Negative Association, Ordering and Convergence of Resampling Methods

Nicolas Chopin

ENSAE, Paristech

(Joint work with Mathieu Gerber and Nick Whiteley, University of Bristol)

Resampling schemes: Informal definition

A resampling scheme is a randomization procedure that takes as an input a weighted sample $\{(X^n, W^n)\}_{n=1}^N$ and returns as an output resampled variables $\{X^{A^n}\}_{n=1}^N$, where A^n is a random index in $\{1, \ldots, N\}$.

A good resampling scheme should be such that

$$\frac{1}{N}\sum_{n=1}^{N} \delta(X^{A^n}) \approx \sum_{n=1}^{N} W^n \delta(X^n)$$

or, in words, the empirical probability measure of the resampled variables should remain close (in some sense) to the weighted empirical measure of the input variables.

Motivation: Resampling, a key element of particle filtering

It is well known that particle filters ‘collapse’ if the particles are not resampled from time to time.

(Other applications of resampling algorithms include e.g. survey sampling and weighted bootstrap.)

Most commonly used resampling methods (in PF)

- **Multinomial resampling:**

$$A^n = F_N^-(U^n), \quad n = 1, \ldots, N, \quad F_N(x) = \sum_{n=1}^{N} W^n \mathbb{I}(n \leq x)$$

  where $\{U^n\}_{n=1}^N$ are i.i.d. $\mathcal{U}(0,1)$ random variables.

- **Stratified resampling** (Kitagawa, 1996):

$$A^n = F_N^-\Big(\frac{n-1+U^n}{N}\Big), \quad n = 1, \ldots, N$$

  where $\{U^n\}_{n=1}^N$ are i.i.d. $\mathcal{U}(0,1)$ random variables.

- **Systematic resampling** (Carpenter et al., 1999):

$$A^n = F_N^-\Big(\frac{n-1+U}{N}\Big), \quad n = 1, \ldots, N$$

  where $U \sim \mathcal{U}(0,1)$.

Inverse CDF plot

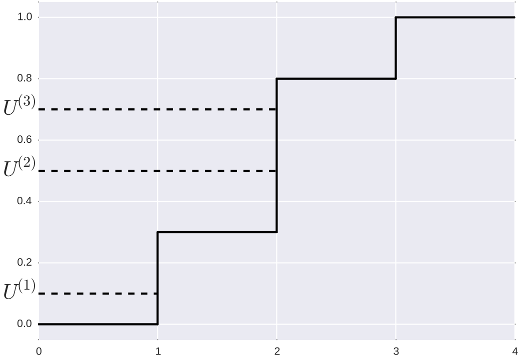


Figure 1: Function $F_N(x) = \sum_{n=1}^{N} W^n \mathbb{I}(n \leq x)$

Schizophrenic Monte Carlo

In practice, We use stratified/systematic (rather than multinomial) resampling, because these schemes are (a) a bit faster, and (b) leads to lower-variance estimates numerically. (See next slide)

In theory, we only consider multinomial resampling, as it is so much easier to study; indeed, resampled particles are IID, from distribution

$$\sum_{n=1}^{N} W^n \delta(X^n).$$

As a result, little is known about stratified/systematic; even wether they are consistent or not.

Numerical comparison of resampling schemes

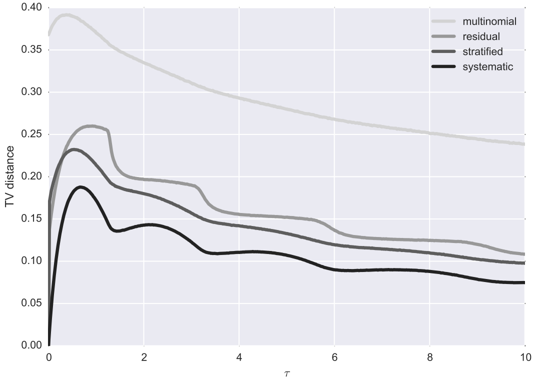


Figure 2: TV distance between empirical distributions of weighted particles, and resampled particles as a function of τ; particles are $\sim N(0,1)$, weight function is $w(x) = exp(-\tau x^2/2)$.

