Faculté d'Ingénierie et Management de la Santé

Mesures et Probabilités Statistiques

Rapport de TP : Simulation de lois de probabilités

Allan De Clercq[1]

[1]Université de Lille

30 septembre 2024

Tuteurs : Assi N'Guessan Assi.Nguessan@polytech-lille.fr

Abstract :

Ce document présente le rapport d'un travail pratique dont l'objectif était d'explorer et d'analyser les propriétés statistiques de différentes distributions de probabilité. La méthodologie consistait à suivre les consignes et le code fournis dans le lien du TP (https://fcorset.github.io/TPs/tp2.html). Les résultats obtenus montrent une concordance significative avec les théories statistiques attendues, confirmant ainsi la validité des méthodes utilisées.

En conclusion, ce travail pratique a permis de renforcer la compréhension des concepts statistiques fondamentaux et de leur application pratique.

Dans ce rapport, seul le code permettant de générer les lois de probabilité est inclus en annexe. Les codes permettant de réaliser les figures et tableaux ne sont pas inclus, car ils ne sont pas pertinents dans une limitation à 2 pages d'annexe de code. L'intégralité du code est cependant disponible sur ce répertoire GitHub : ??

Table des matières

Table des figures

Liste des tableaux

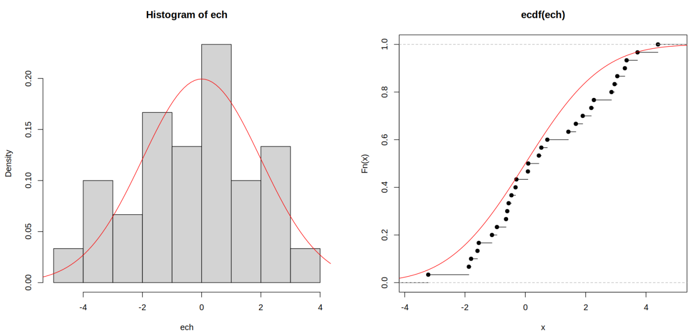
Listings

Introduction

Dans le logiciel R, on peut simuler très simplement à peu près toutes les lois classiques. Dans ce TP, 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 loi, **dmaloi** : permet de calculer la densité de maloi, **pmaloi** : permet de calculer la fonction de répartition de maloi, **qmaloi** : permet de calculer le quantile de maloi.

0.1 Exemple pour la loi normale

Voir code annexe 1 : Code exemple : La loi normale



(a) Histogramme de ech avec courbe de densité

(b) Graphique de la fonction de répartition

FIGURE 1 – Graphiques pour la loi normale

	X5.	X10.	X25.	X50.	X75.	X90.	X95.
quantile.echantillon	-3.41902	-2.48670	-1.65578	-0.49161	0.34339	1.22235	2.01371
quantile.normale	-3.28970	-2.56310	-1.34897	0.0	1.34897	2.56310	3.28970

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

0.2 Analyse des résultats

Dans cet exemple (voir code 1), nous avons simulé un échantillon de taille 30 selon une loi normale d'espérance 0 et d'écart-type 2.

L'histogramme (Figure 1a) montre une divergence entre la densité théorique et la simulation, due à la petite taille de l'échantillon. Cette variabilité est également visible sur le graphique de la fonction de répartition (Figure 1b) et dans le tableau de comparaison des quantiles (Table 1).

En conclusion, un échantillon plus grand est nécessaire pour une meilleure concordance avec la loi normale, conformément à la **loi des grands nombres** [1].

1 Lois discrètes

1.1 Loi uniforme discrète

Soit X une variable aléatoire distribuées selon une loi uniforme discrète sur $\{1, \dots, n\}$ (voir [2]). La distribution est définie comme suit :

$$\forall k \in \{1, \dots, n\},\ p(X = k) = \frac{1}{n} \Leftrightarrow X \to \mathcal{U}(\{1, \dots, n\}) \qquad (1)$$

