Project - Stochastic processes : estimation and prediction

Eliott Van Dieren - 09731900
Nicolas Mil-Homens Cavaco - 03721900

1 Empirical probability density function (PDF)

a) We consider the Lorenz system

$$\begin{cases} \dot{x} &= \sigma(y-x) \\ \dot{y} &= x(\rho-z)-y \\ \dot{z} &= xy-\beta z \end{cases}$$

in the case $\sigma = 10$, $\rho = 28$ and $\beta = \frac{8}{3}$. We simulate the system in time interval $[0, t_{end}]$ for $t_{end} = 100$ $[s]$.

Let $\Omega = [-20, 20] \times [-30, 30] \times [0, 50]$ the study domain. Let $\hat{\Omega} = \bigcup_{e=1}^{N} \hat{\Omega}_e$ the discretized study domain composed of $N = 960$ boxes $\hat{\Omega}_e$ of size $5 \times 5 \times 5$. Figure 1 shows projections of particle positions in xy, yz and xz planes for times $t \in [0, t_{end}]$. Colors represent the probability that the particle is located in a given box at a given time t. This probability was obtained by computing the number of time the particle was inside a given box during the simulation of the system, and then dividing the result by the duration of simulation. In other terms, if we denote by $\underset{\sim}{\boldsymbol{x}} = (x, y, z)$ and $\underset{\sim}{\boldsymbol{x}}_t$ the position of the particle at time t, the empirical pdf is then defined by

$$\hat{p}(\underset{\sim}{\boldsymbol{x}}) := \sum_{e=1}^{N} \omega(\hat{\Omega}_e)\ \delta_{\hat{\Omega}_e}(\underset{\sim}{\boldsymbol{x}}) \approx p(\underset{\sim}{\boldsymbol{x}}); \qquad \text{with} \quad \omega(\hat{\Omega}_e) = \frac{1}{t_{end}} \sum_{t=1}^{t_{end}} \delta_{\hat{\Omega}_e}(\underset{\sim}{\boldsymbol{x}}_t) \text{ and } \delta_{\Omega}(\underset{\sim}{\boldsymbol{x}}) = \begin{cases} 1 & \text{if} \quad \underset{\sim}{\boldsymbol{x}} \in \Omega \\ 0 & \text{else} \end{cases}$$

The pdf projected along the 3 axis in the form of colored matrices has been plotted in the right part of Figure 1. Each one corresponds to a projected view of one box. Firstly, we observe that the pdf takes its maximum values in the boxes $\hat{\Omega}_{e_1} = [0, 5] \times [0, 5] \times [15, 20]$ and $\hat{\Omega}_{e_2} = [0, 5] \times [0, 5] \times [20, 25]$ (in yellow) and that it is "continuous" in the sens that, around these maxima, the pdf decreases progressively (it would not have been the case if the particle could teleport). Secondly, we see that the empirical pdf does not detect correctly the holes (the 2 orbits) in the trajectory when they are smaller than the size of boxes or when they are recovered by 2 adjacent boxes. The selected discretization is thus quite inaccurate in this sens. However, the general look of the trajectory remains correct.

b) Let $X \equiv$ Set of probability distribution on a probability space $\mathcal{E}$. Let P, $Q \in X$, 2 distributions.

We want to calculate the distance between P and Q. As a reminder, a distance (also called metric) on X is a function $\Delta : X \times X \to [0, +\infty[$ which satisfies the following properties $\forall P, Q \in X$:

$$\Delta(P, Q) = 0 \Leftrightarrow P = Q; \quad \Delta(P, Q) = \Delta(Q, P); \quad \forall R \in X : \Delta(P, Q) \leq \Delta(P, R) + \Delta(R, Q).$$

A statistical distance is a more general concept since it does not need to satisfy all the properties of a metric as defined above. In what follows, we will therefore distinguish between the concepts of (statistical) distance and metric.

F-divergence

In order to quantify the similarity between 2 given distributions P and Q, we define f-divergence in the sens of measure theory as

$$D_f(P||Q) = \int_{\Omega} f\left(\frac{dP}{dQ}\right)\ dQ$$

where

- $\frac{dP}{dQ}$ designates the Randon-Nikodyn derivative. This one exists if P is absolutely continuous [1] in respect to Q.
- f is a convex function [2] such that $f(1) = 0$. Therefore, $D_f(P||Q)$ is zero when $P = Q$ and increases when P and Q "move away".

1. cfr. LINMA 1315 - Complément d'analyse for undefined concepts in this document.
2. idem 1.

