoints $\tau_{1:k_n} = (\tau_1, \ldots, \tau_{k_n})$ follow a homogeneous Poisson process with intensity ν
- $(k_n + 1)$ intensities follow independent conjugate priors, $\lambda_{0:k_n} = (\lambda_0, \ldots, \lambda_{k_n}) \overset{\text{iid}}{\sim} \Gamma(\alpha, \beta)$

Intensity on the number of changepoints is assumed to be much smaller than the rate of communications, i.e $\nu \ll \lambda_i$.

Posterior

Given observed data $\mathcal{D}$ over $[t_0, t_n]$, the posterior distribution on the changepoints and intensities $\pi_{[t_0,t_n]}(\tau_{1:k_n}, \lambda_{0:k_n}, k_n|\mathcal{D})$ is known up to proportionality through

$$\gamma_{[t_0,t_n]}(\tau_{1:k_n}, \lambda_{0:k_n}, k_n, \mathcal{D}) = p\left(\mathcal{D}|\tau_{1:k_n}, \lambda_{0:k_n}, k_n\right) p\left(\tau_{1:k_n}, \lambda_{0:k_n}, k_n\right).$$

Integrating out λ gives

$$\gamma_{[t_0,t_n]}(\tau_{1:k_n}, k_n, \mathcal{D}) = \nu e^{-\nu t} \prod_{i=0}^{k_n} \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha + r_i)}{(\beta + \tau_{i+1} - \tau_i)^{r_i+\alpha}}$$

where $\tau_0 = t_0$, $\tau_{k_n+1} = t_n$ and $r_i = \#$ of observations in the i^{th} segment.

For simplification of notation will drop the dependency on $\mathcal{D}$.

Posterior

Given observed data $\mathcal{D}$ over $[t_0, t_n]$, the posterior distribution on the changepoints and intensities $\pi_{[t_0,t_n]}(\tau_{1:k_n}, \lambda_{0:k_n}, k_n|\mathcal{D})$ is known up to proportionality through

$$\gamma_{[t_0,t_n]}(\tau_{1:k_n}, \lambda_{0:k_n}, k_n, \mathcal{D}) = p\left(\mathcal{D}|\tau_{1:k_n}, \lambda_{0:k_n}, k_n\right) p\left(\tau_{1:k_n}, \lambda_{0:k_n}, k_n\right).$$

Integrating out λ gives

$$\gamma_{[t_0,t_n]}(\tau_{1:k_n}, k_n, \mathcal{D}) = \nu e^{-\nu t} \prod_{i=0}^{k_n} \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha + r_i)}{(\beta + \tau_{i+1} - \tau_i)^{r_i+\alpha}}$$

where $\tau_0 = t_0$, $\tau_{k_n+1} = t_n$ and $r_i = \#$ of observations in the i^{th} segment.

For simplification of notation will drop the dependency on $\mathcal{D}$.

- Given a specific time frame we can use RJMCMC to obtain a sample from the target distribution $\pi_{[t_0,t_n]}(\tau_{1:k_n}, k_n)$.
- In a sequential setting, using RJMCMC every time we want a real time update would be computationally slow.
- Can use Sequential Monte Carlo (SMC) ideas to exploit the similarity of $\pi_{[t_0,t_n]}$ and $\pi_{[t_0,t_{n+1}]}$.

- Given a specific time frame we can use RJMCMC to obtain a sample from the target distribution $\pi_{[t_0,t_n]}(\tau_{1:k_n}, k_n)$.
- In a sequential setting, using RJMCMC every time we want a real time update would be computationally slow.
- Can use Sequential Monte Carlo (SMC) ideas to exploit the similarity of $\pi_{[t_0,t_n]}$ and $\pi_{[t_0,t_{n+1}]}$.

SMC

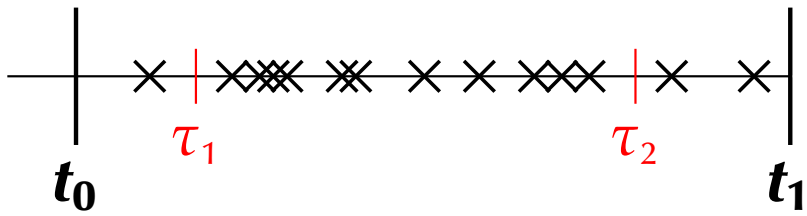
Given a weighted sample $\{(\tau_{1:k_n}, k_n)^{(i)}, w_n^{(i)}\}_{i=1}^N$ at t_n from an IS approximation of $\pi_{[t_0,t_n]}$, at t_{n+1} we seek to

- extend each changepoint vector $\tau_{1:k_n}^{(i)}$ to be a sample $\tau_{1:k_{n+1}}^{(i)}$ from $\pi_{[t_0,t_{n+1}]}$
- revise the weights recursively to obtain $w_{n+1}^{(i)}$

so that $\{(\tau_{1:k_{n+1}}, k_{n+1})^{(i)}, w_{n+1}^{(i)}\}_{i=1}^N$ is a weighted sample from $\pi_{[t_0,t_{n+1}]}$.

SMC: Particle Fusion

Suppose after observing a continuous time process for some initial period we obtain a sample of size N, $\{(\tau_{1:k_1}, k_1)^{(i)}\}_{i=1}^{N}$, where $(\tau_{1:k_1}, k_1)^{(i)} \sim \pi_{[t_0,t_1]}(\tau_{1:k_1}, k_1)$.

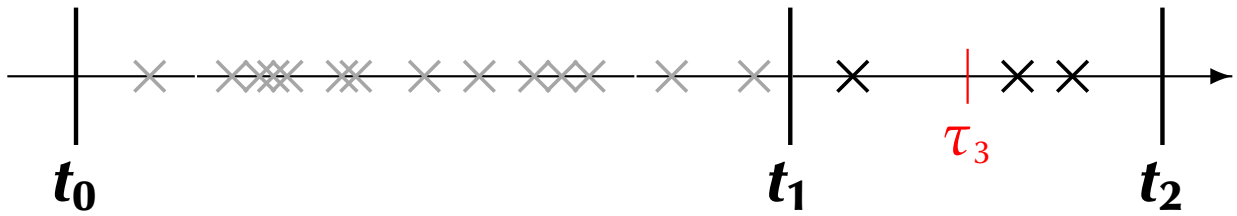


The process is further observed up till t_2.

If we consider the distribution of changepoints in $(t_1, t_2]$ independently from the distribution over $[t_0, t_1]$, then another sample of size M, $\{(\tau_{k_1+1:k_2}, k)^{(i)}\}_{i=1}^{M}$ can be drawn from $\pi_{(t_1,t_2]}(\tau_{k_1+1:k_2}, k)$, where $k = k_2 - k_1$ is the number of changepoints in $(t_1, t_2]$.

