

Probabilités et variables aléatoires

Résumé

Ce chapitre introduit les concepts essentielles des modèles probabilistes afin d'aborder l'inférence statistique : définition d'un événement aléatoire, des probabilités discrètes ou continues, des probabilités conditionnelles et de la notion d'indépendance en probabilités. Après avoir défini la notion de variable aléatoire, celles de lois les plus utilisées sont décrites : discrètes de Bernoulli ; binomiales, géométrique, de Poisson ; continues uniforme, exponentielle, Gamma, normale, du chi-deux, de Student et de Fisher. Espérance et variance d'une variable aléatoires sont définies, avant de signaler les deux théorèmes importants : loi des grands nombre et théorème de central limite.

Retour au plan du cours.

1 Introduction

Dans des domaines très différents comme les domaines scientifique, sociologique ou médical, on s'intéresse à de nombreux phénomènes dans lesquels apparaît l'effet du hasard. Ces phénomènes sont caractérisés par le fait que les résultats des observations varient d'une expérience à l'autre.

Une expérience est appelée "aléatoire" s'il est impossible de prévoir à l'avance son résultat et si, répétée dans des conditions identiques, elle peut donner des résultats différents :

- succession d'appels à un standard téléphonique non surchargé ;
- observation de la durée de vie d'un individu anonyme dans une population ;
- observation de la durée de fonctionnement sans panne d'appareil ;
- jeu de pile ou face.

Voici d'autres exemples de domaines d'applications des probabilités.

Fiabilité On considère un système formé par plusieurs composants. On s'intéresse à la fiabilité du système : on va chercher à calculer la probabilité que le système fonctionne encore à un instant donné. Il faut pour cela connaître la probabilité que chacun des composants fonctionne à cet instant et tenir compte du fait que les composants ne fonctionnent peut-être pas indépendamment les uns des autres.

Fatigue des matériaux Les données de fatigue des matériaux sont très dispersées. On fait alors appel à des modélisations probabilistes et à des méthodes statistiques afin, par exemple, de construire des intervalles de confiance pour le nombre moyen de cycles jusqu'à la rupture.

Télécommunications En télécommunications, on doit souvent tenir compte du "bruit" dans les systèmes. Par exemple, supposons qu'un système émet soit un 0, soit un 1, et qu'il y a un risque p que le chiffre émis soit mal reçu. Il est alors intéressant de calculer la probabilité qu'un 0 ait été émis, sachant qu'un 0 a été reçu, ou encore la probabilité qu'il y ait une erreur de transmission.

2 Notion de probabilité

2.1 événement

DÉFINITION 1. — *On appelle univers associé à une expérience aléatoire l'ensemble Ω de tous les résultats possibles de cette expérience.*

Le choix de l'ensemble Ω comporte une part d'arbitraire. Il dépend de l'idée que l'on a, a priori, sur les résultats de l'expérience aléatoire. Donnons quelques exemples :

1. On lance une pièce de monnaie. Pour l'ensemble Ω , on peut choisir soit $\Omega = \{ \text{pile, face} \}$, soit $\Omega = \{ \text{pile, face, tranche} \}$.
2. On s'intéresse à l'état de fonctionnement d'un système. Dans ce cas $\Omega = \{0, 1\}$ avec la convention 0 si le système est en panne et 1 s'il fonctionne.
3. Le résultat de l'expérience aléatoire est le nombre de tirages nécessaires dans un jeu de pile ou face jusqu'à l'obtention du premier "pile". Dans ce cas, $\Omega = \{1, 2, 3, \dots\} = \mathbb{N}^*$.

4. On considère la succession des appels à un standard téléphonique non surchargé et l'on étudie la répartition des instants où le standard reçoit un appel, à partir d'un instant choisi comme origine (on admet que deux appels ne peuvent se produire rigoureusement au même instant et que le phénomène est limité dans le temps). Une réalisation de cet événement est une suite croissante de nombres réels positifs t_i où t_i désigne l'instant d'enregistrement du i -ème appel : $\Omega = \{0 < t_1 < t_2 < \dots < t_n < t_{n+1} < \dots\}$. L'univers Ω est donc une partie de $(\mathbb{R}^+)^{\mathbb{N}}$.
5. On considère l'expérience aléatoire "durée de vie d'un individu". L'ensemble Ω est soit l'ensemble \mathbb{N} , soit \mathbb{R}^+ selon le procédé discontinu ou continu de cette mesure.

Nous constatons que Ω peut être fini (exemples 1 et 2), dénombrable (exemples 3 et 5) ou non dénombrable (exemples 4 et 5). Lorsque Ω est fini ou dénombrable, on parle d'univers discret. Sinon on parle d'univers continu.

DÉFINITION 2. — *Etant donnée une expérience aléatoire, un événement aléatoire est une partie de l'ensemble des résultats possibles de l'expérience, c'est donc un sous-ensemble A de l'univers Ω . On dit que l'événement A est réalisé si le résultat ω de l'expérience appartient à A .*

On sait que l'événement A est réalisé seulement une fois l'expérience aléatoire réalisée.

Exemples :

- Si l'on s'intéresse à l'événement suivant : "on a obtenu un chiffre pair lors d'un lancer d'un dé à 6 faces", on introduit $A = \{2, 4, 6\}$, qui est un sous-ensemble de $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- Si l'on s'intéresse à l'événement suivant : "la durée de vie du composant est supérieure ou égale à 1000 heures", $A = [1000, +\infty[$ est un sous-ensemble de $\Omega = \mathbb{R}^+$.

L'ensemble \emptyset est appelé l'événement impossible et Ω est appelé l'événement certain.

2.2 Opérations sur les événements

Les événements aléatoires étant des ensembles, introduisons les opérations ensemblistes classiques de la théorie des ensembles.

DÉFINITION 3. — *On appelle événement contraire de A , noté A^C , le complémentaire de A dans Ω :*

$$A^C = \{\omega \in \Omega : \omega \notin A\}.$$

L'événement contraire A^C est réalisé si et seulement si A n'est pas réalisé.

Exemple : Si A est l'événement "la durée de vie du composant est supérieure ou égale à 1000 heures" : $A = [1000, +\infty[$, l'événement contraire est l'événement "la durée de vie du composant est strictement inférieure à 1000 heures" : $A^C = [0, 1000[$.

