DURET Guillaume & EDORH Steve

TP 4 – M3

Probabilités discrètes et continues

Année scolaire 2017/2018

Exercice 1 : Simulation de lois continues par la méthode d'inversion

Introduction :

On nous rappelle le théorème suivant :

Soit X une variable aléatoire continue de fonction de répartition F_X. Alors la variable aléatoire $Y=F_X(X)$ suit la loi uniforme sur [0,1].

Ainsi, indépendamment de sa loi de probabilité, toute variable aléatoire X peut être simulée de à partir d'une variable aléatoire Y soumise à une loi uniforme sur [0,1], via la formule $X=F_X^{-1}(Y)$.

Nous nous proposons donc de vérifier ce théorème pour approximer différentes variables aléatoires.

1) Soit la densité de probabilité f d'une variable aléatoire X :

$$f(x) = \begin{cases} \frac{a}{x+4}, si \ -3 \leq x \leq 2 \\ 0, sinon \end{cases}$$

1a) On cherche à calculer **la valeur de a** et à déterminer l'expression de **la fonction de répartition F_X.**

Dans un premier temps, nous utilisons la propriété d'une densité de probabilité :

$$f \text{ est une densité de probabilité} \Leftrightarrow \int_{-\infty}^{x} f(t)dt = 1$$

$$D'\text{où } l'\text{égalité} \int_{-3}^{x} \frac{a}{t+4} dt = 1$$

$$\Leftrightarrow a \times [ln(x+4)]_{-3}^{x} = a\big(ln(6) - ln(1)\big) = 1$$

$$\Leftrightarrow \boxed{\boldsymbol{a = \frac{1}{ln(6)}}}$$

Il s'avère que l'inconnue a est positive, ainsi f est bien une densité de probabilité.

On peut maintenant **déterminer F_x** :

Pour $x < -3$, $F_X(x) = \int_{-\infty}^{x} f(x)dx = 0$

Pour $-3 \leq x \leq 2$, $\boxed{\boldsymbol{F_X(x) = \int_{-\infty}^{x} f(x)dx = \int_{-3}^{x} \frac{a}{x+4} dx = \frac{1}{ln(6)} ln(x+4)}}$

Et pour $x > 2$, $F_X(x) = \int_{-\infty}^{x} f(x)dx = 1$

1b) On cherche ensuite à tracer sur Matlab l'histogramme la densité correspondant à la **loi simulée**, ainsi que la **courbe représentative de la densité théorique.**

Pour ce faire, on doit tout d'abord calculer la fonction de répartition inverse :

Pour $-3 \leq x \leq 2$, $F_x(x)= \frac{1}{ln(6)} ln(x+4) = Y$

$\Leftrightarrow Y ln(6) = ln(x+4)$

$$\Leftrightarrow e^{ln(6^Y)} = x+4$$

$$\Leftrightarrow x = e^{ln(6^Y)} - 4 = F_X^{-1}$$

On a donc $F_x^{-1}(Y) = e^{ln(6^Y)} - 4$

Finalement, la fonction de répartition inverse est $\boxed{\boldsymbol{F_x^{-1}(Y) = 6^Y - 4 = X}}$

Nous avons donc le script suivant :

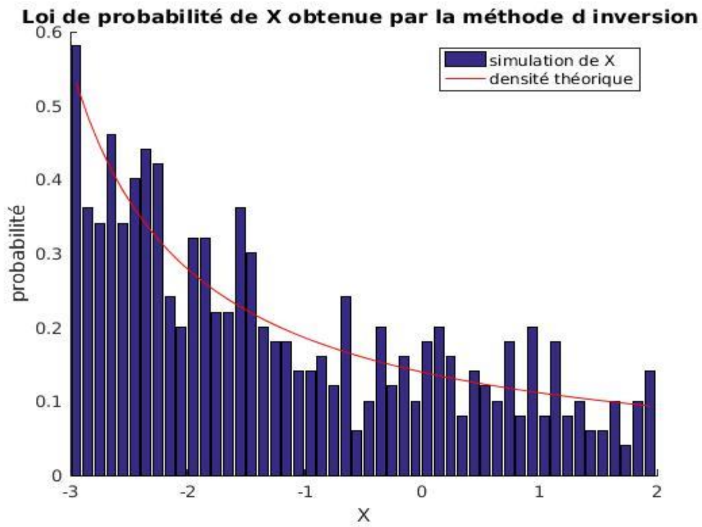
```
N=500;
Y=rand(1,N);% Y suit une loi uniforme
a=1/log(6);
X=exp(Y/a)-4 ;% = Fx-1(Y)
[h,xout]=hist(X,50); % xout sont les abscisses
pas=xout(2)-xout(1);
S=sum(h)*pas; % aire de l'histogramme
figure(1)
hold on;
bar(xout,h/S); % /S pour que l'aire de l'histogramme soit normalisée
f=a./(xout+4);% densité théorique
plot(xout,f,'r');
title('Loi de probabilité de X obtenue par la méthode d inversion ');
xlabel('X');
ylabel('probabilité');
legend('simulation de X','densité théorique');
```