Outline

1. Introduction
2. Consistency results for unordered resampling schemes
3. Analysis of Hilbert-ordered resampling schemes
4. Some numerical results
5. Conclusion

Outline

Resampling schemes: Formal definition

Definition

A resampling scheme is a mapping $\rho : [0,1]^{\mathbb{N}} \times \mathcal{Z} \to \mathcal{P}_f(\mathcal{X})$ such that, for any $N \geq 1$ and $z = (x^n, w^n)_{n=1}^N \in \mathcal{Z}^N$,

$$\rho(u, z) = \frac{1}{N} \sum_{n=1}^{N} \delta(x^{a_N^n(u,z)}),$$

where for each n, $a_N^n : [0,1]^{\mathbb{N}} \times \mathcal{Z}^N \to 1:N$ is a certain measurable function.

Notation:

1. $\mathcal{X} \subseteq \mathbb{R}^d$ is a measurable set.
2. $\mathcal{P}(\mathcal{X})$ is the set probability measures on $\mathcal{X}$.
3. $\mathcal{P}_f(\mathcal{X})$ is the set of discrete probability measures on $\mathcal{X}$.
4. $\mathcal{Z} := \bigcup_{N=1}^{+\infty} \mathcal{Z}^N$ with $\mathcal{Z}^N = \left\{(x, w) \in \mathcal{X}^N \times \mathbb{R}_+^N : \sum_{n=1}^N w_n = 1\right\}$.

Consistent resampling schemes

We consider in this work that a resampling scheme is consistent if it is weak-convergence-preserving.

Definition

Let $\mathcal{P}_0 \subseteq \mathcal{P}(\mathcal{X})$. Then, we say that a resampling scheme $\rho : [0,1]^{\mathbb{N}} \times \mathcal{Z} \to \mathcal{Z}$ is $\mathcal{P}_0$-consistent if, for any $\pi \in \mathcal{P}_0$ and random sequence $(\zeta^N)_{N\geq 1}$ such that $\pi^N \Rightarrow \pi$, $\mathbb{P}$-a.s., one has

$$\rho(\zeta^N) \Rightarrow \pi, \quad \mathbb{P}-\text{a.s.}$$

Remarks:

1. All the random variables are defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$.
2. ζ^N is a r.v. that takes its value in $\mathcal{Z}^N$ and $\pi^N \in \mathcal{P}_f(\mathcal{X})$ is the corresponding probability measure: $\zeta^N = (W^n, X^n)_{n=1}^N$, $\pi^N = \sum_{n=1}^N W^n \delta(X^n)$.
3. It is well known that multinomial resampling is $\mathcal{P}(\mathcal{X})$-consistent.

A general consistency result: Preliminaries

- To simplify the presentation we assume henceforth that $\mathcal{X} = \mathbb{R}^d$.

A general consistency result: Preliminaries

- To simplify the presentation we assume henceforth that $\mathcal{X} = \mathbb{R}^d$.
- A collection of random variables $(Z^n)_{n=1}^N$ are negatively associated (NA) if, for every pair of disjoint subsets I_1 and I_2 of $\{1, \dots N\}$,

$$\mathrm{Cov}\Big(\varphi_1\big(Z^n, n \in I_1\big),\, \varphi_2\big(Z^n, n \in I_2\big)\Big) \leq 0$$

  for all coordinatewise non-decreasing functions φ_1 and φ_2, such that for $k \in \{1, 2\}$, $\varphi_k : \mathbb{R}^{|I_k|} \to \mathbb{R}$ and such that the covariance is well-defined.

A general consistency result: Preliminaries

- To simplify the presentation we assume henceforth that $\mathcal{X} = \mathbb{R}^d$.

- A collection of random variables $(Z^n)_{n=1}^N$ are negatively associated (NA) if, for every pair of disjoint subsets I_1 and I_2 of $\{1, \dots N\}$,

$$\mathrm{Cov}\Big(\varphi_1\big(Z^n, n \in I_1\big),\, \varphi_2\big(Z^n, n \in I_2\big)\Big) \leq 0$$

  for all coordinatewise non-decreasing functions φ_1 and φ_2, such that for $k \in \{1, 2\}$, $\varphi_k : \mathbb{R}^{|I_k|} \to \mathbb{R}$ and such that the covariance is well-defined.

- Let $\widetilde{\mathcal{P}}_b(\mathcal{X}) \subset \mathcal{P}(\mathcal{X})$ be a set of probabilities densities with "not too thin tails" (see paper for exact definition).