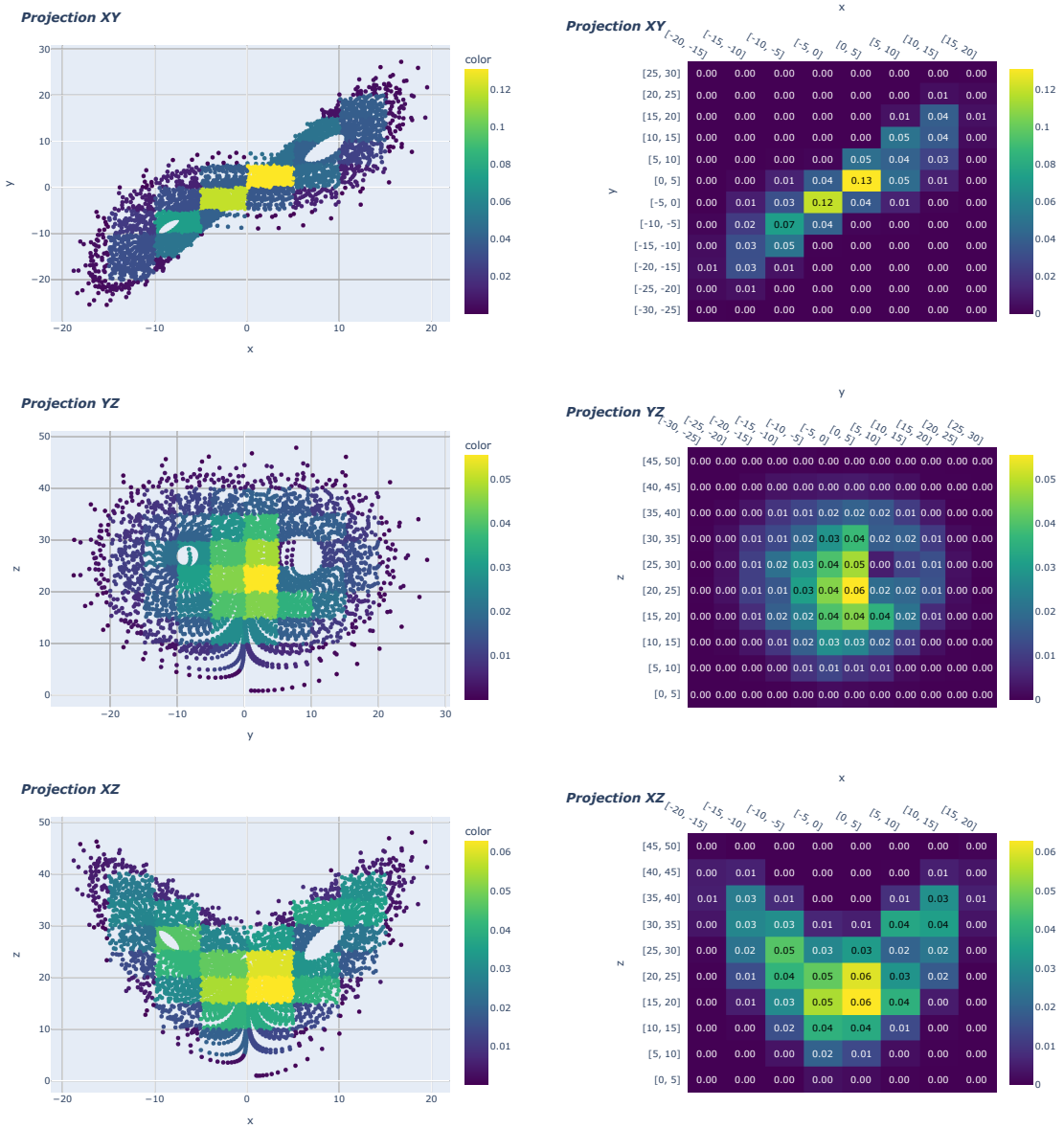

FIGURE 1 – Projections of particle positions along the 3 axes for $t \in [0, t_{end}]$ (left column) and the empirical pdf $\hat{p}$ (right column).

If P, Q are absolutely continuous in respect to a measure [3] μ of $\mathcal{E}$, then

$$D_f(P||Q) = \int_\Omega f\left(\frac{p(x)}{q(x)}\right) q(x)\ d\mu(x)$$

with p and q the probability density function associated to P and Q respectively.
It exists a great variety of f-divergences. We retain here 3 of them :

- The KL-divergence $D_{KL}(P||Q) : KL(t) = t\log(t)$
- The square Hellinger distance $D_H(P||Q) : H(t) = \frac{1}{2}(\sqrt{t}-1)^2$
- The Jensen-Shannon divergence $D_{JS}(P||Q) = \frac{1}{2}D_{KL}(P||R) + \frac{1}{2}D_{KL}(Q||R)$ with $R = \frac{1}{2}(P+Q)$

3. for instance, Lebesgue measure $\lambda^n(.)$ if $\mathcal{E} = \mathbb{R}^n$ or counting measure $|.|$ if $\mathcal{E} = \mathbb{N}^n$, for which integrals become sums.

Bhattacharyya coefficient

We call Bhattacharyya coefficient the quantity $BC(P,Q) := \int_\Omega \sqrt{p(x)q(x)}\, d\mu(x)$. It is interpreted as a measure of the degree of superposition of P and Q, and defines a kind of orthogonality between distributions. It satisfies :

$$0 \leq BC(P,Q) \leq 1; \quad BC(P,Q) = 1 \Leftrightarrow P = Q; \quad BC(P,Q) = 0 \Leftrightarrow P \perp Q$$

Hellinger distance (Δ_H)

The Hellinger distance is defined as $\Delta_H(P,Q) := \sqrt{D_H(P||Q)}$. This distance is a metric and is related to Bhattacharyya coefficient and to usual Euclidean norm :

$$\Delta_H(P,Q) = \sqrt{1 - BC(P,Q)} = \frac{1}{\sqrt{2}} \left|\left| \sqrt{P} - \sqrt{Q} \right|\right|_2$$

Jensen-Shannon distance (Δ_{JS})

The Jensen-Shannon distance is defined by $\Delta_{JS}(P,Q) := \sqrt{D_{JS}(P,Q)}$. This one consists to make KL-divergence symmetrical and ensure that its square root defines a metric on X. We note that contrary to Hellinger distance, Δ_{JS} necessarily involves the ratios $p/(p+q)$ and $q/(p+q)$. By convention, we consider that these ratios are equal to 1 when $p = q = 0$.

Our choice to measure distance between P and Q.

In conclusion, for the first part of this project, we will work with Δ_H and Δ_{JS}. In fact, for our problem, it is important to work with symmetrical distance criterons since P and Q should have the same status. For instance, KL-divergence is not symmetrical and hence is generally used when we want to compare a given distribution Q in respect to a reference distribution P. It is not our case here. However KL-divergence will be more revelant for the second part of the project since we will compare an estimated distribution with the real one. Furthermore, working with metrics is more intuitive in the sens that they satisfy the triangular inequality and thus are closer to the Euclidean metric. Finally, we notice that Δ_H and Δ_{JS} take both their value between 0 and 1. Consequently, we can interpret them as "percentages" of closeness between distributions.

c) The results obtained with parameters $\sigma = 5, \rho = 18$ et $\beta = 8$ are shown in figure 2. We observe that trajectory is quite different compared to the case studied at point a). Statistical indicators [4] leads us to the same conclusion for pdf. (see table of figure 4).