1.1.1 Exemple - Simulation de la loi uniforme discrète

Voir code annexe 2 : Code exemple : Simulation de la loi uniforme discrète

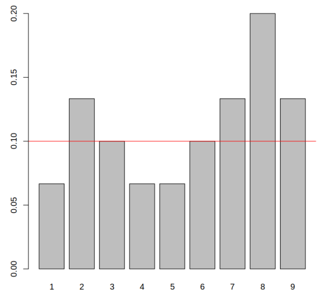


FIGURE 2 – Histogramme de simualtion par loi uniforme discrète avec sa droite de probabilité théorique

1.1.2 Analyse des résultats

Dans cet exemple (voir code 2), nous avons simulé un échantillon de taille 30 selon une loi uniforme discrète $\mathcal{U}(\{1, \ldots, 10\})$ (1). Théoriquement, chaque évenement est équiprobable, donc :

$$\forall k \in \{1, \ldots, 10\},\, p(X = k) = \frac{1}{10}$$

En analysant les résultats de l'histogramme (Figure 3), nous constatons grâce à la droite horizontale rouge représentant la probabilité théorique $p(X) = \frac{1}{10}$, que les fréquences observées varient et ne sont pas toutes égales. Par exemple,

$$p(X = 2) = 0.03 \quad p(X = 4) = 0.23$$

Cette variation est encore une fois due à la taille réduite de l'échantillon. Les fréquences observées devraient se rapprocher des probabilités théoriques avec un échantillon plus important, conformément à la loi des grands nombres [1]. En conclusion, un échantillon plus important est nécessaire pour démontrer la validité de la loi uniforme discrète de manière plus précise.

1.2 Loi de Bernoulli

Cette loi est notée $\mathcal{B}(p)$. On dit que $X \rightsquigarrow \mathcal{B}(p)$ si et seulement si X est une variable binaire telle que (voir [3]) :

$$p(X = 1) = p \quad \text{et} \quad p(X = 0) = 1 - p \tag{2}$$

C'est un cas particulier de la loi binomiale avec $n = 1$. On peut vérifier facilement par le calcul que $\mathbb{E}[X] = p$ et $\mathbb{V}[X] = p(1 - p)$.

1.2.1 Exemple - Simulation de la loi de Bernoulli

Voir code annexe 3 : Code exemple : Simulation de la loi de Bernoulli

0	1
0.73333	0.26667

TABLE 2 – Fréquences observées de l'échantillon simulé par la loi de Bernoulli $\mathcal{B}(0.3)$

1.2.2 Analyse des résultats

Dans cet exemple, nous avons simulé un échantillon de taille 30 selon une loi de Bernoulli $\mathcal{B}(0.3)$. La loi de Bernoulli est une loi binaire qui ne prend que deux valeurs possibles : 0 ou 1. Dans notre cas, les fréquences théoriques sont donc :

$$p(\text{Échec} = 0) = 0.7 \quad p(\text{Succès} = 1) = 0.3$$

En observant les fréquences (Table 2), la fréquence de succès est de 0.27 et celle d'échec est de 0.73. Ces fréquences sont proches des probabilités théoriques, avec une légère variation due à la taille de l'échantillon. Un échantillon plus grand réduirait cette variation, conformément à la loi des grands nombres [1].

1.3 Loi binomiale

On dit que $X \rightsquigarrow \mathcal{B}(n,p)$ si X est la somme de n variables de Bernoulli indépendantes de paramètre p. X représente le nombre de succès parmi n essais indépendants (voir [4]).

Pour résumer, soit $Y \rightsquigarrow \mathcal{B}(p)$ pour $i = 1, \dots, n$ où les Y_i sont indépendantes. On note :

$$X = \sum_{i=1}^{n} Y_i \rightsquigarrow \mathcal{B}(n,p) \tag{3}$$

Les valeurs possibles pour la variable X sont $\{0, \dots, n\}$ et

$$\forall k \in \{0, \dots, n\},\, p_k = p(X = k) = C_n^k p^k (1-p)^{n-k} \tag{4}$$

1.3.1 Exemple - Simulation de la loi de binomiale

Voir code annexe 4 : Code exemple : Simulation de la loi binomiale

5	6	7	8	9	10
0.02	0.06	0.30	0.36	0.16	0.09

TABLE 3 – Tableau partiel $[5, 10]$ des fréquences observées de la simulation binomiale de $p(\text{Succès}) = 0.8$ et $N = 100$ observations

