ON RESAMPLING ALGORITHMS FOR PARTICLE FILTERS

Jeroen D. Hol, Thomas B. Schön, Fredrik Gustafsson

Division of Automatic Control
Department of Electrical Engineering
Linköping University
SE-581 83, Linköping, Sweden
{hol, schon, fredrik}@isy.liu.se

ABSTRACT

In this paper a comparison is made between four frequently encountered resampling algorithms for particle filters. A theoretical framework is introduced to be able to understand and explain the differences between the resampling algorithms. This facilitates a comparison of the algorithms with respect to their resampling quality and computational complexity. Using extensive Monte Carlo simulations the theoretical results are verified. It is found that systematic resampling is favourable, both in terms of resampling quality and computational complexity.

1. INTRODUCTION

The resampling step is a crucial and computationally expensive part in a particle filter [1]. Hence, a well argued choice of resampling method is justified as the entire method benefits from reduced complexity and/or improved quality of the resampling step. In the literature quite a few different resampling methods can be found. The most frequently encountered algorithms are multinomial resampling [2], stratified resampling [1, 3], systematic resampling [3, 4] and residual resampling [5]. Convergence results have been derived for some of them, see e.g., [6, 7]. However, discussions dealing with how and why the resampling algorithms work are scattered among many papers and books and, to the best of the author's knowledge, a detailed overview discussing their principles is missing. This paper aims at filling this gap, by analysing and comparing frequently used resampling algorithms and their implementations. The algorithms are compared with respect to resampling quality and computational efficiency, both theoretically and using simulations.

2. RESAMPLING ALGORITHMS

The resampling step modifies the weighted approximate density p_N to an unweighted density $\hat{p}_N$ by eliminating particles having low importance weights and by multiplying particles having high importance weights. More formally:

$$p_N(x) = \sum_{i=1}^{N} w_i \delta(x - x_i)$$

is replaced by

$$\hat{p}_N(x) = \sum_{k=1}^{N} \frac{1}{N} \delta(x - x_k^*) = \sum_{i=1}^{N} \frac{n_i}{N} \delta(x - x_i)$$

where n_i is the number of copies of particle x_i in the new set of particles $\{x_k^*\}$. Convergence can be proved by assuming that the resampled density is 'close' to the original density [1, 6]. That is, for any function $g(\cdot)$ it holds that

$$\mathrm{E}\left[\left(\int g(x) p_N(x) dx - \int g(x) \hat{p}_N(x) dx \right)^2 \right] \xrightarrow{N \to \infty} 0.$$

There are many different methods to generate the x_k^*. In the particle filter literature four 'basic' resampling algorithms can be identified:

1. **Multinomial resampling**
   Generate N ordered uniform random numbers

   $$u_k = u_{k+1} \tilde{u}_k^{\frac{1}{k}}, \ u_N = \tilde{u}_N^{\frac{1}{N}}, \text{ with } \tilde{u}_k \sim \mathrm{U}[0, 1)$$

   and use them to select x_k^* according to the multinomial distribution. That is,

   $$\begin{aligned} x_k^* &= x(F^{-1}(u_k)) \\ &= x_i \text{ with } i \text{ s.t. } u_k \in \left[\sum_{s=1}^{i-1} w_s, \sum_{s=1}^{i} w_s \right), \end{aligned}$$

   where F^{-1} denotes the generalised inverse of the cumulative probability distribution of the normalised particle weights.

2. **Stratified resampling**
   Generate N ordered random numbers

   $$u_k = \frac{(k-1) + \tilde{u}_k}{N}, \text{ with } \tilde{u}_k \sim \mathrm{U}[0, 1)$$

and use them to select x_k^* according to the multinomial distribution.

3. **Systematic resampling**
Generate N ordered numbers
$$u_k = \frac{(k-1) + \tilde{u}}{N}, \text{ with } \tilde{u} \sim \text{U}[0,1)$$
and use them to select x_k^* according to the multinomial distribution.