DÉFINITION 4. — *Soient A et B deux événements d'un univers Ω .*

- L'événement " A et B " est celui qui est réalisé si A et B sont réalisés. C'est l'intersection

$$A \cap B = \{\omega \in \Omega : \omega \in A \text{ et } \omega \in B\}.$$

- L'événement " A ou B " est celui qui est réalisé si l'un des deux est réalisé ou si les deux sont réalisés. C'est l'union

$$A \cup B = \{\omega \in \Omega : \omega \in A \text{ ou } \omega \in B\}.$$

- L'inclusion $A \subset B$ signifie que l'événement A ne peut être réalisé sans que B le soit.

DÉFINITION 5. — *Deux événements A et B sont dits incompatibles si la réalisation de l'un implique la non-réalisation de l'autre.*

Dans l'espace Ω , deux événements incompatibles sont représentés par deux parties disjointes. Si $A \cap B = \emptyset$, alors A et B sont incompatibles. Il est clair, par exemple que A et A^C sont incompatibles.

2.3 Probabilité

Définition

DÉFINITION 6. — *Soit Ω un univers associé à une expérience aléatoire et soit \mathcal{A} l'ensemble des parties de Ω . Une probabilité \mathbb{P} sur l'espace (Ω, \mathcal{A}) est une application de \mathcal{A} dans $[0, 1]$ telle que*

1. $\mathbb{P}(\Omega) = 1$.
2. Si $(A_n)_{n \geq 1}$ est une famille d'événements de \mathcal{A} 2 à 2 incompatibles,

$$\mathbb{P}\left(\bigcup_{n=1}^{+\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

Le triplet $(\Omega, \mathcal{A}, \mathbb{P})$ est appelé espace de probabilité.

On peut déduire de la définition précédente un certain nombre de propriétés.

PROPOSITION 7. — Soient A et B deux événements aléatoires.

1. $\mathbb{P}(\emptyset) = 0$.
2. $\mathbb{P}\left(\bigcup_{n=1}^N A_n\right) \leq \sum_{n=1}^N \mathbb{P}(A_n)$.
3. Si A_1, \dots, A_N sont deux-à-deux incompatibles,

$$\mathbb{P}\left(\bigcup_{n=1}^N A_n\right) = \sum_{n=1}^N \mathbb{P}(A_n).$$

4. $\mathbb{P}(A^C) = 1 - \mathbb{P}(A)$.
5. Si $A \subset B$, $\mathbb{P}(A) \leq \mathbb{P}(B)$.
6. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.
7. Si Ω est fini ou dénombrable, alors pour tout événement A ,

$$\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\{\omega\}).$$

Exemple : Probabilité uniforme

Soit Ω un ensemble fini : $\Omega = \{\omega_1, \dots, \omega_N\}$. Pour tout $i \in \{1, 2, \dots, N\}$, on pose $\mathbb{P}(\{\omega_i\}) = \frac{1}{N}$. Alors, pour toute partie A de Ω , on a

$$\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\{\omega\}) = \frac{\text{Card}(A)}{N} = \frac{\text{Card}(A)}{\text{Card}(\Omega)}.$$

Dans le cas du lancer de dé à 6 faces, pour tout $\omega \in \{1, 2, \dots, 6\}$, $\mathbb{P}(\{\omega\}) = 1/6$.

Si on note l'événement "on a obtenu un chiffre pair" par $A = \{2, 4, 6\}$, alors

$$\mathbb{P}(A) = 3/6 = 1/2.$$

Remarques : Pour un problème donné, il y a souvent plusieurs modélisations possibles, c'est-à-dire que le choix de l'espace de probabilité n'est pas unique.

Remarque : Choisir un élément au hasard signifie que les divers choix possibles sont équiprobables, donc que l'ensemble Ω est muni de la probabilité uniforme. Dans ce cas, tous les calculs sont simples et se ramènent souvent à des calculs d'analyse combinatoire.

2.4 Probabilités conditionnelles

Dans le chapitre précédent, on a parlé de la probabilité d'un événement sans tenir compte de la réalisation d'autres événements. En pratique, on peut considérer plusieurs événements, certains pouvant avoir une influence sur la réalisation d'autres événements.

Exemple : On lance deux dés. Soient les événements $A = \{\text{la somme est } \geq 11\}$ et $B = \{\text{le lancer du 1er dé donne } 6\}$. Il est clair que la réalisation de B influe sur la réalisation de A .

Supposons que l'on s'intéresse à la réalisation d'un événement A , tout en sachant qu'un événement B est réalisé. Si A et B sont incompatibles, alors la question est réglée : A ne se réalise pas. Mais si $A \cap B \neq \emptyset$, il est possible que A se réalise. Cependant, l'espace des événements possibles n'est plus Ω tout entier, mais il est restreint à B . En fait, seule nous intéresse la réalisation de A à l'intérieur de B , c'est-à-dire $A \cap B$ par rapport à B . Ceci justifie la définition suivante.

DÉFINITION 8. — Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité. Soient A et B deux événements aléatoires tels que $\mathbb{P}(B) \neq 0$. On appelle probabilité conditionnelle de A sachant B la quantité

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Remarque : On a les égalités suivantes :

$$\text{Si } \mathbb{P}(B) > 0, \quad \mathbb{P}(A \cap B) = \mathbb{P}(A|B) \times \mathbb{P}(B).$$

$$\text{Si } \mathbb{P}(A) > 0, \quad \mathbb{P}(A \cap B) = \mathbb{P}(B|A) \times \mathbb{P}(A).$$

PROPOSITION 9. — (formule des probabilités totales) Soit $(A_i)_{i \in I}$ une famille d'événements aléatoires formant une partition de Ω , c'est-à-dire tels que :

- $\cup_{i \in I} A_i = \Omega$;
- $A_i \cap A_j = \emptyset$ pour tout $i \neq j$.

