UNIVERSITY OF MINNESOTA
Driven to Discover℠

Markov chain Monte Carlo sampling

SPiNCOM reading group
Jun. 10th , 2016

Dimitris Berberidis

Problem statement - Motivation

Goal: Draw samples from a given pdf $p(x)$

Impact of sampling :

- Bayesian inference ($x \in \mathcal{X}$:unknowns, $y \in \mathcal{Y}$: data)
  - Normalization
  $$p(x|y) = \frac{p(y|x)p(x)}{\int_{\mathcal{X}} p(y|x')p(x')dx'}$$
  - Marginalization
  $$p(x|y) = \int_{\mathcal{Z}} p(x, z|y)dz$$
  - Expectation
  $$\mathbb{E}_{p(x|y)}[f(x)] = \int_{\mathcal{X}} f(x)p(x|y)dx$$
- Optimization: non-convex multimodal objectives

Our focus

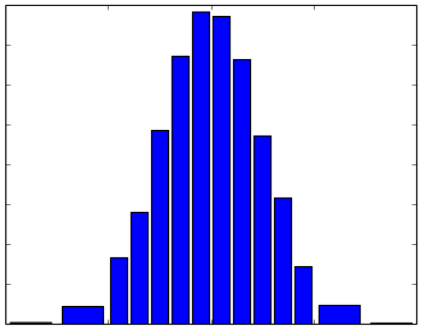
- Statistical mechanics
- Penalized likelihood model selection
- Simulation of physical systems

Roadmap

- **Motivation**
- **Basic Monte Carlo**
- **Rejection Sampling**
- **Marcov chain Monte Carlo**
  - **Metropolis-Hastings**
  - **Gibbs sampling**
- **Importance sampling**
  - **Relation to Rejection Sampling**
  - **Sequential Importance Sampling (Particle Filtering)**
- **Conclusions**

C. Andrieu, N. de Freitas, A. Doucet and M. Jordan, “An Introduction to MCMC for Machine Learning,” *Machine Learning*, pp. 5-43, Jan 2003.

The Monte Carlo principle



- Draw samples $\{x^{(i)}\}_{i=1}^{N}$ i.i.d from $p(x)$
- Approximate $p(x)$ with $p_N(x) = \frac{1}{N}\sum_{i=1}^{N}\delta_{x^{(i)}}(x)$
- Approx. integrals $I(f)$ with tractable sums $I_N(f)$

$$I_N(f) = \frac{1}{N}\sum_{i=1}^{N} f(x^{(i)}) \xrightarrow[N\to\infty]{a.s.} I(f) = \int_{\mathcal{X}} f(x)p(x)dx$$

- $I_N(f)$ unbiased for finite N with $\mathrm{var}(I_N(f)) = \frac{\sigma_f^2}{N}$

$$\sqrt{(N)}(I_N(f) - I(f)) \underset{N\to\infty}{\Longrightarrow} \mathcal{N}(0, \sigma_f^2) \quad \text{with} \quad \sigma_f^2 = \mathbb{E}_{p(x)}[f^2(x)] - I^2(f)$$

- Approx. the maximum of $p(x)$ as $\hat{x} = \arg\max_{i\in\{1,\ldots,N\}} p(x^{(i)})$

Challenge: **What if $p(x)$ does not have a standard form (e.g. Gaussian) ?**

Rejection Sampling

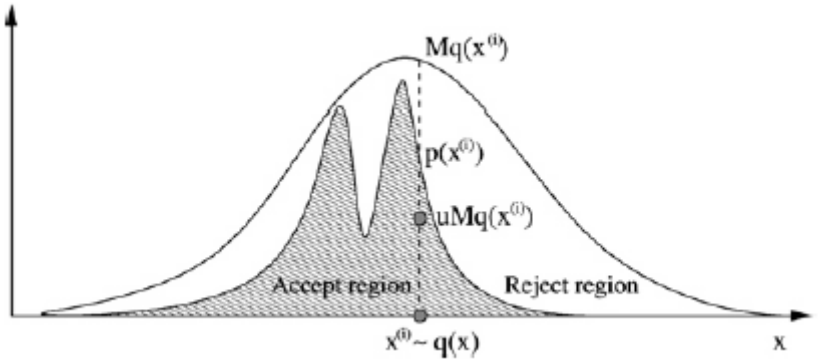
- Instead of $p(x)$, draw i.i.d samples from an "easy" $q(x)$
- Proposal pdf $q(x)$ should satisfy: $p(x) < Mq(x),\ M < \infty$

Rejection Sampling algorithm

Set $i = 1$

Repeat until $i = N$

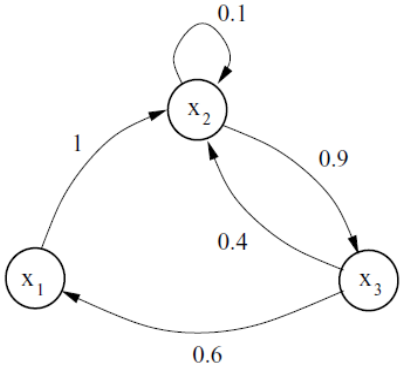
1. Sample $x^{(i)} \sim q(x)$ and $u \sim \mathcal{U}_{(0,1)}$.
2. If $u < \frac{p(x^{(i)})}{Mq(x^{(i)})}$ then accept $x^{(i)}$ and increment the counter i by 1. Otherwise, reject.



- Accepted $x^{(i)}$ sampled according to $p(x)$
- Severe limitation in practice: M can be too large

$$\Pr(x \text{ accepted}) = \Pr\left(u < \frac{p(x)}{Mq(x)}\right) \approx \frac{1}{M} \ll 1$$

Basics of Markov chains



$$T = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0.1 & 0.9 \\ 0.6 & 0.4 & 0 \end{bmatrix}$$

- Discrete stochastic process $x^{(i)}$ is a Marcov chain (MC) if

$$p(x^{(i)}|x^{(i-1)}, \ldots, x^{(1)}) = T(x^{(i)}|x^{(i-1)})$$

- MC is homogeneous if T is time invariant
- After t steps, probability of state $x^{(i)}$ is: $p_t(x^{(i)}) = \sum_{i'} T(x^{(i)}|x^{(i')}) p_{t-1}(x^{(i')})$
- MC reaches stationary distribution $\pi(x)$ if : $p_t(x^{(i)}) = p_{t-1}(x^{(i)}) = \pi(x^{(i)}), \ \forall i$
- MC converges to a stationary distribution if
  - Irreducible: All states are visited (transition graph connected)
  - Aperiodic: Does not get trapped into cycles