0	1	2	3	4	6
0.06	0.29	0.37	0.22	0.05	0.01

TABLE 4 – Tableau partiel $[0, 6]$ des fréquences observées de la simulation binomiale de $p(\text{Succès}) = 0.2$ et $N = 100$ observations

5	6	7	8	9	10
0.026	0.088	0.201	0.301	0.268	0.107

TABLE 5 – Tableau partiel $[5, 10]$ des fréquences théoriques de la loi binomiale $\mathcal{B}(10, 0.8)$, arrondies à 3 décimales

0	1	2	3	4	6
0.107	0.268	0.302	0.201	0.088	0.026

TABLE 6 – Tableau partiel $[0, 6]$ des fréquences théoriques de la loi binomiale $\mathcal{B}(10, 0.2)$, arrondies à 3 décimales

1.3.2 Analyse des résultats

Dans cet exemple (voir code 4), nous avons simulé un échantillon de taille 100 selon deux lois binomiales : $\mathcal{B}(10, 0.8)$ et $\mathcal{B}(10, 0.2)$. Pour chaque loi, nous avons observé le nombre de succès sur 10 essais répétés 100 fois de manière indépendante.

On peut s'apercevoir que pour la loi $\mathcal{B}(10, 0.8)$, le fréquence la plus élevé est $p(X = 8) = 0.36$ et pour la loi $\mathcal{B}(10, 0.2)$, la fréquence la plus élevée est $p(X = 2) = 0.37$. Ce résulat s'explique par la prorpiété d'espérence (moyenne pondérée des valeurs possibles) de la loi binomiale : $\mathbb{E}[X] = np$.

— Pour la loi $\mathcal{B}(10, 0.8)$, on s'attend à avoir $\mathbb{E}[X] = np = 10 * 0.8 = 8$ succès en moyenne.
— Pour la loi $\mathcal{B}(10, 0.2)$, on s'attend à avoir $\mathbb{E}[X] = np = 10 * 0.2 = 2$ succès en moyenne.

En comparant les fréquences observées (Tableaux 3 et 4) avec les fréquences théoriques (Tableaux 5 et 6), nous remarquons une bonne concordance, bien que des variations existent en raison de la taille du nombre de répétitons.

Pour vérifier cela, si nous augmentons la taille de l'échantillon à $N = 10000$, les fréquences observées se rapprochent encore plus des fréquences théoriques :
— Pour la loi $\mathcal{B}(10, 0.8)$: $p(X = 8) = 0.3026$
— Pour la loi $\mathcal{B}(10, 0.2)$: $p(X = 2) = 0.3033$

1.4 Loi Géométrique

On dit que X suit une loi géométrique $\mathcal{G}(p)$ avec $0 < p < 1$ 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 (voir [5]). Les valeurs prises par X sont donc $X(\Omega) = N*$ et les probabilités associés sont :

$$\forall k \in \mathbb{N}^\star,\ p_k = P(X = k) = (1-p)^{k-1}p = q^{k-1}p \tag{5}$$

1.4.1 Exemple - Simulation de la loi géométrique

Voir code annexe 5 : Code exemple : Simulation de la loi géométrique

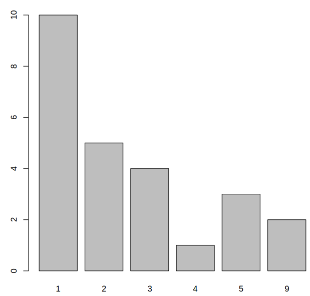


FIGURE 3 – Histogramme de simualtion par loi géométrique avec une probabilité $p(X = \text{Succès}) = 0.3$ et une taille d'échantillon de 25

1.4.2 Analyse des résultats

Dans cet exemple (voir code 5), nous avons simulé un échantillon de taille 25 selon une loi géométrique $\mathcal{G}(0.3)$. Comme nous pouvons le remarquer dans l'histogramme (Figure 3), les fréquences observés sont décroissantes, ce qui est conforme à la loi géométrique (5). Par exemple :

$$\begin{aligned} p(X=1) &= \tfrac{10}{25} = 0.4 \approx (1-0.3)^{1-1}0.3 = 0.3 \\ p(X=2) &= \tfrac{5}{25} = 0.2 \approx (1-0.3)^{2-1}0.3 = 0.21 \end{aligned}$$