On suppose de plus que $\mathbb{P}(A_i) \neq 0$ pour tout $i \in I$. Alors

$$\mathbb{P}(A) = \sum_{i \in I} \mathbb{P}(A|A_i) \mathbb{P}(A_i).$$

PROPOSITION 10. — (formule de Bayes) Sous les mêmes hypothèses que la proposition précédente, on a :

$$\mathbb{P}(A_i|A) = \frac{\mathbb{P}(A|A_i) \mathbb{P}(A_i)}{\sum_{i \in I} \mathbb{P}(A|A_i) \mathbb{P}(A_i)}.$$

La formule de Bayes (publiée après sa mort en 1763) présente un grand intérêt car elle permet de modifier notre connaissance des probabilités en fonction d'informations nouvelles. Cette formule joue donc un rôle très important dans la statistique bayésienne.

2.5 Indépendance

DÉFINITION 11. — Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité, et soient A et B deux événements aléatoires. On dit que A et B sont indépendants si

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \mathbb{P}(B).$$

Remarque : A et B sont indépendants si et seulement si $\mathbb{P}(A|B) = \mathbb{P}(A)$. Cette condition signifie que la probabilité de réalisation de l'événement A n'est pas modifiée par une information concernant la réalisation de l'événement B .

PROPOSITION 12. — Si A et B sont deux événements indépendants alors :

- A^C et B sont également indépendants ;
- A et B^C sont également indépendants ;
- A^C et B^C sont également indépendants.

Nous allons maintenant définir l'indépendance de plus de 2 événements aléatoires.

DÉFINITION 13. — Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité. Pour $n \geq 2$, soient A_1, A_2, \dots, A_n , des événements aléatoires.

- Ces événements sont deux à deux indépendants si pour tout couple (i, j) avec $i \neq j$ on a

$$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i) \mathbb{P}(A_j).$$

- Ces événements sont indépendants (dans leur ensemble) si pour tout $k \in \{2, 3, \dots, n\}$ et tout choix d'indices distincts i_1, \dots, i_k , on a

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \mathbb{P}(A_{i_2}) \dots \mathbb{P}(A_{i_k}).$$

3 Notion de variable aléatoire

3.1 Introduction

Dans de nombreuses expériences aléatoires, on n'est pas intéressé directement par le résultat de l'expérience, mais par une certaine fonction de ce résultat. Considérons par exemple l'expérience qui consiste à observer, pour chacune des n pièces produites par une machine, si la pièce est défectueuse ou non. Nous attribuerons la valeur 1 à une pièce défectueuse et la valeur 0 à une pièce en bon état. L'univers associé à cette expérience est $\Omega = \{0, 1\}^n$. Ce qui intéresse le fabricant est la proportion de pièces défectueuses produites par la machine. Introduisons donc une fonction de Ω dans \mathbb{R} qui à tout $\omega = (\omega_1, \omega_2, \dots, \omega_n)$ de Ω associe le nombre

$$X(\omega) = \sum_{i=1}^n \frac{\omega_i}{n},$$

qui correspond à la proportion de pièces défectueuses associée à l'observation de ω . Une telle fonction X définie sur Ω et à valeurs dans \mathbb{R} s'appelle une variable aléatoire réelle.

3.2 Définitions

Variable aléatoire réelle

DÉFINITION 14. — *Etant donné un univers Ω , une variable aléatoire réelle (v.a.r.) est une application de Ω dans \mathbb{R} :*

$$X : \omega \in \Omega \mapsto X(\omega) \in \mathbb{R}.$$

Loi de probabilité

DÉFINITION 15. — *Soit Ω un univers muni d'une probabilité \mathbb{P} , et soit X une v.a.r. On appelle loi de probabilité de X , notée \mathbb{P}_X , l'application qui à toute partie A de \mathbb{R} associe*

$$\mathbb{P}_X(A) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\}).$$

Remarque : Dans la suite du cours, on utilisera la notation abrégée : $\mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\}) = \mathbb{P}(X \in A)$. De même, on notera $\mathbb{P}(X = x)$ la probabilité $\mathbb{P}(\{\omega \in \Omega : X(\omega) = x\})$.

PROPOSITION 16. — *L'application \mathbb{P}_X définit une probabilité sur \mathbb{R} .*

Fonction de répartition

DÉFINITION 17. — *La fonction de répartition de la v.a.r. X est définie par*

$$F_X(x) = \mathbb{P}(X \leq x), \quad x \in \mathbb{R}.$$

Propriétés de la fonction de répartition :

1. $0 \leq F_X \leq 1$.
2. F_X tend vers 0 en $-\infty$ et vers 1 en $+\infty$.
3. F_X est croissante.
4. F_X est continue à droite.

PROPOSITION 18. — *On a l'identité*

$$\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a), \quad \forall a < b.$$

Remarque : On montre facilement que F_X est continue si et seulement si $\mathbb{P}(X = x) = 0$ pour tout $x \in \mathbb{R}$. On parle alors de loi diffuse ou de v.a.r. continue (voir définition 21).

DÉFINITION 19. — *Soit X une v.a.r. de fonction de répartition F_X supposée strictement croissante de $I \subset \mathbb{R}$ dans $]0, 1[$. Le quantile d'ordre $\alpha \in]0, 1[$ de X est le nombre $x_\alpha \in I$ tel que $F_X(x_\alpha) = \alpha$, ce qui signifie que*

$$\mathbb{P}(X \leq x_\alpha) = \alpha.$$

Remarques :

- $x_{1/2}$ est appelé médiane de X . La médiane vérifie les deux égalités

$$P(X \leq x_{1/2}) = 1/2 = P(X > x_{1/2}).$$

- Dans le cas où F_X n'est pas strictement croissante mais simplement croissante, on définit le quantile d'ordre α par

$$x_\alpha = \inf\{x \in \mathbb{R} : F_X(x) \geq \alpha\}.$$

3.3 Variables aléatoires réelles discrètes

Définition

DÉFINITION 20. — *Une v.a.r. X à valeurs dans un ensemble \mathcal{X} fini ou dénombrable est appelée v.a.r. discrète. Dans ce cas, la loi de X est déterminée par l'ensemble des probabilités :*

$$\mathbb{P}_X(x) = \mathbb{P}(X = x), \quad x \in \mathcal{X}.$$

Ainsi, pour toute partie A de \mathcal{X} , on a alors :

$$\mathbb{P}_X(A) = \mathbb{P}(X \in A) = \sum_{x \in A} \mathbb{P}(X = x) \quad \text{et} \quad \mathbb{P}_X(\mathcal{X}) = \sum_{x \in \mathcal{X}} \mathbb{P}(X = x) = 1.$$