Main result

Theorem ($\mathcal{X} = \mathbb{R}^d$ to simplify)

Let $\rho : [0,1]^{\mathbb{N}} \times \mathcal{Z} \to \mathcal{P}_f(\mathcal{X})$ be an unbiased resampling scheme such that:

1. *For any $N \geq 1$ and $z \in \mathcal{Z}^N$ the collection of random variables $\left\{\#^n_{\rho,z} := \sum_{m=1}^N \mathbb{I}(A^m = n)\right\}_{n=1}^N$ is negatively associated;*
2. *There exists a sequence $(r_N)_{N\geq 1}$ of non-negative real numbers such that $r_N = o(N/\log N)$, and, for N large enough,*

$$\sup_{z\in\mathcal{Z}^N} \sum_{n=1}^N \mathbb{E}\left[(\Delta^n_{\rho,z})^2\right] \leq r_N\, N, \qquad \sum_{N=1}^{\infty} \sup_{z\in\mathcal{Z}^N} \mathbb{P}\left(\max_{n\in 1:N} \left|\Delta^n_{\rho,z}\right| > r_N\right) < +\infty.$$

Then, ρ is $\widetilde{\mathcal{P}}_b(\mathcal{X})$-consistent.

Notation: $\Delta^n_{\rho,z} = \#^n_{\rho,z} - NW^n$.

Definition: ρ is unbiased if $\mathbb{E}\left[\Delta^n_{\rho,z}\right] = 0$ for all n and $z \in \mathcal{Z}$.

Applications

From the previous theorem we deduce the following corollary:

Corollary

- *Multinomial resampling is $\widetilde{\mathcal{P}}_b(\mathcal{X})$-consistent (not new);*
- *Residual resampling is $\widetilde{\mathcal{P}}_b(\mathcal{X})$-consistent (not new);*
- *Stratified resampling is $\widetilde{\mathcal{P}}_b(\mathcal{X})$-consistent (new!);*
- *Residual/Stratified resampling is $\widetilde{\mathcal{P}}_b(\mathcal{X})$-consistent (new!);*
- *SSP resampling is $\widetilde{\mathcal{P}}_b(\mathcal{X})$-consistent (new!, see next slides).*

Remark: This corollary shows that the negative association condition of the theorem holds for resampling schemes producing a set of resampled values $\{X^{A^n}\}_{n=1}^N$ having very different properties.

Resampling as a randomized rounding operation

Definition

For $\xi \in \mathbb{R}_+$, the random variable $Y : \Omega \to \mathbb{N}$ is a randomized rounding of ξ if

$$\mathbb{P}\big(Y = \lfloor \xi \rfloor + 1\big) = \xi - \lfloor \xi \rfloor, \quad \mathbb{P}\big(Y = \lfloor \xi \rfloor\big) = 1 - \big(\xi - \lfloor \xi \rfloor\big).$$

Any randomized rounding technique that takes as input a vector $(\xi_1, \ldots, \xi_N) \in \mathbb{R}_+^N$ returns as output a random vector $(Y^1, \ldots, Y^N) \in \mathbb{N}^N$ verifying

$$\sum_{n=1}^{N} Y^n = \sum_{n=1}^{N} \xi_n, \quad \mathbb{P} - \text{almost surely}$$

may be used to construct an unbiased resampling mechanism.

Resampling as a randomized rounding operation

Definition

For $\xi \in \mathbb{R}_+$, the random variable $Y : \Omega \to \mathbb{N}$ is a randomized rounding of ξ if

$$\mathbb{P}\big(Y = \lfloor \xi \rfloor + 1\big) = \xi - \lfloor \xi \rfloor, \quad \mathbb{P}\big(Y = \lfloor \xi \rfloor\big) = 1 - \big(\xi - \lfloor \xi \rfloor\big).$$

Any randomized rounding technique that takes as input a vector $(\xi_1, \ldots, \xi_N) \in \mathbb{R}_+^N$ returns as output a random vector $(Y^1, \ldots, Y^N) \in \mathbb{N}^N$ verifying

$$\sum_{n=1}^{N} Y^n = \sum_{n=1}^{N} \xi_n, \quad \mathbb{P} - \text{almost surely}$$

may be used to construct an unbiased resampling mechanism.

Systematic resampling as a rounding operation

Systematic resampling is an example of resampling scheme constructed in this way.

However, it is hard to analyse because the underlying rounding process is such that the dependence structure of the output depends on the way we order the input point set.

