KALINKA Quentin

M1 DSS
14/10/2024

Rapport de TP

Simulation de lois de Probabilités

Cours de Mesures et Probabilité Statistique

Mr N'Guessan

Master Data Science en Santé - 1ère année
Année universitaire 2024-2025

Université de Lille
UFR3S – Faculté d'ingénierie et management de la santé (ILIS)
42 rue Ambroise Paré – 59120 LOOS

Université de Lille

Table des matières

Table des Figures

Table des tableaux

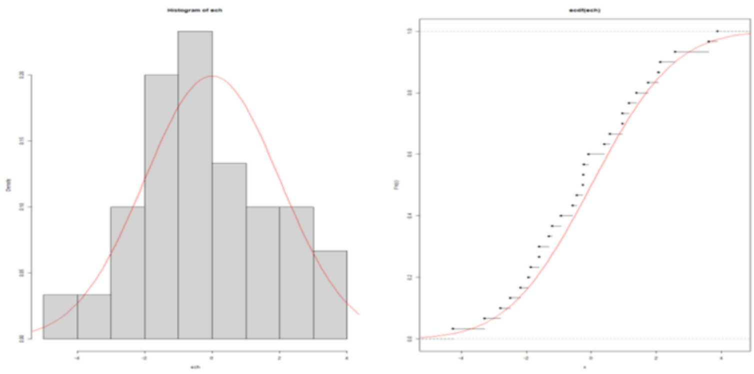
I. Introduction

Dans le logiciel R, on peut simuler très simplement à peu près toutes les lois classiques. Dans ce rapport, on se propose, pour chaque loi usuelle, de simuler des échantillons, et de calculer la moyenne et la variance de ces échantillons, ainsi que faire des graphiques représentant ces lois. Il y a quatre commandes à connaître pour chaque loi : **rmaloi** : permet de simuler selon maloi ; **dmaloi** : permet de calculer la densité de maloi ; **pmaloi** : permet de calculer la fonction de répertition de maloi ; **qmaloi** : permet de calculer le quantile de maloi.

Ce rapport présente une analyse sur les simulations et approximations des lois de probabilités usuelles. Cette étude a été réalisée conformément aux directives et au code fournis dans le travail pratique (*https://fcorset.github.io/TPs/tp2.html*) ayant pour intitulé *« TP 2 : Simulation de lois de probabilités »* [1].

En annexe, il est possible de retrouver le code utilisé lors de la réalisation de ce TP. A noter que seul le code non inclus dans le TP suivi a été ajouté dans ce rapport. L'intégralité du code est disponible dans un dossier associé, sous format '.R'. Ce code est commenté de manière cohérente avec ce rapport, afin de faciliter la suivie et la compréhension de l'ensemble du travail réalisé.

1. Exemple pour la loi normale



(a) - Histogramme de l'échantillon avec courbe de densité théorique

(b) – Graphique de la Fonction de Répartition Empirique

FIGURE 1 - Exemple de représentation graphique pour la loi normale

TABLE 1 - Tableau de comparaison des quantiles de la loi normale

	5%	10%	25%	50%	75%	90%	95%
quantile.echantillon	-2.982250	-2.307173	-1.653392	-0.6537908	0.7997959	1.780225	3.044688
quantile.normale	-3.289707	-2.563103	-1.348980	0.0000000	1.3489795	2.563103	3.289707

Dans cet exemple, nous avons simulé un échantillon de taille n = 30 selon une loi normale d'espérance 0 et d'écart-type 2.

L'histogramme (Figure 1a) met en évidence des fluctuations aléatoires entre la simulation et la densité théorique. Ces écarts apparaissent également sur le graphique de la fonction de répartition empirique (Figure 1b) et le tableau de comparaison des quantiles (Table 1).

Les écarts observées sont attribuables à la taille réduite de l'échantillon. Une augmentation de la taille de l'échantillon n permettrait d'obtenir une approximation plus précise de la distribution théorique (*Loi des grands nombres*) [2].

II. Lois discrètes

1. Loi uniforme discrète

La loi discrète uniforme est une loi de probabilité discrète indiquant une probabilité de se réaliser identique (*équiprobabilité*) à chaque valeur d'un ensemble fini de valeurs possibles.

Soit X une variable aléatoire suivant une loi uniforme discrète sur un ensemble $\{k_1, k_2, \ldots, k_n\}$. La probabilité associée à chaque valeur k_i est donnée par :

$$\forall k \in \{1,\ldots,n\},\ p(X = k_i) = \frac{1}{n} \leftrightarrow X \hookrightarrow U(\{1,\ldots,n\})\text{[3]}.$$

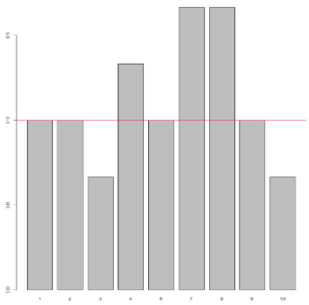


FIGURE 2 - Exemple de représentation graphique d'une variable suivant une loi uniforme discrète

Dans cet exemple, nous avons généré un échantillon de taille $n = 30$ selon une loi uniforme discrète $U(\{1,\ldots,10\})$.

L'histogramme (Figure 2) met en évidence, par la ligne horizontale rouge représentant la probabilité théorique $p(X) = \frac{1}{10}$, que les fréquences observées diffèrent des valeurs théoriques attendues.

Ce résultat illustre un principe statistique fondamental : les échantillons de petite taille tendent à présenter des écarts plus importants par rapport à la distribution théorique, en raison de leur plus grande sensibilité aux fluctuations d'échantillonnages [4]. Ainsi, pour démontrer la validité de la loi uniforme discrète de manière plus précise, une augmentation de la taille de l'échantillon serait nécessaire [2].

2. Loi de Bernoulli

Une épreuve présentant deux issues possibles est appelée une épreuve de Bernoulli, notée $B_{(p)}$.
Les deux issues sont définies comme suit : le succès, noté $X = 1$, se produit avec une probabilité p ; l'échec, noté $X = 0$, se produit avec une probabilité q [5] :

$$p(X = 1) = p \quad \text{et} \quad p(X = 0) = 1 - p = q$$

TABLE 2 - Fréquences observées de l'échantillon simulé par la loi de Bernoulli

```
ech
        0         1
0.6333333 0.3666667
```

Dans cette simulation, nous avons généré un échantillon de taille $n = 30$ suivant une loi de Bernoulli $B(0.3)$. Étant donné que la loi de Bernoulli ne prend que deux valeurs possibles, 0 et 1, les probabilités théoriques associées sont les suivantes :

$$p(\text{Echec} = 0) = 0.7 \qquad p(\text{Succès} = 1) = 0.3$$

En examinant les fréquences observées (Table 2), la fréquence de succès est de 0,37 et celle d'échec est de 0,67. Ces valeurs sont proches des probabilités théoriques attendues, ce qui valide la simulation. Les écarts observés peuvent être attribués à des fluctuations d'échantillonnage dues à la taille de l'échantillon, mais ils restent dans une plage cohérente avec le modèle théorique.

3. Loi binomiale

La loi binomiale est une loi de probabilité définie sur l'ensemble {0,…,n}, qui modélise le nombre de succès obtenus lors de *n* essais indépendants, chacun ayant une probabilité de succès *p*.

