The Probability Hypothesis Density Filter

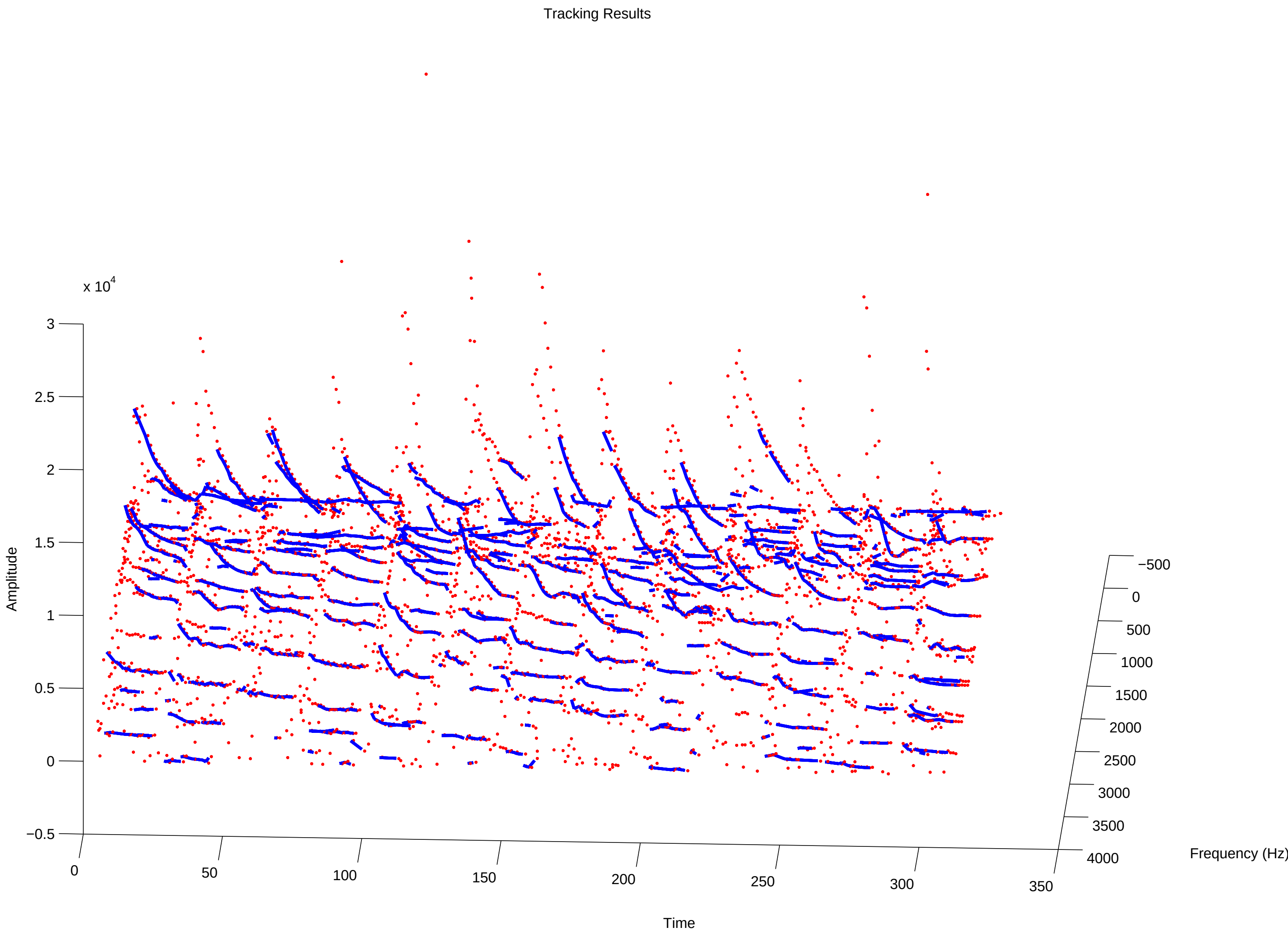
A. Taylan Cemgil

Signal Processing and Communications Lab.

UNIVERSITY OF CAMBRIDGE
Department of Engineering

7 December 2007
NIPS 07 Workshop
Approximate Bayesian Inference in Continuous/Hybrid Systems

Sinusoidal Tracking with Clark, Peeling and Godsill



Outline

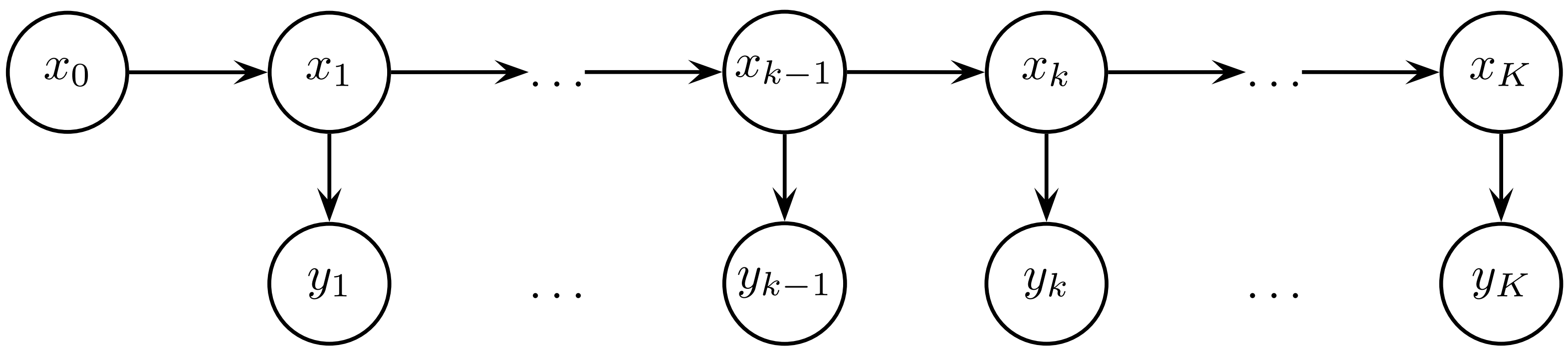
- Introduction to Multi Object Tracking
- The Probability Hypothesis Density Filter
  - A toy model
  - A short summary of point process theory
- Summary

Stochastic Dynamical System

- Generic Model

$$\begin{aligned} x_k &\sim p(x_k|x_{k-1}) && \text{Transition Model} \\ y_k &\sim p(y_k|x_k) && \text{Observation Model} \end{aligned}$$

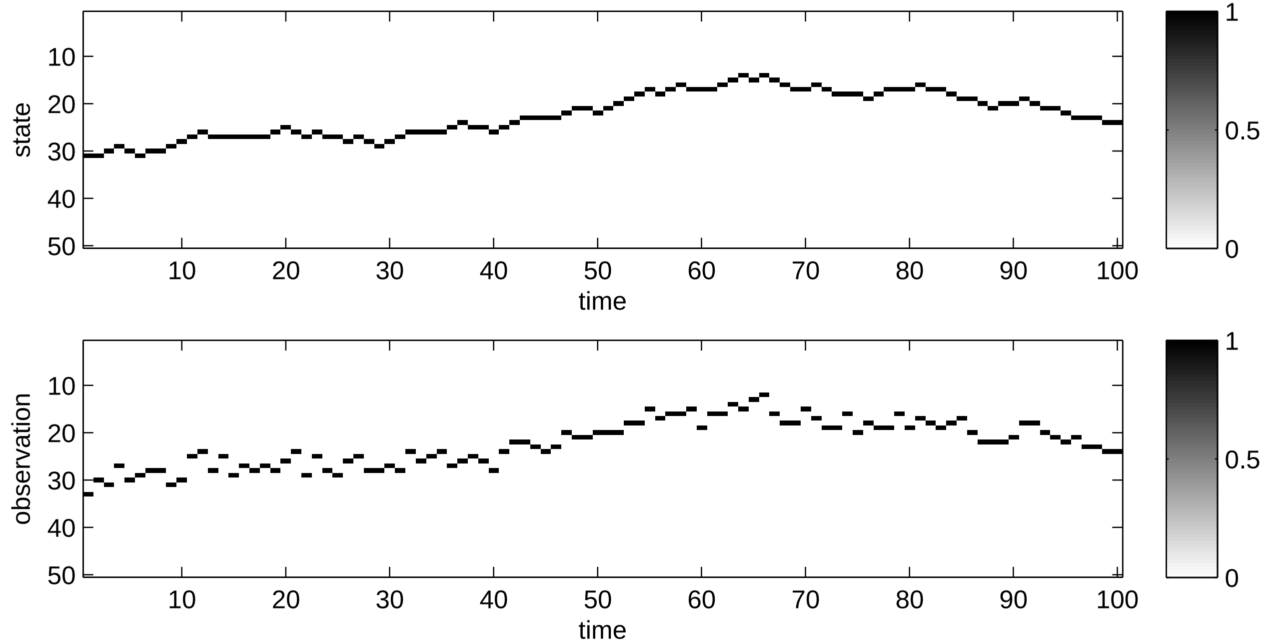
- Examples: Hidden Markov Model, Kalman filter

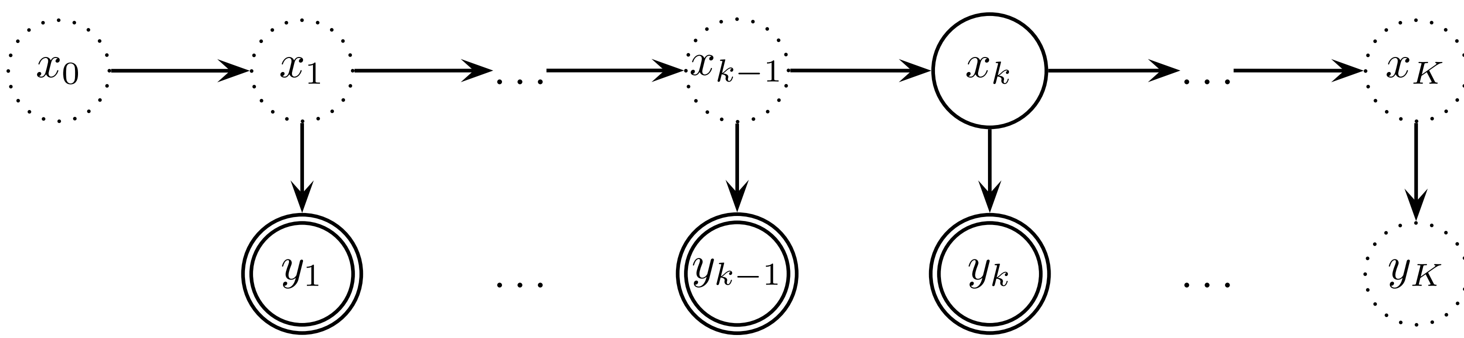


- Observations $y_k \in \mathcal{Y}$ are projections of the latent states $x_k \in$