Dans ce programme, nous mettons en pratique **la méthode d'inversion** pour vérifier la propriété rappelée au début de l'exercice. Nous nous servons de la fonction de répartition inverse de X, pour simuler son comportement en fonction de Y, une variable aléatoire suivant une loi uniforme sur [0,1]. Avec la fonction `hist`, on établit l'histogramme associé à X.

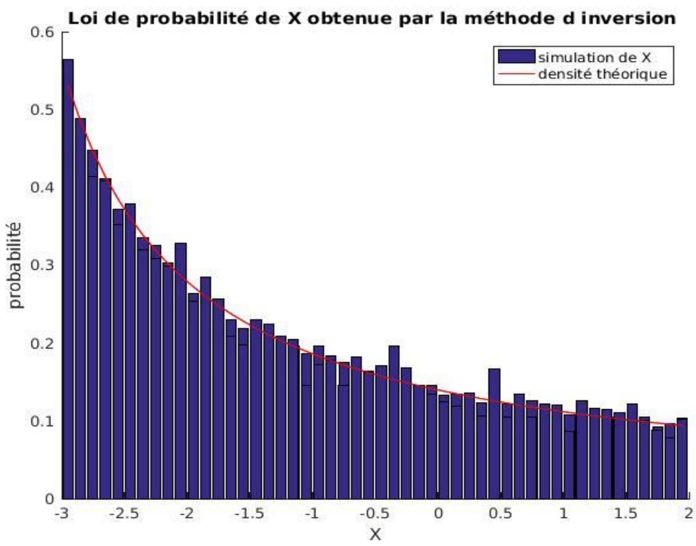
Nous calculons le pas pour pouvoir calculer l'aire totale de l'histogramme. Nous nous en servons pour normaliser l'histogramme obtenu (afin d'avoir des probabilités).

Sur la même figure, nous traçons la densité de probabilité de X.

Pour n=500



Pour N=5000



On observe de manière prévisible que lorsque l'on augmente N la taille de l'échantillon, la simulation devient très précise.

1c) Pour calculer les valeurs empiriques de l'espérance et de l'écart type, nous utilisons respectivement les fonctions `mean` et `std`.

Nous calculons $E(X) = \int_{-\infty}^{+\infty} xf(x)dx$ et $E(X^2) = \int_{-\infty}^{+\infty} x^2 f(x)dx$

afin d'avoir espérance et écart type.

$$E(X) = a\int_{-3}^{2} \frac{x}{x+4}dx$$

$$\Leftrightarrow E(X) = a \times \left(\int_{-3}^{2} \frac{x+4}{x+4} dx - \int_{-3}^{2} \frac{4}{x+4} dx \right)$$

$$\Leftrightarrow E(X) = a \times ([x]_{-3}^{2} - 4 \times [ln(x+4)]_{-3}^{2})$$

$$\Leftrightarrow E(X) = a \times \big(2 + 3 - 4ln(6)\big)$$

Finalement, on a l'espérance qui suit :

$$\boxed{\boldsymbol{E(X) = 5a - 4}}$$

$$E(X^2) = \int_{-3}^{2} \frac{ax^2}{x+4} dx$$

$$\Leftrightarrow E(X^2) = a \int_{-3}^{2} \frac{(x+4)x}{x+4} dx - \int_{-3}^{2} \frac{4ax}{x+4} dx$$

$$\Leftrightarrow E(X^2) = \left[\frac{ax^2}{2}\right]_{-3}^{2} - 4(5a - 4)$$

$$\Leftrightarrow E(X^2) = \frac{4a - 9a}{2} - 20a + 16$$

On a donc :

$$\boxed{\boldsymbol{E(X^2) = 16 - \frac{45a}{2}}}$$

Finalement, on a plus qu'à directement calculer l'écart type **(*X*) =** $\sqrt{\boldsymbol{E(X^2) - E(X)^2}}$

