# 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)

### 2.1.3 Resampling (rejuvenation)

**<u>Fact</u>:**

- **Variance of $w_t$ increases (stochastically) as $t$ increases**
- **SMC does not allow to go back to 'correct' early samples**
- **Carrying samples with small weight forward wastes computational resources**

**Solution: duplicate the 'important' samples and remove the 'unimportant' samples.**

**Simple resampling:**

**At time $t$, a set of samples $S_t = \{(\boldsymbol{x}_t^{(j)}, w_t^{(j)})\}_{j=1}^m$**

**Simple Resampling Step:**

**(A) Sample a new set of streams $S_t^{'}$ from $S_t$ according to $w_t^{(j)}$, with replacement.**

**(B) All sampled samples are assigned weight 1.**

**The resampled samples behaves as identical (but not independent) samples from $\pi_t(\boldsymbol{x}_t)$.**

**Homework: Show the new sample is still properly weighted with respect to $\pi_t(\boldsymbol{x}_t)$.**

**Residual resampling:**

**At time $t$, a set of samples $S_t = \{(\boldsymbol{x}_t^{(j)}, w_t^{(j)})\}_{j=1}^m$. $\sum_{j=1}^m w_t^{(j)} = 1$.**

- **Make $\lfloor m w_t^{(j)} \rfloor$ copies of $x_t^{(j)}$.**
- **Let $m^* = m - \sum_{j=1}^m \lfloor m w_t^{(j)} \rfloor$ and $w_t^{*(j)} = m w_t^{(j)} - \lfloor m w_t^{(j)} \rfloor$, $j = 1, ...m$.**
- **Resample $m^*$ samples from $S_t$ with probability proportional to $w_t^{*(j)}$ with replacement.**

**Prune-and-Enriched Rosenbluth Method (Grassberger 1997):**

- **(Sequentially) Replacing each zero weight sample with the sample of highest weight.**
- **The weight of both the original sample and the duplicated sample are set to half of the original weight.**

**Homework: Show the new sample is still properly weighted with respect to $\pi_t(\boldsymbol{x}_t)$.**

**Remarks:**

- **Resampling provides more efficient samples of future states**
- **Resampling increases sampling variation in the past states**
- **Resampling reduces the number of distinctive samples in the past states**
- **Frequent resampling can be** *shortsighted*
- **(online estimation) Resampling should be done after estimation.**

**Resampling Schedule:**

- **deterministic: resampling at time** $t_0, 2t_0, 3t_0, ...$
- **dynamic: monitoring the weight variance**

**A simulated example:**

$$y_t = x_t + 0.8x_{t-1} - 0.4x_{t-2} + \varepsilon_t$$

**with $x_t$ i.i.d from $\{0, 1, 3\}$ and SNR=15dB.**

- **The coefficients $\phi$ are integrated out with a normal prior.**
- **200 simulated sequences. Sample size $T = 200$.**
- **Number of streams $m = 1000$.**
- **Delayed estimation: $\hat{x}_t = MAP(\pi_{t+3}(x_t))$**
- **simple random sampling (s) versus residual sampling (r)**
- **Deterministic schedule: $t_0, 2t_0, 3t_0, \ldots$**
  **Dynamic schedule: when the effective sample size is less than 3.**

| error | Deterministic Resampling Schedule $t_0$ | | | | | | | | | | | | dynamic schedule | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | | 5 | | 20 | | 50 | | 100 | | 200 | | | |
| | s | r | s | r | s | r | s | r | s | r | s | r | s | r |
| 0-2 | 11 | 5 | 7 | 13 | 13 | 13 | 7 | 10 | 1 | 0 | 0 | 0 | 11 | 12 |
| 3-5 | 49 | 49 | 46 | 53 | 61 | 65 | 53 | 49 | 28 | 28 | 7 | 7 | 69 | 58 |
| 6-8 | 41 | 43 | 50 | 52 | 72 | 70 | 57 | 58 | 59 | 58 | 12 | 12 | 66 | 67 |
| 9-11 | 23 | 20 | 27 | 30 | 38 | 38 | 52 | 48 | 43 | 44 | 47 | 47 | 29 | 41 |
| 12-15 | 10 | 9 | 13 | 7 | 8 | 6 | 17 | 20 | 33 | 32 | 44 | 44 | 16 | 8 |
| 16-25 | 11 | 10 | 14 | 11 | 8 | 8 | 14 | 15 | 35 | 35 | 84 | 84 | 6 | 11 |
| 16-50 | 4 | 10 | 8 | 9 | 0 | 0 | 0 | 0 | 1 | 3 | 6 | 6 | 1 | 1 |
| $>$50 | 51 | 54 | 35 | 25 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 2 |

## Why resampling?

**The asymptotic variance ($\times m$) (estimating $\mu = \int h(\boldsymbol{x}_n)\pi_n(\boldsymbol{x}_n)d\boldsymbol{x}_n$)**

- **No resampling:**

$$\int \frac{\pi_n^2(\boldsymbol{x}_n)(h(\boldsymbol{x}_n)-\mu)^2}{g(\boldsymbol{x}_n)} d\boldsymbol{x}_n$$

- **Resampling (Del Moral 2004, Chopin, 2004)**

$$\int \frac{\pi_n^2(x_1)(\mu_1(x_1)-\mu)^2}{g_1(x_1)} dx_1 + \sum_{t=2}^{n} \frac{\pi_n^2(\boldsymbol{x}_t)(\mu_t(\boldsymbol{x}_t)-\mu)^2}{\pi_{t-1}(\boldsymbol{x}_{t-1})g_t(x_t \mid \boldsymbol{x}_{t-1})} d\boldsymbol{x}_t$$

**where**

$$\mu_t(\boldsymbol{x}_t) = \int h(\boldsymbol{x}_t)\pi_n(\boldsymbol{x}_{t+1:n} \mid \boldsymbol{x}_t) d\boldsymbol{x}_{t+1:n}$$

**(a much smoother function and 'closer' to $\mu$)**

  - **e.g. the first term: same as sample from $g_1(x_1)\pi_n(\boldsymbol{x}_{2:n})$**

## Flexible Resampling Schemes

**The resampling trick:**

- **Suppose** $\boldsymbol{x}_t^{(1)}, \ldots, \boldsymbol{x}_t^{(m)}$ **following** $g(\boldsymbol{x}_t)$
- **Sample** $m$ **samples (with replacement) from the set** $\{\boldsymbol{x}_t^{(1)}, \ldots, \boldsymbol{x}_t^{(m)}\}$ **with probability proportional to** $\alpha_t(\boldsymbol{x}_t^{(j)})$, $j = 1, \ldots, m$.
- **The resulting set asymptotically follow the distribution**

$$g(\boldsymbol{x}_t)\alpha(\boldsymbol{x}_t)$$

- **e.g.** $\alpha(\boldsymbol{x}_t) = \pi_t(\boldsymbol{x}_t)/g_t(\boldsymbol{x}_t)$

**One can choose different (and better)** $\alpha(\boldsymbol{x}_t)$ **to serve different purposes.**

## Flexible Resampling Schemes

- **The square-root of weights (Liu, 2001)**

$$\alpha_{t-1}(\boldsymbol{x}_{t-1}) = \sqrt{w_{t-1}(\boldsymbol{x}_{t-1})}$$

- **Auxiliary particle filter (Pitt & Shephard, 1999)**

$$\alpha_{t-1}(\boldsymbol{x}_{t-1}) = w_{t-1}(\boldsymbol{x}_{t-1})\gamma_t(\hat{x}_t \mid \boldsymbol{x}_{t-1})$$

**where $\hat{x}_t$ is a (global) prediction of $x_t$.**

- **Incremental-Weight Spreading. (Neil Shephard, private conversation)**

$$\alpha_{t-1}(\boldsymbol{x}_{t-1}) = [\prod_{\ell=1}^{L} u_{t-\ell}(\boldsymbol{x}_{t-\ell})]^{1/L} = \prod_{\ell=1}^{L} \left[ \frac{\pi_{t-\ell}(\boldsymbol{x}_{t-\ell})}{\pi_{t-\ell-1}(\boldsymbol{x}_{t-\ell-1}) g_{t-\ell}(x_{t-\ell} \mid \boldsymbol{x}_{t-\ell})} \right]^{1/L}$$