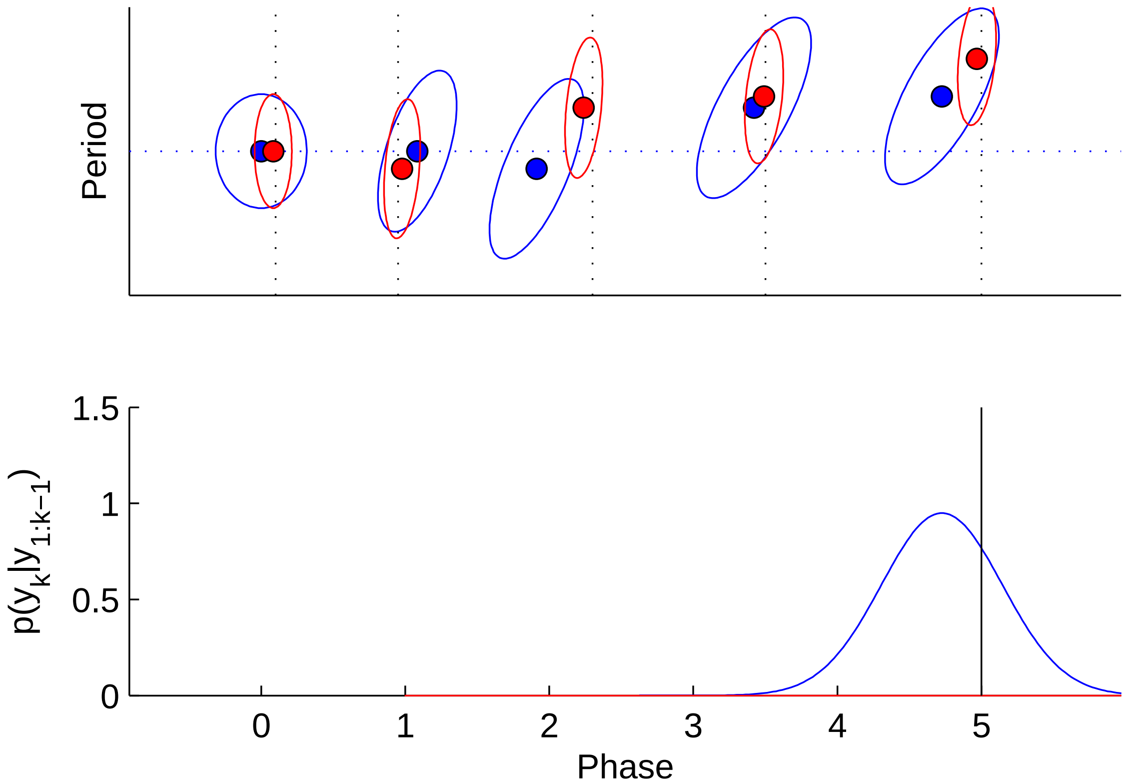
- Exact inference possible only for few cases

Tracking - Filtering



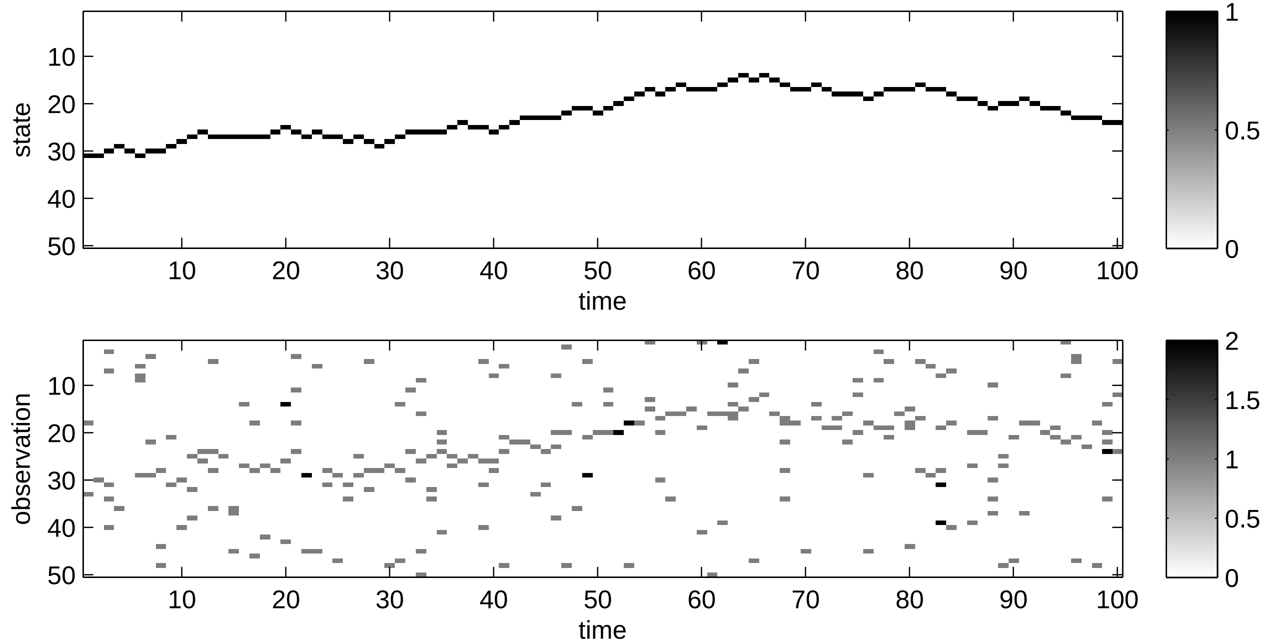


Tracking - Kalman Filtering



- Propagate exact sufficient statistics of the filtering density

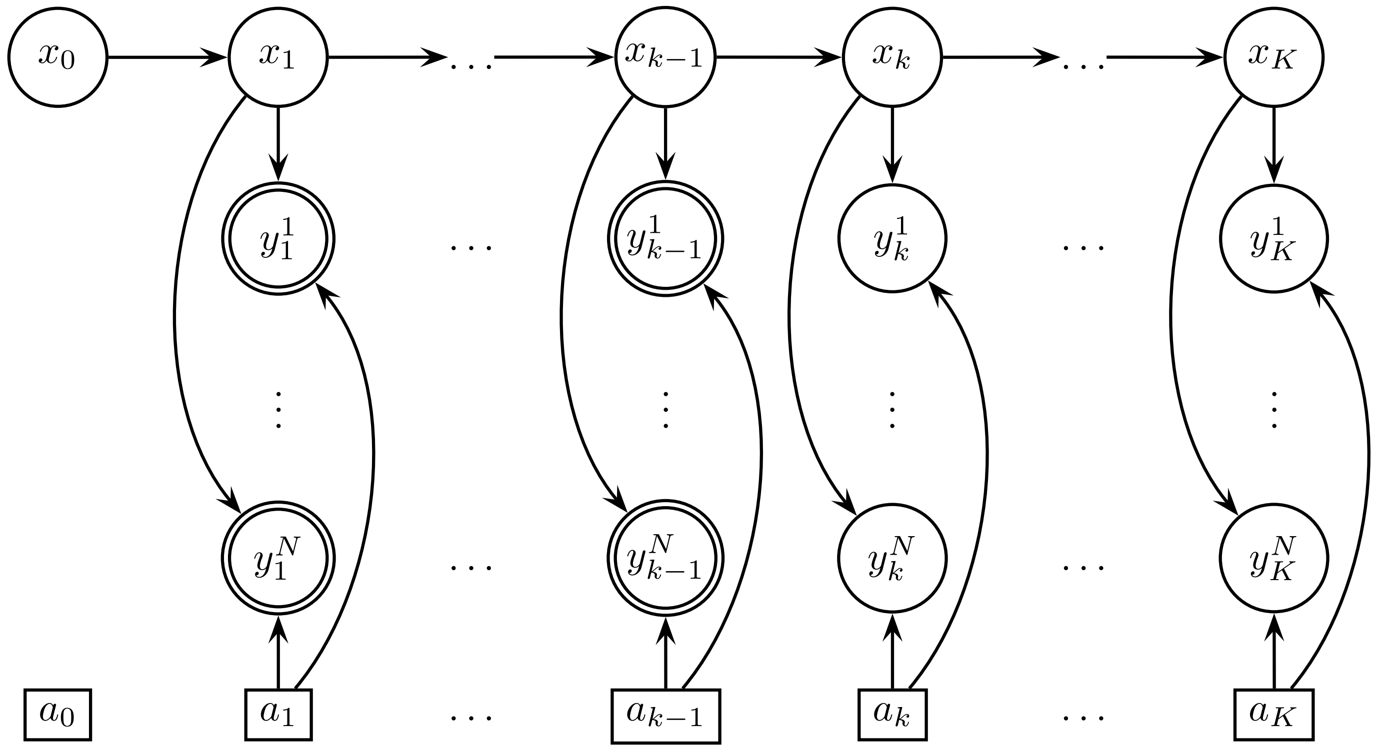
A harder scenario – clutter and missing detections



- At each time k, we observe n_k observations, at most one corresponding to the true target

- Have to solve the association problem – Combinatorial

A harder scenario – clutter and missing detections

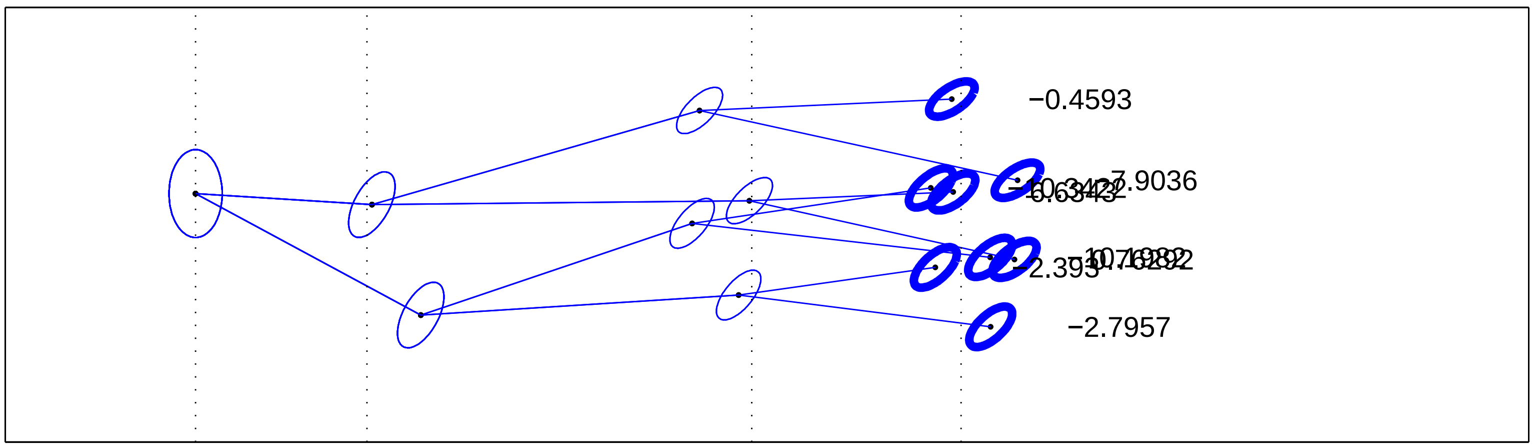


- The latent discrete switch variables a_k denote which of the observations is the true one

Inference in Switching State Space models

- Unlike HMM's or KFM's, summing over indicators a_k does not simplify the filtering density.