(Formule de Koenig-Huygens)

Nous rajoutons ainsi les lignes suivantes au script :

```
esp_empi=mean(X)

esp_theo=5*a-4

ecart_empi=std(X)

ecart_theo=sqrt((16-45*a/2)-(5*a-4)^2)
```

On obtient les résultats suivants pour N=5000 :

```
esp_empi = -1.2329

esp_theo = -1.2094

ecart_empi = 1.3990

ecart_theo = 1.4070
```

Ces résultats témoignent de **l'efficacité de la méthode d'inversion**. La proximité entre théorie et pratique s'observe de plus en **plus lorsque la taille de l'échantillon augmente**.

2) Nous simulons cette fois une loi de Poisson de paramètre λ=3. Sa densité est la fonction f définie par :

$$f(x) = \begin{cases} 3e^{-3x}, si\ x \geq 0 \\ 0\ \ sinon \end{cases}$$

Sa fonction de répartition associée est :

$$F_X(x) = \begin{cases} 1 - e^{-3x}, si\ x \geq 0 \\ 0\ \ sinon \end{cases}$$

Comme précédemment, nous utilisons la méthode d'inversion sur cette nouvelle variable aléatoire. Nous devons donc exprimer F_X^{-1}.

Soit $Y = \mathrm{F}_X(\mathrm{x})$

$$\Leftrightarrow Y = 1 - e^{-3x}$$

$$\Leftrightarrow \cdots$$

$$\Leftrightarrow \boxed{\boldsymbol{x = -\frac{\ln(1-Y)}{3} = F_X^{-1}(Y)}}$$

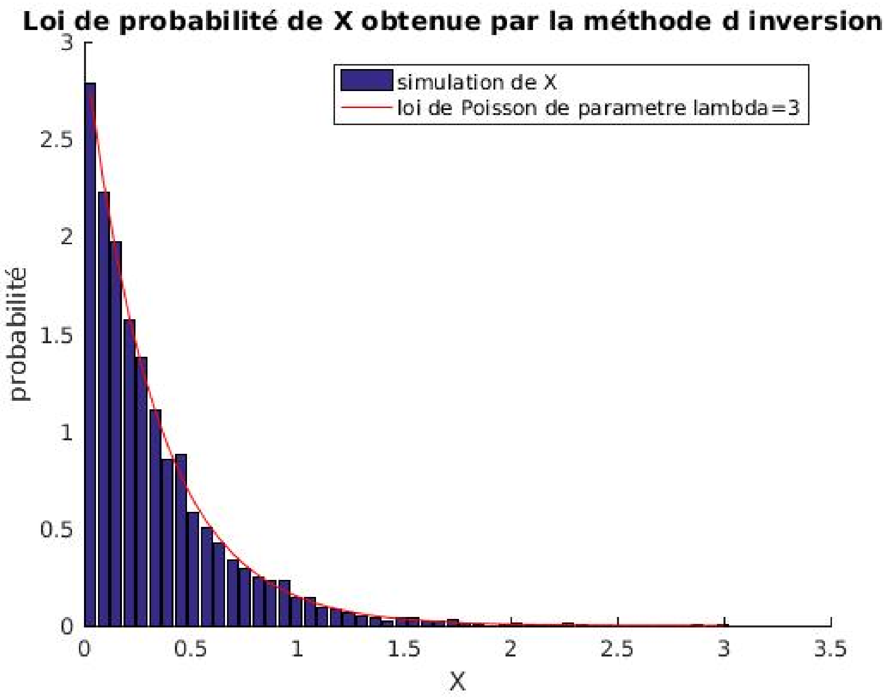
Nous rajoutons donc les lignes suivantes à notre code :

```
lambda=3;
X2=log(-Y+1)/(-lambda) ; % fonction de répartition inverse
[h2,xout2]=hist(X2,50); % xout sont les abscisses
pas2=xout2(2)-xout2(1);
S2=sum(h)*pas2; % aire de l'histogramme
figure(2)
hold on;
bar(xout2,h2/S2); % /S normalisation de l'aire
f2=lambda*exp(-lambda*xout2);
plot(xout2,f2,'r');
xlabel('X');
ylabel('probabilité');
legend('simulation de X','loi de Poisson de paramètre lambda=3');
title('Loi de probabilité de X obtenue par la méthode d inversion ');


esp_empi2=mean(X2)
esp_theo2=1/lambda
ecart_empi2=std(X2)
```

```
ecart_theo2=1/lambda
```

Le raisonnement est identique à celui adopté lors de la question précédente.