Markov chain Monte Carlo

- **Goal**: Construct MC with target $p(x)$ as stationary distribution

- Sufficient condition: The detailed balance condition (DBC)

$$p(x^{(i)})T(x^{(i')}|x^{(i)}) = p(x^{(i')})T(x^{(i)}|x^{(i')}), \quad \forall i, i'$$

- Continuous states
  - Transition kernel: $p_t(x^{(i)}) = \int_{\mathcal{X}} K(x^{(i)}|x^{(i')})p_{t-1}(x^{(i')})dx^{(i')}$
  - DBC remains the same

- Run MC to convergence and obtain non i.i.d samples
  - Design $K(\cdot|\cdot)$ to achieve fast convergence (e.g. small mixing time)

$$t_{mix} = \min t \quad s.t. \quad |\Pr(x_t \in \mathcal{C}) - \Pr(\mathcal{C})| \leq 1/4 \quad \forall \mathcal{C} \subseteq \mathcal{X}$$

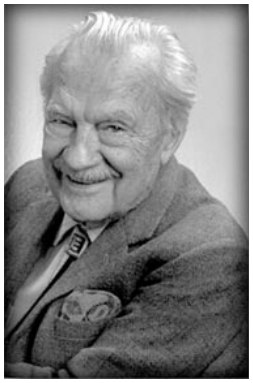
The Metropolis-Hastings sampler

1. Initialise $x^{(0)}$.
2. For $i = 0$ to $N - 1$
   - Sample $u \sim \mathcal{U}_{[0,1]}$.
   - Sample $x^\star \sim q(x^\star|x^{(i)})$.
   - If $u < \mathcal{A}(x^{(i)}, x^\star) = \min\left\{1, \frac{p(x^\star)q(x^{(i)}|x^\star)}{p(x^{(i)})q(x^\star|x^{(i)})}\right\}$

     $x^{(i+1)} = x^\star$

     else

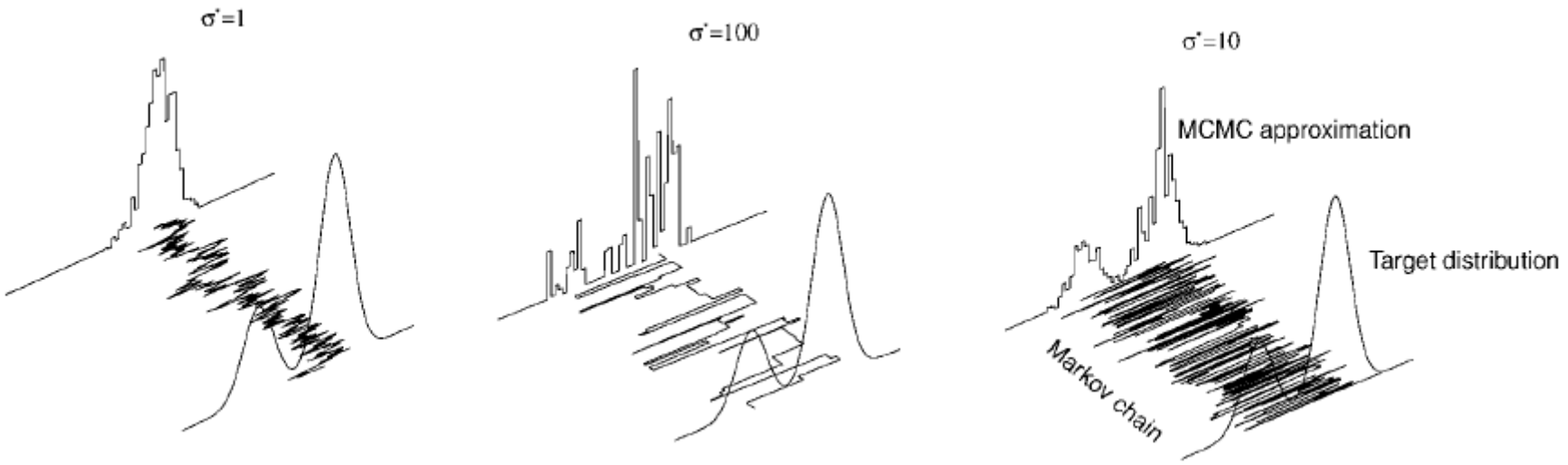
     $x^{(i+1)} = x^{(i)}$



Rejection probability

- MH transition kernel: $K_{MH}(x^{(i+1)}|x^{(i)}) = q(x^{(i+1)}|x^{(i)})\mathcal{A}(x^{(i)}, x^{(i+1)}) + \delta_{x^{(i)}} r(x^{(i)})$
- $K_{MH}(\cdot|\cdot)$ satisfies DBC $\longrightarrow$ Admits $p(x)$ as stationary dist.
- Scale of $p(x)$ not needed! (recall $p(x|y) \propto p(y|x)p(x)$)
- MH always aperiodic; irreducible if support of $q(\cdot)$ includes support of $p(\cdot)$
- Special cases of MH
  - Independent sampler: $q(x^*|x^{(i)}) = q(x^*)$
  - Metropolis sampler: $q(x^*|x^{(i)}) = q(x^{(i)}|x^*)$

Example of MH sampling



- Three different Gaussians as proposal distributions
- Choice of proposal distribution is critical!