Si une variable aléatoire X suit une loi binomiale $B(n,p)$, elle peut prendre les valeurs entières de 0 à n. Pour résumer, soit $Y_i \hookrightarrow B(p)$ pour i = 1,…,n, où les Y_i sont indépendants. On note :

$$X = \sum i = 1nYi \rightsquigarrow B(n,p)$$

Les valeurs possibles pour la variable X sont {0,…,n} et [6] :

$$\forall k \in \{1,\dots,n\},\ p(X = k_i) = C^k{}_n p^k\ (1-p)^{n-k}$$

TABLE 3 - Fréquences observées des échantillon simulé suivant la loi binomiale B(n,p)

ech

5	6	7	8	9	10
0.06	0.16	0.16	0.36	0.15	0.11

(a) – Table [5,10] des fréquences observées de l'échantillon simulé par la loi binomiale de B(10,0.8)

ech

0	1	2	3	4	5	6	7	8
0.01	0.04	0.16	0.23	0.25	0.15	0.10	0.03	0.03

(b) – Table [0,8] des fréquences observées de l'échantillon simulé par la loi binomiale de B(10,0.4)

Dans cet exemple, nous avons simulé un échantillon de taille n = 100 suivant deux loi binomiale : B(10,0.4) et B(10,0.8). Pour chaque distribution, nous avons observé le nombre de succès de 10 essais indépendants répétés 100 fois.

On peut mettre en évidence que pour la simulation B(10,0.4), la fréquence la plus élevée est $p(X = 4) = 0.25$, tandis que pour B(10,0.8), la fréquence la plus élevée est $p(X = 8) = 0.36$. Ces résultats se corrèlent avec la propriété d'espérance de la loi binomiale : $E[X] = np$. En effet :

- Pour B(10,0.4), $E[X] = 10\times0.4 = 4$, ce qui correspond à la valeur la plus fréquente.
- Pour B(10,0.8), $E[X] = 10\times0.8 = 8$, ce qui correspond également à la valeur la plus fréquente.

TABLE 4 - Fréquences théoriques des échantillons simulé suivant la loi binomiale B(n,p)

0	1	2	3	4	5	6	7	8	9	10
0.000	0.000	0.000	0.001	0.006	0.026	0.088	0.201	0.302	0.268	0.107

(a) - Table des fréquences théoriques de l'échantillon de loi binomiale B(10,0.8)

0	1	2	3	4	5	6	7	8	9	10
0.006	0.040	0.121	0.215	0.251	0.201	0.111	0.042	0.011	0.002	0.000

(b) - Table des fréquences théoriques de l'échantillon de loi binomiale B(10,0.4)

En comparant les fréquences théoriques (Table 4) avec les fréquences observées lors des simulations (Table 3), on observe on peut ressortir une concordance générale dans les résultats. Bien que des variations mineures soient présentes, elles peuvent être attribuées à la taille finie de l'échantillon, soit le nombre de répétitions.

En modifiant les paramètres p et n lors de différentes simulations suivant la loi binomiale, plusieurs observations peuvent également être faites :

- Lorsque n est élevé, la loi binomiale tend à devenir symétrique lorsque p est proche de 0,5. En revanche, pour des valeurs p proches de 0 ou de 1, la distribution reste asymétrique.
- Si p est proche de 0, la distribution se concentre vers 0 succès ; si p est proche de 1, elle se concentre vers n succès.

4. Loi géométrique

La loi géométrique, notée *G(p)* avec $0 < p < 1$, est une loi de probabilité discrète qui modélise l'observation du nombre d'épreuves de Bernoulli identiques et indépendantes devant se succéder pour espérer un premier succès [7].

Cette loi correspond au schéma suivant. On répète, de façon indépendante, une même expérience (qui a une probabilité p de réussite et une probabilité $q = 1 - p$ d'échec) et on note X le rang du premier succès.

Les valeurs prises par X sont donc $X(\Omega)=N^\star$ et les probabilités associées sont [8] :

$$\forall k \in \mathbb{N}^\star,\ p_k = P(X = k) = (1-p)^{k-1}p = q^{k-1}p$$

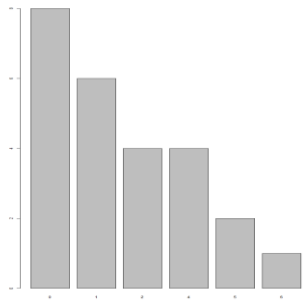


FIGURE 3 - Exemple de représentation graphique d'une variable suivant une loi géométrique

Dans cet exemple, nous avons généré un échantillon de taille $n = 25$ suivant une loi géométrique $G(0.3)$. L'histogramme des fréquences observées (Figure 3) montre une tendance décroissante, correspondant aux fréquences théoriques attendues : $P(X = 0) = 7.5$; $P(X = 1) = 5.25$; $P(X = 2) = 3.675$.

Cette observation met en évidence une propriété de la loi géométrique, disant que la probabilité de succès diminue de manière exponentielle à chaque essai supplémentaire.

5. Loi de Poisson

Cette loi décrit la probabilité qu'un événement se réalise durant un intervalle de temps donné, lorsque la probabilité de réalisation d'un événement est très faible et que le nombre d'essais est très grand.

On dit que X suit une loi de Poisson, notée $P(\lambda)$ avec $\lambda > 0$.

On peut voir cette loi comme la loi des événements rares. Cette loi est une approximation de la loi binomiale, $B(n,p)$ lorsque n est grand et p petit (i.e. pour des événements rares). On pose alors $\lambda = np$. En pratique, on dit que cette approximation est bonne dès que $n > 20$ et $p \leq 0.1$ et $np \leq 5$.

Les valeurs possibles pour X sont dans N et les probabilités associées [9] :

$$\forall k \in \mathbb{N},\ p_k = P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}$$

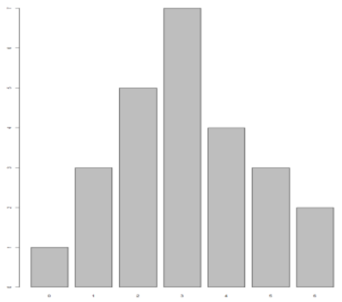


FIGURE 4 - Exemple de représentation graphique d'une variable suivant une loi de Poisson

Dans cet exemple, nous avons généré un échantillon de taille $n = 25$ suivant une loi de Poisson $P(3)$, conformément aux conditions $n > 20$ et $p \leq 0.1$ et $np \leq 5$.

L'histogramme des fréquences observées (Figure 4) montre une concordance avec les fréquences théoriques attendues :

- $P(X = 3) = \frac{3^3 e^{-3}}{3\,!} \approx 0.224$
- $P(X = 4) = \frac{3^4 e^{-3}}{4\,!} \approx 0.168$

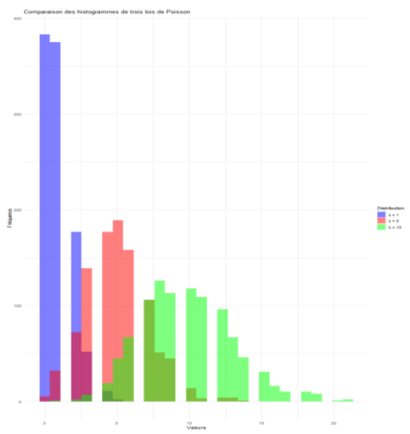


FIGURE 5 - Représentation graphique simulant différentes lois de Poisson en faisant varier λ

Lors de la variation du paramètre λ de la loi de Poisson, une influence significative sur la forme et les propriétés de la distribution est observée (Figure 5). En effet, λ, représentant l'espérance de la loi de Poisson, détermine à la fois la moyenne et la variance de la distribution, ce qui influe sur la probabilité d'observer un certain nombre d'événements dans un intervalle donné.