SMC: Particle Fusion

Suppose after observing a continuous time process for some initial period we obtain a sample of size N, $\{(\tau_{1:k_1}, k_1)^{(i)}\}_{i=1}^{N}$, where $(\tau_{1:k_1}, k_1)^{(i)} \sim \pi_{[t_0,t_1]}(\tau_{1:k_1}, k_1)$.



The process is further observed up till t_2.

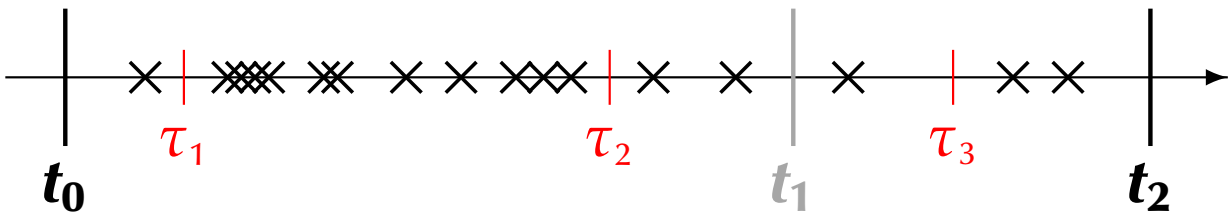
If we consider the distribution of changepoints in $(t_1, t_2]$ independently from the distribution over $[t_0, t_1]$, then another sample of size M, $\{(\tau_{k_1+1:k_2}, k)^{(i)}\}_{i=1}^{M}$ can be drawn from $\pi_{(t_1,t_2]}(\tau_{k_1+1:k_2}, k)$, where $k = k_2 - k_1$ is the number of changepoints in $(t_1, t_2]$.

The sample size for the new interval can be chosen so that $M \leq N$; if $M < N$, we create replicates of that sample so that there are equal size samples from the posteriors on each interval.

Joining the samples drawn from $\pi_{[t_0,t_1]}(\tau_{1:k_1}, k_1)$ and $\pi_{(t_1,t_2]}(\tau_{k_1+1:k_2}, k)$ will give an approximate sample

$$\{(\tau_{1:k_2}, k_2)^{(i)} = ((\tau_{1:k_1}, k_1)^{(i)}, (\tau_{k_1+1:k_2}, k)^{(1+\mathrm{mod}(i-1,M))})\}_{i=1}^{N}$$

from $\pi_{[t_0,t_2]}(\tau_{1:k_2}, k_2)$ under an assumption of independence between the two intervals.



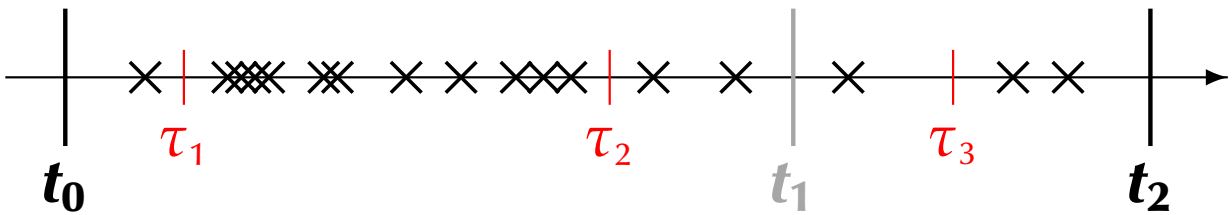
Having $M < N$ will decrease computational effort.

The sample size for the new interval can be chosen so that $M \leq N$; if $M < N$, we create replicates of that sample so that there are equal size samples from the posteriors on each interval.

Joining the samples drawn from $\pi_{[t_0,t_1]}(\tau_{1:k_1}, k_1)$ and $\pi_{(t_1,t_2]}(\tau_{k_1+1:k_2}, k)$ will give an approximate sample

$$\{(\tau_{1:k_2}, k_2)^{(i)} = ((\tau_{1:k_1}, k_1)^{(i)}, (\tau_{k_1+1:k_2}, k)^{(1+\mathrm{mod}(i-1,M))})\}_{i=1}^{N}$$

from $\pi_{[t_0,t_2]}(\tau_{1:k_2}, k_2)$ under an assumption of independence between the two intervals.



Having $M < N$ will decrease computational effort.

This assumption of independence on each new interval observed over time provides a proposal/importance distribution at t_n which we know up to proportionality

$$q_n(\tau_{1:k_n}, k_n) = \prod_{j=1}^{n} \gamma_{(t_{j-1},t_j]}(\tau_{k_{j-1}+1:k_j}, k) = q_{n-1}(\tau_{1:k_{n-1}}, k_{n-1})\gamma_{(t_{n-1},t_n]}(\tau_{k_{n-1}+1:k_n}, k).$$

The unnormalised importance weights given $\{(\tau_{1:k_n}, k_n)^{(i)}\}_{i=1}^{N}$ can then be computed sequentially. For $n \geq 2$

$$w_n^{(i)} = \frac{\gamma_{[t_0,t_n]}((\tau_{1:k_n}, k_n)^{(i)})}{q_n((\tau_{1:k_n}, k_n)^{(i)})}$$

$$= w_{n-1}^{(i)} \frac{\gamma_{[t_0,t_n]}((\tau_{1:k_n}, k_n)^{(i)})}{\gamma_{[t_0,t_{n-1}]}((\tau_{1:k_{n-1}}, k_{n-1})^{(i)})\gamma_{(t_{n-1},t_n]}((\tau_{k_{n-1}+1:k_n}, k)^{(1+\mathrm{mod}(i-1,M))})}$$

and $w_1^{(i)} = 1$.

This assumption of independence on each new interval observed over time provides a proposal/importance distribution at t_n which we know up to proportionality

$$q_n(\tau_{1:k_n}, k_n) = \prod_{j=1}^{n} \gamma_{(t_{j-1},t_j]}(\tau_{k_{j-1}+1:k_j}, k) = q_{n-1}(\tau_{1:k_{n-1}}, k_{n-1})\gamma_{(t_{n-1},t_n]}(\tau_{k_{n-1}+1:k_n}, k).$$

The unnormalised importance weights given $\{(\tau_{1:k_n}, k_n)^{(i)}\}_{i=1}^{N}$ can then be computed sequentially. For $n \geq 2$

$$\begin{aligned} w_n^{(i)} &= \frac{\gamma_{[t_0,t_n]}((\tau_{1:k_n}, k_n)^{(i)})}{q_n((\tau_{1:k_n}, k_n)^{(i)})} \\ &= w_{n-1}^{(i)} \frac{\gamma_{[t_0,t_n]}((\tau_{1:k_n}, k_n)^{(i)})}{\gamma_{[t_0,t_{n-1}]}((\tau_{1:k_{n-1}}, k_{n-1})^{(i)})\gamma_{(t_{n-1},t_n]}((\tau_{k_{n-1}+1:k_n}, k)^{(1+\mathrm{mod}(i-1,M))})} \end{aligned}$$

and $w_1^{(i)} = 1$.

