Short Course

State Space Models, Generalized Dynamic Systems
and
Sequential Monte Carlo Methods,
and
their applications
in Engineering, Bioinformatics and Finance

Rong Chen
Rutgers University
Peking University

Part Two: Sequential Monte Carlo Methods – the Framework and Implementation

2.1 A Framework

2.1.1 (Optional) Intermediate Distributions

2.1.2 Propagation: Sampling Distribution

2.1.3 Resampling/Rejuvenation

2.1.4 Inference: Rao-Blackwellization

2.2 Some Theoretically Results

2.3 Some Applications (in detail)

Sequential Importance Sampling (SIS)

SIS Step: for $j = 1, \ldots, m$:

(A) Draw $x_{t+1}^{(j)}$ from $g_{t+1}(x_{t+1} \mid \boldsymbol{x}_t^{(j)})$ and form $\boldsymbol{x}_{t+1}^{(j)} = (\boldsymbol{x}_t^{(j)}, x_{t+1}^{(j)})$

(B) Compute the incremental weight

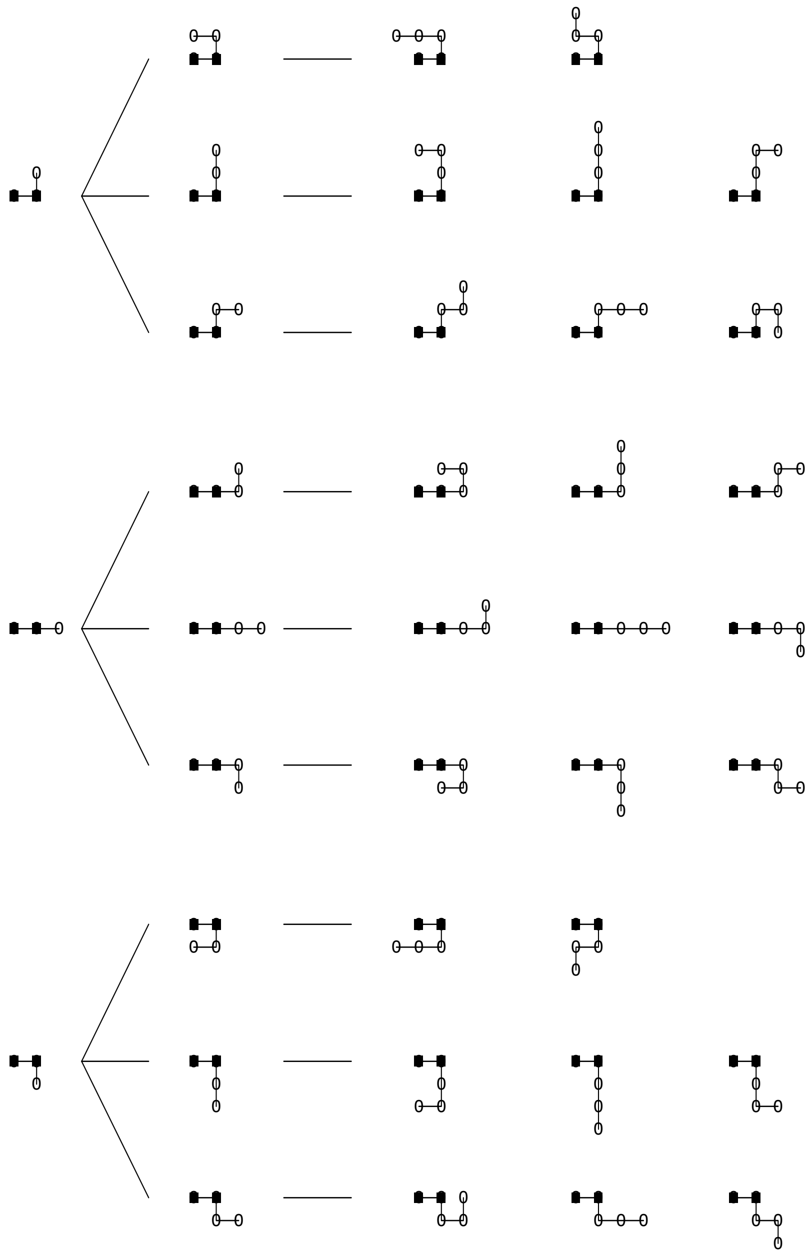
$$u_{t+1}^{(j)} = \frac{\pi_{t+1}(\boldsymbol{x}_{t+1}^{(j)})}{\pi_t(\boldsymbol{x}_t^{(j)}) g_{t+1}(x_{t+1}^{(j)} \mid \boldsymbol{x}_t^{(j)})}$$