- **Delayed resampling and block sampling (Wang et al, 2002, Doucet et al 2006)**

$$\alpha_{t-1}(\boldsymbol{x}_{t-1}) = \frac{\pi_{t+\delta}(\boldsymbol{x}_{t-1})}{\pi_{t+\delta-1}(\boldsymbol{x}_{t-2}) g_t(x_{t-1} \mid \boldsymbol{x}_{t-1})}$$

- **Resampling with backward pilots, (Lin et al, 2009)**

$$\alpha_{t-1}(\boldsymbol{x}_{t-1}) = \frac{\hat{\pi}_n(\boldsymbol{x}_{t-1})}{\hat{\pi}_n(\boldsymbol{x}_{t-2}) g_{t-1}(x_{t-1} \mid \boldsymbol{x}_{t-2})}$$

- **Resampling with function consideration (Zhang et al, 2003)**

$$\alpha_{t-1}(\boldsymbol{x}_{t-1}) = \| \hat{\mu}_{t-1}(\boldsymbol{x}_{t-1}) w_{t-1}(\boldsymbol{x}_{t-1}) \|$$

**where** $\hat{\mu}_{t-1}(\boldsymbol{x}_{t-1})$ **is an estimate of**

$$\int |h(\boldsymbol{x}_n) \pi_n(\boldsymbol{x}_n)| d\boldsymbol{x}_n$$

**At times $t = 2, \ldots, n$,**

**(0) Construct $\boldsymbol{\alpha}_{t-1} = \{\alpha(\boldsymbol{x}_{t-1}^{(1)}), \ldots, \alpha(\boldsymbol{x}_{t-1}^{(m)})\}$**

**(A) Sample $A_{t-1}^{(j)}$ with prob $\{\alpha_{t-1}^{(1)}, \ldots, \alpha_{t-1}^{(m)}\}$**

**(B) sample $x_t^{(j)} \sim g_t(\cdot \mid \boldsymbol{x}_{t-1}^{A_{t-1}^{(j)}})$ and set $\boldsymbol{x}_t^{(j)} := (\boldsymbol{x}_{t-1}^{A_{t-1}^{(j)}}, x_t^{(j)})$, and**

**(C) compute and normalize the weights**

$$
\begin{aligned}
u_t(\boldsymbol{x}_{1:t}^{(j)}) &= \frac{\pi_t(\boldsymbol{x}_t^{(j)})}{\pi_{t-1}(\boldsymbol{x}_{t-1}^{A_{t-1}^{(j)}})g_t(x_t^{(j)} \mid \boldsymbol{x}_{n-1}^{A_{t-1}^{(j)}})} \\
w_t(\boldsymbol{x}_{1:t}^{(j)}) &= u_t(\boldsymbol{x}_{1:t}^{(j)})\frac{W_{t-1}(\boldsymbol{x}_{n-1}^{A_{n-1}^{(j)}})}{\alpha_{t-1}(\boldsymbol{x}_{t-1}^{A_{t-1}^{(j)}})} = \frac{\pi_t(\boldsymbol{x}_t^{(j)})}{g_1(x_1^{A_1^{(j)}}) \prod_{i=1}^{t} g_i(x_i^{(j)} \mid \boldsymbol{x}_{i-1}^{A_{i-1}^{(j)}}) \prod_{i=1}^{t-1} \alpha_i(\boldsymbol{x}_{i-1}^{A_{i-1}^{(j)}})}
\end{aligned}
$$

**and**

$$
W_t^{(j)} = \frac{w_t(\boldsymbol{x}_t^{(j)})}{\sum_{j=1}^{m} w_t(\boldsymbol{x}_t^{(j)})}
$$

**Why is it beneficial?**

**In fact, flexible resampling is nothing but changing the intermediate distribution.**

**Under flexible resampling scheme, the new intermediate distribution is**

$$\pi_t^*(\boldsymbol{x}_t) \propto \prod_{i=1}^{t-1} [g_i(x_i \mid \boldsymbol{x}_{i-1}) \alpha_i(\boldsymbol{x}_{i-1})]$$

**When**

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

**we get back** $\pi_t^*(\boldsymbol{x}_t) = \pi_t(\boldsymbol{x}_t)$**.**

- **Often, there are natural intermediate distributions.**
  - **In state space model,** $\pi_t(\boldsymbol{x}_t) = p(\boldsymbol{x}_t \mid y_1, \ldots, y_t)$**.**
- **Often, the intermediate distributions guides the design of** $g_t(x_t \mid \boldsymbol{x}_{t-1})$
  - $g_t(x_t \mid \boldsymbol{x}_{t-1})$ **close to** $\pi_t(x_t \mid \boldsymbol{x}_{t-1})$
- **The design of** $\alpha_t(\boldsymbol{x}_t)$ **can depend on the current samples of** $\boldsymbol{x}_t$**. Adaptivity.**
  - $\alpha_t(\boldsymbol{x}_t) = w_t^{\beta_t}(\boldsymbol{x}_t)$ **where** $\beta_t$ **depends on the variance of the current weight (for example).**

**Optimal intermediate distribution:** $\pi_t(\boldsymbol{x}_t) = \pi_n(\boldsymbol{x}_t)$ **(the true marginal)**

**The variance becomes**

$$\int \frac{\pi_n^2(x_1)(\mu_1(x_1) - \mu)^2}{g_1(x_1)} dx_1 + \sum_{t=2}^{n} \int \frac{\pi_n(x_t \mid \boldsymbol{x}_{t-1})}{g_t(x_t \mid \boldsymbol{x}_{t-1})} \pi_n(\boldsymbol{x}_{t-1})(\mu_t(\boldsymbol{x}_t) - \mu)^2 dx_{1:t}$$

**Almost like each step is from the true distribution** $\pi_n(\boldsymbol{x}_t)$**.**

**Delayed resampling:**

$$\pi^*(\boldsymbol{x}_t) = \pi_{t+\delta}(\boldsymbol{x}_t)$$

**with**

$$\alpha_t(\boldsymbol{x}_t) = \frac{\pi_{t+\delta}(\boldsymbol{x}_t)}{g_t(x_t \mid \boldsymbol{x}_{t-1})\alpha_{t-1}(\boldsymbol{x}_{t-1})}$$

**Or an approximated delayed resampling**

$$\pi^*(\boldsymbol{x}_t) = \hat{\pi}_{t+\delta}(\boldsymbol{x}_t)$$

**If**

$$\alpha_t(\boldsymbol{x}_t) = \frac{\hat{\pi}_n^{(t)}(\boldsymbol{x}_t)}{g_t(x_t \mid \boldsymbol{x}_{t-1})\alpha_{t-1}(\boldsymbol{x}_{t-1})}, t = 1, \ldots, n-1$$

**(achievable in certain cases, e.g. backward pilot) then**

$$\int \frac{\pi_n^2(x_1)(\mu_1 - \mu)^2}{g_1(x_1)} dx_1 + \sum_{t=2}^{n} \int \frac{\pi_n^2(\boldsymbol{x}_t)(\mu_t - \mu)^2}{\hat{\pi}_n^{(t)}(\boldsymbol{x}_{t-1}) g_t(x_t \mid \boldsymbol{x}_{t-1})} dx_{1:t}$$

**The difference is between** $\pi_{t-1}(\boldsymbol{x}_{t-1})$ **and** $\hat{\pi}_n^{(t)}(\boldsymbol{x}_{t-1})$

## Combined sampling and resampling scheme: (discrete state space)

- **If $x_t$ takes values in $\{a_1, \ldots, a_k\}$**
- **Evaluate $\alpha_t(\boldsymbol{x}_{t-1}^{(j)}, a_i)$, $i = 1, \ldots, k$, $j = 1, \ldots, m$.**
- **Sample $m$ distinct samples from $\{(\boldsymbol{x}_{t-1}^{(j)}, a_i), i = 1, \ldots, k, j = 1, \ldots, m\}$ with probability proportional to $\alpha_t(\boldsymbol{x}_{t-1}^{(j)}, a_i)$.**
- **Update weights**

