Statistical Analysis of Networks and Systems
Assignment 1

teacher: [redacted]
student: Bertille Temple

Delivered the 1st December, 2022
MIRI master, FIB, UPC

Contents

1 Multivariate Gaussian random variables

[redacted]

Figure 1: First statement

In the first case Σ is a diagonal matrix, which means that variables X1 and X2 are uncorrelated/independent.

1.1 Independent variables

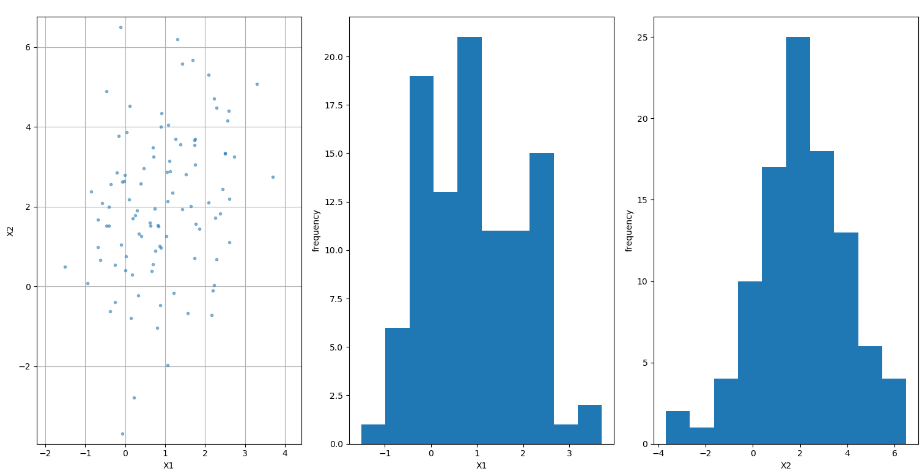

Figure 2: Multivariate gaussian variables with n = 100 and null covariance

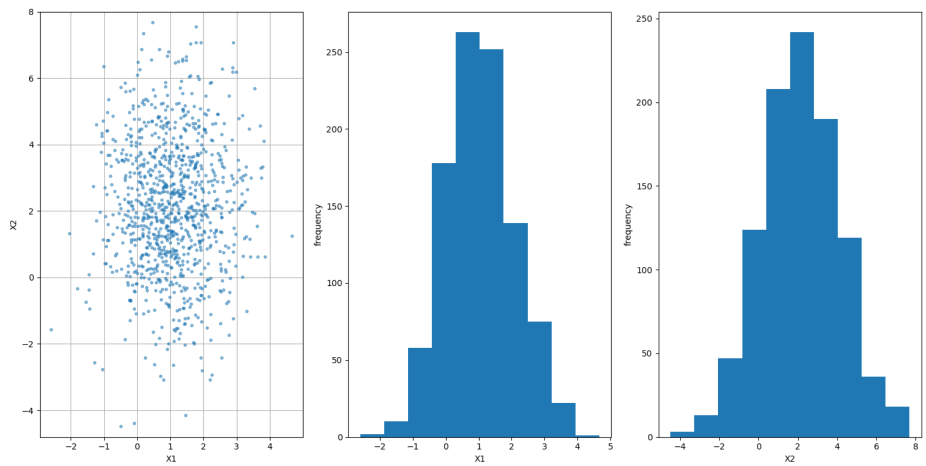

Figure 3: Multivariate gaussian variables with n = 1000 and null covariance

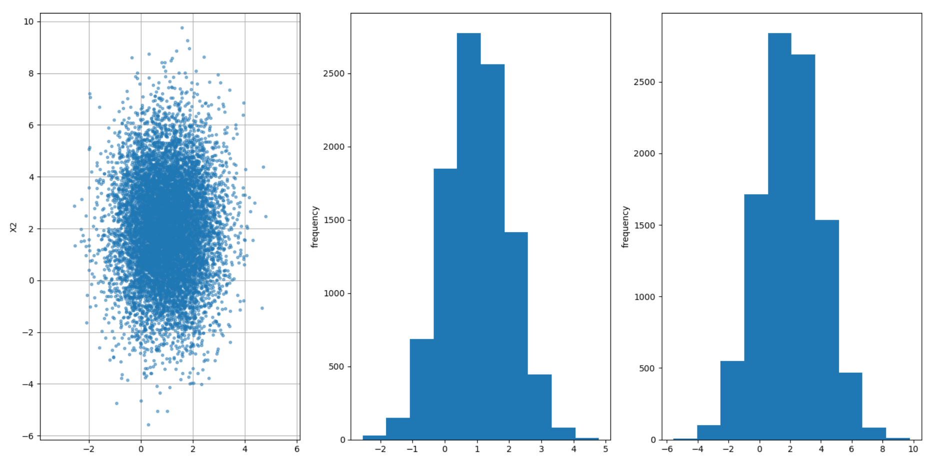

Figure 4: Multivariate gaussian variables with n = 10000 and null covariance

The histograms point out that, for all the samples sizes, both X1 and X2 follow a Normal distribution. We can explain that by the fact that a n-dimensional Gaussian with a diagonal covariance matrix is the same as a collection of n independent Gaussian random variables.

1.2 Negative covariances

Let us now generate a sample with the same amount of points using:

$$\Sigma = \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix} \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}$$

