Implicit sampling for particle filters

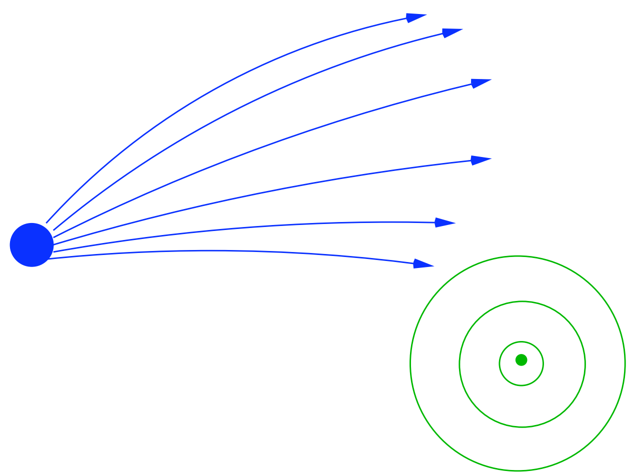
Alexandre Chorin and Xuemin Tu
University of California at Berkeley

Example:

Try to rescue people in a boat in the middle of the ocean from approximate knowledge of their initial position and from uncertain observations of a later position.

Equations of motion : $dx = f(x, \omega)dt + gdW$, or, in a discrete approximation, $x_{n+1} = x_n + \delta F(x_n) + GdW$,

Observations: $b^n = b(n\delta) = g(x^n) + V$, V random.



Special case: equations linear, pdf Gaussian, $\Rightarrow$ Kalman filter.

Extensions: extended Kalman filter, ensemble Kalman filter, try to fit a non-Gaussian situation into a Gaussian framework.

Simple particle filter:

Follow a bunch of "particles" (samples, replicas) whose empirical density at time $t = n\delta$ approximates a pdf P_n of a set of systems moved by the equations of motion constrained by the observations.

Given P_n:

First evolve the particles by the equations of motion alone: (generates a "prior" density).

Take the observations into account by weighting the particles. (generates a "posterior" density).

To avoid following irrelevant particles, resample, so that you have again a bunch of particles with equal weights. (for $\theta_k \sim [0,1]$, pick $\widehat{x}^{n+1} = x^{n+1}$ such that

$A^{-1} \sum_1^{i-1} W_j < \theta_k < A^{-1} \sum_1^{i} W_j$, where $A = \sum W_j$.

Final (important) step: go back, smooth the past, and resample.

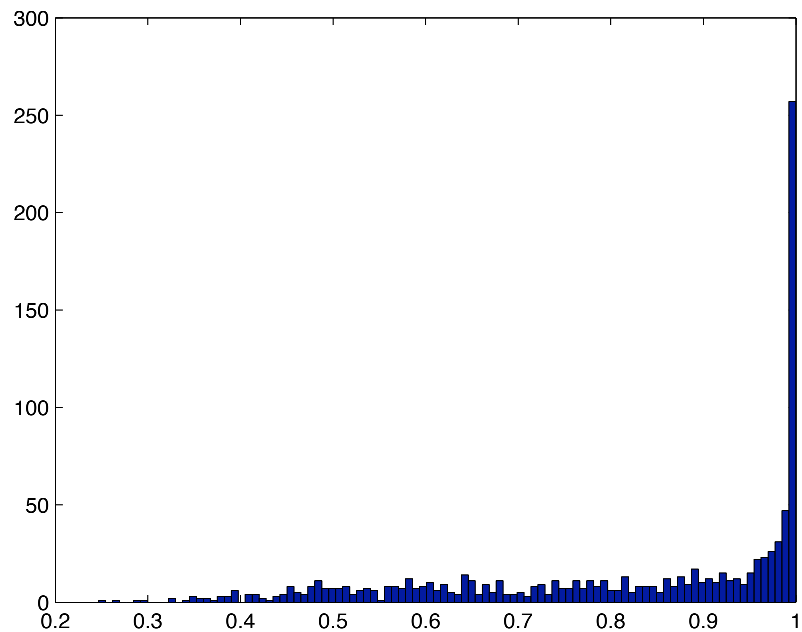
Bayes theorem:

$$P(x^{n+1}|x^n, b^{n+1}) = \frac{P(x^{n+1}|x^n)P(b^{n+1}|x^{n+1})}{P(b^{n+1}|b^n)}$$