MCMC with mixture of transition kernels

- Key property
  - Let $K_1(\cdot|\cdot)$ and $K_2(\cdot|\cdot)$ trans. kernels converge $p(x)$
  - $K_{mix}(\cdot|\cdot) = \lambda K_1(\cdot|\cdot) + (1-\lambda)K_2(\cdot|\cdot),\ 0 \leq \lambda \leq 1$ also converges to $p(x)$

1. Initialise $x^{(0)}$.
2. For $i = 0$ to $N-1$
   - Sample $u \sim \mathcal{U}_{[0,1]}$.
   - If $u < \nu$

     Apply the MH algorithm with a global proposal.
   - else

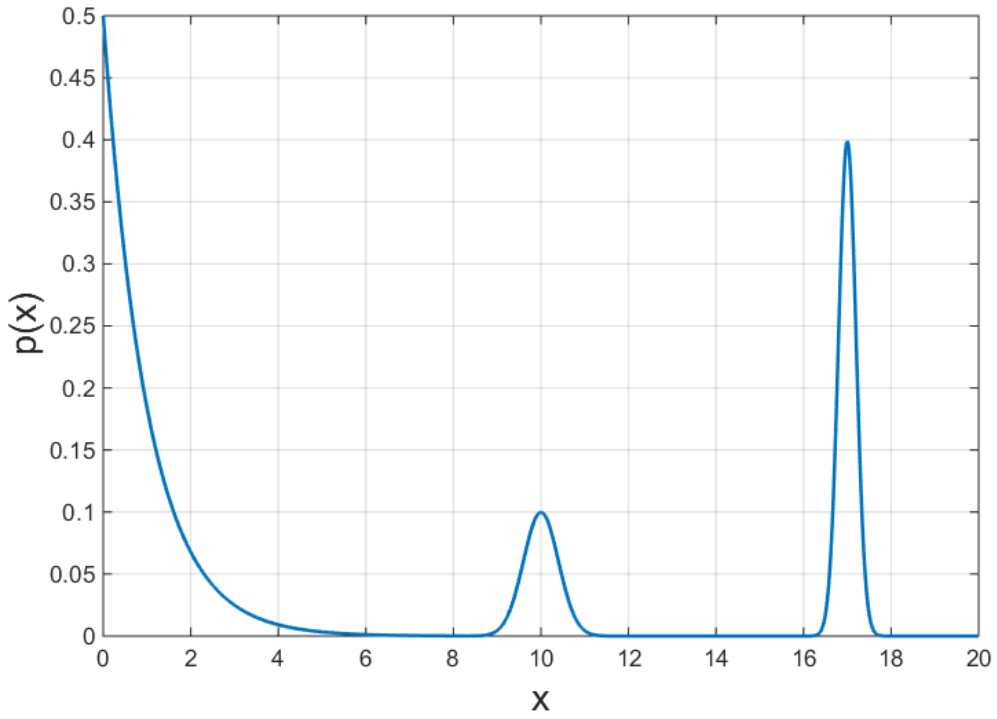
     Apply the MH algorithm with a random walk proposal.

- Intuition
  - Local random walk reduces the number of rejections
  - Global proposal helps discover other modes

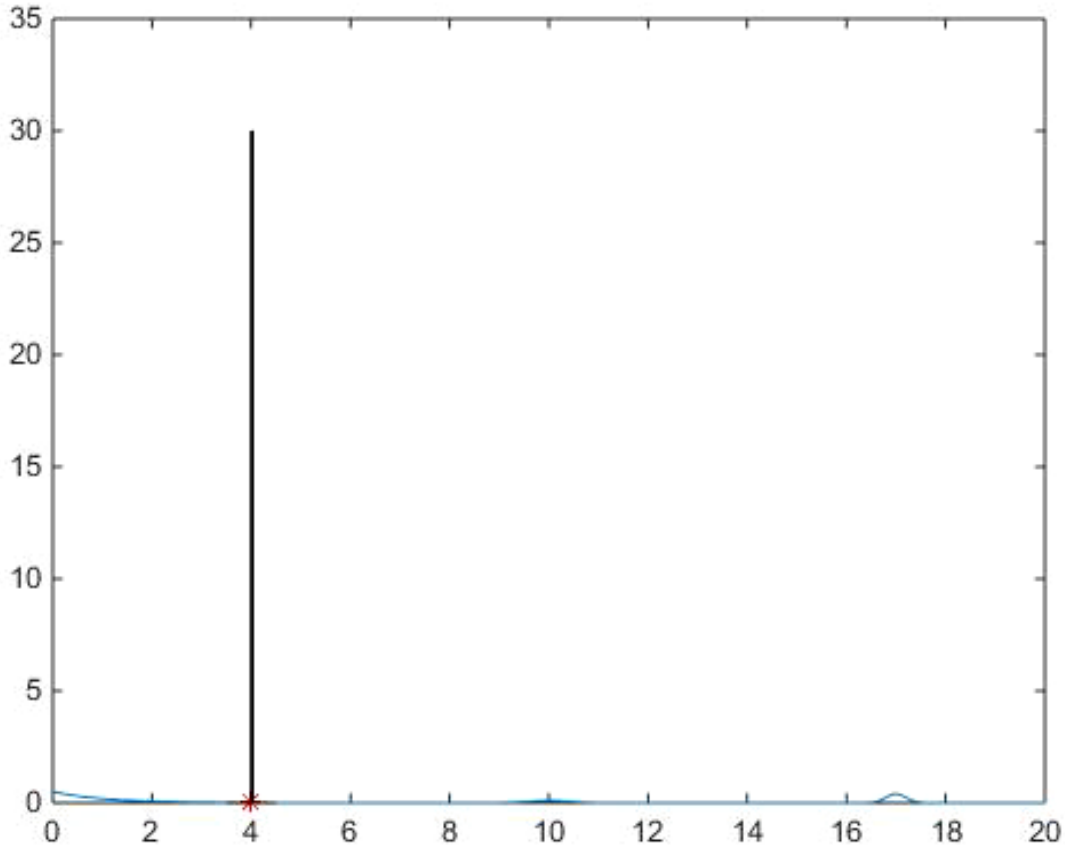
Example of MH with mixture of Kernels

Target: $$p(x) = 0.6 \times \exp(x; 1) + 0.15 \times \mathcal{N}(x; 10, 0.4) + 0.25 \times \mathcal{N}(x; 17, 0.2)$$

Proposal: $$q(x^*|x^{(i)}) = \begin{cases} \mathcal{N}(x^*; x^{(i)}, 0.25), & \text{w.p. } 0.97 \\ \text{Unif}[0, 20], & \text{w.p. } 0.03 \end{cases}$$