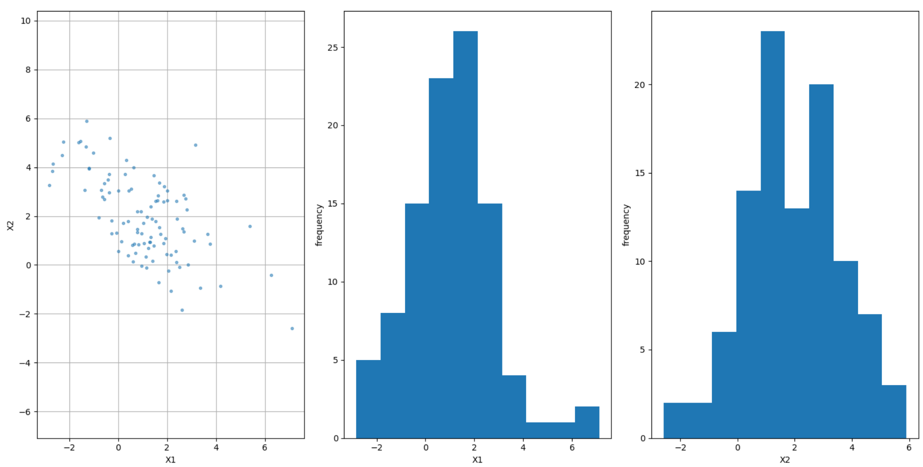

Figure 5: Multivariate gaussian variables with n = 100 and negative covariance

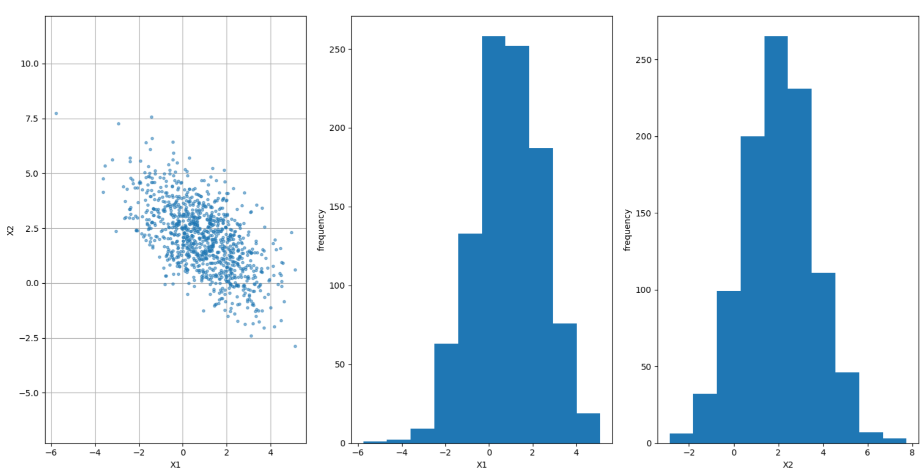

Figure 6: Multivariate gaussian variables with n = 1000 and negative covariance

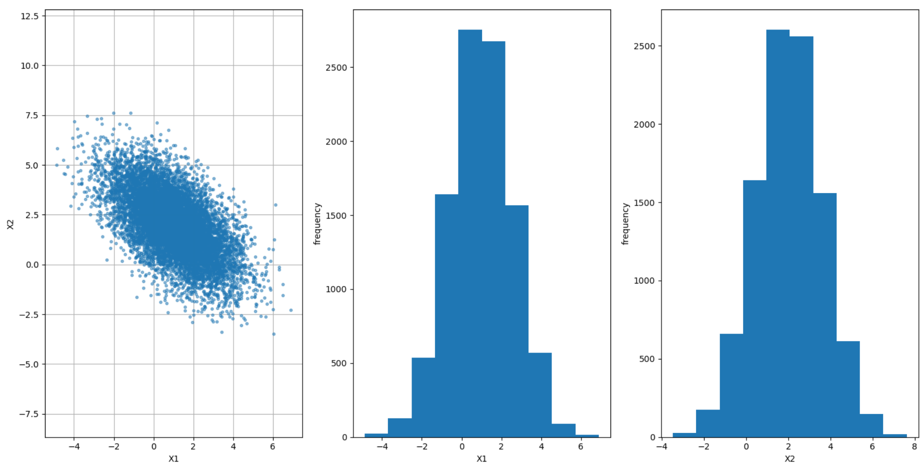

Figure 7: Multivariate gaussian variables with n = 10000, negative covariance

The scatter plots of the new generated samples show the same ellipse as previously, but rotated to the left because the covariance values are negatives. Indeed, we can compute easily by hand that:

$$\Sigma = \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix} \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}$$