Fails, in particular when there are many variables: (example from Bickel at al.):

Model: $\boldsymbol{x} \sim N(0,1)$ Observation: $\boldsymbol{b} = \boldsymbol{x} + N(0,1)$

$\dim(\boldsymbol{x}) = 100$, 1000 particles, 1000 runs for max weight distributions

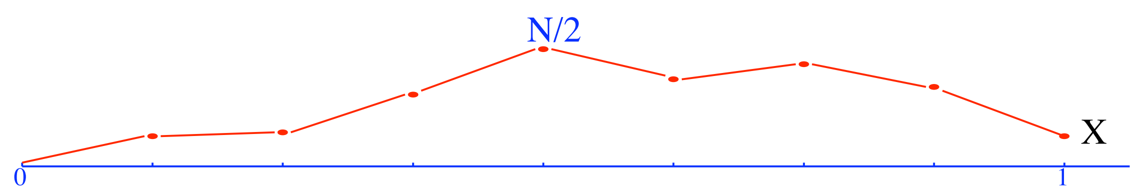


Partial remedy: better choice of prior in linear problems (special case of what we offer).

Our remedy: sampling by interpolation and iteration, special case of chainless sampling.

Simple example of sampling by interpolation and iteration: Non-linear Brownian bridge.

$dx = f(x,t)dt + \sqrt{\beta}dw$, $x(0) = 0$, $x(1) = X$.



Discretize: $x^{n+1} = x^n + f(x^n,t^n)\delta + (x^{n+1} - x^n)f'(x^n,t^n)\delta + W^{n+1}$, where $f'(x^n,t^n) = \frac{\partial f}{\partial x^n}(x^n,t^n)$

The joint probability density of the variables $x^1,\ldots,x^{N-1}$ is

$Z^{-1}\exp(-\sum_0^{N-1} V_n)$, where Z is the normalization constant and

$$V_n = \frac{\Big((1-\delta f')(x^{n+1}-x^n)-\delta f\Big)^2}{2\beta\delta}$$
$$= \frac{\Big(x^{n+1}-x^n-\delta f/(1-\delta f')\Big)^2}{2\beta_n},$$

where f, f' are functions of the x^n, t^n, and $\beta_n = \beta\delta/(1-\delta f')^2$.

Let $a_n = f(x^n, t^n)\delta/(1 - \delta f'(x^n, t^n))$.

Special case $f(x,t) = f(t)$, $f' = 0$. Each $x^{n+1} - x^n = N(a_n, \beta/N)$, with the $a_n = f(t^n)\delta$ known. $N = 2^k$.

Consider $x^{N/2}$. $x^{N/2} = \sum_1^{N/2}(x^n - x^{n-1}) \sim N(A_1, V_1)$, where $A_1 = \sum_1^{N/2} a_n, V_1 = \beta/2$.

On the other hand, $X = x^{N/2} + \sum_{N/2+1}^{N}(x^n - x^{n-1})$, so that $x^{N/2} \sim N(A_2, V_2)$, with

$$A_2 = X - \sum_{N/2}^{N-1} a_n, \quad V_2 = V_1.$$

The pdf of $x^{N/2}$ is the product of the two pdfs;

$$\begin{aligned}&\exp\left(-\frac{(x-A_1)^2}{2V_1}\right)\exp\left(-\frac{(x-A_2)^2}{2V_2}\right)\\=&\exp\left(-\frac{(x-\bar{a})^2}{2\bar{v}}\right)\exp(-\phi),\end{aligned}$$

where $\bar{v}=\frac{V_1V_2}{V_1+V_2}$, $\bar{a}=\frac{V_2A_1+V_1A_2}{V_1+V_2}$, and $\phi=\frac{(A_2-A_1)^2}{2(V_1+V_2)}$; $e^{-\phi}$ is the probability of getting from the origin to X, up to a normalization constant.

Pick a sample $\xi_1 \sim N(0,1)$; obtain a sample of $x^{N/2}$ by $x^{N/2} = \bar{a} + \sqrt{\bar{v}}\xi_1$.