Experiment with mixture of Kernels



Simulated Annealing

- Simple modification of the MH algorithm for global optimization

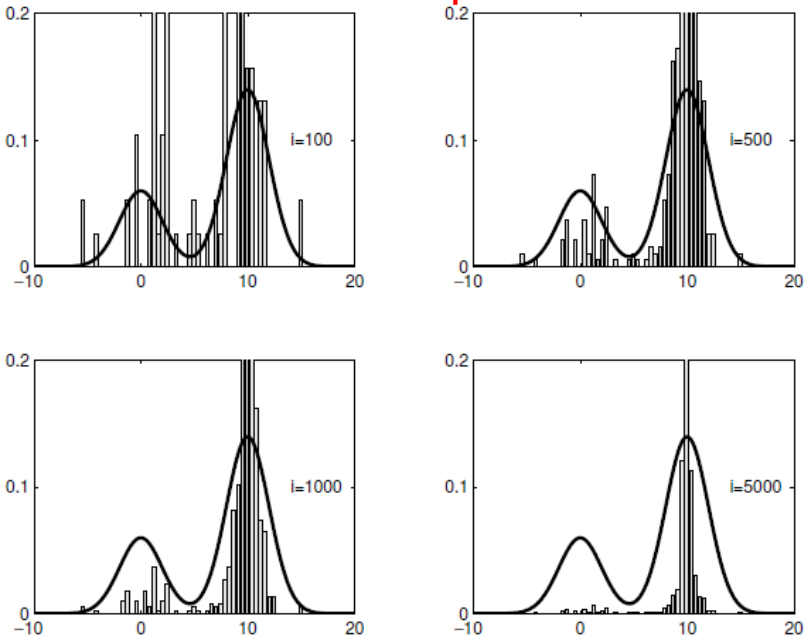
1. Initialise $x^{(0)}$ and set $T_0 = 1$.
2. For $i = 0$ to $N - 1$
   - Sample $u \sim \mathcal{U}_{[0,1]}$.
   - Sample $x^\star \sim q(x^\star | x^{(i)})$.
   - If $u < \mathcal{A}(x^{(i)}, x^\star) = \min\left\{1, \frac{p^{\frac{1}{T_i}}(x^\star) q(x^{(i)} | x^\star)}{p^{\frac{1}{T_i}}(x^{(i)}) q(x^\star | x^{(i)})}\right\}$

     $$x^{(i+1)} = x^\star$$

     else

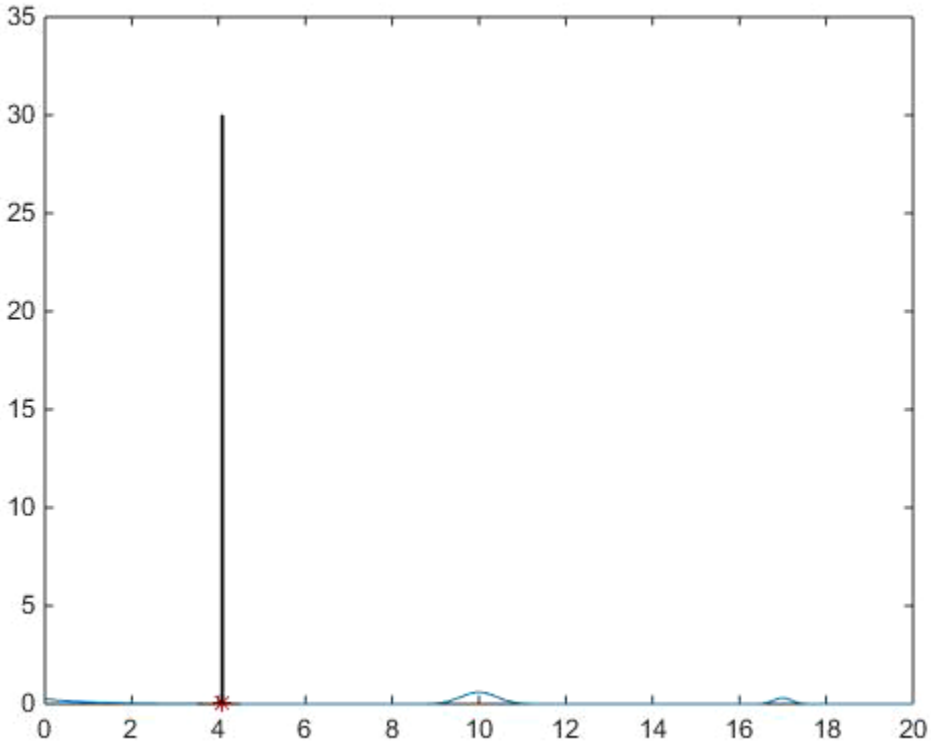
     $$x^{(i+1)} = x^{(i)}$$
   - Set T_{i+1} according to a chosen cooling schedule.

Example



- Simulates a non-homogeneous MC with $p_i(x) \propto p^{1/T_i}(x)$
- Intuition: $p^{1/T_i}(x)$ concentrates around global max. of $p(x)$ as $T_i \to 0$

Experiment with Simulated Annealing



Cycles of MH kernels

1. Initialise $x^{(0)}$.
2. For $i = 0$ to $N - 1$
   - Sample the block $x_{b_1}^{(i+1)}$ according to an MH step with proposal distribution $q_1(x_{b_1}^{(i+1)}|x_{-[b_1]}^{(i+1)}, x_{b_1}^{(i)})$ and invariant distribution $p(x_{b_1}^{(i+1)}|x_{-[b_1]}^{(i+1)})$.
   - Sample the block $x_{b_2}^{(i+1)}$ according to an MH step with proposal distribution $q_2(x_{b_2}^{(i+1)}|x_{-[b_2]}^{(i+1)}, x_{b_2}^{(i)})$ and invariant distribution $p(x_{b_2}^{(i+1)}|x_{-[b_2]}^{(i+1)})$.
   - ⋮
   - Sample the block $x_{b_{n_b}}^{(i+1)}$ according to an MH step with proposal distribution $q_{n_b}(x_{b_{n_b}}^{(i+1)}|x_{-[b_{n_b}]}^{(i+1)}, x_{b_{n_b}}^{(i)})$ and invariant distribution $p(x_{b_{n_b}}^{(i+1)}|x_{-[b_{n_b}]}^{(i+1)})$.

- Multivariate state is split into n_b blocks
  - Each block is updated separately
- Transition Kernel

$$K_{MH-Cycle}\left(x^{(i+1)}|x^{(i)}\right) = \prod_{j=1}^{n_b} K_{MH(j)}\left(X^{(i+1)_{b_j}}|x_{b_j}^{(i)}, x_{-b_j}^{(i+1)}\right)$$

- Block correlated variables together for fast convergence
- Trade-off on block size
  - Small block size: Chain takes long time to explore space
  - Large block size: Acceptance probability is small

Gibbs sampling

1. Initialise $x_{0,1:n}$.
2. For $i = 0$ to $N-1$
   - Sample $x_1^{(i+1)} \sim p(x_1|x_2^{(i)}, x_3^{(i)}, \ldots, x_n^{(i)})$.
   - Sample $x_2^{(i+1)} \sim p(x_2|x_1^{(i+1)}, x_3^{(i)}, \ldots, x_n^{(i)})$.
   - $\vdots$
   - Sample $x_j^{(i+1)} \sim p(x_j|x_1^{(i+1)}, \ldots, x_{j-1}^{(i+1)}, x_{j+1}^{(i)}, \ldots, x_n^{(i)})$.
   - $\vdots$
   - Sample $x_n^{(i+1)} \sim p(x_n|x_1^{(i+1)}, x_2^{(i+1)}, \ldots x_{n-1}^{(i+1)})$.