Nous obtenons ainsi la figure suivante, toujours avec N=5000 :



Pour espérances et écarts types sont les suivants :

```
esp_empi2 = 0.3308

esp_theo2 = 0 .3333

ecart_empi2 = 0.3300

ecart_theo2 = 0.3333
```

Nous pouvons une fois de plus noter l'efficacité de la méthode d'inversion pour un échantillon suffisamment grand.

<u>Conclusion :</u>

Cet exercice nous a permis d'appréhender et d'utiliser la méthode d'inversion afin de simuler des variables aléatoires suivant des lois de probabilités différentes. Nous avons pu constater que cette approximation est correcte, notamment lorsque l'échantillon est suffisamment grand.

Nous pouvons donc imaginer les applications possibles de cette méthode dans le domaine des statistiques.

Exercice 2 : Simulation de la loi normale (méthode de Box-Muller)

Introduction :

Nous allons voir dans cet exercice comment simuler la loi normale, en effet celle-ci n'ayant pas une fonction de répartition usuel, il est impossible d'utiliser directement la méthode d'inversion comme dans l'exercice précédent.

Cependant pour pouvoir quand même simuler cette loi nous allons utiliser la méthode de Box-Muller qui permet de simuler 2 variables aléatoire normales centrées réduites.

Tout d'abord sachant que X et Y sont les lois centrées réduites que l'on cherche à simuler, nous introduisons deux autres variables aléatoires R et θ qui correspondent aux coordonnées polaires du point de coordonné (X, Y) de tel sorte que l'on ait :

$$\begin{cases} X = R * cos(\theta) \\ Y = R * \sin(\theta) \end{cases}$$

Nous allons chercher à déterminer les lois de R et θ afin d'avoir une expression des loi X et Y.

1) Pour ce faire nous calculons la matrice jacobienne de R et θ afin de déterminer la densité conjointe de R et θ grâce à la formule suivante :

$$f(r,\theta) = |J(r,\theta)| * f_{X,Y}(x,y)$$

De plus les loi X et Y sont indépendantes on a donc :

$$f(r,\theta) = |J(r,\theta)| * f_X(x) * f_Y(y)$$

On calcule : $J(r,\theta) = \begin{vmatrix} \frac{dX}{dR} & \frac{dX}{d\theta} \\ \frac{dY}{dR} & \frac{dY}{d\theta} \end{vmatrix} = \begin{vmatrix} \cos(\theta) & -r * \sin(\theta) \\ \sin(\theta) & r * \cos(\theta) \end{vmatrix} = r$

On utilise donc la formule ci-dessus sachant que $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ et $f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}$ (densité des lois normales centré réduite),

nous trouvons finalement : $f_{R,\theta}(r,\theta) = r\frac{1}{2\pi} e^{-\frac{(x^2+y^2)}{2}} = \frac{1}{2\pi} r e^{-\frac{r^2}{2}}$.

On peut remarquer que la densité conjointe ne dépend pas du paramètre θ, on a donc nécessairement une densité marginale qui est constante pour la variable θ ce qui implique par définition que les variables aléatoires R et θ sont indépendantes ($f_{R,\theta}(r,\theta) = f_R(r) * f_\theta(\theta)$).

On détermine ensuite classiquement les densités marginales de R et θ à l'aide de la définition :

$$f_\theta(\theta) = \int_0^{2\pi} \frac{1}{2\pi} r e^{-\frac{r^2}{2}} d\theta = \frac{1}{2\pi} \text{ et } f_R(r) = \int_0^{+\infty} \frac{1}{2\pi} r e^{-\frac{r^2}{2}} dr = r e^{\frac{r^2}{2}}$$

On remarque facilement que la variable θ suit une loi uniforme

On peut constater que $f_\theta(\theta)$ et $f_R(r)$ sont bien des densités de probabilité car leurs aires font bien 1.

