SMC

Organization

- SMC on product spaces
- Transforming other problems into product spaces (sequential change of measure)

Motivations for SMC on product spaces

- Sequential predictions / streaming data / HMM / state space models
  - latent state from noisy observation
  - change point
- Time series where ‘time’ is not time
  - genomics: ‘time’ = position on genome
    - observations: SNP
    - latent: haploblock (chunk shared by several individuals)

Common feature: the latent space is a product space $F_t = E_1 \times E_2 \times \cdots \times E_t$ indexed by the integers $t \in \{1, \dots, n\}$.

Sequence of targets

- As in PT we now have a sequence of targets
  - but: with different dimensionality now vs. fixed dimensionality for PT
- In the product space context, sometimes we care about all targets (real time predictions), sometimes, we care only about the last one
- Typical problems:
  - integrating test functions
  - + computing normalization Z (e.g. for model selection, where Z = P(data)

Building block: sequential importance sampling

- Rewrite self-normalized importance sampling so that it can be done with a sequence of targets
- Use the following identities:

$$\gamma(x_{1:n}) = \frac{\gamma(x_{1:n})}{\gamma(x_{1:n-1})}\frac{\gamma(x_{1:n-1})}{\gamma(x_{1:n-2})}\cdots\frac{\gamma(x_{1:1})}{\gamma(x_{\emptyset})}, \qquad q(x_{1:n}) = q(x_1|x_{\emptyset})q(x_2|x_{1:1})q(x_3|x_{1:2})\ldots q(x_n|x_{1:n-1}),$$

- Yields the recursions

$$\begin{aligned} x_t^i &\sim q(\cdot|x_{1:t-1}^i) \\ x_{1:t} &= (x_{1:t-1}^i, x_t^i), \end{aligned} \qquad w_t^i = w_{t-1}^i \frac{\gamma(x_{1:t})}{\gamma(x_{1:t-1})}\frac{1}{q(x_t|x_{1:t-1})}.$$

- Does not work! (Why?) But forms basis of SMC

Fix: resampling

- Intuition: prune particles with low normalized weights
- Constraints: we still want consistency
- Idea: resample N times according to the normalized weights
  - *multinomial resampling*

Notation for our goals

Given a model (joint)...: $\gamma_t(\mathbf{x}_t) = p(\mathbf{x}_t, \mathbf{y}_t)$

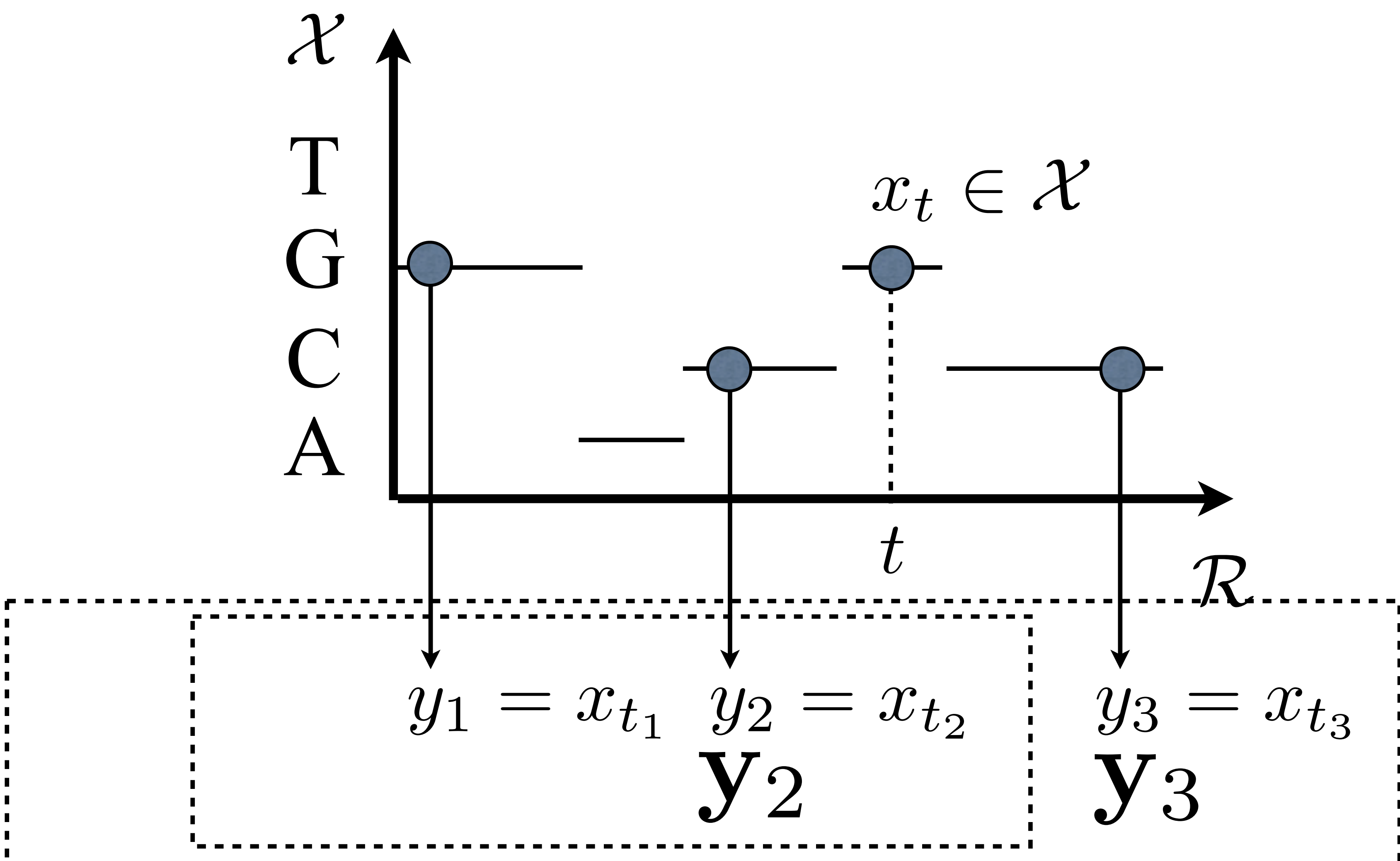
Sample from a *target distribution*: $\pi_t(\mathbf{x}_t) = p(\mathbf{x}_t | \mathbf{y}_t)$