$$\Sigma = \frac{1}{4} \times \begin{bmatrix} 10 & \mathbf{-6} \\ \mathbf{-6} & 10 \end{bmatrix}$$

As a last remark we can point out that the ellipse is rotated from -$\pi/4$, and that cos(-$\pi/4$) $= \frac{\sqrt{2}}{2}$ and sin(-$\pi/4$) $= -\frac{\sqrt{2}}{2}$. These values seem to be related to the values included in the Σ matrix.

2 Bayesian estimation of 2D Gaussian random variables

[redacted]

[redacted]

Figure 8: Second statement

2.1 Problem introduction

The parameter we want to estimate is the mean: $\theta = \mu$.
We make 10 samples: m = 10.

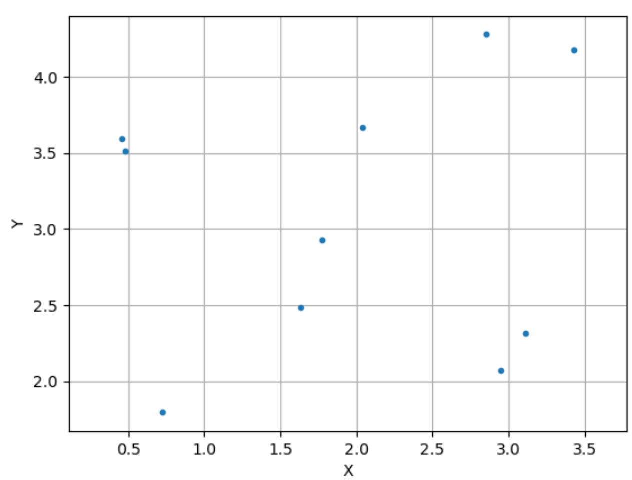

Figure 9: scatter plot of the 10 measures of the point location located at (2,3)