## Application: SALs

- **Starting and Ending at** $(0, 0)$
- **Intermediate distributions**

  $\pi_t(\boldsymbol{x}_t)$**: uniform of all SAW such that** $d(\boldsymbol{x}_t) < n - t$ **(support)** **where** $d(\boldsymbol{x}_t) = |x_{t,1}| + |x_{t,2}|$
- **Combined sampling and resampling**
  - **Freedom:** $\delta(\boldsymbol{x}_t) = n - t - d(\boldsymbol{x}_t)$
  - **Flexibility:**

$$\beta(\boldsymbol{x}_t) = \frac{|x_{t,1}|}{d(\boldsymbol{x}_t)} \frac{|x_{t,2}|}{d(\boldsymbol{x}_t)} (\delta(\boldsymbol{x}_t) + 1)$$

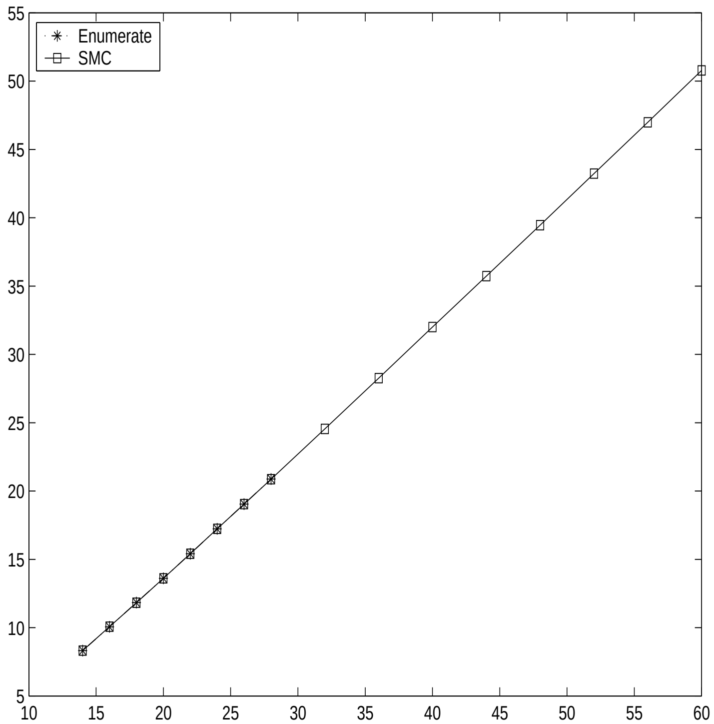
  - **Priority score**

$$\alpha_t(\boldsymbol{x}_t) = w_{t-1} exp \left\{ - \left[ c_1 + \frac{\delta(\boldsymbol{x}_t)^{-c}}{T_{1t}} + \frac{\beta(\boldsymbol{x}_t)}{T_{2t}} \right] \right\}.$$

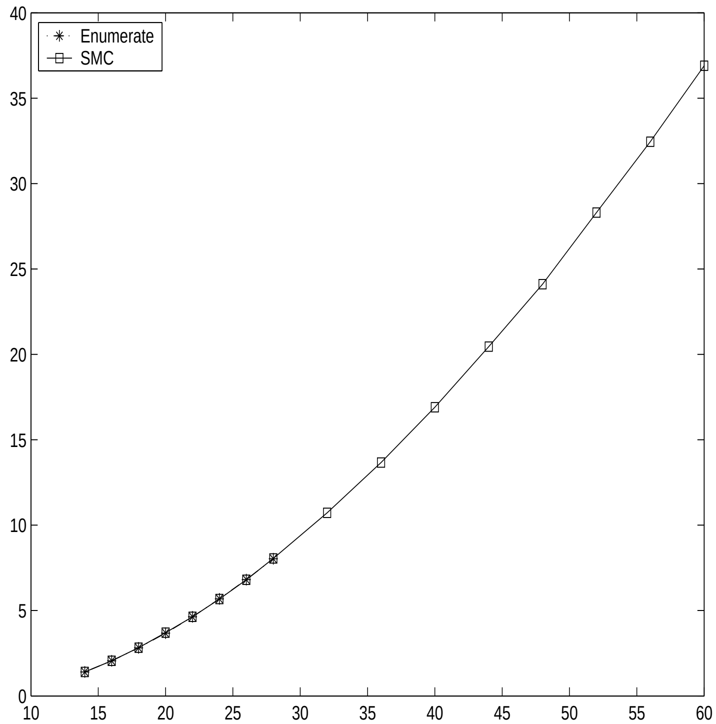
    **with temperature sequences** $T_{1t}$ **and** $T_{2t}$**.**

## SAL:

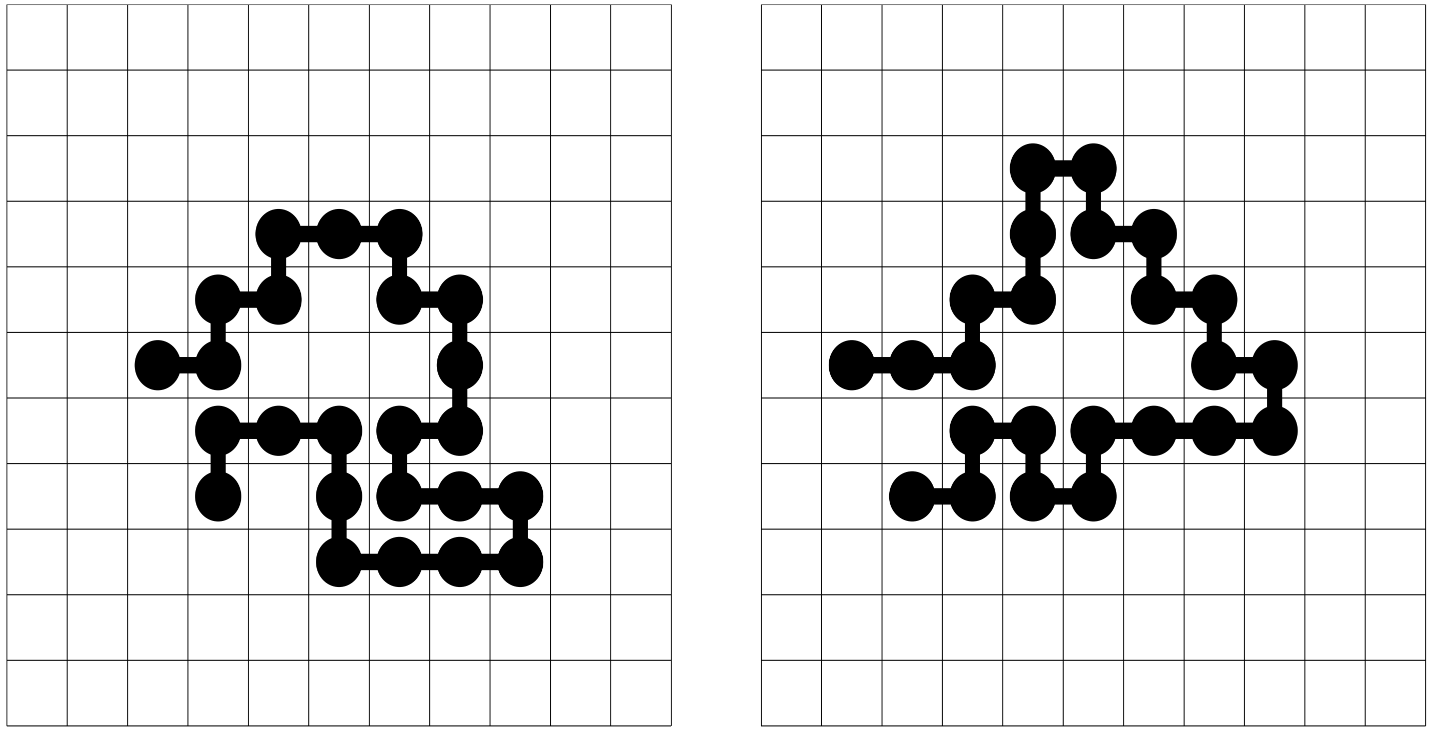
**Log(total number)**




**# of inside voids (mean)**



# Example: SAW with shape-specific void

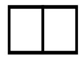


- **Let $\Omega$ be the set of all length-$n$ SAWs.**
- **Let $C_\nu$ be the set of all length-$n$ conformation with void $\nu$**
- **Estimate:**

