

LINMA1731 - Stochastic processes: estimation and prediction

Project : Tracking a Lorenz particle

1 - Theoretical Questions

If we have the Lorenz system defined by the following equations:

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

where x , y and z stand for the position coordinates of the point in the system and σ , ρ and β are positive real parameters. Lorenz used the values $\sigma = 10$, $\rho = 28$ and $\beta = \frac{8}{3}$ for which the system shows a chaotic behaviour.

1.1 - Empirical probability density function

Starting point at $(x, y, z) = (1, 1, 1)$ at time $t = 0$. The total time is $100s$ and we make step of time of $0.02s$. We have the following restriction for the domain of respectively x, y, z : $[-20, 20] \times [-30, 30] \times [0, 50]$.

a) To be able to compute the empirical PDF, we are going to divide the spatial domain into 3D cubic box of length $l = 5$. For exemple the interval x is then going to be : $x \in [-20, -15] \times [-15, -10] \times [-10, -5] \times [-5, 0] \times [0, 5] \times [5, 10] \times [10, 15] \times [15, 20]$. We will then have 8 intervals for x . We will then start from the point $(x_0, y_0, z_0) = (1, 1, 1)$ and for each step of time

$$t_k = k \times 0.02$$

with $k = 0, 1, 2, \dots, 5000$. We have the point

$$(x_k, y_k, z_k)$$

. The next step is computing in wich box is the point and increment the counter of the box. We will then compute the empirical PDF by dividing the number of points in each box ($N_{i,j,k}$) by the total number of points (N_{total}) iid :

$$P_{i,j,k} = \left(\frac{N_{i,j,k}}{N_{\text{total}}} \right)$$

If we project the empirical PDF on the different plan we obtain the following result :

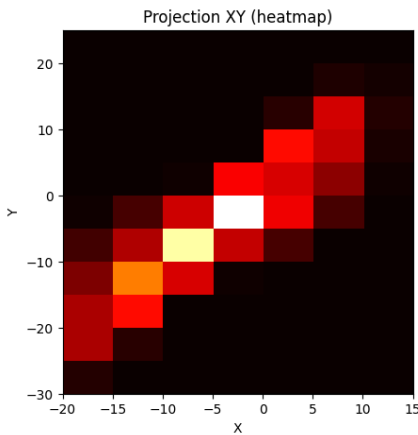


Fig. 1. – Plan XY

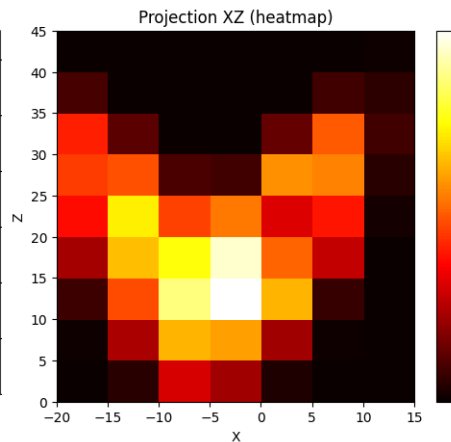


Fig. 2. – Plan XZ

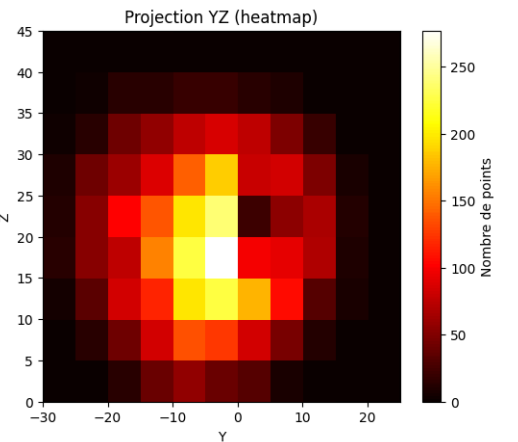


Fig. 3. – Plan XY

What we can see is that there is a majority of point in the middle of the domain. In the plan XY it's seems like we have a normal distribution. We can clearly observe the general shape of the system just by looking at the different plan.

b) To measure the distance between two empirical PDF, we made the choice to measure it by using the Kullback-Leibler divergence. The Kullback-Leibler divergence is a measure of how one probability distribution diverges from a second expected probability distribution. The Kullback-Leibler divergence is defined as follows¹ :

For two discrete probability distributions P and Q defined on the space X , the Kullback-Leibler divergence of P from Q is defined as:

$$D_{KL}(P\|Q) = \sum_{x \in X} P(x) \log \frac{P(x)}{Q(x)}$$

where $P(x)$ and $Q(x)$ are the probabilities of the event x according to the distributions P and Q , respectively.