- For $\mathcal{X} \subseteq \mathbb{R}^n$ assume that we know

$$p(x_j|x_{-j}) = p(x_j|x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n)$$

- Gibbs sampling proposal distribution

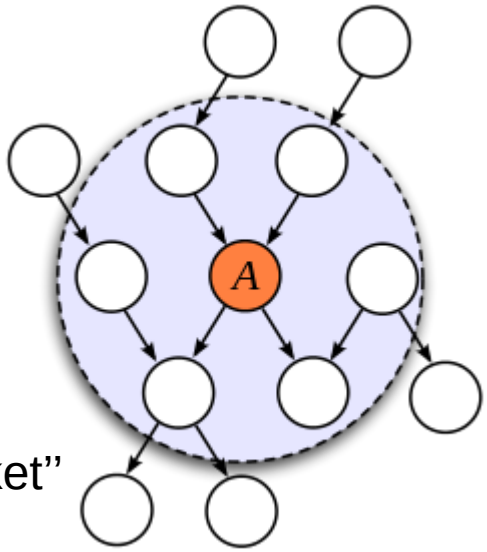
$$q(x^*|x^{(i)}) = \begin{cases} p(x_j^*|x_{-j}^{(i)}), & x_{-j}^* = x_{-j}^{(i)} \\ 0, & \text{Otherwise} \end{cases}$$

- Acceptance probability =1

- Combined with MH if $p(x_j|x_{-j})$ not easy

- To sample Markov networks, condition on ``Markov Blanket''

$$p(x_j|x_{-j}) = p(x_j|x_{\text{pa}(j)}) \prod_{k \in \text{ch}(j)} p(x_k|x_{\text{pa}(k)})$$



Importance sampling - Basics

- Key idea: sample from $q(x)$ and weight with $w(x) := \frac{p(x)}{q(x)}$

$$I(f) = \int_{\mathcal{X}} f(x)p(x)dx = \int_{\mathcal{X}} f(x)w(x)q(x)dx$$

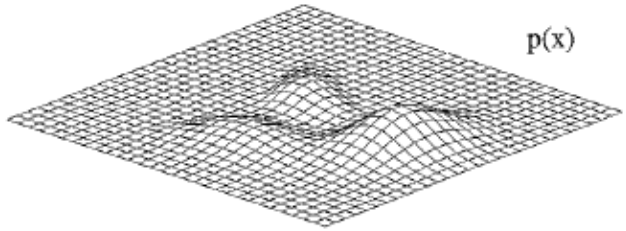
- Draw $\{x^{(i)}\}_{i=1}^{N}$ i.i.d from $q(x)$ to obtain: $\hat{I}_N(f) = \frac{1}{N}\sum_{i=1}^{N} f(x^{(i)})w(x^{(i)})$
- Target $p(x)$ is approximated by $\hat{p}_N(x) = \frac{1}{N}\sum_{i=1}^{N} w(x^{(i)})\delta_{x^{(i)}}(x)$
- Estimate $\hat{I}_N(f)$ is unbiased and: $\hat{I}_N(f) \xrightarrow[N\to\infty]{a.s.} I(f)$
- If scale of $p(x)$ unknown, set $w(x) \propto \frac{p(x)}{q(x)}$ and normalize $\{w(x^{(i)})\}_{i=1}^{N}$

$$I(f) = \frac{\int f(x)w(x)q(x)dx}{\int w(x)q(x)dx} \approx \frac{\frac{1}{N}\sum_{i=1}^{N} f(x^{(i)})w(x^{(i)})}{\frac{1}{N}\sum_{i=1}^{N} w(x^{(i)})} = \sum_{i=1}^{N} f(x^{(i)})\tilde{w}(x^{(i)})$$

Efficiency of importance sampling

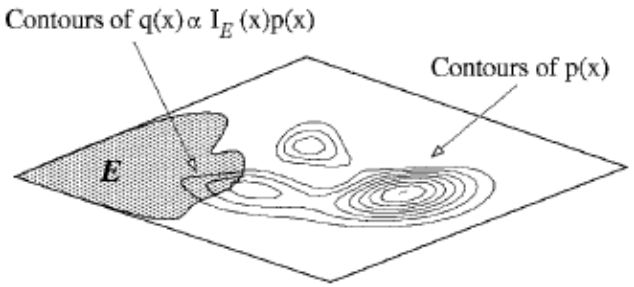
- Proposal pdf $q(x)$ selected to minimize variance

$$\mathrm{var}_{q(x)}(\hat{I}_N(f)) = \frac{\mathbb{E}_{q(x)}[f^2(x)w^2(x)] - I^2(f)}{N}$$



- Variance lower bound (using Jensen's ineq.)

$$\begin{aligned}\mathbb{E}_{q(x)}\left[f^2(x)w^2(x)\right] &\geq \left(\mathbb{E}_{q(x)}\left[|f(x)|w(x)\right]\right)^2 \\ &= \left(\int_{\mathcal{X}} |f(x)|p(x)dx\right)^2\end{aligned}$$



- Optimum importance distribution $q^*(x)$

$$q^*(x) = \frac{|f(x)|p(x)}{\int_{\mathcal{X}} |f(x)|p(x)dx} \neq p(x)$$

- IS can be super efficient!
  - Generally difficult to sample $q^*(x)$

RS as a special case of IS

- Recall the rejection sampling method

> Set $i = 1$
>
> Repeat until $i = N$
>
> 1. Sample $x^{(i)} \sim q(x)$ and $u \sim \mathcal{U}_{(0,1)}$.
> 2. If $u < \frac{p(x^{(i)})}{Mq(x^{(i)})}$ then accept $x^{(i)}$ and increment the counter i by 1. Otherwise, reject.