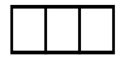
$$P(\boldsymbol{x}_n \in C_\nu \mid \boldsymbol{x}_n \in \Omega) = \frac{\sum_{\boldsymbol{x}_n \in C_\nu} 1}{\sum_{\boldsymbol{x}_n \in \Omega} 1}$$

- **Problem: Grow a SAW of length-$n$ in $C_\nu$**
- **One possible solution: rejection method – too inefficient**

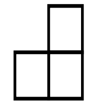
- **Intermediate distributions: order of growth**
  - **Select the monomers on the void wall first**
  - **Then grow the segments between the monomers on the wall**
- **Sampling distribution:**
  - **Self-avoiding**
  - **Shrinking support – distance, connectivity**
  - **Lookahead**
- **Resampling score: freedom and flexibility**



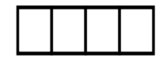
void 2.1



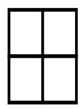
void 3.1



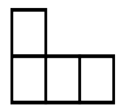
void 3.2



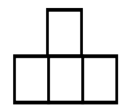
void 4.1



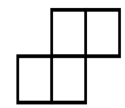
void 4.2



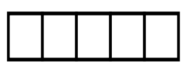
void 4.3



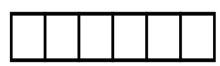
void 4.4



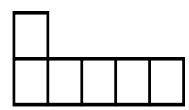
