TS115 Rapport de TP

Oumaima Najib

May 2021

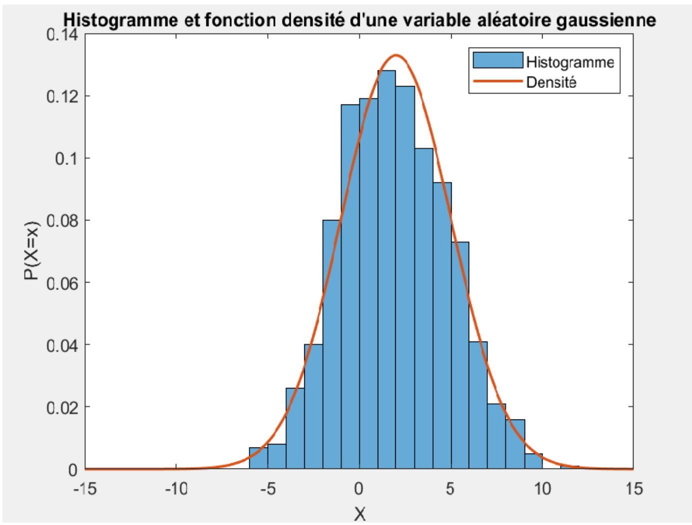
1 Variable aléatoire gaussienne univariée

1.1 Manipulation 1

La variable aléatoire gaussienne Y de moyenne μ et de variance σ est paramétrée en utilisant la formule : $\mu + \sigma \times X$, avec $X \sim N(0, 1)$.

On a utilisé la normalisation de type 'pdf'. c_i est le nombre d'élément dans chaque bin, c'est à dire le nombre de fois ou ont été générés des x appartenant à $[x_i\ ,\ x_i + 1]$, w_i correspond à la largeur du bin. Et N correspond au nombre de réalisations : La valeur assigné à chaque bin est égal à :

$$\hat{f}_X(x) = \frac{c_i}{N \times w_i}.$$



Avec cette normalisation, la hauteur de chaque bin v_i est égal à la probabilité de sélectionner une observation se trouvant dans l'intervalle du bin, la somme des aires sous les barres est donc égal à 1.

1.2 Manipulation 2

L'expression de l'entropie dans le cas d'une variable aléatoire continue est :

$$H(X) = -\int_{\infty} f_X(x) log(f_X(x))\, dx$$

avec f_X la densité de probabilité de X.

Pour obtenir l'entropie de la variable aléatoire générée, on calcule l'aire sous les bins obtenue dans l'histogramme, avec multiplication par $-log(v_i)$. La somme continue est remplacée par une somme discrète sur les valeurs estimées de f_X, ici les v_i. On multiplie cette somme par la largeur des bins de l'histogramme pour obtenir le bon résultat. On obtient une valeur proche de 2.51, qui est la valeur théorique de l'entropie d'une variable aléatoire gaussienne de variance 9.

$$H = ln(\sigma\sqrt{2\pi} * e) = 2,51.$$

2 Variable multidimensionnelle aléatoire gaussienne

2.1 Question 1

Soit Σ la matrice , au vu de l'indépendance, diagonale de covariance de la variable aléatoire gaussienne à P composantes, et μ le vecteur représentant son centre.

On note X le vecteur gaussien dans R^p : $X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}$ avec $X_1, X_2, \cdots, X_p$ les composantes de X.

Son espérance s'écrit : $\mu = \begin{bmatrix} E[X_1] \\ E[X_2] \\ \vdots \\ E[X_p] \end{bmatrix}$ et sa matrice de covariance :

$$\Sigma = E[(X-E[X])(X-E[X])^\top] = \begin{bmatrix} Var(X_1) & Cov(X_1, X_2) & \cdots & Cov(X_1, X_p) \\ Cov(X_2, X_1) & Var(X_1, X_2) & \cdots & Cov(X_2, X_p) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(X_p, X_1) & Cov(X_p, X_2) & \cdots & Var(X_p) \end{bmatrix}$$

Comme les variables aléatoires sont indépendantes et de même variance :

$$\Sigma = \begin{bmatrix} \sigma & 0 & \cdots & 0 \\ 0 & \sigma & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma \end{bmatrix}$$

avec σ la variance de X_1.