and the new weight

$$w_{t+1}^{(j)} = u_{t+1}^{(j)} w_t^{(j)}$$

Note: $\sum_{j=1}^{m} w_t^{(j)}/m$ is an unbiased estimate of the likelihood function.

t		$t+1$		$t+1$
$\pi_t(\boldsymbol{x}_t)$		$\pi_t(\boldsymbol{x}_t)g_{t+1}(x_{t+1} \mid \boldsymbol{x}_t)$		$\pi_{t+1}(\boldsymbol{x}_{t+1})$
	$\searrow$ $g_{t+1}(\cdot \mid \boldsymbol{x}_t)$ $\nearrow$		$\searrow$ w_{t+1} $\nearrow$	
$\cdots\cdots\cdots$	$\cdots$	$\cdots\cdots\cdots\cdots\cdots$	$\cdots$	$\cdots\cdots\cdots$
$(\boldsymbol{x}_t^{(1)}, w_t^{(1)})$	$\longrightarrow$	$x_{t+1}^{(1)}$	$\longrightarrow$	$(\boldsymbol{x}_{t+1}^{(1)}, w_{t+1}^{(1)})$
$\vdots$		$\vdots$		$\vdots$
$(\boldsymbol{x}_t^{(m)}, w_t^{(m)})$	$\longrightarrow$	$x_{t+1}^{(m)}$	$\longrightarrow$	$(\boldsymbol{x}_{t+1}^{(m)}, w_{t+1}^{(m)})$



Efficiency! Efficiency! Efficiency!

The variance of the estimator (iid samples):

$$Var = \frac{1}{m} \int \left[\frac{h(x)\pi(x)}{g(x)} \right]^2 g(x)dx - \textbf{mean}^2$$

When $h(x) = 1$, it is the weight variance with respect to g.

<u>Efficiency (rule of thumb):</u>

$$\textbf{effective sample size} = \frac{m}{1 + cv^2(w)}$$

2.1 Sequential Monte Carlo (SMC) in a Nutshell

SMC = *Sequential Importance Sampling*

Importance Sampling

$+$

Sequential Sampling
(of the components of each sample)

Importance Sampling

Target distribution $\pi(\cdot)$**. Available an iid sample** $\{x_1, \ldots, x_m\}$ **from a trial distribution** $g(\cdot)$

$$E_\pi(h(X)) = \int h(x)\pi(x)dx = \int h(x)\frac{\pi(x)}{g(x)}g(x)dx = E_g(h(X)w(X))$$

where $w(x) = \pi(x)/g(x)$**.**

We have

$$\frac{1}{\sum w_i}\sum_{i=1}^{m} w_i h(x_i) \approx E_\pi(h(x))$$

The sample $(x_j, w_j), j = 1, \ldots, m$ **is said to be** <u>**properly weighted**</u> **with respect to distribution** π**.**

Fact: The <u>same</u> sample can be properly weighted by <u>different</u> sets of weights, with respect to the <u>same</u> target distribution π.

