

Petit Dictionnaire des Probabilités et de la
Statistique Élémentaires
L2 S4 Licence Informatique 2019-2020
Cours: Probabilités et Statistique

S. De Bièvre

22 mars 2020

Introduction

Ce qui suit est une liste, en ordre alphabétique, de la plupart des notions introduites dans le cours, avec des explications plus ou moins détaillées. Comme un dictionnaire ou une encyclopédie, il ne faut pas lire ce texte dans l'ordre, mais plutôt commencer par une notion que vous souhaitez apprendre, approfondir, revisiter, puis lire d'autres parties au fur et à mesure que vous en avez besoin. J'ai systématiquement mis les mots repertoriés en italique. Donc si une notion utilisée vous est inconnue, et qu'elle est en italique, vous pouvez la trouver expliquée ailleurs dans le texte.

La terminologie et les notations utilisées, même dans les probabilités et les statistiques élémentaires, sont relativement sophistiquées et il est important de ne pas se laisser dérouter par cette technicité apparente mais de saisir les notions intuitivement, c'est à dire, de comprendre. Ce qui n'est pas si simple non plus parce que même les probabilités et les statistiques élémentaires cachent des notions relativement sophistiquées. Les mathématiques utilisées, elles, sont par contre plutôt simples et d'un niveau qui ne dépasse pas de beaucoup un Bac S. La difficulté dans un premier cours de probabilités et de statistique n'est donc pas tellement de nature mathématique, contrairement à ce que certains prétendent.

Pour avancer dans la compréhension, il est essentiel et indispensable d'étudier de multiples exemples. Il y en a quelques-uns dans ce qui suit. Ce

document ne peut être utile qu'en complément de vos notes de cours, des livres que je vous ai recommandés (voir Moodle), des fiches de TD et de vos notes de TD. Vous y trouverez notamment d'autres exemples.

Samedi 21 mars 2020. J'ai jeté un coup d'oeil sur les premières réponses au sondage que nous avons lancé hier. On en a déjà une trentaine.

À la demande de certains, j'ai maintenant inclus un Index (page suivante). Cela vous permettra de mieux vous repérer. Et de voir rapidement quelles notions sont traitées dans le document. J'ai aussi commencé une nouvelle page pour chaque rentrée.

Certains se plaignent que certaines entrées sont trop longues : je m'efforce de donner des détails, c'est un choix.

Quelqu'un dit qu'il y a "trop de définitions". Cela me fait sourire : si elles sont là, c'est qu'elles sont essentielles. Désolé.

Certains disent que "beaucoup" de notions ne sont pas expliquées. Cela me paraît exagéré. Ceci dit, si vous me donnez des exemples, je les rajouterai. Je n'ai toutefois pas l'intention décrire un livre de probastat. Il y en a beaucoup et je vous ai donné deux références du bon niveau pour vous. Il faut les consulter (si vous y avez accès...).

Certains disent qu'ils n'arrivent pas à se concentrer sur un texte, qu'ils veulent voir des vidéos. Désolé, je n'ai pas le temps pour produire des vidéos. Et puis, quand les cours reprendront, il suffira de venir... Si je trouve des ressources video sur internet, je vous les indiquerai. Ceci dit, je suis de la vieille école et je pense que les livres (et donc les textes) sont des outils formidables pour apprendre.

Certains disent que le niveau mathématique est trop élevé. Comme j'ai dit à plusieurs reprises, il ne dépasse pas celui d'une terminale S. Et donc pour un L2 d'info, ce n'est pas trop.

À peu près deux tiers des répondant disent que c'est bien et utile. Merci à eux.

Index

Aléatoire, 4
Classe modale (d'un histogramme), 5
Densité de probabilité, 6
Écart type, 9
Écart type empirique, 11
Échantillon, 12
Effectif, 13
Espérance, 14
Évènement, 15
Évènement contraire, 17
Évènements indépendants, 18
Formule de König, 21
Formule de transfert, 19
Fréquence, 22
Histogramme, 23
Indépendance de variables aleatoires, 30
Indépendance d'événements, 29
Inégalité de Tchebychev, 30
Intervalle de confiance, 31
Loi d'une variable, 42
Loi de Pareto, 40
Loi de probabilité, 41
Loi exponentielle, 33
Loi forte des grands nombres, 35
Loi géométrique, 38
Loi normale, 39
Marche aléatoire, 44
Médiane, 45
Médiane empirique, 45
Mode, 45
Moyenne empirique, 46
Moyenne pondérée, 47
Ordinateur, 50
Population, 51
Probabilité, Une, 52
Probabilités, Les, 54
Processus stochastique, 56
Simuler, 57
Somme de variables aléatoires, 59
Statistique descriptive, 61
Statistique inférentielle, 61
Statistique, La, 61
Théorème Central Limite, 62
Théorème de Tchebychev, 64
Univers, 66
Variable aléatoire, 66
Variables aléatoires indépendantes, 66

Aléatoire

Dans la vie courante, on dit d'une chose, d'un événement, qu'il est aléatoire s'il est partiellement imprévisible ou incertain. L'exemple type est le lancer d'un dé : on ne sait pas quel chiffre apparaîtra lors du lancer d'un dé. On s'imagine néanmoins que, lorsqu'on lance le dé un grand nombre de fois, la fréquence d'apparition de l'un des six chiffres sur ses faces est la même pour tous, c'est à dire, une fois sur six ou donc à peu près dans 16% des cas. Ceci représente une situation d'équiprobabilité : toutes les issues possibles ont la même fréquence de réalisation et on dit qu'ils sont équiprobables, qu'elles ont la même probabilité. Lors d'une élection, l'issue est incertaine aussi. Mais les issues ne sont pas forcément équiprobables. En effet, supposons qu'il y ait 5 candidats, certains seront alors plus connus et plus soutenus que d'autres, et on attribue des probabilités de succès différentes aux différents candidats. Pour cela, on organise des sondages d'opinion, qu'on utilise pour prédire au mieux l'issue de l'élection : cela revient à essayer de réduire l'incertitude sur l'issue de l'élection, c'est à dire le caractère aléatoire du scrutin. L'issue d'un match de foot, ou d'un championnat de foot, ou de la Champions League est aussi incertaine. Dans ce cas, les experts de tout bord – journalistes sportifs et consultants – vous expliqueront tout de même quelle équipe a plus de chance que telle autre. En étudiant les résultats passés des équipes, en suivant les mercatos, en lisant les revues spécialisés, vous pouvez essayer de réduire l'incertitude sur les issues de ces événements sportifs et espérer gagner de l'argent en jouant sur les sites de pari en ligne qui vous proposent des gains si vous prédisez correctement les résultats sportifs. À noter que les exploitants de ces sites doivent eux-mêmes se faire une idée de ces mêmes probabilités afin d'éviter de perdre de l'argent dans l'opération. L'étude quantitative des phénomènes et situations aléatoires est le sujet d'une branche des mathématiques qu'on appelle “les *probabilités*”.

Classe modale (d'un histogramme)

La classe modale d'un *histogramme* est la classe ayant la plus grande *fréquence*. À comparer avec le mode de l'*échantillon*. Ce n'est pas tout à fait la même chose, mais l'idée est semblable : c'est la classe dans la quelle la caractéristique a la plus grande probabilité de tomber. Un histogramme peut être bi-modale ou multi-modale, tout comme un échantillon, bien sûr. Lorsqu'on connaît l'histogramme mais pas l'échantillon qui a servi à sa construction, on peut utiliser comme estimation du *mode* de l'échantillon, le centre de la classe modale.

Densité de probabilité

Une *densité de probabilité* est une fonction positive ρ , définie sur un interval I , et telle que

$$\int_I \rho(x) dx = 1.$$

Cette définition est simple et courte. Elle permet de répondre à des questions typiques d'interrogation écrite ou d'examen, comme : "Vérifier si les fonctions ci-dessous sont des densités de probabilité : $\rho(x) = \sin(x) \exp(-x)$, sur $x \geq 0$, et $\rho(x) = x \exp(-x)$ sur $x \geq 0$." Le seul intérêt de ces exercices est de vous permettre de gagner quelques points faciles aux contrôles. Mais elles ne permettent pas de voir si vous comprenez vraiment la notion introduite. La difficulté est conceptuelle et consiste à comprendre pourquoi on appelle "densité de probabilité" toute fonction positive dont l'intégrale vaut 1. En d'autres termes, que viennent faire les mots "densité" et "probabilité" ici ?

Étant donné une densité de probabilités sur un intervalle I , on peut définir, pour tout ensemble J contenu dans I ,

$$P(J) = \int_J \rho(x) dx.$$

On a $0 \leq P(J) \leq 1$ et il est clair que, si $J_1, J_2 \subset I$, et que $J_1 \cap J_2 = \emptyset$, alors $P(J_1 \cup J_2) = P(J_1) + P(J_2)$. La propriété analogue pour une famille J_n de sous-ensembles disjoints de I est également vraie. Il en résulte que P a les propriétés d'une *probabilité* sur I . Étant donnée une densité de probabilités, on peut donc construire une *probabilité*. Remarquez qu'il est donc facile de construire de multiples exemples de probabilités puisqu'il est facile de construire des fonctions positives d'intégrale 1. On connaît les densités des *lois uniformes, exponentielles et normales*, par exemple, mais d'autres densités sont faciles à construire : voir par exemple les *lois de Pareto* ou la Fig. 1.

Il est important d'associer des images à ces propos. Quand on vous présente une densité de probabilité, il faut d'abord la dessiner. Remarquons que l'aire sous son graphe est toujours égale à 1. Et que $P(J)$ est l'aire sous ce graphe au dessus de J . Considérons une famille d'intervalles $J_a = [a, a + \Delta]$, avec Δ un nombre fixé et pas trop grand, et $a \in I$, alors $P(J_a)$ changera quand vous déplacez a . Typiquement, $P(J_a)$ sera grand quand $\rho(a)$ est grand et petit quand $\rho(a)$ est petit. En d'autres termes, la probabilité est concentré là où la densité de probabilité est grande. Ceci est un élément d'interprétation important et utile. Plus précisément on peut dire ceci :

$$P(J_a) = \int_a^{a+\Delta} \rho(x) dx \simeq \rho(a) \Delta. \quad (1)$$

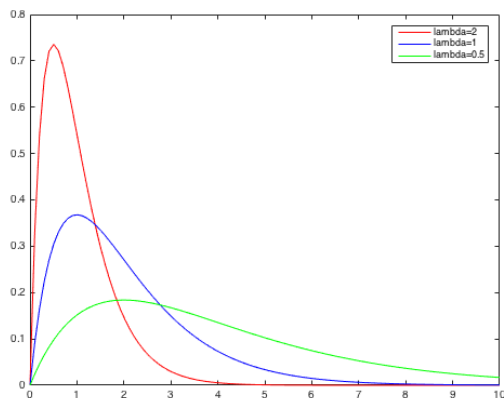


FIGURE 1 – Les graphes des densités de probabilité $\rho_\lambda(x) = \lambda^2 x \exp(-\lambda x)$ pour λ comme indiqué.

Ceci résulte de votre cours d'analyse de première année. Si $R(x)$ est une primitive de ρ , on a

$$\begin{aligned} P(J_a) &= \int_a^{a+\Delta} \rho(x) dx = R(a+\Delta) - R(a) \\ &= [R(a) + \rho(a)\Delta + O(\Delta^2)] - R(a) \\ &= \rho(a)\Delta + O(\Delta^2) \simeq \rho(a)\Delta. \end{aligned}$$

Ici j'ai utilisé que l'approximation affine de $R(x)$ au point $x = a$ est $R(x) = R(a) + R'(a)(x - a)$. On retiendra de (1) que la probabilité d'un petit intervalle J_a de longueur Δ est approximativement égale à $\rho(a)\Delta$. Si vous avez oublié votre cours d'analyse de première année, on peut aussi comprendre (1) en remarquant que $P(J_a)$ est l'aire sous le graphe de ρ entre a et $a + \Delta$, qui est approximativement égale à $\rho(a)\Delta$, c'est à dire l'aire du rectangle de base Δ et de hauteur $\rho(a)$.

Dans la Fig. 1 on voit trois densités de probabilité différentes. Imaginez-vous qu'elles modélisent le nombre d'heures qu'un patient attend dans un service d'urgence de trois hopitaux différents. Dans quel service préféreriez-vous aller ? Quel est le nombre moyen d'heures qu'on attend approximativement dans chaque service ? Dans quel service les fluctuations des nombres d'heures d'attente sont les plus prononcées ? On peut répondre à ces questions sans calculer quoi que ce soit. Des réponses précises nécessitent le calcul

de *l'espérance* et la *variance* des lois correspondantes, pas bien difficiles dans ce cas précis.

Écart type d'une variable aléatoire

L'écart type d'une *variable aléatoire* est la racine carrée de sa *variance*. Fort bien, mais avec cela on n'a rien dit tant qu'on n'a pas défini sa variance. On peut penser à l'écart type comme étant la distance moyenne qui sépare les valeurs de la variable aléatoire de leur moyenne. Techniquement, mathématiquement, ceci n'est pas tout à fait correct, mais peu importe, dans l'esprit c'est bien cela. Pour une variable aléatoire discrète qui prend des valeurs x_k avec une probabilité p_k , la *variance*, notée σ_X^2 ou simplement σ^2 si la confusion n'est pas possible est donnée par

$$\sigma_X^2 = \sum_k p_k (x_k - \mu)^2.$$

Ici

$$\mu = \mu_X = \sum_k p_k x_k = \mathbb{E}(X)$$

est l'*espérance* de X . On trouve aussi la notation

$$\mathbb{V}(X) = \sigma_X^2.$$

On remarquera que σ_X^2 est la *moyenne pondérée* des écarts $|x_k - \mu|$ entre les valeurs x_k de la variable aléatoire et sa son espérance μ , élevés au carré. L'analogie avec la *variance empirique* est évidente. L'écart type est donc notée σ_X ou σ . Lorsqu'on jette un dé, les valeurs x_k sont 1, 2, 3, 4, 5, 6 et $\mu_X = 3.5$. On a alors $p_k = 1/6$ et donc

$$\begin{aligned} \sigma_X^2 = \frac{1}{6} [(1 - 3.5)^2 + (2 - 3.5)^2 + (3 - 3.5)^2 + (4 - 3.5)^2 \\ + (5 - 3.5)^2 + (6 - 3.5)^2] = 2.92. \end{aligned} \quad (2)$$

Par conséquent

$$\sigma_X = 1.71.$$

On peut dire ceci : "Lorsqu'on jette un dé, l'*écart* entre le chiffre obtenu et le résultat moyen, qui est 3.5, est *typiquement* de l'ordre de 1.7." Si, au contraire, vous tirez un entier au hasard entre 1 et 10, la moyenne sera $\mu = \mathbb{E}(X) = 5.5$ et la variance

$$\begin{aligned} \sigma_X^2 = \frac{1}{10} [(1 - 5.5)^2 + (2 - 5.5)^2 + (3 - 5.5)^2 + \dots \\ + (8 - 5.5)^2 + (9 - 5.5)^2 + (10 - 5.5)^2] = 8.25. \end{aligned} \quad (3)$$

Par conséquent $\sigma_X = 2.87$. L'écart type est plus grand cette fois-ci, parce que les valeurs 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 sont plus éparpillées, elles sont en moyenne plus loin de la moyenne 5.5 que ne le sont les chiffres du dé de leur moyenne, qui est 3.5.

