# Sequential Bayesian estimation and Particle filters.

Dmitriy Shutin

dshutin@tugraz.at.

Signal Processing and Speech Communication Laboratory

www.spsc.tugraz.at

Graz University of Technology

# Content of the talk

- Concept of Sequential Bayesian estimation.
  - Linear-Gaussian case: Kalman filter.
  - Nonlinear case: EKF, UKF.
- Particle filters.
  - Monte Carlo methods and Importance sampling.
  - MC for sequential estimation: SIS algorithm.
  - Resampling: SIR algorithm.
- Conclusions.

# Sequential Bayesian estimation

- A system is described by a state vector $\boldsymbol{x}_n$
- The evolution of the state sequence $\{\boldsymbol{x}_n, n \in \mathbb{N}\}$

$$\boldsymbol{x}_n = \boldsymbol{f}_n(\boldsymbol{x}_{n-1}, \boldsymbol{v}_{n-1}) \quad \boldsymbol{x}_n \in \mathbb{C}^{N_x}$$

- Observation of the system output $\boldsymbol{z}_n$ is a function of the system state $\boldsymbol{x}_n$

$$\boldsymbol{z}_n = \boldsymbol{h}_n(\boldsymbol{x}_n, \boldsymbol{w}_n), \quad \boldsymbol{z}_n \in \mathbb{C}^{N_z}$$

- $\boldsymbol{w}_n \in \mathbb{C}^{N_w}$ and $\boldsymbol{v}_n \in \mathbb{C}^{N_v}$, $n \in \mathbb{N}$, are i.i.d noise sequences.
- Functions $\boldsymbol{h}_n(\cdot)$ and $\boldsymbol{f}_n(\cdot)$ are not necessarily linear.

# Sequential Bayesian estimation, cont'd

Our task is to estimate the current state $\boldsymbol{x}_n$ at time $n$ given measurements $\boldsymbol{z}_{1:n}$ up to time $n$:

Compute $p(\boldsymbol{x}_n|\boldsymbol{z}_{1:n})$, assuming $p(\boldsymbol{x}_0|\boldsymbol{z}_0) \equiv p(\boldsymbol{x}_0)$

Let us assume $p(\boldsymbol{x}_{n-1}|\boldsymbol{z}_{n-1})$ is available. Then,

- Prediction Step:

$$p(\boldsymbol{x}_n|\boldsymbol{z}_{1:n-1}) = \int p(\boldsymbol{x}_n|\boldsymbol{x}_{n-1})p(\boldsymbol{x}_{n-1}|\boldsymbol{z}_{1:n-1})d\boldsymbol{x}_{n-1}$$

- Update Step:

$$p(\boldsymbol{x}_n|\boldsymbol{z}_{1:n}) = \frac{p(\boldsymbol{z}_n|\boldsymbol{x}_n)p(\boldsymbol{x}_n|\boldsymbol{z}_{1:n-1})}{p(\boldsymbol{z}_n|\boldsymbol{z}_{1:n-1})}$$

# Sequential Bayesian estimation, cont'd

$$1) \quad p(\boldsymbol{x}_n|\boldsymbol{z}_{1:n-1}) = \int p(\boldsymbol{x}_n|\boldsymbol{x}_{n-1})p(\boldsymbol{x}_{n-1}|\boldsymbol{z}_{1:n-1})d\boldsymbol{x}_{n-1}$$

Prediction eq.

$$p(\boldsymbol{x}_0|\boldsymbol{z}_0) = p(\boldsymbol{x}_0)$$ Prior

$$\boldsymbol{x}_n = \boldsymbol{f}_n(\boldsymbol{x}_{n-1}, \boldsymbol{v}_{n-1})$$ State trans.

$$\boldsymbol{z}_n = \boldsymbol{h}_n(\boldsymbol{x}_n, \boldsymbol{w}_n)$$

Measurement eq.

$$2) \quad p(\boldsymbol{x}_n|\boldsymbol{z}_{1:n}) = \frac{p(\boldsymbol{x}_n|\boldsymbol{z}_{1:n-1})p(\boldsymbol{z}_n|\boldsymbol{x}_n)}{p(\boldsymbol{z}_n|\boldsymbol{z}_{1:n-1})}$$

Update eq.