- Number of Gaussian kernels to represent exact filtering density increases exponentially

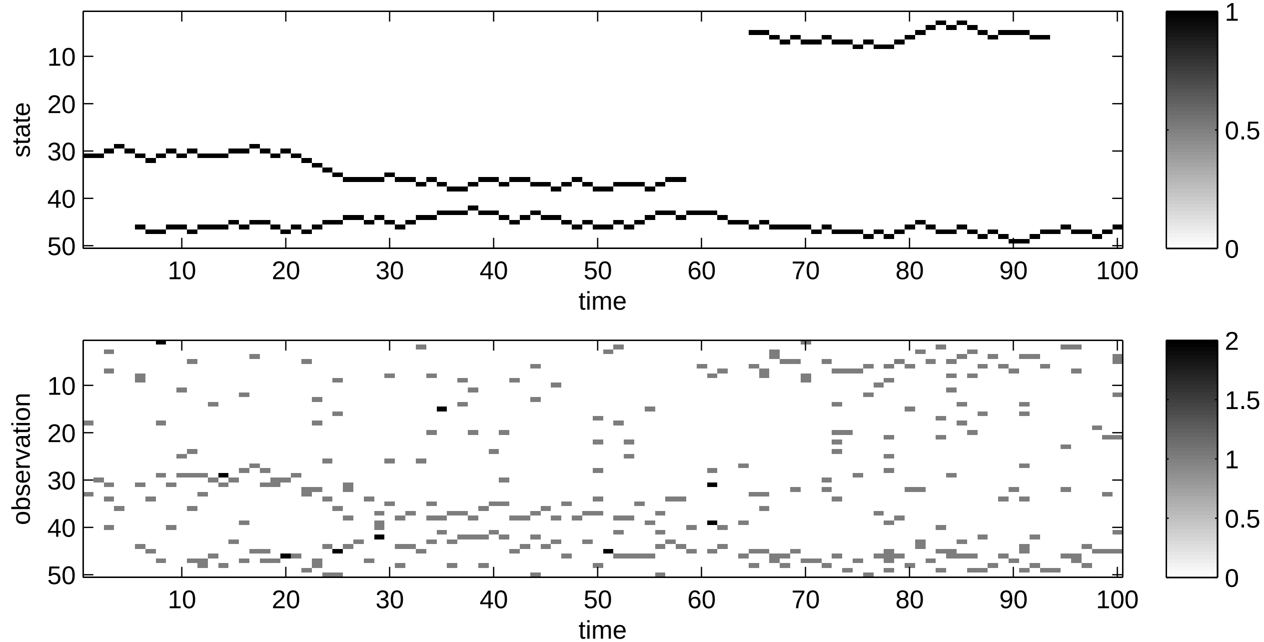


- Bad news: exact inference problem is shown to be NP hard

Approximate Inference

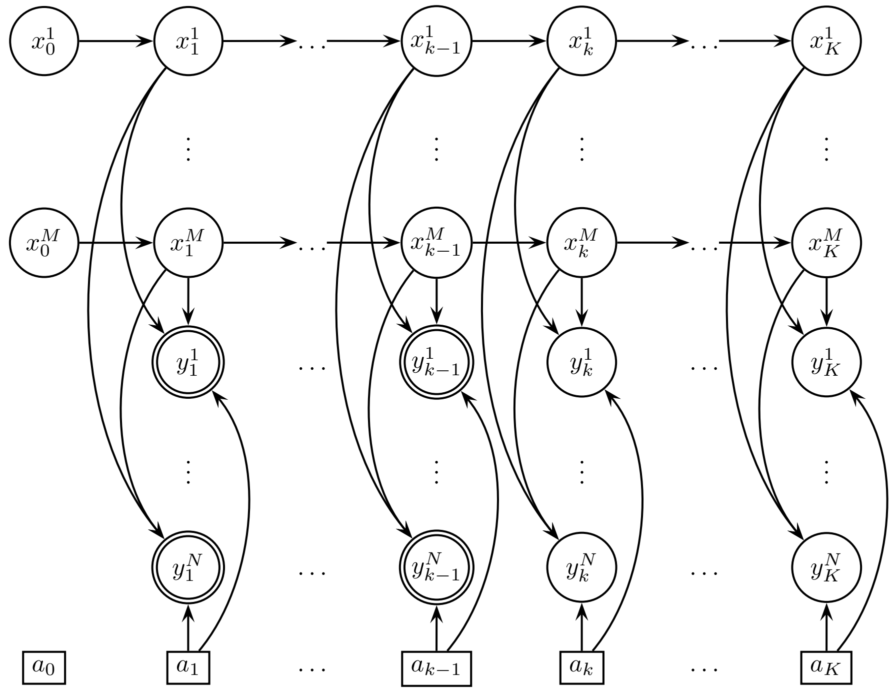
- Sequential Monte Carlo (Particle Filtering)
  - Sample branches with a probability propotional to the evidence, Mixture Kalman filter (Chen and Liu 2001)
- Deterministic Approximations
  - Assumed density filter (ADF) : Project the filtering density by moment matching onto a tractable family,
- See *"Bayesian inference in dynamic models – an overview"* by Tom Minka

An even harder scenario



- At each time slice, a chain can cease to exist, or a new chain is born
- Clutter and misdetections

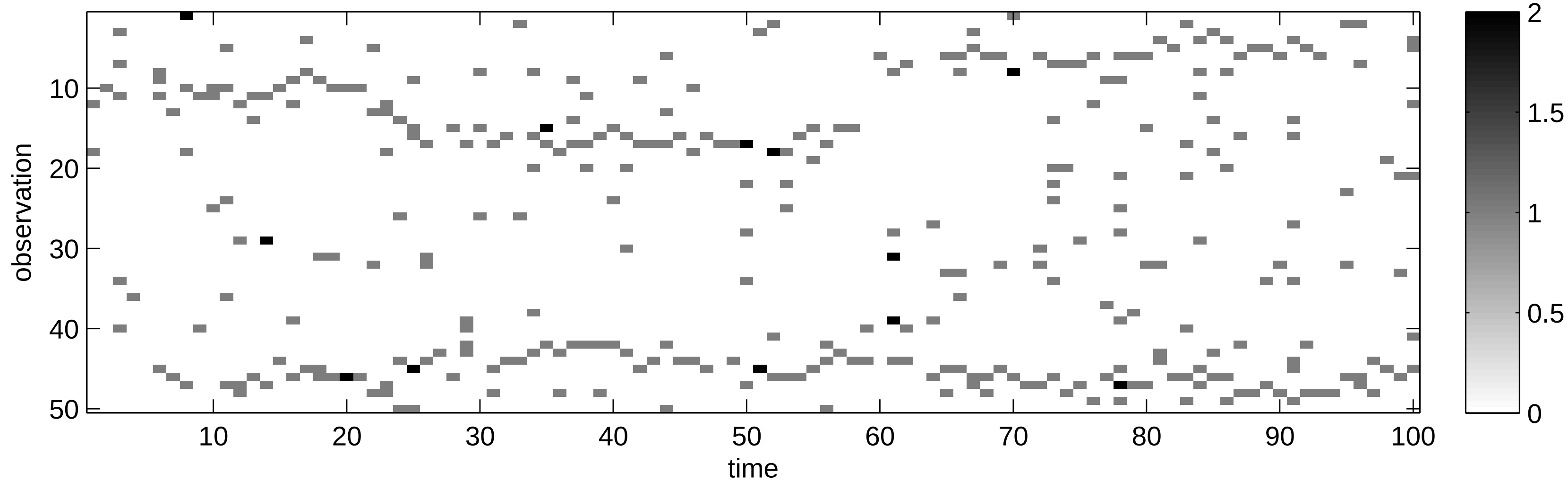
An even harder scenario



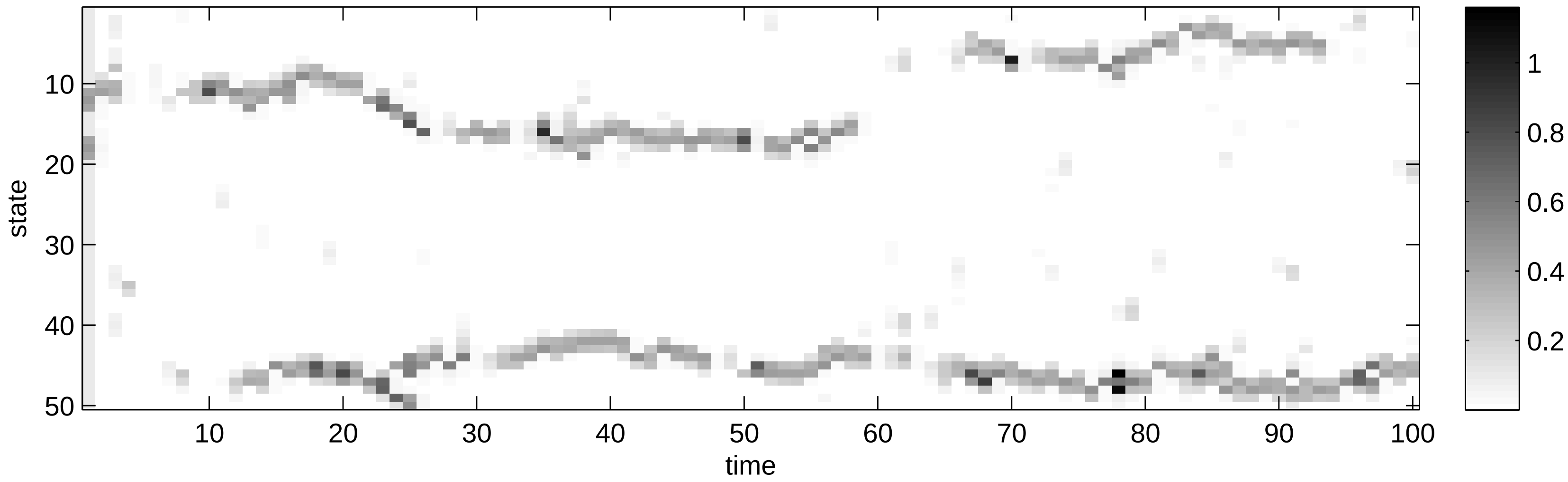
- Factorial dynamic model with changing number of latent chains (not modeled here)