Considérons une variable aléatoire X qui suit une loi binomiale de paramètres n et p . On calcule alors facilement

$$\mathbb{E}(X) = np, \quad \mathbb{V}(X) = np(1-p), \quad \sigma_X = \sqrt{np(1-p)}.$$

Fixons $p = 0.33$ et varions n . Donc, par exemple, on jette un dé n fois et on considère qu'un succès correspond à l'obtention de 1 ou 2, ce qui correspond bien à $p = 1/3 = 0.33$. Le nombre moyen de succès, $\mathbb{E}(X)$ vaut np , ce qui est conforme à notre intuition : le nombre de succès moyen est proportionnel au nombre d'essais. Mais le nombre de succès peut être plus grand ou plus petit que np . L'écart type nous renseigne sur la taille de cette variation : elle donne une mesure de la fenêtre autour de $\mathbb{E}(X)$ dans laquelle le nombre de succès tombe le plus souvent. Par exemple, avec $n = 10$, $\mathbb{E}(X) = 3.3$ et $\sigma_X = 1.5$ tandis que, avec $n = 100$, $\mathbb{E}(X) = 33$ et $\sigma_X = 4.7$. En regardant les graphes de la Fig. 2, on constate que les probabilités p_k de k succès sont plus étalées dans le cas $n = 100$ que $n = 10$: observez bien l'échelle sur l'axe horizontal. Une autre façon de voir les choses est de dire que, plus la variance est grande, plus il y a de l'incertitude sur les valeurs de X . L'*Inégalité de Tchebycheff* donne un sens précis à ces différentes idées.

Pour une variable aléatoire à densité, on a

$$\sigma_X^2 = \int_{\mathbb{R}} (x - \mu)^2 \rho_X(x) dx, \quad \text{où } \mu = \mu_X = \int_{\mathbb{R}} x \rho_X(x) dx$$

Comme pour *l'espérance*, on peut lire approximativement la valeur de l'écart type sur un graphe. Par exemple, dans la Fig. 1, on voit que le graphe de la densité est de plus en plus étalé au fur et à mesure que λ augmente. Ceci se manifeste dans une croissance de l'écart type σ_{X_λ} des variables aléatoires X_λ en fonction de λ . On trouve

$$\mathbb{E}(X_\lambda) = \lambda^2 \int_0^{+\infty} x^2 \exp(-\lambda x) dx = \frac{2}{\lambda}, \quad \mathbb{E}(X_\lambda^2) = \lambda^2 \int_0^{+\infty} x^3 \exp(-\lambda x) dx = \frac{6}{\lambda^2},$$

$$\text{et donc } \mathbb{V}(X_\lambda) = \frac{2}{\lambda^2}, \quad \sigma(X_\lambda) = \frac{\sqrt{2}}{\lambda}.$$

Voir aussi : *Formule de König, Inégalité de Tchebycheff, variance empirique, variance d'une somme de variables aléatoires indépendantes.*

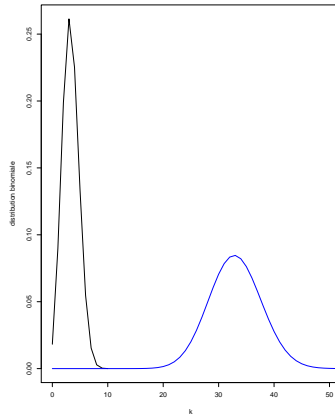


FIGURE 2 – Distributions binomiales avec $p = 0.33$ et $n = 10$ (en noir, à gauche) et $n = 100$ (en bleu, à droite). On remarque que la distribution pour $n = 100$ est nettement plus étalée autour de sa moyenne que celle avec $n = 10$, qui est plus resserrée.

Écart type empirique d'un échantillon

Notée s , c'est la racine carrée positive de la *variance empirique* de l'échantillon.

Échantillon

Un échantillon de taille N est un sous-ensemble de N individus d'une *population*. Pour chaque individu de l'échantillon, on s'intéresse à une ou plusieurs de ses caractéristiques. Ces dernières peuvent être quantitatives, comme l'âge, le revenu, la taille, le nombre d'enfants, le poids, ... Ou catégorielles (aussi appelées qualitatives) : sexe, nationalité, opinions politiques, état civil, ... Lorsque la caractéristique est quantitative, elle peut être caractérisée notamment par son *mode*, sa *moyenne empirique*, son *écart type* et sa *variance empirique*. Une représentation graphique concise et parlante d'une caractéristique quantitative d'un échantillon est son *histogramme*.

Effectif

L'effectif N_k d'une classe I_k d'un *échantillon* est le nombre d'observations qu'il contient. S'il y a K classes, on a donc $\sum_{k=1}^K N_k = N$, où N est la taille de l'échantillon.

Voir aussi : *histogramme, fréquence absolue et relative.*

Espérance

Pour une variable aléatoire X discrète, prenant les valeurs $\{x_1, x_2, \dots, x_K\}$, l'espérance est définie par

$$\mathbb{E}(X) = \sum_{k=1}^K x_k P(X = x_k).$$

On peut avoir $K = +\infty$. Dans ce cas la somme se transforme en série. Il s'agit donc d'une moyenne pondérée qu'on peut aussi penser comme un barycentre, pour ceux qui savent ce que cela signifie. Dans ce cas, les x_k sont les positions des masses et les $P(X = x_k)$ sont remplacées par m_k/M , où les m_k sont les différentes masses et M est la mass totale : $M = \sum_{k=1}^M m_k$.

Lorsque une variable aléatoire X a une densité ρ , son espérance est définie par l'expression

$$\mu_X = \mathbb{E}(X) = \int_{\mathbb{R}} x \rho(x) dx.$$

Dans cette vision, il faut penser ρ comme un densité de masse.

Par exemple, pour des variables aléatoires X_λ ayant comme densité ρ_λ de la Fig. 1, on trouve

$$\mathbb{E}(X_\lambda) = \lambda^2 \int_0^{+\infty} x^2 \exp(-\lambda x) dx = \frac{2}{\lambda}.$$

Remarquez que, au fur et à mesure que λ diminue, le “centre” de la figure se déplace vers la droite ; ceci est conforme avec le fait que $\mathbb{E}(X_\lambda) = \frac{2}{\lambda}$ augmente au fur et à mesure que λ diminue.

Attention, il est complètement faux de dire ou de penser que l'espérance de X correspond au maximum de sa densité ρ_X . Les contre-exemples abondent. Pour la loi ρ_λ par exemple, le maximum se trouve à $x = \frac{1}{\lambda}$.

Évènement

On appelle “évènement” associé à une expérience aléatoire tout sous-ensemble de son *univers* Ω . Lorsqu’on lance deux dés, l’univers est $\Omega = \llbracket 1, 6 \rrbracket \times \llbracket 1, 6 \rrbracket$. Un exemple d’évènement est le sous-ensemble des issues défini par “la somme des dés vaut 7”. Désignons-le par A . On a donc

$$A = \{(1, 6), (6, 1), (2, 5), (5, 2), (3, 4), (4, 3)\}.$$

Alternativement, il est donné par

$$A = \{\omega \in \Omega \mid \omega_1 + \omega_2 = 7\}.$$

En général, on désigne par des lettres majuscules A, B, C, \dots les sous-ensembles (aussi appelés parties) de Ω . Ils sont dans les applications le plus souvent déterminés par une propriété portant sur les issues ω appartenant à Ω . Un second exemple est

$$B = \{\omega \in \Omega \mid \omega_1 \omega_2 = 12\}.$$

Explicitement, on a

$$B = \{(2, 6), (6, 2), (3, 4), (4, 3)\}.$$

Et en mots : “ B est l’évènement correspondant à tous les lancer pour lesquels le produit des chiffres vaut 12.” En toute généralité, un évènement se présente donc sous la forme

$$A = \{\omega \in \Omega \mid P(\omega)\},$$

où $P(\omega)$ exprime une propriété de ω . L’énumération explicite de toutes les issues pour lesquelles la propriété est vérifiée est rarement possible, puisqu’elles peuvent être trop nombreuses. Par exemple, suppose qu’on lance N fois une pièce. Prenons $N = 2000$. Quelles sont les issues où il y a au moins une fois face ? Les énumérer serait fastidieux, voire impossible. On peut identifier l’univers correspondant à cette expérience comme étant $\Omega = \{0, 1\}^N$, avec $N = 2000$, en s’accordant que 1 correspond à “pile” et 0 à “face”. Appelons F l’évènement considéré. Une façon très concise de le décrire est

$$F = \{\omega \in \Omega \mid \omega_1 \omega_2 \dots \omega_N = 0\}.$$

En effet, un “face” au n -ème lancer correspond à $\omega_n = 1$. Donc on aura au moins une fois face si et seulement si le produit $\omega_1 \omega_2 \dots \omega_N$ est nul.

Supposons maintenant qu'on a deux évènements A et B , déterminés respectivement par des propriétés $P(\omega)$ et $Q(\omega)$:

$$A = \{\omega \in \Omega \mid P(\omega)\}, \quad B = \{\omega \in \Omega \mid Q(\omega)\}.$$

On a alors

$$A \cap B = \{\omega \in \Omega \mid P(\omega) \text{ et } Q(\omega)\}, \quad A \cup B = \{\omega \in \Omega \mid P(\omega) \text{ ou } Q(\omega)\}.$$

On se rappellera que ω appartient à $A \cap B$ si et seulement si ω appartient à A ET à B . Donc si et seulement si $P(\omega)$ ET $Q(\omega)$ sont vraies. Par ailleurs, ω appartient à $A \cup B$ si et seulement si ω appartient à A OU à B . Donc si et seulement si $P(\omega)$ OU $Q(\omega)$ sont vraies.

Évènement contraire

Lorsque Ω est un *univers* et $A \subset \Omega$ un évènement, on appelle $\overline{A} = \Omega \setminus A$ l'évènement contraire de A . On a

$$\overline{A} = \{\omega \in \Omega \mid \omega \notin A\}.$$

Une autre notation fréquemment rencontrée est $\overline{A} = A^c$. Lorsque $A = \{\omega \in \Omega \mid P(\omega)\}$, on a

$$\overline{A} = \{\omega \in \Omega \mid \text{not} P(\omega)\}.$$

Si P est une *probabilité* sur Ω , on a $P(\overline{A}) = 1 - P(A)$.

Événements indépendants.

Soit Ω l'univers associé à une expérience aléatoire et P une probabilité sur Ω . On dit que deux événements A et B sont indépendants lorsque

$$P(A \cap B) = P(A)P(B).$$

Cette définition très brève et d'apparence très simple est censée capter l'idée d'indépendance de la vie courante. Voici un exemple. Si on considère toute la population française, on peut constater que 70% des adultes portent des lunettes. Si on considère seulement les personnes entre 20 et 25 ans, ce pourcentage est nettement moins élevé : 45%. Les événements $A = \{\text{“porter des lunettes”}\}$ et $B = \{\text{“avoir entre 20 et 25 ans”}\}$ ne sont donc pas indépendants. En effet,

$$\frac{P(A \cap B)}{P(B)} = 0.45, \quad \text{et} \quad P(A) = 0.74$$

et donc

$$P(A \cap B) \neq P(A)P(B).$$

Dans ce cas précis, ceci signifie que la probabilité de porter des lunettes augmente avec l'âge, c'est à dire qu'elle n'est pas indépendante de l'âge. La définition mathématique semble donc bien capter une intuition évidente : tout le monde sait que, avec l'âge, on voit moins bien.

Lorsqu'on a plusieurs événements A_1, \dots, A_n , la définition est plus complexe. Regardons le cas où $n = 3$. On dit que A_1, A_2, A_3 sont indépendants si

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3), \quad P(A_i) \cap P(A_j) = P(A_i)P(A_j),$$

pour tout $1 \leq i < j \leq 3$. En d'autres termes, il faut A_1 et A_2 soient indépendants, ainsi que A_1 et A_3 et A_2 et A_3 . Mais aussi que $P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3)$.

Voir aussi : *variables aléatoires indépendantes.*

Formule de transfert.

Lorsque X est une variable aléatoire, on est souvent amené à calculer $\mathbb{E}(f(X))$, où f est une fonction. L'exemple le plus simple est le cas où $f(x) = x^2$, donc $\mathbb{E}(X^2)$, qui intervient dans le calcul de la *variance*. A priori, si on regarde la définition de l'*espérance*, pour calculer $\mathbb{E}(f(X))$, il faut connaître la loi de $f(X)$. Mais, même si on connaît la loi de X , il n'est pas toujours facile de déterminer la loi de $f(X)$, par contre. À titre d'exemple, c'est un bon exercice (pas tout à fait facile) de construire la loi de $Y = X^2$, lorsque X suit une loi uniforme sur $[-a, a]$. On peut montrer que Y est une variable aléatoire ayant une loi à densité $\rho_Y(y)$, donnée par

$$\rho_Y(y) = \frac{1}{2a\sqrt{y}}, \quad 0 \leq y \leq a^2, \quad \rho_Y(y) = 0 \quad \text{sinon.}$$

Une façon empirique de mettre en évidence la loi de Y et de confirmer numériquement cette expression est de la *simuler*. Demander à R de vous donner N valeurs aléatoires uniformément distribuées entre $-a$ et a : en d'autres termes, de *simuler* une variable aléatoire X de loi uniforme sur $[-a, a]$. Calculer, toujours avec R, le carré de ces valeurs : elles fournissent une simulation de la variable Y . Vous pouvez alors faire un histogramme de ces valeurs. Voici le code :

```
N=100;
a=6;
X<-runif(N, -a,a);
Y<-X^2
hist(Y, freq=FALSE)
curve((2*a)^(-1)*(sqrt(x))^( -1), add=TRUE)
```

Si vous testez ce code pour différentes valeurs de a et de N , vous pourrez apprécier comment les valeurs de $Y = X^2$ sont distribuées : déjà sûrement pas uniformément, puisqu'il y en a bien plus proche de zéro que loin. On constate également que la densité de probabilité ρ_Y ci-dessus épouse en effet très bien l'histogramme lorsque N est pris de plus en plus grand. Cette expérience donne donc une confirmation empirique de la formule pour ρ_Y donnée ci-dessus.