In addition, it is easy to show that the collection of offspring numbers $\{\#^n_{\rho,z}\}_{n=1}^N$ produced by systematic resampling is in general not NA.

SSP resampling as a “valid” version systematic resampling

The SSP (for *Srinivasan Sampling Process*) resampling scheme is based on Srinivasan’s (2001) randomized rounding technique.

This resampling mechanism requires $N-1$ uniform random variables in $[0,1]$ and $\mathcal{O}(N)$ operations (as all the other resampling schemes mentioned above).

SSP resampling satisfies the negative association condition of our general consistency result (as well as the other conditions) and is therefore $\widetilde{\mathcal{P}}_b(\mathcal{X})$-consistent.

Basic idea behind of SSP

Start with $Y^n = NW^n$ for $n = 1, \ldots, N$. Take a pair, say $Y^1 = 3.7$, $Y^2 = 2.2$.

- With probability p, increase both, by amount 0.3: then Y^1 is 4.
- With probablity $(1 - p)$, decrease both, by amount 0.2; then Y^2 is 2.

(Choose p so that the scheme remains unbiased: $p = 2/5$.)

Pair the particle with a fractional weight with another particle, and start over.

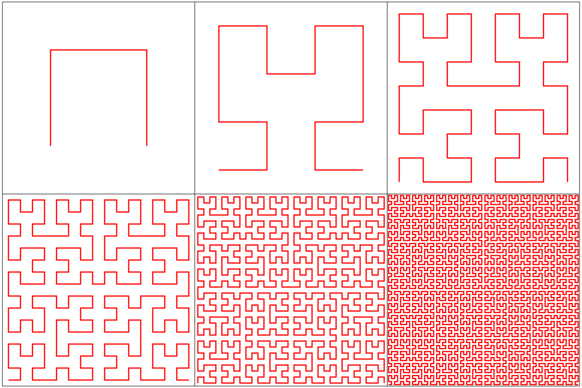
Very vague sketch of the proof

In a first step, we show that consistency is equivalent to a certain condition on the set of points, when ordered through the Hilbert curve.
In a second step, we use the NA condition to show that the same technical condition holds whatever the order of the input points.

The Hilbert space filling curve

The Hilbert space filling curve $H : [0,1] \to [0,1]^d$ is a continuous and surjective mapping.

It is defined as the limit of a sequence $(H_m)_{m \geq 1}$



First six elements of the sequence $(H_m)_{m \geq 1}$ for $d = 2$ (source: Wikipedia)

Outline

Motivation: Kitagawa's Conjecture

Using simulations, Kitagawa (1996) noticed the following.

Conjecture (Kitagawa, 1996)

Assume $\mathcal{X} = \mathbb{R}$ (i.e. $d = 1$). Then, if the point $\{X^n\}_{n=1}^N$ are ordered before the resampling, The approximation error of stratified resampling is of size $\mathcal{O}_{\mathbb{P}}(N^{-1})$;

Is this true? does it generalize to $d > 1$?

Variance of Hilbert-ordered stratified resampling

Theorem ($\mathcal{X} = \mathbb{R}^d$ to simplify)

Let $\pi \in \widetilde{\mathcal{P}}_b(\mathcal{X})$ be such that $\pi(\mathrm{d}x) = p(x)\lambda_d(\mathrm{d}x)$ where $p(x)$ is strictly positive on $\mathcal{X}$, and $(\zeta^N)_{N\geq 1}$, with $\zeta^N \in \mathcal{Z}^N$, such that

$$\sum_{n=1}^{N} W_N^n \delta_{X_N^n} \Rightarrow \pi, \quad \lim_{N\to+\infty} \left(\max_{n\in 1:N} W_N^n \right) = 0, \quad \mathbb{P} - a.s.$$

Then, the Hilbert ordered stratified resampling scheme ρ^*_{strat} is such that

1. *For any $\varphi \in \mathcal{C}_b(\mathcal{X})$, $\mathrm{Var}\left(\rho^*_{\mathrm{strat}}(\zeta^N)(\varphi)\right) = o(1/N)$.*
2. *If $\varphi : \mathcal{X} \to \mathbb{R}$ is such that, for constants $C_{\varphi,\psi} < +\infty$ and $\gamma \in (0,1]$,*

$$\left|\varphi \circ \psi^{-1}(x) - \varphi \circ \psi^{-1}(y)\right| \leq C_{\varphi,\psi} \|x - y\|_2^{\gamma}, \quad \forall (x,y) \in (0,1)^d$$

   we have $\mathrm{Var}\left(\rho^*_{\mathrm{trat}}(\zeta^N)(\varphi)\right) \leq \left(2\sqrt{d+3}\right)^{2\gamma} C^2_{\varphi,\psi} N^{-1-\frac{\gamma}{d}}$.
3. *For φ as in 2., $\mathrm{Var}\left(\rho^*_{\mathrm{strat}}(\zeta^N)(\varphi)\right) = o\left(N^{-1-\frac{\gamma}{d}}\right)$ if $\mathcal{X} = [a,b] \subset \mathbb{R}^d$.*