- Define a new target distribution in $\bar{\mathcal{X}} = \mathcal{X} \times [0, 1]$

$$\bar{p}(x, z) = \begin{cases} Mq(x), & \text{for } x \in \mathcal{X}, z \in \left[0, \frac{p(x)}{Mq(x)}\right] \\ 0, & \text{Otherwise} \end{cases}$$

- IS with target $\bar{p}(x, y)$ and proposal $\bar{q}(x, y) = q(x), \ \forall (x, y) \in \bar{\mathcal{X}}$

$$\bar{w}^{(i)} = \frac{\bar{p}(x^{(i)}, z^{(i)})}{\bar{q}(x^{(i)}, z^{(i)})} = \begin{cases} M, & \text{for } x^{(i)} \in \mathcal{X}, z^{(i)} \in \left[0, \frac{p(x^{(i)})}{Mq(x^{(i)})}\right] \\ 0, & \text{Otherwise} \end{cases}$$

- Equivalent to RS if samples are used to obtain $\hat{I}_N(f)$
  - IS generally (and provably) more efficient for this purpose

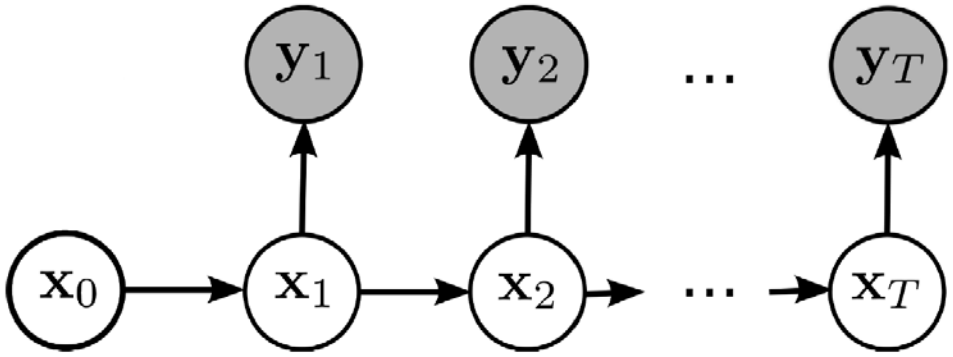
Y. Chen, "Another look at rejection sampling through importance sampling," *Statistic & Probability Letters,* pp. 277-283, May 2005.

Hidden markov model

- The hidden Marcov model

State transition model: $p(x_t|x_{0:t-1}, y_{1:t-1}) = p(x_t|x_{t-1})$

Observation model: $p(y_t|x_{0:t}, y_{1:t-1}) = p(y_t|x_t)$



- **Goal of filtering**: Approximate $p(x_t|y_{1:t})$ and $I(f_t) = \mathbb{E}_{p(x_t|y_{t-1})}[f(x_t)]$

Sequential Importance Sampling (particle filtering)

- Target density: $p(x_{0:t}|y_{1:t})$
- Importance density: $q(x_{0:t}|y_{1:t}) = q(x_t|x_{0:t-1}, y_{1:t})q(x_{0:t-1}|y_{1:t-1})$
- How to sample from $q(x_{0:t}|y_{1:t})$?

Leave the past unchanged

- At time $t-1$ we have: $x_{0:t-1}^{(1)}, x_{0:t-1}^{(2)}, \ldots, x_{0:t-1}^{(N)} \sim q(x_{0:t-1}|y_{1:t-1})$
- Sample for $i \in \{1, \ldots, N\}$: $x_t^{(i)} \sim q(x_t|x_{0:t-1}^{(i)}, y_{1:t})$
- Importance weights : $w_t^{(i)} \propto \frac{p(y_t|x_t^{(i)})p(x_t^{(i)}|x_{0:t-1}^{(i)}, y_{0:t-1})}{q(x_t^{(i)}|x_{0:t-1}^{(i)}, y_{0:t})}$
- Augment $x_{0:t}^{(i)} := \left[x_{0:t-1}^{(i)} \; x_t^{(i)}\right]$ without changing the past (filtering)

Particle degeneracy – How to fix it

Theorem: The unconditional variance of the weights (with $y_{1:k}$ interpreted as r.v.'s) increases with time.

Proof. The weight sequence $W_t^{(i)}$ is a Martingale random process

Martingale definition: $\mathbb{E}[W_t|W_{0:t-1}] = W_{t-1}$

Variance of a martingale is always non-decreasing

$$\mathrm{var}(X) = \mathrm{var}(\mathbb{E}[X|Y]) + \mathbb{E}[\mathrm{var}(X|Y)] \quad \leftarrow \text{Rao-Blackwell}$$

$$\mathrm{var}(W_t) = \underbrace{\mathrm{var}(\mathbb{E}[W_t|W_{0:t-1}])}_{W_{t-1}} + \underbrace{\mathbb{E}[\mathrm{var}(W_t|W_{0:t-1})]}_{\geq 0}$$

- **Theoretical fix:** Sample from optimal
- **Practical fix**: Resample particles after each iteration

A. Kong, J. S. Liu, and W. H. Wong, "Sequential imputations and Bayesian missing data problems," *J. of the American Statistical Association,* pp. 278-288, March 1994.

The particle filter with resampling

Sequential importance sampling step

– For $i = 1, ..., N$, sample from the transition priors

$$\widetilde{x}_t^{(i)} \sim q_t\left(\widetilde{x}_t | x_{0:t-1}^{(i)}, y_{1:t}\right)$$

and set

$$\widetilde{x}_{0:t}^{(i)} \triangleq \left(\widetilde{x}_t^{(i)}, x_{0:t-1}^{(i)}\right)$$

– For $i = 1, ..., N$, evaluate and normalize the importance weights

$$w_t^{(i)} \propto \frac{p\left(y_t | \widetilde{x}_t^{(i)}\right) p\left(\widetilde{x}_t^{(i)} | x_{0:t-1}^{(i)}, y_{1:t-1}\right)}{q_t\left(\widetilde{x}_t^{(i)} | x_{0:t-1}^{(i)}, y_{1:t}\right)}.$$

Selection step

– Multiply/Discard particles $\left\{\widetilde{x}_{0:t}^{(i)}\right\}_{i=1}^{N}$ with high/low importance weights $w_t^{(i)}$ to obtain N particles $\left\{x_{0:t}^{(i)}\right\}_{i=1}^{N}$.