4. **Residual resampling**
Allocate $n_i' = \lfloor Nw_i \rfloor$ copies of particle x_i to the new distribution. Additionally, resample $m = N - \sum n_i'$ particles from $\{x_i\}$ by making n_i'' copies of particle x_i where the probability for selecting x_i is proportional to $w_i' = Nw_i - n_i'$ using one of the resampling schemes mentioned earlier.

All these algorithms are unbiased and can be implemented in $\mathcal{O}(N)$ as the random numbers are ordered, but have different computational complexities. The methods apply different sample generation methods as illustrated in Fig. 1. Hence,
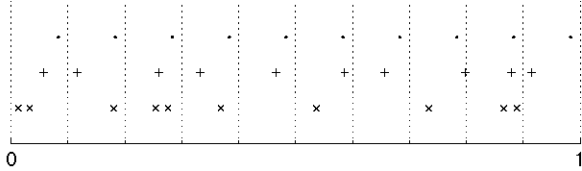

Fig. 1. Ten standard uniform samples generated using multinomial resampling (x), stratified resampling (+) and systematic resampling (·).

the different algorithms have different resampling qualities, as illustrated by the following analysis.

3. RESAMPLING QUALITY

As mentioned in the introduction, the quality of resampling is implicitly defined in the sense of the distance between the integrals

$$I(g) = \text{E}_{p_N}[g] = \int g(x)p_N(x) = \sum_{i=1}^{N} g(x_i)w_i \quad (1\text{a})$$

$$\hat{I}(g) = \text{E}_{\hat{p}_N}[g] = \int g(x)\hat{p}_N(x) = \frac{1}{N}\sum_{k=1}^{N} g(x_k^*) \quad (1\text{b})$$

for arbitrary functions g. Variance is natural measure for this distance, but other measures can be used as well. Depending on the particular resampling algorithm it might be possible to rewrite (1b) to

$$\hat{I}(g) = \sum_{i=1}^{N} \frac{n_i}{N}g(x_i), \quad \hat{I}(g) = \frac{1}{N}\sum_{k=1}^{N} g(x(F^{-1}(u_k)). \quad (2)$$

Integrals such as (1) are the subject of Monte Carlo integration theory [8]. The simplest method is to draw $u_k \sim \text{U}[0,1)$. This case,

$$\hat{I}^u(g) = \frac{1}{N}\sum_{k=1}^{N} g(x_k^*), \quad x_k^* \sim p_N, \quad (3)$$

has the properties that it is unbiased, $\text{E}[\hat{I}^u(g)] = I(g)$, and $\text{Var}\,\hat{I}^u(g) = N^{-1}\,\text{Var}_{P_N}\,g$. There exist several tools to reduce the distance between the integrals (1a) and (1b).

3.1. Set restriction

Set restriction is a powerful method to reduce integration variance [8]. The reasoning is rather simple: variance is defined by

$$\text{Var}\,\hat{I} = \text{E}_{\{x_k^*\}}[\hat{I}(g) - I(g)]^2$$
$$= \text{E}_{\{n_i\}}\Big[\sum_{i=1}^{N} \frac{n_i - Nw_i}{N}g(x_i)\Big]^2, \quad (4)$$

where (2) has been used. In the case that the n_i are distributed according to the multinomial distribution the set of possible values for each n_i is given by $S_i = \{0,\ldots,N\}$. By restricting this set to values which lie closer to Nw_i the variance is reduced.