Given $x^{N/2}$, sample $x^{N/4}, x^{3N/4}$, then $x^{N/8}$, $x^{3N/8}$, etc. Define $\xi = (\xi_1, \xi_2, \ldots, \xi_{N-1})$; for each ξ find a sample $\mathbf{x} = (x^1, \ldots, x^{N-1})$ such that

$$\begin{aligned} &\exp\left(-\frac{\xi_1^2 + \cdots + \xi_{N-1}^2}{2}\right) \exp\left(-\frac{(X - \sum_n a_n)^2}{2\beta}\right) \\ = &\exp\left(-\frac{(x^1 - x^0 - a_0)^2}{2\beta/N} - \frac{(x^2 - x^1 - a_1)^2}{2\beta/N} \right. \\ &\left. - \cdots - \frac{(x^N - x^{N-1} - a_{N-1})^2}{2\beta/N}\right), \end{aligned}$$

the factor $\exp\left(-\frac{(X - \sum_n a_n)^2}{2\beta}\right)$ on the left is the probability of the fixed end value X up to a normalization constant.

In linear problem, this factor is the same for all the samples and harmless. The Jacobian $J = \partial(x^1, \ldots, x^{N-1})/\partial(\xi^1, ..., \xi^{N-1})$ is a constant independent of the sample. Each sample is independent of any previous samples.

General case: iterate. First pick $\xi = (\xi_1, \xi_2, \ldots, \xi_{N-1})$, where each ξ_l, $l = 1, \ldots, N-1$, is drawn independently from the $N(0,1)$ density (this vector remains fixed during the iteration).

Make a first guess $\mathbf{x}_0 = (x_0^1, x_0^2, \ldots, x_0^{N-1})$ Evaluate the functions f, f' at $\mathbf{x}_j$, sample, repeat. The vectors converge to $\mathbf{x} = (x^1, \ldots, x^{N-1})$. Both a_n, β_n are functions of the final $\mathbf{x}$. The factor $F|J|$, $F = \exp\left(-\frac{(X - \sum_n a_n)^2}{2\sum_n \beta_n}\right)$, $J =$ Jacobian is a sampling weight.

Iteration converges if

$$KL < 1,$$

where K is the Lipshitz constant of f, and L is the length of the interval (here $L = 1$).

Application to filtering:

Start from P_n, pdf at time $n\delta$.

1. Pick $\xi \sim e^{-\xi^*\xi/2}/(2\pi)^{n/2}$, (reference variable)

2. Write the (unnormalized) pdf $P(b^{n+1}|x^{n+1})P(x^{n+1}|x^n)$ in the form $\exp(-(x^{n+1}-a_{n+1})^* H(x^{n+1})(x^{n+1}-a_{n+1}))\exp(-\Phi(n+1))$ (a "pseudo-Gaussian").

(remembering $\exp(-(x-a_1)^2/(2V_1))\exp(-(x-a_2)^2/(2V_2)) = \exp((x-\bar{a})^2/(2\bar{V})\exp(-\Phi)$,

with

$\bar{a} = \frac{V_2 a_1 + V_1 a_2}{V_1+V_2}, \bar{V} = \frac{V_1 V_2}{V_1+V_2}, \Phi = \frac{(a_1-a_2)^2}{2(V_1+V_2)}$.

3. Equate arguments of exponentials.

(example: sample $y \sim \exp(-(y-a)^2/(2v))/\sqrt{2\pi v}$ as a function of $\xi \sim \exp(-\xi^2/2)/\sqrt{2\pi}$; equating arguments gives $y = a + \sqrt{v}\xi$).