Lorsque λ augmente, la distribution présente une augmentation du nombre d'événements observés, accompagnée d'une variabilité accrue des résultats. Cela se traduit par une concentration plus marquée des probabilités autour des valeurs élevées, indiquant une fréquence d'occurrence plus élevée des événements.

III. Lois continues

1. Loi uniforme sur [0,1], notée $U_{[0,1]}$

Une variable aléatoire X suit la loi uniforme sur l'intervalle [0,1], notée $X \sim U([0,1])$, si et seulement si, pour tout intervalle I inclus dans [0,1], la probabilité de l'événement « $x \in I$ » est l'aire de l'ensemble des points $M(x, y)$ tels que $x \in I$ et $0 \leq y \leq 1$ [10].

On dit qu'une X suit une loi uniforme sur [0,1] si et seulement si sa densité est la fonction définie sur [0,1] :

$$\forall x \in \mathbb{R},\ f(x) = \begin{cases} 1 & \text{si } x \in [0,1] \\ 0 & \text{sinon} \end{cases}$$

La fonction de répartition d'une loi uniforme sur [0,1] est :

$$\forall x \in \mathbb{R},\ F(x) = \int_{-\infty}^{x} f(t)\,dt = \begin{cases} 0 & \text{si } x < 0 \\ x & \text{si } x \in [0,1] \\ 1 & \text{si } x > 1 \end{cases}$$

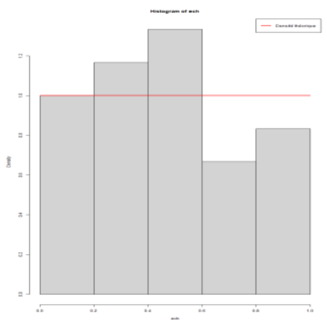


FIGURE 6 - Exemple de représentation graphique d'une variable suivant une loi uniforme sur [0,1]

Dans cet exemple, nous avons généré un échantillon de taille $n = 30$ suivant une loi uniforme sur [0,1].

L'histogramme (Figure 6) met en évidence, par la ligne horizontale rouge représentant la probabilité théorique, que les fréquences observées diffèrent légèrement des valeurs théoriques attendues. Cette ligne souligne l'hypothèse d'équiprobabilité de chaque issue, indiquant que la probabilité de chaque résultat est identique dans le modèle théorique.

Pour démontrer la validité de la loi uniforme continue de manière plus précise, une augmentation de la taille de l'échantillon serait nécessaire.

2. Loi uniforme sur [a,b], notée $U_{[a,b]}$, avec (a < b)

On dit que la variable aléatoire X suit une loi uniforme sur un segment $[a,b]$ si sa densité de probabilité $f(x)$ est une constante C sur $[a,b]$ et est nulle en dehors du segment $[a,b]$.

On dit que $X \hookrightarrow U[a,b]$ si et seulement si sa fonction densité est [11] :

$$\forall x \in \mathbb{R},\ f(x) = \begin{cases} \dfrac{1}{b-a} & \text{si } x \in [a,b] \\ 0 & \text{sinon} \end{cases}$$

La fonction de répartition d'une loi uniforme sur [a,b] est :

$$\forall x \in \mathbb{R},\ F(x) = \int_{-\infty}^{x} f(t)\,dt = \begin{cases} 0 & \text{si } x < a \\ \dfrac{x-a}{b-a} & \text{si } x \in [a,b] \\ 1 & \text{si } x > b \end{cases}$$

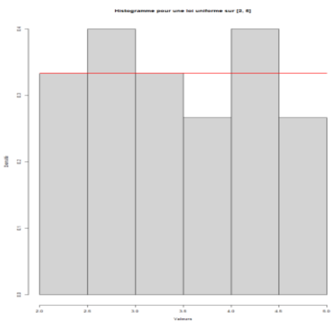


FIGURE 7 - Exemple de représentation graphique d'une variable suivant une loi uniforme sur [2,5]

Dans cet exemple, nous avons généré un échantillon de taille $n = 30$ suivant une loi uniforme sur [2,5].

L'histogramme (Figure 6) montre que les fréquences observées convergent vers la densité théorique de la loi uniforme sur [2,5], représentée par la ligne horizontale rouge. Cette densité théorique est donnée par $f(x) = \frac{1}{5-2} = \frac{1}{3}$, indiquant une distribution uniforme des probabilités sur cet intervalle.

Cependant, des variations mineures sont présentes dans les fréquences observées, dues à la taille limitée de l'échantillon.

3. Loi normale centrée réduite

On dit qu'une variable aléatoire X suit la loi normale centrée réduite, ce que l'on note $X \hookrightarrow N(0,1)$, si elle admet comme fonction densité de X [12] :

$$\forall x \in \mathbb{R},\ f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Une telle variable aléatoire X admet alors une espérance $E(X) = 0$ et une variance $V(X) = 1$.

Le premier paramètre de la loi normale est l'espérance de la loi et le second correspond à la variance de la loi. On peut remarquer que cette fonction est paire, i.e. $\forall x \in R,\ f(-x) = f(x)$, ce qui implique une symétrie par rapport à l'axe des ordonnées (Figure 8).

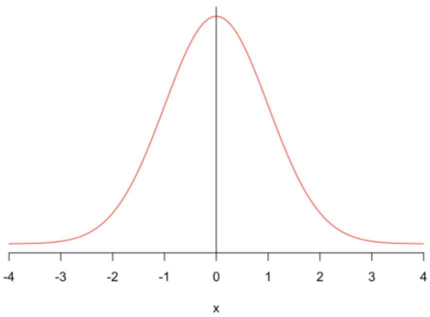


FIGURE 8 - Densité de la loi Normale centrée réduite

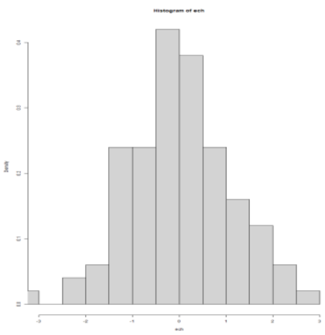


FIGURE 9 - Exemple de représentation graphique d'une variable suivant une loi normale centrée réduite

Dans cet exemple, nous avons généré un échantillon de taille $n = 100$ selon une loi normale centrée réduite $N(0,1)$.

L'histogramme des fréquences observées (Figure 9) montre une concordance proche avec les fréquences théoriques attendues :

- $f(1) = f(-1) = \frac{1}{\sqrt{2\pi}}\, e^{\frac{(-1)^2}{2}} \approx 0.242$
- $f(2) = f(-2) = \frac{1}{\sqrt{2\pi}}\, e^{\frac{(-2)^2}{2}} \approx 0.054$

Ces valeurs sont proches des probabilités théoriques attendues, ce qui valide la simulation. Les écarts observés peuvent être attribués à des fluctuations d'échantillonnage dues à la taille de l'échantillon, mais ils restent dans une plage cohérente avec le modèle théorique.

4. Loi normale $N(\mu, \sigma^2)$

On dit que $X \hookrightarrow N(\mu,\sigma 2)$ si sa fonction densité, ayant comme représentation graphique une courbe en cloche, est :

$$\forall x \in \mathbb{R},\ f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

On pourra vérifier que l'intégrale de cette fonction sur R vaut bien 1. μ représente l'espérance de la loi et σ^2 la variance.