3.2. Stratification

Stratification is a method that originated from survey sampling [9]. The domain of the random variable is partitioned into different strata, that is $\mathcal{D} = \bigcup_{j=1}^{p} \mathcal{D}_j$ where $\mathcal{D}_k \cap \mathcal{D}_l = \emptyset$ for $k \neq l$. By drawing N_j samples x_{jk}^* from the normalised restricted density in each strata p_j an estimator is given by

$$\hat{I}^s(g) = \sum_{j=1}^{p} \frac{\rho_j}{N_j}\sum_{k=1}^{N_j} g(x_{jk}^*), \quad x_{jk}^* \sim p_j, \quad (5)$$

where ρ_j is the probability of region $\mathcal{D}_j$. This estimate is unbiased, $\text{E}[\hat{I}^s(g)] = I(g)$. Using proportional allocation, that is, $N_j = N\rho_j$, the integral (5) has the property that $\text{Var}\,\hat{I}^s(g) = N^{-1}\text{E}[\text{Var}_{P_j}\,g] \leq N^{-1}x\,\text{Var}_{P_N}\,g$, see e.g., [7, 10]. Hence, the variance of (5) does not increase compared to (3). On the contrary, a decrease in variance is quite possible.

3.3. Theory of uniform distributions

The theory of uniform distributions, see for instance [11], provides an intuitive method. This theory is based on the Koksma-Hlawka inequality [12]

$$|\hat{I}(g) - I(g)| \leq D_N^*(u_1,\ldots,u_N)V_{HK}(g), \quad (6)$$

which separates the effects of the random numbers u_k from that of the function g. Here $V_{HK}(g)$ is the total variation of g in the sense of Hardy and Krause. Note that V_{HK} only depends on g. The star discrepancy D_N^* is defined as

$$D_N^*(\{u_i\}) = \sup_{a \in [0,1)^d} \left| \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{(0,a]}(u_i) - |[0,a)| \right| \qquad (7)$$

where $\mathbb{1}$ is the indicator function defined by

$$\mathbb{1}_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases} \qquad (8)$$

and $| \cdot |$ denotes volume. The star discrepancy D_N^* is a numerical measure of how uniform a set of points is distributed in the unit cube. It compares the fraction of points in a box to the volume of this box. Clearly this difference will be smaller when the points are more uniform distributed. Now, the Koksma-Hlawka inequality (6) relates a smaller discrepancy to better integration accuracy, implying that more uniform samples have better integrating properties.

3.4. Discussion

The resampling algorithms discussed in Section 2 differ in which of the methods discussed above they apply. *Multinomial resampling* is the basic approach of (3). *Stratified resampling* applies, as its name implies, stratification. More precisely, the interval $[0,1)$ is partitioned into N regions from which one sample is drawn. This partitioning results in a variance reduction. An alternative explanation for this improved quality is provided by inspecting the 'uniformity' of the samples. As illustrated by Fig. 1 the samples are more uniformly distributed for stratified resampling than for multinomial resampling. Hence, the quality is improved. This view has been extended further in *systematic resampling* which has the lowest possible discrepancy. Systematic resampling can be interpreted by set reduction as well: a line segment of length ℓ always contains $\lfloor N\ell \rfloor$ points placed a distance N^{-1} apart and at most one point more. Hence, systematic resampling restricts the set values of n_i from $\{0, 1, \ldots, N\}$ to $\{\lfloor Nw_i \rfloor, \lfloor Nw_i \rfloor + 1\}$. Due to the fact that systematic resampling produces its samples dependently is it hard to conduct a proper variance analysis of the algorithm. An artificial example showing an increased variance is given in [7]. *Residual resampling* uses set restriction to improve the variance. The probability space is modified in such a way that now $n_i \in \{\lfloor Nw_i \rfloor \ldots, N\} \subset \{0, 1, \ldots, N\}$.

The previous theoretical analysis shows that the resampling quality can be improved by using a different algorithm than multinomial resampling. Variance results confirm that residual and stratified resampling have lower variances. Although not confirmed by a variance analysis, systematic resampling is better than stratified resampling as it has the lowest discrepancy.
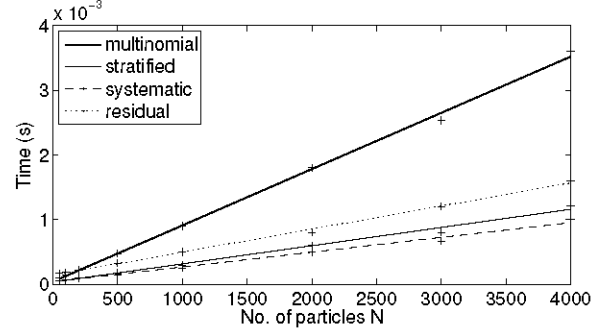

Fig. 2. Computational effort as a function of the number of particles.