- **If** (X, Y, w) **is properly weighted with respect to** $\pi(x, y)$**, then** (X, w) **is properly weighted with respect to** $\pi(x) = \int \pi(x, y)dy$**, because**

$$\begin{aligned} \frac{1}{\sum w_i} \sum_{i=1}^{m} w_i h(x_i) &\approx E_{\pi(x,y)}(h(x)) \\ &= \int \int h(x)\pi(x, y)dydx \\ &= \int h(x)\pi(x)dx \end{aligned}$$

- $(x_1, y_1), \ldots, (x_m, y_m)$ **from** $g(x, y)$**. Let**

$$w_i^{(1)} = \frac{\pi(x_i, y_i)}{g(x_i, y_i)}; \qquad w_i^{(2)} = \frac{\pi(x_i)}{g(x_i)}$$

where $\pi(x) = \int \pi(x, y)dy$ **and** $g(x) = \int g(x, y)dy$**. Then both** $(x_i, w_i^{(1)})$ **and** $(x_i, w_i^{(2)})$ **are properly weighted with respect to** $\pi(x)$**.**

  - **Note: Works for any** $\pi(x, y)$ **such that** $\int \pi(x, y)dy = \pi(x)$**.**
  - **e.g.** $\pi(x, y) = \pi(x)\pi(y)$ **for any density** $\pi(y)$**.**
  - **e.g.** $\pi(x, y) = \pi(x)\pi(y \mid x)$ **for any conditional density** $\pi(y \mid x)$**.**
  - **Which one is more efficient?**

Sequential Sampling:

- **Target distribution** $\pi(\boldsymbol{x}) = \pi(x_1, \ldots, x_n)$ **is of high dimensional.**
- **Need to construct (and sample from) a high dimensional trial distribution** $g(\boldsymbol{x})$**.**
- **Solution: sequential build-up**

$$g(\boldsymbol{x}) = g_1(x_1) g_2(x_2 \mid x_1) \ldots g_n(x_n \mid x_1, \ldots, x_{n-1})$$

  with each g_i **easy to sample from.**

- **The importance weight is**

$$w(\boldsymbol{x}) \propto \frac{\pi(x_1, \ldots, x_n)}{g_1(x_1) g_2(x_2 \mid x_1) \ldots g_n(x_n \mid x_1, \ldots, x_{n-1})}$$

- **Inference**

$$E_\pi(h(\boldsymbol{x})) \approx \frac{\sum_i h(\boldsymbol{x}^{(i)}) w(\boldsymbol{x}^{(i)})}{\sum_i w(\boldsymbol{x}^{(i)})}$$

Efficiency! Efficiency! Efficiency!

Design Issues

- (optional) The choice of the intermediate distributions
- How to propagate efficiently? — Use more information
- How to handle small weights? — Resampling
- How to make inference efficiently? — Rao-Blackwellization

2.1.1 Design Issues – Intermediate Distributions

Setting up a series of targets, starting with an easy one, and gradually move from one target to the next, eventually reach the final target.

$$\pi_1 \Rightarrow \pi_2 \Rightarrow \pi_3 \Rightarrow \ldots \Rightarrow \pi_n$$

The Growth Principle

Decompose a complex problem into a sequence of simpler problems.

Recall:

$$w(\boldsymbol{x}_n) = \frac{\pi(\boldsymbol{x}_n)}{g_1(x_1)g_2(x_2 \mid \boldsymbol{x}_1) \ldots g_n(x_n \mid \boldsymbol{x}_{n-1})}$$

Where is the intermediate distribution $\pi_t(\boldsymbol{x}_t)$?