Remarks

- The weights are quick to calculate, equivalent to a RJMCMC death move at t_{n-1}.
- Preferably the interval $[t_n, t_{n+1}]$ would be small enough that *a priori* we will only have 0 or 1 changepoints i.e. $\Pr(1 \text{ changepoint}) \approx \nu(t_{n+1} - t_n)$
  - MCMC sampling from $[t_n, t_{n+1}]$ then very straightforward
  - Allows for a smaller M, the size of the samples drawn from each interval.
- Standard resampling techniques can be incorporated into the algorithm to deal with weight degeneracy

Remarks

- The weights are quick to calculate, equivalent to a RJMCMC death move at t_{n-1}.
- Preferably the interval $[t_n, t_{n+1}]$ would be small enough that *a priori* we will only have 0 or 1 changepoints i.e. Pr(1 changepoint) $\approx \nu(t_{n+1} - t_n)$
  - MCMC sampling from $[t_n, t_{n+1}]$ then very straightforward
  - Allows for a smaller M, the size of the samples drawn from each interval.
- Standard resampling techniques can be incorporated into the algorithm to deal with weight degeneracy

Remarks

- The weights are quick to calculate, equivalent to a RJMCMC death move at t_{n-1}.
- Preferably the interval $[t_n, t_{n+1}]$ would be small enough that *a priori* we will only have 0 or 1 changepoints i.e. Pr(1 changepoint) $\approx \nu(t_{n+1} - t_n)$
  - MCMC sampling from $[t_n, t_{n+1}]$ then very straightforward
  - Allows for a smaller M, the size of the samples drawn from each interval.
- Standard resampling techniques can be incorporated into the algorithm to deal with weight degeneracy

Limitation

When extending the particles for $(t_{n-1}, t_n]$ the independence assumption prevents a genuine changepoint near t_{n-1} being properly explored.

Time of last changepoint

At t_{n-1}, let t^*_{n-1} be the time of the most recent changepoint. The posterior expectation of t^*_{n-1} can then be calculated using the particle approximation $\{(\tau_{1:k_{n-1}}, k_{n-1})^{(i)}, w^{(i)}_{n-1}\}_{i=1}^N$ for $\pi_{[t_0,t_{n-1}]}$

$$\mathbb{E}_{\pi_{[t_0,t_{n-1}]}}[t^*_{n-1}] \approx \sum_{i=1}^{N} \tau^{(i)}_{k_{n-1}} W^{(i)}_{n-1}$$

where $W^{(i)}_{n-1}$ are the normalised weights.

When sampling changepoints for the interval $(t_{n-1}, t_n]$, use the data from $(t^*_{n-1}, t_n]$. However, only sample changepoints in $(t_{n-1}, t_n]$.

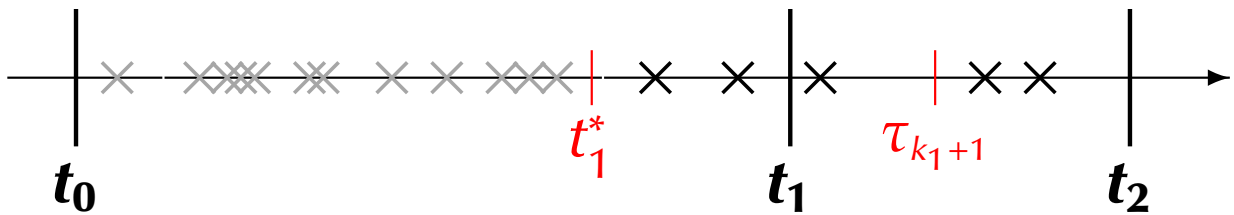
Time of last changepoint

At t_{n-1}, let t^*_{n-1} be the time of the most recent changepoint. The posterior expectation of t^*_{n-1} can then be calculated using the particle approximation $\{(\tau_{1:k_{n-1}}, k_{n-1})^{(i)}, w^{(i)}_{n-1}\}_{i=1}^N$ for $\pi_{[t_0,t_{n-1}]}$

$$\mathbb{E}_{\pi_{[t_0,t_{n-1}]}}[t^*_{n-1}] \approx \sum_{i=1}^{N} \tau^{(i)}_{k_{n-1}} W^{(i)}_{n-1}$$

where $W^{(i)}_{n-1}$ are the normalised weights.

When sampling changepoints for the interval $(t_{n-1}, t_n]$, use the data from $(t^*_{n-1}, t_n]$. However, only sample changepoints in $(t_{n-1}, t_n]$.



Function of interest

As a measure of anomaly at time t a natural function of interest $g(t)$ would be distance to the nearest changepoint to the left

$$g(t) = t - \tau_{i^*}$$
$$i^* = \max_i \{\tau_i \leq t\}$$

For $t > 0$ the prior expectation of $g(t)$ is non-zero and increasing with t, so standardising gives a revised function of interest

$$h(t) = \frac{g(t) - \mathbb{E}[g(t)]}{\sqrt{\mathbb{V}\text{ar}[g(t)]}}$$

Function of interest

As a measure of anomaly at time t a natural function of interest $g(t)$ would be distance to the nearest changepoint to the left