$$\pi_t(\mathbf{x}_t) = \frac{\gamma_t(\mathbf{x}_t)}{Z_t}$$

.. and/or evaluate the normalization: $Z = p(\mathbf{y}_t)$

Notation

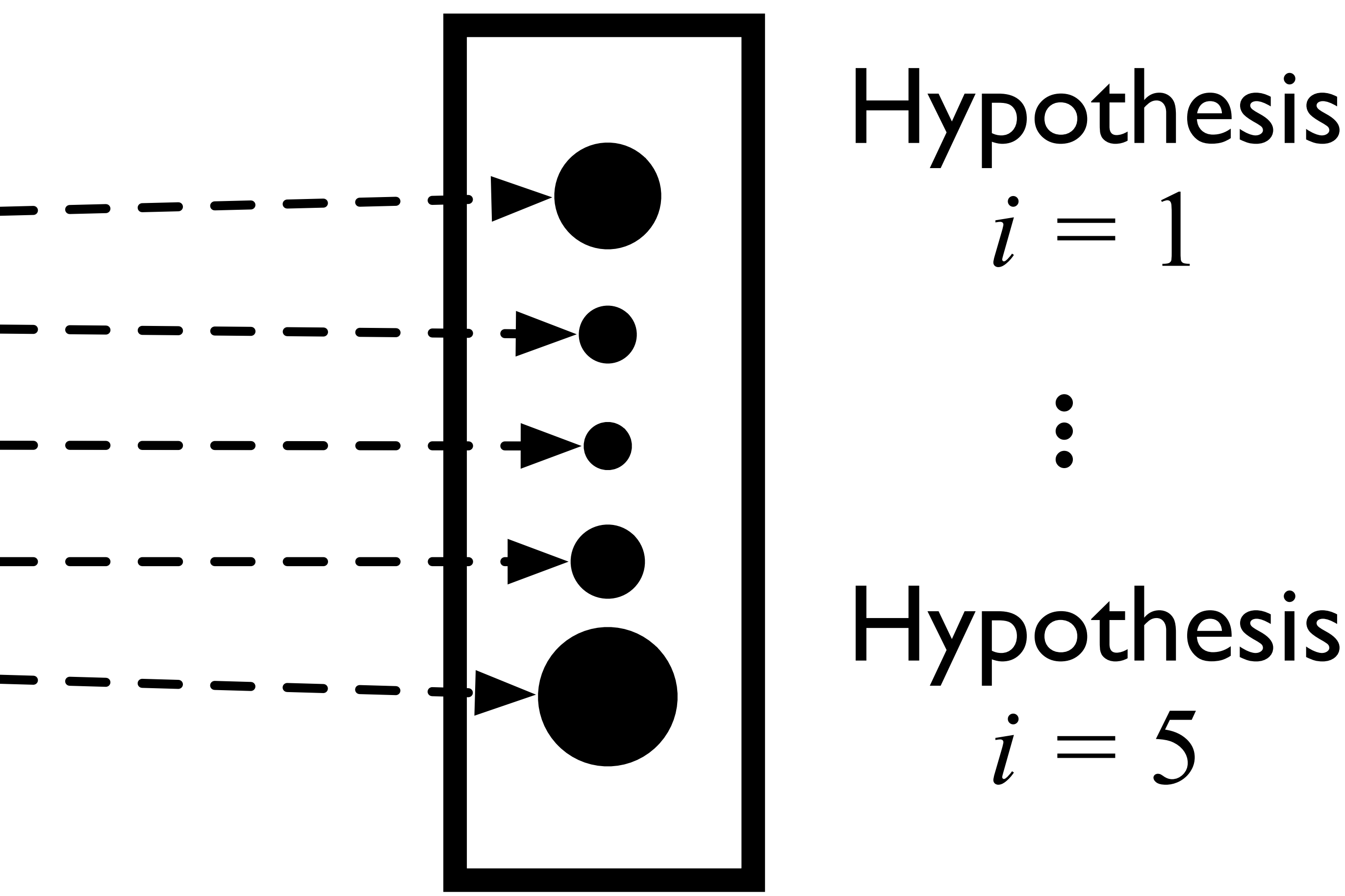
$\mathcal{X}$	State space
$x_t \in \mathcal{X}$	Point in that space
	Subscript: process index
$\mathbf{x}_t$	Many points in the state space
$\mathbf{y}_t$	Many observations