La loi gaussienne est donnée par :

$$f_X(x) = \frac{1}{(2\pi)^{P/2}\ |\Sigma|^{1/2}} exp[-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)]$$

2.2 Question 2

On le démontre en comparant les fonctions caractéristiques de $\Sigma_y^{1/2}X + \mu$ et Y où $X \sim N(0, I)$. En effet, la fonction caractéristique d'une variable aléatoire mutlivariée est donnée par : $\phi_{\mu,\Sigma}(t) = exp(it^\top \mu - \frac{1}{2}t^\top \Sigma t)$.

En appliquant la transformation :

$$\begin{aligned} \phi_Y(t) &= E[exp(it^\top X)] \\ &= E[exp(it^\top(\Sigma_y^{1/2}X + \mu)] \\ &= exp(it^\top \mu) \times E[exp(it^\top(\Sigma_y^{1/2}X)] \\ &= exp(it^\top \mu) \times \phi_X(\Sigma_y^{1/2}t) \\ &= exp(it^\top \mu) \times E[exp(-\frac{1}{2}t^\top \Sigma t))] \\ &= exp(it^\top \mu - \frac{1}{2}t^\top \Sigma t). \end{aligned}$$

Le résultat correspond bien à la fonction caractéristique d'une variable aléatoire $Y \sim N(\mu, \Sigma)$.

2.3 Manipulation 3

On génére 1000 échantillons de la variable aléatoire Y, on obtient une matrice 2x1000, dont la première ligne correpond aux échantillons de la variable aléatoire Y_1 et la deuxième à ceux d'Y_2. À l'aide de mean et cov, on obtient des valeurs des espérance et covariances expérimentales proches aux valeurs théoriques ~ 0.2, $\sim \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$.

2.4 Manipulation 4

On obtient les histogrammes représentés sur les figures ci-dessous. Les histogrammes sont bien centrés selon les deux axes en leurs moyennes respectives, (0_2 pour X et μ pour Y). De plus, on remarque que l'histogramme du vecteur gaussien Y forme une ellipse et celui de X forme un cercle ; c'est les covariances non nulles, qui représentent la non indépendance entre Y_1 et Y_2, qui sont à l'origine de cette différence.

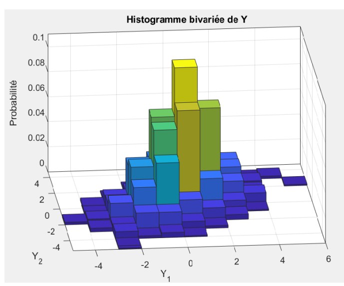


FIGURE 1 – Histogramme 2D de Y

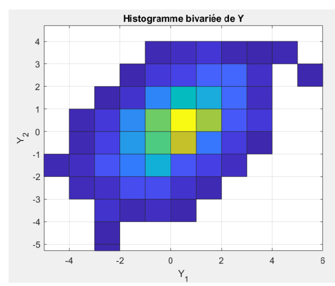


FIGURE 2 – Vu de haut

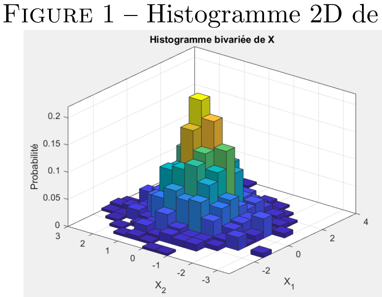


FIGURE 3 – Histogramme 2D de X

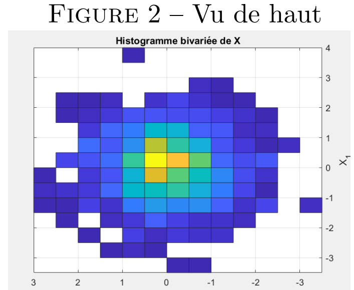


FIGURE 4 – Vu de haut

2.5 Manipulation 5

On obtient la matrice suivante : $\sim \begin{bmatrix} 1 & 0.96 \\ 0.96 & 1 \end{bmatrix}$ On en déduit donc que le coefficient de corrélation du couple $(X, Y) \approx 0.96$: les deux variables aléatoires sont donc corrélées.

3 Mesures quantitatives

3.1 Manipulations 6 et 7

X_1 est le vecteur représentant les enregistrements des hauteurs des précipitations relevées dans la ville de Bordeaux, X_2 dans Nantes et X_3 dans San Tiago.