$$g(t) = t - \tau_{i^*}$$
$$i^* = \max_i \{\tau_i \leq t\}$$

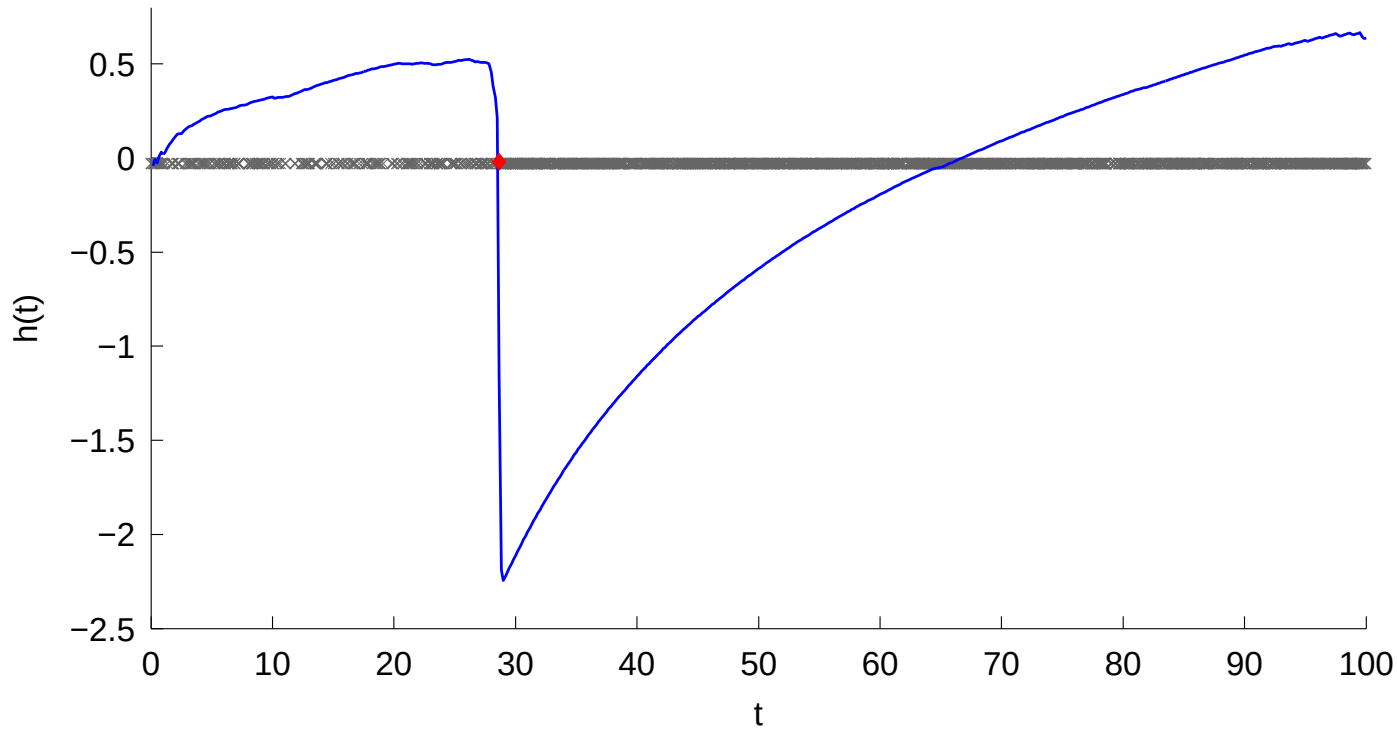
For $t > 0$ the prior expectation of $g(t)$ is non-zero and increasing with t, so standardising gives a revised function of interest

$$h(t) = \frac{g(t) - \mathbb{E}\left[g(t)\right]}{\sqrt{\mathbb{V}\text{ar}\left[g(t)\right]}}$$

We could now classify an anomaly at t if, given the data,

$$\Pr(h(t) < 0) > \alpha$$

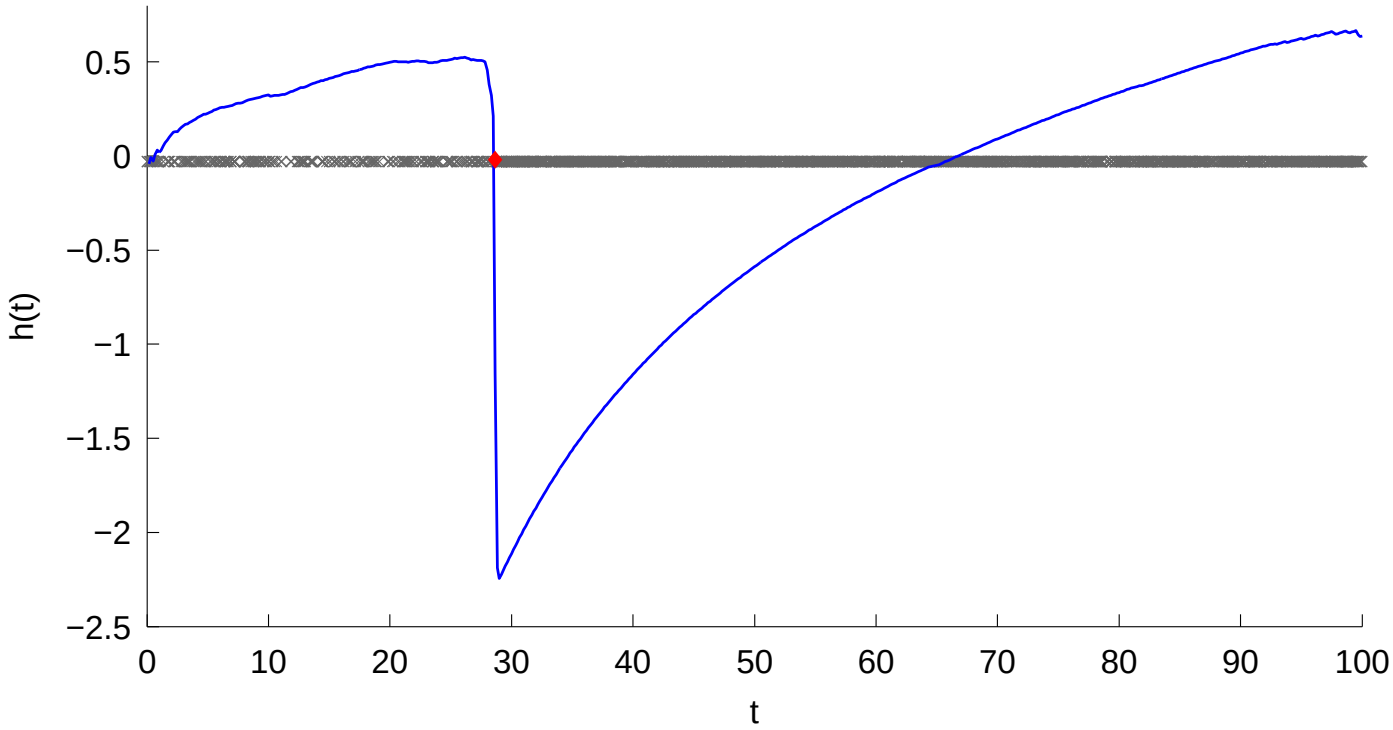
for some α, say .95.