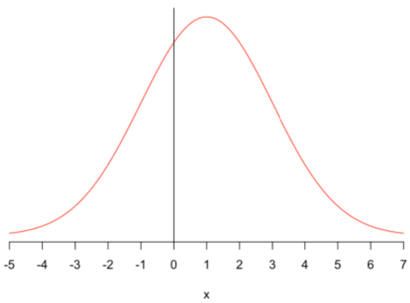


FIGURE 10 - Densité de la loi N(1,4)

Dans cet exemple (Figure 10), nous avons généré une courbe de densité d'une loi normale $N(1,4)$, qui est une distribution normale avec une moyenne $\mu = 1$ et une variance $\sigma^2 = 4$.

Les distributions normales $N(0,1)$ et $N(1,4)$ diffèrent principalement en termes de leur moyenne et de leur variance, ce qui affecte la position et l'étalement de leurs courbes de densité : $N(0,1)$ présente un pic centré à $x = 0$, avec une courbe étroite ; $N(1,4)$ présente un pic décalé *à* $x = 4$, avec une courbe plus large.

$$\textbf{Théorème : Si } \mathbf{X} \rightsquigarrow \mathcal{N}(\mu, \sigma^2) \textbf{ alors } \frac{\mathbf{X}-\mu}{\sigma} \rightsquigarrow \mathcal{N}(\mathbf{0}, \mathbf{1})$$

Le théorème indique que, pour une variable $X \sim N(\mu,\sigma^2)$, on peut obtenir une loi normale centrée réduite en utilisant la transformation standardisée $\frac{X-\mu}{\sigma} \sim N(0,1)$. Cela signifie que toute distribution normale peut être transformée en une distribution normale standard, avec une moyenne de 0 et une variance de 1.

IV. Exercices sur la loi normale

1. Distribution des valeurs

Pour simuler *n* valeurs aléatoirement suivant la loi normale de moyenne *m* et d'écart-type *s* on utilise sous R la commande **rnorm**.

Pour chacun des échantillons générés *N(0,1)*, *N(1,1)*, *N(2,1)*, *N(0,2)*, *N(0,4)*, les résultats ont été compilés dans un tableau (Table 5).

TABLE 5 - Distribution des valeurs pour n = 20

N = 20	**m = 0 et s = 1**	**m = 1 et s = 1**	**m = 2 et s = 1**	**m = 0 et s = 2**	**m = 0 et s = 4**
< -2	1	0	0	4	5
< 0	11	3	0	7	5
= 0	0	0	0	0	0
> 2	1	0	10	3	5

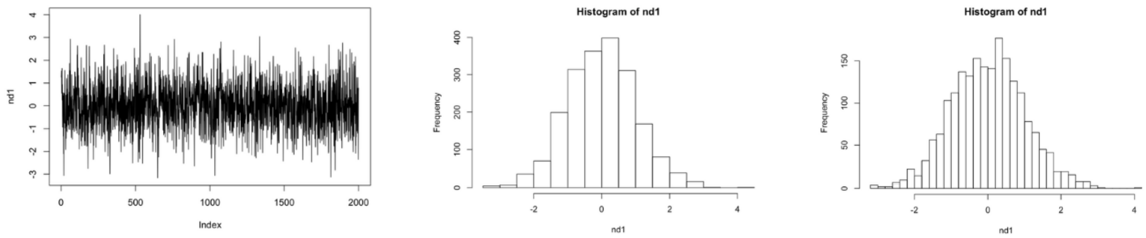
a) Question 1

Avec les mêmes valeurs des moyennes et des écart-types, on simule maintenant des échantillons de taille n = 2000 individus. Pour chacun des échantillons générés les résultats ont été compilés dans un tableau (Table 6).

TABLE 6 - Distribution des valeurs pour n = 2000

n= 2000	**m = 0 et s = 1**	**m = 1 et s = 1**	**m = 2 et s = 1**	**m = 0 et s = 2**	**m = 0 et s = 4**
< -2	(nb= 54) 2.70%	(nb= 5) 00.25%	(nb= 0) 0%	(nb= 308) 15.40%	(nb= 614) 30.70%
< 0	(nb= 997) 49.85%	(nb= 313) 15.65%	(nb= 30) 1.50%	(nb= 969) 48.45%	(nb= 992) 49.60%
= 0	(nb= 0) 0%	(nb= 0) 0%	(nb= 0) 0%	(nb= 0) 0%	(nb= 0) 0%
> 2	(nb= 0) 2.25%	(nb= 0) 15.85%	(nb= 0) 50.60%	(nb= 0) 16.35%	(nb= 0) 30.90%

b) Question 2



(a) Graphique en ligne *(b) Histogramme 1* *(c) Histogramme 2*

FIGURE 11 - Exemple de représentation graphique pour l'échantillon nd1

La fonction *plot()* trace un graphique en ligne (Figure a) des valeurs contenues dans l'objet *nd1*. Lorsque le graphique est tracé, les axes *x* et *y* correspondent aux éléments suivants : *x* représente les indices des valeurs dans le vecteur *(1,2,3,…,n) ; y* représente les valeurs elles-mêmes du vecteur *nd1*.

L'histogramme, avec un nombre de segments ajusté via la fonction *`breaks()`*, permet une interprétation plus précise des distributions par rapport au nuage de points. Pour l'échantillon étudié, un histogramme à 50 segments (Figure c) offre une meilleure approximation d'une distribution normale, améliorant ainsi la clarté et la précision de l'analyse des variations de valeurs.

c) Question 3

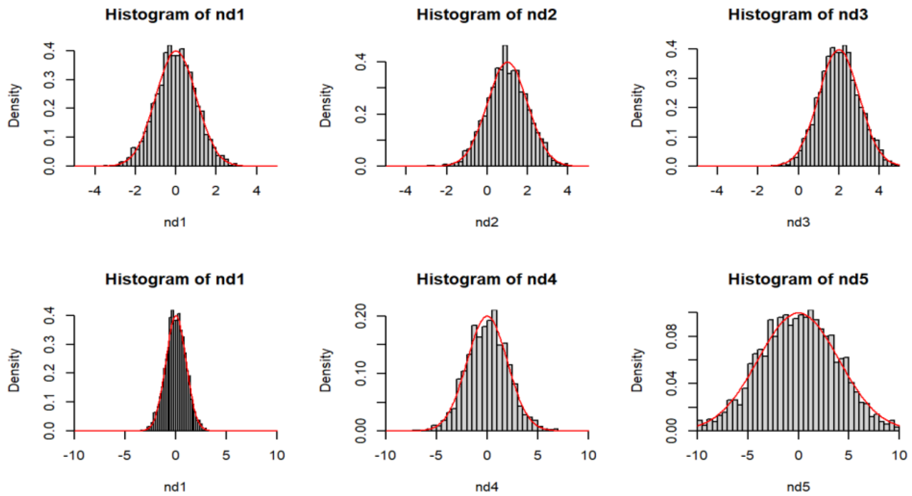


FIGURE 12 - Histogrammes des 5 échantillons

<u>Comparaison des courbes *nd1*, *nd2*, *nd3* :</u>

L'analyse des courbes *nd1*, *nd2*, et *nd3* révèle une variation minime de la densité entre elles. Toutefois, il est notable que l'augmentation de la moyenne μ entraîne un déplacement vers la droite de la courbe correspondante.