The second choice of distance is the distance of Bhattacharyya. The Bhattacharyya distance is a measure of the similarity of two probability distributions. It is defined as the negative logarithm of the Bhattacharyya coefficient, which is defined as follows² :

For two discrete probability distributions P and Q defined on the same probability space X , the Bhattacharyya distance between P and Q is defined as:

$$D_B(P, Q) = -\log \left(\sum_{x \in X} \sqrt{P(x)Q(x)} \right)$$

where $P(x)$ and $Q(x)$ are the probabilities of the event x according to the distributions P and Q , respectively.

c) To measure the impact of the parameters σ, ρ, β on the distribution. We have made a function where we can observe the evolution of the system of Lorenz by adjusting the parameters.

Influence of parameter σ :

- **Small** : The system tends to have a more regular behaviour. After a certain value of σ the system starts to have stable behaviour.
- **Big** : The general behaviour of the system is chaotic. After a certain value of σ the system tends to have a special behaviour

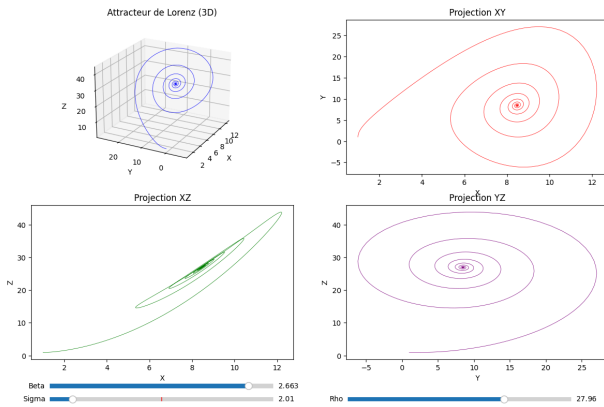


Fig. 4. – Small σ

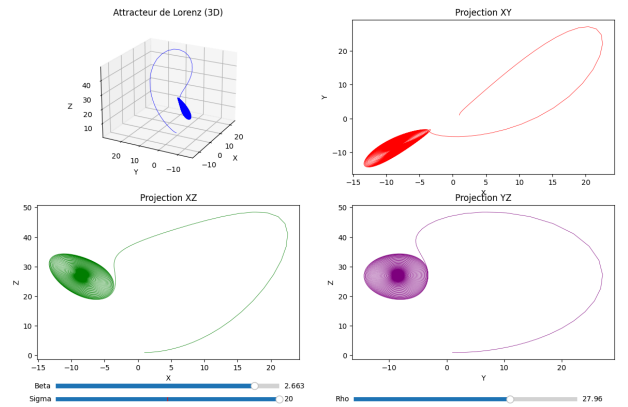


Fig. 5. – Big σ

Influence of parameter ρ :

- **Small** :

The system converges to a stable point. The system is not chaotic.

- **Middle** :

The system is oscillating between two points. The system is not chaotic.

¹https://fr.wikipedia.org/wiki/Divergence_de_Kullback-Leibler

²https://fr.wikipedia.org/wiki/Distance_de_Bhattacharyya

- **Big :**

The system is chaotic. We can not predict the future position of the system.

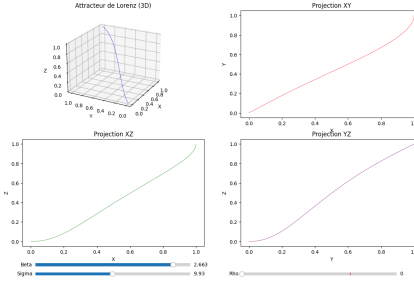


Fig. 6. – Small ρ

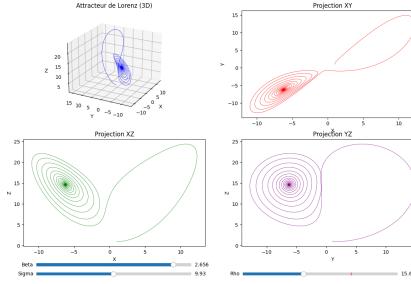


Fig. 7. – Middle ρ

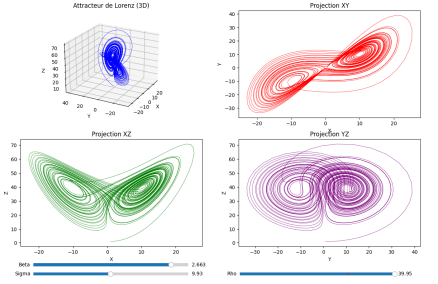


Fig. 8. – Big ρ

Influence of parameter β :

- **Small :**

The system tend to be less chaotic.