La formule de transfert permet de calculer $\mathbb{E}(f(X))$ SANS calculer d'abord la loi de $f(X)$. Pour une variable aléatoire discrète, elle s'écrit

$$\mathbb{E}(f(X)) = \sum_k f(x_k)P(X = x_k),$$

et pour une variable à densité

$$\mathbb{E}(f(X)) = \int f(x)\rho(x)dx.$$

La preuve du cas discret a été donnée en cours. Celle du cas où la variable admet une densité se comprend par analogie avec la précédente, à condition d'avoir compris ce qu'est la *loi d'une variable aléatoire* et une *densité de probabilité*.

C'est un bon exercice de calculer $\mathbb{E}(Y)$ avec $Y = X^2$ et $X \sim \mathcal{U}(-a, a)$ d'une part en utilisant la densité ρ_Y ci-dessous et de l'autre avec la formule de transfert. On obtient $\mathbb{E}(Y) = \frac{a^2}{3}$.

Formule de König

ATTENTION : en cours, le vendredi 13 mars, j'ai appelé ceci la "Formule de Kramer." C'est faux. Désolé.

La formule de König dit que

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2.$$

Elle est parfois utile dans les calculs. Elle est facile à démontrer, par exemple pour une *variable aléatoire discrète* X , qui prend les valeurs distinctes x_1, \dots, x_K , avec probabilité $p_k = P(X = x_k)$. En effet, avec $\mu = \mathbb{E}(X)$, on a

$$\begin{aligned} \mathbb{V}(X) &= \sum_{k=1}^K (x_k - \mu)^2 p_k = \sum_{k=1}^K (x_k^2 p_k - 2\mu p_k + \mu^2 p_k) \\ &= \mathbb{E}(X^2) - 2\mu \mathbb{E}(X) + \mu^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2. \end{aligned} \quad (4)$$

Ici, on a utilisé successivement, la définition de $\mathbb{V}(X)$, le développement du carré, la *formule de transfert* et la définition de l'espérance.

On remarque que, comme $\mathbb{V}(X) \geq 0$ (c'est une somme de termes positifs), on a toujours que

$$\mathbb{E}(X^2) \geq \mathbb{E}(X)^2.$$

On peut se demander s'il est possible pour $\mathbb{V}(X)$ d'être nulle. Voici la réponse :

$$\mathbb{V}(X) = 0 \Leftrightarrow P(X = \mu) = 1.$$

En d'autres termes, si $\mathbb{V}(X)$ est nulle, la variable aléatoire ne prend plus qu'une seule valeur μ , qui est alors automatiquement son espérance. Ceci signifie qu'une variable aléatoire dont la variance est nulle n'est, de ce fait, pas aléatoire : sa valeur est sûre. Pour démontrer ceci, il suffit de remarquer que, si

$$0 = \mathbb{V}(X) = \sum_{k=1}^K (x_k - \mu)^2 p_k,$$

alors chaque terme est nul, puisque tous les termes sont positifs. Donc, pour tout k ,

$$(x_k - \mu)^2 p_k = 0.$$

Par conséquent, pour tout k , soit $p_k = 0$, soit $x_k = \mu$. En conclusion, $p_{k_*} \neq 0$ pour une seule valeur k_* de k , celle où $x_{k_*} = \mu$. Comme $\sum_{k=1}^K p_k = 1$, on a $p_{k_*} = 1$. On conclut que, en effet, X prend la valeur $\mu = x_{k_*}$ avec probabilité 1.

Fréquence absolue et relative

La fréquence absolue d'une classe I_k d'un *échantillon* est son *effectif* N_k . Sa fréquence relative est $f_k = \frac{N_k}{N}$.

Voir aussi. *histogramme*.

Histogramme (d'un échantillon)

Supposons que vous disposez d'une liste L de nombres réels,

$$L = (\ell_1, \ell_2, \ell_3, \ell_4, \dots, \ell_N).$$

Typiquement, on pensera à une liste de données, comme par exemple la taille en centimètres de tous les hommes de 25 ans résidant en France :

172, 184, 161, 167, 194, 183, 172, 173, 196, 161, 174, 176, 168, 169,
184, 193, 181, 156, 199, 174, ...

Dans ce cas N serait de l'ordre de quelques millions si on disposait de la liste complète. En pratique, on ne dispose que d'un *échantillon*, c'est à dire une liste incomplète, qui peut contenir quelques dizaines, quelques centaines ou plusieurs milliers de données. Ci-dessus, j'ai pris $N = 20$ dans un premier temps pour illustrer mes propos, mais normalement, N est bien plus grand. On verra des exemples ci-dessous. En quoi une telle liste peut-elle être intéressante ? Si vous souhaitez, par exemple, vendre des pantalons d'homme sur le marché français, il est clairement important de savoir combien de tels pantalons de chaque taille fabriquer. Mais l'inspection de la liste des tailles des hommes donnée ci-dessus ne vous donne pas directement une information utile à cette fin. Ce qu'il faudrait savoir est le nombre d'hommes qui mesurent entre 155 et 160 cm, entre 160 et 165 cm, entre 165 et 170 cm, et ainsi de suite. En d'autres termes, il faudrait ranger les hommes dont vous connaissez la taille dans des boîtes : une boîte pour ceux qui mesurent entre 155 et 160 cm, une pour ceux entre 160 et 165 cm, une pour ceux entre 165 et 170 cm, et ainsi de suite. Une bonne dizaine de boîtes, donc. Dans le jargon de la statistique descriptive, on appelle ces boîtes des "classes". On désigne par I_k l'intervalle correspondant à la k -ème boîte/classe. En général, lorsqu'il y a K telles classes, désignons par N_k le nombre de valeurs dans la liste qui sont dans I_k , pour $k = 1 \dots K$. On appelle N_k l'*effectif de la classe* I_k ou encore sa *fréquence absolue*¹. Par exemple, pour la liste de 20 tailles donnée ci-dessus, on a

$$N_1 = 1, N_2 = 2, N_3 = 3, N_4 = 5, N_5 = 1, N_6 = 3, N_7 = 1, \\ N_8 = 2, N_9 = 2, N_{10} = 0.$$

L'information contenue dans ces $K = 10$ nombres est très bien résumée dans la représentation graphique qui consiste à indiquer sur l'axe des abscisses les bords des classes, et sur l'axe des ordonnées les N_k , comme dans la

1. La terminologie varie d'un livre à l'autre, d'un site web à l'autre, malheureusement.

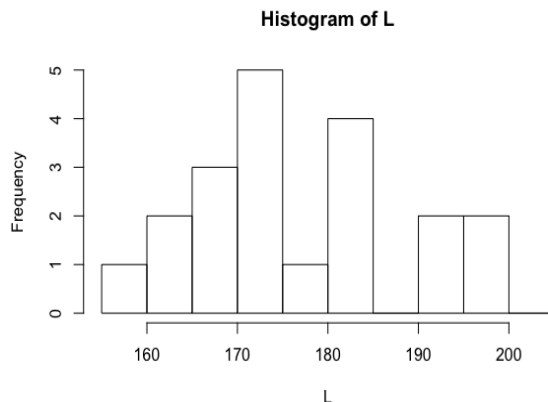


FIGURE 3 – Histogramme des effectifs ou des fréquences absolues d’un échantillon des tailles de $N = 20$ hommes de 25 ans résidant en France. Remarquer que la somme des hauteurs fait bien $N = 20$.

Fig. 3 où l’on voit au dessus de la k ème boîte un rectangle de hauteur N_k . C’est ce qu’on appelle l’histogramme des *effectifs* (ou des *fréquences absolues*) de l’échantillon. L’histogramme fournit une représentation graphique des données, qui se veut parlante. On verra comment en extraire des informations utiles et intéressantes sur l’échantillon d’abord, sur la *population* ensuite (*statistique inférentielle*).

On notera qu’on peut faire plusieurs histogrammes différents à partir de la liste L initiale. En effet, il y a une certaine liberté dans le choix des classes dont le nombre et les tailles peuvent être choisis librement. L’idée est de choisir les classes de façon à ce qu’il y ait assez d’éléments de la suite L dans chaque classe, mais pas trop. Les classes doivent donc n’être ni trop larges ni trop étroites. On notera que typiquement K est beaucoup plus petit que N . Dans l’exemple ci-dessous, l’échantillon n’est intéressant que si N est grand, disons au moins de quelques dizaines ou centaines. Mais K est fixé, typiquement entre 5 et 15 à peu près. En pratique, on variera un peu K , pour voir quel choix semble donner l’information la plus claire sur l’échantillon. Avec un ordinateur, c’est facile à faire, voir ci-dessous.

Comment décider concrètement du nombre de classes, c’est à dire de K ? Il existe quelques règles pour le faire, dont celle-ci :

$$K = 1 + \log_2 N, \quad \text{ou} \quad 2^{K-1} = N.$$

Donc, si $N = 100$, comme $2^7 = 128$, on peut prendre $K = 8$. Pour quelques explications simples et complémentaires sur le choix de K , voir <https://fr.wikipedia.org/wiki/Histogramme>.

Ci-dessus, j'ai déterminé les N_k (appelés les *fréquences*) "à la main". Cela se fait facilement :

- Rangez d'abord la liste dans l'ordre croissante ;
- Comptez combien de nombres de la liste se trouvent dans chaque classe.

Si N est grand, c'est forcément long et fastidieux. Ce n'est pas étonnant, le rangement, c'est toujours fastidieux et ennuyeux !

Heureusement, un ordinateur équipé avec un logiciel adapté peut faire le travail pour vous en un clin d'oeil. Le logiciel R, par exemple, propose une commande `hist(L)` qui vous donne immédiatement l'histogramme de la liste L . Vous pouvez indiquer, en utilisant l'argument `breaks`, la taille des boîtes que vous souhaitez qu'il utilise mais si vous ne dites rien du tout, il se débrouillera pour faire un choix raisonnable de K . Voici par exemple un échantillon de $N = 400$ tailles d'homme adulte :

```
[1] 158 161 161 175 172 158 193 179 175 165 163 161 158
175 161 185 173 161 176 157 175 183
[23] 174 162 158 161 157 166 163 167 171 169 174 160 158 161
163 159 161 164 157 175 190 161
[45] 173 176 170 175 162 184 163 172 170 171 164 173 164
152 168 179 180 170 169 158 155 173
[67] 167 157 159 165 162 174 163 163 179 167 157 175
175 153 169 170 164 163 165 185 158 156
[89] 172 175 162 166 166 155 165 157 167 161 164 170
177 184 168 188 172 160 170 161 173 179
[111] 171 159 169 158 163 164 162 168 168 175 165 158
174 168 164 182 161 166 165 151 167 172
[133] 171 179 179 173 165 187 173 163 162 172 183 164
172 158 187 168 159 174 159 159 167 181
[155] 173 175 161 178 180 160 182 158 170 161 184 157
166 167 189 161 189 169 179 175 164 156
[177] 158 161 184 172 167 168 161 174 170 162 162 165
167 159 180 179 166 168 167 188 178 165
[199] 168 166 172 157 175 170 159 172 158 160 154 182
183 167 169 162 177 174 158 175 169 161
[221] 175 175 177 171 155 178 167 180 172 172 168 154
158 157 158 183 182 182 158 167 171 163
```

[243] 167 177 157 177 176 153 180 163 176 180 164 184
 174 164 157 178 179 183 178 177 162 177
 [265] 158 162 167 165 169 159 165 164 178 168 164 171
 163 188 160 169 157 176 165 176 184 189
 [287] 167 159 151 162 174 158 156 161 179 170 171 184
 171 166 180 162 181 167 168 186 158 169
 [309] 176 166 188 165 178 174 163 180 169 170 161 183
 172 156 181 168 176 176 158 186 178 157
 [331] 169 167 165 180 170 172 174 159 181 177 172 164
 160 158 184 182 159 165 178 162 173 176
 [353] 174 161 178 182 156 159 182 162 168 169 168
 166 169 168 169 177 159 173 160 160 172 157
 [375] 169 170 159 173 161 164 173 186 177 176 156
 176 155 165 172 157 170 165 170 168 165 161
 [397] 178 157 158 164

Il est clair que contempler cette liste n'est pas particulièrement instructif. Si j'avais donné le nom de chaque homme, on aurait su que Jean mesure 172 cm et Mohamed 176 cm, mais cette information très détaillée ne nous sert à rien sauf si Jean ou Mohamed achètent effectivement un pantalon. Ce qu'il nous faut est une information sur le nombre d'hommes mesurant 172 cm ou 176 cm.

Faire l'histogramme à la main prendrait énormément de temps, clairement. L'histogramme correspondant, fait avec R, est à gauche dans la Fig. 4. J'ai imposé des `breaks` par `breaks=seq(from=155, to=200, by=5)`. En d'autres termes, j'ai fixé la taille des classes, pas leur nombre. On peut alternativement fixer leur nombre en écrivant par exemple `breaks=5` : R se débrouillera pour faire à peu près 5 breaks alors. À droite, on trouve l'histogramme lorsque $N = 1200$. On remarquera que ces deux histogrammes se ressemblent beaucoup. Dans une boîte donnée, on trouve à peu près trois fois plus de données à droite qu'à gauche, ce qui semble indiquer que la fraction de données $f_k = N_k/N$ (appelées *fréquences relatives*) varie très peu avec N , si N est grand. Ceci explique que la forme globale de l'histogramme est très semblable dans les deux cas.

On peut encore remarquer que $\sum_{k=1}^K f_k = 1$, ce qui signifie que les f_k forment une *probabilité*. On peut alors conclure qu'il est raisonnable de considérer que f_k représente la probabilité qu'un homme de 25 ans, choisi au hasard parmi les hommes de 25 ans habitant en France, ait une taille dans I_k .