1.5 Loi de Poisson

On dit que X suit une loi de Poisson $\mathcal{P}(\lambda)$ avec $\lambda > 0$. On peut voir cette loi comme la loi des événements rares. Cette loi est une approximation de la loi binomiale, $\mathcal{B}(n, p)$ lorsque n est grand et p petit. On pose alors $\lambda = np$. En pratique, on dit que cette approximation est bonne dès que $n > 20$ et $p <= 0.1$ et $np <= 5$ (voir [6]).

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

$$\forall k \in \mathbb{N},\, p_k = P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!} \tag{6}$$

1.5.1 Exemple - Simulation de la loi Poisson

Voir code annexe ?? : Code exemple : Simulation de la loi Poisson

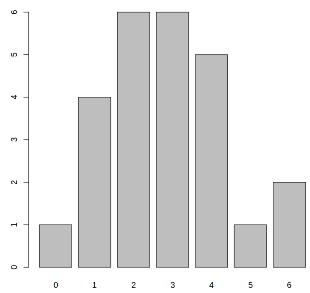


FIGURE 4 – Histogramme de simualtion par loi poisson avec un paramètre $\lambda = 3$ et une taille d'échantillon de 25

1.5.2 Analyse des résultats

Dans cet exemple, nous avons simulé un échantillon de taille $n = 25$ selon une loi de Poisson $\mathcal{P}(3)$. Ce qui signifie $\lambda = 3 = np$ donc cet exemple à une probabilité de succès de $p = 0.12$. On se trouve bien dans le cas où $n > 20$ et $p <= 0.1$ et $np <= 5$.

En observant l'histogramme (Figure 4), nous constatons que les fréquences observées sont conformes à la loi de Poisson (6). Par exemple :

$$\begin{aligned} p(X = 0) &= \tfrac{1}{25} = 0.04 \approx e^{-3}\frac{3^0}{0!} = 0.049 \\ p(X = 1) &= \tfrac{4}{25} = 0.16 \approx e^{-3}\frac{3^1}{1!} = 0.149 \\ p(X = 2) &= \tfrac{6}{25} = 0.24 \approx e^{-3}\frac{3^2}{2!} = 0.224 \end{aligned}$$

Nous pouvons observer grâce à la Figure 5 ci-dessous que la loi Poisson converge est une approximation de la loi binomiale pour n grand et p petit.

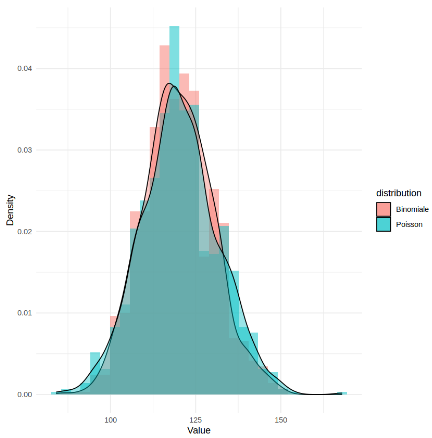


FIGURE 5 – Comparaison de la loi de Poisson et de la loi binomiale pour $n = 1000$ et $p = 0.12$

Lorsque l'on fait varier le paramètre λ de la loi de Poisson, on peut observer que la distribution se déplace vers la droite (Figure 6). Cela s'explique par le fait que λ est l'espérance de la loi de Poisson, dans le tableau ci-dessous Table 7, on peut voir que pour $\lambda = 2$ la médiane et la moyenne sont égales à 2 (ou très proche en fonction de la taille de l'échantillon n) Alors que pour $\lambda = 5$, la moyenne est de 5 et la médiane est de 5.