<u>Comparaison des courbes nd1,nd4,nd5 :</u>

Pour les courbes *nd1*, *nd4*, et *nd5*, il est observé qu'elles demeurent centrées autour de 0, en raison de l'égalité de leurs moyennes. Néanmoins, on constate que l'augmentation de l'écart-type σ^2 se traduit par un élargissement de la courbe, ce qui accentue son amplitude sur l'axe des abscisses.

d) Question 4

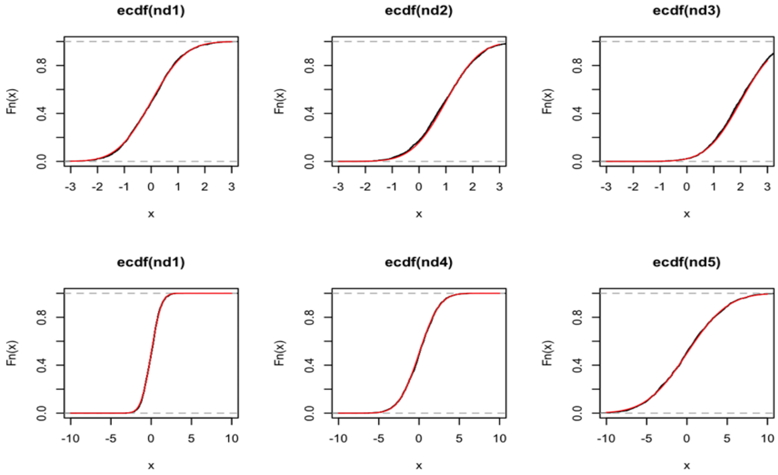


FIGURE 13 - Fonctions de répartition des 5 échantillons

<u>Comparaison des courbes *nd1*, *nd2*, *nd3* :</u>
L'analyse des courbes *nd1*, *nd2*, et *nd3* révèle un décalage des distributions entre elles. Plus précisément, l'augmentation de la moyenne μ entraîne un déplacement vers la droite des distributions correspondantes. Cela indique que chaque distribution reste symétrique autour de sa moyenne respective, mais que les moyennes μ croissantes se traduisent par un décalage progressif des courbes sur l'axe des abscisses.

<u>Comparaison des courbes *nd1*, *nd4*, *nd5* :</u>
Pour les courbes *nd1*, *nd4*, et *nd5*, il est observé qu'elles demeurent centrées autour de 0, en raison de l'égalité de leurs moyennes μ. Cependant, l'augmentation de l'écart-type σ^2 entre les courbes entraîne un élargissement progressif de la courbe, ce qui accentue sa dispersion sur l'axe des abscisses. Cela illustre que, bien que les distributions aient la même moyenne μ, un écart-type σ^2 plus grand augmente la dispersion et la variabilité des valeurs autour de cette moyenne.

e) Question 5

On veut montrer le lien entre la densité et la fonction de répartition.

La densité de la loi normale de moyenne $\mu = 0$ et de variance $\sigma^2 = 1$, que l'on note $X \hookrightarrow N(0,1)$, est donné par la formule suivante :

$$\forall x \in \mathbb{R}, f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Le lien entre la densité et la fonction de répartition est donné de la manière suivante [13] :

$$f(x) = \frac{1}{s\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2s^2}}$$

$$F(t) = P(X \leq t) = \int_{-\infty}^{t} f(x)\, dx$$

La valeur au point t de la fonction de répartition (théorique) $F(t)$ d'une loi gaussienne de moyenne m et d'écart-type s est donnée par la commande *pnorm(t,m,s)* de *R*.

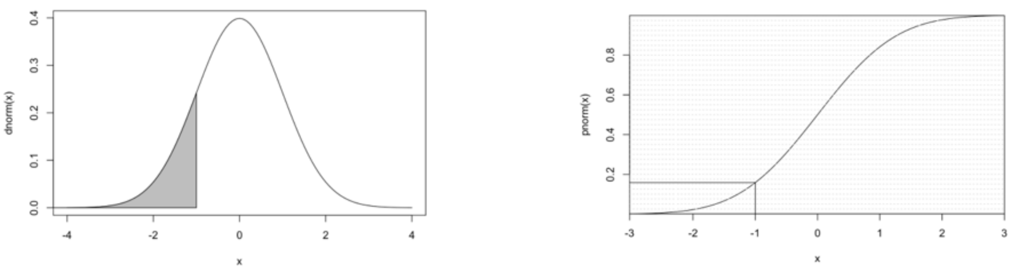


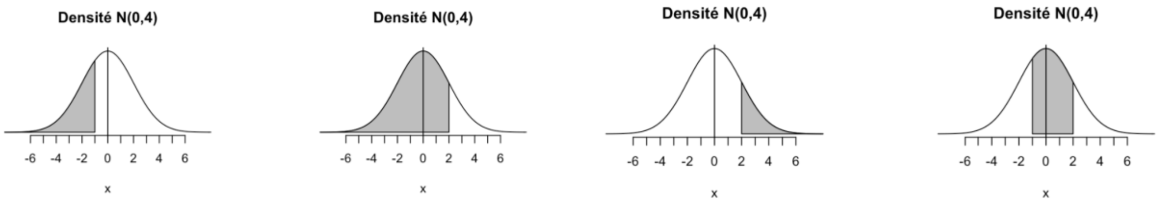
FIGURE 14 - Courbe de densité et Courbe cumulative simulé selon le script R associé au TP

Le script R associé permet de visualiser et de calculer la probabilité qu’une variable normale standard soit inférieure à -1.

Ce script montre visuellement le lien entre la densité et la fonction de répartition en coloriant l'aire sous la courbe de la densité, qui correspond à la probabilité cumulée. La fonction de répartition *F(t)* à un point t correspond précisément à l'intégrale de la densité de probabilité de -∞ à t, soit l'aire sous la courbe jusqu'à ce point.

Le script permet ainsi de visualiser la probabilité cumulée et de comparer cette valeur théorique avec l'aire sous la courbe de densité. Cela met en évidence le rôle crucial de l'intégrale dans le calcul de la fonction de répartition et illustre comment cette dernière est directement reliée à la densité de probabilité.

f) Question 6



Aire de la figure	Proba représentée par l’aire	Ecriture avec *F(t)*	Valeur numérique *pnorm*
1	*P(X < -1)*	*F(-1)*	*pnorm(-1) ≈ 0.3085*
2	*P(X < 2)*	*F(2)*	*pnorm(2) ≈ 0.8413*
3	*P(X > 2)*	1 - *F(2)*	*1 – pnorm(2) ≈ 0.1587*
4	*P(-1 < X < 2)*	*F(2) - F(-1)*	*pnorm(2) - pnorm(-1) ≈ 0.5328*

Soit X une variable aléatoire gaussienne de moyenne 15 et de variance 9.