- **Big :**

The chaos can be amplified. The system is more chaotic.

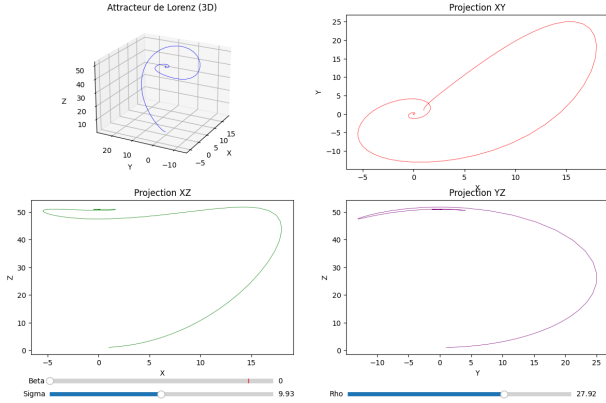


Fig. 9. – Small β

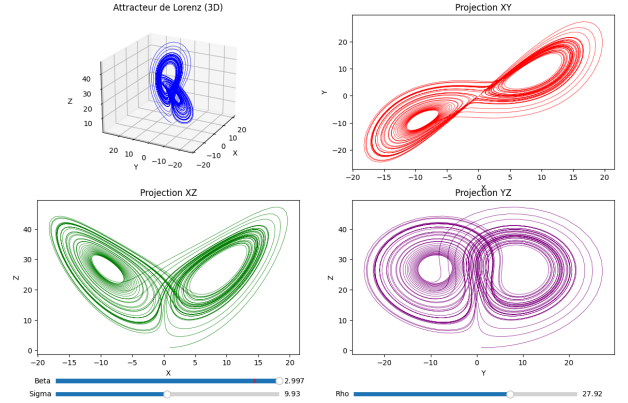


Fig. 10. – Big β

If we mesure the distance between our empirial PDF and a new one calculated with different paramters ($\sigma = 5, \rho = 20, \beta = 2$) we can see that the distances are not the same. In term of number we can see that the distance of KL divergence is 18.13 and the distance of Bhattacharyya is 4.51. The distance of KL divergence is bigger than the distance of Bhattacharyya. The distance of Bhattacharyya is smaller because she is not impacted as mush as the KL divergence by the presence of 0 in one distribution and not in the other one. As the twee distance are big with respect to the value that they can take, we can say that the two empirial PDF are not really close.

d) If we mesure the distance between our empirial PDF and the same one but with an other start point $((x_0, y_0, z_0) = (10, 10, 10))$ The distance is bigger for the Kullback-Leibler divergence than for the Bhattacharyya distance. In term of number we have a distance of 0.896 for the KL Divergence and a distance of 0.0448 for the Bhattacharyya distance. As the twee distance are small with respect to the value that they can take, we can say that the two empirial PDF are really close.

In conclusion for this part we can see that the initial state does not impact as mush as the paramter of the system. The distance between the two empirial PDF is smaller when we change the initial state than when we change the parameters of the system.

1.2 - Particle Filter

If we have a dynamic state space model :

$$\begin{cases} x_t = f(x_{t-1}, v_t) \\ y_t = g(x_t, w_t) \end{cases}$$

where $x_t \in \mathbb{R}, y_t \in \mathbb{R}$ and v_t, w_t are the noise. Our goal is to estimate x_t using observations $y_{1:t} = \{y_1, y_2, \dots, y_t\}$. Since we don't know exactly $p(x_{0:t}|y_{1:t})$, we are going to approximate it using a set of weighted samples. That bring some problems as over time, some particles have very low weights, while a few dominate. This leads to weight degeneracy, where most particles contribute very little to the estimate. To fix this, we perform resampling, which eliminates low-weight particles and duplicates high-weight ones. Here we are going to compare the performance of three type of resampling based on there complexity and variance. We are then going to compute $var(N_t^i)$

- Multinomial Resampling :

As the N_t^i are computed from a multinomial distribution with paramters N and \tilde{w}_t^i we have directly that the variance is given by : $var(N_t^i) = N * \tilde{w}_t^i * (1 - \tilde{w}_t^i)$.