4. COMPUTATIONAL COMPLEXITY

The multinomial, stratified and systematic resampling algorithms are very similar. They only differ in the how the ordered sequence of numbers is generated. All the algorithms are $\mathcal{O}(N)$, which impies that it is sufficient to compare them based on the complexity of the operations for one element. Since fractional power and random number generation are more complex operations than addition/subtraction or multiplication/division, multinomial resampling is the most expensive resampling algorithm, followed by stratified resampling and finally systematic resampling.

Residual resampling is more difficult to place. Experiments show that approximately $N/2$ particles are determined deterministically, leaving the other half to be determined using one of the algorithms discussed before. This complexity reduction is cancelled by the recalculation of the weights and other preparations. Hence, simulations have to point out which position residual resampling has.

5. SIMULATIONS

In the previous sections four resampling algorithms are reviewed and, based on the briefly reviewed theoretical framework, a comparison is made in terms of resampling quality and computational complexity. The results of this comparison are validated using simulations.

The computational complexity of the algorithms is investigated by measuring the time required to perform resampling of a random weight sequence. Fig. 2 shows the measured times each resampling algorithm requires. Thus, to reduce computational complexity, stratified resampling and systematic resampling are favourable, where the latter is slightly better.

The multinomial likelihood function [13] is given by

$$P(N_1 = n_1, \ldots, N_n = n_n) = \frac{N!}{n_1! \cdots n_n!} w_1^{n_1} \cdots w_n^{n_n}, \quad (9)$$

where $\sum_{i=1}^{n} n_i = N$. It attains its global maximum at $n_i = Nw_i$ for $i = 1, \ldots, N$. Resampling algorithms applying set reduction will on average have their n_i closer to Nw_i. Hence, their average likelihood values will also be higher. Fig. 3
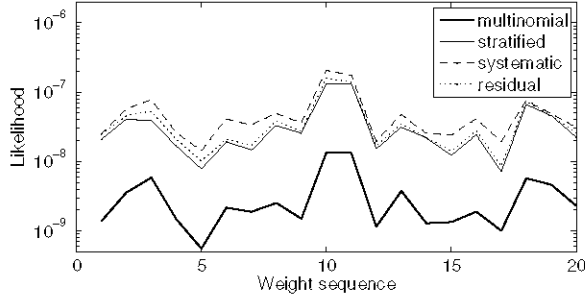

Fig. 3. The mean value of the multinomial likelihood over 100 simulations is shown for 20 random weight sequences of 20 particles.

shows effects of the resampling algorithms on the mean likelihood value. Similar results are observed using sequences of 10, 40 or 80 particles (not shown). Multinomial resampling has always the lowest likelihood value, illustrating the presence of set reduction with the other resampling algorithms.

The effects of the resampling algorithms on the root mean square error (RMSE) has been investigated by considering the following 2D tracking model of an aircraft

$$x_{t+1} = \begin{bmatrix} 1 & 0 & T & 0 & T^2/2 & 0 \\ 0 & 1 & 0 & T & 0 & T^2/2 \\ 0 & 0 & 1 & 0 & T & 0 \\ 0 & 0 & 0 & 1 & 0 & T \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} x_t + n_t, \qquad (10a)$$

$$y_t = \begin{bmatrix} \sqrt{p_x^2 + p_y^2} \\ \arctan(p_y/p_x) \end{bmatrix} + \nu_t \qquad (10b)$$

where the state, $x = \begin{bmatrix} p_x & p_y & v_x & v_y & a_x & a_y \end{bmatrix}^T$, contains position, velocity and acceleration. Range and bearing are measured and $n_t \sim N(0, Q)$ and $\nu_t \sim N(0, R)$ are mutually independent Gaussian noise sequences. The simulation parameters are given in Table 1. The simulations show that the