- $P(X \leq 17) = pnorm(\frac{17-15}{3}) = pnorm(\frac{2}{3}) \approx 0.5879$
- $P(X > 7) = 1 - P(X \leq 7) = pnorm(\frac{7-15}{3}) = 1 - pnorm(-\frac{8}{3}) \approx 0.8129$
- $P(7 < X < 15) = P(X < 15) - P(X \leq 7) = pnorm(\frac{15-15}{3}) - pnorm(-\frac{8}{3}) \approx 0.3129$

V. Lien entre lois normales

Soit X une variable aléatoire distribuée suivant la loi normale d'espérance 0 et d'écart-type 1. Les variables $X_1 = X+1$, $X_2 = X-1$, $X_3 = 2X$ et $X_4 = X/2$, où $X \sim N(0,1)$, suivent aussi une loi normale comme l'illustre les simulations réalisées (Figure 15) :

- $X_1 = X + 1$ suit une loi normale d'espérance $0 + 1 = 1$ et d'écart-type 1, donc $X_1 \sim N(1,1)$
- $X_2 = X - 1$ suit une loi normale d'espérance $0 - 1 = -1$ et d'écart-type 1, donc $X_2 \sim N(-1,1)$
- $X_3 = 2X$ suit une loi normale d'espérance 0 et d'écart-type 2, donc $X_3 \sim N(0,2^2) = N(0,4)$
- $X_4 = X/2$ suit une loi normale d'espérance 0 et d'écart-type ½, donc $X4 \sim N(0,(½)^2) = N(0,0.25)$

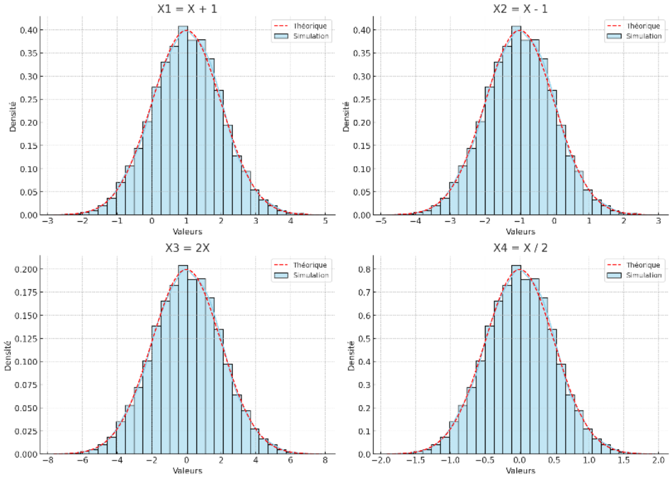


FIGURE 15 - Distribution des variables X_1, X_2, X_3, X_4

Dans cet exemple, nous avons généré un échantillon de taille $n = 10\,000$ selon la distribution des variables X_1, X_2, X_3 et X_4.

Les courbes de distribution obtenues (Figure 15) présentent des concordances proches avec les distributions théoriques attendues, indiquant que ces variables suivent approximativement une loi normale.

Soit maintenant Y une variable aléatoire de moyenne 4 et de variance 3, $Y(4,3)$, alors :

- $Z = \frac{Y-2}{\sqrt{3}} \sim N(\frac{2}{\sqrt{3}}, 1)$

  où $E[Z] = E[\frac{Y-2}{\sqrt{3}}] = \frac{E[Y-2]}{\sqrt{3}} = \frac{2}{\sqrt{3}}$; $Var[Z] = Var(\frac{Y-2}{\sqrt{3}}) = \left(\frac{1}{\sqrt{3}}\right)^2 Var(Y-2) = \frac{1}{3}.3 = 1$

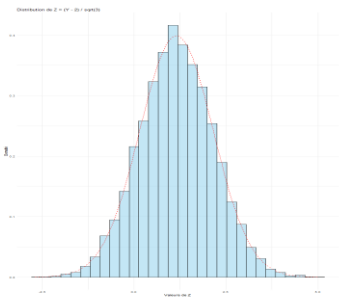


FIGURE 16 - Distribution de la variable Z

Dans cet exemple, nous avons généré un échantillon de taille $n = 10\,000$ selon la distribution de la variable Z.

La courbe de distribution obtenue (Figure 16) présente une concordance proche avec la distribution théorique attendue, indiquant que la variable Z suit approximativement une loi normale.

VI. Sommes de variables aléatoires gaussiennes

En se basant sur les résultats obtenus via les différentes simulations (Figure 17), *si* $X \hookrightarrow N(m_1,\sigma^2_1)$ *et* $Y \hookrightarrow N(m_2,\sigma^2_2)$, alors :

- $X + Y \rightsquigarrow N(m_1 + m_2,\ \sigma^2_1 + \sigma^2_2)$
- $X + Y \rightsquigarrow N(m_1 - m_2,\ \sigma^2_1 + \sigma^2_2)$

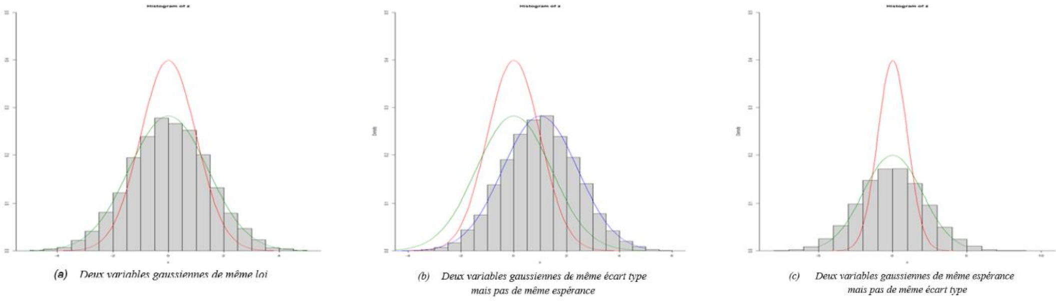


FIGURE 17 - Représentation graphique de variables aléatoires gaussiennes

Les variables X_i étant toutes des variables aléatoires gaussiennes centrées réduites, elles suivent toutes la loi $N(0,1)$. Ainsi, chaque $X_i \sim N(0,1)$, ce qui signifie que $\mu_i = 0$ et $\sigma^2_i = 1$ pour tout i.

La variable $\bar{X}$ est définie comme la moyenne de n variables gaussiennes centrées réduites, dont l'espérance $E[\bar{X}]$ et de variance $Var(\bar{X})$:

$$\mathbb{E}[\overline{X}] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \frac{1}{n}\sum_{i=1}^{n} 0 = 0 \qquad \mathrm{Var}(\overline{X}) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) = \frac{1}{n^2}\sum_{i=1}^{n} 1 = \frac{1}{n^2}\cdot n = \frac{1}{n}$$

Ainsi, on peut conclure que :

$$\bar{X} = \frac{1}{n}(X_1 + X_2 + X_3 + .. + X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i \quad o\grave{u}\ \bar{X} \sim N(0,\frac{1}{n})$$

VII. Somme de carrés de variables gaussiennes centrées réduites

Si X_1, X_2, …, X_n sont des variables aléatoires indépendantes et identiquement distribuées suivant une loi normale standard *N(0,1)*, alors la somme des carrées de ces variables suit une loi du khi-deux avec *n* degrés de liberté.

La formule peut être complété alors de la manière suivante :

$$X_1^2 + X_2^2 + X_3^2 + \ldots + X_n^2 \sim X^2(n)$$

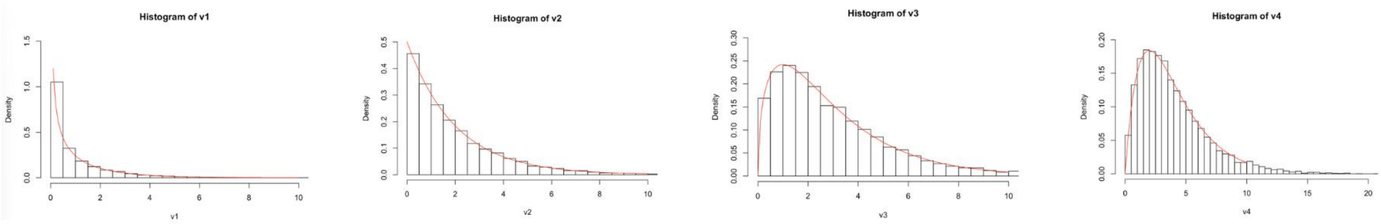


FIGURE 18 - Représentation graphique de somme de carrés de variables gaussiennes centrées réduites

VIII. Exercice

Pour répondre à ces différentes questions, nous allons utiliser les propriétés de la loi normale. Étant donné que la taille X des hommes suit une loi normale de moyenne μ = 172 cm et de variance σ^2=196 cm^2, nous pouvons alors en déduire que l'écart-type σ est $\sigma = \sqrt{196} = 14$ cm.