Yields $\xi^*\xi = (x^{n+1} - a_{n+1})^* H(x^{n+1})(x^{n+1} - a_{n+1})$

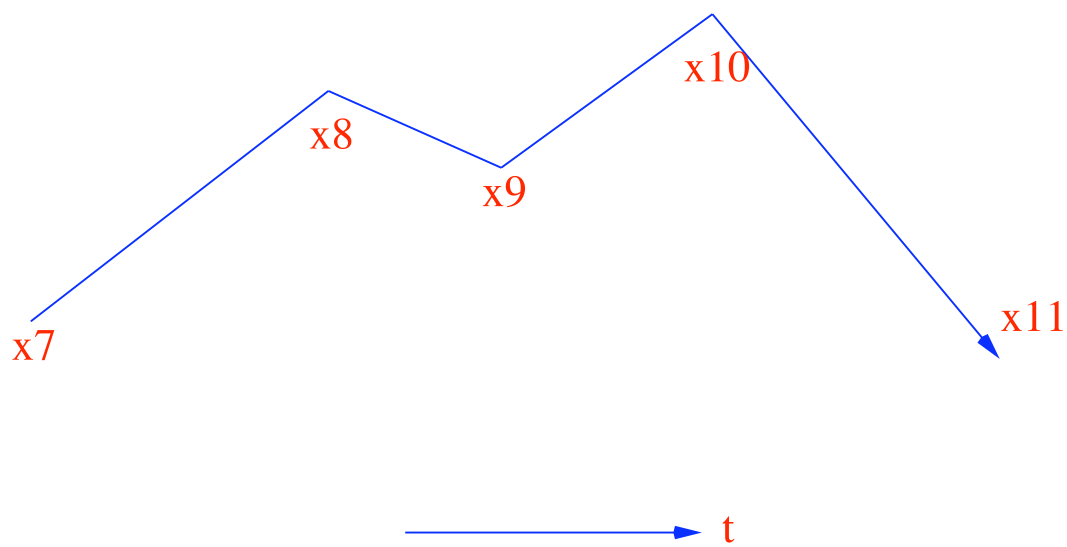
Solve by iteration: $x_j \to x^{n+1}$, $x_0 = 0$, Define $a_j = a(x_j), H(j) = H(x_j)$, and factor $H(x_j) = LL^*$, L lower triangular. Then

$x_{j+1} = a_j + L^{-1}\xi, \Phi_j \to \Phi$.

4. Estimate J, the Jacobian.

Each value of x^{n+1} appears with probability (up to constant factors) $\exp(-\xi^*\xi/2)/|J|$, and its value is $\exp(-\xi^*\xi/2)\exp(-\Phi)$, so the integration weight is $\exp(-\Phi)|J|$ (the variability in $\exp(-\xi^*\xi/2)$ has been discounted in advance).

Backward sampling:



Analogous to the Brownian bridge, same steps, $H(x_n)$ takes into account the known positions at time $(n+1)\delta$ to update x^n; a new forward step is needed.

Special cases: Linear equations and Gaussian densities- a single particle and a single iteration produce the Kalman variance and means.

Linear observation function: a single iteration converges.

Diagonal observations: algebra simplifies.

Example 0: (The Bickel Gaussian) all particles have equal weights.

Example 1: The ship azimuth problem.

$$x^{n+1} = x^n + u^{n+1},$$
$$y^{n+1} = y^n + v^{n+1},$$

$$u^{n+1} = N(u^n, \beta),$$
$$v^{n+1} = N(v^n, \beta),$$

$$b^n = \text{arctan}(y^n/x^n) + N(0, s),$$

Initial data given.

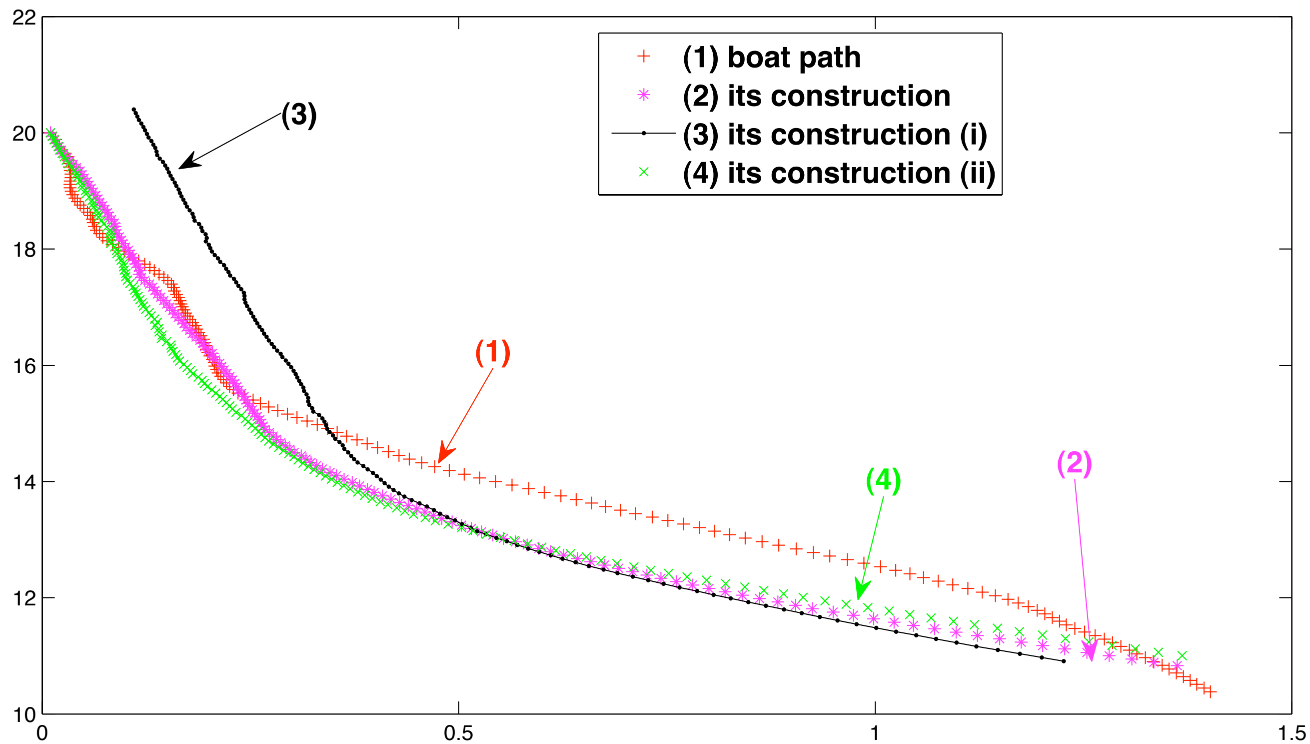
Make a run, generate synthetic observations, use them to reconstruct the path.