- Residual Resampling :

The first syep is to compute the deterministic part, so for each particle i , we are going to compute a integer $\tilde{N}_t^i = \lfloor N \tilde{w}_t^i \rfloor$. By doing this we ensure that particle with a high weight gets a certain number of copies without any randomness. After having computed the deterministic part, there are still $\bar{N}_t = N - \sum_{i=1}^N \tilde{N}_t^i$ particles that don't have a weight. For these particles we are going to apply a procedure called SIR (Sampling Importance Resampling) wich give us the following new weights : $w_t'^{(i)} = \bar{N}_t^{-1} (\tilde{w}_t^{(i)} N - \tilde{N}_t^i)$ The total number N_t is then given by the sum of the deterministic part and the random part. As the deterministic part is deterministic, we have that the variance of the deterministic part is 0. The variance is then given by the variance of the random part which is given by : $var(N_t^{(i)}) = \bar{N}_t w_t'^{(i)} (1 - w_t'^{(i)})$ By doing some more computation we can see that the variance is smaller than the one of the multinomial resampling.

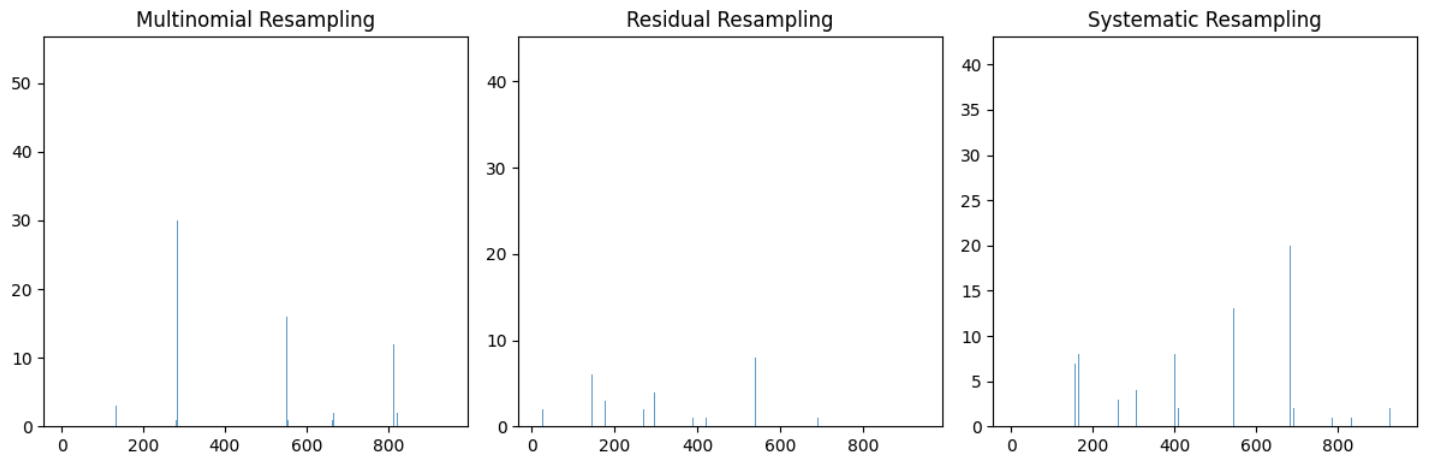
- Systematic Resampling :

If we define the cumulative sequence : $C_i = \sum_{j=1}^i \tilde{w}_t^{(j)}$ with $C_0 = 0$ and $C_M = 1$. We have that the cumulative sums partition the interval $[0, 1]$ into M subintervals: $(C_0, C_1], (C_1, C_2], \dots, (C_{M-1}, C_M]$. Each of these interval have a length of $\tilde{w}_t^{(i)}$. As we sample a set $U = \{u_1, u_2, \dots, u_N\}$ of N points in $[0, 1]$, if we count the number of points $N_t^{(i)}$ iid : $N_t^{(i)} = \#\{u_k \in U : C_{i-1} < u_k \leq C_i\}$ In a pure multinomial resampling scheme, the variance in the number of offspring $N_t^{(i)}$ would be : $var(N_t^{(i)}) = N * \tilde{w}_t^{(i)} (1 - \tilde{w}_t^{(i)})$.

However, because systematic or stratified sampling forces an even spread of points, the variance is further reduced. For this strategy, the variance becomes $var(N_t^{(i)}) = \bar{N}_t w_t'^{(i)} (1 - \bar{N}_t w_t'^{(i)})$.

Type of sampling	Variance	Time
Multinomial Resampling	3.6303	0.000070s
Residual Resampling	3.1031	0.000060s
Systematic Resampling	3.0735	0.000040s

By applying the three differents method of resampling we can see that the systematic resampling is clearly the best one in term of variance and time. To compute these data we have used our sample from point 1.2. We can also plot the different repartition of the particles after resampling.



It is much clear on a random set of weights :