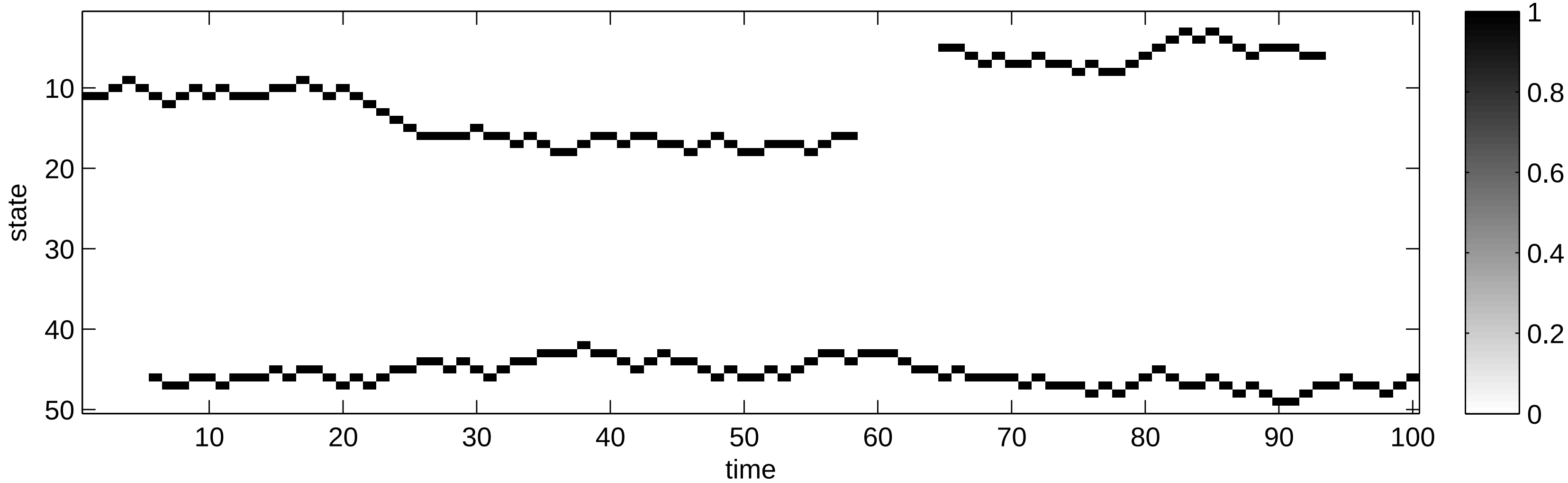
- Association problem: combinatorial explosion in the number of switch variables

- Multi-hypothesis tracker (MHT), Joint Probabilistic Data Association filter (JPDA), Harmonic analysis on finite groups (Kondor, Howard, Jebara (2007))




```
p_sur = 0.98;  % Survival probability
b = 0.01;        % Birth intensity
p_det = 0.5;  % Detection probability
c = 1/30;        % Clutter intensity
lam0 = 1;    % Prior inensity
```

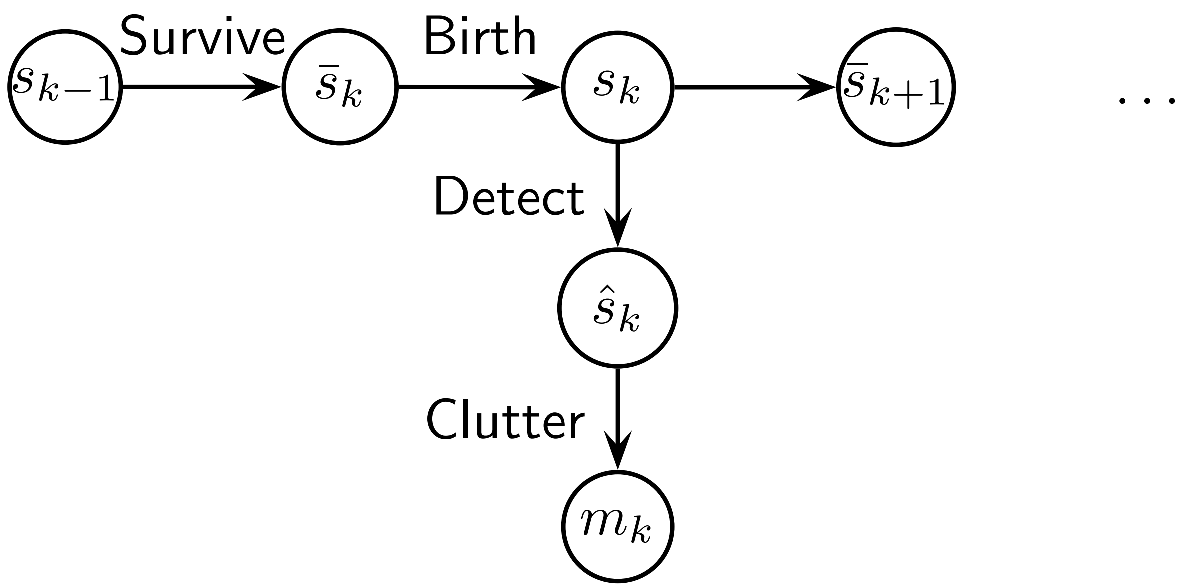

state
10
20
30
40
50
10
20
30
40
50
60
70
80
90
100
time
1
0.8
0.6
0.4
0.2



state
10
20
30
40
50
time
10
20
30
40
50
60
70
80
90
100
1
0.8
0.6
0.4
0.2
0

A simplified Model

- Suppose we just want to **track the number of objects** s_k



- We want to design a filter to track the number of objects, i.e. want to compute sequentially the filtering density $p(s_k|y_{1:k})$

Notation

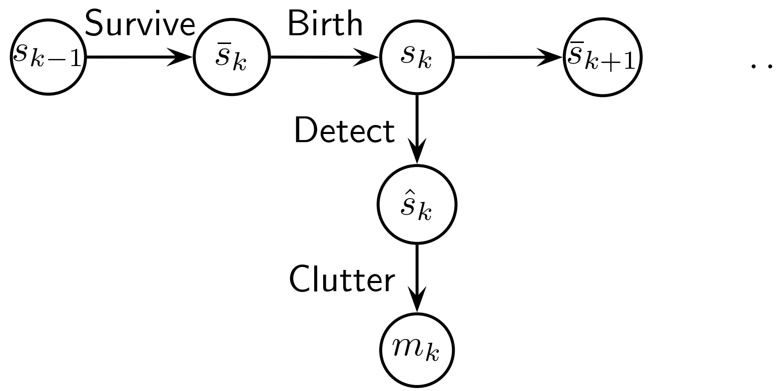
- Poisson Distribution

$$\mathcal{PO}(s;\lambda) = \frac{e^{-\lambda}}{s!}\lambda^s$$

- Binomial distribution – Number of succesful outcomes in n independent trials with success probability π

$$\mathcal{BI}(s;n,\pi) \quad = \quad \binom{n}{s}\pi^s(1-\pi)^{n-s}$$

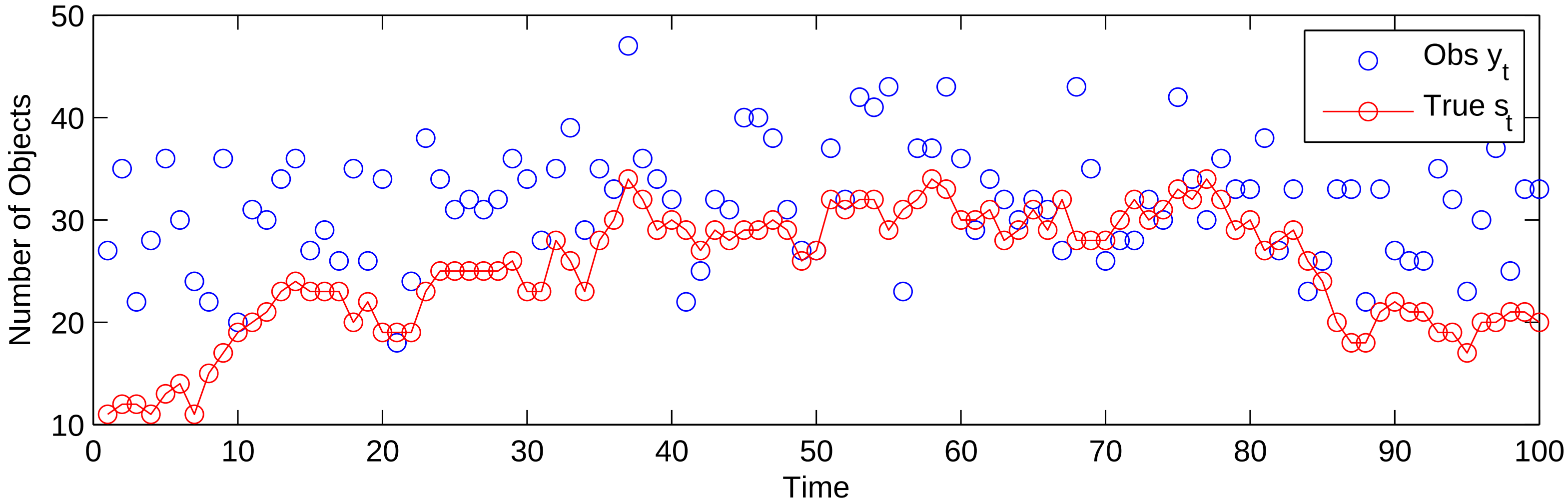
Basic Model



Survive	$\bar{s}_k \vert s_{k-1} \sim \mathcal{BI}(\bar{s}_k; s_{k-1}, \pi_{sur})$	
Birth	$s_k = \bar{s}_k + v_k$	$v_k \sim \mathcal{PO}(v_k; b)$
Detect	$\hat{s}_k \vert s_k = \mathcal{BI}(\hat{s}_k; s_k, \pi_{det})$	
Observe in Clutter	$y_k \vert \hat{s}_k = \hat{s}_k + e_k$	$e_k \sim \mathcal{PO}(e_k; c)$