First question: How accurately can one reconstruct the boat path from the data?

160 steps and 160 observations. The empirical s.d. of the observation noise has mean s and s.d $.11s$.

Many runs made, those inconsistent with the data rejected, the mean and variance of those that are consistent tabulated. A smaller error in a single run is accidental.

Sample boat path and its reconstruction (100 particles and 160 observations)



Construction (i): with perturbation of the initial positions
Construction (ii): with perturbation of the system variances

Table I

Intrinsic uncertainty in the azimuth problem

step	x component	y component
40	.0005	.21
80	.004	.58
120	.010	.88
160	.017	.95

Table II
The mean of the discriminant D as a function of $\sigma_{\text{assumed}}/\sigma$, 30 particles

$\sigma_{\text{assumed}}/\sigma$	3000 runs	200 runs
.5	$1.15 \pm .01$	$1.15 \pm .06$
.6	$1.07 \pm .01$	$1.07 \pm .06$
.7	$1.07 \pm .01$	$1.07 \pm .05$
.8	$1.04 \pm .01$	$1.04 \pm .05$
.9	$1.02 \pm .01$	$1.02 \pm .05$
1.0	$1.01 \pm .01$	$1.00 \pm .05$
1.1	$.95 \pm .01$	$1.01 \pm .05$
1.2	$.95 \pm .01$	$.95 \pm .04$
1.3	$.94 \pm .01$	$.96 \pm .05$
1.4	$.90 \pm .01$	$.88 \pm .04$
1.5	$.89 \pm .01$	$.88 \pm .04$
2.0	$.85 \pm .01$	$.83 \pm .04$

Example 3: The ecological dynamics

$$\begin{aligned}
\frac{dP}{dt} &= \frac{N}{0.2+N}\gamma P - 0.1P - 0.6\frac{P}{0.1+P}Z + N(0,\sigma_P^2)\\
\frac{dZ}{dt} &= 0.18\frac{P}{0.1+P}Z - 0.1Z + N(0,\sigma_Z^2)\\
\frac{dN}{dt} &= 0.1D + 0.24\frac{P}{0.1+P}Z - \gamma P\frac{N}{0.2+N} + 0.05Z + N(0,\sigma_N^2)\\
\frac{dD}{dt} &= -0.1D + 0.1P + 0.18\frac{P}{0.1+P}Z + 0.05Z + N(0,\sigma_D^2)
\end{aligned}$$

$$\gamma_t = 0.14 + 3\Delta\gamma_t, \quad \Delta\gamma_t = 0.9\Delta\gamma_{t-1} + N(0,\sigma_\gamma^2)$$

$$\log P_n^{\text{obs}} = \log P_n + N(0,\sigma_{\text{obs}}^2)$$

Observation: the concentration of plant pigments in the eastern tropical Pacific from late 1997 to mid 2002 (NASA's SeaWiFS)

The filter results with 10 particles