Standard SMC

Output: competing 'hypotheses' $\mathbf{x}_t^i$

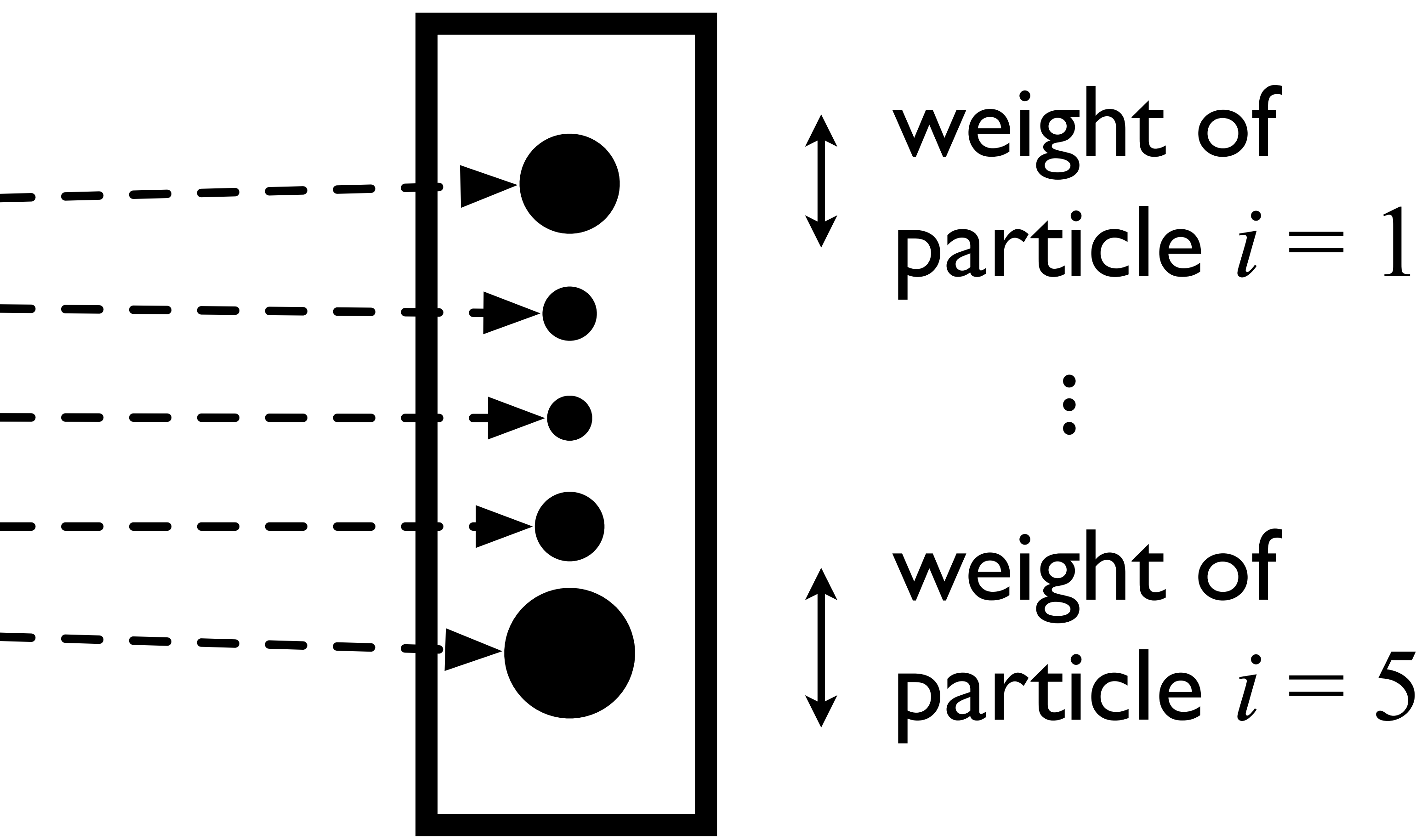
$t =$ last time observed



Standard SMC

Output: competing 'hypotheses' $\mathbf{x}_t^i$

weight for each of these w_t^i

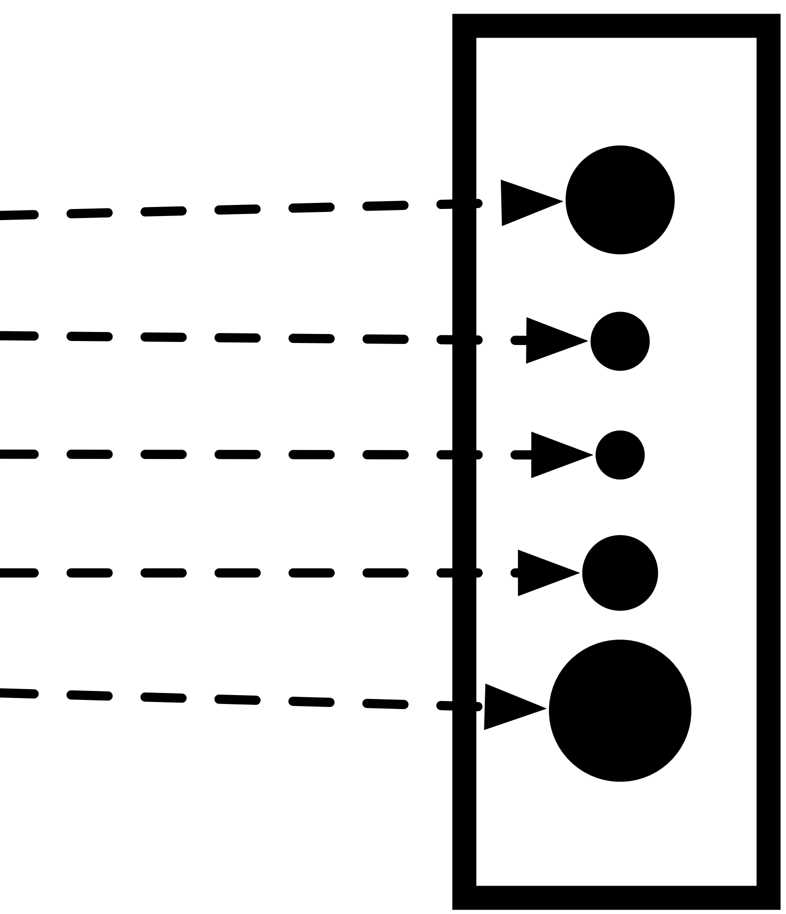