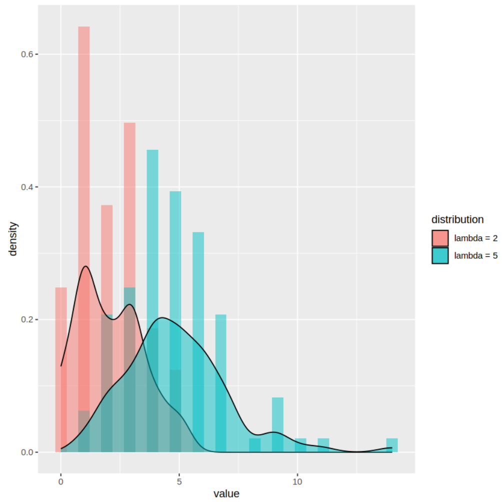


FIGURE 6 – Comparaison de la loi de Poisson selon le paramètre λ

λ	n	Moyenne	Médiane
2	100	2.05	2.00
5	100	4.87	5.00

TABLE 7 – Comparaison des paramètres de la loi de Poisson pour $\lambda = 2$ et $\lambda = 5$

2 Lois continues

2.1 Loi Uniforme sur $[0, 1]$, notée $\mathcal{U}([0, 1])$

2.2 Loi Uniforme sur $[a, b]$, notée $\mathcal{U}([a, b])$ avec $a < b$

2.3 Loi Normale centrée réduite, notée $\mathcal{N}(0, 1)$

2.4 Loi Normale, notée $\mathcal{N}(\mu, \sigma^2)$

2.5 Exercices sur la loi normale

2.5.1 Question 1

2.5.2 Question 2

2.5.3 Question 3

2.5.4 Question 4

2.5.5 Question 5

2.5.6 Question 6

2.5.7 Question 7

2.6 Lien entre les lois normales

2.7 Sommes de variables aléatoires gaussiennes

2.8 Somme de carrés de variables gaussiennes centrées réduites

2.9 Exercice

2.10 Loi Exponentielle

Code annexe

```
#------------------------ Code exemple : La loi normale ------------------------
# Args :
#   n: taille de l'échantillon -> int
#   ech: simulation de l'échantillon -> vecteur
#   quantile.normale: quantile de la loi normale -> vecteur
#   quantile.echantillon: quantile de l'échantillon -> vecteur
#-------------------------------------------------------------------------------

n<-30   #Taille de l'échantillon 30
# Simulation d'une loi normale d'espérance 1 et de variance 4 (écart-type = 2)
ech<-rnorm(n,mean = 0,sd = 2)

# Graphique 1
    # Création de l'histogramme de ech
    hist(ech, freq=F)
    # Ajout de la courbe de densité de a loi normale
    curve(dnorm(x,mean=0,sd=2),from=-6,to=6,ylab="densité",add=T,col="red")
# Graphique 2
    # Tracer le graphique de ech
    plot(ecdf(ech))
    # Tracer la fonction de répartition de ech
    curve(pnorm(x,mean=0,sd=2),from=-6,to=6,ylab="Fonction de répartition",add=T,col=
  "red")

# Création du quantile de la loi normale d'espérance 1 et d'écart-type = 2
quantile.normale<-qnorm(p=c(0.05,0.1,0.25,0.5,0.75,0.9,0.95),mean = 0,sd = 2)
# Création du quantile de la simulation "ech"
quantile.echantillon<-quantile(x=ech, probs=c(0.05,0.1,0.25,0.5,0.75,0.9,0.95))
# Tableau de comparaison des 2 quantiles
(rbind(quantile.echantillon,quantile.normale))
```

Listing 1 – Code exemple : La loi normale

```
#------------------- Code exemple : La loi uniforme discrete -------------------
# Args :
#   k: taille de l'échantillon -> int
#   ech: simulation de l'échantillon -> vecteur
#-------------------------------------------------------------------------------

 k <- 30  # Création de la taille de l'échantillon
 # Simulation d'un échantillon de taille n et de valeurs {1,...,10}.
 ech<-sample(1:10,k,replace=T)

 # Graphique 1
   # Création d'une graphique bar
   barplot(table(ech)/k)
   # Trace de la probabilite de la loi uniforme discrete sur {1,...,10}.
   abline(h=1/10,col="red")
```