La justification d'une telle hypothèse nécessite davantage de travail, mais

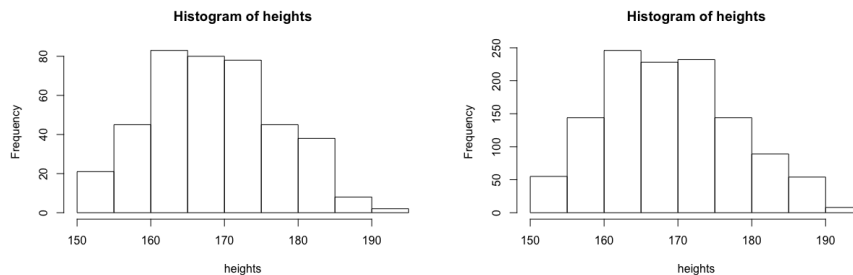


FIGURE 4 – Histogramme des effectifs d’un échantillon des fréquences des tailles de $N = 400$ (à gauche) et de $N = 1200$ (à droite) hommes de 25 ans résidant en France.

elle est intuitive et paraît a priori raisonnable. Pour mieux la justifier, il faut notamment expliquer ce qu’on entend par “choisir un homme au hasard” et aussi s’interroger sur la façon dont on a constitué l’échantillon utilisé pour l’analyse. Si, par exemple, tous les hommes de l’échantillon sont membres d’un club de basket, la distribution de leurs tailles ne sera pas représentative de celle de la population entière.

On entre alors dans le domaine de la *statistique inférentielle*. Le but est de choisir intelligemment un échantillon, puis d’utiliser les données de cette échantillon pour déduire (=inférer) des propriétés de la population toute entière. Ce n’est pas une mince affaire.

Pour le moment, acceptons cette interprétation des f_k . Il est clair que l’information dans l’histogramme est utile au fabricant de pantalons. S’il considère que les hommes qui achèteront ses pantalons sont pris au hasard dans la population, il doit fabriquer des pantalons de différentes tailles en nombre proportionnel au f_k .

Finalement, supposons que je veux connaître la probabilité qu’un homme de 25 ans, choisi au hasard parmi les hommes de 25 ans habitant en France, mesure ℓ cm. Il est alors naturel de procéder ainsi. Supposons par exemple que $\ell = 167$, qui est compris entre 165 et 170 cm. Comme $\ell \in I_3 = [165, 170[$, on peut penser que la réponse est approximativement $f_3/5$, si N est assez grand. Plus généralement :

$$\begin{aligned} d_k &= P(\text{“un homme de 25 ans résidant en France mesure } \ell \text{ cm”}) \\ &= \frac{\text{“nombre d’hommes de taille } \ell\text{”}}{\text{“nombre total d’hommes”}} \simeq \frac{f_k}{|I_k|}, \end{aligned}$$

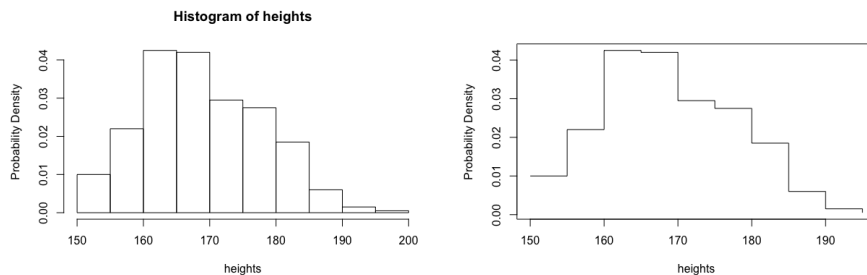


FIGURE 5 – À gauche, l’histogramme de la densité de probabilité $f_k/|I_k|$ pour des échantillons de la taille de $N = 400$ hommes de 25 ans résidant en France. À droite, la fonction en escalier correspondante. Observer que l’aire des rectangles dans la figure de gauche et l’aire sous le graphe dans la figure de droite valent 1.

où I_k est l’intervalle auquel appartient ℓ .

Si on exécute en R la commande `hist(L, freq=FALSE)`, le logiciel trace la fonction qui à chaque ℓ associe la valeur de $\frac{f_k}{|I_k|}$ correspondante. C’est ce qui a été fait dans le panneau de gauche de la Fig. 5. On remarquera, en examinant la figure, que l’aire sous le graphe de cette figure vaut 1. On comprend facilement pourquoi. En effet, cette aire est la somme des aires des rectangles, qui est donnée par

$$\sum_k d_k |I_k| = \sum_{k=1}^K \frac{f_k}{|I_k|} |I_k| = 1.$$

Le graphe montre donc une fonction positive, d’intégrale 1, c’est à dire une *densité de probabilité*. Pour mieux le mettre en évidence, on voit dans le panneau de droite le graphe de cette densité qui est une fonction en escalier obtenue en effaçant les lignes verticales de l’histogramme. C’est pour cela que dans le `help` de R, on lit que, si on introduit `freq=FALSE` comme argument dans `hist(L)`, alors on obtient une *densité de probabilité*.

Voir aussi : *moyenne pondérée*.

Indépendance d'événements.

Voir : *événements indépendants.*

Indépendance de variables aléatoires.

On dit que une famille X_1, X_2, \dots, X_n de variables aléatoires à valeurs dans \mathbb{R}^d et définies sur un même espace probabilisé (Ω, P) est indépendante si pour tout $A_1, \dots, A_n \in \mathbb{R}^d$ on a

$$P(X_1 \in A_1, X \in A_2, \dots, X_n \in A_n) = P(X_1 \in A_1)P(X \in A_2) \dots P(X_n \in A_n).$$

La définition est donc une adaptation évidente de la notion d'événements indépendants. Pour bien la comprendre, il est essentiel de regarder un grand nombre d'exemples de familles de variables aléatoires indépendantes et dépendantes afin de se forger une intuition sur la notion.

Inégalité de Tchebychev

Une entrée à venir.

Intervalle de confiance

Il ne sera question ici que des intervalles de confiance pour la moyenne de la population μ . On peut en construire pour d'autres caractéristiques de la *population*, également, mais on ne le fera pas ici. Voici la définition : étant donné donc un échantillon de taille n , ayant une moyenne empirique m_n et un écart type s_n , l'intervalle de confiance (pour la moyenne) de seuil (ou de niveau) $(1 - \alpha)$ est l'intervalle

$$\left[m_n - z \frac{s_n}{\sqrt{n}}, m_n + z \frac{s_n}{\sqrt{n}} \right].$$

Ici $z \geq 0$ est lié à $0 \leq \alpha \leq 1$ de la façon suivante :

$$\frac{\alpha}{2} = \frac{1}{\sqrt{2\pi}} \int_z^{+\infty} \exp\left(-\frac{x^2}{2}\right) dx.$$

En utilisant une table de la loi normale, on peut trouver α étant donné z et vice versa. Par exemple $\alpha = 0.05$ et donc $(1 - \alpha) = 0.95$ correspond à $z = 1.96 \simeq 2$. Et $\alpha = 0.10$ et donc $(1 - \alpha) = 0.90$ correspond à $z = 1.645$. Avec ces informations, on sait maintenant calculer n'importe quel intervalle de confiance (pour la moyenne d'une population). C'est un tout petit peu technique, à cause du lien entre z et α , mais pas trop tout de même. Par contre, cette définition ne nous dit absolument rien sur la signification de cet intervalle. À quoi est-il bon, quel information nous fournit-il ? La réponse est celle-ci :

« Pour $100(1 - \alpha)\%$ des échantillons de taille n assez grande, la moyenne μ de la population se trouve dans l'intervalle de confiance de cet échantillon. »

Supposons par exemple qu'on ait pris $\alpha = 0.05$. Dans ce cas, cela signifie que, si vous choisissez un grand nombre d'échantillons dans votre population, alors dans 95% des cas, la moyenne de la population (que vous ne connaissez pas !) se trouve dans l'intervalle de confiance correspondant. On peut aussi dire, de façon parfaitement équivalente :

« Pour $100(1 - \alpha)\%$ des échantillons de taille n assez grande, l'intervalle de confiance de cet échantillon contient la moyenne μ de la population. »

Les formulations suivantes se trouvent souvent dans les textes et manuels divers :

« La probabilité que la moyenne μ de la population se trouve dans l'intervalle de confiance est $(1 - \alpha)$. »

« La probabilité que l'intervalle de confiance contient la moyenne μ de la population est $(1 - \alpha)$. »

Ce sont des formulations plus courtes, mais moins claires. En effet, on ne voit pas trop de quelle probabilité, de quelle expérience aléatoire on parle. Il s'agit de l'ensemble de tous les échantillons possibles ! La probabilité fait donc référence à une fraction de ces échantillons, ce que les formulations précédentes rendent plus claires. Finalement, attention à ceci : la moyenne μ de la population n'est pas du tout aléatoire, elle a une valeur précise et unique. Par contre, une phrase qui commence par « La probabilité que μ ... » donne fortement l'impression que c'est μ qui est aléatoire, ce qui est donc faux. Ce sont les intervalles de confiance qui sont aléatoires, puisqu'elles dépendent des échantillons.

Si vous pensez que vous avez compris ceci à la première lecture, vous vous trompez vraisemblablement : c'est bien plus subtil qu'il n'y paraît et demande un certain temps de réflexion afin d'être compris. Si, au contraire, vous pensez que c'est incompréhensible, vous vous trompez également. En ce donnant le temps de la pratique et de la réflexion, c'est tout à fait compréhensible et d'ailleurs fort intéressant.

Pour finir, un petit exemple, qui révélera la subtilité. Supposons que vous voulez savoir combien d'heures par jour un étudiant universitaire en France passe à pianoter sur son téléphone, en moyenne. Vous faites une étude sur $n = 150$ étudiants et vous trouvez que $m_n = 6h30$, $s_n = 0h30$. L'intervalle de confiance à 95% est donc $[6h25, 6h35]$. Selon ce qui précède, cela signifie que, si vous prenez au hasard un échantillon de 150 étudiants, vous pouvez affirmer que dans 95% des cas, la moyenne μ recherchée se trouve dans l'intervalle de confiance correspondant. Bon, mais qu'est-ce que cela m'apprend sur VOTRE échantillon ? A priori, rien ! Sauf que vous pouvez vous dire ceci : « Puisque pour 95% des échantillons, la moyenne de la population est dans l'intervalle de confiance correspondant, la probabilité que ce soit le cas pour MON échantillon est de 0.95. Dans ce sens, je suis confiant à 95% que la moyenne de la population se trouve bien dans mon intervalle de confiance. »

Reste à comprendre si tout ceci est bien vrai ! Pour cela, il faut des preuves et des hypothèses précises. Ceci étant un dictionnaire et non pas un encyclopédie, je ne les donnerai pas. Mais sachez que la justification des affirmations qui précèdent utilise intensément le Théorème Central Limite et qu'elle nécessite des hypothèses qu'il convient de vérifier avant d'utiliser les conclusions dans les applications. Une précaution pas toujours respectée en pratique. ...

Loi exponentielle

On dit d'une variable aléatoire X qu'elle suit une loi exponentielle de paramètre $\lambda > 0$ si X admet une densité $\rho(x)$ donnée par $\rho(x) = \lambda \exp(-\lambda x)$. On écrit alors $X \sim \mathcal{E}(\lambda)$.

Dans quel(les) situations cette loi apparaît-elle ?

Comme on le verra, la loi exponentielle est, dans un sens qui deviendra clair, la version continue de la *loi géométrique*. Supposons qu'on étudie les appels téléphoniques arrivant dans un centre d'appels. On sait qu'il y a, en moyenne, λ appels par unité de temps. Par exemple, 240 appels par heure : $\lambda = 240$ donc dans ce cas. On peut, alternativement, s'intéresser au nombre de photons qui sortent par unité de temps d'un laser. Ou au nombre de noyaux d'une substance radio-active qui se désintègrent par unité de temps. J'appellerai "événement" l'arrivée d'un appel téléphonique ou d'un photon, ou encore la désintégration d'un noyau atomique, selon la situation étudiée. Dans tous ces cas, il est raisonnable de postuler que les événements sont indépendants. Ce n'est pas évident, mais en examinant chaque situation, on peut conclure que ceci est une hypothèse raisonnable. Qui plus est, elle permet, comme on le verra, de progresser très simplement dans l'analyse du phénomène étudié. Il conviendra, bien sûr, de tester la théorie construite sur la base de cette hypothèse à la réalité. Si l'accord n'est pas bon, il faudra conclure que l'hypothèse était trop simpliste et affiner la théorie. Appelons T_1 la variable aléatoire "instant auquel a lieu le premier événement." T_1 prend des valeurs dans $]0, +\infty[$. On veut déterminer $P(t \leq T_1 < t + \Delta t)$, pour Δt petit. Posons, à cet effet $t = n\Delta t$. On prendra plus loin n grand et Δt petit, tout en gardant le produit des deux constant et égal à t . C'est ce qui permet de "discrétiser" le problème, c'est à dire de revenir à un problème de probabilités discrètes. Sachant que le nombre moyen d'événements par unité de temps est λ , il est raisonnable de postuler que la probabilité qu'un événement ait lieu dans un intervalle de petite taille Δt vaut $\lambda\Delta t$. Pour que cela ait du sens, il faut clairement que $\lambda\Delta t < 1$, ce qu'on supposera toujours. On partage alors l'intervalle $]0, t]$ en n intervalles de taille Δt . On peut donc introduire la variable S_k qui vaut 1 lorsqu'il y a un événement dans $[(k-1)\Delta t, k\Delta t]$ et 0 sinon. On a donc $P(S_k = 1) = \lambda\Delta t$. Les variables S_k sont tout simplement des variables de Bernoulli avec une probabilité de succès donnée par $\lambda\Delta t$. Il en suit que

$$P(t \leq T_1 \leq t + \Delta t) = P(S_1 = 0 = S_2 \cdots = S_n, S_{n+1} = 1).$$

On a dit qu'on considèrerait les événements dans les différents intervalles $[(k-1)\Delta t, k\Delta t]$ comme indépendants. Mais cela signifie que nous avons réduit la

question à la suivante : quelle est la probabilité d'avoir $k - 1$ échecs, suivi d'un succès, dans une suite de k épreuves de Bernouilli indépendantes ? Et nous connaissons la réponse : la loi du premier succès est donnée par une *loi géométrique*. Donc

$$P(S_1 = 0 = S_2 \cdots = S_n, S_{n+1} = 1) = (1 - \lambda \Delta t)^n \lambda \Delta t.$$

Or, en utilisant $n \Delta t = t$ on trouve que

$$\lim_{n \rightarrow +\infty} (1 - \lambda \Delta t)^n = \lim_{n \rightarrow +\infty} \left(1 - \frac{\lambda t}{n}\right)^n = \exp(-\lambda t)$$

On en conclut que, lorsque Δt est petit (et donc n grand)

$$P(t \leq T_1 \leq t + \Delta t) \simeq \lambda \exp(-\lambda t) \Delta t.$$

On en conclut que T_1 est une variable aléatoire à densité $\rho(t)$ avec $\rho(t) = \lambda \exp(-\lambda t)$. On voit ici comment la loi exponentielle “émerge” d’une suite de *lois géométriques* de paramètre $p = \lambda t/n$, lorsque $n \rightarrow +\infty$.