When the data follows a multi Gaussian distribution, the Maximum Likelihood Estimate(MLE) tells us that the estimation of μ is equal to the arithmetic mean of all the vector coordinates from the samples, ie:

$$\hat{\mu} = \frac{1}{m}\sum_{i=1}^{m} X_i$$

Yet, from the second exercise title, we guessed that the goal of the exercise was not to use a frequentist approach but a **Bayesian** one.
With Bayes formula, we know that:

$$p(\theta|D) = \frac{p(D|\theta) \times p(\theta)}{p(D)}$$

As the evidence p(D) is only a normalising constant that do not depend of θ, we can remove it.

$$p(\theta|D) \;\propto\; p(D|\theta) \times p(\theta)$$

The posterior $p(\theta|D)$ is proportional to the likelihood $p(D|\theta)$ multiplied by the prior $p(\theta)$. We want to find θ for which the posterior is maximum (Maximum A Posteriori estimate called MAP). First, let us take a look at the likelihood.

2.2 The likelihood: $p(D|\theta)$ with $D \hookrightarrow \mathcal{N}(\theta, \Sigma_D)$

$$\theta \text{ is unknown and } \Sigma_D = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$D = [\begin{bmatrix} x_1 \\ y_1 \end{bmatrix}, \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}, ..., \begin{bmatrix} x_{10} \\ y_{10} \end{bmatrix}]$$

Assuming that the likelihood of a vector is the product of its individual components, we have:

$$p(D|\theta) = p(\begin{bmatrix} x_1 \\ y_1 \end{bmatrix}|\theta) \times p(\begin{bmatrix} x_2 \\ y_2 \end{bmatrix}|\theta) \times ... \times p(\begin{bmatrix} x_{10} \\ y_{10} \end{bmatrix}|\theta)$$

Each vector $d_i = \begin{bmatrix} x_i \\ y_i \end{bmatrix}$ from the data D follows a 2D Gaussian distribution too, in other words for i $\in [\![1, 10]\!]$:

$$d_i \hookrightarrow \mathcal{N}(\theta, \Sigma_D)$$

To obtain the likelihood, we have to compute the expression below for i $\in [\![1, 10]\!]$

$$p(\begin{bmatrix} x_i \\ y_i \end{bmatrix}|\theta) = \frac{1}{(2\pi)^{\frac{m}{2}}|\Sigma_D|^{\frac{1}{2}}} e^{-\frac{1}{2}[x_i - \boldsymbol{\theta_0} \; y_i - \boldsymbol{\theta_1}]\Sigma_D^{-1}\begin{bmatrix} x_i - \boldsymbol{\theta_0} \\ y_i - \boldsymbol{\theta_1} \end{bmatrix}}$$

and then make the product of the 10 factors computed. Note that we wrote the unknown parameter in bold.
Regarding the program, we use the logarithm function because if we do not use it, the computer rounds the likelihood value to 0 or Inf because it is obtained by multiplying a lot of little probabilities.

2.3 The uniform prior: $p(\boldsymbol{\theta})$ with $\boldsymbol{\theta} \hookrightarrow \boldsymbol{\mathcal{U}}(0, 1)$

The pdf(probability density function) of a a β-distribution is:

$$f(x, \alpha, \beta) = x^{\alpha-1}(1-x)^{\beta-1}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$$

With $\Gamma(n)$ = (n-1)! for positive integer n.
As the uniform distribution is a β-distribution with $\alpha = \beta = 1$, we obtain that:

$$p(\theta) = \theta^0(1-\theta)^0\frac{1!}{0!0!}$$

$$p(\theta) = 1$$

As a result, the posterior is equal to the likelihood and MAP is equivalent to a Maximum Likelihood Estimation(MLE) problem when the prior is uniform. This can be seen in the figure 10: indeed, we can not distinguish the purple points associated to the posterior from the red one, which are associated to the likelihood.