Table 1. Simulation parameters

Parameter	Value	Description
T	1	sample time
x_0	$[2000, 2000, 20, 20, 0, 0]^T$	initial position
P_0	$\mathrm{diag}[4, 4, 16, 16, 0.04, 0.04]$	$\mathrm{Cov}\, x_0$
Q	$\mathrm{diag}[4, 4, 4, 4, 0.01, 0.01]$	$\mathrm{Cov}\, n_t$
R	$\mathrm{diag}[100, 10^{-6}]$	$\mathrm{Cov}\, \nu_t$

estimates slightly change depending on which resampling algorithm is used, shown in Fig. 4. However, the effects are not significant.
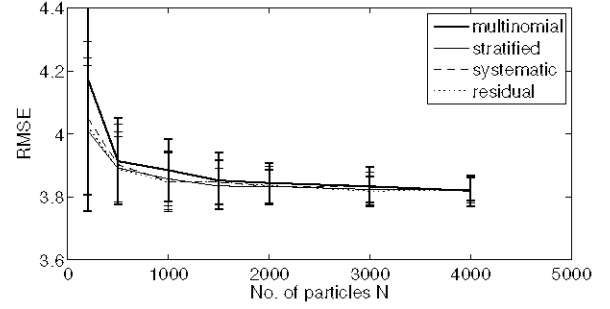

Fig. 4. Average RMSE values for velocity and their standard deviations as a function of the number of particles for the aircraft model.

6. CONCLUSIONS

Considering resampling quality and computational complexity, strafied and systematic resampling are favourable over multinomial resampling. They reduce the computational complexity while giving identical or perhaps slightly improved particle filter estimates. Aditionally, from a uniform distribution perspective systematic resampling is theoretically superior.

7. REFERENCES

[1] A. Doucet, N. de Freitas, and N. Gordon, Eds., *Sequential Monte Carlo methods in practice*, Springer Verlag, New York, 2001.

[2] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *IEE Proceedings-F*, vol. 140, no. 2, pp. 107–113, 1993.

[3] G. Kitagawa, "Monte Carlo filter and smoother for non-Gaussian nonlinear state space models," *Journal of Computational and Graphical Statistics*, vol. 5, no. 1, pp. 1–25, 1996.

[4] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Proceedings on Signal Processing*, vol. 50, no. 2, pp. 174–188, 2002.

[5] J. S. Liu and R. Chen, "Sequential Monte Carlo methods for dynamic systems," *Journal of the American Statistical Association*, vol. 93, no. 443, pp. 1032–1044, 1998.

[6] D. Crisan and A. Doucet, "A survey of convergence results on particle filtering methods for practitioners," *IEEE Transactions on Signal Processing*, vol. 50, no. 3, pp. 736–746, 2002.

[7] R. Douc, O. Cappé, and E. Moulines, "Comparison of resampling schemes for particle filtering," in *Proc. of Image and Signal Processing and Analysis (ISPA)*, Sept. 2005, pp. 64–69.

[8] C. P. Robert and G. Casella, *Monte Carlo statistical methods*, Springer, New York, 1999.

[9] W. G. Cochran, *Sampling Techniques*, John Wiley and Sons, New York, 3rd edition, 1977.

[10] J. D. Hol, "Resampling in particle filters," Internship report, Linköping University, 2004, LiTH-ISY-EX-ET-0283-2004.

[11] A. B. Owen, "Monte Carlo variance of scrambled net quadrature," *SIAM Journal on Numerical Analysis*, vol. 34, no. 5, pp. 1884–1910, 1997.

[12] E. Hlawka, "Funktionen von beschränkter Variation in der Theorie der Gleichverteilung," *Annali di Mathematica Pura ed Applicata*, vol. 54, pp. 325–333, 1961.

[13] W. W. Hines and D. C. Montgomery, *Probability and statistics in engineering and management science*, John Wiley & Sons, New York, 1972.