Loi forte des grands nombres

On considère une famille de variables aléatoires *indépendantes* $X_i, i \in \mathbb{N}$, ayant tous la même loi d'espérance $\mu \in \mathbb{R}$. Donc notamment, pour tout $i \in \mathbb{N}$, $\mathbb{E}(X_i) = \mu$. On suppose également que toutes ces variables aléatoires sont définies sur le même univers Ω , avec probabilité P . On s'intéresse alors à

$$S_N = \sum_{i=1}^N X_i. \quad (5)$$

Un exemple simple mais pertinent d'une telle situation est celui-ci. On jette de façon répétée un dé. Alors $\Omega = \llbracket 1, 6 \rrbracket^{\mathbb{N}^*}$. Une issue de cette expérience est de la forme $\omega = (\omega_1, \omega_2, \omega_3, \dots, \omega_i, \dots)$, où chaque ω_i appartient à $\llbracket 1, 6 \rrbracket$. Par exemple, $\omega = (1, 3, 2, 4, 5, 1, 4, \dots)$ correspond à des lancer donnant respectivement 1, puis 3, puis 2, et ainsi de suite. Imaginons que, à chaque lancer, on fait autant de pas en avant que le nombre apparaissant sur le dé. On se trouvera alors, après N lancer, à

$$S_N = \omega_1 + \omega_2 + \dots + \omega_N$$

pas de son point de départ. En posant $X_i(\omega) = \omega_i$, on voit que les X_i sont indépendants (puisque les lancer sont supposés l'être) et nous sommes donc bien dans la situation de (5). Clairement, S_N est également une variable aléatoire définie sur Ω , en tant que somme de variables aléatoires. Dans l'exemple, si on jette à nouveau N fois un dé, les résultats seront différents et donc la position finale S_N sera différente. Dans le panneau de gauche de la Fig. 6 on voit ainsi la trajectoire $N \rightarrow S_N$ pour onze expériences aléatoires différentes et pour N allant de $N = 1$ à $N = 20$. Il faut s'imaginer qu'onze personnes différentes ont chacune lancé successivement 20 dés. On voit sur la figure que leurs positions finales s'échelonnent entre 61 et 81.² La moyenne de ces onze positions finales est à peu près égale à 70. Ceci est attendu. En effet, à chaque lancer, on avance entre 1 et 6 pas. Le nombre moyen de pas avec lequel on avance est donc 3.5. Par conséquent, en moyenne, après 20 lancer, on avance de $20 \times 3.5 = 70$ pas. En termes mathématiques, on a $\mathbb{E}(X_i) = \mu = 3.5$ et, par linéarité de l'espérance,

$$\mathbb{E}(S_N) = \mu N.$$

On observe que les valeurs de S_N fluctuent autour de cette valeur moyenne μN . La taille et la nature de ces fluctuations est étudié dans le *Théorème*

2. Un code en R qui produit le panneau gauche de la figure est donné ci-dessous. Il vous permet de varier les paramètres et de mieux comprendre ce qui se passe.

Centrale Limite. Pour l'instant, contentons-nous d'étudier plutôt S_N/N . On peut considérer cette quantité, dans le cas de l'exemple avec les dés, comme la vitesse moyenne de la personne. En effet, c'est la distance parcourue S_N en N unités de temps (les temps entre les lancers), divisés par ce temps. Le graphe de cette quantité est faite pour les 11 graphes du panneau gauche de la figure dans son panneau de droite. On observe que, pour toutes les trajectoires, S_N/N semble s'approcher de $\mu = 3.5$, lorsque N augmente. À $N = 20$, elle sont toutes entre 3 et 4. Au fait, si on trace des trajectoires pour de plus grandes valeurs de N , on observe que ce phénomène s'accroît. Dans la Fig. 7, N va jusqu'à 1000 et on voit que S_N/N devient, pour toutes les trajectoires tracées, petite. La formulation précise de cette observation est donnée par la loi forte des grands nombres :

Theorem. *Soit une famille de variables aléatoires indépendantes $X_i, i \in \mathbb{N}$, définies sur un univers Ω et ayant toutes la même loi d'espérance $\mu \in \mathbb{R}$. Alors, la probabilité que*

$$\lim_{N \rightarrow +\infty} \frac{S_N}{N} = \mu$$

vaut 1.

ajouter un mot sur la probabilité 1. donnons un exemple où N n'est pas le temps. taille de l'échantillon par exemple. nombre de mesures...

Le code en R pour la Fig. 6. Panneau de gauche.

```
Nmax<-20;#le nombre maximal de lancer de des
K<-10# le nombre de realisations; ou le nombre de trajectoires creees
N=1:Nmax;
y=floor(runif(Nmax, 1,7)); #Nmax lancer de de
SN=cumsum(y);
plot(N, SN, type='l', ylim=c(0,4.5*Nmax)) #une premiere trajectoire tracee
for (i in 1:K)
{y=floor(runif(N, 1,7)); #ceci donne Nmax lancer de des
SN=cumsum(y)
points(N,SN, type="l") #'points' ajoute les K trajectoires a la premiere
}
```

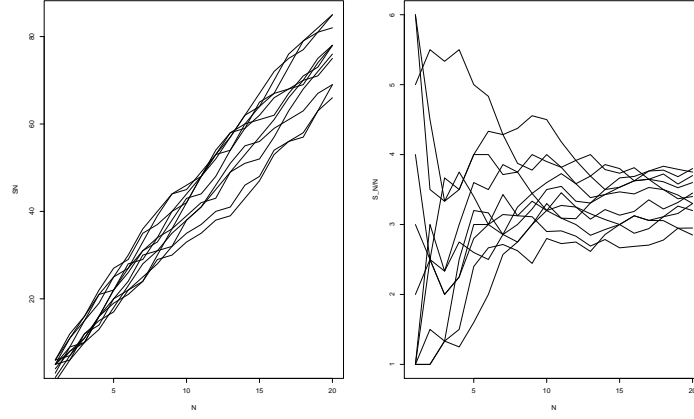


FIGURE 6 – À gauche. Onze trajectoires, S_N en fonction de N , pour N allant de 1 à 20. À chaque valeur de N , la trajectoire avance de 1, 2, 3, 4, 5, 6 pas, selon le résultat d'un lancer de dé. À droite. Pour quatre de ces trajectoires, le graphe de S_N/N en fonction de N .

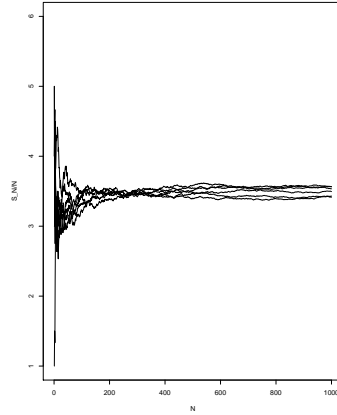


FIGURE 7 – Pour huit trajectoires, le graphe de S_N/N en fonction de N , avec N allant jusqu'à $N = 1000$. On remarque que l'écart entre S_N/N et 3.5 est petit pour toutes les trajectoires, lorsque N devient grand.

Loi géométrique

La loi géométrique est une loi de probabilité sur \mathbb{N}_* , donnée par

$$p_k = (1 - q)^{k-1} q,$$

où $0 \leq q < 1$. Elle apparaît naturellement lorsqu'on pose la question : “Quel est la probabilité que le 3 apparaîtra pour la première fois au k -ème lancer, lorsqu'on lance successivement un dé ?” En posant $q = \frac{1}{6}$, la réponse est $p_k = (\frac{5}{6})^{k-1} \frac{1}{6}$. En effet, on a $k-1$ fois un chiffre autre que 3, puis finalement 3 au k -ème lancer. L'espérance et la variance de la loi géométrique se calculent facilement.

Loi normale

On dit d'une *variable aléatoire* X qu'elle suit une loi normale d'*espérance* μ et de *variance* σ si elle admet une *densité de probabilité* $\rho(x)$ donnée par

$$\rho(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right).$$

On écrit alors $X \sim \mathcal{N}(\mu, \sigma)$. Nous avons tracé ces densités pour différentes valeurs de μ et de σ dans le premier TP. Vous pouvez apprendre quelques propriétés importantes de la loi normale en regardant ici <https://www.youtube.com/watch?v=E4EdnaP-iCg>. La raison principale de l'importance des lois normales se trouve dans le *Théorème Central Limite*.

Lire aussi : *Théorème Central Limite*.

Loi de Pareto

Une loi de Pareto est une loi de probabilités à densité, déterminée par deux paramètres $a, k > 1$, de la façon suivante :

$$\begin{aligned}\rho_k(x) &= (k-1)a^{k-1}x^{-k}, \quad x \geq a \\ &= 0.\end{aligned}$$

On vérifie facilement que $\int_0^{+\infty} \rho_k(x) = 1$, de façon que ρ_k est bien une *densité de probabilité*. On remarque que, plus k est petit, moins $\rho_k(x)$ décroît vite lorsque $x \rightarrow +\infty$. Plus précisément, la *probabilité de survie* d'une Loi de Pareto est, pour $x \geq a$

$$P(X \geq x) = \left(\frac{a}{x}\right)^{k-1}.$$

Pour $x > a$ fixé, ceci est d'autant plus petit que k est grand. Mais c'est toujours plus grand que pour une loi exponentielle ou normale. On parle de "queues de la distribution lourdes".

Si maintenant X est une variable aléatoire ayant comme loi une Loi de Pareto, on peut se demander si X a une espérance et/ou un écart type.

À quoi servent les Loïs de Pareto ?

Loi de probabilité

Le terme “loi de probabilité” est essentiellement synonyme de “*probabilité*” mais souvent réservé au cas où la loi est sur l’axe des réels. Donc la *loi exponentielle*, la *loi normale*, les *lois de Pareto*, les *lois normales*, sont des *lois de probabilités*.

Loi d'une variable aléatoire

Considérons le cas d'une *variable aléatoire* X discrète prenant des valeurs x_k . Introduisons $p_k = P(X = x_k)$, “la probabilité que X vaut x_k ”. On remarquera que

$$\sum_k p_k = 1.$$

On peut par conséquent définir une probabilité P_X sur \mathbb{R} en posant, pour tout $A \subset \mathbb{R}$,

$$P_X(A) = \sum_{x_k \in A} p_k = P(X \in A).$$

On appelle P_X la loi de X ; c'est la probabilité induite par X sur \mathbb{R} . Remarquons qu'elle est complètement déterminée par les p_k et on dit par conséquent aussi souvent que les p_k sont la loi de X .

Lorsque X n'est pas une variable discrète, on définit la loi de X directement par

$$P_X(A) = P(X \in A). \quad (6)$$

Un cas particulièrement utile/fréquent/important est celui où la loi de X , P_X est donnée par *une densité de probabilité*, que l'on note alors ρ_X :

$$P_X(A) = \int_A \rho_X(x) dx.$$

On dirait alors des choses comme “ X admet une loi à densité” ou encore “ X est une variable à densité” ou toute autre variante raisonnable. On écrit aussi ρ plutôt que ρ_X quand il ne peut pas y avoir confusion.

Dans de nombreux cas, on ne s'intéresse pas à Ω et P , mais à une ou plusieurs variables aléatoires et à leur loi. Par exemple, considérons le cas d'une marche aléatoire dans laquelle un mobile se déplace sur une droite en faisant, à des instants régulièrement espacés, un déplacement soit vers la droite avec probabilité p , soit vers la gauche avec une probabilité $1 - p$. Supposons que le point fasse N déplacements. On peut considérer alors que Ω est l'ensemble de toutes les suites de N déplacements. Un élément de Ω est donc de la forme

$$\omega = (\omega_1, \dots, \omega_N),$$

par exemple

$$\omega = (1, 1, -1, 1, -1, -1, -1, 1, 1, 1, 1, -1, 1, \dots 1).$$

Ce qui nous intéresse dans ce cas est de savoir la position $X(\omega)$ du point après N déplacements. En supposant que le point démarre à l'origine, on a

$$X(\omega) = \sum_{k=1}^N \omega_k.$$

La loi de X , si on sait la calculer, nous renseigne par exemple sur la probabilité que le point soit retourné à l'origine après N déplacements. Ou sur la probabilité qu'il se trouve à une distance au moins D de son point de départ. La loi de X nous renseigne donc sur la distribution des positions du point après N déplacements, mais pas sur le chemin suivi pour arriver à ces positions, qui est codé dans Ω . Donc, quand on ne s'intéresse uniquement à la position finale du mobile et pas au chemin qu'il a suivi pour y arriver, on se contentera d'étudier X et sa loi.

Marche aléatoire

Une marche aléatoire est un *processus stochastique* S_t , $t \in \mathbb{N}$, définie comme une somme de variables aléatoires *indépendantes* et identiquement distribuées X_t :

$$S_t = S_0 + \sum_{s=1}^t X_s.$$

La marche aléatoire est donc une *somme de variables aléatoires* indépendantes, ce qui simplifie grandement son étude. En effet, en utilisant les propriétés de telles sommes, on voit déjà tout de suite que

$$\mathbb{E}(S_t) = \mathbb{E}(S_0) + t\mathbb{E}(X_1), \quad \mathbb{V}(S_t - S_0) = t\mathbb{V}(X_1).$$

Lire aussi : *Somme de variables aléatoires, Théorème Central Limite.*

Médiane (d'une variable aléatoire)

Soit une variable aléatoire X , on appelle médiane de X tout réel m ayant la propriété

$$P(X \leq m) = \frac{1}{2} = P(X > m).$$

Pour les variables X_λ de la Fig. 1 on trouve facilement que la médiane se trouve entre $\sqrt{2}/\lambda$ et $\mathbb{E}(X_\lambda)$. Détailler. La distribution est biaisée. Bcp de masse vers les grandes valeurs.

Médiane empirique (d'un échantillon)

Étant donné un *échantillon* de taille N , sa médiane empirique, souvent désigné par m , est définie de la façon suivante. La médiane m est un nombre tel qu'il y a autant de valeurs de l'échantillon plus grand que plus petit que m . On peut définir de la même façon la médiane de n'importe quelle suite de nombres finie. Par exemple, la médiane de 1, 2, 3, 4, 5, 6, 7 est $m = 4$. La médiane n'est pas définie de façon unique si le nombre N des éléments de la suite est paire. En effet, une médiane de 1, 2, 3, 4, 5, 6 est 3.5. Mais le nombre 3.8 satisfait aussi à la définition. En général, si $N = 2M$, et si les éléments de l'échantillon sont notés x_1, \dots, x_N , on choisit $m = \frac{1}{2}(x_M + x_{M+1})$. Si vous utilisez un logiciel pour déterminer la médiane d'une suite de nombres, il utilisera cette formule.