Realisation from the process

```
p_sur = 0.9;   % Survival probability
b = 3;         % Birth intensity
p_det = 0.5;   % Detection probability
c = 20;        % Clutter intensity
lam0 = 10;     % Prior inensity
```



Superposition of Poisson random variables

$$\begin{aligned} s &\sim \mathcal{PO}(s;\lambda) \\ e &\sim \mathcal{PO}(e;\nu) \\ y &= s+e \end{aligned}$$

$$p(y) = \mathcal{PO}(s;\lambda+\nu)$$

Thinning Poisson Random variables

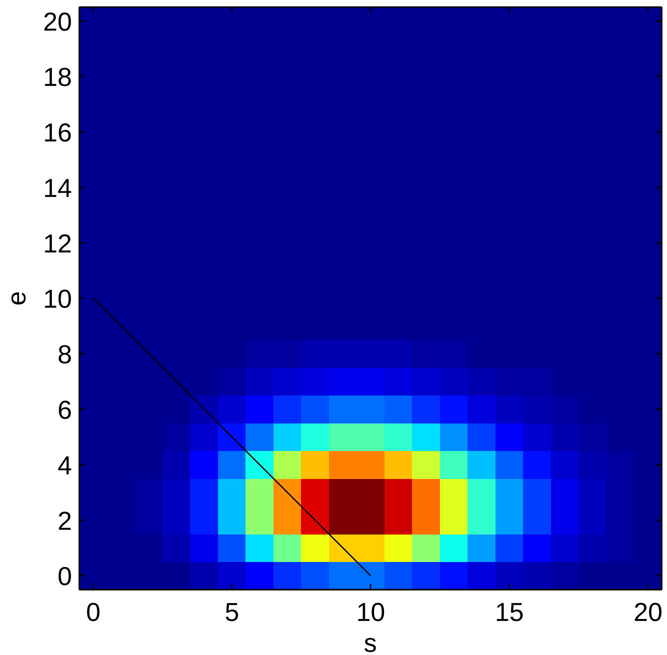
- Suppose we have n objects. Each object survives independently with probability π.

$$
\begin{aligned}
s|n &\sim \mathcal{BI}(s; n, \pi) \\
n &\sim \mathcal{PO}(n; \lambda)
\end{aligned}
$$

$$
p(s) = \sum_n \mathcal{BI}(s; n, \pi)\mathcal{PO}(n; \lambda) = \mathcal{PO}(s; \lambda\pi)
$$

Observing the sum of two Poisson Random variables

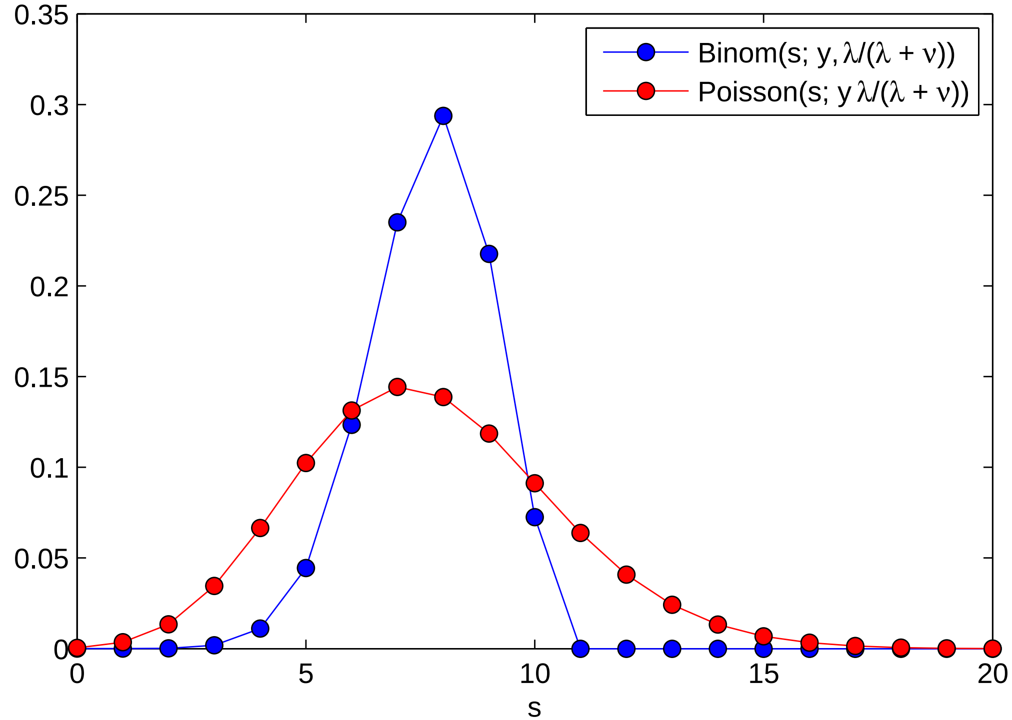
$$s \sim \mathcal{PO}(s;\lambda) \qquad e \sim \mathcal{PO}(e;\nu) \qquad y = s + e$$



$$p(s|y) = \mathcal{BI}(s; y, \lambda/(\lambda+\nu))$$

Moment matching

$$\lambda^* = \operatorname*{argmin}_{\lambda} KL(\mathcal{BI}(s; m, \pi)||\mathcal{PO}(s; \lambda)) = m\pi$$



$$\mathcal{BI}(x; m, \pi) \quad \approx \quad \mathcal{PO}(x; m\pi)$$

Probability generating functions (z-transforms)

- For a discrete random variable $n \sim p(n)$

$$
\begin{aligned}
G(z) &= \sum_{n=0}^{\infty} p(n) z^n \\
G(1) &= 1 \qquad\qquad G'(1) = \langle n \rangle
\end{aligned}
$$

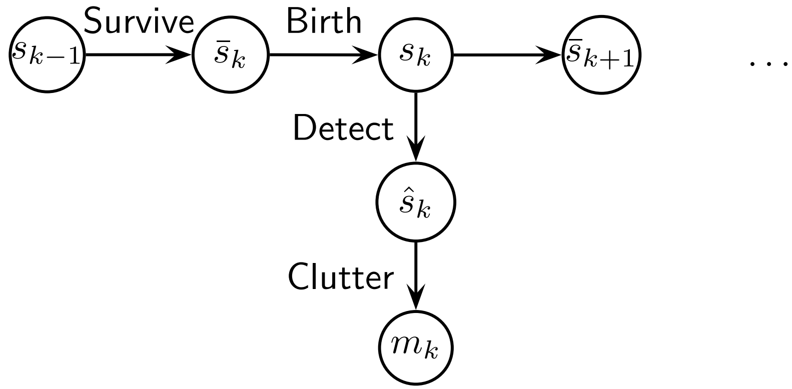
$$
\begin{aligned}
\mathcal{BI}(s; N, \pi) &\Leftrightarrow G_{\mathcal{BI}}(z) = (1 - \pi + \pi z)^N \\
G'_{\mathcal{BI}}(z) &= N\pi(1 - \pi + \pi z)^{N-1} \Rightarrow \langle s \rangle = N\pi
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{PO}(s; \lambda) &\Leftrightarrow G_{\mathcal{PO}}(z) = \exp(\lambda(z - 1)) \\
G'_{\mathcal{PO}}(z) &= \lambda \exp(\lambda(z - 1)) \Rightarrow \langle s \rangle = \lambda
\end{aligned}
$$

First moment approximation to $\mathcal{BI}$.

Sketch of the derivation of the ADF

- Start : $p(s_{k-1}|m_{1:k-1}) = \mathcal{PO}(s_{k-1}; \lambda_{k-1|k-1})$



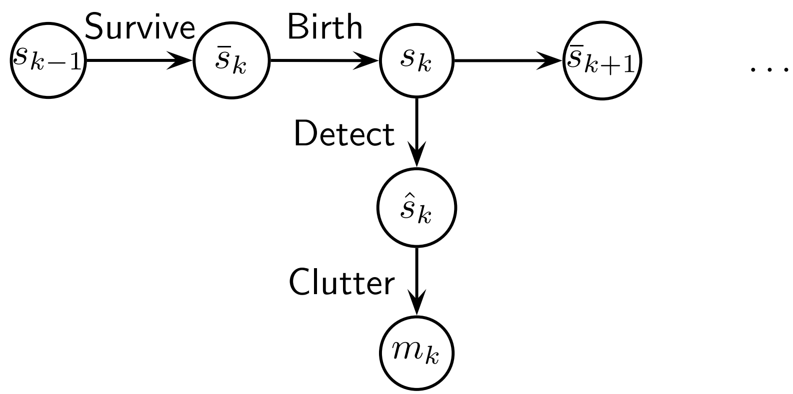
- Prediction Step (Survive + Birth) :

$$
\begin{aligned}
p(s_k|m_{1:k-1}) &= \mathcal{PO}(s_k; \lambda_{k|k-1}) \\
\lambda_{k|k-1} &= b + \pi_{sur}\lambda_{k-1|k-1}
\end{aligned}
$$

Sketch of the derivation of the ADF

- Update and Project Step (Observe in Clutter) $(\pi_{det} = 1)$

$$\begin{aligned} p(s_k|m_{1:k}) &\propto \mathcal{BI}(s_k; m_k; \lambda_{k|k-1}/(c + \lambda_{k|k-1})) \\ &\approx \mathcal{PO}(s_k; \lambda_{k|k}) \end{aligned}$$