Exemple : Supposons que l'on observe la durée de vie T d'une ampoule électrique et que cette durée de vie T , exprimée en heures, satisfait pour tout $0 < a < b$,

$$\mathbb{P}(a < T \leq b) = \exp(-a/100) - \exp(-b/100).$$

On note X le nombre de périodes complètes de 100 heures que dure l'ampoule. Les valeurs possibles de X étant entières, la v.a.r. X est donc discrète. Calculons la fonction de répartition de X . Comme X est positive, on a

$$F_X(x) = \mathbb{P}(X \leq x) = 0, \quad \forall x < 0.$$

De plus, pour tout $n \in \mathbb{N}$,

$$\mathbb{P}(X = n) = \mathbb{P}(100n \leq T < 100(n+1)) = \exp(-n) - \exp(-(n+1)).$$

Ainsi, on a donc pour tout $x \geq 0$:

$$\begin{aligned} \mathbb{P}(X \leq x) &= \sum_{n=0}^{[x]} \mathbb{P}(X = n) \\ &= 1 - \exp(-([x] + 1)). \end{aligned}$$

On notera que la fonction F_X est une fonction en escalier.

Exemples de variables discrètes

Soit X une v.a.r. discrète prenant ses valeurs dans un ensemble $\{x_1, x_2, \dots, x_n\}$, éventuellement infini. Alors la loi de X est caractérisée par l'ensemble des probabilités $\mathbb{P}(X = x_i)$, c'est-à-dire les nombres réels positifs p_i tels que

$$\mathbb{P}(X = x_i) = p_i \quad \text{avec} \quad 0 \leq p_i \leq 1 \quad \text{et} \quad \sum_{i=1}^n p_i = 1.$$

Loi de Bernoulli

On dit qu'une v.a.r. X à valeurs dans $\{0, 1\}$ suit une loi de Bernoulli de paramètre $p \in]0, 1[$, notée $\mathcal{B}(p)$, si

$$\mathbb{P}(X = 1) = 1 - \mathbb{P}(X = 0) = p.$$

Par exemple, cette loi intervient lorsque l'on modélise l'état de fonctionnement d'un système. La probabilité que le système fonctionne vaut p et la probabilité que le système ne fonctionne pas vaut $1 - p$. Cette loi s'applique aussi aux jeux de hasard de type binaire comme pile ou face ...

Loi binomiale

On dit qu'une v.a.r. X à valeurs dans $\{0, 1, \dots, n\}$ suit une loi binomiale de paramètres (n, p) , notée $\mathcal{B}(n, p)$, si

$$\mathbb{P}(X = k) = C_n^k p^k (1-p)^{n-k}, \quad 0 \leq k \leq n.$$

Cette loi intervient par exemple pour modéliser le nombre de pièces défectueuses dans un lot de n pièces, qui ont chacune une probabilité p d'être défectueuse, indépendamment les unes des autres.

Loi géométrique

On dit qu'une v.a.r. X à valeurs dans \mathbb{N}_* suit une loi géométrique de paramètre $p \in]0, 1[$, notée $\mathcal{G}(p)$, si

$$\mathbb{P}(X = k) = p(1-p)^{k-1}, \quad k \in \mathbb{N}^*.$$

Cette loi permet de modéliser le nombre de réalisations indépendantes d'une expérience à 2 issues (succès-échec), jusqu'à l'obtention du premier succès, si à chaque réalisation la probabilité de succès est p .

Loi de Poisson

On dit qu'une v.a.r. X à valeurs dans \mathbb{N} suit une loi de Poisson de paramètre $\lambda > 0$, notée $\mathcal{P}(\lambda)$, si

$$\mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad k \in \mathbb{N}.$$

Cette loi intervient comme comportement limite de la loi binomiale lorsque $n \rightarrow +\infty$ et $np \rightarrow \lambda$.

Elle intervient également pour modéliser des "événements rares". Soit N la variable aléatoire comptant le nombre d'occurrences d'un événement pendant une période donnée T . On suppose qu'un seul événement arrive à la fois, que le nombre d'événement se produisant pendant T ne dépend que de la durée de cette période et que les événements sont indépendants.

Si le nombre moyen d'événements (i.e. accidents) par unité de temps (i.e. semaine) est c , alors on démontre que la probabilité d'obtenir n événements

pendant un temps T est :

$$\mathbb{P}(N = n) = \exp(-cT) \frac{(cT)^n}{n!}.$$

3.4 Variables aléatoires réelles continues

Définition

DÉFINITION 21. — Soit X une v.a.r. qui prend un nombre infini non dénombrable de valeurs. Si F_X est une fonction continue, on dit que X est une v.a.r. continue. Dans ce cas, la loi de X est déterminée par l'ensemble des probabilités $\mathbb{P}(a < X < b)$, pour tout $a < b$.

Remarque : Notons que l'on peut mettre $<$ ou \leq dans ce qui précède car la variable étant continue, on a $\mathbb{P}(X = x) = 0$ pour tout $x \in \mathbb{R}$. **Exemple :** Soit $\lambda > 0$. Une v.a.r. X de fonction de répartition

$$F_X(x) = \begin{cases} 1 - \exp(-\lambda x) & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$$

est continue.

DÉFINITION 22. — Si l'on peut écrire la fonction de répartition d'une variable continue sous la forme

$$F_X(t) = \int_{-\infty}^t f_X(x) dx,$$

où f_X est une fonction de \mathbb{R} dans \mathbb{R} , alors on dit que f_X est la densité de probabilité de la v.a.r. X .

Ceci implique que l'on a pour tout $a < b$:

$$\mathbb{P}(a < X < b) = F_X(b) - F_X(a) = \int_a^b f_X(x) dx.$$

Cee intégrale étant positive pour tout $a < b$, il en résulte que $f_X \geq 0$. De plus, puisque $\lim_{t \rightarrow +\infty} F_X(t) = 1$, on a

$$\int_{-\infty}^{+\infty} f_X(x) dx = 1.$$

Une densité de probabilité est donc une fonction positive ou nulle, d'intégrale 1, et qui caractérise la loi d'une v.a.r. continue. De plus, en tout point $x_0 \in \mathbb{R}$ où F_X est dérivable, on a $f_X(x_0) = F'_X(x_0)$.