We could now classify an anomaly at t if, given the data,

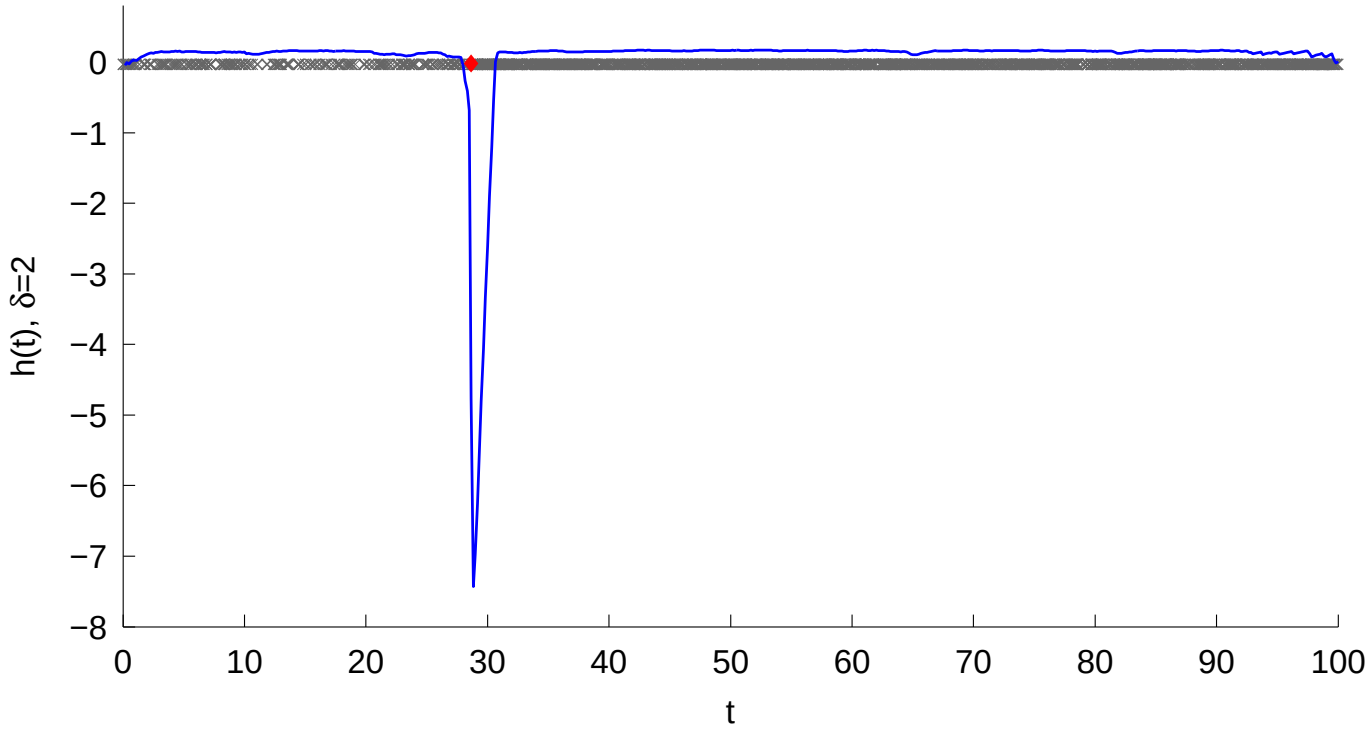
$$\Pr(h(t) < 0) > \alpha$$

for some α, say .95.



Alternatively, to control the length of time for which the change is considered anomalous, only look back over a time δ

$$g_\delta(t) = \min(t - \tau_{i^*}, \delta)$$



For any t satisfying $\Pr(h(t) < 0) > \alpha$, an estimate for the location of the most recent changepoint is then given by

$$\hat{\tau}_{i^*} = t - \mathbb{E}_{\pi}[g(t)].$$

We can now approximate $\Pr(h(t) < 0)$ given the weighted sample:

$$\Pr(h(t) < 0) \approx \sum_{i=1}^{N} \mathbb{I}(h^{(i)}(t) < 0) W_n^{(i)}$$

- If only interested in recent changepoints then consider calculating it for $t_{n-1} < t < t_n$ (online estimate)
- We might also choose to revise the estimates over $t_0 < t < t_n$ in light of the updated weights, this will give you a sequence of curves over time (restrospective estimate).

500 simulated Poisson process datasets each over 100 time units.

Total number of simulated changepoints = 983.

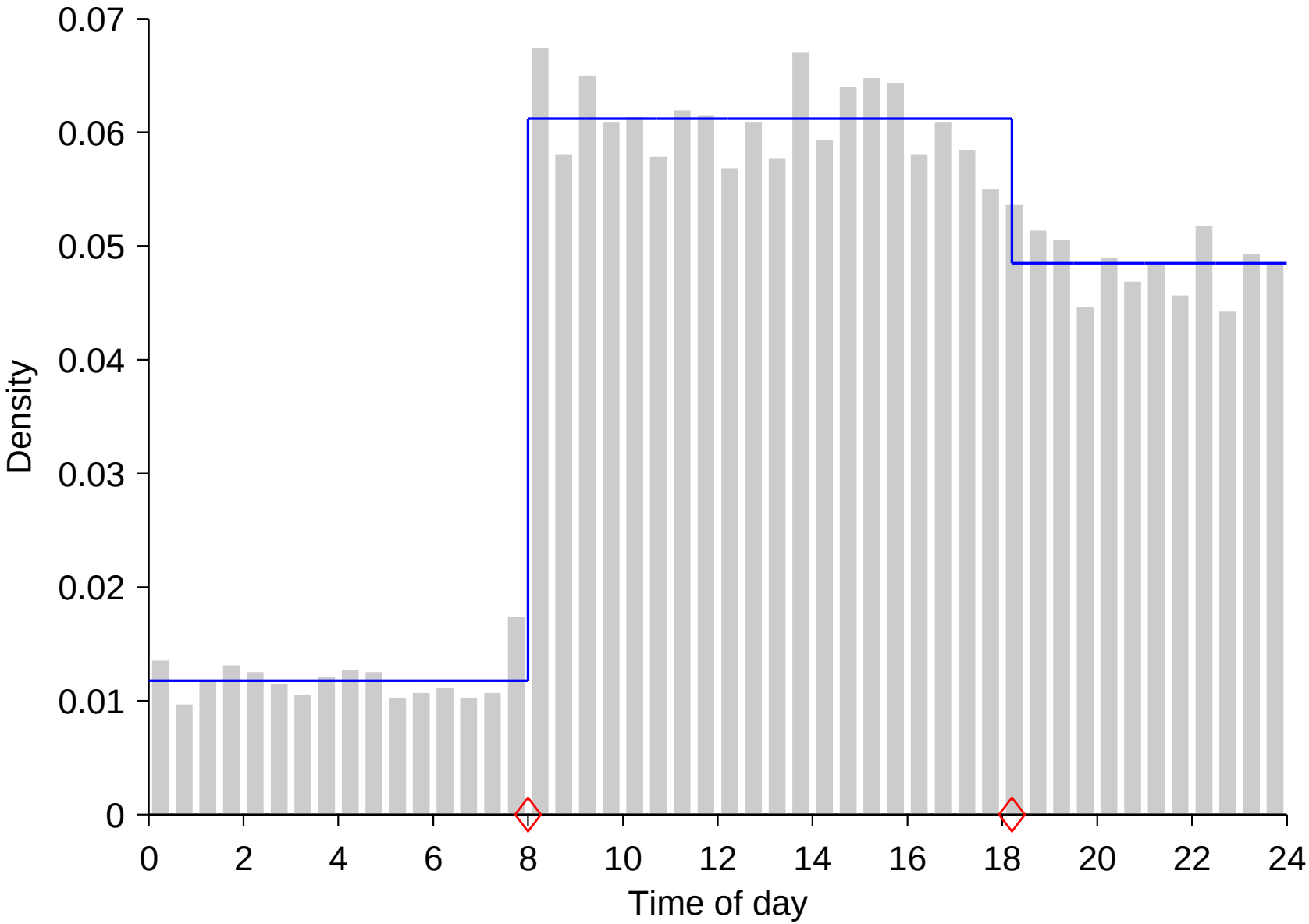
Using $N = 5{,}000$, $M = 500$; update windows of two.