void 4.5



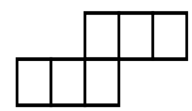
void 5.1



void 6.1

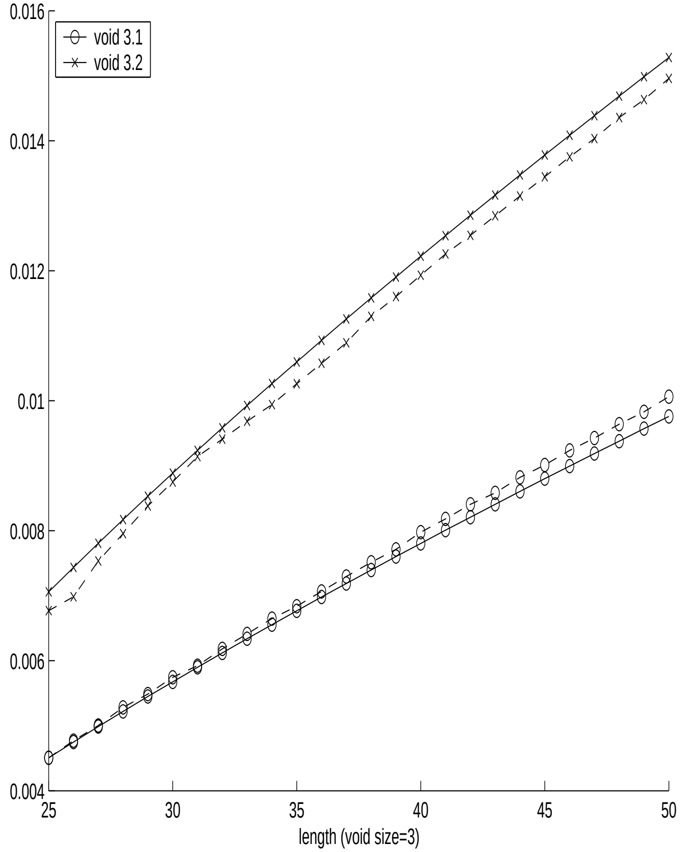


void 6.2

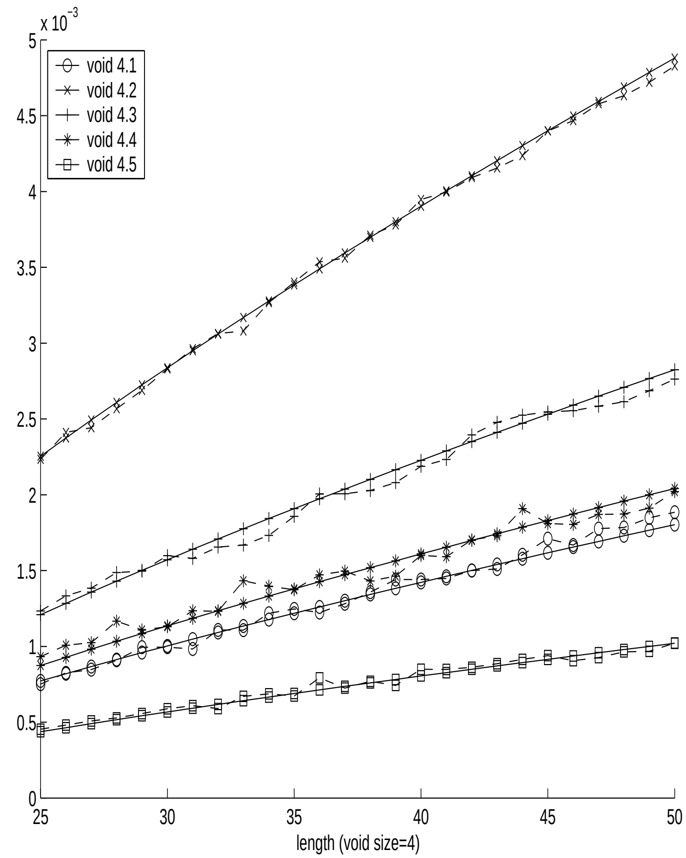


void 6.3

# Fraction of conformations:




,



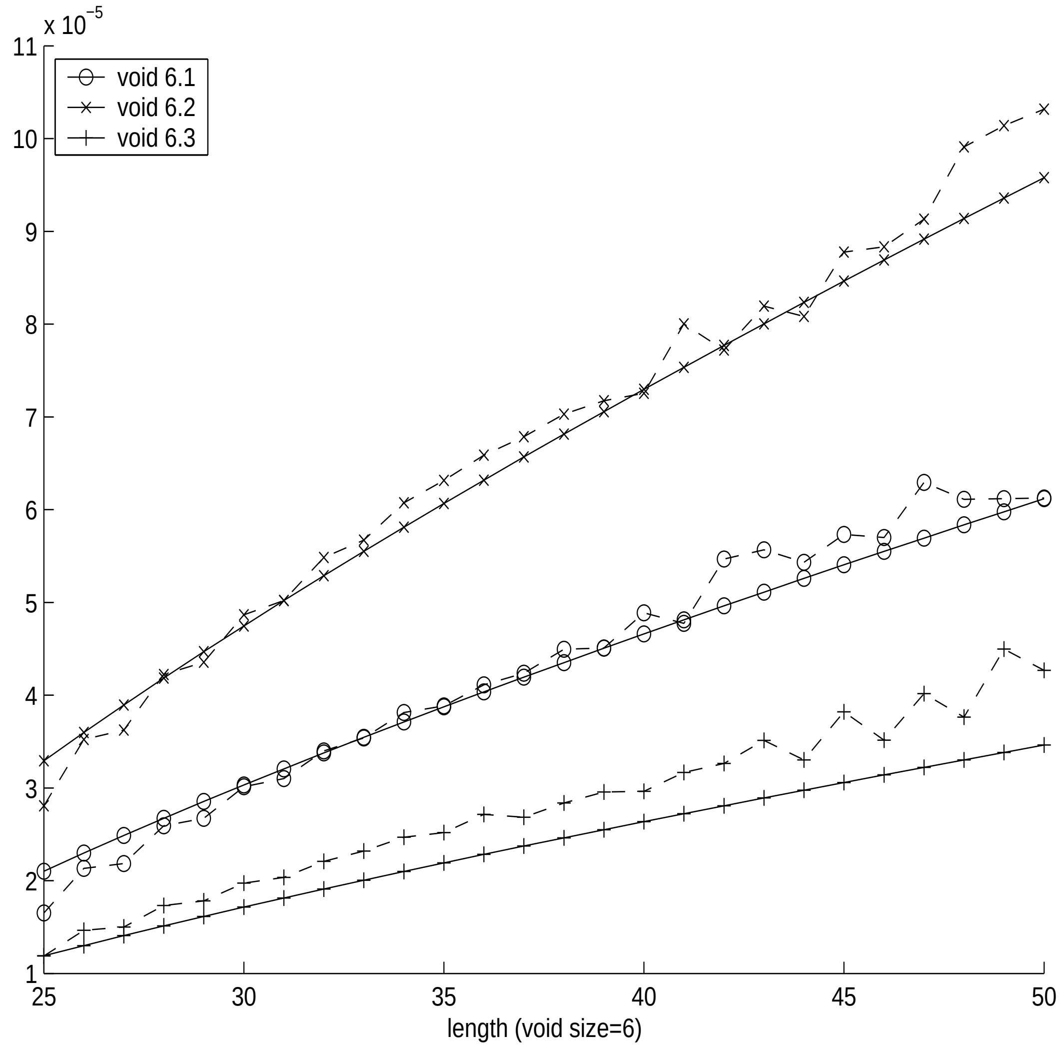
**Fraction of conformation**

$$\hat{f}(\nu, n) = c_1 r(\nu)[(1 - c_2 e(\nu)) c_3^{-w(\nu)+14} (n - w(\nu) + 1)^{c_4}]$$

**where**

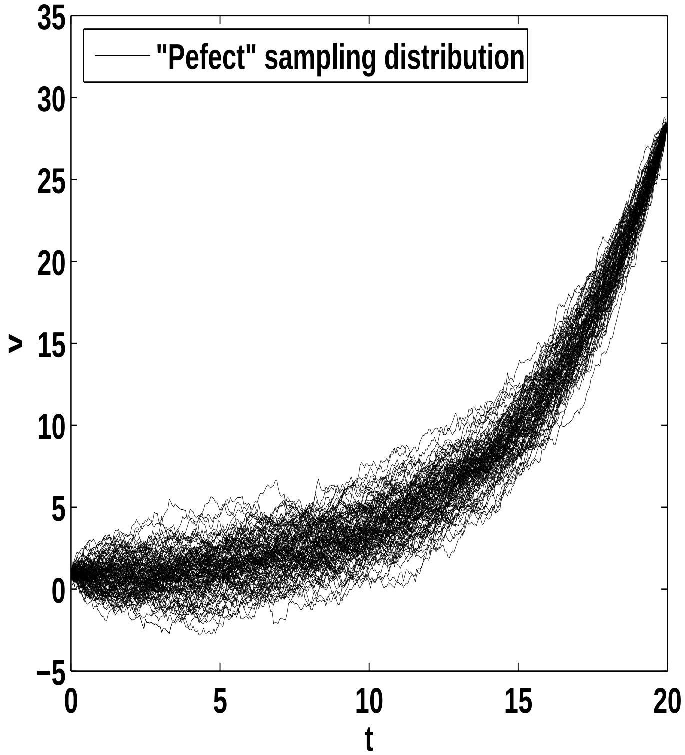
- $w(\nu)$: **wall size**
- $e(\nu)$: **number of outer corners**
- $r(\nu)$: **number of different rotational transformations**

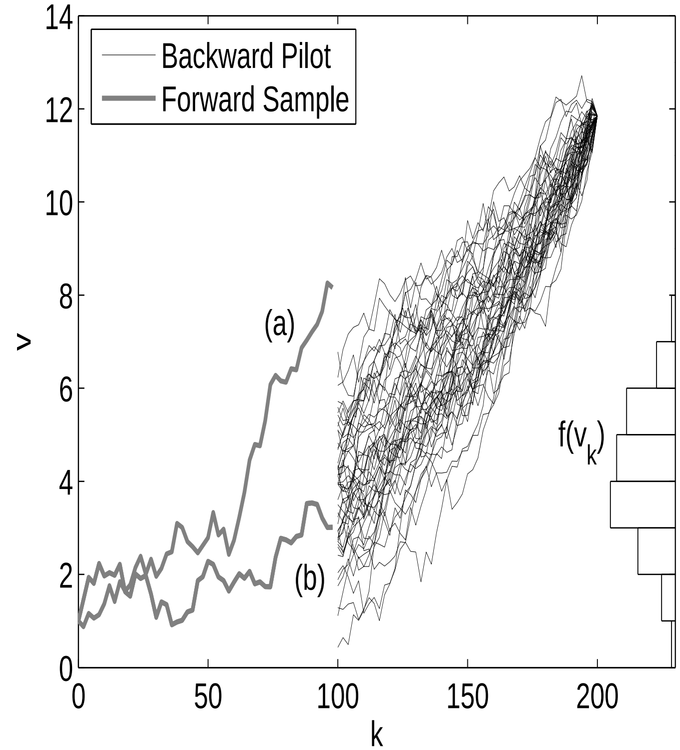
# Out-sample prediction



## Example: Generating Samples of Diffusion Bridges

- **Generate** $p(x_1, \ldots, x_{n-1} \mid x_0 = a, x_n = b)$
- **Sequential:** $p(x_t \mid x_0, x_n, \boldsymbol{x}_{t-1}) \propto p(x_t \mid x_{t-1}) p(x_n \mid x_t)$
- **Use backward pilots to estimate** $p(x_n \mid x_t)$**.**
- **resampling according to** $\alpha(x_t) = w_t(x_t)\hat{\pi}_n(x_t \mid x_n)$



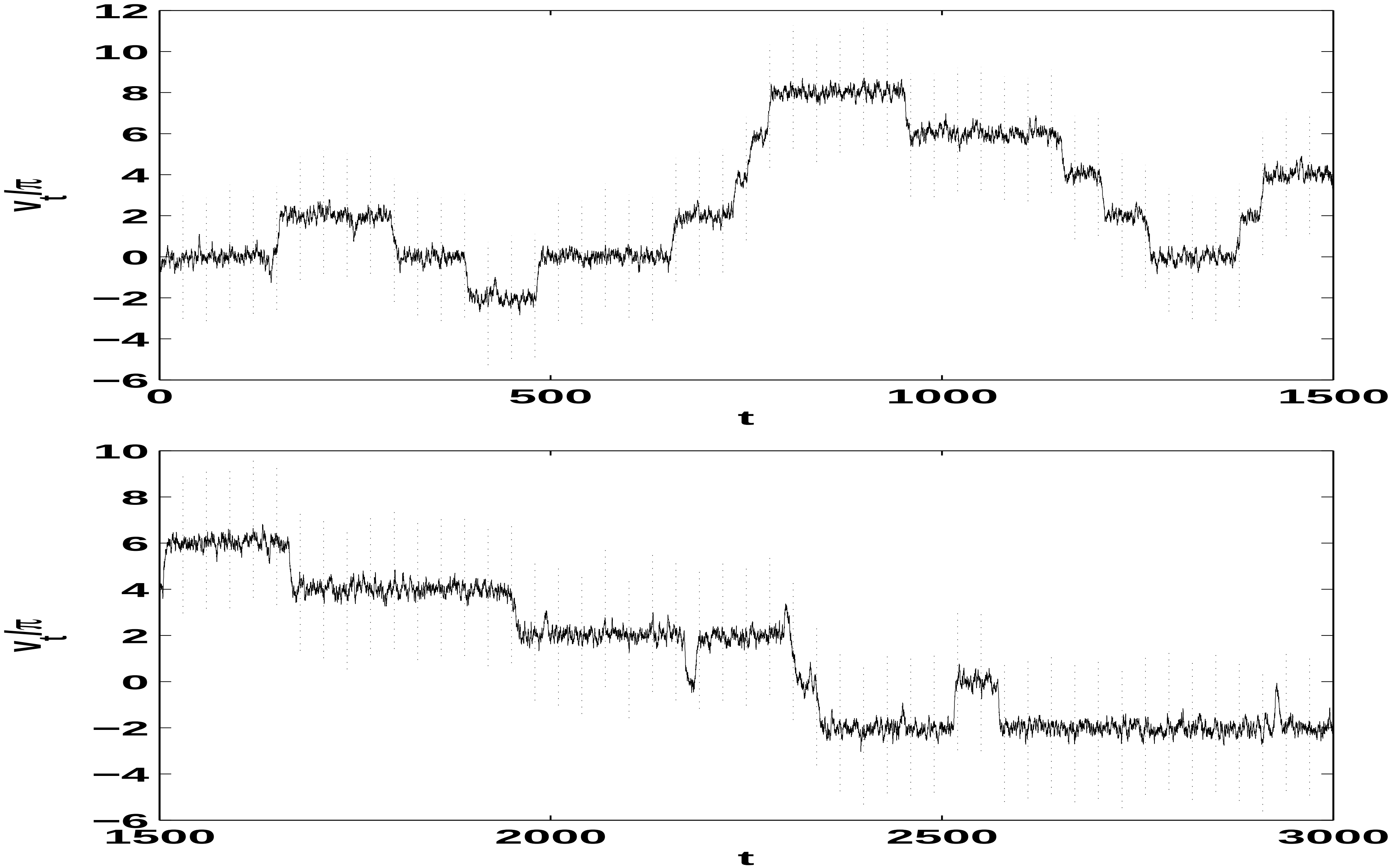
## Example (Beskos et al. 2006)