# Linear-Gaussian case: Kalman filter.

It is generally assumed that:

- $\boldsymbol{w}_n$ and $\boldsymbol{v}_n$, are multivariate Gaussian random variables with known means and covariance matrices.
- $\boldsymbol{h}_n(\cdot)$ is a linear function of $\boldsymbol{x}_n$ and $\boldsymbol{w}_n$.
- $\boldsymbol{f}_n(\cdot)$ is a linear function of $\boldsymbol{x}_{n-1}$ and $\boldsymbol{v}_{n-1}$

That can be rewritten as

$$\boldsymbol{x}_n = \mathbf{F}_n \boldsymbol{x}_{n-1} + \boldsymbol{v}_{n-1}$$

$$\boldsymbol{z}_n = \mathbf{H}_n \boldsymbol{x}_n + \boldsymbol{w}_n$$

In this case, Prediction and Update equations can be evaluated analytically! $\rightarrow$ Kalman filter.

# Nonlinear case.

If $\boldsymbol{h}_n(\cdot)$ or $\boldsymbol{f}_n(\cdot)$ are nonlinear then no closed-form solution exists for Prediction and Update equations.

Solution: Local linearization of $\boldsymbol{h}_n(\cdot)$ and $\boldsymbol{f}_n(\cdot)$:

$$\hat{\mathbf{F}}_n = \left.\frac{d\boldsymbol{f}_n(\boldsymbol{x})}{d\boldsymbol{x}}\right|_{\boldsymbol{x}=\boldsymbol{m}_{n-1|n-1}}$$

$$\hat{\mathbf{H}}_n = \left.\frac{d\boldsymbol{h}_n(\boldsymbol{x})}{d\boldsymbol{x}}\right|_{\boldsymbol{x}=\boldsymbol{m}_{n|n-1}}$$

where $\boldsymbol{m}_{\cdot|\cdot}$ is last state estimate. Then, application of Kalman filter is straightforward.

- Unscented Kalman filter.

# Particle filters.

What if the approximations are not satisfactory?

- Nonlinearity is to strong.
- The densities are multimodal, i.e. Gaussian approximation is not appropriate.
- No closed-form derivative of the nonlinear transformations can be computed.

To approximate unknown PDF, Monte Carlo methods can be employed.

> The key idea: approximate the required PDF with a set of random samples with associated weights and compute estimates based on those samples.

# Monte Carlo methods.

Monte Carlo are computational techniques that approximates a desired density by means of drawing random samples from the corresponding distribution.

Two basic MC problems:

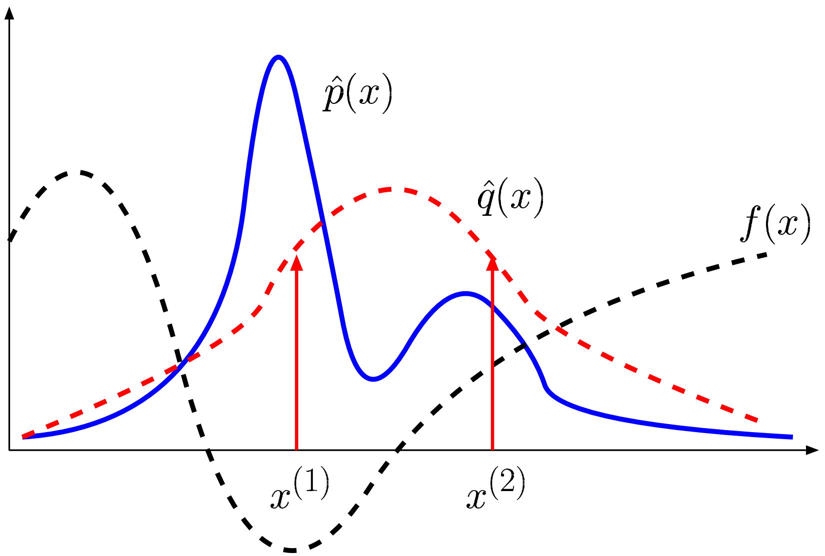
- **Problem 1**: Generate samples $\{\boldsymbol{x}^{(r)}\}_{r=1}^{R}$ from a given distribution $p(\boldsymbol{x})$
- **Problem 2**: Estimate expectation of functions under this distribution, i.e.

$$\boldsymbol{f} = \langle f(\boldsymbol{x}) \rangle = \int f(\boldsymbol{x}) p(\boldsymbol{x}) d\boldsymbol{x}, \quad \rightarrow \quad \boldsymbol{f} \approx \frac{1}{R} \sum_{r=1}^{R} f(\boldsymbol{x}^{(r)})$$

# Importance sampling.

Importance sampling solves **Problem 2**.