	SMC			**MCMC**		
Significance level α	50%	70%	90%	50%	70%	90%
Detected changepoints $[\Pr(h_{\delta=2}(t) < 0) > \alpha]$	813	777	718	819	790	720
False detections	50	12	0	44	12	0
Average update intervals till detection	0.70	0.79	1.08	0.70	0.79	1.11
Run times	15min 40secs			32min 45secs		

VAST Data

Synthetic data set of phone calls made between 400 individuals on a fictitious island over a ten day period.
(`http://www.cs.umd.edu/hcil/VASTchallenge08`)

- Anomalous activity is known to occur on the start of the eighth day involving 11 individuals
- Daily cyclical effects present in data


0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
Density
0
2
4
6
8
10
12
14
16
18
20
22
24
Time of day

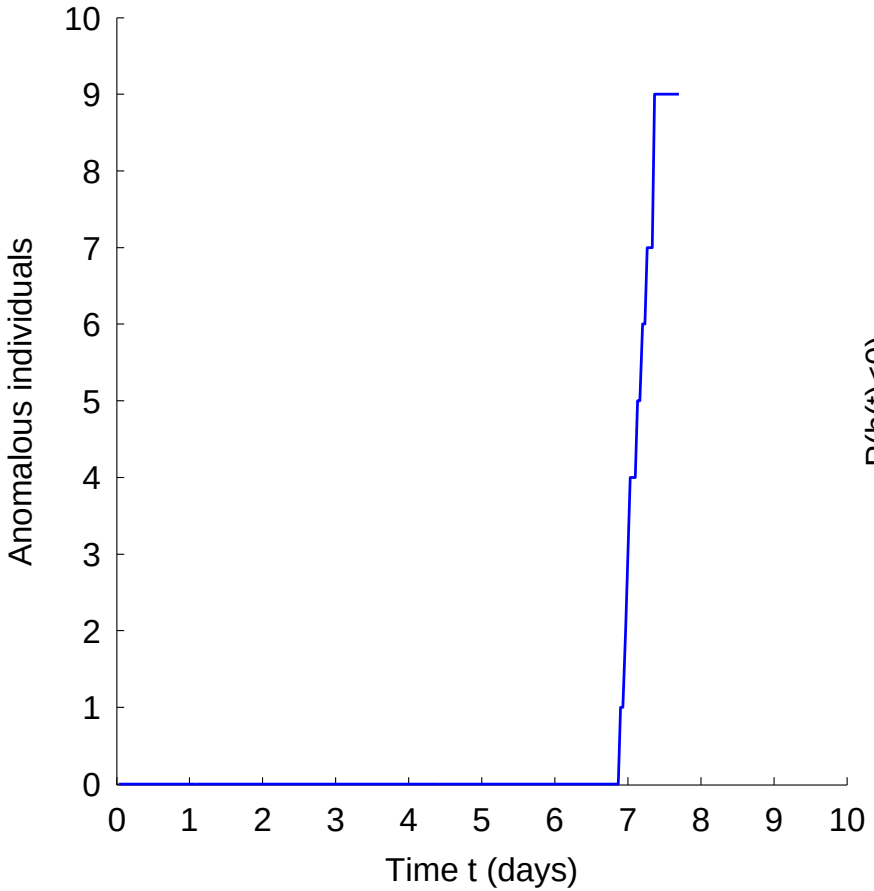
Seasonal component

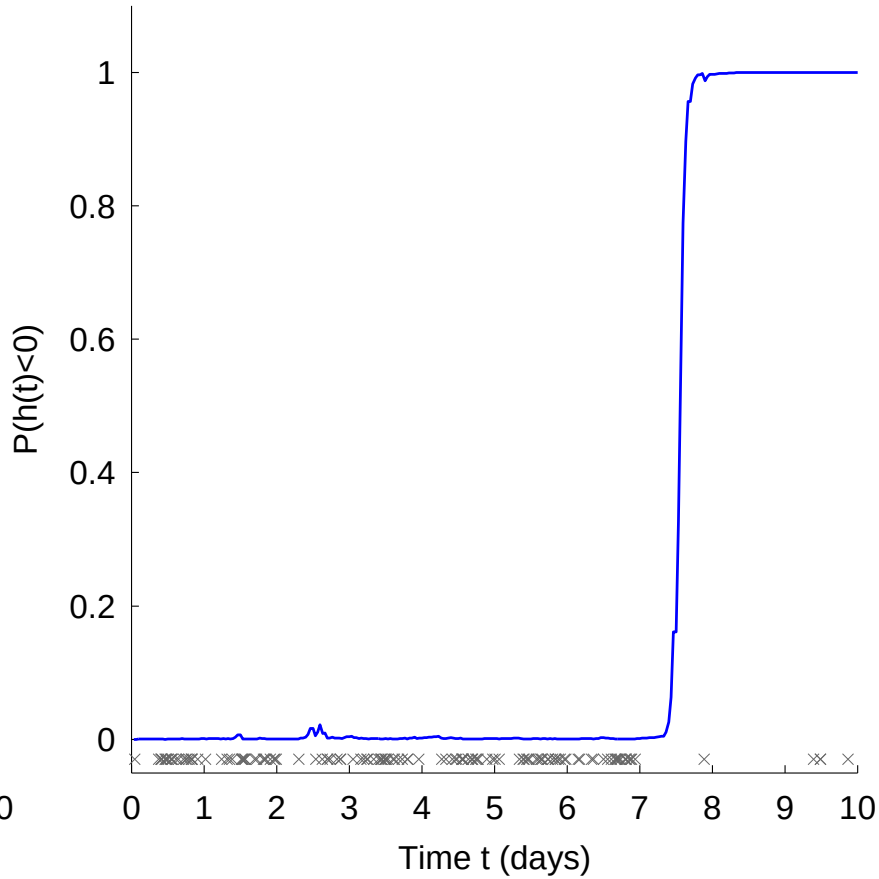
- Ignoring seasonality would lead to lots of "false" changepoints
- Use MCMC to identify the seasonal changepoints and corresponding intensities given the data across all individuals
- Globally apply a set of seasonal changepoints $s = (s_1, \ldots, s_l)$ and intensity multipliers $\mu = (\mu_0, \ldots, \mu_l)$ so that the intensity at any time will be $\lambda_i \mu_j$ for some i, j. For identifiability, w.l.o.g. set $\mu_0 = 1$
- Can integrate out λ as before so that our target distribution is $\pi_{[t_0,t_n]}(\tau_{1:k_n}, k_n | \mathcal{D}, s, \mu)$