Exemple : Dans l'exemple de la durée de vie T d'une ampoule électrique, T a pour densité de probabilité

$$f(x) = \begin{cases} \exp(-x/100)/100 & \text{pour tout } x \geq 0 \\ 0 & \text{pour tout } x < 0. \end{cases}$$

Enfin, établir que deux v.a.r. (discrètes ou continues) X et Y ont même loi, c'est démontrer que l'on a l'égalité suivante :

$$\mathbb{P}(a < X \leq b) = \mathbb{P}(a < Y \leq b), \quad a, b \in \mathbb{R}.$$

Ainsi, en faisant tendre a vers $-\infty$, on obtient le résultat suivant :

THÉORÈME 23. — Deux v.a.r. à valeurs dans le même ensemble d'arrivée ont la même loi si et seulement si leurs fonctions de répartition sont égales.

Exemples de variables continues

Soit X une v.a.r. continue. Alors la loi de X est caractérisée par l'ensemble des probabilités

$$\mathbb{P}(a < X < b) = \int_a^b f_X(x) dx,$$

où f_X est la densité de probabilité de X et a et b sont deux nombres réels, éventuellement infinis. Comme nous l'avons vu plus haut, il suffit de connaître cette densité pour connaître la loi de X .

Loi uniforme

La loi uniforme sur un intervalle est la loi des "tirages au hasard" dans cet intervalle. Si $a < b$ sont deux réels, la loi uniforme sur l'intervalle $[a, b]$ est notée $\mathcal{U}(a, b)$. Elle a pour densité :

$$\frac{1}{b-a} 1_{[a,b]}(x).$$

Loi exponentielle

On dit que X suit une loi exponentielle de paramètre $\lambda > 0$, notée $\mathcal{E}(\lambda)$, si la loi de X a pour densité

$$f_X(x) = \begin{cases} \lambda \exp(-\lambda x) & \text{si } x \geq 0, \\ 0 & \text{si } x < 0. \end{cases}$$

La loi exponentielle est utilisée en fiabilité. Le paramètre λ représente le taux moyen de défaillance alors que son inverse $\theta = 1/\lambda$ est “le temps moyen de bon fonctionnement”. La loi exponentielle s’applique bien aux matériels électroniques ou aux matériels subissant des défaillances brutales.

Loi Gamma

La loi exponentielle est un cas particulier de la famille des lois Gamma. Soient $a > 0$ et $\lambda > 0$. On dit que X suit une loi Gamma de paramètres (a, λ) , notée $\gamma(a, \lambda)$, si la loi de X a pour densité

$$f_X(x) = \begin{cases} \frac{\lambda^a}{\Gamma(a)} x^{a-1} \exp(-\lambda x) & \text{si } x \geq 0, \\ 0 & \text{si } x < 0, \end{cases}$$

où pour tout $a > 0$, la fonction gamma est donnée par $\Gamma(a) = \int_0^{+\infty} x^{a-1} \exp(-x) dx$. Le paramètre a est un paramètre de forme alors que le paramètre λ est un paramètre d’échelle. Pour n entier, $a = n/2$ et $\lambda = 1/2$, la loi $G(n/2, 1/2)$ est appelée loi du chi-deux à n degrés de liberté, et notée $\chi^2(n)$. Elle joue un rôle important en statistique, c’est la loi de la somme des carrés de n variables aléatoires indépendantes de loi $\mathcal{N}(0, 1)$. On l’utilise pour les variances empiriques d’échantillons gaussiens. La loi $G(1, \lambda)$ est la loi exponentielle $\mathcal{E}(\lambda)$.

Loi normale de paramètres (μ, σ^2)

Soient $\mu \in \mathbb{R}$ et $\sigma > 0$. On dit que X suit une loi normale de paramètres (μ, σ^2) , notée $\mathcal{N}(\mu, \sigma^2)$, si la loi de X a pour densité

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \quad x \in \mathbb{R}.$$

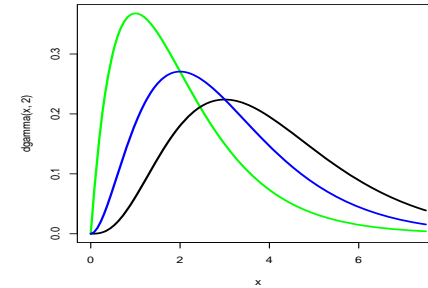


FIGURE 1 – Graphe de la densité de la loi Gamma de paramètre de forme $a = 2, 3, 4$.

la densité de la loi normale présente un axe de symétrie vertical pour $x = \mu$; il n’existe pas d’expression analytique de la fonction de répartition de X qui est approchée par le calcul numérique de l’intégrale de la densité. La loi normale s’applique à de nombreux phénomènes, en physique, en économie (erreurs de mesure), biologie; c’est une conséquence du théorème central limite vue dans la section 5.3, elle est la forme limite de nombreuses lois discrètes. Ainsi, toute grandeur résultat d’un ensemble ou d’une “somme” de plusieurs variables indépendantes et de même loi se distribue approximativement suivant une loi normale. Il faut cependant remarquer que les variables utilisées dans les domaines technologique, économique, biologique sont bien souvent positives. Pour que la loi normale puisse être représentative d’un tel phénomène, il faut que la probabilité théorique d’obtenir des valeurs négatives de la variable soit très faible. Il faut en particulier éviter d’utiliser cette modélisation pour les queues des distributions.