Listing 2 – Code exemple : Simulation de la loi uniforme discrète

```
#---------------------- Code exemple : La loi de Bernoulli ---------------------
# Args :
#   n: taille de l'échantillon -> int
#   p: paramètre de la loi de Bernoulli -> float
#   esp: espérance théorique -> float
#   var: variance théorique -> float
#   ech: simulation de l'échantillon -> vecteur
#-------------------------------------------------------------------------------

n<- 30 # Taille de l'échantillon
```

```
p<-0.3  # paramètre de la loi de Bernoulli
esp<-p # Espérence théorique
var<-p*(1-p) # espérance théorique

# Simulation d'un échantillon de taille n selon une loi de Bernoulli
ech<-rbinom(n,size = 1,prob = p) # size = 1 correspond à la loi de Bernoulli
table(ech)/n # Affichage des fréquences observées
```

Listing 3 – Code exemple : Simulation de la loi de Bernoulli

```
#--------------------- Code exemple : La loi de Bernoulli ---------------------
# Args :
#   N: nombre d'observations -> int
#   n: nombre d'essais -> int
#   p_1: probabilité succès du premier test -> float
#   p_2: probabilité de succès du deuxième test -> float
#   ech_1: simulation de l'échantillon pour le premier test -> vecteur
#   ech_2: simulation de l'échantillon pour le deuxième test -> vecteu
#   ech_3: simulation de l'échantillon pour le troisième -> vecteur
#   ech_4: simulation de l'échantillon pour le quatrième -> vecteur
#------------------------------------------------------------------------------

N <- 100 # taille de l'échantillon
n <- 10  # premier paramètre de la loi

p_1<- 0.8 # deuxième paramètre de la loi : probabilité de succès
p_2<- 0.2 # deuxième paramètre de la loi : probabilité de succès

ech_1<-rbinom(N,n,p_1) # Simulation d'un échantillon pour p=0.8
ech_2<-rbinom(N,n,p_2) # Simulation d'un échantillon pour p=0.2

# Affichage des la table de propotion de la simulation
table(ech_1)/N ; table(ech_2)/N

dbinom(x=0:n,size = n,prob = p_1) # Simulation de la densité de la loi binomiale pour
    p=0.8
dbinom(x=0:n,size = n,prob = p_2) #  Simulation de la densité de la loi binomiale
   pour p=0.2

ech_3<-rbinom(10000,n,p_1) # Simulation d'un échantillon pour p=0.8 et N=1000
ech_4<-rbinom(10000,n,p_2) # Simulation d'un échantillon pour p=0.2 et N=1000

```

Listing 4 – Code exemple : Simulation de la loi binomiale

```
#---------------------- Code exemple : La loi Géométrique ---------------------
# Args :
#   n: nombre d'essais -> int
#   p: probabilité succès du premier test -> float
#------------------------------------------------------------------------------

n <- 25 # taille de l'échantillon
p <- 0.3 # paramètre de la loi
ech<-rgeom(n,p)+1 # simulation de la loi
barplot(table(ech))
```

Listing 5 – Code exemple : Simulation de la loi géométrique

Références

1. P. L. HSU, Robbins Herbert. Complete Convergence and the Law of Large Numbers [https://doi.org/10.1073/pnas.33.2.25]. 1947.
2. *Loi uniforme discrète*. Aussi disponible à l'adresse : https://bibmath.net/dico/index.php?action=affiche&quoi=.%2Fl%2Floiunifdisc.html. Accessed : 2024-30-09.
3. *Loi de Bernoulli*. Aussi disponible à l'adresse : https://www.bibmath.net/dico/index.php?action=affiche&quoi=./l/loibernoulli.html. Accessed : 2024-30-09.
4. *Loi binomiale*. Aussi disponible à l'adresse : https://www.bibmath.net/dico/index.php/index.php?action=affiche&quoi=./l/loibinomiale.html. Accessed : 2024-30-09.
5. *Loi géométrique*. Aussi disponible à l'adresse : https://www.bibmath.net/dico/index.php?action=affiche&quoi=./l/loigeometrique.html. Accessed : 2024-30-09.
6. *Loi de Poisson*. Aussi disponible à l'adresse : https://www.bibmath.net/dico/index.php?action=affiche&quoi=./l/loipoisson.html. Accessed : 2024-30-09.