Finding anomalous individuals

- Use the SMC algorithm with 1 hour time increments (240 update windows) with $N = 10{,}000$ and $M = 5{,}000$
- Declare an individual as anomalous at time t if $\Pr(h(t) < 0) > 0.95$
- At each time step t_n we can recalculate the Monte Carlo estimate of $\Pr(h(t) < 0)$ for all $0 < t < t_n$ in light of the updated SMC weights, to identify the individuals we currently regard as having behaved anomalously now or in the past

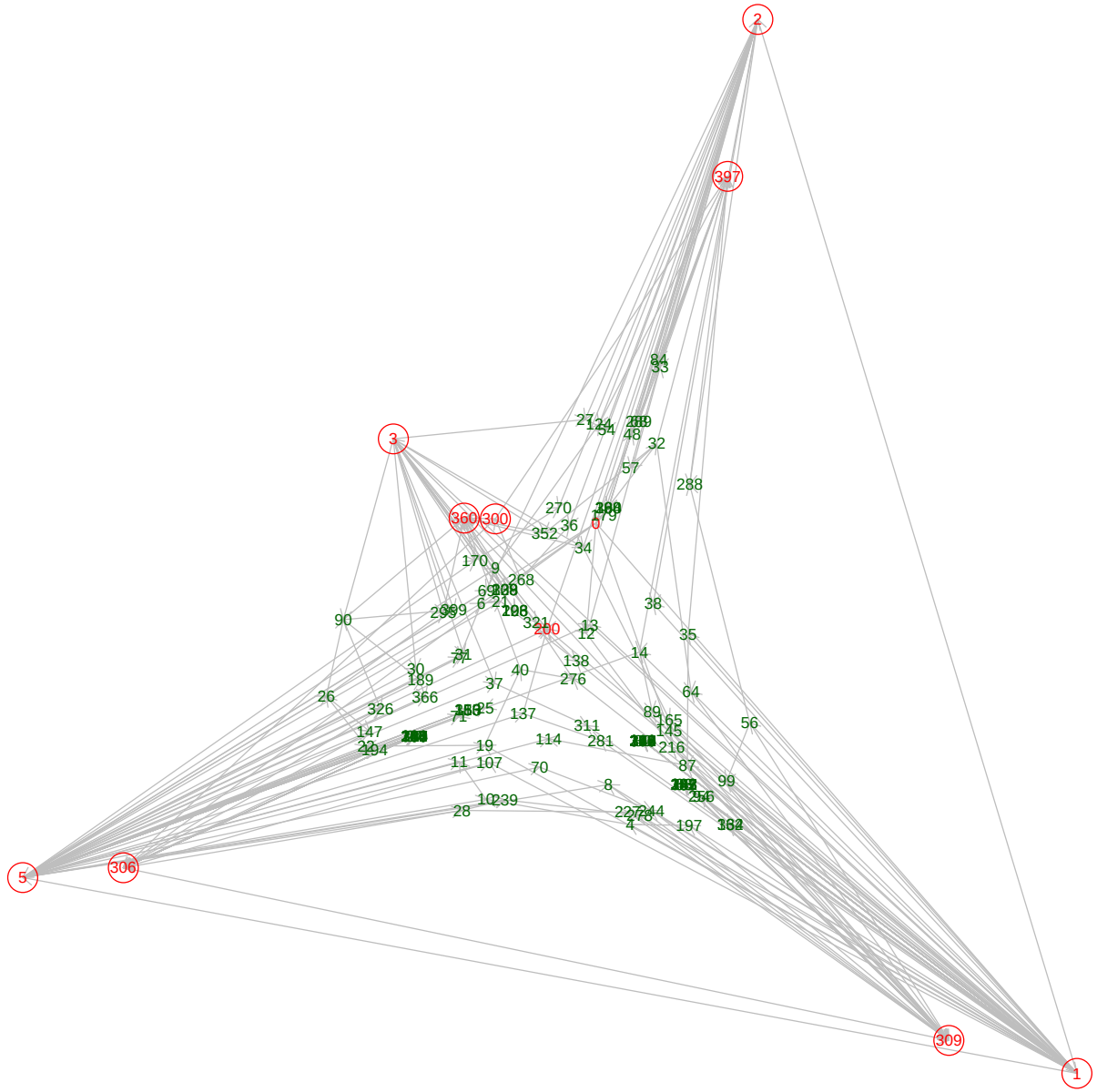
Most anomalous time period involves 9 individuals at the start of the eighth day.





- Each individual took 20 secs to run, which is approximately 1/10th of a second per update window.
- Completely parallelizable across all individuals in the network.

Calls between cell phones of nodes who were in contact with the anomalous individuals before and during the most anomalous time period.



Remarks

- Can extend method to non-conjugate models
- Not limited to poisson processes - have applied algorithm to a piecewise AR process and Markov Chains

Further Work

- Varying the number of particles sampled on each update window over time according to the complexity of the process
- If running more than one process allows you to allocate computational resources to processes that are “harder”

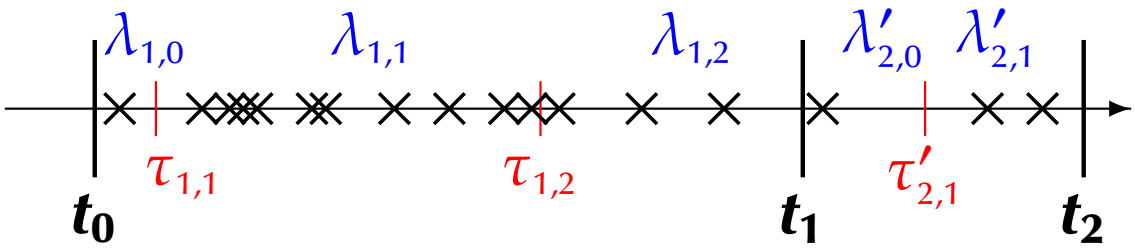
Bibliography

Heard, N. A., Weston, D. J., Platanioti, K. and Hand, D. J. (2010) Bayesian Anomaly Detection Methods for Social Networks. *Annals of Applied Statistics*.

Non-conjugate models

Suppose we have non-conjugate priors for the intensities. Then the posterior of interest is $\pi_{[t_0,t_n]}(\tau_n, \lambda_n)$ where $\lambda_n = (\lambda_{n,0}, \ldots, \lambda_{n,k})$.

Problem: Sample from the proposal distribution is over-parameterised.

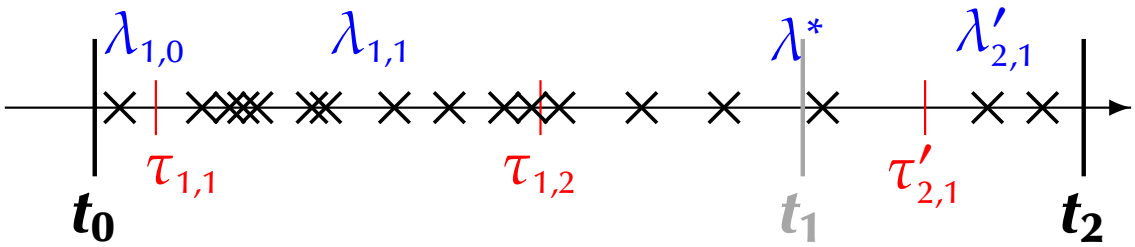


Need to replace the intensity pair $(\lambda_{1,2}, \lambda'_{2,0})$ with a single intensity λ^* to cover the interval $(\tau_{1,2}, \tau'_{2,1}]$.