Lois du χ^2 , de Student et de Fisher

Par définition, la variable aléatoire, somme des carrés de ν variables indépendantes $\mathcal{N}(0, 1)$ suit une loi du χ^2 à ν degrés de liberté. Deux autres lois jouent des rôles important en statistique. La loi de Student à n degrés de liberté, $\mathcal{T}(n)$ est la loi du rapport $X/(\sqrt{Y/n})$, où les variables aléatoires X et Y sont indépendantes, X de loi $\mathcal{N}(0, 1)$, Y de loi $\chi^2(n)$. Elle a pour densité :

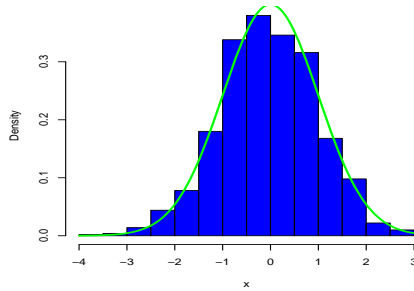


FIGURE 2 – Densité théorique et histogramme de la simulation de la loi normale.

$$\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)}\left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}.$$

Elle décrit la distribution de la moyenne empirique d'un échantillon gaussien.

La loi de Fisher de paramètres n et m (entiers positifs), est la loi du rapport $(X/n)/(Y/m)$, où X et Y sont deux variables aléatoires indépendantes, de lois respectives $\chi^2(n)$ et $\chi^2(m)$. Elle caractérise la distribution de rapports de variances et est très présente en théorie des tests (analyse de variance et modèle linéaire). L'expression de sa densité est définie par un rapports de fonctions $\Gamma(x)$.

4 Caractéristiques des variables aléatoires

4.1 Espérance

Définition

DÉFINITION 24. — Soit X une v.a.r. et h une application de \mathbb{R} dans \mathbb{R} . Donc $h(X)$ est elle aussi une v.a.r.

- Si X est discrète à valeurs dans un ensemble \mathcal{X} , l'espérance de $h(X)$ est

la quantité

$$\mathbb{E}(h(X)) = \sum_{x \in \mathcal{X}} h(x)\mathbb{P}(X = x),$$

pourvu que cette série converge (dans le cas où \mathcal{X} est infini).

- Si X est continue et admettant une densité f_X , l'espérance de $h(X)$ est la quantité

$$\mathbb{E}(h(X)) = \int_{-\infty}^{+\infty} h(x)f_X(x)dx,$$

à condition que cette intégrale soit convergente.

Notons que si $h(x) = x$, on obtient $\mathbb{E}(X)$ appelée espérance mathématique (ou moyenne) de la v.a.r. X . Par ailleurs, si l'on définit la v.a.r. suivante :

$$1_{\{X \in A\}} = \begin{cases} 1 & \text{si } X \in A \quad (A \subset \mathbb{R}) \\ 0 & \text{sinon} \end{cases}$$

qui est appelée fonction caractéristique de l'événement $\{X \in A\}$, alors l'espérance de cette v.a.r. est :

$$\mathbb{E}(1_{\{X \in A\}}) = \mathbb{P}(X \in A) = \mathbb{P}_X(A),$$

d'où le lien étroit entre probabilité et espérance.

Propriétés

1. L'espérance est linéaire : pour tout $\alpha, \beta \in \mathbb{R}$, et pour toutes v.a.r. X et Y

$$\mathbb{E}(\alpha X + \beta Y) = \alpha\mathbb{E}(X) + \beta\mathbb{E}(Y).$$

2. Si X est une v.a.r. constante égale à $a \in \mathbb{R}$, c'est-à-dire pour tout $\omega \in \Omega$, $X(\omega) = a$, alors $\mathbb{P}(X = a) = 1$ et $\mathbb{E}(X) = a$.

3. L'espérance d'une v.a.r. positive est positive. En particulier, si $X \geq Y$ (ce qui signifie que pour tout $\omega \in \Omega$, $X(\omega) \geq Y(\omega)$), alors $\mathbb{E}(X - Y) \geq 0$ donc $\mathbb{E}(X) \geq \mathbb{E}(Y)$.

L'espérance d'une v.a.r. X est un indicateur de "localisation" de sa loi :

$$\mathbb{E}(X) \simeq \text{"valeur moyenne de } X\text{"}.$$

Néanmoins, la connaissance de l'espérance mathématique donne peu de renseignements sur cette v.a.r. Ainsi, il faut étudier "l'étalement" de sa loi, c'est-à-dire la dispersion de la v.a.r. X autour de sa moyenne $\mathbb{E}(X)$.

4.2 Variance et écart-type

Définitions

Pour rendre positifs les écarts entre X et son espérance $\mathbb{E}(X)$, un autre outil plus facile à manipuler que la valeur absolue, est à notre disposition : la mise au carré. On ne va donc pas calculer la moyenne des écarts mais la moyenne des écarts au carré. C'est ce qu'on appelle la variance.

DÉFINITION 25. — La variance de la v.a.r. X est la quantité :

$$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}(X))^2].$$

Propriétés :

- $\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$.
- $\text{Var}(aX+b) = a^2\text{Var}(X)$ pour tout $a, b \in \mathbb{R}$. En particulier, $\text{Var}(X+b) = \text{Var}(X)$.

Afin d'être en mesure de comparer, en termes d'ordre de grandeur, variance et espérance, il faut prendre la racine carrée de la variance. C'est ce qu'on appelle l'écart-type.

DÉFINITION 26. — La racine carrée de $\text{Var}(X)$, notée σ_X , est appelée écart-type de X .

Remarques :

- Si X est une v.a.r. telle que $\mathbb{E}(X) = \mu$ et $\text{Var}(X) = \sigma^2$, alors la variable $Y = (X - \mu)/\sigma$ est d'espérance nulle et de variance 1. On dit que Y est centrée (d'espérance nulle) et réduite (de variance 1).
- Le moment d'ordre k est défini par

$$m_k = \int_{-\infty}^{\infty} (x - \mu)^k f(x) dx.$$

- $\gamma_1 = \frac{m_3}{\sigma^3}$ est le coefficient d'asymétrie (*Skewness*).
- $\gamma_2 = \frac{m_4}{\sigma^4} - 3$ est le coefficient d'aplatissement (*Kurtosis*).

Exemples :

- Loi uniforme $\mathcal{U}[a, b]$: $\mathbb{E}(X) = \frac{a+b}{2}$ et $\text{Var}(X) = \frac{(b-a)^2}{12}$.