Standard SMC

Output: competing 'hypotheses' $\mathbf{x}_t^i$

weight for each of these w_t^i

Can view these as a (random) distribution

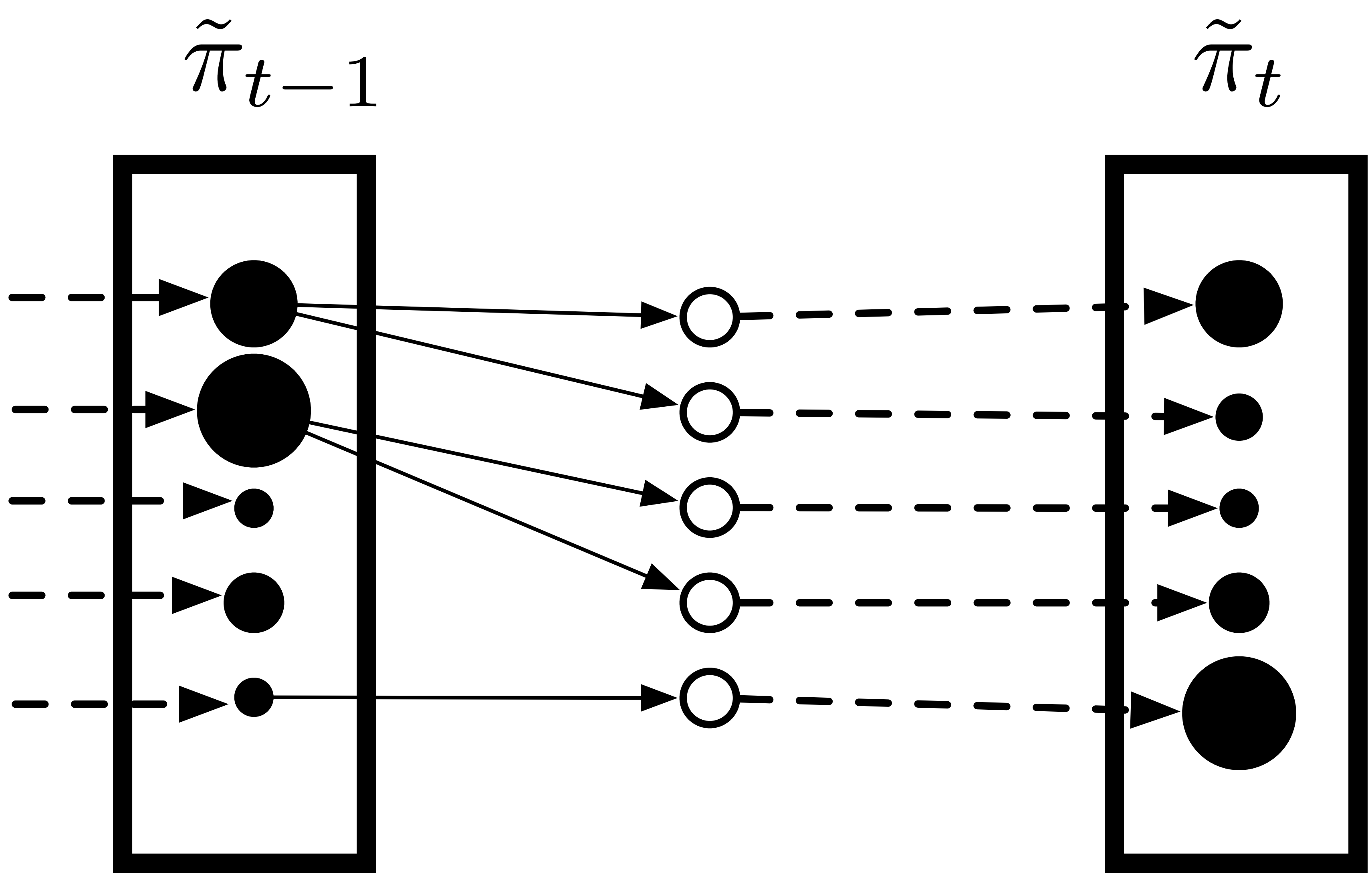


$$\tilde{w}_t^i = \frac{w_t^i}{\sum_j w_t^j}$$

$$\tilde{\pi}_t(\cdot) = \sum_i \tilde{w}_t^i \delta_{\mathbf{x}_t^i}(\cdot)$$