$$dv_t = sin(v_t - \theta)dt + dw_t$$

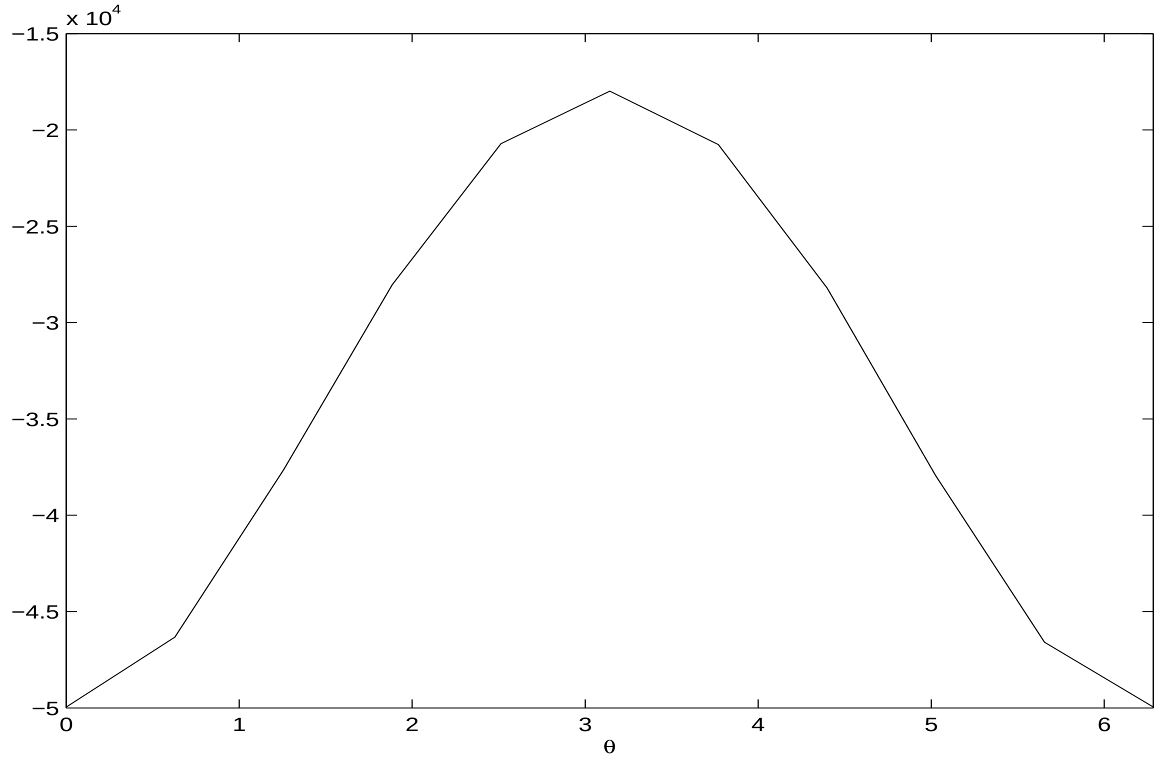
- **Comparison between 'exact sampling' (Beskos et al. 2006), SMC-0 and SMC-1.**
- **100 realizations. Stepsize 0.001.**
- **Performance measure $\tilde{L}(\theta)$: exact sampling, 10M samples.**

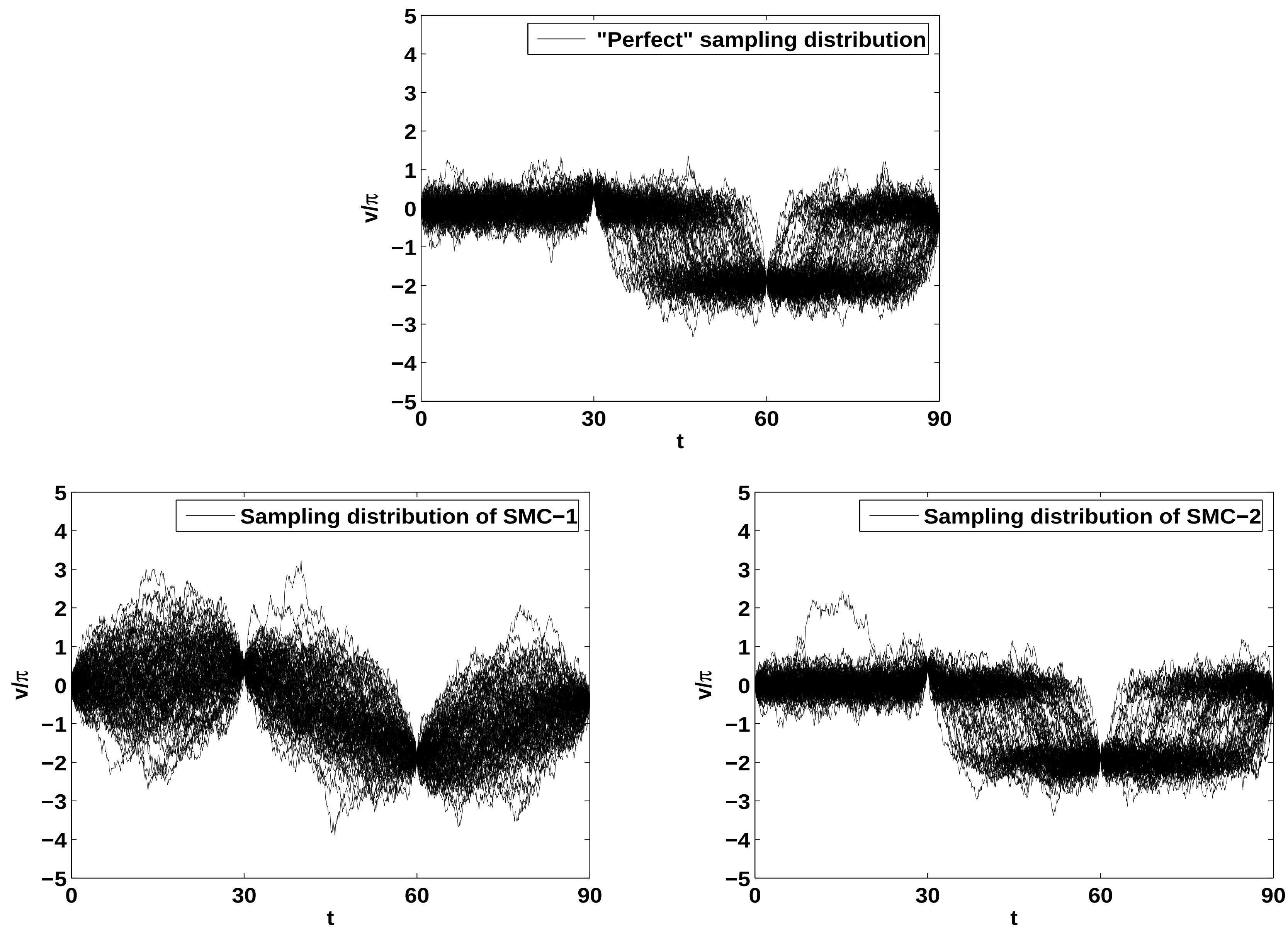
$$\mathbf{RMSE}(\theta) = \left[ \frac{1}{100} \sum_{i=1}^{100} (\hat{L}_i(\theta) - \tilde{L}(\theta))^2 \right]^{1/2}$$