1. $p(x) = \hat{p}(x)/Z$ is known up to a normalizing constant $Z$ and $\hat{p}(x)$ can be evaluated at all $x$.
2. $p(x)$ is two complicated to directly sample from.
3. There is $q(x) = \hat{q}(x)/Z_q$ (*Importance density*) from which we can sample and $\hat{q}(x)$ can be evaluated at all $x$.




- $w_r = \hat{p}(x^r)/\hat{q}(x^r)$
- Compute the required expectation

$$\langle f(\boldsymbol{x}) \rangle = \frac{\sum_r w_r f(x^r)}{\sum_r w_r}$$

# From IS to sequential estimation.

- Our goal is to get $p(\boldsymbol{x}_{0:n}|\boldsymbol{z}_{1:n})$ and from it $\langle \boldsymbol{x}_n \rangle$.
- Drawing samples from $p(\boldsymbol{x}_{0:n}|\boldsymbol{z}_{1:n})$ directly is difficult.
- We select easier importance density $q(\boldsymbol{x}_{0:n}|\boldsymbol{z}_{1:n})$ to draw samples $\boldsymbol{x}_{0:n}^r$, $r = 1 \ldots R$, from it.

$$w_n^r = \frac{p(\boldsymbol{x}_{0:n}^r|\boldsymbol{z}_{1:n})}{q(\boldsymbol{x}_{0:n}^r|\boldsymbol{z}_{1:n})}$$

- In sequential estimation we want to obtain $p(\boldsymbol{x}_{0:n}|\boldsymbol{z}_{1:n})$ based on $p(\boldsymbol{x}_{0:n-1}|\boldsymbol{z}_{1:n-1})$
- How to incorporate previous weights $w_{n-1}^r$ in computation of the new weights $w_n^r$?

# From IS to sequential estimation, cont'd.

- First, we factorize importance density $q(\boldsymbol{x}_{0:n}|\boldsymbol{z}_{1:n})$ as

$$q(\boldsymbol{x}_{0:n}|\boldsymbol{z}_{1:n}) = q(\boldsymbol{x}_n|\boldsymbol{x}_{0:n-1}, \boldsymbol{z}_{1:n})q(\boldsymbol{x}_{0:n-1}|\boldsymbol{z}_{1:n-1})$$

- Then, we consider density of interest $p(\boldsymbol{x}_{0:n}|\boldsymbol{z}_{1:n})$:

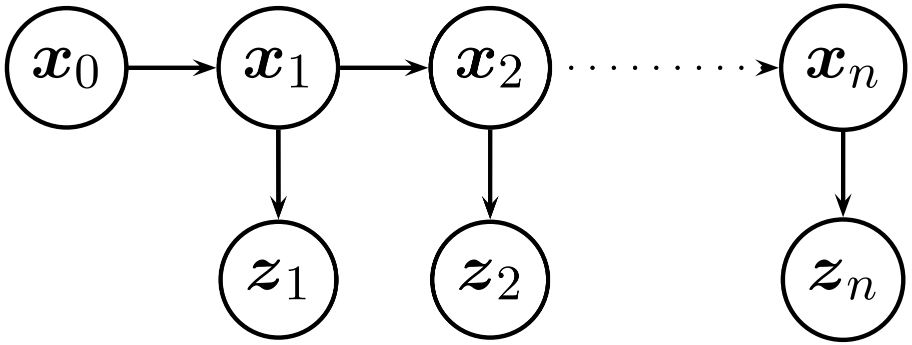
$$p(\boldsymbol{x}_{0:n}|\boldsymbol{z}_{1:n}) = p(\boldsymbol{x}_{0:n}|\boldsymbol{z}_n, \boldsymbol{z}_{1:n-1}) = \frac{p(\boldsymbol{z}_n|\boldsymbol{x}_{0:n}, \boldsymbol{z}_{1:n-1})p(\boldsymbol{x}_{0:n}|\boldsymbol{z}_{1:n-1})}{\int(\cdots)d\boldsymbol{x}_{0:n}} = \cdots$$