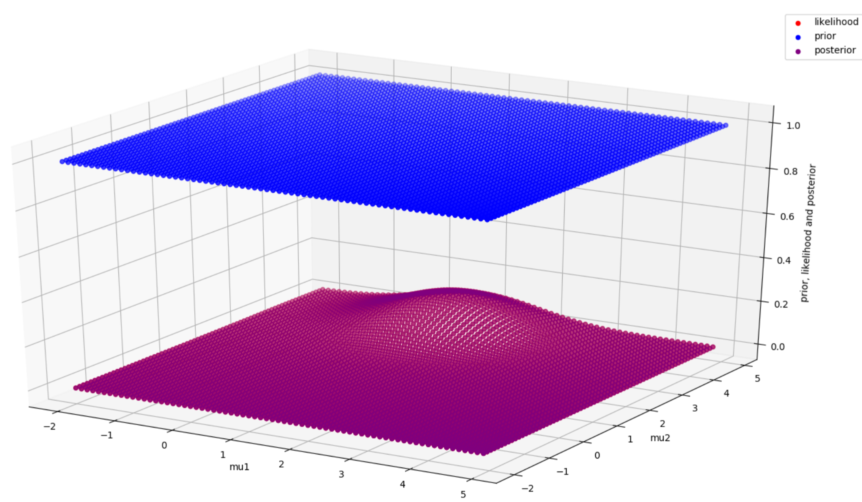

Figure 10: posterior with uniform prior

In order to find μ for which the posterior is maximum, we created various μ with coordinates between -2 and 5 by step of 0.1. For each of them, we computed the posterior, and finally selected the mu coordinates that had the maximum posterior. As a result with uniform prior, we get results as

$$\hat{\theta} = (2.3, 2.8)$$

2.4 The multivariate Gaussian prior: $p(\boldsymbol{\theta})$ with $\boldsymbol{\theta} \hookrightarrow \mathcal{N}(\boldsymbol{\mu_\theta}, \boldsymbol{\Sigma_\theta})$

As the prior follows the same distribution as the posterior, we have a conjugate prior.

$$p(\theta) = \frac{1}{(2\pi)^{\frac{m}{2}} |\Sigma_\theta|^{\frac{1}{2}}} e^{-\frac{1}{2}[\boldsymbol{\theta_0} - \mu_1 \ \boldsymbol{\theta_1} - \mu_2] \Sigma_\theta^{-1} \begin{bmatrix} \boldsymbol{\theta_0} - \mu_\theta \\ \boldsymbol{\theta_1} - \mu_\theta \end{bmatrix}}$$

In the expression above of the prior, μ_θ and Σ_θ are fixed (we should choose them). Again, note that we wrote the unknown parameter in bold.
We tested different values of μ with μ_1 taking values between -5 and 5, and μ_2 taking value between -3 and 7, with step of 0.1. (values chosen to get better plots) Figure 11 plots the posterior, likelihood and prior of p(θ) when we set : $\mu_\theta = (1, 1)$ and $\Sigma_\theta = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$ With those values, we get: $\hat{\theta} = (2.2, 3.4)$