Others cases were studied but given that the space in this document is limited, we could not show the plots. However, we analyze how the trajectory changes when one parameter varies :

- Variation of β : when $\beta = 0.7$, the trajectory as the same appearance than in the case shown in figure 1 but the "butterfly's wings" are much thinner. For $\beta = 1.0, 2.0$, the plots have exactly the same look than in question 1. a). The holes becomes smaller when β increases. Finally, for $\beta = 4.0$, the particle seems to be attracted by an equilibrium point around which it rotates and go closer and closer.
- Variation of ρ : When $\rho < 18$, the plots look like those plotted on the figure 2. For $\rho = 30.0$, the right hole becomes much greater while the left one fills in, compared witt the figure 1. When $\rho = 50.$, every holes are filled in and when $\rho = 100$, it seems that the particle left the study domain.
- Variation of σ : When $\sigma = 5$, we observe the same phenomenon than when $\beta = 4.0$. For $\sigma = 15$, the right holes opens widely.

d) The results obtained with parameters initial conditions $(x_0, y_0, z_0) = (10, 10, 10)$ are shown in figure 3. We observe that trajectory is quite similar compared to the case studied at point a). Statistical indicators leads us to the same conclusion for pdf. (see table of figure 4).

As previously, we discuss briefly what we observe when initial conditions change. We test when x_0 goes from -20 to 10 (from left to right). The maximum of the pdf moves from right to the left along de x-axis. For variations of y_0 and z_0, we observe the same phenomenon but their is no straightforward conclusion that can be drawn as for x_0. Furthermore, we study the impact of small variation of amount 0.1 around the initial position $(1, 1, 1)$. Despite the system is chaotic, we see that the pdf hardly change. This shows that studying this system in a probabilistic way with pdf is more revelant than trying to determine the particle position at each time in a deterministic way.

4. We note that chosen indicators are symmetrical (excepted KL-divergence which is not used), which means that they will give the same result whether they are evaluated for (P_a, P_c) or (P_c, P_a).

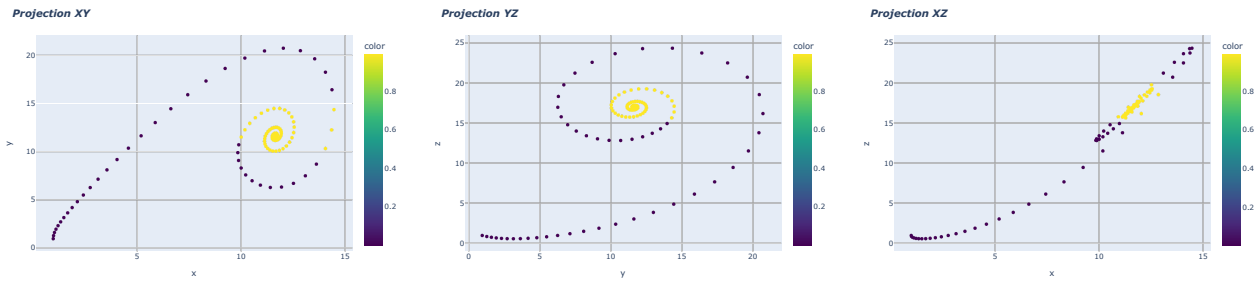

FIGURE 2 – Trajectory of the particle and pdf (see colors) for the parameters $\sigma = 5, \rho = 18$ et $\beta = 8$.

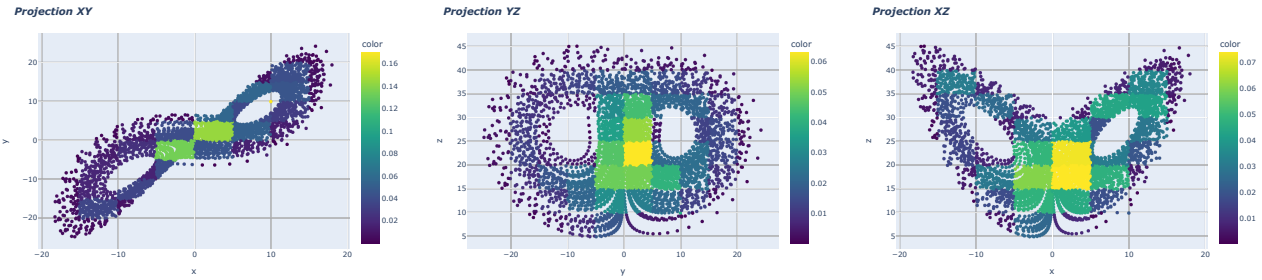

FIGURE 3 – Trajectory of the particle and pdf (see colors) for the initial conditions $(x_0, y_0, z_0) = (10, 10, 10)$.

	BC	Δ_H	Δ_{JS}
(P_a, P_c)	0.011	0.994	0.829
(P_a, P_d)	0.956	0.209	0.187

FIGURE 4 – Statistical distances comparing P_a with P_c (first line) and P_a with P_d (second line), where P_i designates probability distribution at point $i \in \{a, c, d\}$. Bhattacharyya coefficient is given for information.