# From IS to sequential estimation, cont'd.

$$\frac{p(\boldsymbol{z}_n|\boldsymbol{x}_{0:n}, \boldsymbol{z}_{1:n-1})p(\boldsymbol{x}_{0:n}|\boldsymbol{z}_{1:n-1})}{\int(\cdots)d\boldsymbol{x}_{0:n}} =$$

$$\frac{p(\boldsymbol{z}_n|\boldsymbol{x}_{0:n}, \boldsymbol{z}_{1:n-1})p(\boldsymbol{x}_n, \boldsymbol{x}_{0:n-1}|\boldsymbol{z}_{1:n-1})}{\int(\cdots)d\boldsymbol{x}_{0:n}} =$$

$$\frac{p(\boldsymbol{z}_n|\boldsymbol{x}_{0:n}, \boldsymbol{z}_{1:n-1})p(\boldsymbol{x}_n|\boldsymbol{x}_{0:n-1}, \boldsymbol{z}_{1:n-1})p(\boldsymbol{x}_{0:n-1}|\boldsymbol{z}_{1:n-1})}{\int(\cdots)d\boldsymbol{x}_{0:n}}$$




$$p(\boldsymbol{z}_n|\boldsymbol{x}_{0:n}, \boldsymbol{z}_{1:n-1}) = p(\boldsymbol{z}_n|\boldsymbol{x}_n)$$
$$p(\boldsymbol{x}_n|\boldsymbol{x}_{0:n-1}, \boldsymbol{z}_{1:n-1}) = p(\boldsymbol{x}_n|\boldsymbol{x}_{n-1})$$

$$p(\boldsymbol{x}_{0:n}|\boldsymbol{z}_{1:n}) \propto p(\boldsymbol{z}_n|\boldsymbol{x}_n)p(\boldsymbol{x}_n|\boldsymbol{x}_{n-1})p(\boldsymbol{x}_{0:n-1}|\boldsymbol{z}_{1:n-1})$$

# From IS to sequential estimation, cont'd.

Now we put this expression into the formula for importance coefficients

$$w_n^r = \frac{p(\boldsymbol{x}_{0:n}^r|\boldsymbol{z}_{1:n})}{q(\boldsymbol{x}_{0:n}^r|\boldsymbol{z}_{1:n})} = \frac{p(\boldsymbol{z}_n|\boldsymbol{x}_n^r)p(\boldsymbol{x}_n^r|\boldsymbol{x}_{n-1}^r)p(\boldsymbol{x}_{0:n-1}^r|\boldsymbol{z}_{1:n-1})}{q(\boldsymbol{x}_n^r|\boldsymbol{x}_{0:n-1}^r, \boldsymbol{z}_{1:n})q(\boldsymbol{x}_{0:n-1}^r|\boldsymbol{z}_{1:n-1})} =$$

$$\frac{p(\boldsymbol{z}_n|\boldsymbol{x}_n^r)p(\boldsymbol{x}_n^r|\boldsymbol{x}_{n-1}^r)}{q(\boldsymbol{x}_n^r|\boldsymbol{x}_{0:n-1}^r, \boldsymbol{z}_{1:n})} \cdot w_{n-1}^r$$

If we are only interested in $p(\boldsymbol{x}_n|\boldsymbol{z}_{1:n})$, then it makes sense to select importance density as $q(\boldsymbol{x}_n^r|\boldsymbol{x}_{0:n-1}^r, \boldsymbol{z}_{1:n}) = q(\boldsymbol{x}_n^r|\boldsymbol{x}_{n-1}^r, \boldsymbol{z}_n)$.