2) On veut se ramener à une fonction usuelle pour R, pour cela on va étudier R²
On détermine sa fonction de répartition grâce à la formule du cours :

$$F_{R^2}(\rho) = P(R^2 < \rho) = P(R < \sqrt{\rho}) = \int_0^{\sqrt{\rho}} re^{-\frac{r^2}{2}} dr = 1 - e^{-\frac{\rho}{2}}$$

On remarque donc que la variable aléatoire R² suit une loi exponentielle de paramètre $\lambda = ½$

3) La variable aléatoire R² suit une loi exponentielle et a donc une fonction de répartition usuelles, afin de pouvoir la simuler nous pouvons donc utiliser la méthode d'inversion.

D'après le théorème d'inversion on a $F_{R^2}(r^2) \sim \mathrm{U1}$, Avec U1 une variable aléatoire uniforme.

Ainsi U1=$1 - e^{-\frac{R^2}{2}}$ et donc $R^2 = -2\ln(-U1)$ ce qui nous permet de pouvoir simuler R tel que

$$R = \sqrt{-2\ln(1 - U1)}$$

4) On fait la même chose avec la loi uniforme θ :

$F_{\theta}(\theta) = \int_0^{\theta} \frac{1}{2\pi} dt = \frac{\theta}{2\pi} = \mathrm{U2}$ nous permet donc de trouver : $\theta = 2\pi\mathrm{U2}$

5) Maintenant que l'on a des expressions de R et θ en fonction de variables uniforme on peut aussi obtenir X et Y en fonction de variables uniformes ce qui va nous permet de simuler ces deux loi normales centrées réduites avec les formules :

$$X = \sqrt{-2\ln(1 - U1)}\cos(2\pi\mathrm{U2})$$

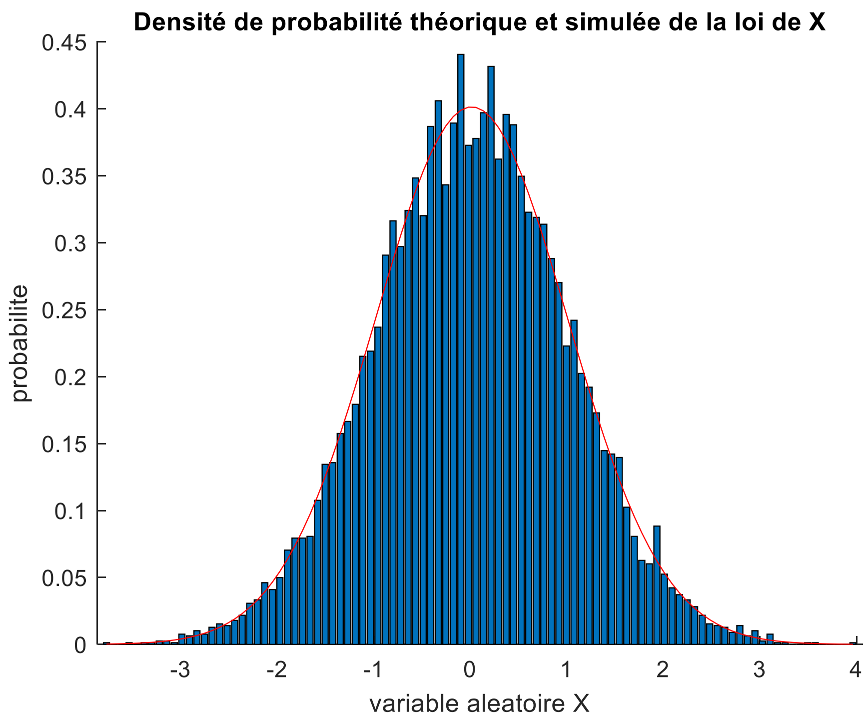
$$Y = \sqrt{-2\ln(1 - U1)}\sin(2\pi\mathrm{U2})$$