Mode (d'un échantillon), Le

Le mode d'un échantillon est la valeur ayant la plus grande fréquence absolue. Un échantillon peut être bi-modal ou même multi-modal : le mode n'est pas forcément unique. Par exemple, le mode de

158 161 161 175 172 158 193 179 175 165 163 161

158 175 161 185 173 161 176 157 175 183

est 161. Cette valeur intervient 5 fois. Mais la suite

158 161 161 175 172 175 193 179 175 165 163 161

158 175 161 158 175 161 176 157 181 183

a deux modes 161 et 175. Le logiciel R n'a pas de commande "mode". Pour de petits échantillons vous pouvez utiliser `sort()`, qui range les valeurs dans l'ordre, et puis lire le mode sur l'échantillon ainsi réordonné.

Un échantillon peut être bimodal parce qu'il est obtenu à partir de deux populations distinctes. Si vous considérez par exemple 100 hommes de 25 ans, dont 50 font du basket dans un club, et 50 sont spéléologues, vous risquez de trouver deux modes parce que vous aurez deux populations de tailles très différentes.

Voir aussi : *classe modale d'un histogramme.*

Moyenne empirique (d'un échantillon)

La moyenne empirique d'un échantillon x_1, x_2, \dots, x_N est

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i.$$

Moyenne pondérée

Je donnerai la définition générale ci-dessous, mais posons-nous d'abord la question suivante. Supposons que nous avons accès à un *histogramme* d'un *échantillon*, mais pas à l'échantillon lui-même. Pouvons-nous, en utilisant les informations obtenues à partir de l'histogramme, déterminer la *moyenne empirique* \bar{x} de l'échantillon ? La réponse est : "Non, pas exactement, mais tout de même approximativement." Je m'explique. On a donc un histogramme, dont on connaît les classes I_k et les effectifs N_k , et donc aussi l'effectif total $N = \sum_{k=1}^K N_k$. On ne connaît donc pas les x_i , mais on peut s'imaginer qu'ils soient ordonnés :

$$x_1 < x_2 < \dots < x_N.$$

En particulier, on a x_1, x_2, x_{N_1} qui appartiennent tous à I_1 . Et puis $x_{N_1+1}, \dots, x_{N_1+N_2}$ qui appartiennent à I_2 et ainsi de suite. On ne connaît pas les valeurs exactes de x_1, \dots, x_{N_1} , mais si on désigne par c_1 le centre de la première classe I_1 , on peut faire l'approximation qui consiste à dire que $x_1 \simeq c_1, x_2 \simeq c_1, \dots, x_{N_1} \simeq c_1$. Elle sera d'autant meilleure que la classe I_1 est petite. On a alors

$$x_1 + \dots + x_{N_1} \simeq N_1 c_1.$$

C'est à dire : la somme des valeurs x_i dans la classe I_1 est approximativement égale au produit du centre c_1 de la classe avec l'effectif N_1 de la classe. On peut faire le même raisonnement pour chaque classe. Pour la deuxième, cela donne par exemple

$$x_{N_1+1} + \dots + x_{N_1+N_2} \simeq N_2 c_2,$$

où c_2 est le centre de la deuxième classe I_2 . On a, en fin de compte,

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \simeq \frac{1}{N} \sum_{k=1}^K N_k c_k \simeq \sum_{k=1}^K f_k c_k.$$

Il faut remarquer que la dernière somme ne contient plus que K termes, et nous avons vu qu'en règle générale, K est beaucoup plus petit que N : la dernière somme est donc plus simple à calculer que la première ! Et on peut lire les valeurs des *fréquences relatives* $f_k = \frac{N_k}{N}$ directement sur l'histogramme, donc on dispose de tous les éléments pour calculer cette somme.

Pour résumer, je viens donc de démontrer que la moyenne empirique \bar{x} d'un échantillon peut être approchée par

$$\bar{x} \simeq \sum_{k=1}^K f_k c_k, \quad \text{où} \quad \sum_{k=1}^K f_k = 1.$$

Le membre de droite dans la première équation ci-dessus est ce qu'on appelle une "moyenne pondérée" des c_k , $k = 1, \dots, K$, avec "poids" f_k . On "pondère" chacune des valeurs c_1, c_2, \dots, c_K par un "poids" f_k et on fait la somme. Si tous les f_k sont égaux, on a forcément $f_k = \frac{1}{K}$, puisque leur somme vaut 1. Dans ce cas précis,

$$\frac{1}{K} \sum_{k=1}^K c_k;$$

la moyenne pondérée est alors simplement la moyenne empirique des c_1, c_2, \dots, c_K . Cela correspond au cas où l'histogramme serait parfaitement "plat" puisqu'alors $N_k = \frac{N}{K}$: les effectifs de toutes les classes sont identiques.

Quand on fait des statistiques, on interprète les fréquences relatives f_k comme des probabilités : si l'échantillon est représentatif (ce qui implique que N doit être relativement grand), on interprète f_k comme la probabilité qu'un individu tiré au hasard dans la population ait une valeur x_i dans la classe I_k , et donc proche de c_k .

On peut de la même façon estimer *l'écart type empirique* d'un échantillon à partir de la seule connaissance de l'histogramme de fréquences correspondant. On calcule pour cela de façon approchée la *variance empirique* s^2 de l'échantillon en calculant la moyenne pondérée suivante :

$$s^2 \simeq \frac{1}{N} \sum_{k=1}^K N_k (c_k - \bar{x})^2 = \sum_{k=1}^K f_k (c_k - \bar{x})^2.$$

La moyenne pondérée intervient en d'autres circonstances. À un certain nombre de concours ou de diplômes, différentes matières ont différents coefficients. Disons qu'un candidat ou étudiant doit passer cinq épreuves et qu'il reçoit une note sur 20 pour chaque épreuve : r_1, r_2, r_3, r_4, r_5 . Mais les épreuves peuvent être considérées plus ou moins importantes : un plus grand coefficient pour les mathématiques par exemple, et pour le français, que pour l'histoire-géo. Supposons que ces coefficients soient t_1, t_2, t_3, t_4, t_5 . On calcule alors une moyenne pondérée sur 20 de la façon suivante :

$$\sum_{k=1}^5 \frac{t_i}{T} r_i, \quad \text{où} \quad T = \sum_{i=1}^5 t_i.$$

On remarquera que ici aussi, la somme des poids $p_i = \frac{t_i}{T}$ vaut 1.

En toute généralité, lorsqu'on dispose d'une famille de nombres c_1, \dots, c_L et de poids $0 \leq p_1, \dots, p_L \leq 1$ tels que $\sum_{\ell=1}^L p_\ell = 1$, alors la moyenne

pondérée des c_ℓ est donnée par

$$\sum_{\ell=1}^L p_\ell c_\ell.$$

Voir aussi : *espérance*.

Ordinateur

Les ordinateurs nous fournissent des outils de calcul rapides. En utilisant par exemple le logiciel R, on peut très rapidement produire des *histogrammes*, calculer des *moyennes*, *médianes*, *écarts types*, *simuler des lois*, faire des regressions linéaires, calculer de *intervalles de confiance*, et bien d'autres choses encore. Il est important, à chaque fois que vous utilisez un ordinateur pour effectuer un calcul, de vous poser la question : "Saurais-je faire cette opération moi-même, avec un stylo et un bout de papier, fût-ce beaucoup plus lentement ?" Si la réponse est "non", cela signifie que vous ne comprenez pas ce que vous faites. Ce qui n'est pas une bonne chose. Si la réponse est "oui", une deuxième question se pose : "Suis-je capable de correctement interpréter le résultat ?" C'est encore plus important et nettement plus difficile. Beaucoup d'études statistiques sont faites par des gens qui manipulent des logiciels, et appliquent des règles aveuglement, sans jamais faire une analyse critique des résultats qu'ils obtiennent. Ne les rejoignez pas.

Population

Une population est un ensemble d'individus dont on souhaite étudier certaines caractéristiques. Les individus ne sont pas forcément des humains : ils peuvent être des animaux ou des objets, des pays, de régions, des institutions, etc. Les populations d'intérêt sont normalement très grandes, et les caractéristiques intéressantes difficilement accessibles. Par exemple, un fabricant de bureaux voudra connaître la qualité du mécanisme d'ouverture et de fermeture de leurs tiroirs. Dans ce cas, la population de bureaux inclut tous les bureaux qu'il construira dans les mois et années à venir, et qu'il ne pourra donc pas tester aujourd'hui, puisqu'ils n'existent pas encore. Il ne pourra en tester qu'une petite partie – un *échantillon* – dans le but de prédire la qualité de la production entière. Si vous voulez connaître la consommation d'alcool hebdomadaire des étudiants de l'Université de Lille, vous n'avez aucun moyen d'obtenir cette information pour chaque étudiant : elle n'existe nullepart et elle est d'ailleurs de nature privée. Mais vous pouvez proposer un questionnaire à un certain nombre d'étudiants volontaires et ainsi constituer un *échantillon*.

Voir aussi : inférence statistique.

Probabilité, Une

Une probabilité sur un ensemble Ω est une fonction P qui associe à chaque sous-ensemble A de Ω un nombre $P(A)$, appelé “probabilité de A ” et qui a les propriétés suivantes :

$$0 \leq P(A) \leq 1, \quad P(\Omega) = 1,$$

et

$$P(\cup_n A_n) = \sum_n P(A_n) \quad \text{si} \quad A_n \cap A_m = \emptyset, \forall n \neq m.$$

Cette définition d'apparence simple a été proposée par le mathématicien Russe A. Kolmogorov (1903-1987) en 1933. Comme on le verra, elle est plus sophistiquée qu'il n'y paraît et capte très bien la notion intuitive de “probabilité d'un événement. L'ensemble Ω doit être pensée ici comme contenant tous les issus possibles d'une expérience aléatoire. La condition $P(\Omega) = 1$ exprime le fait qu'on n'a pas oublié des issus, c'est à dire que l'une d'elles se réalisera avec certitude. Si, par exemple, vous modéliser un jeu de fléchettes, vous pouvez prendre comme Ω le disque formé par la cible. Mais cela signifie que vous considérez que les joueurs sont assez bons pour ne jamais rater complètement la cible ! De même, lorsque vous lancez un dé, vous considérez que les issus possibles sont 1, 2, 3, 4, 5, 6 et vous excluez donc la possibilité que le dé finit sa course coincé contre un bord du jeu dans une position oblique.

Les hypothèses impliquent facilement que, lorsque $A \subset B$, $P(A) \leq P(B)$. Ceci exprime l'idée que, comme B contient plus d'issus possibles que A , sa probabilité est plus grande.

Je signale brièvement une subtilité mathématique, pour aussitôt l'oublier. Lorsque l'ensemble Ω n'est pas dénombrable, on ne peut pas, dans la plupart des cas, définir P sur *tous* les sous-ensembles de Ω , mais seulement sur une très large sous-famille. Lorsque $\Omega = \mathbb{R}$, par exemple, on la définira sur tous les intervalles et sur tous les unions et intersections dénombrables d'intervalles. Mais il existe des ensembles qui ne sont pas de cette forme. En général, quelque soit Ω , les probabilités que vous rencontrerez pourront toujours être définies sur tous les sous-ensembles que vous rencontrerez, donc vous pourrez sans danger ignorer cette difficulté.

Comment construire une probabilité ? Deux cas se présentent souvent.

Lorsque Ω est fini ou dénombrable, on peut l'identifier avec la suite fini ou infini $(\omega_n)_n$ de ses éléments. On a alors

$$P(A) = \sum_{n; \omega_n \in A} p_n$$

où

$$p_n = P(\{\omega_n\})$$

est la probabilité de l'événement élémentaire $\{\omega_n\}$. Une probabilité sur Ω est donc complètement déterminée par la donnée d'une suite finie ou infinie de $0 \leq p_n$ satisfaisant $\sum_n p_n = 1$.

Lorsque $\Omega = \mathbb{R}$, on peut construire des probabilités en utilisant une *densité de probabilité*.

Probabilités, Les

Dans la vie courante, les mots “probable” ou “probabilité” sont utilisés pour décrire une situation incertaine ou imprévisible, dans laquelle la chance ou la malchance jouent un rôle important. “Il est peu probable que le PSG gagne la Ligue des Champions.” “Il est plus probable que ce soit lui que moi qui vient demain.” “La probabilité d’une issue favorable à ce conflit n’est pas bien élevée.” La plupart du temps, nous utilisons les probabilités pour décrire des situations incertaines qui doivent se produire dans l’avenir, mais pas toujours. “Je parie qu’il n’y soit pas allé.” “Il n’est pas probable qu’il t’ait compris.” Dans tous les cas, il y a une incertitude sur les conditions entourant la situation décrite, ou sur le fonctionnement du système observé, ou sur les deux. Prédire les cours en bourse est difficile parce qu’on ne connaît pas les lois qui gèrent leur évolution et nous ne sommes même pas sûr que de telles lois existent. Sans parler du fait que ces cours sont influencés par des acteurs extérieurs que non seulement nous ne maîtrisons pas, mais dont nous ne comprenons pas tout à fait l’influence sur les cours. On dit alors que le hasard joue un rôle pour indiquer que plusieurs issues sont possibles, sans qu’on sache laquelle se produira, ni pourquoi.

Pour décrire le mouvement des planètes autour du soleil, par contre, nous ne faisons pas appel aux probabilités, parce que, depuis Newton, nous connaissons bien les lois qui règlent leur mouvement, ainsi que les positions et vitesses initiales des planètes. Quand nous jetons deux dés sur une table, nous connaissons aussi les lois, mais nous maîtrisons mal la condition initiale, c’est à dire la vitesse de déplacement et de rotation que nous donnons à chacun des dés au moment de les lancer. L’issue est donc incertaine ou, de façon plus sophistiquée, *aléatoire*. On dit que c’est le hasard qui décide. Mais on peut se poser la question de savoir quelles issues sont plus probables que d’autres. Quand on jette deux dés, la somme peut valoir 2,3,4,..., 11,12. Est-il plus probable que la somme vaut 7 ou 8 plutôt que 2 ou 12 ? Et si oui, de combien ?