Then,

$$w_n^r = w_{n-1}^r \cdot \frac{p(\boldsymbol{z}_n|\boldsymbol{x}_n^r)p(\boldsymbol{x}_n^r|\boldsymbol{x}_{n-1}^r)}{q(\boldsymbol{x}_n^r|\boldsymbol{x}_{n-1}^r, \boldsymbol{z}_n)}$$

# Sequential Importance Sampling (SIS).

- **function** $[\{\boldsymbol{x}_n^r, w_n^r\}_{r=1}^R] = \mathrm{SIS}\Big(\{\boldsymbol{x}_{n-1}^r, w_{n-1}^r\}_{r=1}^R, \boldsymbol{z}_n\Big)$
  - **for** $r = 1 : R$
    - Draw $\boldsymbol{x}_n^r$ from $q(\boldsymbol{x}_n^r|\boldsymbol{x}_{n-1}^r, \boldsymbol{z}_n)$
    - Compute $w_n^r$

$$w_n^r = w_{n-1}^r \cdot \frac{p(\boldsymbol{z}_n|\boldsymbol{x}_n^r)p(\boldsymbol{x}_n^r|\boldsymbol{x}_{n-1}^r)}{q(\boldsymbol{x}_n^r|\boldsymbol{x}_{n-1}^r, \boldsymbol{z}_n)}, \qquad w_n^r = w_n^r \Big/ \sum_r (w_n^r)$$

  - **end**
  - The sought posterior $p(\boldsymbol{x}_n|\boldsymbol{z}_{1:n})$ is approximated

$$p(\boldsymbol{x}_n|\boldsymbol{z}_{1:n}) \approx \sum_{r=1}^R w_n^r \delta(\boldsymbol{x}_n - \boldsymbol{x}_n^r)$$

# Sequential Importance Sampling (SIS).

Simple example:

- System equations:

$$x_n = f_n(x_{n-1}) + v_{n-1}$$
$$z_n = h_n(x_n) + e_n.$$

- $v_n \sim \mathcal{N}(0, \sigma_v^2)$, $e_n \sim \text{Laplace}(\mu_e, \sigma_e)$
-

$$p(z_n|x_n^r) = \frac{1}{\sigma_e} \exp\left\{ - \frac{|(z_n - h_n(x_n^r)) - \mu_e|}{\sigma_e} \right\}$$
$$p(x_n^r|x_{n-1}^r) = \frac{1}{\sqrt{2\pi}\sigma_v} \exp\left\{ - \frac{|x_n^r - f_n(x_{n-1}^r)|^2}{2\sigma_v^2} \right\}$$

# Sequential Importance Sampling (SIS).

A common problem with SIS algorithm is degeneracy phenomenon: all but several weights negligibly small.

To mitigate degeneracy problem one can

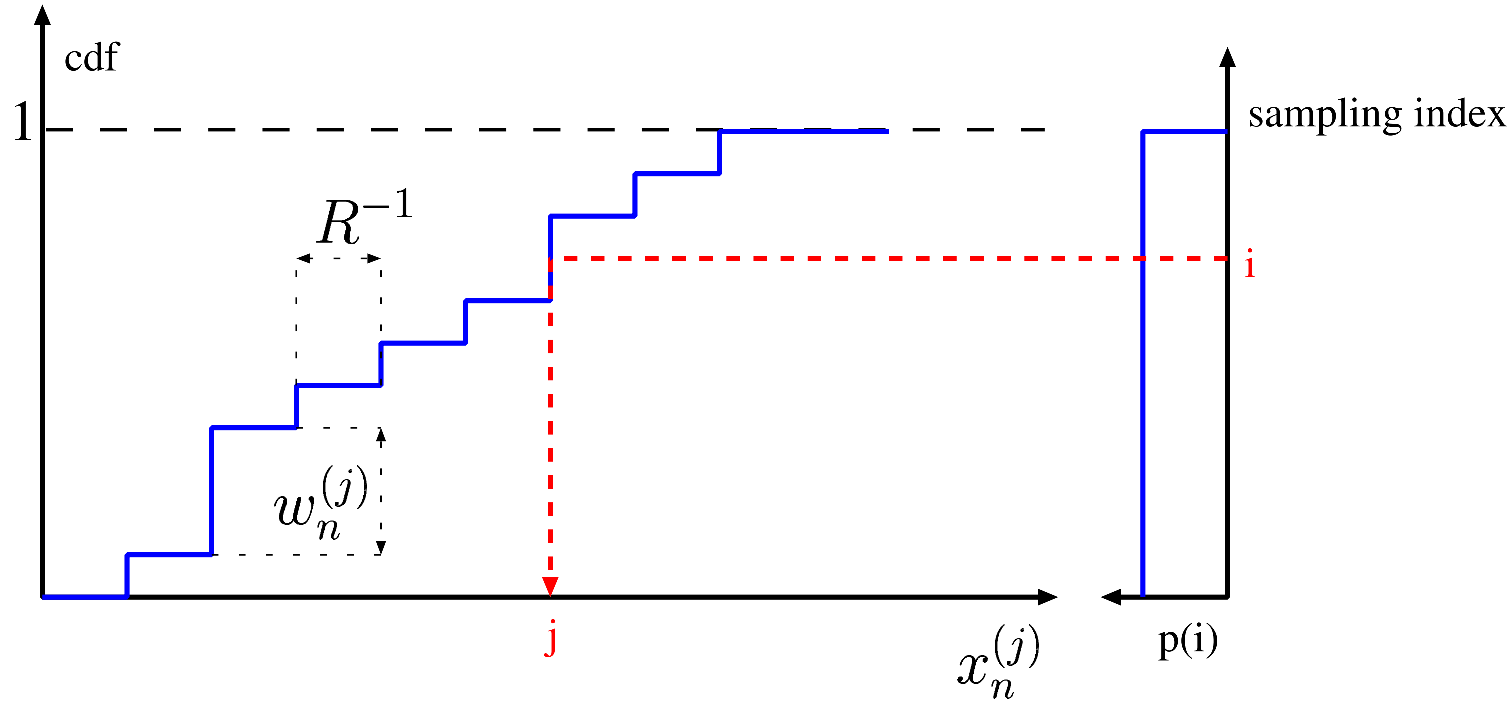
- Increase the number of particles $R$ to a very large number.
- Properly choose the importance density $q(\cdot)$. This is a crucial point in PF design!
  - $q(\boldsymbol{x}_n^r|\boldsymbol{x}_{n-1}^r, \boldsymbol{z}_n) = p(\boldsymbol{x}_n^r|\boldsymbol{x}_{n-1}^r)$