- Loi de Bernoulli $\mathcal{B}(p)$: $\mathbb{E}(X) = p$ et $\text{Var}(X) = p(1-p)$.
- Loi binomiale $\mathcal{B}(n, p)$: $\mathbb{E}(X) = np$ et $\text{Var}(X) = np(1-p)$.
- Loi géométrique $\mathcal{G}(p)$: $\mathbb{E}(X) = \frac{1}{p}$ et $\text{Var}(X) = \frac{1-p}{p^2}$.
- Loi de Poisson $\mathcal{P}(\lambda)$: $\mathbb{E}(X) = \text{Var}(X) = \lambda$.
- Loi normale $\mathcal{N}(\mu, \sigma^2)$: $\mathbb{E}(X) = \mu$ et $\text{Var}(X) = \sigma^2$.
- Loi exponentielle $\mathcal{E}(\lambda)$: $\mathbb{E}(X) = \frac{1}{\lambda}$ et $\text{Var}(X) = \frac{1}{\lambda^2}$.

4.3 Inégalité de Chebychev

THÉORÈME 27. — (inégalité de Chebychev)

Soit $\epsilon > 0$ et soit X une v.a.r. admettant une variance. Alors on a :

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \epsilon) \leq \frac{\text{Var}(X)}{\epsilon^2}.$$

Cette inégalité permet de comprendre la signification de l'écart-type $\sigma_X = \sqrt{\text{Var}(X)}$, au sens où il caractérise la dispersion de la v.a.r. autour de son espérance mathématique :

$$\begin{aligned} \mathbb{P}(|X - \mathbb{E}(X)| \geq \epsilon\sigma_X) &\leq \frac{1}{\epsilon^2} \\ \mathbb{P}(|X - \mathbb{E}(X)| < \epsilon\sigma_X) &\geq 1 - \frac{1}{\epsilon^2}. \end{aligned}$$

Supposons que $\epsilon = 10$. Alors l'événement $\{|X - \mathbb{E}(X)| \geq 10\sigma_X\}$ a peu de chances de se réaliser car on a

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq 10\sigma_X) \leq \frac{1}{100}.$$

Supposons maintenant que $\sigma_X = 0$. Alors nous obtenons pour tout $\epsilon > 0$

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \epsilon) \leq 0.$$

Par conséquent

$$\mathbb{P}(|X - \mathbb{E}(X)| > 0) = 0,$$

et donc X est presque sûrement égale à $\mathbb{E}(X)$.

On doit cependant remarquer que, malgré son intérêt théorique certain, l'inégalité de Chebychev présente peu d'intérêt en pratique, car ne faisant pas intervenir la loi de probabilité suivie par la v.a.r. considérée, elle donne une majoration de la probabilité beaucoup trop grande.

4.4 Indépendance de variables aléatoires

DÉFINITION 28. — Deux v.a.r. X et Y sont dites indépendantes si et seulement si

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B), \quad \forall A, B \subset \mathbb{R}.$$

On peut montrer que l'indépendance est équivalente à

$$\mathbb{P}(X \leq a, Y \leq b) = \mathbb{P}(X \leq a)\mathbb{P}(Y \leq b), \quad \forall (a, b) \in \mathbb{R}^2,$$

ou encore en termes de fonctions de répartition :

$$F_{X,Y}(a, b) = F_X(a)F_Y(b), \quad \forall (a, b) \in \mathbb{R}^2.$$

THÉORÈME 29. — Soient X et Y deux v.a.r.

- Cas discret : X et Y sont indépendantes si et seulement si pour tout couple $(x, y) \in \mathcal{X} \times \mathcal{Y}$, on a $p_{X,Y}(x, y) = p_X(x)p_Y(y)$.
- Cas continu : X et Y sont indépendantes si et seulement si pour tout couple $(s, t) \in \mathbb{R}^2$, on a $f_{X,Y}(s, t) = f_X(s)f_Y(t)$.

5 Théorèmes limites

5.1 Introduction

Deux théorèmes mathématiques ont une place particulière en théorie des probabilités et en statistiques : la loi des grands nombres et le théorème central limite. Ils interviennent dans l'étude de phénomènes aléatoires comportant un grand nombre de v.a.r. indépendantes de même loi. Par exemple, pour le premier cité, il apparaît lorsque l'on étudie la proportion de "pile" dans un jeu de pile ou face, ou encore la moyenne de lancers de dé successifs. Quant au second, il nous donne de façon informelle une estimation précise de l'erreur que l'on commet en approchant l'espérance mathématique par la moyenne arithmétique.

5.2 Loi (faible) des grands nombres

THÉORÈME 30. — (LGN) Soient X_1, \dots, X_n des v.a.r. indépendantes, de même loi, et admettant une variance. On note $\mu = \mathbb{E}(X_1)$. Alors, pour tout

$\epsilon > 0$,

$$\mathbb{P}\left(\left|\frac{X_1 + \dots + X_n}{n} - \mu\right| > \epsilon\right) \xrightarrow{n \rightarrow \infty} 0.$$

Dans ce cas, on dit que la moyenne arithmétique $\frac{X_1 + \dots + X_n}{n}$ converge en probabilité vers l'espérance mathématique μ lorsque n tend vers $+\infty$.

5.3 Théorème central limite

On a vu que deux v.a.r. ont la même loi si et seulement si leur fonctions de répartition sont égales. Ainsi, la fonction de répartition est souvent utilisée en pratique afin de démontrer l'égalité en loi. On est donc amené à définir la convergence en loi comme la convergence des fonctions de répartition associées.

DÉFINITION 31. — Soit $(Y_n)_{n \in \mathbb{N}}$ une suite de v.a.r. et soit Y une v.a.r. On dit que $(Y_n)_{n \in \mathbb{N}}$ converge en loi vers Y si pour tout x_0 point de continuité de la fonction de répartition F_Y de Y ,

$$F_{Y_n}(x_0) = \mathbb{P}(Y_n \leq x_0) \xrightarrow{n \rightarrow +\infty} F_Y(x_0) = \mathbb{P}(Y \leq x_0).$$

On note la convergence en loi $Y_n \xrightarrow{\mathcal{L}} Y$.

La convergence en loi est réalisée aux points de continuité de F_Y . C'est la convergence simple de la suite de fonctions de répartition F_{Y_n} .