2 Particle Filter : Resampling

In this section, we will compare and analyse different resampling methods, so we assume that the weights $\tilde{\omega}_i^{(t)}, i \in \{1, \dots, M\}$, where M is the number of observed particles, are known. N is the number of resampled particles. Let $Q_t^{(i)} = \sum_{k=1}^{i} \tilde{\omega}_t^{(k)}$ and $Q_t^{(-1)} = 0$, the cumulative sums from the first i weights

2.1 Multinomial Resampling

As we know that $N_t^{(i)} \sim \text{Mult}(N, \tilde{\omega}_t^{(i)})$, variance of multinomial resampling method is given by :

$$\text{Var}(N_t^{(i)}) = N\tilde{\omega}_t^{(i)}(1 - \tilde{\omega}_t^{(i)}).$$

In practice, we use the bootstrap method in order to sample a multinomial random variable. The principle is to generate a uniform r.v. $u_t \sim U[0, 1)$ and apply the inverse CDF of $N_t^{(i)}$ to it. Equivalently, we generate N random numbers $u_t^{(n)} \sim U[0, 1)$ and for the n^{th} selection of our resampling, the particle $x_t^{(i)}$ is chosen if $Q_t^{(i-1)} \leq u_t^{(n)} \leq Q_t^{(i)}$.

The lower and upper bonds for $N_t^{(i)}$ are respectively 0 (not selected at all) and N which means that $\tilde{x}_t^{(i)}$ would be the only selected particle. Furthermore, the time complexity of the basic algorithm is $\mathcal{O}(N \log M)$ where N is the number of resampled particles and M is the number of available particles. This complexity comes from the use of binary search $(\log M)$ to find i from the condition $Q_t^{(i-1)} \leq u_t^{(n)} \leq Q_t^{(i)}$. In order to use the binary search, we have to assume that all $\tilde{\omega}_t^{(i)}$ are positive so that $Q_t^{(i-1)} \leq Q_t^{(i)}$. As we are doing this computation N times, it gives us the complexity stated above. Note that there are ways of reducing this time complexity to $\mathcal{O}(N)$ by using the inverse CDF the normalized particle weights but it will not be discussed in this report.

2.2 Residual Resampling

The residual resampling method can be split in two distinct parts. The first part consists of replicating $\tilde{x}_t^{(i)}$, $\lfloor N\tilde{\omega}_t^{(i)} \rfloor$ times. For the remaining places available, $\hat{N}_t = N - \sum_{i=1}^{N} \lfloor N\tilde{\omega}_t^{(i)} \rfloor$, we will compute the residual weights $\hat{w}_t^{(i)}$ and reuse the Multinomial resampling method with parameters $\hat{N}_t$ and $\hat{w}_t^{(i)}$. The variance is thus given by :

$$\text{Var}(N_t^{(i)}) = \hat{N}_t \hat{w}_t^{(i)}(1 - \hat{w}_t^{(i)}) \quad \text{where } \hat{w}_t^{(i)} = \frac{N\tilde{\omega}_t^{(i)} - \lfloor N\tilde{\omega}_t^{(i)} \rfloor}{N - \sum_{j=1}^{N} \lfloor N\tilde{\omega}_t^{(j)} \rfloor}.$$

The time complexity of this method will be $\mathcal{O}(M) + \mathcal{O}(\hat{N}_t \log M)$ where the second term comes from the multinomial method on the residual part and can therefore also be reduced to $\mathcal{O}(\hat{N}_t)$ as stated before.

2.3 Systematic Resampling

For this method, we will sample an initial element $u_1 \sim U[0, 1/N)$ and then place $N-1$ elements $u_2, u_3, \ldots, u_N$ whose distances are exactly $1/N$ apart starting from u_1. Then, to compute $N_t^{(i)}$, we find the number of elements lying within the interval $\left[Q_t^{(i-1)}, Q_t^{(i)}\right], \forall i$. The expression of $\text{Var}(N_t^{(i)})$ is obtained by making the following observation : the length of each interval i is exactly equal to $\tilde{\omega}_t^{(i)} = \frac{\lfloor N\tilde{\omega}_t^{(i)} \rfloor}{N} + r_t^{(i)} \in \left[\frac{\lfloor N\tilde{\omega}_t^{(i)} \rfloor}{N}, \frac{\lfloor N\tilde{\omega}_t^{(i)} \rfloor}{N} + \frac{1}{N}\right)$, where $r_t^{(i)} = \frac{N\tilde{\omega}_t^{(i)} - \lfloor N\tilde{\omega}_t^{(i)} \rfloor}{N}$. Since $u_n = u_1 + \frac{n-1}{N} \in \left[\frac{n-1}{N}, \frac{n-1}{N} + \frac{1}{N}\right)$, the number of elements $N_t^{(i)}$ in the i^{th} interval is equal to $A = \lfloor N\tilde{\omega}_t^{(i)} \rfloor$ or $B = \lfloor N\tilde{\omega}_t^{(i)} \rfloor + 1$. This is a Brenoulli trial where the probability of a success B is $p_t^{(i)} = \frac{r_t^{(i)}}{\frac{1}{N}}$.

Therefore, $N_t^{(i)} \sim \text{Bernoulli}(p_t^{(i)})$ and closed form of $\text{Var}(N_t^{(i)})$ is :

$$\text{Var}(N_t^{(i)}) = N r_t^{(i)} \left(1 - N r_t^{(i)}\right) = N \left(\tilde{\omega}_t^{(i)} - \frac{\lfloor N\tilde{\omega}_t^{(i)} \rfloor}{N}\right) \left(1 - N\tilde{\omega}_t^{(i)} + \lfloor N\tilde{\omega}_t^{(i)} \rfloor\right)$$

The time complexity of this method will be $\mathcal{O}(N)$ as we only have to loop once on each element $Q_t^{(i)}$ as all $u_t^{(i)}$'s are computed in advance.