$$\begin{aligned}
\pi_n(\boldsymbol{x}_n) &= \frac{\pi_n(\boldsymbol{x}_n)}{\pi_{n-1}(\boldsymbol{x}_{n-1})}\frac{\pi_{n-1}(\boldsymbol{x}_{n-1})}{\pi_{n-2}(\boldsymbol{x}_{n-2})}\cdots\frac{\pi_2(\boldsymbol{x}_2)}{\pi_1(\boldsymbol{x}_1)}\pi_1(\boldsymbol{x}_1) \\
&= \left[\pi_n(x_n \mid \boldsymbol{x}_{n-1})\frac{\pi_n(\boldsymbol{x}_{n-1})}{\pi_{n-1}(\boldsymbol{x}_{n-1})}\right]\left[\pi_{n-1}(x_{n-1} \mid \boldsymbol{x}_{n-2})\frac{\pi_{n-1}(\boldsymbol{x}_{n-2})}{\pi_{n-2}(\boldsymbol{x}_{n-2})}\right] \\
&\quad \ldots \left[\pi_2(x_2 \mid \boldsymbol{x}_1)\frac{\pi_2(\boldsymbol{x}_1)}{\pi_1(\boldsymbol{x}_1)}\right]\pi_1(\boldsymbol{x}_1)
\end{aligned}$$

where $\pi_t(\boldsymbol{x}_{t-1}) = \int \pi_t(\boldsymbol{x}_{t-1}, x_t)dx_t$**.**

Hence

$$w_n(\boldsymbol{x}_n) = \frac{\pi_1(x_1)}{g_1(x_1)}\prod_{t=2}^{n}\frac{\pi_t(x_t \mid \boldsymbol{x}_{t-1})}{g_t(x_t \mid \boldsymbol{x}_{t-1})}\frac{\pi_t(\boldsymbol{x}_{t-1})}{\pi_{t-1}(\boldsymbol{x}_{t-1})}$$

And the intermediate weight is

$$w_t(\boldsymbol{x}_t) = w_{t-1}(\boldsymbol{x}_{t-1})\frac{\pi_t(x_t \mid \boldsymbol{x}_{t-1})}{g_t(x_t \mid \boldsymbol{x}_{t-1})}\frac{\pi_t(\boldsymbol{x}_{t-1})}{\pi_{t-1}(\boldsymbol{x}_{t-1})}$$

SMC – The Algorithm

- **Construct the intermediate distributions $\pi_t(\boldsymbol{x}_t)$.**
  - $\pi_{t-1}(\boldsymbol{x}_{t-1}) \approx \pi_t(\boldsymbol{x}_{t-1}) = \int \pi_t(\boldsymbol{x}_t) dx_t$ **and** $\pi_n(\boldsymbol{x}_n) = \pi(\boldsymbol{x})$
- **Construct the sampling distributions $g(x_t \mid \boldsymbol{x}_{t-1})$.**
  - $\pi_t(x_t \mid \boldsymbol{x}_{t-1}) \approx g_t(x_t \mid \boldsymbol{x}_{t-1})$
  - **better: also compensate for** $\pi_t(\boldsymbol{x}_{t-1})/\pi_{t-1}(\boldsymbol{x}_{t-1})$

SMC Step: for $j = 1, \ldots, m$:

(A) Draw $x_t^{(j)}$ from $g(x_t \mid \boldsymbol{x}_{t-1}^{(j)})$. **Let $\boldsymbol{x}_t^{(j)} = (\boldsymbol{x}_{t-1}^{(j)}, x_t^{(j)})$.**

(B) Compute the incremental weight

$$u_t^{(j)} \propto \frac{\pi_t(x_t^{(j)} \mid \boldsymbol{x}_{t-1}^{(j)})}{g_t(x_t^{(j)} \mid \boldsymbol{x}_{t-1}^{(j)})} \frac{\pi_t(x_{t-1}^{(j)})}{\pi_{t-1}(\boldsymbol{x}_{t-1}^{(j)})}$$

and the new weight $w_{t+1}^{(j)} = u_{t+1}^{(j)} w_t^{(j)}$

2.1.2 Propagation – choosing $g_t(x_t \mid \boldsymbol{x}_{t-1})$ close to $\pi_t(x_t \mid \boldsymbol{x}_{t-1})$

$$\textbf{state equation:} \quad x_t = s_t(\boldsymbol{x}_{t-1}, \varepsilon_t) \ \textbf{ or } \ x_t \sim q_t(\cdot \mid \boldsymbol{x}_{t-1})$$
$$\textbf{observation equation:} \quad y_t = h_t(x_t, e_t) \ \textbf{ or } \ y_t \sim f_t(\cdot \mid x_t)$$

Intermediate distribution: $\pi_t(\boldsymbol{x}_t \mid \boldsymbol{y}_t) \propto \prod_{s=1}^{t} f_s(y_s \mid x_s) q_s(x_s \mid \boldsymbol{x}_{s-1})$.