- Many available methods for selection (resampling)
  - Simplest is to ``clone '' $\tilde{x}_t^{(i)}$ w.p. $\propto w_t^{(i)}$
  - Particles that are not cloned are ``killed"

The bootstrap particle filter

- Simple, non-adaptive proposal distribution

$$q(x_t^{(i)}|x_{0:t-1}^{(i)}|y_{1:t}) = q(x_t^{(i)}|x_{0:t-1}^{(i)}|y_{1:t-1}) = q(x_t^{(i)}|x_{t-1}^{(i)})$$

$$w_t^{(i)} \propto p(y_t|x_t^{(i)})$$

- Convenient for non-linear models with additive Gaussian noise
  - Transition prob. and likelihood are both Gaussian (easy to sample)

$$x_t = f(x_{t-1}) + u_t$$
$$y_t = h(x_t) + v_t$$

- Simple to implement; Modular structure; Adheres parallelization
- Resampling is very critical!
  - Ensures that the particles 'follow' the target

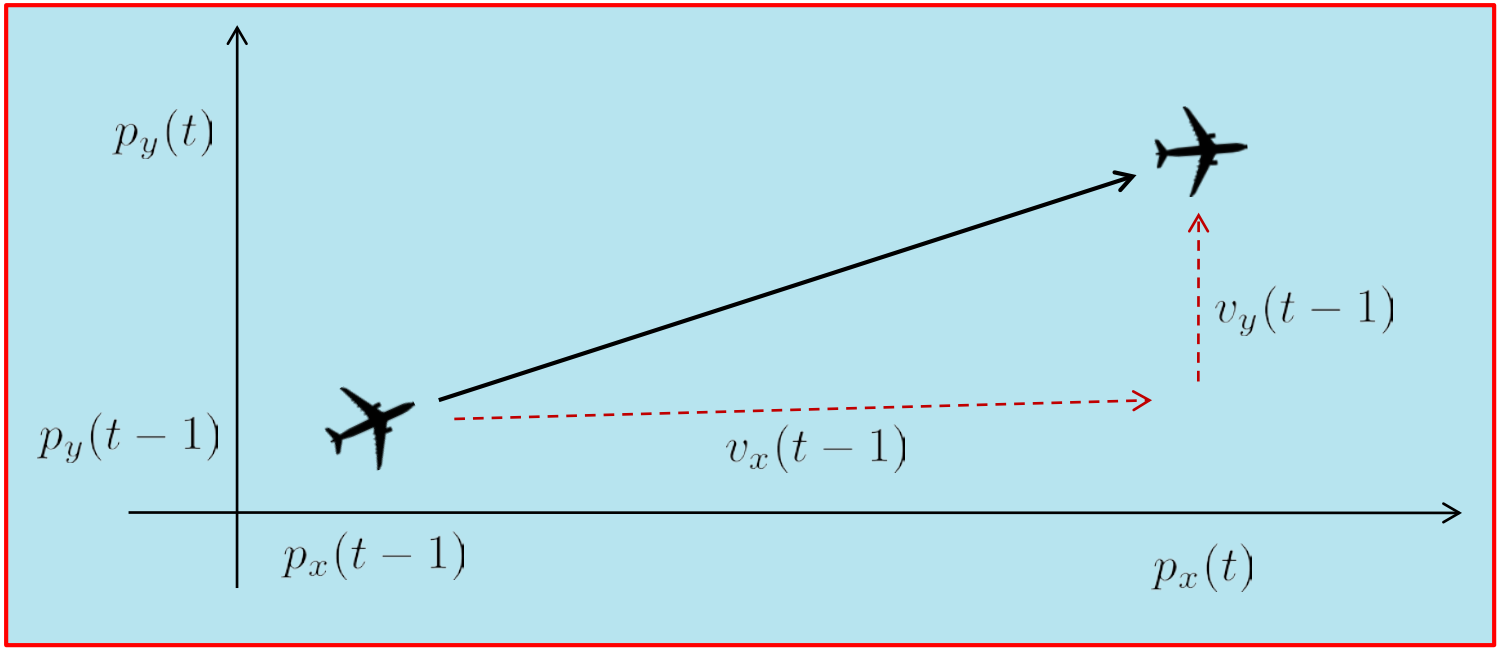
A. Doucet, N. de Freitas and N. Gordon, "Sequential Monte Carlo Methods in Practice," *Springer*, 2001.

Example: target tracking

- State: position $(p_x(t), p_y(t))$ and constant velocity $(v_x(t), v_y(t))$

$$p_x(t) = p_x(t-1) + T_s v_x(t-1), \qquad v_x(t) = v_x(t-1) + u_x(t)$$
$$p_y(t) = p_y(t-1) + T_s v_y(t-1), \qquad v_y(t) = v_y(t-1) + u_y(t)$$

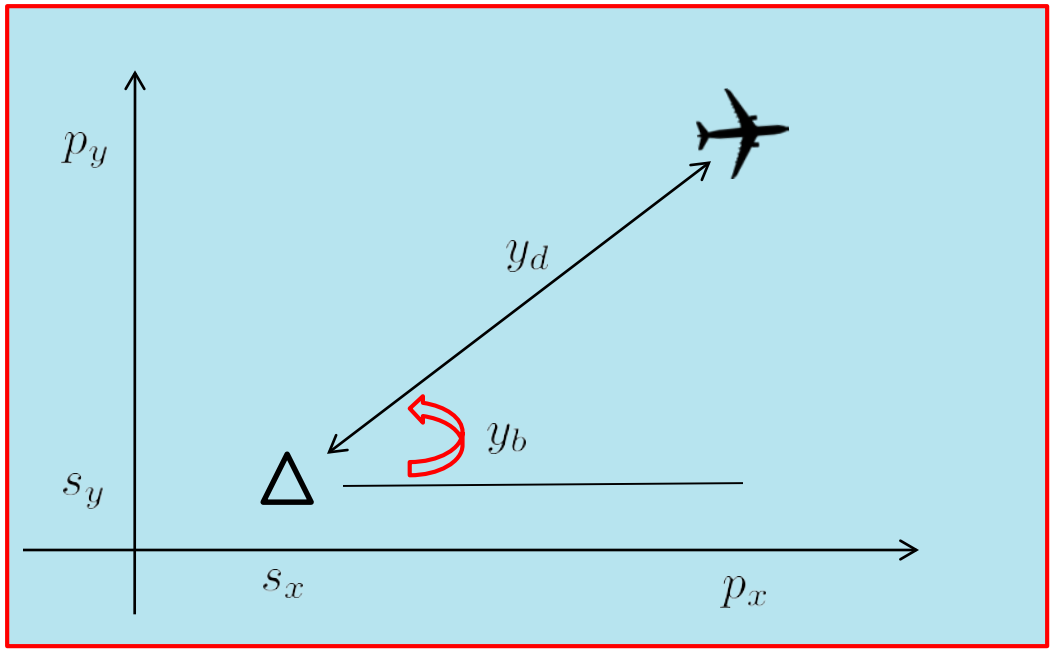
Speed corrections
(Gaussian noise with cov. Q)