Standard SMC inner working: 1. Assume inductively that we have computed approximation for:

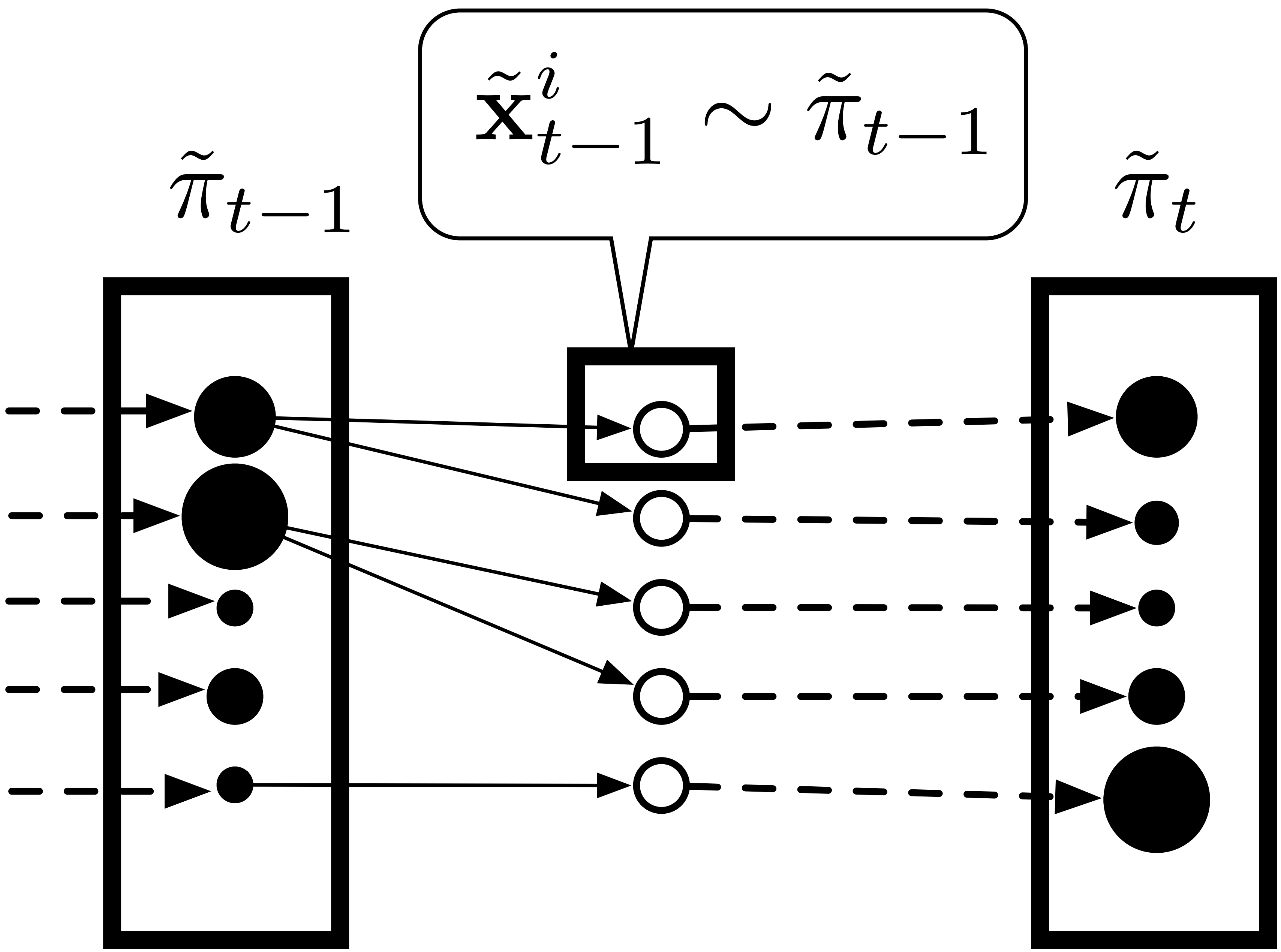
$$\pi_{t-1}(\mathbf{x}_{t-1}) = p(\mathbf{x}_{t-1}|\mathbf{y}_{t-1})$$



Standard SMC inner working:

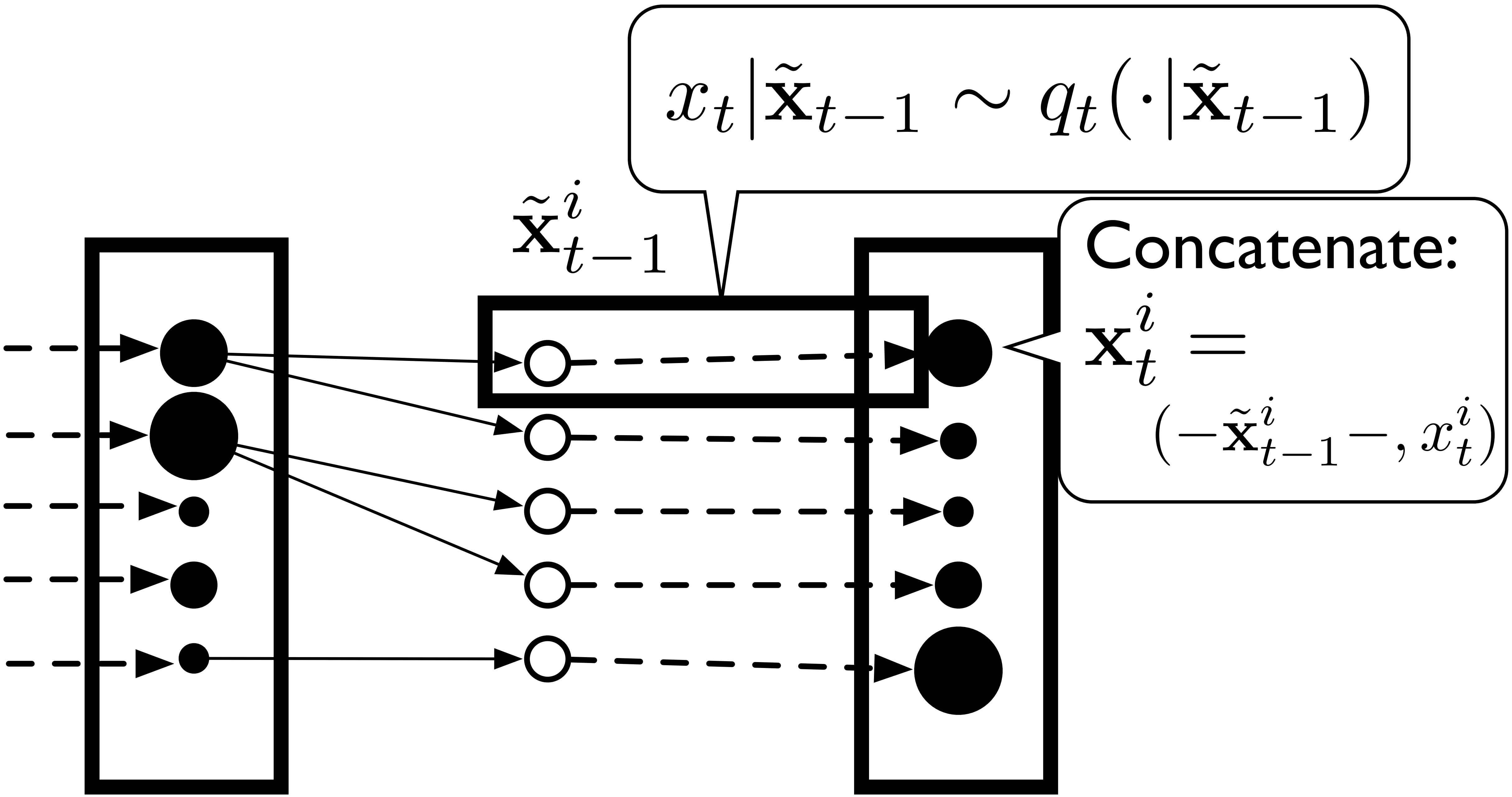
1. Assume inductively...

2. Sample from $\tilde{\pi}_{t-1}$



Standard SMC inner working:

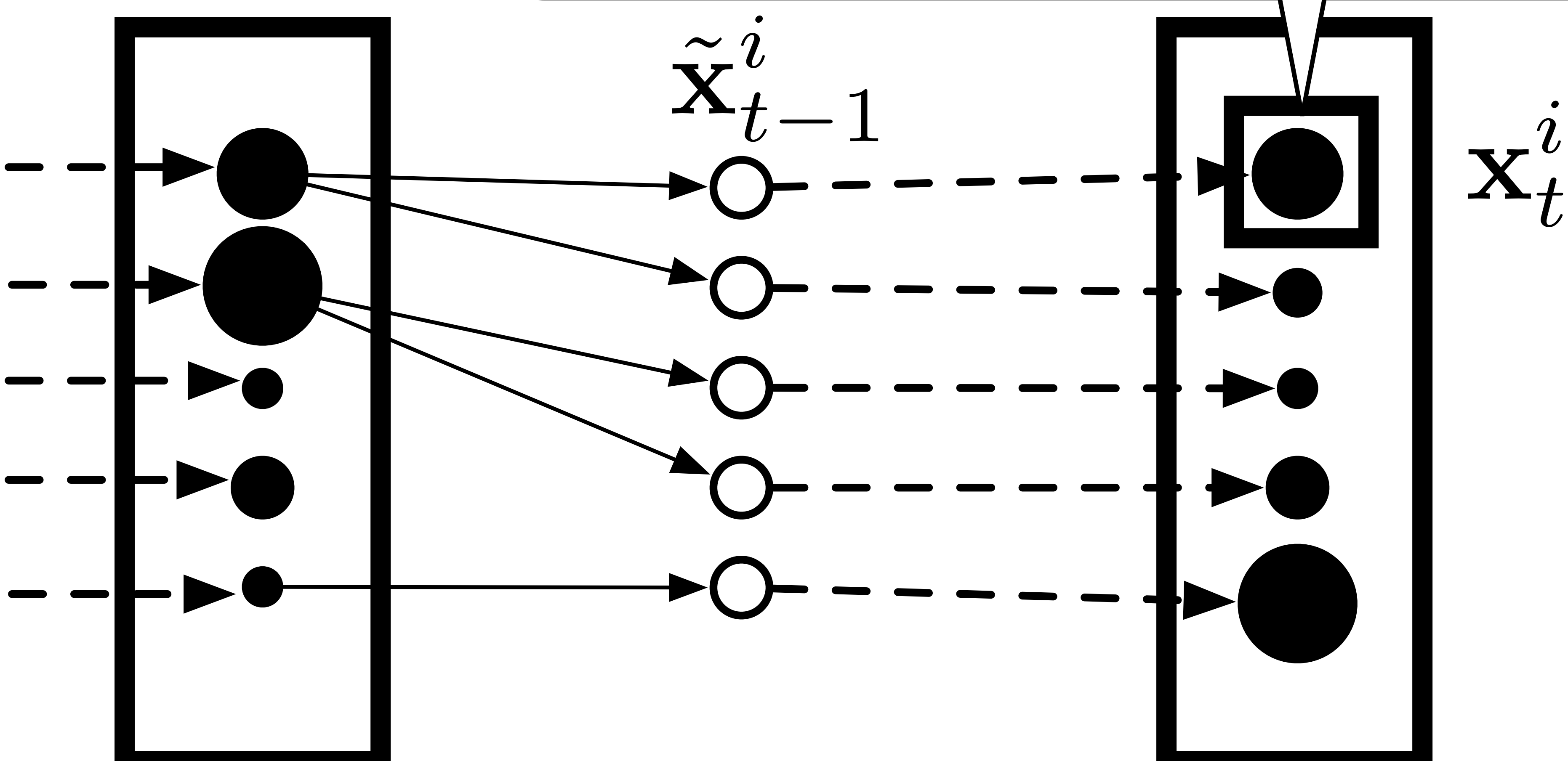
1. Assume inductively...
2. Sample from $\tilde{\pi}_{t-1}$
3. Propose (extend):



Standard SMC inner working:

1. Assume inductively...
2. Sample from $\tilde{\pi}_{t-1}$
3. Propose (extend)
4. Reweigh:

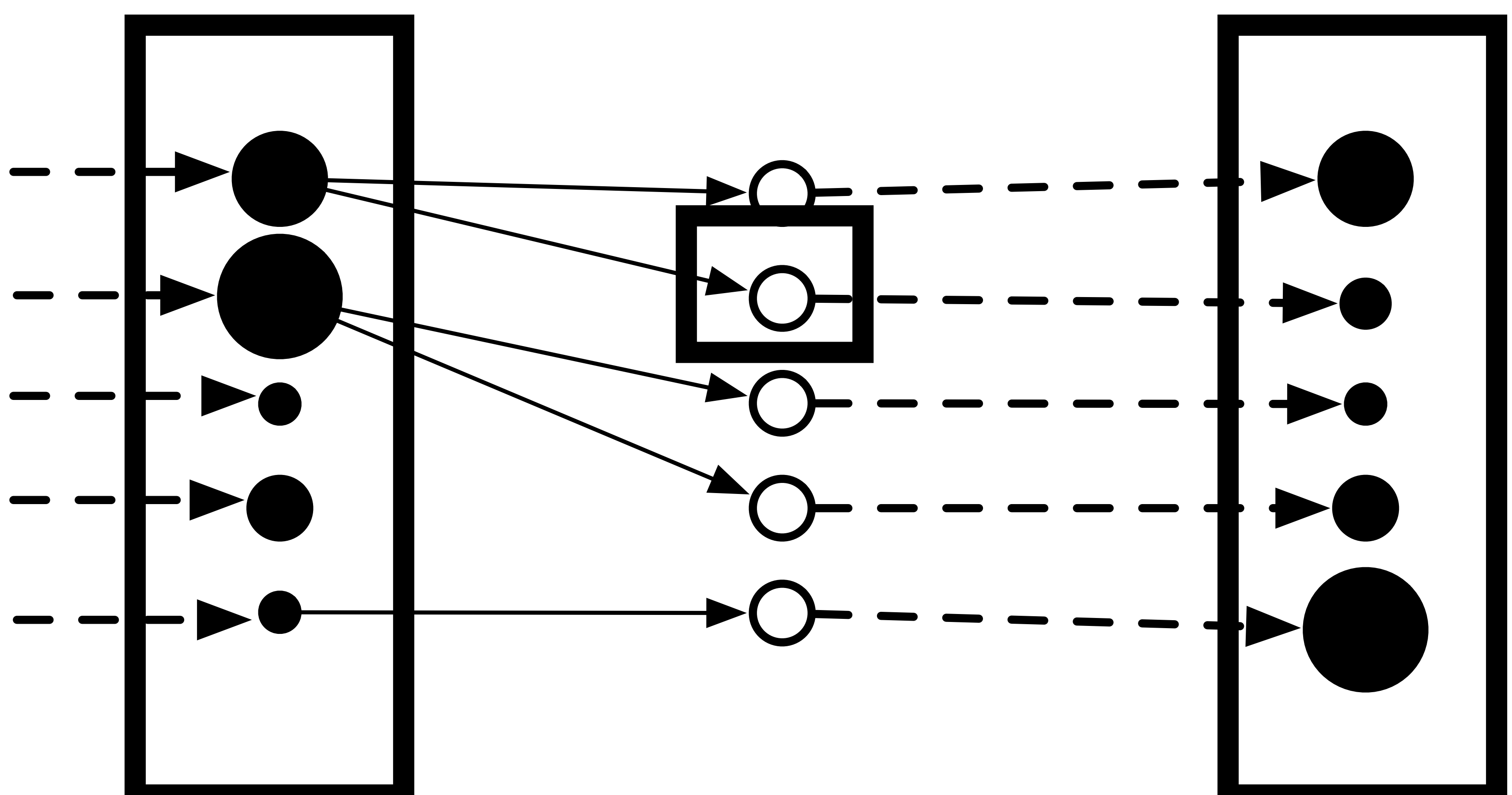
$$w_t^i = \frac{\pi_t(\mathbf{x}_t^i)}{\pi_{t-1}(\tilde{\mathbf{x}}_{t-1}^i)} \frac{1}{q_t(x_t^i|\tilde{\mathbf{x}}_{t-1}^i)}$$



Standard SMC inner working:

1. Assume inductively...
2. Sample from $\tilde{\pi}_{t-1}$
3. Propose (extend)
4. Reweigh

Repeat for each particle (5 times)



Normalization constant estimate

- Unbiased estimate provided by the product of the unnormalized weight averages:

$$\hat{C} = \prod_t \frac{1}{5} \sum_{i=1}^{5} w_t^i$$

Some pointers

- Theory: see Del Moral, 2013 for LLN, CLT
- How to build MC intervals: see J. Olsson, R. Douc (2018)
- Proposals:
  - sometimes, forced to pick dynamics
  - else, various options, e.g. *lookahead proposal*

Resampling

- Efficient implementation: can be done in linear time in the number of particles (via spacings of a Poisson process, see Devroye's book on random generation)
- Often important not perform resampling at every step
  - Monitor *relative ESS* after each proposal round

$$\frac{(E_q[\tilde{W}])^2}{E_q[\tilde{W}^2]} \approx \frac{(\frac{1}{n}\sum \tilde{W}^{(i)})^2}{\frac{1}{n}\sum (\tilde{W}^{(i)})^2}$$

  - Resample when it drops under a threshold (0.5) typically
- Finally, alternatives to multinomial resampling exist, see Mathieu Gerber, Nicolas Chopin, Nick Whiteley, 2017 for recent analysis of those

Organization

- SMC on product spaces
- **Transforming other problems into product spaces (sequential change of measure)**

AIS / Jarzynski's trick

- Target spaces F_t, not product spaces,
  - important e.g. $F_t = S$ (change of measure)
- Auxiliary spaces:

$$S_{1:n} = S \times S \times \ldots S$$

- Distribution on those? Use a *backward kernel B*

$$\pi_{1:n}(x_{1:n}) = \pi_n(x_n) \prod_{m<n} B_m(x_m | x_{m+1})$$

- Get weight update:

$$\tilde{w}(x_{1:n-1}, x_{1:n}) = \frac{\gamma_n(x_n)}{\gamma_{n-1}(x_{n-1})} \frac{B_{n-1}(x_{n-1}|x_n)}{K_n(x_n|x_{n-1})}$$

Example

- Setup: change of measure on annealed distributions
- K_n: π_n invariant kernel (from MH)
- Problem: cannot compute weight in general

$$\tilde{w}(x_{1:n-1}, x_{1:n}) = \frac{\gamma_n(x_n)}{\gamma_{n-1}(x_{n-1})} \frac{B_{n-1}(x_{n-1}|x_n)}{K_n(x_n|x_{n-1})}$$

- Idea: use fact we are free to pick B as we wish; use

$$B_{n-1}(x_{n-1}|x_n) = \frac{\pi_n(x_{n-1})K_n(x_n|x_{n-1})}{\pi_n(x_n)}$$

- Weight update simplifies (check)

$$\tilde{w} = \frac{\gamma_n(x_{n-1})}{\gamma_{n-1}(x_{n-1})}$$