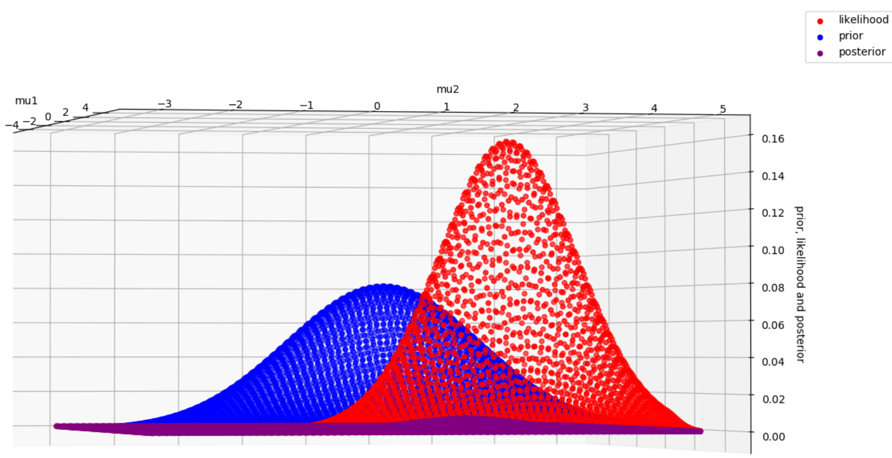

Figure 11: Posterior, likelihood and prior of p(θ) with multi Gaussian prior

The likelihood height is twice higher than the prior height. It is related to the variances values on the diagonal of Σ equal to 2. If they are set to 4 instead of 2, then the likelihood becomes 4th times higher than the prior. It can be seen that the prior mean has been set to (1,1) because the coordinates of the pic of the blue "mountain" is (1,1).
As the posterior curve is almost non visible in figure 11 , we plot it alone in the figure 12:

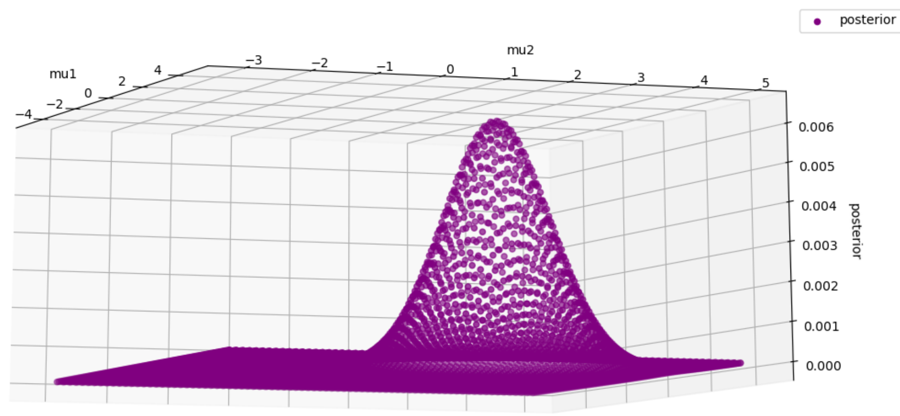

Figure 12: Posterior of p(θ) with multi Gaussian prior

Figure 13 plots the posterior, likelihood and prior of p(θ) with : $\mu_\theta = (2.5, 3.5)$ and $\Sigma_\theta = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.
With those values, we get: $\hat{\theta} = (2.1, 2.9)$, which are the mu coordinates that are the closest to the pic of the purple "mountain", among all mu coordinates tested.

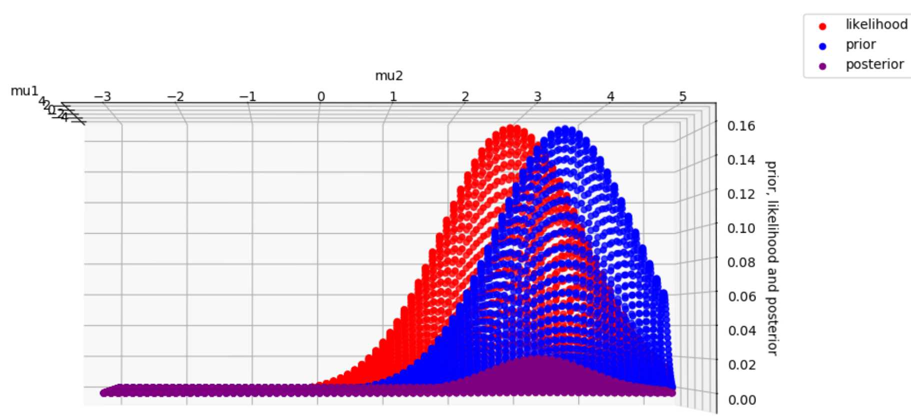

Figure 13: Posterior, likelihood and prior of p(θ) with multi Gaussian prior with different parameters

As the variance is 1, the likelihood and the prior have the same height.