- **Observation step size $\Delta = 30$**
- **Eular approximation**
- **Roughly same CPU time**


12
10
8
6
4
2
0
−2
−4
−6
0
500
1000
1500
t
$v_t/\pi$
10
8
6
4
2
0
−2
−4
−6
1500
2000
2500
3000
t
$v_t/\pi$

## (Average) Likelihood function:

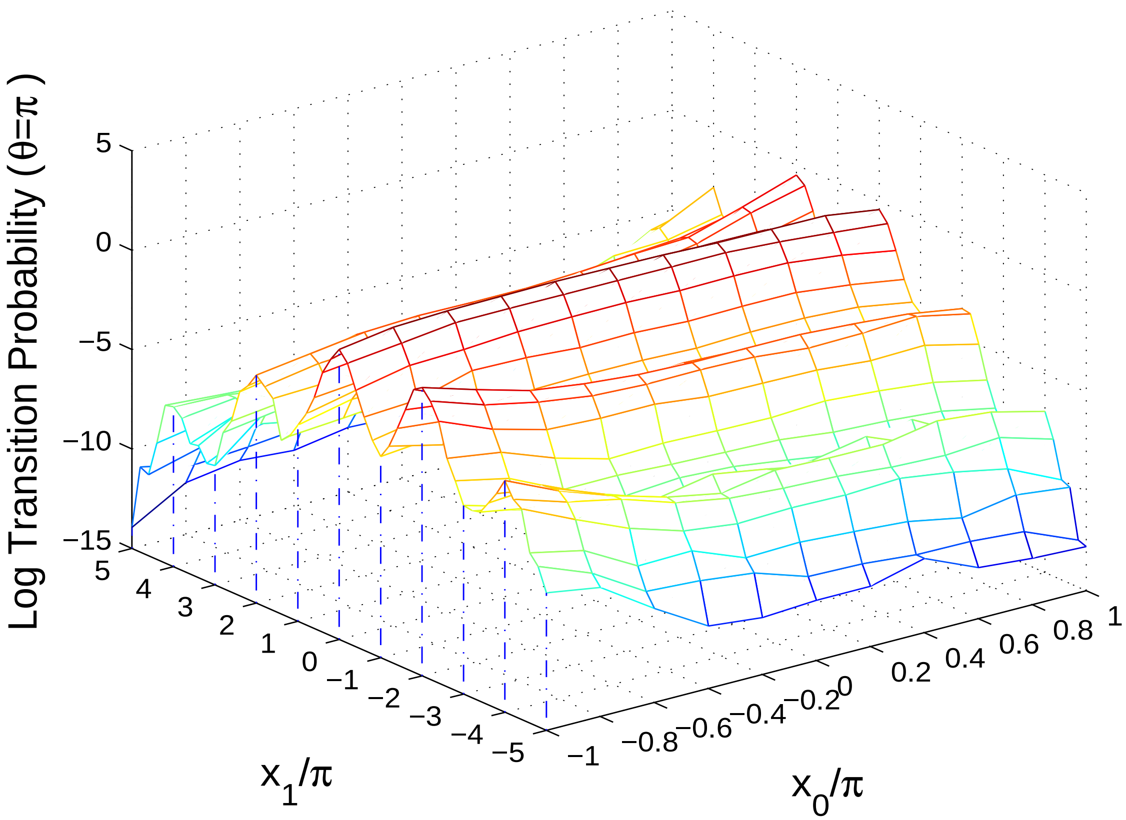



"Perfect" sampling distribution
Sampling distribution of SMC−1
Sampling distribution of SMC−2
v/π
t
5
4
3
2
1
0
−1
−2
−3
−4
−5
0
30
60
90

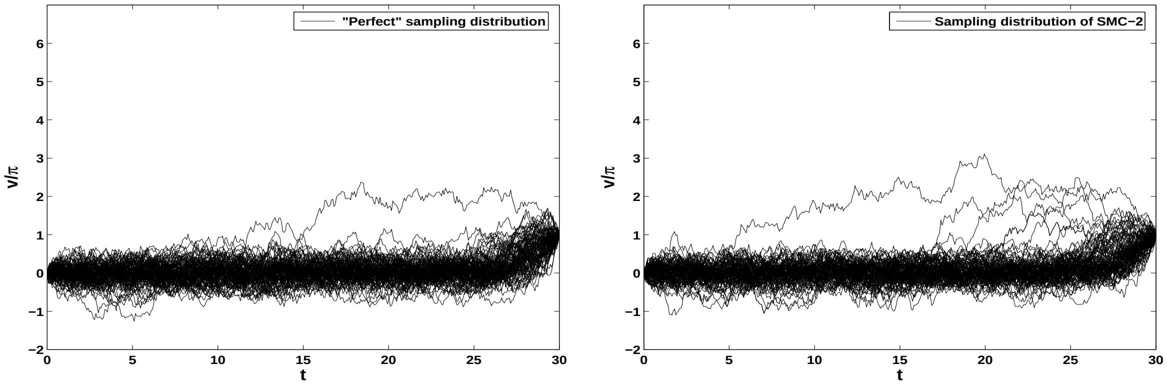
**Estimation of the likelihood function:**

| $RMSE$ | **exact** | **SMC-0** | **SMC-1** |
|---|---|---|---|
| $m$ | **80,000** | **3,500** | **1,000** |
| $\theta = 0.0\pi$ | **1.719** | **0.519** | **0.325** |
| $\theta = 0.2\pi$ | **1.488** | **0.497** | **0.291** |
| $\theta = 0.4\pi$ | **1.211** | **0.433** | **0.214** |
| $\theta = 0.6\pi$ | **0.901** | **0.397** | **0.157** |
| $\theta = 0.8\pi$ | **0.648** | **0.347** | **0.136** |
| $\theta = 1.0\pi$ | **0.588** | **0.331** | **0.122** |
| $\theta = 1.2\pi$ | **0.671** | **0.356** | **0.135** |
| $\theta = 1.4\pi$ | **0.870** | **0.399** | **0.165** |
| $\theta = 1.6\pi$ | **1.217** | **0.452** | **0.227** |
| $\theta = 1.8\pi$ | **1.573** | **0.507** | **0.299** |
| **time(sec.)** | **0.490** | **0.478** | **0.470** |

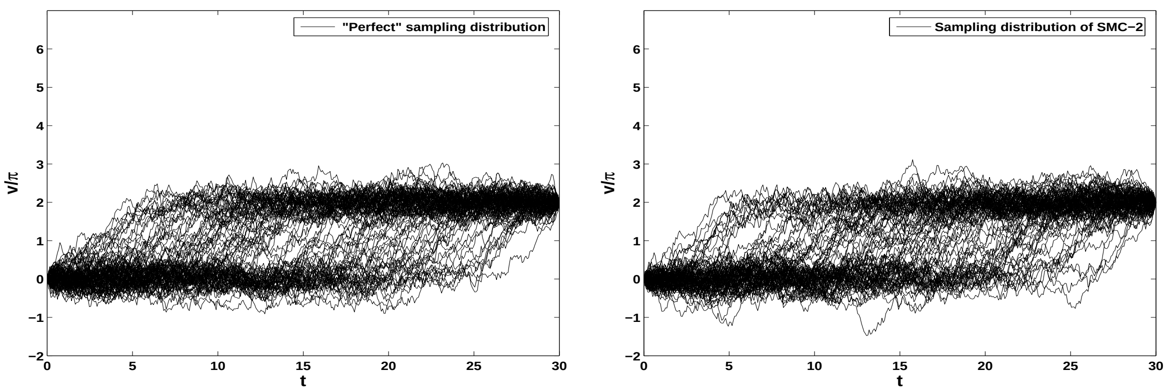
## Estimation of the log-transition density