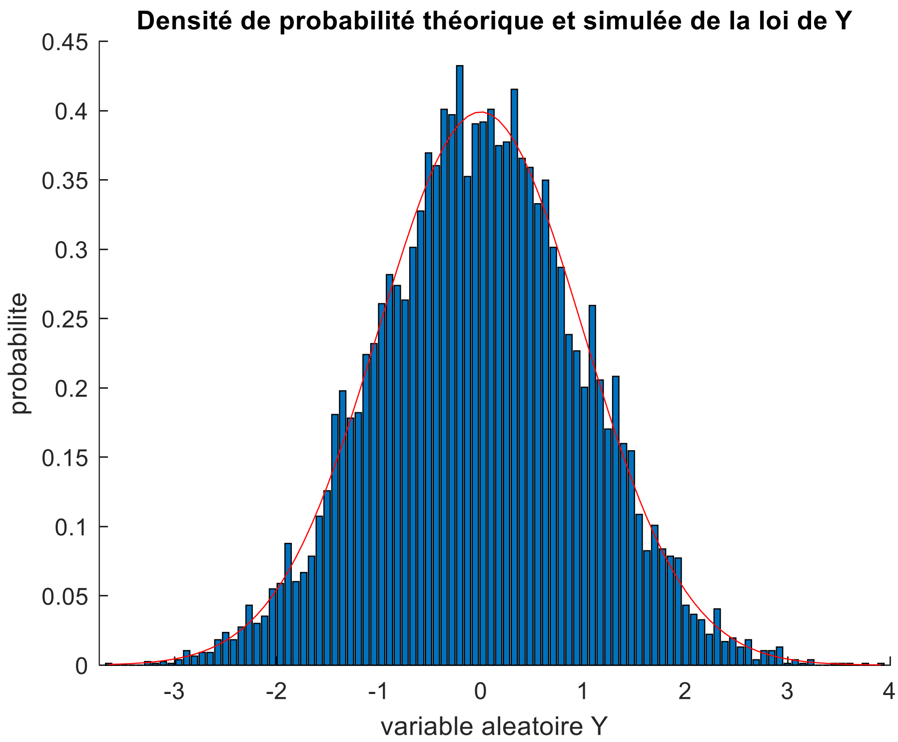
6) Nous allons maintenant simuler ces variables aléatoires avec le programme suivant :

```
N=10000;

U1=rand(1,N);
U2=rand(1,N);
X=sqrt(-2*log(1-U1)).*cos(2*pi*U2);

[h,xout]=hist(X,100); % xout sont les abscisses
pas=xout(2)-xout(1);
S=sum(h)*pas; % aire de l'histogramme
figure(1)
hold on;
bar(xout,h/S); % /S pour que l'aire des barres fasse 1
esp_empiX=mean(X)
ecart_empiX=std(X)
f=(1/(ecart_empi*sqrt(2*pi)))*exp(-(xout-esp_empi).^2/(2*ecart_empi.^2));
% Tracer de l'histogramme et de la courbe théorique
plot(xout,f,'r');
```

```
title('Densité de probabilité théorique et simulée de la loi de X');
xlabel('variable aleatoire X');
ylabel('probabilite');



Y=sqrt(-2*log(1-U1)).*sin(2*pi*U2);
[h2,xout2]=hist(Y,100); % xout sont les abscisses
pas2=xout2(2)-xout2(1);
S2=sum(h)*pas2; % aire de l'histogramme
figure(2)
hold on;
bar(xout2,h2/S2); % /S pour que l'aire des barres fasse 1
esp_empiY=mean(Y)
ecart_empiY=std(Y)
f2=(1/(ecart_empi2*sqrt(2*pi)))*exp(-(xout2-esp_empi2).^2/(2*ecart_empi2.^2));
% Tracer de l'histogramme et de la courbe théorique
plot(xout2,f2,'r');
title('Densité de probabilité théorique et simulée de la loi de Y');
xlabel('variable aleatoire Y');
ylabel('probabilite');
```

Nous commençons tout d’abord à générer les lois uniformes afin de les utiliser dans les expressions des lois de X et de Y, à l’aide de la commande hist on calcule l’histogramme que l’on normalise afin d’avoir des probabilités. On superpose à cet histogramme les courbe de la densité théorique. On obtient les figures ci-dessous :





$Eemp(X)$ = 0.0039 et σemp (X) = 1.0032

$Eemp(Y)$ = 0.0084 et $\sigma emp(Y)$ = 0.9954

On remarque donc que la simulation empirique suit bien la densité de probabilité théorique d'une variable aléatoire centrée réduite.

De plus les moyenne et les écarts types calculés empirique sont très proche de la théorie(E(X)=E(Y)=0 et σ(X=σ(Y)=1)

7) On cherche ensuite à utiliser cette simulation de loi normale centrée réduite pour la simulation une loi normale quelconque.