La théorie mathématique des probabilités propose une quantification de l’incertitude inhérente à toutes ces situations. Elle trouve ses origines dans le désir de quantifier les divers degrés de probabilité des issues possibles dans les jeux de cartes notamment. Bien évidemment, si elle ne s’occupait que des gains et des pertes au casino, elle n’aurait pas suscité l’intérêt qu’on lui connaît et on ne l’enseignerait pas aux lycées et dans toutes les formations universitaires, y inclus dans les sciences humaines et sociales. Le fait est

que les probabilités, ainsi que *la statistique*, qui y est étroitement liée, sont présentes dans toutes les sciences, en médecine et pharmacologie, et dans la gestion publique comme dans les entreprises. Elles sont aujourd’hui au coeur de l’intelligence artificielle. Elles ont pénétré dans la vie quotidienne au travers des lotteries, des paris sportifs et des sondages, qui tentent de comparer la probabilité relative de diverses issues électorales mais aussi par le biais de la météo, qui nous propose des probabilités de précipitation notamment.

En conclusion, les probabilités sont une branche des mathématiques appliquées. Elles utilisent ou développent des outils de l’analyse mathématique afin d’analyser des problèmes issus des sciences physiques ainsi que des sciences humaines et sociales ou des questions de société dans lesquels le hasard joue un rôle. La modélisation mathématique de ces problèmes est une partie intégrante des probabilités.

Les probabilités, sont-elles un sujet intéressant, qui devrait intriguer, voir passionner, tout jeune étudiant universitaire ? Sont-elles utiles ? Les réponses à ces questions sont forcément subjectives puisque les notions d’intérêt et d’utilité le sont. Toujours est-il que les probabilités et les statistiques ne servent pas qu’à satisfaire notre curiosité sur l’issue d’une élection ou notre plaisir à parier sur l’avenir. Et encore moins à compter des boules rouges et blanches, avec ou sans remise, dans des urnes. Elles jouent depuis longtemps un rôle important dans les assurances, depuis quelques décennies dans le monde de la finance, et depuis quelques années dans le développement des “data science”, l’intelligence artificielle et l’apprentissage machine. Elles contribuent donc à façonner notre monde et on peut espérer que cela suffit pour qu’elles suscitent l’intérêt de tout jeune étudiant universitaire, notamment en informatique.

Pour etymologie du mot “probable”, qui vient bien sûr du latin, consulter par exemple <https://www.dictionnaire-academie.fr/article/A9P4372>.

Processus stochastique

Un processus stochastique est une famille de variables aléatoires X_t , indexées par t , tous définies sur un même espace probabilisé Ω . Ici t peut être soit un entier, soit un réel. Dans les applications, il représente souvent le temps, d'où la notation. Les processus stochastiques servent à modéliser différents phénomènes qui manifestent un comportement temporel erratique : le mouvement d'une feuille ballotée par le vent, d'une molécule dans un gaz, du cours d'une action en bourse, Le processus stochastique le plus simple est sans doute la *marche aléatoire*. On appelle trajectoire ou réalisation du processus toute fonction $t \rightarrow X_t(\omega)$, pour un $\omega \in \Omega$ donné. Comprendre un processus stochastique signifie tout d'abord comprendre l'évolution dans le temps de la loi de X_t en fonction de t . Mais ce n'est pas tout. On souhaite aussi comprendre les corrélations qui existent entre X_t et $X_{t'}$ pour deux valeurs de t différentes. Ce n'est pas simple en général. On doit se contenter parfois de comprendre seulement l'espérance et la variance des X_t , plutôt que leur loi plus généralement. Et l'étude des corrélations est souvent très complexe. Une famille importante où ces questions peuvent être bien comprises sont les processus markoviens.

Simuler une loi de probabilité

Supposons qu'on s'intéresse à une *variable aléatoire réelle* X de loi \mathcal{L} . Considérons le cas où X est une variable discrète. On désigne par x_k les valeurs que prend X et on écrit $P(X = x_k) = p_k$. On suppose qu'un connaît la loi de X , c'est à dire qu'on connaît les p_k .

Simuler X , ou simuler \mathcal{L} , signifie réaliser un grand nombre N de fois une expérience aléatoire dont les issues possibles sont les valeurs x_k de X . On désigne par N_k le nombre de fois que la valeur x_k se réalise dans les N expériences. On appelle aussi N_k la fréquence de x_k . On dit qu'on a simulé X ou la loi de X si les fréquences relatives d'apparition $f_k = \frac{N_k}{N}$ des différents x_k sont proches de $p_k = P(X = x_k)$. Il en résulte que si on réalise l'histogramme de fréquences relatives des N issues des expériences, on obtient un graphe qui est très proche du graphe de p_k en fonction de x_k .

On peut faire ce type d'expérience "à la main", mais aujourd'hui, pour pouvoir prendre N grand, on fait appel à un ordinateur. En utilisant R, une telle suite d'expériences peut être réalisée facilement, le résultat étant une suite de N valeurs x_k . On peut bien évidemment prendre N très grand.

Si par exemple, X est une variable aléatoire dont la loi est la loi binomiale de paramètres $n = 30, p = 0.33$, on peut procéder de la façon suivante. On se rappelle que le nombre de succès dans un schéma de Bernoulli de $n = 30$ essais et avec une probabilité de réussite $p = 0.3$, suit une loi binomiale. On lance 30 fois un dé et on considère qu'un succès correspond à l'apparition des chiffres 1 ou 2, tandis que sinon on enregistre un échec. On enregistre à la fin de cette expérience le nombre de succès, qui sera un entier entre 0 et 30. Puis on recommence la même expérience un grand nombre N de fois. On obtient ainsi une liste de N entiers entre 0 et $n = 30$, qui représente le nombre de succès dans le i -ème schéma réalisé. Si $N = 20$, on aura par exemple :

12 5 14 15 6 8 10 10 12 10 6 6 7 8 9 9 10 10 14 5

On introduit alors N_k , le nombre de fois qu'il y a eu k succès. Ici, $N_1 = 0, N_5 = 2, N_9 = 2, N_{10} = 5, N_{14} = 1$ et ainsi de suite. Et on introduit aussi les fréquences relatives correspondantes : $f_k = \frac{N_k}{N}$.

L'*histogramme* des f_k est donné dans la Fig. 8 ci-dessous pour deux valeurs de N , $N = 50$ et $N = 5000$ ³. Les N expériences aléatoires sont réalisées à l'aide du logiciel R. On fait confiance aux "help" du logiciel qui dit que, lorsque vous exécutez la commande `rbinom(size=30, n=50, p=0.3)`,

3. On remarquera qu'ici $x_k = k$ et que l'histogramme est construit avec des boîtes de taille 1.

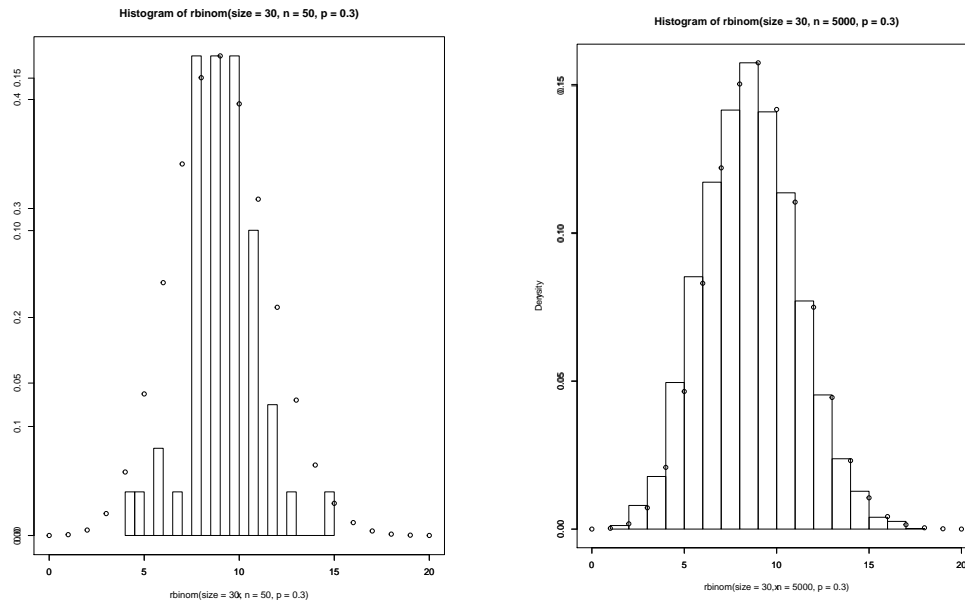


FIGURE 8 – Histogramme d’une simulation d’un schéma de Bernoulli et loi binomiale.

l’ordinateur produit un vecteur de 50 entiers entre 0 et 30, qui sont obtenus en exécutant un schéma de Bernoulli de taille 30 et avec une probabilité de succès $p = 0.3$ à chaque essai. On remarque que quand on ne fait que 50 expériences aléatoires, l’histogramme n’est pas très proche du graphe des p_k , tandis que lorsque $N = 5000$, l’accord est excellent. On en conclut qu’on a bien simulé la loi binomiale.

Les graphes sont produits avec les commandes suivantes :

```
hist(rbinom(size=30,n=5000, p=0.3), freq=FALSE, xlim=c(0,20), breaks=20)
x<-0:20
y<-dbinom(x, size=30, p=0.3)
#computes binomial distribution for elements of vector x
par(new=TRUE)#keeps the graph
plot(x,y)#plot of binomial distribution
```

Somme de variables aléatoires (indépendantes) : espérance, variance, loi

ATTENTION. Cette entrée sera encore étoffée. 16 Mars.

Soient deux variables aléatoires X_1 et X_2 , définies sur le même univers Ω . On peut considérer leur somme $S_2 = X_1 + X_2$. Est-il facile d'exprimer, de calculer, de déterminer la loi de S_2 quand on connaît les lois de X_1 et de X_2 ? La question est légitime et la réponse "NON". Pour le comprendre, on va établir qu'il est facile de trouver l'espérance $\mathbb{E}(S_2)$ et déjà moins facile de déterminer la variance $\mathbb{V}(S_2)$ de S_2 . Seulement lorsque X_1 et X_2 sont indépendantes devient-il relativement simple de calculer la loi de la somme en termes de lois des deux termes X_1 et X_2 .

On a, avec des notations évidentes,

$$\mathbb{E}(S_2) = \mu_1 + \mu_2,$$

donc l'espérance d'une somme est la somme des espérances. Toujours, sans autres conditions. Par contre

$$\mathbb{V}(S_2) = \sigma_1^2 + \sigma_2^2 + 2\mathbb{E}((X_1 - \mu_1)(X_2 - \mu_2)).$$

Le dernier terme ne peut en général pas être déterminé en termes des lois de X_1, X_2 . Une exception notable est celle de deux *variables aléatoires indépendantes*, au quel cas on a

$$\mathbb{E}((X_1 - \mu_1)(X_2 - \mu_2)) = \mathbb{E}((X_1 - \mu_1)\mathbb{E}(X_2 - \mu_2)) = 0.$$

Donc, pour des variables aléatoires indépendantes (et cette condition est importante!) on trouve que la variance de la somme est la somme des variances :

$$\mathbb{V}(S_2) = \sigma_1^2 + \sigma_2^2.$$

Dans le cas de N variables X_1, \dots, X_N , on a, avec $S_N = \sum_{n=1}^N X_n$

$$\mathbb{E}(S_N) = \sum_{n=1}^N \mu_n$$

et, lorsque les X_n sont indépendantes,

$$\mathbb{E}(S_N) = \sum_{n=1}^N \sigma_n^2$$

Ces quelques observations, très simples, sont essentielles si l'on veut comprendre la Loi Forte des Grands Nombres et le Théorème Central Limite. Dans ces cas, on suppose non seulement que les variables X_n sont indépendantes mais aussi qu'elles ont toutes la même loi et donc la même espérance μ et variance σ . Il en résulte que, sous ces hypothèses, on a

$$\mathbb{E}(S_N) = N\mu, \quad \mathbb{V}(S_N) = N\sigma^2, \quad \text{et donc} \quad \sigma_N = \sqrt{N}\sigma.$$

C'est de ce simple calcul qu'émerge la racine carrée de N qu'on retrouve dans le *Théorème Central Limite*.

Statistique, La

Un exemple frappant de l'utilisation des probabilités et de la statistique sont les assurances. Combien y aura-t-il des accidents de voiture l'année prochaine? D'incendies d'immeubles? De tempêtes violentes, d'inondations? On ne le sait pas précisément, mais une analyse de la probabilité de ces événements permet d'en évaluer le coût attendu et donc de fixer les primes d'assurance. Dans nos sociétés modernes, les assurances maladie, de responsabilité civile, de nos voitures et logement sont un élément essentiel de notre niveau de vie. Donc l'utilité des probabilités et de la statistique ne laisse pas de tout de ce point de vue déjà.

Depuis les années 80, les techniques probabilistes ont pénétré le monde de la finance. Associées à l'informatique, elles y jouent aujourd'hui un rôle très important, partiellement néfaste. De nombreux livres, films et documentaires existent qui expliquent en termes simples les mécanismes mises en oeuvre. Pour une première lecture, vous pouvez consulter http://math.univ-lille1.fr/~debievre/MathFi_EC.html

Statistique descriptive

A VENIR

Statistique inférentielle

A VENIR

Théorème Central Limite, Le

Ce résultat donne une information cruciale sur la loi de la somme $S_N = \sum_{n=1}^N X_n$ de N variables aléatoires réelles X_1, \dots, X_N indépendantes et toutes de même loi. Ce type de somme est important dans une variété d'applications.

Par exemple, supposons que vous jouiez un jeu d'argent et que vous puissiez gagner ou perdre une certaine quantité d'argent à chaque fois que vous jouez. Appelons X_i votre gain ou votre perte au i -ème jeu. Alors votre gain ou perte après N jeux est S_N . Il y a un intérêt évident à connaître la loi de S_N , c'est à dire la probabilité que vous gagnerez ou perdrez telle ou telle somme d'argent. Un autre exemple est celui de la *marche aléatoire* : ici les X_i sont les déplacements d'un mobile le long d'un axe et donc S_N sa position finale, si $X_0 = 0$. Un troisième exemple est celui de mesures expérimentales d'une quantité donnée. Par exemple, si vous mesurez un grand nombre N de fois la hauteur de votre table de travail, vous obtiendrez des valeurs différentes X_i à chaque coup. On considère alors qu'une bonne estimation de la vraie hauteur de la table est donnée par $\frac{S_N}{N}$: c'est la moyenne de toute les mesures prises. Dans ces trois exemples, à chaque fois qu'on fait N jeux, ou N pas ou N mesures, on obtient une valeur différente de S_N . En d'autres termes, pour chaque N , S_N est une variable aléatoire. Le Théorème Central Limite nous informe sur la loi de S_N lorsque N est grand. Ou, en termes plus simples, elle nous informe sur la distribution des valeurs de S_N lorsque N est grand.