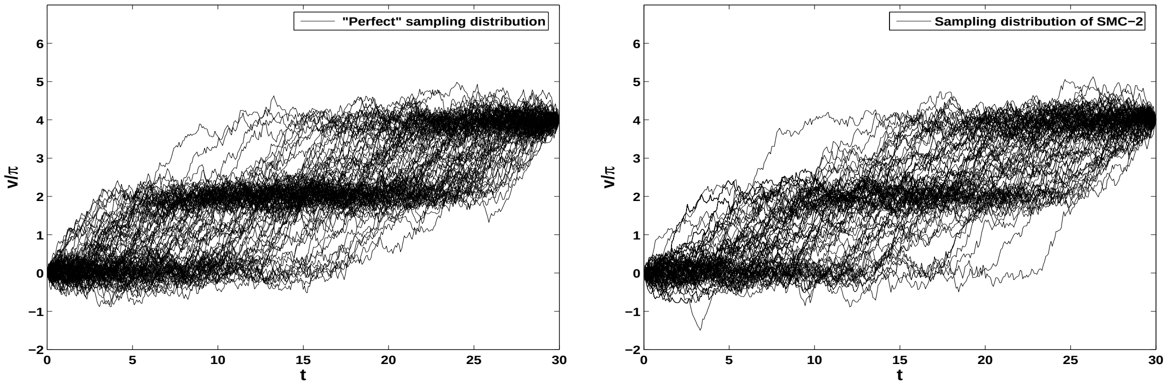
$y_0 = 0, y_n = \pi$



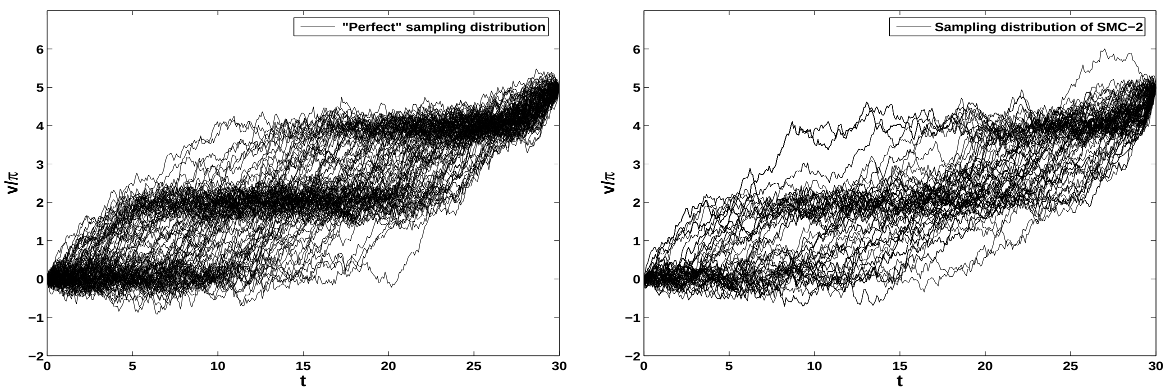

$y_0 = 0, y_n = 2\pi$



$$y_0 = 0, y_n = 4\pi$$




$$y_0 = 0, y_n = 5\pi$$



## Questions

- **What are the principles of designing the intermediate distributions or the equivalent resampling scheme?**
- **How do we know one is better than another?**
- **Trade-off between better intermediate distributions and complexity**
- **Rationalize some of the existing resampling schemes**

### 2.1.4 Inference

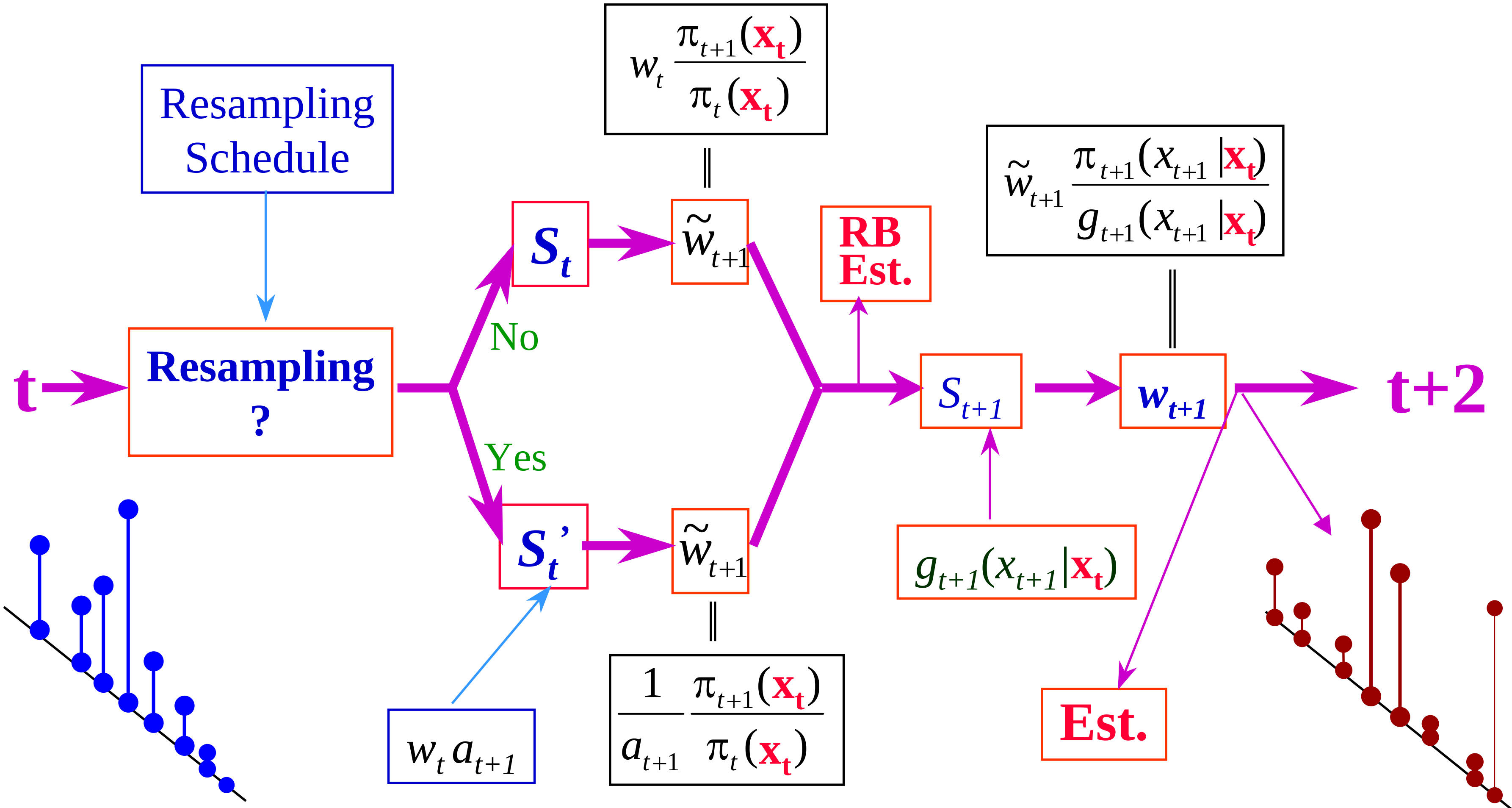
**Inference:**

$$\hat{E}_{\pi_t} h(\boldsymbol{x}_t) = \frac{\sum_{j=1}^m w_t^{(j)} h(\boldsymbol{x}_t^{(j)})}{\sum_{j=1}^m w_t^{(j)}}$$

- **Estimation should be done before a resampling step**
- **Rao-Blackwellization: For example, if $w_{t+1}$ does not depend on $x_{t+1}$, then**

$$\hat{E}_{\pi_{t+1}} h(x_{t+1}) = \frac{\sum_{j=1}^m w_{t+1}^{(j)} E_{\pi_{t+1}}(h(x_{t+1}) \mid \boldsymbol{x}_t^{(j)})}{\sum_{j=1}^m w_{t+1}^{(j)}}$$

- **Delayed estimation (i.e. $E_{\pi_t} h(x_{t-k})$ at time $t$) is usually more accurate since the estimation is based on more information.**
- **Frequent resampling may have adverse effect.**




A discrete approx.of prob. dist. at time t.

A discrete approx.of prob. dist. at time *t+1*.