Conditional distribution: $\pi_t(x_t \mid \boldsymbol{x}_{t-1}, \boldsymbol{y}_t) \propto q_t(x_t \mid \boldsymbol{x}_{t-1}) f_t(y_t \mid x_t)$

(1) Standard Particle Filters: (use the state equation only)

$$g_t(x_t \mid x_{t-1}, y_t) = q_t(x_t \mid x_{t-1})$$

with weight $w_t = w_{t-1} f_t(y_t \mid x_t)$.

(2) Independent Particle Filters: (use the observation equation only) (Lin et al, 2004)

$$g_t(x_t \mid x_{t-1}, y_t) \propto f_t(y_t \mid x_t)$$

with weight $w_t = q_t(x_t \mid x_{t-1})$.

(3) Full information (Liu and Chen 1998)

$$g_t(x_t \mid x_{t-1}, y_t) \propto q_t(x_t \mid x_{t-1}) f_t(y_t \mid x_t)$$

with weight

$$w_t = w_{t-1} \int q_t(x_t \mid x_{t-1}) f_t(y_t \mid x_t) dx_t$$

(4) Auxiliary Particle Filters (Pitt & Shepherd, 1998):

$$g_{t+1}(x_{t+1} \mid \boldsymbol{x}_t) \propto q_{t+1}(x_{t+1} \mid x_t) \hat{f}_{t+1}(y_{t+1} \mid x_{t+1})$$

with weight

$$w_{t+1} \propto w_t \frac{f_{t+1}(y_{t+1} \mid x_{t+1})}{\hat{f}_{t+1}(y_{t+1} \mid x_{t+1})} \int q_{t+1}(x_{t+1} \mid x_t) \hat{f}_{t+1}(y_{t+1} \mid x_{t+1}) dx_{t+1}$$

where $\hat{f}_{t+1}$ is an approximation of f_{t+1}.

Delay (Look-ahead) Methods – use more information

- **Dynamic systems often process strong ’memory’**
- **Future observations can reveal substantial information on the current state**
- **Slight delay is tolerable**
- **Make inference on the state x_t at time $t+d$, with information $y_1, \ldots, y_{t+d}$ available.**
- **The intermediate distribution becomes**