Le Théorème Central Limite dit que, lorsque les variables aléatoires X_i sont indépendantes et ont une espérance et une variance respectivement données par μ et σ , alors la somme $S_N = \sum_{i=1}^N X_i$ a une loi proche d'une *loi normale* d'espérance μN et de variance $\sigma\sqrt{N}$. On peut écrire

$$S_N \approx \mathcal{N}(\mu N, \sigma\sqrt{N}) \quad (7)$$

ou alternativement, si l'on s'intéresse plutôt à S_N/N ,

$$\frac{S_N}{N} \approx \mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{N}}\right). \quad (8)$$

Ce qui est étonnant ici, c'est l'universalité du résultat : quelque soit la loi des X_i (binomiale, géométrique, Pareto, exponentielle, ou n'importe quelle autre loi), la loi de $S_N = \sum_{i=1}^N X_i$ s'approche *toujours* d'une *loi normale*. Laquelle précisément ? Eh bien, les *lois normales* étant caractérisées par leur espérance et leur variance, bien évidemment, il s'agit de LA *loi normale* dont l'espérance et la variance sont celles de S_N . Donc la *loi normale* d'espérance μN et d'écart type $\sigma\sqrt{N}$.

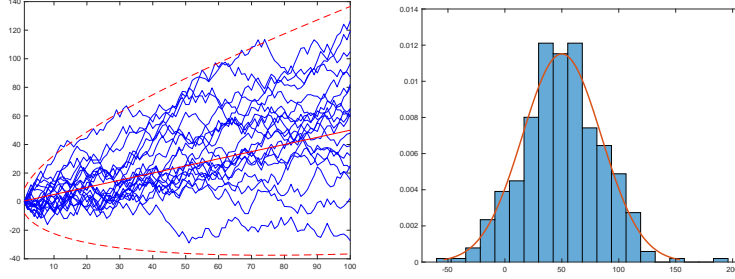


FIGURE 9 – À gauche. 20 trajectoires de la marche aléatoire décrite dans le texte. À droite l’histogramme des positions S_N à $N = 100$ pour 200 telles trajectoires. Ainsi que la courbe de la loi normale d’espérance 50 et d’écart type $\sqrt{1210} \approx 35$.

Une formulation un peu plus précise de (7) est la suivante. Pour $a' < b'$,

$$P(a' \leq S_N \leq b') \approx \frac{1}{\sqrt{2\pi\sigma^2N}} \int_{a'}^{b'} \exp\left(-\frac{(s - \mu N)^2}{2\sigma^2N}\right) ds. \quad (9)$$

On reconnaît dans le membre de droite de cette équation la densité d’une *loi normale* d’espérance μN et de variance $\sigma\sqrt{N}$.

Une reformulation de (7) qui est instructive est celle-ci. Pour tout $a < b$, on a

$$P(\mu N + a\sigma\sqrt{N} \leq S_N \leq \mu N + b\sigma\sqrt{N}) \xrightarrow{N \rightarrow +\infty} \frac{1}{\sqrt{2\pi}} \int_a^b \exp\left(-\frac{t^2}{2}\right) dt. \quad (10)$$

Cela signifie que, pour N grand, la probabilité que la somme S_N se trouve dans la fenêtre $[\mu N + a\sigma\sqrt{N}, \mu N + b\sigma\sqrt{N}]$ est proche de la probabilité qu’une variable normale standard soit dans $[a, b]$.

La Fig. 9 donne une illustration de ce résultat. Ici, les X_i sont uniformément distribuées dans l’intervalle $[-5.5, 6.5]$. On peut penser à X_i comme étant le i -ème pas d’une *marche aléatoire*. On a $-5.5 \leq X_i \leq 6.5$ et le pas moyen vaut donc $\mathbb{E}(X_i) = 0.5$. Un petit calcul montre que la variance du pas vaut $\mathbb{V}(X_i) = (2\sqrt{3})^2$. Les lignes bleues du panneau gauche de la figure représentent chacune une trajectoire $N \rightarrow S_N$, pour N allant de 1 à 100. Elles sont produites avec un code MatLab, voir ci-dessous. On s’imagine tirer au hasard pour chaque i allant de 1 à N un pas X_i entre -5.5 et 6.5 et ainsi progresser : $S_{i+1} = S_i + X_{i+1}$. Après N pas, la position moyenne des marcheurs est approximativement $\mathbb{E}(S_N) = N/2$. Cette moyenne est

indiquée par la droite rouge sur la figure dont on observe qu'elle a bien la pente 1/2. Les deux courbes rouges en pointillés correspondent aux fonctions $N \rightarrow 0.5N - (2.5)2\sqrt{3}\sqrt{N}$ et $N \rightarrow 0.5N + (2.5)2\sqrt{3}\sqrt{N}$. On constate que, pour N grand, les 20 courbes bleues se trouvent bien entre ses deux bornes. Ceci est conforme au résultat (10). En effet, on a

$$\int_{-a}^a \exp(-\frac{t^2}{2}) dt = 2\Phi(a) - 1,$$

où

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp(-\frac{t^2}{2}) dt$$

est la fonction de répartition de la loi normale, dont la table se trouve sur la fiche de TD 3, dernière page. En posant $a = 2.5$, on trouve $\Phi(2.5) = 0.994$ donc

$$\int_{-2.5}^{2.5} \exp(-\frac{t^2}{2}) dt = 2\Phi(2.5) - 1 = 0.988.$$

Finalement, (10) devient

$$P(0.5N - 2.5(2\sqrt{3})\sqrt{N} \leq S_N \leq 0.5N + 2.5(2\sqrt{3})\sigma\sqrt{N}) \xrightarrow{N \rightarrow +\infty} 0.988.$$

Donc, en gros, la probabilité qu'une des marches aléatoires sorte de cette fenêtre lorsque N est suffisamment grand est petite : $1 - 0.998 = 0.012$. Une chance sur cent donc, à peu près. Donc, si on ne trace que 20 trajectoires, on s'attend à ce qu'elles soient toutes dans cette fenêtre. On remarque que ceci n'est pas le cas lorsque N est plus petit que 80 : quelques trajectoires sont alors en dehors de cette fenêtre. Mais lorsque N devient plus grand, elles sont toutes dans la fenêtre.

Pour bien mettre en évidence la convergence de la loi de S_N vers une loi normale, on peut regarder dans le panneau droit de la Fig. 9 l'histogramme de 200 valeurs de S_N , pour $N = 100$. En rouge, on y voit superposée la densité de la loi normale d'espérance $0.5N = 50$ et d'écart type $2\sqrt{3}\sqrt{N}$. On voit que l'accord est déjà très bien, même si on n'a utilisé que 200 trajectoires pour simuler la loi de S_N et que $N = 100$ n'est pas très grand.

De façon analogue, si l'on s'intéresse à S_N/N plutôt que à S_N lui même, la formulation précise de (8) est :

$$P(\mu + a\frac{\sigma}{\sqrt{N}} \leq \frac{1}{N}S_N \leq \mu + b\frac{\sigma}{\sqrt{N}}) \xrightarrow{N \rightarrow +\infty} \frac{1}{\sqrt{2\pi}} \int_a^b \exp(-\frac{t^2}{2}) dt. \quad (11)$$

En prenant par exemple $a = 3 = -b$, on conclut cette fois-ci que les valeurs de S_N/N se trouvent dans un intervalle de diamètre $6\frac{\sigma}{\sqrt{N}}$ centré sur son

espérance μ avec une probabilité égale à $2\Phi(3) - 1 = 0.997$, qui est très proche de 1. Cet intervalle devient de plus en plus petit au fur et à mesure que N croît grâce au facteur $N^{-1/2}$. Cette information complète donc celle donnée par la *loi forte des grands nombres*. Celle-ci dit que, avec probabilité égale à 1, on a

$$\lim_{N \rightarrow +\infty} \frac{S_N}{N} = \mu.$$

Le Théorème Central Limite donne donc une information sur la variation des valeurs de S_N/N autour de μ : elle sont concentrées dans un intervalle autour de μ dont la taille décroît comme $N^{-1/2}$. C'est une des leçons centrales à retenir du Théorème Central Limite.

L'énoncé du Théorème Central Limite est relativement simple. Son interprétation, ses diverses formulations et ses applications sont multiples et riches. Comprendre le Théorème Central Limite prend du temps.

Code matlab de la Fig. 9 Ce qui suit est le code MatLab qui a permis de faire le Fig. 9. On peut le transcrire facilement en R, je vous laisse faire. Il faut remarquer qu'il sera commode d'utiliser la commande `runif(Lmax, amin, amax)` pour simuler les X_i . On a $\text{amin} = a - d$, $\text{amax} = a + d$. Cela fait une boucle en moins dans le code R que dans le code MatLab. Une fois que votre code fonctionne, vous pouvez jouer avec les paramètres pour illustrer différentes applications du Théorème Central Limite : varier Lmax et N, notamment. Mais aussi changer la loi uniforme pour les X_i en une autre loi, et observer les changements.

```
%sum of iid random variables
clear all;
close all;
Nmax=5000;%temps final. en nombre de pas de la marche
d=6;% taille typique du pas
a=0.5;%taille de la derive. valeur moyenne du pas.
Figure1=figure(1);
Lmax=1000;%nombre de realisations du processus.
Finalposition=zeros(1, Lmax);
var=0; %initialiser l'ecart quadratique
for L=1:1:Lmax
    S=zeros(1, Nmax);%vector with positions after N steps, for realization L
    S(1,1)=0;%everybody starts at 0
    sum=0;
    for N=2:1:Nmax
```

```

        sum=sum+d*2*(0.5-rand)+a;
        S(1,N)=sum; %position after N steps for realization L
    end %end of N loop.
    Finalposition(1,L)=S(1,Nmax);
    plot(1:Nmax, S, 'b') %plot of the trajectory of Lth realization
        hold all;
        var=var+(S(1,Nmax)-Nmax*a)^2;% variance de S_Nmax (times Lmax!).
    end %end of L loop
    sqrt(var/Lmax)/(d*sqrt(1/3)*sqrt(Nmax))%ecart quadratique de la position finale calculé
    %divise par l'ecart quad theorique
    x=1:1:Nmax;
    xlim([1 Nmax]);
    plot(x, a*x+ 2.5*d*sqrt(1/3)*sqrt(x), '--r', x,a*x -3*d*sqrt(1/3)*sqrt(x), '--r')
    plot(x, a*x,'-r')
    Figure2=figure(2);
    hist(Finalposition, 20)

```

Théorème de Tchebychev, Le

Ce résultat s'applique à n'importe quelle suite de données numériques et donc à n'importe quel échantillon. Sans formules, il dit la chose suivante :

La fraction des données d'un échantillon quelconque qui se trouvent à moins que k écarts type de la moyenne empirique \bar{x} vaut au moins $1 - k^{-2}$.

Ici k est un nombre positif plus grand que 1. Par exemple, en prenant $k = 2$, on constate qu'au moins $3/4$ des données se trouvent à moins que deux écarts type de la moyenne empirique.

Vous trouverez une application ici : <https://www.youtube.com/watch?v=uMgK000XFhA>. Regarder cela avant d'étudier la preuve.

La démonstration de ce résultat est astucieuse mais très simple. La voici. Soient x_1, x_2, \dots, x_N un échantillon, \bar{x} sa moyenne empirique et s son écart type. Considérons l'ensemble d'indices

$$E_k = \{1 \leq i \leq N \mid |x_i - \bar{x}| \geq ks\}.$$

L'ensemble E_k est l'ensemble des indices i pour lesquels la donnée correspondante x_i se situe à *plus* que k écarts type de \bar{x} . On a

$$\#E_k = \sum_{i \in E_k} 1.$$

C'est à dire que le nombre d'éléments de E_k vaut $1 + 1 + \dots + 1$, où il y a autant de 1 dans la somme que d'éléments dans E_k . Mais remarquons maintenant que, pour chaque $i \in E_k$, on a $1 \leq \frac{|x_i - \bar{x}|}{ks}$ et donc $1 \leq \frac{|x_i - \bar{x}|^2}{k^2 s^2}$: cela suit directement de la définition de E_k . On peut donc écrire

$$\#E_k = \sum_{i \in E_k} 1 \leq \sum_{i \in E_k} \frac{|x_i - \bar{x}|^2}{k^2 s^2} = \frac{1}{k^2 s^2} \sum_{i \in E_k} |x_i - \bar{x}|^2.$$

On y est presque, maintenant. En effet, on a bien évidemment que

$$\sum_{i \in E_k} |x_i - \bar{x}|^2 \leq \sum_{i=1}^N |x_i - \bar{x}|^2 = N s^2,$$

puisque pour obtenir la seconde somme on a rajouté des termes à la première. On peut alors conclure que

$$\#E_k = \sum_{i \in E_k} 1 \leq N \frac{1}{k^2},$$

et donc que $\frac{\#E_k}{N^2} \leq \frac{1}{k^2}$: la fraction des données qui se situe à *plus* que k écarts type de la moyenne empirique est donc *inférieure* à k^{-2} . Cela implique bien évidemment que la fraction des données qui se situe à *moins* que k écarts type de la moyenne empirique est *supérieure* à $1 - k^{-2}$.

Univers

On appelle “univers” l’ensemble des issues possibles d’une expérience aléatoire.

On le désigne le plus souvent par Ω et ses éléments par ω .

Voir aussi : *évènement*

Variable aléatoire

Une variable aléatoire (réelle) sur un ensemble Ω muni d’une probabilité P est une fonction $X : \Omega \rightarrow \mathbb{R}$. On désigne par $X(\Omega)$ les valeurs prises par X . On dit que la variable est discrète lorsque ces valeurs forment un ensemble fini ou dénombrable. On écrit alors $X(\Omega) = \{x_1, x_2, \dots, x_k, \dots\}$. Si elle n’est pas discrète, elle est continue.

Voir aussi : loi d’une variable aléatoire.

Variables aléatoires indépendantes

Voir *indépendance de variables aléatoires*.

Variance d’une variable aléatoire

Voir sous *écart type* d’une variable aléatoire.

Variance empirique d’un échantillon

La variance empirique d’un échantillon $x_1 \dots x_N$ est le nombre positif s^2 donné par

$$s^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2,$$

où \bar{x} est la *moyenne empirique de l’échantillon*. Sa racine carrée s est l’*écart type empirique*. La variance et l’écart type sont des caractéristiques de dispersion de l’échantillon, ce qui signifie tout simplement qu’elles mesurent la distribution des valeurs de l’échantillon autour de leur moyenne. Si les données sont très étalées, l’écart type est grand. Si elles sont très resserrées autour de la moyenne empirique, l’écart type est petit. À noter que s a la même dimension que les valeurs x_i : si ces derniers sont en euros, s est exprimé en euros, s’ils sont en mètres, s est en mètres etc.