Distance and bearing measurements

$$\begin{bmatrix} y_d(t) \\ y_b(t) \end{bmatrix} = \begin{bmatrix} \sqrt{(p_x(t)-s_x)^2+(p_y(t)-s_y)^2} \\ \tan^{-1}\left(\frac{p_y(t)-s_y}{p_x(t)-s_x}\right) \end{bmatrix} + \begin{bmatrix} v_d(t) \\ v_b(t) \end{bmatrix}$$

Uncorrelated Gaussian noise



Tracking

- Bootstrap PF with $N = 300$ particles: $\mathbf{x}_t^{(i)} = [p_x^{(i)}(t)\ p_y^{(i)}(t)\ v_x^{(i)}(t)\ v_y^{(i)}(t)]^T$

- Sampling step (propagation of particles)

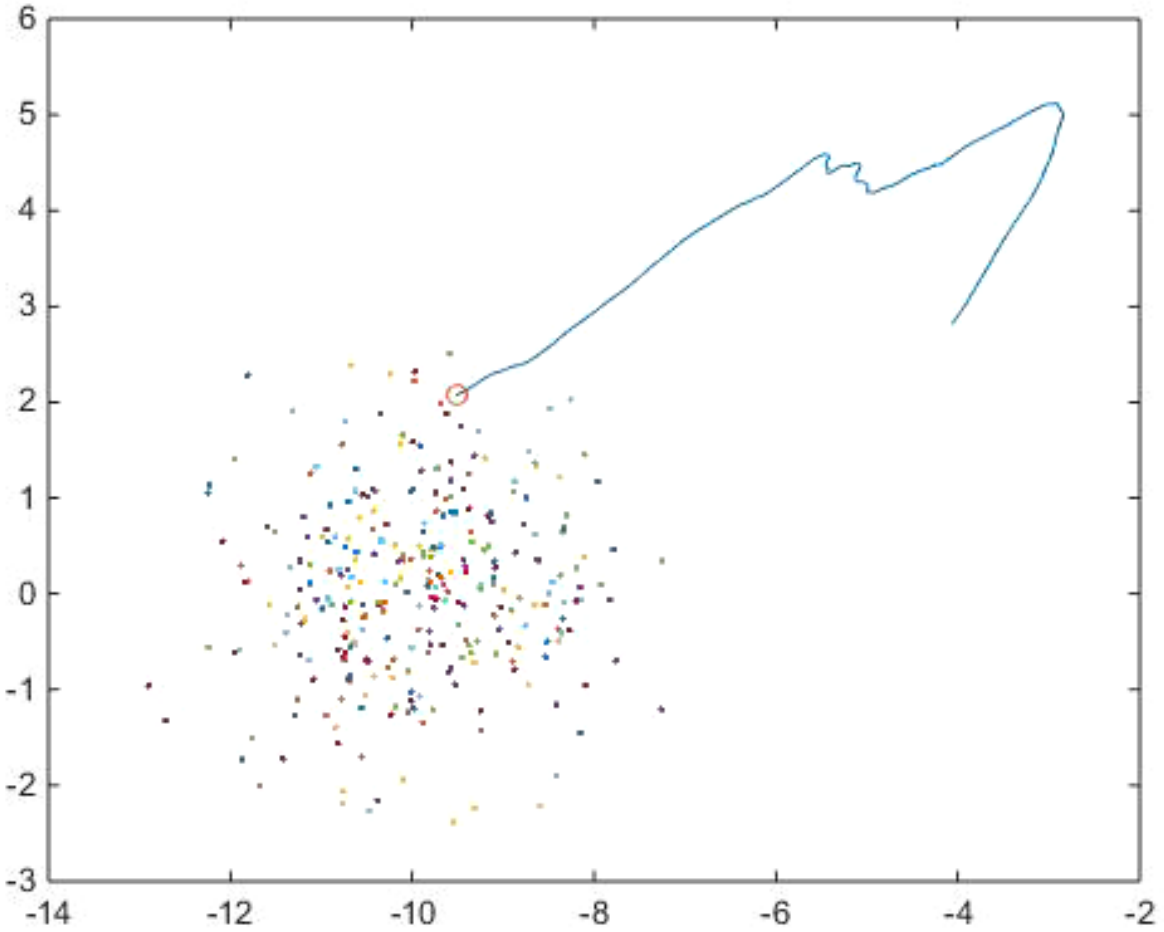
$$p_x^{(i)}(t) = p_x^{(i)}(t-1) + T_s v_x^{(i)}(t-1)$$
$$p_y^{(i)}(t) = p_y^{(i)}(t-1) + T_s v_y^{(i)}(t-1)$$
$$\begin{bmatrix} v_x^{(i)}(t) \\ v_y^{(i)}(t) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} v_x^{(i)}(t-1) \\ v_y^{(i)}(t-1) \end{bmatrix}, \mathbf{Q} \right)$$

- Evaluation of weights (likelihood of particles)

$$w_t^{(i)} = \mathcal{N}(y_d(t); \sqrt{(p_x^{(i)}(t) - s_x)^2 + (p_y^{(i)}(t) - s_y)^2}, \sigma_d^2)$$
$$\times \mathcal{N}\left(y_b(t); \tan^{-1}\left((p_y^{(i)}(t) - s_y)/(p_x^{(i)}(t) - s_x)\right), \sigma_b^2\right)$$

- Randomized resampling w.p. $\propto w_t^{(i)}$

Result



Conclusions

- MCMC and IS: powerful, all-around tools for Bayesian inference
- Applicable to any problem if tuned properly
  - Proposal distributions
  - Resampling schemes (in PF)
- Other MCMC derivatives
  - MCMC expectation-maximization algorithms
  - Hybrid MC
  - Slice sampler
  - Reversible jump MCMC for model selection