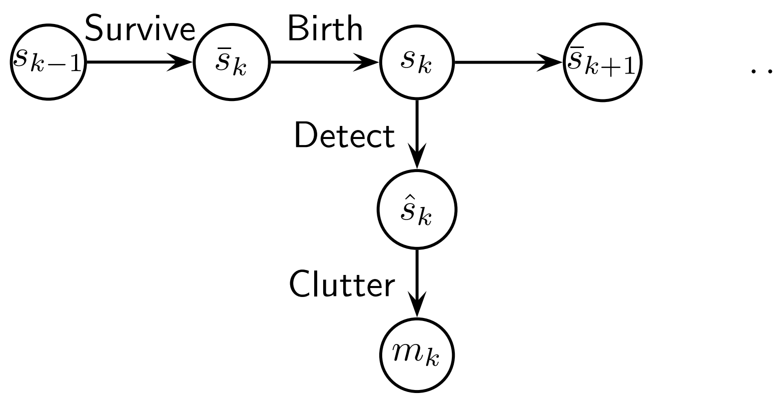


$$\lambda_{k|k} = m_k \frac{\lambda_{k|k-1}}{c + \lambda_{k|k-1}}$$

Sketch of the derivation of the ADF

- Update and Project Step (Observe in Clutter) ($\pi_{det} < 1$)

$$p(s_k|m_{1:k}) \quad \propto \quad \mathcal{PO}(s_k; \lambda_{k|k})$$

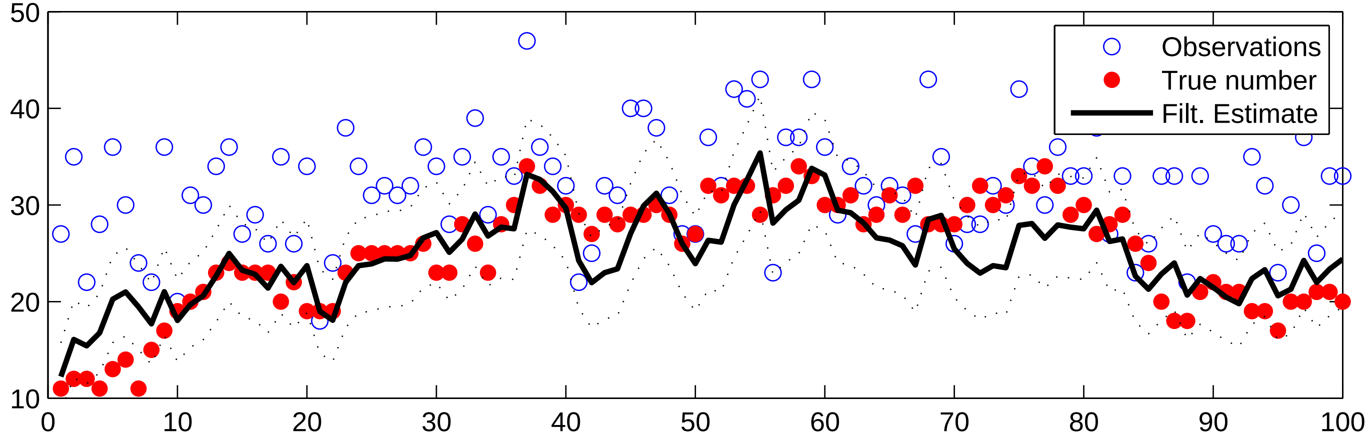


$$\lambda_{k|k} \quad = \quad (1 - \pi_{det})\lambda_{k|k-1} + m_k \frac{\pi_{det}\lambda_{k|k-1}}{c + \pi_{det}\lambda_{k|k-1}}$$

Assumed Density Filter

$$\lambda_{k|k-1} = b + \pi_{sur}\lambda_{k-1|k-1}$$

$$\lambda_{k|k} = (1-\pi_{det})\lambda_{k|k-1} + m_k \frac{\pi_{det}\lambda_{k|k-1}}{c + \pi_{det}\lambda_{k|k-1}}$$



Point Process, Definition

- A random countable set
- Realizations are from the state space

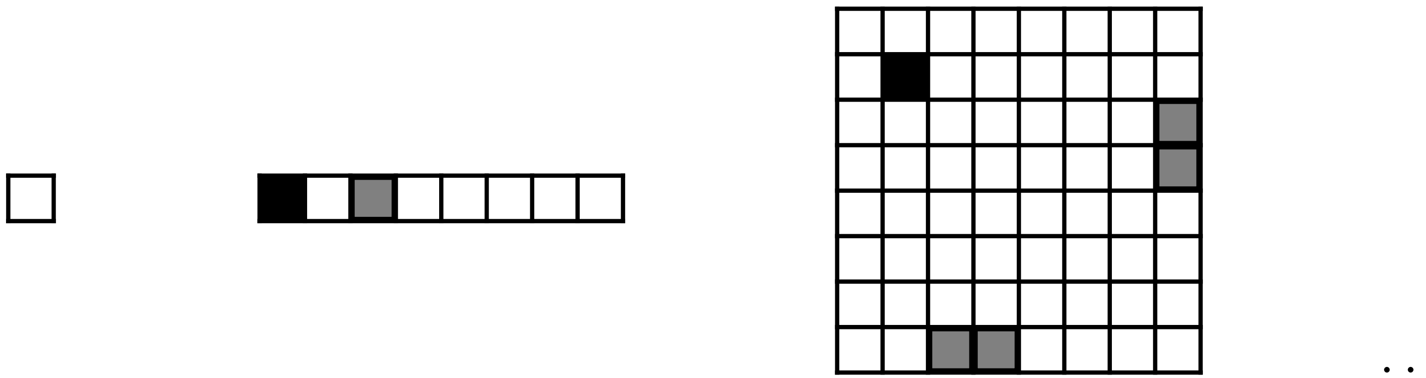
$$\mathcal{X}^{\cup} = \bigcup_{n=0}^{\infty} \mathcal{X}^n \qquad\qquad \mathcal{X}^0 \equiv \emptyset$$

- Defined by restrictions $P^{(n)}$ on $\mathcal{X}^n$, where the probability observing n points in A is denoted by

$$P^{(n)}(A \times \cdots \times A)$$

Point Process

$P^{(0)}$ $\quad$ $P^{(1)}(dx_1)$ $\quad$ $P^{(2)}(dx_1, dx_2)$



$\mathcal{X}^0$ $\quad$ $\mathcal{X}^1$ $\quad$ $\mathcal{X}^2$

- All $P^{(n)}$ are symmetric, e.g. $P^{(2)}(dx_1, dx_2) = P^{(2)}(dx_2, dx_1)$
- Normalisation