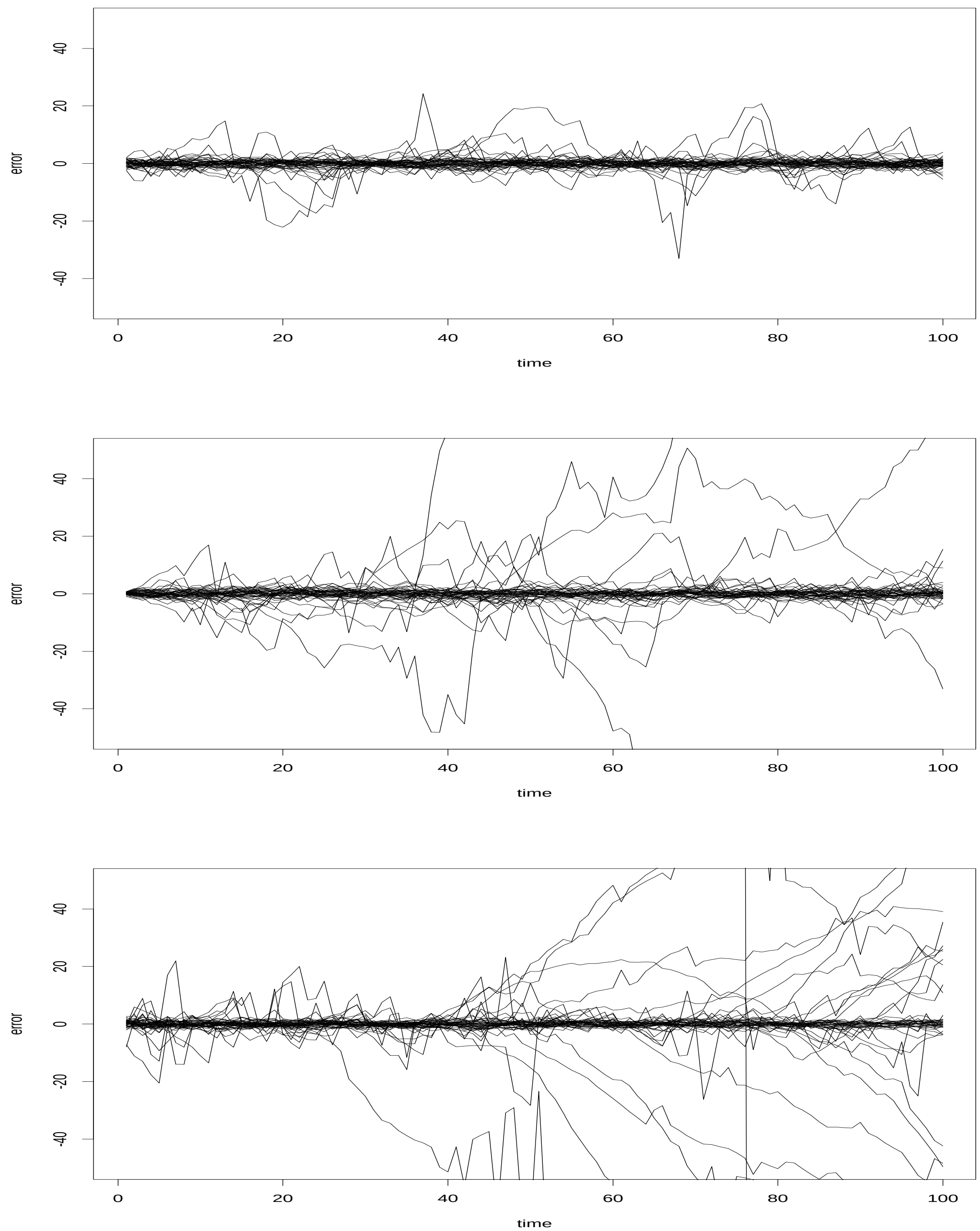
$$\pi_t^*(\boldsymbol{x}_t) = \int \pi_{t+d}(\boldsymbol{x}_t, x_{t+1}, \ldots, x_{t+d}) dx_{t+1} \ldots x_{t+d}$$

In state space model,

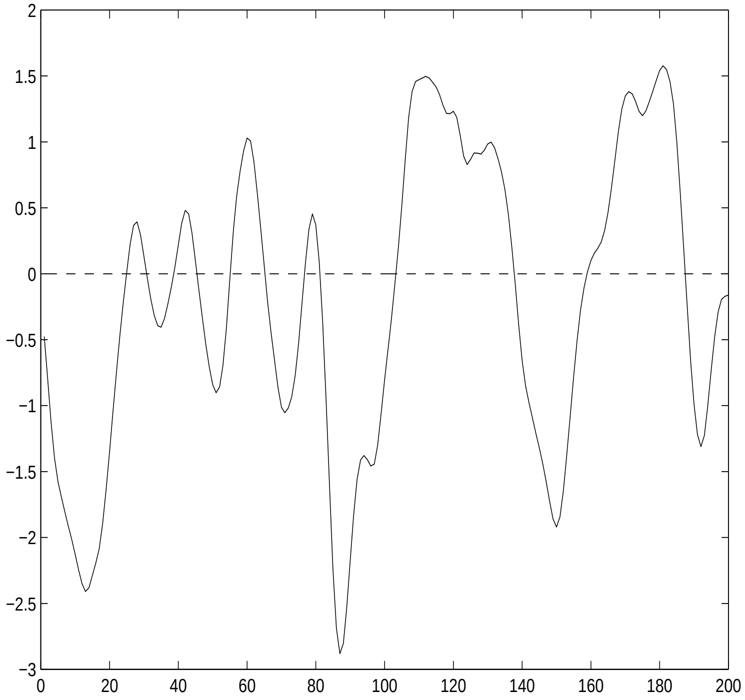
$$\pi_t(x_t) = p(x_t \mid y_1, \ldots, y_t, y_{t+1}, \ldots, y_{t+d})$$