Propriété d'additivité de la loi normale : si X_1, \dots, X_n sont des v.a.r. indépendantes et de même loi $\mathcal{N}(\mu, \sigma^2)$, alors la v.a.r. $X_1 + \dots + X_n$ suit la loi $\mathcal{N}(n\mu, n\sigma^2)$. Ce résultat implique que la v.a.r. centrée réduite $\sum_{i=1}^n \frac{X_i - \mu}{\sigma\sqrt{n}}$ suit la loi normale $\mathcal{N}(0, 1)$. Que se passe-t-il dans le cas général où les v.a.r. X_i ne sont pas nécessairement normales ? Le résultat ci-dessus se transforme alors en un résultat de convergence en loi.

THÉORÈME 32. — (TCL) Soient X_1, \dots, X_n des v.a.r. indépendantes, de même loi, et admettant une variance. On note $\mu = \mathbb{E}(X_1)$ et $\sigma^2 = \text{Var}(X_1)$. Alors

$$\frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1) \quad \text{lorsque } n \rightarrow +\infty.$$

5.4 Approximation d'une loi binomiale

On contrôle n pièces et on introduit les v.a.r. X_1, \dots, X_n définies par $X_i = 1$ si la i -ème pièce contrôlée est défectueuse, et 0 sinon. On note $Y = X_1 + \dots + X_n$ le nombre total de pièces défectueuses dans le lot. Alors la v.a.r. Y suit une loi binomiale de paramètres (n, p) où p est la probabilité qu'une pièce soit défectueuse.

Approximation par une loi normale

Puisque les v.a.r. X_i sont indépendantes, de même loi, et de variance finie :

$$\mathbb{P}(X_i = 1) = p, \quad \mathbb{P}(X_i = 0) = 1-p, \quad \mathbb{E}(X_i) = p \quad \text{et} \quad \text{Var}(X_i) = p(1-p),$$

on peut appliquer le TCL :

$$\frac{X_1 + \dots + X_n - np}{\sqrt{np(1-p)}} \xrightarrow{\mathcal{L}} Z \sim \mathcal{N}(0, 1) \quad \text{lorsque} \quad n \rightarrow +\infty.$$

On peut donc approcher la loi de la v.a.r. $(Y - np)/\sqrt{np(1-p)}$ par une loi normale $\mathcal{N}(0, 1)$. Ceci revient à approcher la loi de Y par une loi $\mathcal{N}(np, np(1-p))$. En pratique, on utilise cette approximation si

$$\inf(np, n(1-p)) \geq 5.$$

Ainsi, si $p = 1/2$, l'approximation est correcte pour $n \geq 10$, par contre si $p = 1/100$, il faut $n \geq 500$ pour pouvoir l'utiliser.

Attention, une v.a.r. binomiale est une variable discrète à valeurs dans $\{1, \dots, n\}$, alors qu'une v.a.r. normale est continue et à valeurs dans \mathbb{R} . Par ailleurs, dans le cas d'une loi binomiale, un point a une probabilité non nulle alors que dans le cas d'une loi normale, un point est un ensemble de probabilité nulle. Pour ces deux raisons, il faut faire une "correction de continuité" quand on utilise l'approximation d'une loi binomiale par une loi normale.

ou encore

Approximation par une loi de Poisson

Lorsque p est très petit, on utilise plutôt l'approximation de la loi binomiale par une loi de Poisson, qui est satisfaisante pour $p < 0,1$ et $n > 50$.

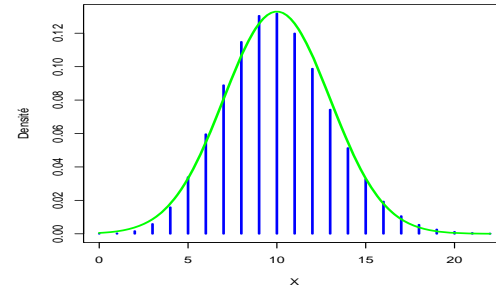


FIGURE 3 – Approximation d'une loi $\mathcal{B}(100, 0.1)$ par une loi $\mathcal{N}(10, 9)$.

PROPOSITION 33. — Soit Y_n une v.a.r. binomiale de paramètres (n, p) . On suppose que $n \rightarrow +\infty$ et $p = \lambda/n$, où $\lambda > 0$. Alors, pour tout $k \in \mathbb{N}$,

$$\lim_{n \rightarrow +\infty} P(Y_n = k) = \exp(-\lambda) \frac{\lambda^k}{k!}.$$

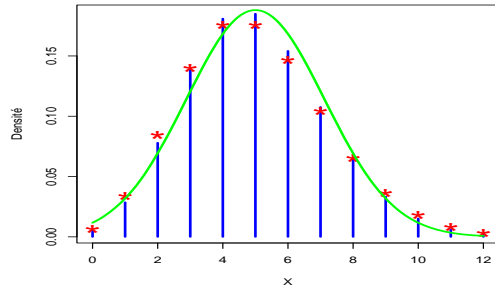


FIGURE 4 – Approximation d'une loi $\mathcal{B}(50, 0.1)$ (bleue) par une loi de Poisson(5) (rouge) et une loi $\mathcal{N}(5, 4.5)$ (verte).

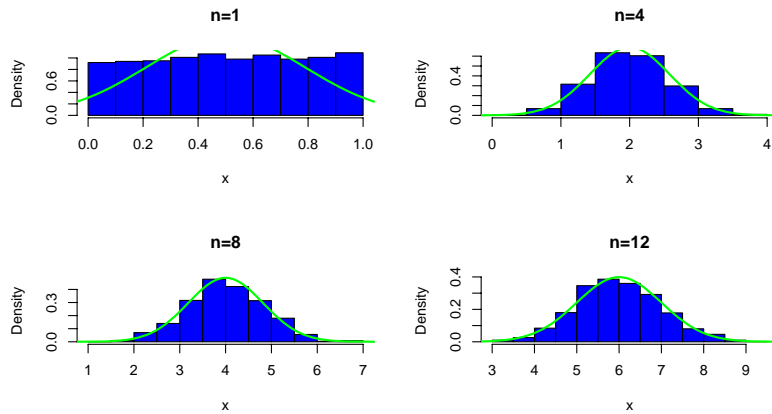


FIGURE 5 – Simulation d'une v.a.r. gaussienne par la somme de n v.a.r. uniformes.