$$\sum_{n=0}^{\infty} P^{(n)}(\mathcal{X}^n) = 1$$

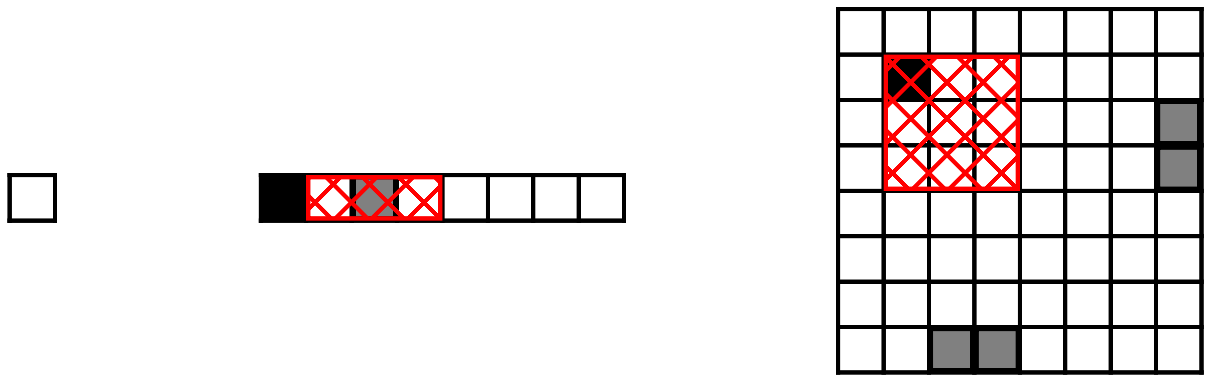
Realisations from a point process

- Choose a set A

- Generate the number of points in A

$$N_A \sim p(n) = P^{(n)}(A^n)$$

$P^{(0)}$ $\quad$ $P^{(1)}(dx_1)$ $\quad$ $P^{(2)}(dx_1, dx_2)$

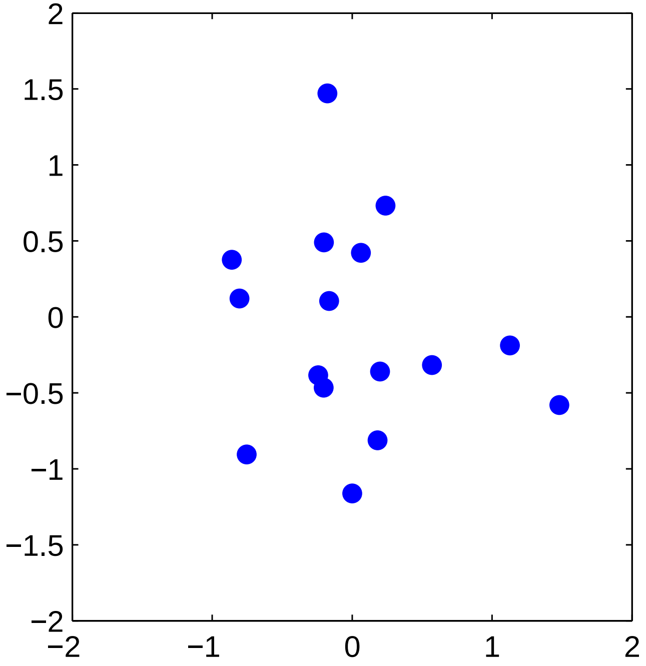


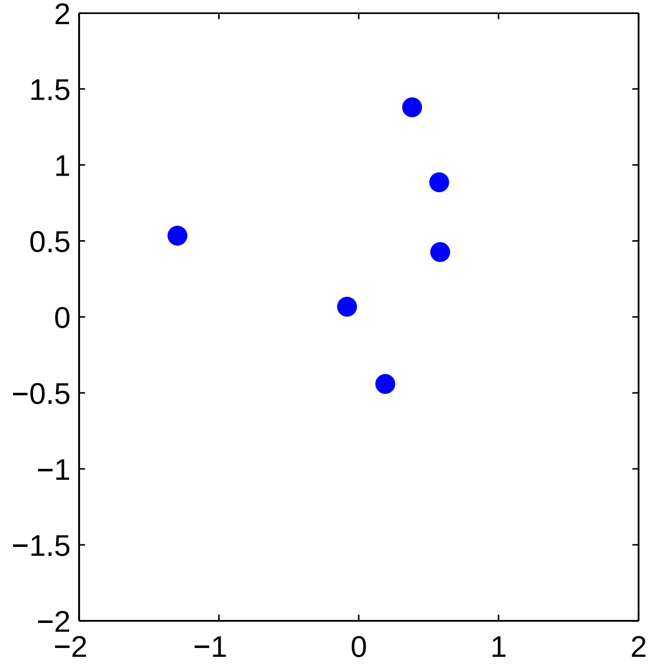
$\ldots$

Realisations from a point process

- Generate the coordinates from the joint distribution

$$(x_1, \ldots, x_n) \sim P^{(n)}(dx_1, \ldots, dx_n)/P^{(n)}(A^n)$$





Poisson Point Process

- For any $A \subset \mathcal{X}$, we denote a poisson point process by

$$X \sim \mathcal{PP}_A(X;\lambda) \qquad\qquad X \subset A$$

- The number of points to be observed in A is distributed by

$$|X \cap A| = N_A \sim \mathcal{PO}(N_A;\Lambda_A)$$

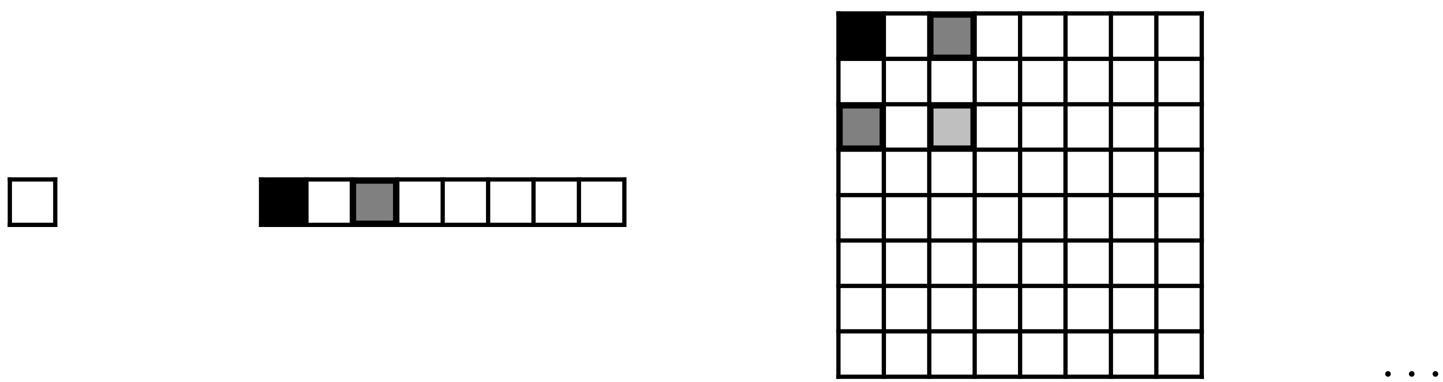
Intensity function	$\lambda(x) : \mathcal{X} \to \mathbb{R}^+$
Intensity measure	$\Lambda_A = \int_A \lambda(x)dx$

Sampling from a Poisson Point Process

- Each $P^{(n)}$ is fully factorised

$$P^{(n)} \propto \prod_{i=1}^{n} \lambda(x_i)$$

$P^{(0)}$ $\quad$ $P^{(1)}(dx_1)$ $\quad$ $P^{(2)}(dx_1, dx_2)$



- Need to only represent $\lambda(x)$, a positive function

Sampling from a Poisson Point Process

- Generate number of points on A

$$n \sim P^{(n)}(A^n) = \mathcal{PO}(n; \Lambda_A) = \frac{\exp(-\Lambda_A)}{n!}\Lambda_A^n$$

- For $i = 1 \dots n$, generate the coordinates **independently** from

$$x_i \sim \lambda(x)/\Lambda_A$$

Superposition of Poisson Processes

$$X_i \sim \mathcal{PP}_A(X_i; \lambda_i(x)) \qquad i = 1 \dots L$$

$$X \sim \mathcal{PP}_A(X; \sum_i \lambda_i(x))$$

$$X = X_1 \cup X_2 \cup \cdots \cup X_L$$

Thinning a Poisson Process

$$X \sim \mathcal{PP}_A(X; \lambda(x))$$

Each point $x_i \in X$ survives with probability $0 < \pi_{sur}(x_i) < 1$

$$X_{thin} \sim \mathcal{PP}_A(X; \pi_{sur}(x)\lambda(x))$$

Displacement of a Poisson Process

- $X_t \sim \mathcal{PP}_{A_t}(X_t; \lambda(x_t))$
- Suppose each point independently moves according to a transition kernel $p(x_{t+1}|x_t)$

$$
\begin{aligned}
X_{t+1} &\sim \mathcal{PP}_{A_{t+1}}(X_{t+1}; \lambda(x_{t+1})) \\
\lambda(x_{t+1}) &= \int_A dx_t p(x_{t+1}|x_t)\lambda(x_t)
\end{aligned}
$$

Partially observing a Poisson Point Process

$$\begin{aligned} X &\sim \mathcal{PP}_A(X; \lambda(x)) \\ C &\sim \mathcal{PP}_A(C; c(x)) \\ Y &= X \cup C \end{aligned}$$

- The posterior process $p(X|Y)$ is in general **not** a Poisson process
- Find the **Probability generating functional** of the posterior process
- Find the first (functional) derivative and calculate the first moment
- Moment matching : This gives the intensity function of the "nearest" Poisson process in the KL sense

Probability generating functional Daley and Vere-Jones 2003

- Generalises the probability generating function

$$G[z] = \sum_{n=0}^{\infty} \int_{\mathcal{X}^n} P_X^{(n)}(dx_1, \dots, dx_n) z(x_1) \cdots z(x_n)$$

- The functional derivative

$$G^{(1)}[z;\zeta] = \lim_{\epsilon \to 0} \frac{G[z+\epsilon\zeta] - G[z]}{\epsilon}$$

- The intensity is recovered by

$$\Lambda_A = \lim_{z \to 1} G^{(1)}[z; \mathbb{I}_A]$$

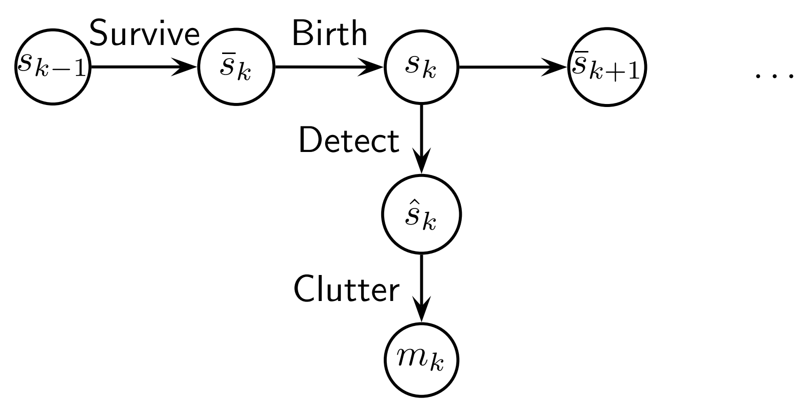
Probability Hypothesis Density Filter – The intensity recursion

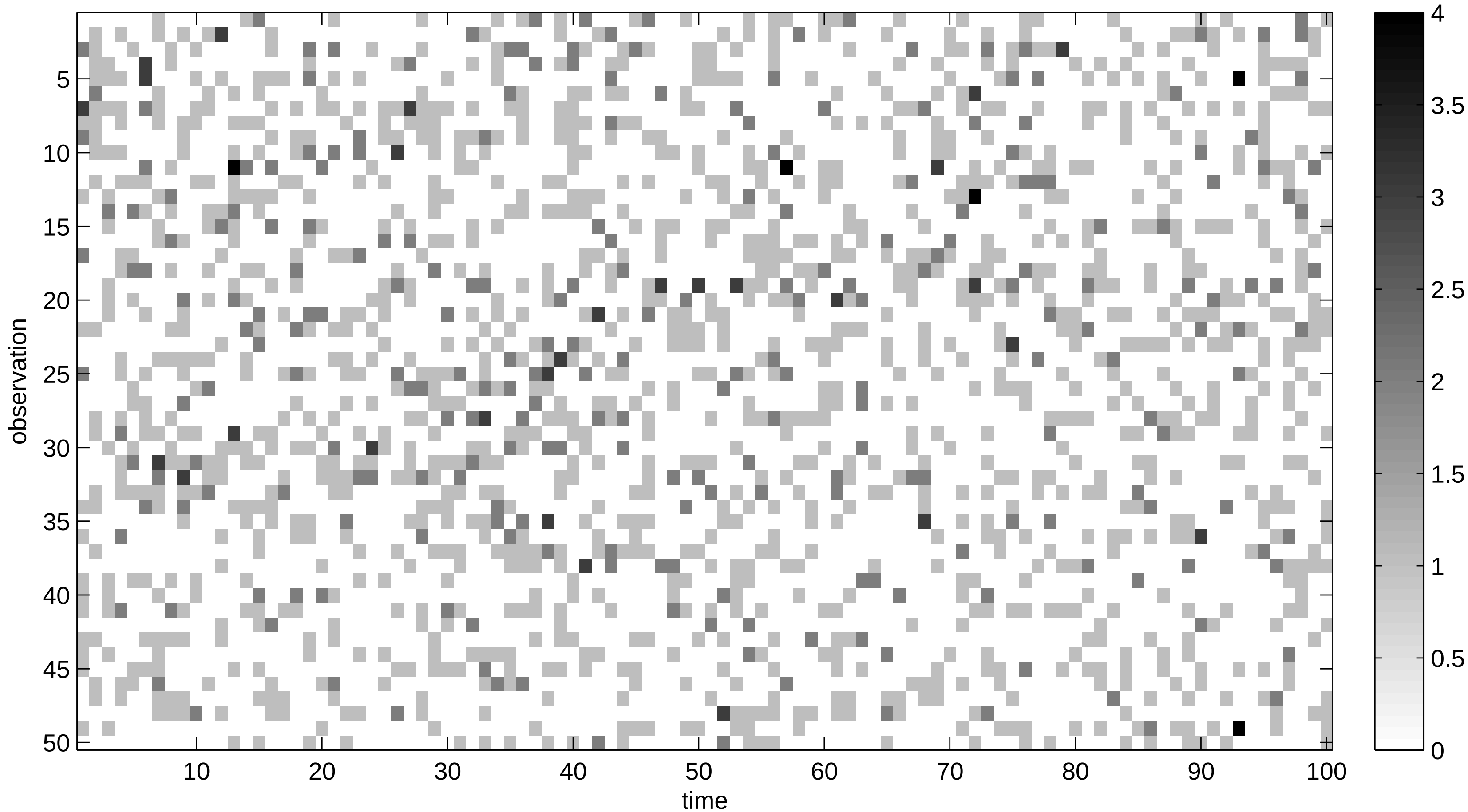
- Predict

$$\lambda_{t|t-1}(x_t) \quad = b_t(x_t) + \pi_{sur}(x_t) \int dx_{t-1} p(x_t|x_{t-1}) \lambda_{t-1}(x_{t-1})$$