- Probabilité que X soit inférieur à 160 cm ?

$$Z = \frac{X - \mu}{\sigma} = \frac{160 - 172}{14} = \frac{-12}{14} \approx -0.857 \quad ; \quad P(X < 160) = P(Z < -0.857) \approx 0.1967 \; (ou\ 19.67\ \%)$$

- Probabilité qu'un homme tiré au hasard mesure plus de deux mètres ?

$$Z = \frac{200 - 172}{14} = \frac{28}{14} = 2 \quad ; \quad P(X > 200) = P(Z > 2) \approx 0.0228 \; (ou\ 2.28\ \%)$$

- Calcul de *P(165 < X < 185)*

$$Z_{165} = \frac{165 - 172}{14} \approx -0.5 \quad et \quad Z_{185} = \frac{185 - 172}{14} \approx 0.9286$$

$$P(165 < X < 185) = P(Z < 0.9286) - P(Z < -0.5) \approx 0.5149 \; (ou\ 51.49\ \%)$$

- Probabilité qu'un homme soit plus grand qu'une femme choisie au hasard

$$Y = X_{homme} - X_{femme} \rightarrow \mu_Y = \mu_h - \mu_f = 172 - 162 = 10cm \; ; \; \sigma_Y = \sqrt{\sigma_h + \sigma_f} = \sqrt{196 + 144} \approx 18.44\ cm$$

$$P(Y > 0) = P\left(Z > \frac{0 - \mu_Y}{\sigma_Y}\right) = P\left(Z > \frac{0 - 10}{18.44}\right) \quad ; \quad P(Y) \approx 0.7062 \; (ou\ 70.62\ \%)$$

IX. Loi Exponentielle

On dit que X suit une loi exponentielle de paramètre $\lambda > 0$, ce que l'on note $X \hookrightarrow E(\lambda)$ si et seulement si sa fonction densité est :

$$\forall x \in \mathbb{R},\ f(x) = \begin{cases} \lambda \exp(-\lambda x) & \text{si } x \geq 0 \\ 0 & \text{sinon} \end{cases}$$

Pour vérifier que f est bien une fonction densité, il faut vérifier les deux conditions suivantes :

- Non-négativité

$$pour\ x \geq 0, f(x) = \lambda . e^{-\lambda x} \geq 0 \quad ; \quad pour\ x < 0, f(x) = 0 \quad ; \quad donc\ \forall x, f(x) \geq 0$$

- Intégration à 1

$$\int_{-\infty}^{+\infty} f(x)\, dx = \int_{0}^{+\infty} \lambda . e^{-\lambda x}\, dx = 1 \quad ; \quad \int_{0}^{-\infty} e^{-\lambda x} (-\frac{1}{\lambda})\, dx = 1$$

Puisque *f(x)* est non-négatif pour tout x et que son intégrale sur R est égale à 1, il est possible de conclure que *f(x)* est bien une fonction densité de probabilité pour la loi exponentielle de paramètre $\lambda > 0$.

Bibliographie

[1] « TP 2 : Simulation de lois de probabilités ». [En ligne]. Disponible sur: https://fcorset.github.io/TPs/tp2.html
[2] « Loi des grands nombres ».[En ligne]. Disponible sur: https://www.bibmath.net/dico/index.php?action=affiche&quoi=./l/lgn.html
[3] « Loi uniforme discrète ». [En ligne]. Disponible sur: https://bibmath.net/dico/index.php?action=affiche&quoi=./l/loiunifdisc.html
[4] « Stat_Cours5.pdf ». [En ligne]. Disponible sur: https://ernst.r.perso.math.cnrs.fr/cours/Stat_Cours5.pdf
[5] « Loi de Bernoulli ». [En ligne]. Disponible sur: https://www.bibmath.net/dico/index.php?action=affiche&quoi=./l/loibernoulli.html
[6] « Loi binomiale ». [En ligne]. Disponible sur: https://www.bibmath.net/dico/index.php?action=affiche&quoi=./l/loibinomiale.html
[7] « Loi géométrique », *Wikipédia*. 9 juin 2024. [En ligne]. Disponible sur: https://fr.wikipedia.org/w/index.php?title=Loi_g%C3%A9om%C3%A9trique&oldid=215823436
[8] « Loi géométrique ». [En ligne]. Disponible sur: https://www.bibmath.net/dico/index.php?action=affiche&quoi=./l/loigeometrique.html
[9] « Loi de Poisson ». [En ligne]. Disponible sur: https://www.bibmath.net/dico/index.php?action=affiche&quoi=./l/loipoisson.html
[10] « 332.pdf ». [En ligne]. Disponible sur: https://www.mathemathieu.fr/component/attachments/download/332
[11] « Loi uniforme continue ». [En ligne]. Disponible sur: https://www.bibmath.net/dico/index.php?action=affiche&quoi=./l/loiunifcont.html
[12] « Loi normale ». [En ligne]. Disponible sur: https://www.bibmath.net/dico/index.php?action=affiche&quoi=./l/loinormale.html
[13] « Variables aléatoires à densité ». [En ligne]. Disponible sur: https://www.bibmath.net/ressources/index.php?action=affiche&quoi=capes/cours/vadensite.html