En modélisant ces données comme étant des variables aléatoires, on peut trouver numériquement la corrélation entre les précipitations de ces trois villes respectives.

On calcule la matrice de covariance de X_1 et X_2 :

$$\begin{bmatrix} var(X_1) = 65.0000 & cov(X_1, X_2) = 35.0000 \\ cov(X_1, X_2) = 35.0000 & var(X_2) = 65.0000 \end{bmatrix}$$

En comparant les valeurs de variance et de covariance, on en déduit que les hauteurs des précipitations de Bordeaux et Nantes sont fortement corrélées, ce qui est logique au vu du fait que les deux villes sont proches. ($\sim 350km$).

En calculant la matrice de covariance de X_1 et X_3 on obtient :

$$\begin{bmatrix} 65.0000 & 0.1090 \\ 0.1090 & 67.0723 \end{bmatrix}$$

On trouve dans ce cas que $cov(X_1, X_3) \approx 0$, ce qui veut dire que ces données sont peu correlées, si ce n'est pas correlées. La distance entre la ville de Bordeaux et San Tiago est de $9,273km$, c'est ce qui explique l'indépendance entre X_1 et X_2.

On retrouve à peu près le même résultat pour X_2 et X_3.

3.2 Manipulation 8

L'expression de l'information mutuelle de deux variables aléatoires bivariées s'écrit :

$$I(X, Y) = H(X) + H(Y) - H(X, Y)$$

Il nous faut alors trouver un moyen de calculer l'entropie conjointe du couple (X, Y). Dans le cas de X_3 et X_2, X_2 et X_3, $I(X)$ sera nulle au vu de l'indépendance.

```matlab
close all
clear
clc

%% Manipulations 1,2

%_______________Paramètres
    N = 1000;
    u=2 ; sigma =3;
    Realisations = u + sigma * randn(N,1);
    y = -15:0.1:15;
    f = exp(-(y-u).^2./(2*sigma^2))./(sigma*sqrt(2*pi));

%_______________Figures
figure, hist = histogram(Realisations, 'Normalization','pdf');
hold on
plot(y, f, 'Linewidth', 1.5); xlabel('X'), ylabel('P(X=x)');
title("Histogramme et fonction densité d'une variable aléatoire gaussienne");
legend('Histogramme', 'Densité');
H= calcul_entropie(hist);

%% Manipulations 3,4,5

%_______________Paramètres

       X = randn(2,N);
       A = [2 1; 1 2]; [U,D] = eig(A); B = U*D.^(0.5)*U';
        mu = [0.2 0.2]';
       Y = B*X + mu;
       m_1 = mean(Y(1,:));m_2 = mean(Y(2,:)); covv = cov(Y(1,:),Y(2,:)); %Vérification
       XT = X.'; YT = Y.';
       corrcoefXY = corrcoef(YT(:,1),XT(:,1));

%_____________Figures
figure,  h = histogram2(Y(1,:),Y(2,:),'Normalization','pdf');
h.FaceColor = 'flat';
xlabel('Y_1'), ylabel('Y_2');
zlabel('Probabilité') ;
title("Histogramme bivariée de Y");
figure,
 h1 = histogram2(X(1,:), X(2,:),'Normalization','pdf');
xlabel('X_1'), ylabel('X_2');
zlabel('Probabilité') ;
title("Histogramme bivariée de X");
h1.FaceColor = 'flat';

%% Manipulation 6,7,8

%______________Paramètres
    load X_pluv.mat
    X1 = X_pluv(:,1);    %   Bordeaux
    X2 = X_pluv(:,2);    %   Nantes
    X3 = X_pluv(:,3);    %   Santiago

    Y1 = [X1, X2]; Y2 = [X1, X3]; Y3 = [X2, X3];
    cov1 = cov(Y1); cov2 = cov(Y2); cov3 = cov(Y3);

%% Functions
%_______________Entropie

function H=calcul_entropie(hist)
      H = 0;
      number_bins = hist.NumBins;
      values = hist.Values;
      width = hist.BinWidth;
      for i=1:number_bins
        if(values(i) ~= 0)
          H = H - values(i)*width*log(values(i));
        end
      end
end
```