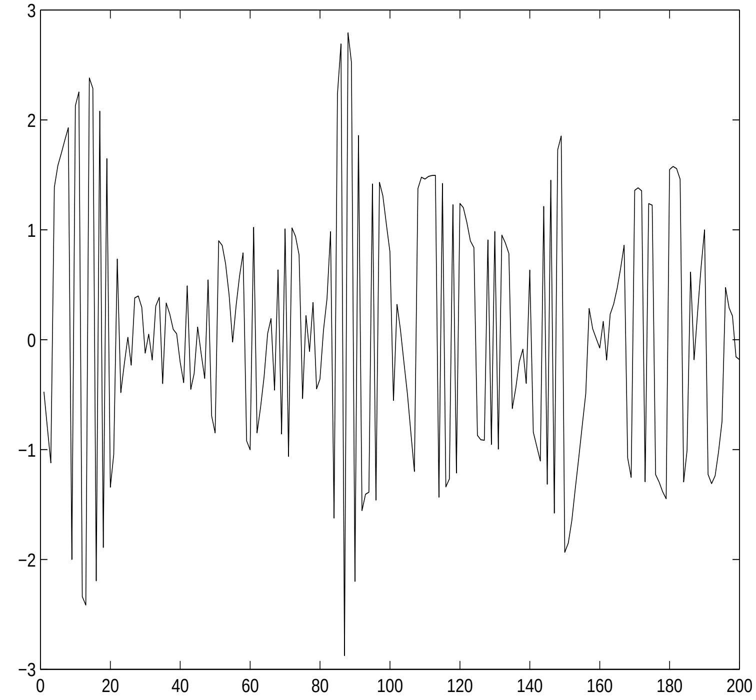
- **Closer to the ultimate target distribution:** $p(x_t \mid y_1, \ldots, y_n)$

Tracking in clutter:

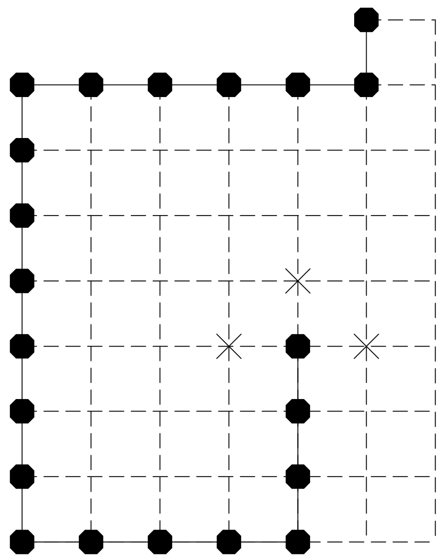


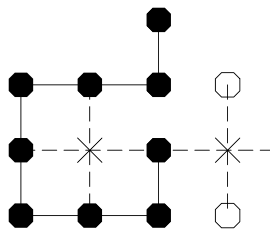
Example – Fading Channels

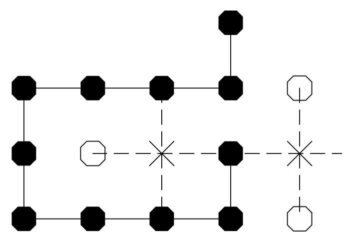




Example: SAW



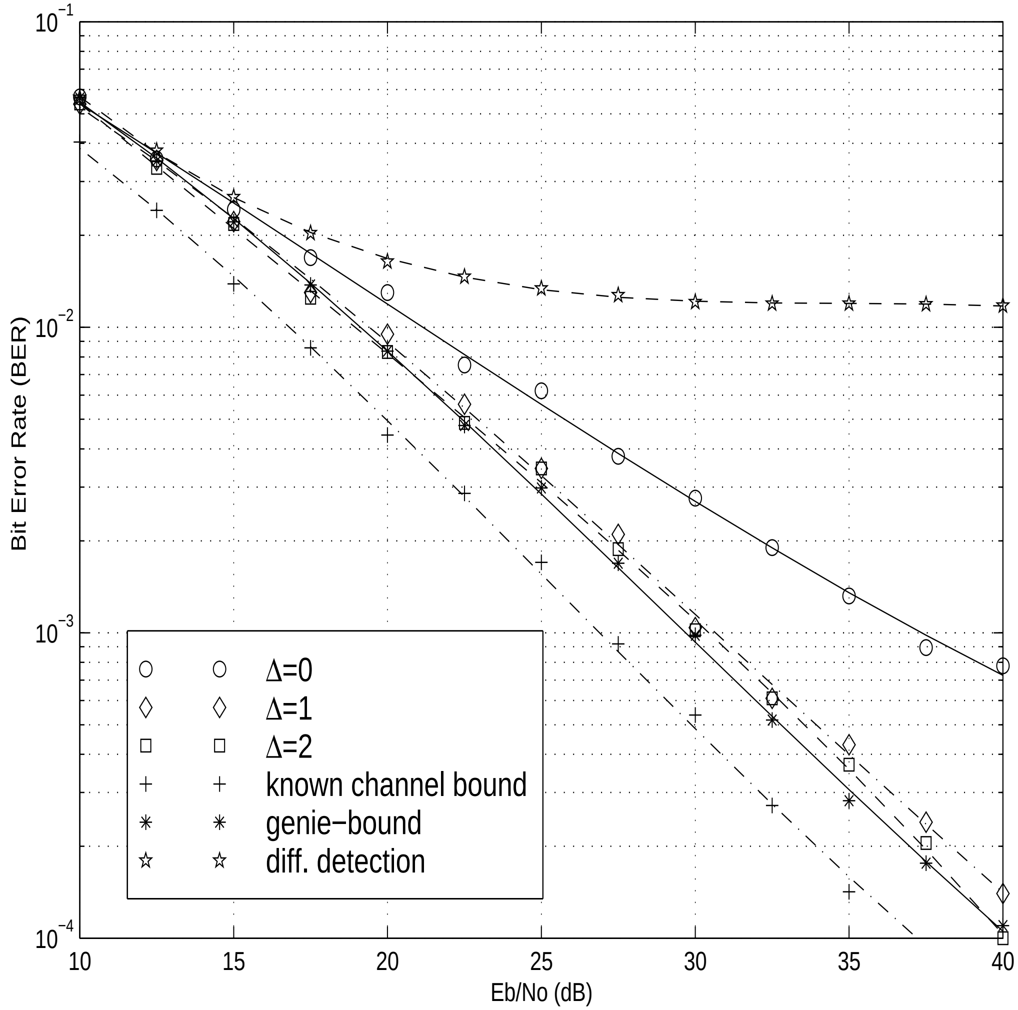




Different Delay Methods – various computational cost

- **Delay Weight Method**
- **Delay Sample Method: (exact method)**
- **Hybrid method – combining of delay weight and delay sample**
- **Pilot sampling method – sending pilot to partially explore the future space**
- **Mutli-level method – coarsen (reconstruct) the state space and use delayed SMC on the simpler state space, refine results in the original state space**
- **Adaptive delay method – long delay when information is week and short/no delay when information is strong.**

Example – Fading Channels – Uncoded – Gaussian Noise



Example – Fading Channels – Coded – Hybrid Delay Pilot, Sample and Weight