CLT

$$N^{1/2}\left(\sum_{n=1}^{N} W_t^n\varphi(X_t^n) - \mathbb{Q}_t(\varphi)\right) \Rightarrow N\left(0, \mathcal{V}_t(\varphi)\right)$$

where the asymptotic variances are defined recursively:

$$\begin{aligned}
\mathcal{V}_t\left[\varphi\right] &= \frac{1}{\ell_t^2}\widetilde{\mathcal{V}}_t\left[G_t\left\{\varphi - \pi_t(\varphi)\right\}\right] \\
\widehat{\mathcal{V}}_t\left[\varphi\right] &= \mathcal{V}_t\left[\varphi\right] + R_t\left(\rho, \varphi\right) \\
\widetilde{\mathcal{V}}_{t+1}\left[\varphi\right] &= \widehat{\mathcal{V}}_t\left[M_{t+1}(\varphi)\right] + \pi_t\left[V_{t+1}(\varphi)\right]
\end{aligned}$$

We proved that $R_t(\varphi) = 0$ for the Hilbert-ordered version of stratified resampling. (It is > 0 for multinomial and residual, see C, 2004).

Note: also optimality results for the auxiliary weight function of the APF, where the optimal function depends on the resampling scheme.

Outline

Set-up

Following Guarniero et al. (2016), we consider the linear Gaussian state-space models where $X_0 \sim N_d(0, I_d)$, and, for $t = 1, \ldots, T$,

$$X_t = FX_{t-1} + V_t, \qquad V_t \sim N_d(0, I_d),$$
$$Y_t = X_t + W_t, \qquad W_t \sim N_d(0, I_d),$$

with $F = (\alpha^{|i-j|+1})_{i,j=1:d}$, and $\alpha = 0.4$.

We compare below particle filter algorithms based on (i) stratified resampling, (ii) Hilbert-ordered stratified resampling and (iii) SSP resampling.

Results are presented for the bootstrap particle filter and for the particle filter based on a guided proposal distribution.

We take $d = 5$, $T = 500$ and $N = 2^{13}$.

Results (log-likelihood estimation)

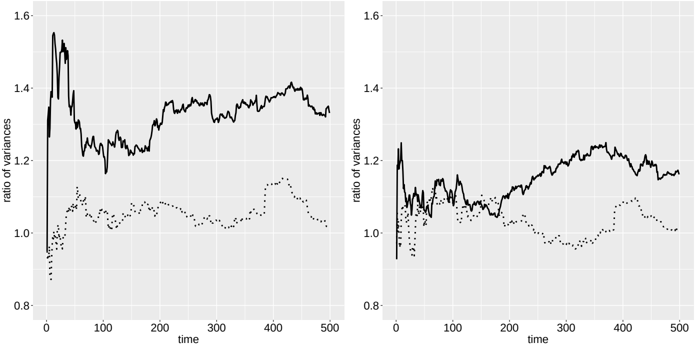


Figure 3: Estimation of the log-likelihood function as a function of t. The left (resp. right) plot gives the variance of SMC based on stratified resampling divided by that of SMC based on Hilbert-ordered stratified resampling (resp. unordered SSP resampling). Continuous lines are for SMC based on the guided proposal while the dotted line is for the bootstrap particle filter. Results are based on 1 000 independent runs of the algorithms.

Outline

Summary

We established a general consistency result for resampling schemes satisfying a negative association condition.

We showed how this result can be used to prove the validity of some of the most commonly used resampling mechanisms in SMC.

However, the consistency of systematic resampling applied on randomly ordered point set remains an open problem (but it is worth addressing it?).

Lastly, we showed that sorting the particles along the Hilbert curve allows to build (i) resampling schemes that converge faster than $N^{-1/2}$ and (ii) convergent deterministic resampling mechanisms.