2.4 Comparison of performances

Firstly, we compare the variance of each method as follow :

$$\begin{aligned}
\text{Var}(N_{t,mult}^{(i)}) &= N\tilde{\omega}_t^{(i)}(1 - \tilde{\omega}_t^{(i)}) \\
\text{Var}(N_{t,res}^{(i)}) &= N \left(\tilde{\omega}_t^{(i)} - \frac{\lfloor N\tilde{\omega}_t^{(i)} \rfloor}{N}\right) \left(1 - \frac{\tilde{\omega}_t^{(i)} - \frac{\lfloor N\tilde{\omega}_t^{(i)} \rfloor}{N}}{1 - \sum_{j=1}^{N} \frac{\lfloor N\tilde{\omega}_t^{(j)} \rfloor}{N}}\right) \\
\text{Var}(N_{t,sys}^{(i)}) &= N \left(\tilde{\omega}_t^{(i)} - \frac{\lfloor N\tilde{\omega}_t^{(i)} \rfloor}{N}\right) \left(1 - N\left[\tilde{\omega}_t^{(i)} - \frac{\lfloor N\tilde{\omega}_t^{(i)} \rfloor}{N}\right]\right) \\
&= N\tilde{\omega}_t^{(i)}(1 - \tilde{\omega}_t^{(i)}) - \left(N(N-1)(\tilde{\omega}_t^{(i)})^2 - 2N\tilde{\omega}_t^{(i)} \lfloor N\tilde{\omega}_t^{(i)} \rfloor + \lfloor N\tilde{\omega}_t^{(i)} \rfloor (1 + \lfloor N\tilde{\omega}_t^{(i)} \rfloor)\right)
\end{aligned}$$

We observe directly that :

$$\begin{aligned}
&\text{Var}(N_{t,res}^{(i)}) \geq \text{Var}(N_{t,sys}^{(i)}) \quad \text{if } 1 - \sum_{j=1}^{N} \frac{\lfloor N\tilde{\omega}_t^{(j)} \rfloor}{N} \geq \frac{1}{N} \\
&\text{Var}(N_{t,mult}^{(i)}) \geq \text{Var}(N_{t,sys}^{(i)}) \quad \text{if } N(N-1)(\tilde{\omega}_t^{(i)})^2 - 2N\tilde{\omega}_t^{(i)} \lfloor N\tilde{\omega}_t^{(i)} \rfloor + \lfloor N\tilde{\omega}_t^{(i)} \rfloor (1 + \lfloor N\tilde{\omega}_t^{(i)} \rfloor) \geq 0
\end{aligned}$$

Empirically, by drawing N weights $\sim U[0, 1)$ and by normalizing it, we observe that the variance does not particularly grow w.r.t N but we can observe that, for uniformly distributed weights, the multinomial variance was greater and more volatile w.r.t N than the two others, as seen on Figure 6. Please note that the analysis was made on $N_t^{(1)}$ but that for other $N_t^{(i)}$'s, the same pattern occurs but the value of the variances fluctuate depending which particle one chooses to pick. However, the time complexity did not match the theoretical expectations, as one could thought that multinomial

resampling would be less efficient than systematic and residual due to its complexity (see Figure 6). This can be due to a lot of factors : the usage of numpy functions for multinomial which might be better implemented than our algorithm for the systematic or residual, the machine on which we ran the code or the method of computation time. Lasly, we see on Figure 5 that our closed form variances matches the empirical variances pretty well. [5]

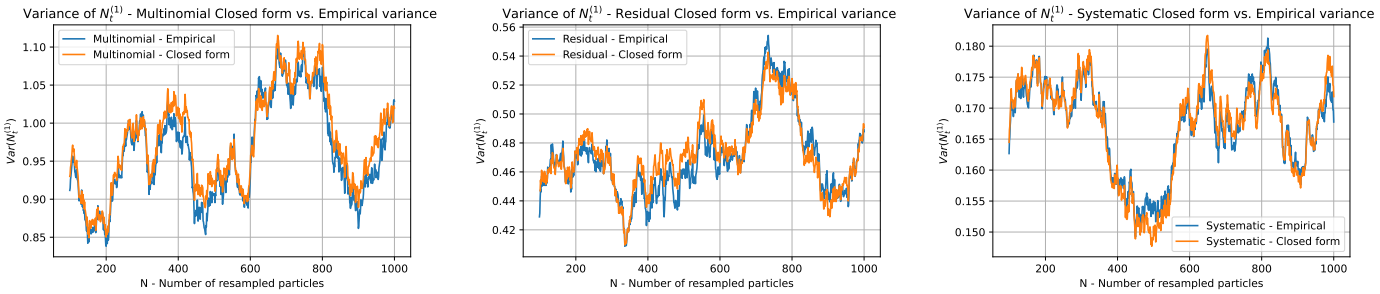

FIGURE 5 – Evolution of $N_t^{(1)}$ w.r.t N for empirical and closed form variances, weights uniformly distributed

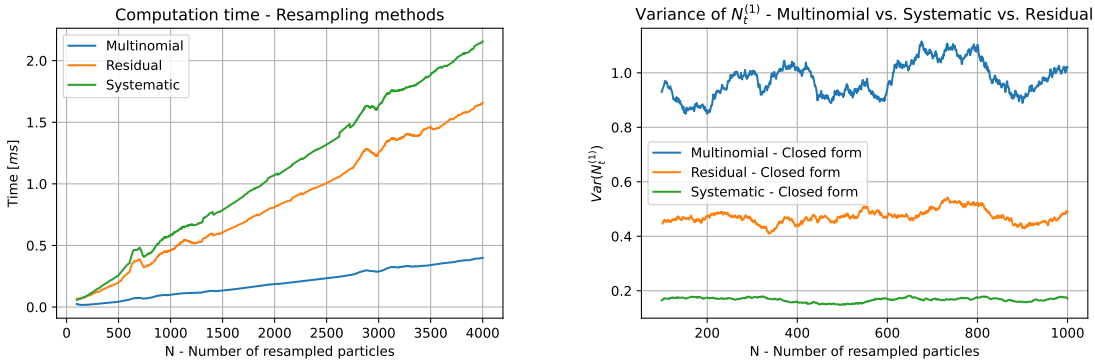

FIGURE 6 – Comparison of $\text{Var}(N_t^{(1)})$ from multiple resampling methods (closed form variance), weights uniformly distributed and time complexity

3 Particle Filter Implementation for the Lorenz System

3.1 Parameters discussion

In this subsection, we will discuss the various output with respect to certain parameters such as Σ (for the approximation error), the frequency of observations h and the different resampling methods described in the previous section. Let x_{mean}^t be the predicted trajectory which is the mean of the particles that we created at time t and x_{true}^t be the real $3D$ trajectory for a time t.