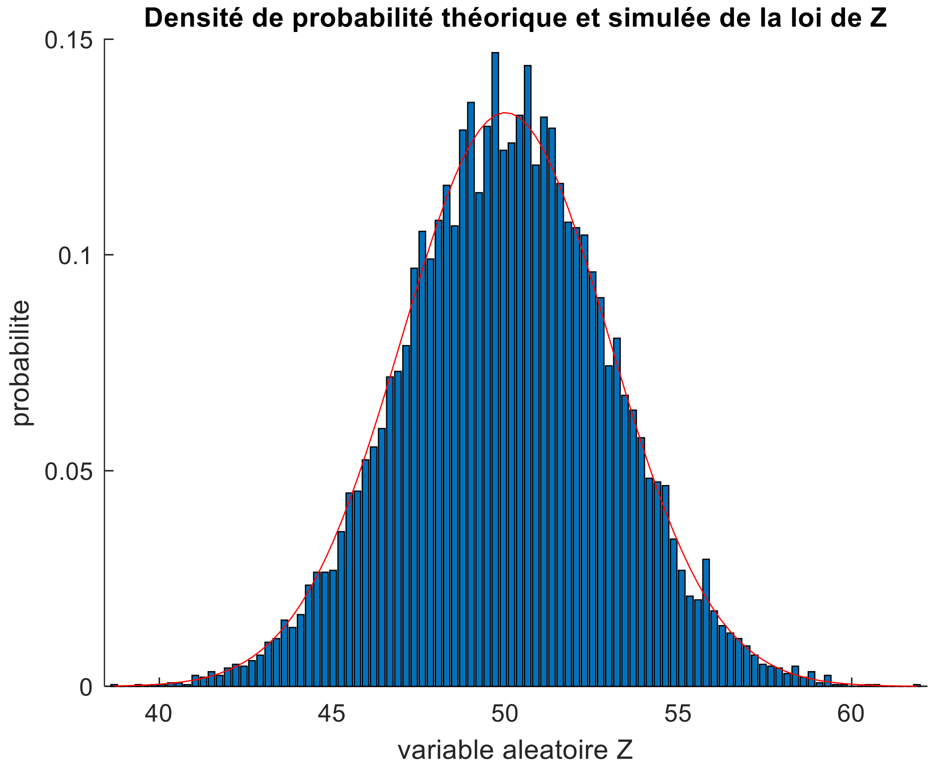
Pour cela on va utiliser le fait que pour une variable aléatoire normal Z quelconque de moyenne m et d'écart type σ on a $X = \frac{Z-m}{\sigma}$ suit une loi normale centrée réduite. On exprime donc Z en fonction d'une variable aléatoire centrée réduite : $\mathrm{Z} = \sigma\mathrm{X} + \mathrm{m}$

On obtient donc le code suivant :

```
% paramètre de la loi normale que l'on veut simuler
Moyenne=50;
ecart_type=3;


Z=ecart_type*X+Moyenne;
[h3,xout3]=hist(Z,100); % xout sont les abscisses
pas3=xout3(2)-xout3(1);
S3=sum(h3)*pas3; % aire de l'histogramme
figure(3)
hold on;
bar(xout3,h3/S3); % /S pour que l'aire des barres fasse 1
```

```
esp_empi3=mean(Z)
ecart_empi3=std(Z)
f3=(1/(ecart_type*sqrt(2*pi)))*exp(-(xout3-Moyenne).^2/(2*ecart_type.^2));
% Tracer de l'histogramme et de la courbe théorique
plot(xout3,f3,'r');
title('Densité de probabilité théorique et simulée de la loi de Z');
xlabel('variable aleatoire Z');
ylabel('probabilite');
```

Tout comme précédemment on trouve la courbe associée :



$Eemp$(Z) = 50.0201 et σemp(Z) = 3.0162

On remarque donc encore une fois que la simulation empirique suit bien la densité de probabilité théorique d'une variable aléatoire centrée quelconque de paramètre m et σ.

De plus les moyenne et les écarts types calculés empirique sont très proche de la théorie(E(Z)= 50 et σ(Z)=3

<u>Conclusion</u> : le but de cet exercice était de simuler une loi normale qui à la particularité de ne pas pouvoir être simulable avec la méthode d'inversion. Nous avons donc exprimé les lois que l'on cherche en fonction de 2 autres lois qui sont usuel et sur laquelle on peut utiliser la méthode d'inversion.

Enfin on peut imaginer la méthode qui consiste à 'introduire de nouvelles variables aléatoires peut permettre de simuler d'autre lois que la loi normale qui ne sont pas directement simulable avec la méthode d'inversion.