Non-conjugate models

Suppose we have non-conjugate priors for the intensities. Then the posterior of interest is $\pi_{[t_0,t_n]}(\tau_n, \lambda_n)$ where $\lambda_n = (\lambda_{n,0}, \ldots, \lambda_{n,k})$.

Problem: Sample from the proposal distribution is over-parameterised.



Need to replace the intensity pair $(\lambda_{1,2}, \lambda'_{2,0})$ with a single intensity λ^* to cover the interval $(\tau_{1,2}, \tau'_{2,1}]$.

At t_n let $s_1(\lambda_{n-1,k}, \lambda'_{n,0})$ be a suitably chosen function to combine the intensities. For example,

$$\lambda^* = s_1(\lambda_{n-1,k}, \lambda'_{n,0}) = \frac{(t_{n-1} - \tau_{n-1,k})\lambda_{n-1,k} + (\tau'_{n,1} - t_{n-1})\lambda'_{n,0}}{\tau'_{n,1} - \tau_{n-1,k}}.$$

Marginal distribution of λ^* is unlikely to be analytically available, instead must settle for a joint change of variable

$$(\lambda^*, u_n) = s(\lambda_{n-1,k}, \lambda'_{n,0}) = (s_1(\lambda_{n-1,k}, \lambda'_{n,0}), s_2(\lambda_{n-1,k}, \lambda'_{n,0}))$$

where s_2 is some other transformation of $(\lambda_{n-1,k}, \lambda'_{n,0})$ so that we have a one to one mapping $(\lambda_{n-1,k}, \lambda'_{n,0}) \mapsto (\lambda^*, u_n)$.

Let $|J_n|$ be the Jacobian for the transformation s.

We now have a proposal distribution $q_n(\tau_n, \lambda_n, u_{2:n})$ which generates suitable intensities, λ^*, for each of the combined regions but also the nuisance parameters $u_{2:n}$

$$q_n(\tau_n, \lambda_n, u_{2:n}) = \prod_{j=1}^{n} \gamma_{(t_{j-1},t_j]}(\tau'_j, \lambda'_j)|J_j|$$
$$= q_{n-1}(\tau_{n-1}, \lambda_{n-1}, u_{2:n-1})\gamma_{(t_{n-1},t_n]}(\tau'_n, \lambda'_n)|J_n|.$$

Proposal distribution is still of too high a dimension.

Solution: Extend the target distribution.

$$\pi_{[t_0,t_n]}(\tau_n, \lambda_n, u_{2:n}) = \pi_{[t_0,t_n]}(\tau_n, \lambda_n) \prod_{i=2}^{n} f(u_i|\tau_i, \lambda_i)$$

where f can be *any* density with the correct support for u_i.

As the target $\pi_{[t_0,t_n]}(\tau_n, \lambda_n)$ is a marginal of $\pi_{[t_0,t_n]}(\tau_n, \lambda_n, u_{2:n})$ we can use IS to obtain estimates from this distribution where the importance weights are now expressed as

$$w_n^{(i)} = w_{n-1}^{(i)} \frac{\gamma_{[t_0,t_n]}(\tau_n^{(i)}, \lambda_n^{(i)}) f(u_n|\tau_n, \lambda_n)}{\gamma_{[t_0,t_{n-1}]}(\tau_{n-1}^{(i)}, \lambda_{n-1}^{(i)}) \gamma_{(t_{n-1},t_n]}(\tau_n^{\prime(1+\mathrm{mod}(i-1,M))}, \lambda_n^{\prime(1+\mathrm{mod}(i-1,M))}) |\mathbb{J}_n|}.$$

Solution: Extend the target distribution.

$$\pi_{[t_0,t_n]}(\tau_n, \lambda_n, u_{2:n}) = \pi_{[t_0,t_n]}(\tau_n, \lambda_n) \prod_{i=2}^{n} f(u_i|\tau_i, \lambda_i)$$

where f can be *any* density with the correct support for u_i.

As the target $\pi_{[t_0,t_n]}(\tau_n, \lambda_n)$ is a marginal of $\pi_{[t_0,t_n]}(\tau_n, \lambda_n, u_{2:n})$ we can use IS to obtain estimates from this distribution where the importance weights are now expressed as

$$w_n^{(i)} = w_{n-1}^{(i)} \frac{\gamma_{[t_0,t_n]}(\tau_n^{(i)}, \lambda_n^{(i)}) f(u_n|\tau_n, \lambda_n)}{\gamma_{[t_0,t_{n-1}]}(\tau_{n-1}^{(i)}, \lambda_{n-1}^{(i)}) \gamma_{(t_{n-1},t_n]}(\tau_n'^{(1+\mathrm{mod}(i-1,M))}, \lambda_n'^{(1+\mathrm{mod}(i-1,M))}) |\mathbb{J}_n|}.$$

Choosing s

The parameter transformation $s = (s_1, s_2)$ should be chosen so that if $\lambda_{n-1,k}$ and $\lambda'_{n,0}$ are samples from their own conditional posterior distributions:

- $\lambda^* = s_1(\lambda_{n-1,k}, \lambda'_{n,0})$ should be close to a draw from the posterior for the intensity on the joined segment.
- $u_n = s_2(\lambda_{n-1,k}, \lambda'_{n,0})$ has a distribution which we might be able to loosely identify and can guide how to extend the target.

For example, $u_n = \lambda_{n-1,k} - \lambda'_{n,0}$ so that the domain of u_n is $\mathbb{R}$ and should be near zero if the merger is a good match.

Then $f(u_n|\tau_n, \lambda_n) = \phi(u_n)$, the density of a standard normal could be a good choice.

Choosing s

The parameter transformation $s = (s_1, s_2)$ should be chosen so that if $\lambda_{n-1,k}$ and $\lambda'_{n,0}$ are samples from their own conditional posterior distributions:

- $\lambda^* = s_1(\lambda_{n-1,k}, \lambda'_{n,0})$ should be close to a draw from the posterior for the intensity on the joined segment.
- $u_n = s_2(\lambda_{n-1,k}, \lambda'_{n,0})$ has a distribution which we might be able to loosely identify and can guide how to extend the target.

For example, $u_n = \lambda_{n-1,k} - \lambda'_{n,0}$ so that the domain of u_n is $\mathbb{R}$ and should be near zero if the merger is a good match.

Then $f(u_n|\tau_n, \lambda_n) = \phi(u_n)$, the density of a standard normal could be a good choice.