3.1.1 Σ variation

In order to analyze the Σ variation, we simulated the trajectories of the Lorenz system with $\Sigma \in \{0, 1, 10, 100\}$. As we can see on Figure 7, the euclidean distance between x_{mean} and x_{true} globally grows when Σ increases, as expected because there will be more variance in the approximation error v_t. However, when $\Sigma = 0$, we might have expected that the particle filter x_{mean} would exactly match x_{true} but in reality, this is only the case for the first time steps (here, when $t \leq 16s$). This can be shown on the euclidean distance graph below. After $16s$, the estimation seems diverging from x_{true}, as the euclidean distance grows to ≈ 40 and when comparing with $\Sigma = 1$, we see that this distance is way higher. When we look at much higher values of Σ, the filter can not track the x_{true} anymore and becomes completely useless which euclidean distances ≈ 80.

5. All the code related to this question are at the end of our notebook

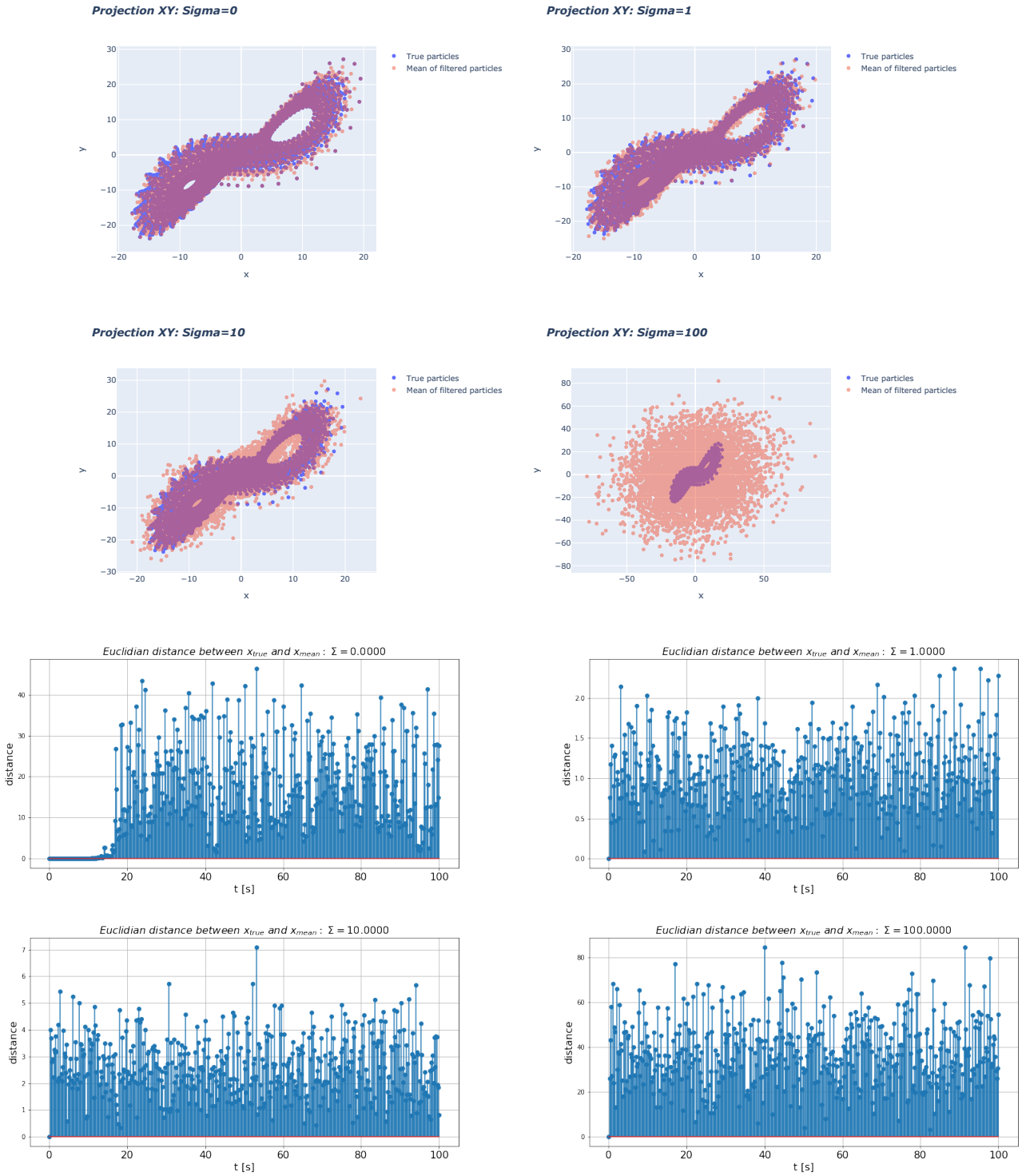

FIGURE 7 – Comparison of real and estimated XYZ trajectories and euclidean distance for $\Sigma = \{0, 1, 10, 100\}$

3.1.2 h variation

In order to analyze the h variation, we simulated the trajectories of the Lorenz system with $h \in \{0.008, 0.02, 0.08, 0.1, 0.15, 0.2\}$. As expected from theory, from a certain big enough time step, the Runge-Kutta Method does not converge anymore and therefore does not lead to insightful and relevant results if time step values $h \geq 0.15s$. On Figure 8 we observe that the trajectory is less precise as time step increases as expected and we see that the value of the euclidean distance between the real trajectory x_{true} and the estimated trajectory x_{mean} does not particularly changes with respect to $h \leq 0.15s$.