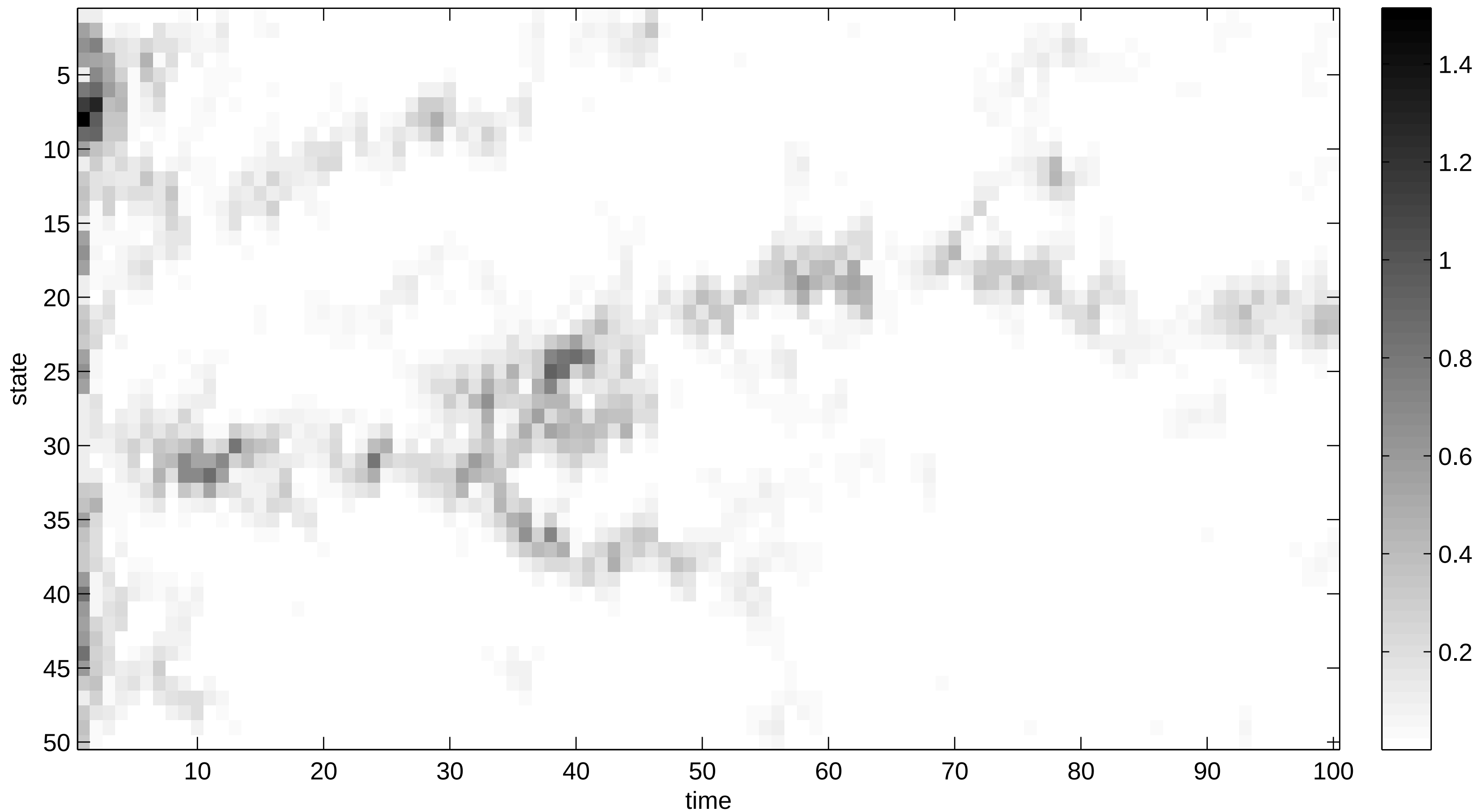
- Update

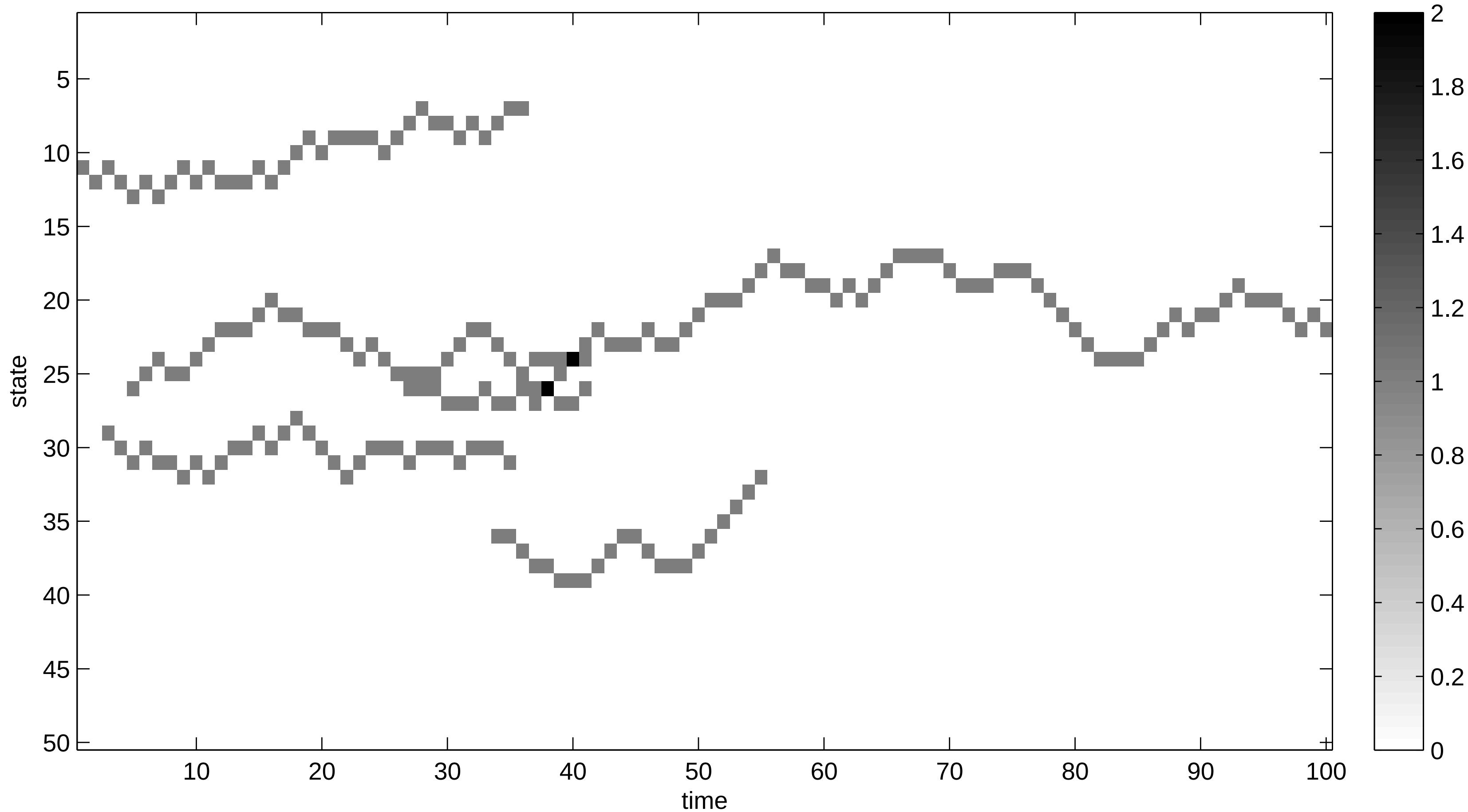
$$\lambda_t(x_t) \quad = (1 - \pi_{det}(x_t))\lambda_{t|t-1}(x_t) + \sum_{y_t \in \mathcal{Y}_t} \frac{\pi_{det}(x_t) p(y_t|x_t) \lambda_{t|t-1}(x_t)}{c_t(y_t) + \int dx' \pi_{det}(x') p(y_t|x') \lambda_{t|t-1}(x')}$$






```
p_sur = 0.98;  % Survival probability              b = 0.01;       % Birth intensity
p_det = 0.5;  % Detection probability             c = 1/3;       % Clutter intensity
lam0 = 1;    % Prior inensity
```


state
time
5
10
15
20
25
30
35
40
45
50
10
20
30
40
50
60
70
80
90
100
1.4
1.2
1
0.8
0.6
0.4
0.2



state
5
10
15
20
25
30
35
40
45
50
10
20
30
40
50
60
70
80
90
100
time
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0

Summary

- The PHD filter is an ADF
  - Retains only the first moment (the intensity)
  - A second moment approximation exists but is costly (Singh et. al. 2007)
- The PHD filter **does not solve** the association problem – it is a bypass
  - Tracks the intensity field over time
- Implemented in practice via sequential Monte Carlo
  - A stochastic approximation to a variational approximation
- Future ideas
  - A PHD Smoother is not known
  - Use as an efficient proposal for a factorial model
  - Filtering for the true posterior point process

Bibliography

D. J. Daley and D. Vere-Jones. An Introduction to the Theory of Point Processes, volume I: Springer, New York, second edition, 2003.

R. P. S. Mahler. Multitarget bayes filtering via first-order multitarget moments. IEEE Transactions on Aerospace and Electronic Systems, October 2003.

S.S. Singh, B-N Vo., A. Baddeley, and S. Zuyev. Filters for spatial point processes. Technical Report CUED/F-INFENG/TR.591, University of Cambridge, 2007.

B. Vo, S. Singh, and A. Doucet. Sequential Monte Carlo methods for multitarget filtering with random finite sets. IEEE Transactions on Aerospace and Electronic Systems, October 2005.

D. Clark, A. T. Cemgil, P. Peeling, and S. Godsill. Multi-object tracking of sinusoidal components in audio with the gaussian mixture probability hypothesis density filter. Proc. of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, October 2007.