Annexes

```r

# Introduction
##########################################################

# Lois discretes
##########################################################

## 1. Loi uniforme discrete
## 2. Loi de Bernoulli
## 3. Loi binomiale
## 4. Loi geometrique
## 5. Loi de poisson

# Simulez différentes lois de Poisson en faisant varier λ ainsi que la taille de l'échantillon
lambda1 <- 1
lambda2 <- 5
lambda3 <- 10
sample1 <- rpois(1000, lambda1)
sample2 <- rpois(1000, lambda2)
sample3 <- rpois(1000, lambda3)

data <- data.frame(
  value = c(sample1, sample2, sample3),
  group = factor(rep(c(paste("λ =", lambda1),
                       paste("λ =", lambda2),
                       paste("λ =", lambda3)),
                     each = 1000),
                 levels = c(paste("λ =", lambda1),
                            paste("λ =", lambda2),
                            paste("λ =", lambda3)) ))

ggplot(data, aes(x = value, fill = group)) +
  geom_histogram(position = "identity", alpha = 0.5, bins = 30) +
  labs(title = "Comparaison des histogrammes de trois lois de Poisson",
       x = "Valeurs",
       y = "Fréquence",
       fill = "Distribution") +
  scale_fill_manual(values = c("blue", "red", "green")) +
  theme_minimal()

# Lois continues
##########################################################

## Loi uniforme sur [0,1], notee U[0,1]
## Loi uniforme sur [a,b], notee U[a,b] avec (a<b)
## Loi normale centree reduite, notee N(0,1)
## Loi normale N(u,o^2)


# Exercices sur la loi normale
##########################################################

## Distribution des valeurs
## Question 1
## Question 2
## Question 3
## Question 4
## Question 5
## Question 6

moyenne <- 15 # Définition moyenne
sigma <- 9 # Définition sigma
prob_X_le_17 <- pnorm(17, mean = moyenne, sd = sigma) # Probabilité que X <= 17
prob_X_gt_7 <- 1 - pnorm(7, mean = moyenne, sd = sigma) # Probabilité que X > 7
prob_7_lt_X_lt_15 <- pnorm(15, mean = moyenne, sd = sigma) - pnorm(7, mean = moyenne, sd = sigma) # Probabilité que 7 < X < 15
# Résultats
prob_X_le_17
prob_X_gt_7
prob_7_lt_X_lt_15

# Liens entre lois normales
##########################################################

## Partie 1
n <- 10000  # Nombre de simulations
set.seed(123)
X <- rnorm(n, mean = 0, sd = 1) # Génération de X ~ N(0, 1)

X1 <- X + 1
X2 <- X - 1
X3 <- 2 * X
X4 <- X / 2

# Données pour le graphique
data <- data.frame(
  value = c(X1, X2, X3, X4),
  variable = factor(rep(c("X1 = X + 1", "X2 = X - 1", "X3 = 2X", "X4 = X / 2"), each = n)))

# Tracé des graphiques de densité avec une boucle
ggplot(data, aes(x = value)) +
  geom_histogram(aes(y = after_stat(density)), bins = 30, color = "black", fill = "skyblue", alpha = 0.5) +
  facet_wrap(~ variable, scales = "free") +
  # Fonction pour X1
  stat_function(data = subset(data, variable == "X1 = X + 1"), aes(x = value),
                fun = dnorm, args = list(mean = 1, sd = 1), color = "red", linetype = "dashed") +
  # Fonction pour X2
  stat_function(data = subset(data, variable == "X2 = X - 1"), aes(x = value),
                fun = dnorm, args = list(mean = -1, sd = 1), color = "red", linetype = "dashed") +
  # Fonction pour X3
  stat_function(data = subset(data, variable == "X3 = 2X"), aes(x = value),
                fun = dnorm, args = list(mean = 0, sd = 2), color = "red", linetype = "dashed") +
  # Fonction pour X4
  stat_function(data = subset(data, variable == "X4 = X / 2"), aes(x = value),
                fun = dnorm, args = list(mean = 0, sd = 0.5), color = "red", linetype = "dashed") +
  labs(title = "Comparaison des distributions simulées et théoriques",
       x = "Valeurs", y = "Densité") +
  theme_minimal() +
  theme(legend.position = "none")

## Partie 2
n <- 10000  # Nombre de simulations
set.seed(123)  # Pour la reproductibilité
Y <- rnorm(n, mean = 4, sd = sqrt(3))  # Génération de Y ~ N(4, 3)

Z <- (Y - 2) / sqrt(3) # Calcul de la variable Z

# Données pour le graphique
data <- data.frame(value = Z)

# Tracé de l'histogramme pour Z
ggplot(data, aes(x = value)) +
  geom_histogram(aes(y = after_stat(density)), bins = 30, color = "black", fill = "skyblue", alpha = 0.5) +
  # Fonction pour Z
  stat_function(aes(x = value),
                fun = dnorm, args = list(mean = 2 / sqrt(3), sd = 1), color = "red", linetype = "dashed") +
  labs(title = "Distribution de Z = (Y - 2) / sqrt(3)",
       x = "Valeurs de Z", y = "Densité") +
  theme_minimal()
```

```r

# Sommes de variables aleatoires gaussiennes
##########################################################

old_par <- par()
par(xaxs = "i", yaxs = "i")

# Deux variables gaussiennes de meme loi
x <- rnorm(10000, 0, 1)
y <- rnorm(10000, 0, 1)

z <- x + y
hist(z, breaks = 20, freq = FALSE, ylim = c(0, 0.5))
curve(dnorm, from = -4, to = 4, add = TRUE, col = "red") # la loi N(0,1)

x <- seq(-5,5,0.1)
lines(x, dnorm(x, 0, sqrt(2)), col = "green4") # la loi proposee

# Deux variables gaussiennes de meme ecart type mais pas de meme esperance
x <- rnorm(10000, 0, 1)
y <- rnorm(10000, 1, 1)

z <- x + y
hist(z, breaks = 20, freq = FALSE, ylim = c(0,0.5))
curve(dnorm, from = -4, to = 4, add = TRUE, col = "red") # la loi N(0,1)

x <- seq(-5, 5, 0.1)
lines(x, dnorm(x, 0, sqrt(2)), col = "green4") # la loi proposee
lines(x, dnorm(x, 1, sqrt(2)), col = "blue")

# Deux variables gaussiennes de meme esperance mais pas de meme ecart type
x <- rnorm(10000, 0, 1)
y <- rnorm(10000, 0, 2)

z <- x + y
hist(z, breaks = 20, freq = FALSE, ylim = c(0, 0.5))
curve(dnorm, from = -4, to = 4, add = TRUE, col = "red")

x <- seq(-5, 5, 0.1)
lines(x, dnorm(x, 0, sqrt(4)), col = "green4") # la loi proposee


## Ecrire un script pour expliquer la loi d'une moyenne de lois gaussiennes

# Définir les paramètres
mu <- 5    # Espérance de la distribution gaussienne
sigma <- 2    # Ecart-type de la distribution gaussienne
n <- 30    # Taille de chaque échantillon
repeats <- 1000    # Nombre de répétitions pour estimer la distribution de X_bar

# Générer les échantillons de la moyenne de n variables gaussiennes
set.seed(123)    # Pour reproductibilité
sample_means <- replicate(repeats, mean(rnorm(n, mean = mu, sd = sigma)))

# Tracer l'histogramme de la distribution de X_bar
hist(sample_means, breaks = 50, probability = TRUE, main = expression(paste("Distribution de ", bar(X), " avec n = ", n)),
     xlab = expression(bar(X)), col = "lightgreen", border = "black")

# Superposer la densité de la loi normale attendue N(mu, sigma^2/n)
x_vals <- seq(min(sample_means), max(sample_means), length = 100)
y_vals <- dnorm(x_vals, mean = mu, sd = sigma / sqrt(n))
lines(x_vals, y_vals, col = "blue", lwd = 2)
legend("topright", legend = c("Histogramme des moyennes", expression(paste("Densité de ", N(mu, sigma^2/n)))),
       col = c("lightgreen", "blue"), lwd = 2, bty = "n")

# Sommes de carres de variables gaussiennes centrees reduites
##########################################################

# Exercices
##########################################################

# Parametres pour la loi normale des hommes
mu_men <- 172 # Moyenne
sigma_men <- 14 # Écart-type

# 1. Probabilité que X < 160 cm
prob_1 <- pnorm(160, mean = mu_men, sd = sigma_men)

# 2. Probabilité que X > 200 cm
prob_2 <- 1 - pnorm(200, mean = mu_men, sd = sigma_men)

# 3. Probabilité que 165 < X < 185 cm
prob_3 <- pnorm(185, mean = mu_men, sd = sigma_men) - pnorm(165, mean = mu_men, sd = sigma_men)

# 4. Probabilité que X_h > X_f
mu_women <- 162 # Moyenne
sigma_women <- 12  # Écart-type
mu_diff <- mu_men - mu_women
sigma_diff <- sqrt(sigma_men^2 + sigma_women^2)

prob_4 <- 1 - pnorm(0, mean = mu_diff, sd = sigma_diff)

# Affichage des résultats
cat("P(X < 160) =", prob_1, "\n")
cat("P(X > 200) =", prob_2, "\n")
cat("P(165 < X < 185) =", prob_3, "\n")
cat("P(X_m > X_f) =", prob_4, "\n")

# Loi exponentielle
##########################################################

```