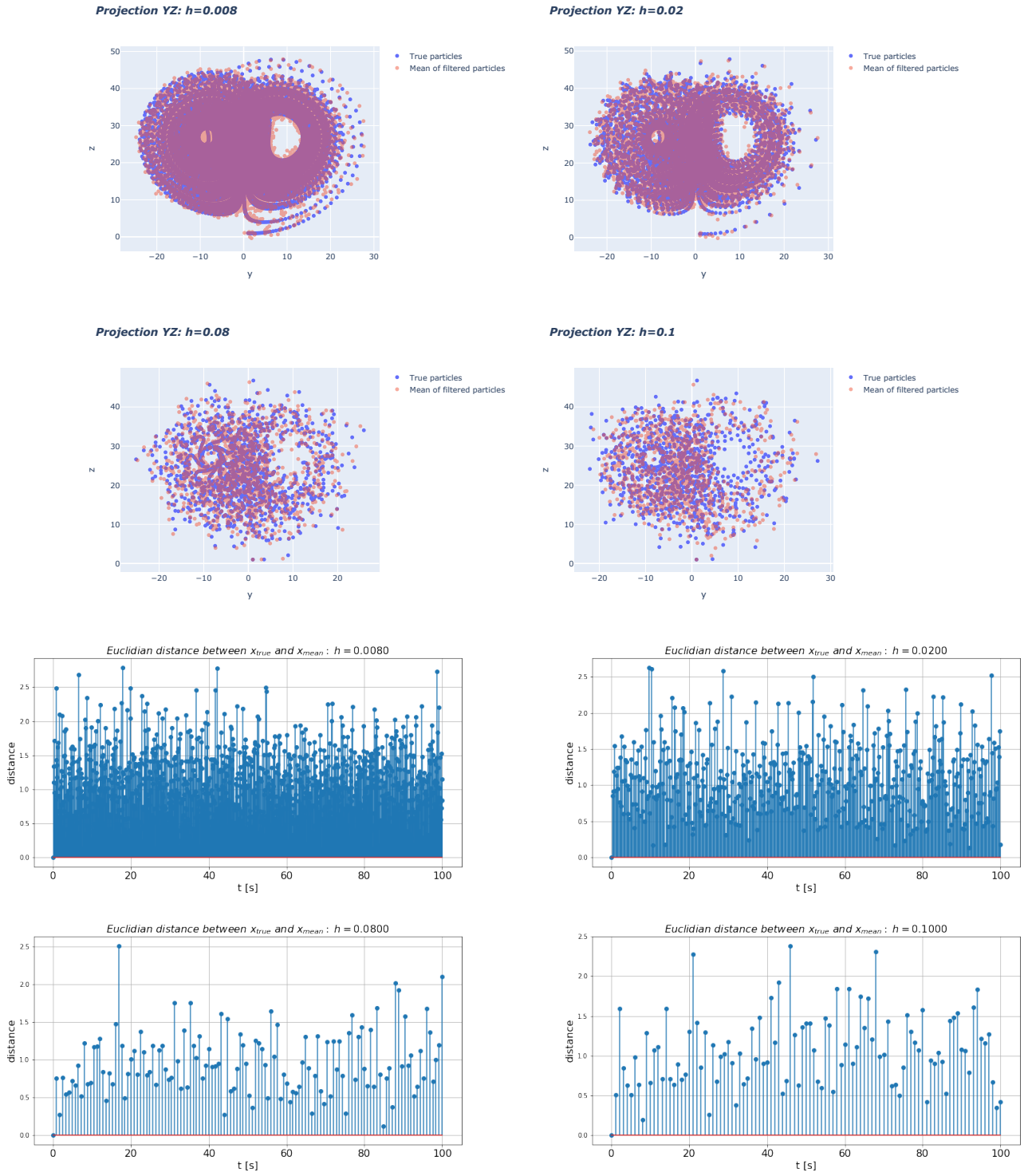

FIGURE 8 – Comparison of real and estimated XYZ trajectories and euclidean distance for $h = \{0.008, 0.02, 0.08, 0.1\}$

3.1.3 Resampling Methods

An interesting observation is that running several times particle filter algorithm for a given set of parameters does not give the same output. The reason is that the resampling of particles is not a deterministic operation. In order to compare resampling methods previously detailed, we simulate a large number of times the particle filter for each resampling method and we take the mean of total Euclidean distance between x_{mean} and x_{true}. Since the number of simulations is large, the mean should be a good criterion of comparison in order to reduce the randomness effect by computing $\sum_{t\in\{h,2h,\dots,t_{end}\}} ||x_{true}(t) - x_{mean}(t)||_2$. We conclude, as seen on Figure 9, that the multinomial resampling method has higher Euclidean distance than the two other resampling methods and that the systematic method has, most of the time,

better results than the residual resampling method.