$$w_n^r = w_{n-1}^r \cdot p(\boldsymbol{z}_n|\boldsymbol{x}_n^r)$$

  - Markov Chain Monte Carlo.
- Employ resampling schemes.

# Resampling schemes.

Resampling is aimed at amending degeneracy problem.



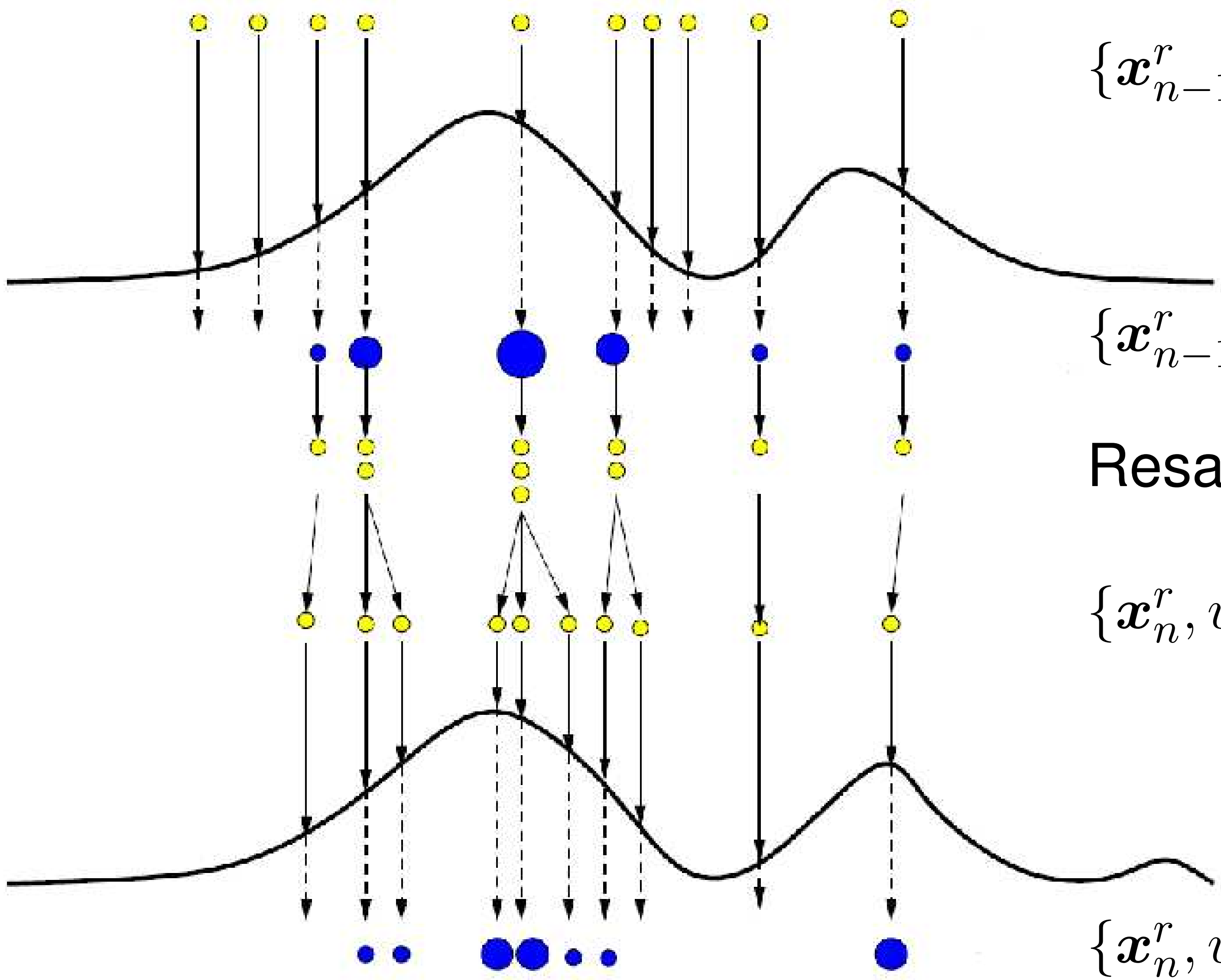

Example: $R = 10$, $w_n^r = 0.3$, $\{\boldsymbol{x}_n^r\}$.

After resampling this will result (on average) in

$R_j = 0.3 \times 10 = 3$ copies of $\boldsymbol{x}_n^r$, each having weight $1/R$

$E\{R_j\} = R \cdot w_n^j$

# Resampling schemes.



# Sequential Importance Resampling (SIR).

- **function** $[\{\boldsymbol{x}_n^r, w_n^r\}_{r=1}^R] = \mathrm{SIS}\Big(\{\boldsymbol{x}_{n-1}^r, w_{n-1}^r\}_{r=1}^R, \boldsymbol{z}_n\Big)$
  - **for** $r = 1 : R$
    - Draw $\boldsymbol{x}_n^r$ from $q(\boldsymbol{x}_n^r|\boldsymbol{x}_{n-1}^r, \boldsymbol{z}_n)$
    - Compute $w_n^r$

$$w_n^r = w_{n-1}^r \cdot \frac{p(\boldsymbol{z}_n|\boldsymbol{x}_n^r)p(\boldsymbol{x}_n^r|\boldsymbol{x}_{n-1}^r)}{q(\boldsymbol{x}_n^r|\boldsymbol{x}_{n-1}^r, \boldsymbol{z}_n)}, \qquad w_n^r = w_n^r \Big/ \sum_r (w_n^r)$$

    - Resample particles $\{\boldsymbol{x}_n^r\}, w_n^r$.
  - **end**
  - The sought posterior $p(\boldsymbol{x}_n|\boldsymbol{z}_{1:n})$ is approximated $p(\boldsymbol{x}_n|\boldsymbol{z}_{1:n}) \approx \sum_{r=1}^R \delta(\boldsymbol{x}_n - \boldsymbol{x}_n^r)/R$

# To conclude...

- Monte Carlo assumption must hold: Posterior is well represented by Dirac point-mass approximation.
- Importance Sampling assumption: it is possible to obtain samples from the posterior by sampling from a suitable importance distribution.
- Markov Chain Monte Carlo is often employed ('Smoothing step').
- Suitable importance density is crucial.
- Resolution is determined by $R$.
- Check out other PF algorithms.

# To conclude...

Literature:

- Sanjeev Arulampalam, **Simon Maskell**, **Neil Cordon**.
  ’A Tutorial on Particle Filters for On-line Non-linear/Non-Gaussian Bayesian Tracking’.
- **Rudolph van der Merwe**, **Arnaud Doucet**, Nando de Freitas, **Eric Wan**.
  ’The Unscented Particle Filters’.