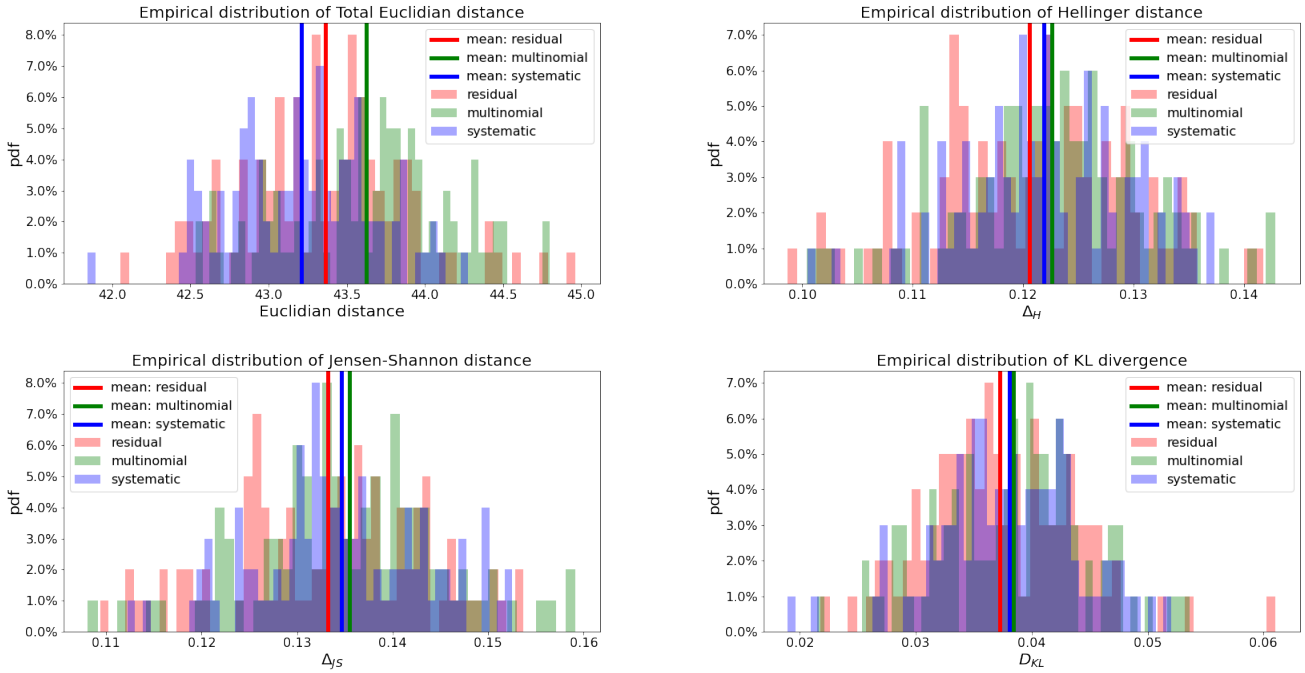

FIGURE 9 – Comparison of the resampling methods by using Euclidean and statistical distances between x_{mean} and x_{true} for the set of parameters : $\sigma = 10,\ \rho = 28,\ \beta = 8/3,\ (x_0, y_0, z_0) = (1, 1, 1), t_{end} = 10\ [s],\ h = 0.02\ [s],\ \Sigma = 1$.

Furthermore, as discussed in Section 1, we could also use statistical distances on empirical pdf of the real trajectory and the filtered one in order to see the distance on a global aspect of the trajectory, instead of the euclidean method which gives us the precision error at each time step. If we only look at the averages, figure 9 shows that Hellinger and Jensen-Shannon distances lead us to the same conclusion that the multinomial resampling method is a less effective resampling method. As for the residual and systematic methods, we see that the residual gives us better results with respect to statistical distances than the systematic method.

We can also compute the weight degeneracy and compare it with respect to the chosen resampling method, as seen on Figure 1. First, we observe that for each resampling method, certain weights are boosted and others are lowered, because we want that the sum of weights are equal to one. For the multinomial resampling method, we observe that most important weights have been boosted, except for $\tilde{\omega}_6^{(2)}$,$\tilde{\omega}_6^{(3)}$ where their importance has been lowered and for $\tilde{\omega}_6^{(48)}$,$\tilde{\omega}_6^{(49)}$ where it has been set to zero. For the residual resampling, we see that the importance of the highest initial weights are conserved (even boosted) and that the remaining weights are following the same pattern as the multinomial resampling, with boosting and nullifying some weights. Lastly, the systematic resampling method keeps more or less the same importance for all weights, except some which are set to zero ($\tilde{\omega}_6^{(9)}$).

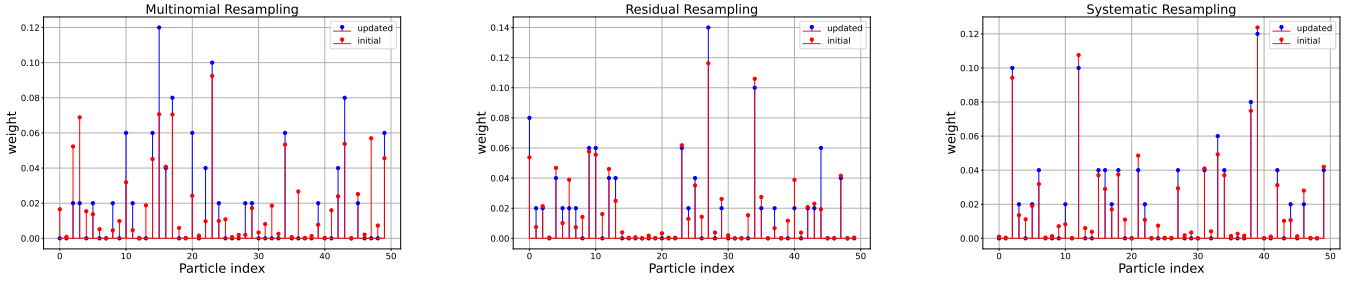

FIGURE 10 – Comparison of weights before and after resampling at $t = 6s$.

Bibliography :

- Sampling Theory| Chapter 11 | Systematic Sampling | Shalabh, IIT Kanpur
- Comparison of Resampling Schemes for Particle Filtering, Douc - Cappé - Moulines, Proceedings of the 4^{th} International Symposium on Image and Signal Processing and Analysis (2005)
- Resampling Methods for particle filtering, Li - Bolic - Djuric, IEEE Signal Magazine (2015)
- https ://en.wikipedia.org/wiki/Statistical_distance
- https ://en.wikipedia.org/wiki/F-divergence
- https ://encyclopediaofmath.org/wiki/Bhattacharyya_distance
- https ://en.wikipedia.org/wiki/Bhattacharyya_distance
- https ://en.wikipedia.org/wiki/Jensen%E2%80%93Shannon_diverence
- https ://en.wikipedia.org/wiki/Hellinger_distance
- https ://nbviewer.org/github/rlabbe/Kalman-and-Bayesian-Filters-in-Python/blob/master/12-Particle-Filters.ipynb