Lycée Pierre-Gilles de Gennes **BCPST2**
Mathématiques **2019-2020**

Notes de cours 02

Bases des probabilités

Table des matières

1 Modèles probabilistes généraux- la théorie

L'existant

Nous avons revu les concepts de base du calcul des probabilités de première année, qui concerne les calculs sur des modèles « finis ».

On a notamment revu la définition d'un espace probabilisé $(\Omega, \mathbb{P})$ où Ω est un ensemble fini, dont la famille des parties $\mathscr{P}(\Omega)$ est appelée la famille des événements et $\mathbb{P} : \mathscr{P}(\Omega) \to [0,1] \subset \mathbb{R}$ est une application qui à chaque événement A associe sa probabilité $\mathbb{P}(A)$.

Le plus important pour la modélisation, c'est l'introduction et la manipulation des variables aléatoires. Le cadre de première année rattache la théorie des probabilités aux mathématiques « standard ». Il permet de ne traiter que le cas où un nombre fini de variables prenant un nombre fini de valeurs interviennent.

J'espère avoir démontré dans la partie précédente que l'on peut progressivement se désintéresser de l'hypothèse de finitude et mener beaucoup de calculs sans se préoccuper de la forme exacte de l'ensemble Ω.

1.1 L'axiomatique de KOLMOGOROV

Disclaimer

On donne maintenant les définitions générales d'espace probabilisé, de tribu (d'événements), de probabilité qui vont permettre de traiter le cas de variables prenant possiblement une infinité de valeurs et le cas d'un nombre infini de variables.

C'est une théorie très abstraite, ces définitions sont au programme mais, comme on le verra assez rapidement, on va peu s'en préoccuper. Un modèle probabiliste est construit par la donnée des variables intéressantes, de leurs distributions et de leurs relations et permet de calculer probabilité d'événements, espérance de variable, etc...Il s'agit fondamentalement, pour faire des calculs fondés mathématiquement, de savoir qu'il existe *au moins* un ensemble Ω, une tribu d'événements $\mathscr{T}$ et une probabilité $\mathbb{P}$ sur cette tribu « supportant » le modèle, *i.e.* les variables aléatoires voulues. Les résultats attendus doivent être indépendants du choix de Ω, $\mathscr{T}$, $\mathbb{P}$.

Cette question d'existence est logiquement extrêmement délicate et en conséquence...

Pour chacun des modèles probabilistes que nous rencontrerons, nous admettrons qu'il existe un $(\Omega, \mathscr{T}, \mathbb{P})$ abstrait permettant de faire des calculs sensés.

Au niveau du théorème 10, nous donnons quelques explications autour de ce fait.

A savoir donc, mais sans en faire une fixation. LAPLACE, BERNOULLI,.. faisaient au $XVIII^e$ et XIX^e du calcul des probabilités sans cet arsenal logique [1] qui date des années 1930.

Tribu d'événements

Définition 1. *Soit Ω un ensemble non vide. $\mathscr{T}$ un sous-ensemble [2] de $\mathscr{P}(\Omega)$. On dit que $\mathscr{T}$ est une* tribu *(d'événements) si*

1. $\emptyset \in \mathscr{T}$, $\Omega \in \mathscr{T}$, *respectivement l'*événement impossible *et l'*événement certain
2. *$\mathscr{T}$ est stable par passage au complémentaire.*

$$\text{Si } A \in \mathscr{T}, \text{ alors } \overline{A} \in \mathscr{T}$$

3. *$\mathscr{T}$ est stable par $\cup$ dénombrable [3] : Si $(A_n)_{n\in\mathbb{N}}$ est une famille d'éléments de $\mathscr{T}$ alors $\cup_{n\in\mathbb{N}} A_n \in \mathscr{T}$ où*

$$\cup_{n\in\mathbb{N}} A_n = \{\omega \in \Omega, \exists n \in \mathbb{N}, \omega \in A_n\}$$

$\mathscr{T}$ est stable par $\cap$ dénombrable [4] : Si $(A_n)_{n\in\mathbb{N}}$ est une famille d'éléments de $\mathscr{T}$ alors $\cap_{n\in\mathbb{N}} A_n \in \mathscr{T}$ où

$$\cap_{n\in\mathbb{N}} A_n = \{\omega \in \Omega, \forall n \in \mathbb{N}, \omega \in A_n\}$$

Avec le vocabulaire des événements, ces axiomes se lisent comme suit

1. Un élément $\omega \in \Omega$ est appelé un élément de hasard (en anglais 'random element'). On trouve aussi la locution 'configuration élémentaire'. J.P KAHANE a proposé de l'appeler un *randon*.
2. Un événement, c'est ce qui peut avoir lieu ou pas quand un ω est choisi. C'est un ensemble de randons qui ont une propriété logique commune. Il a lieu si le ω choisi fait partie de l'événement, il n'a pas lieu sinon.
3. L'événement impossible, c'est ce qui est toujours faux, n'a jamais lieu. L'événement certain c'est ce qui est toujours vrai, a toujours lieu.
4. Si A est un événement, l'événement contraire $\overline{A}$ est aussi un événement.

1. Cet arsenal n'est pas inutile, il a permis de lever et/ou résoudre certains paradoxes que les pratiques antérieures ont pu mettre à jour
2. Les éléments de $\mathscr{T}$ sont donc des parties de Ω.
3. union dénombrable
4. intersection dénombrable

5. Si A_1, A_2 sont deux événements, $A_1 \cup A_2 = (A_1 \text{ ou } A_2)$ est un événement. C'est l'ensemble des randons ω faisant partie de l'événement A_1 ou de l'événement A_2. C'est l'événement « alternative de A_1 et A_2 ».
6. Si A_1, A_2 sont deux événements, $A_1 \cap A_2 = (A_1 \text{ et } A_2)$ est un événement. C'est l'ensemble des randons ω faisant partie de l'événement A_1 et de l'événement A_2. C'est l'événement « conjonction de A_1 et A_2 ».
7. Les deux derniers axiomes montrent que l'on peut faire une conjonction ou une alternative d'une infinité *dénombrable* d'événements et obtenir encore un événement.
8. Un événement A implique un événement B si $A \subset B$. Cela signifie que si A a lieu lorsque ω est choisi, alors B a lieu.
9. Deux événements A et B sont dits « équivalents » si $A = B$. Cela signifie que lorsque ω est choisi A a lieu si et seulement si B a lieu.
10. Deux événements sont dits « incompatibles »si leur conjonction est impossible, *i.e.* si $A \cap B = \emptyset$. Cela signifie que si A a lieu lorsque ω est choisi, alors B n'a pas lieu et réciproquement.

Probabilité sur une tribu

Définition 2. *Soit $\mathscr{T}$ une tribu sur un ensemble Ω. Une application $\mathbb{P} : \mathscr{T} \to [0,1] \subset \mathbb{R}$ est appelée une* (mesure de) probabilité. *Si*

1. $\mathbb{P}(\emptyset) = 0$ *et* $\mathbb{P}(\Omega) = 1$.
2. *Si $A_1, \dots, A_n, \dots$ est une famille (au plus dénombrable) d'événements incompatibles alors*

$$\begin{aligned} \mathbb{P}(\cup_{n\in\mathbb{N}} A_n) &= \mathbb{P}(A_1 \cup \dots \cup A_n \cup \dots) \\ &= \mathbb{P}(A_1) + \dots + \mathbb{P}(A_n) + \dots \\ &= \lim_{N\to+\infty} \sum_{n=1}^{N} \mathbb{P}(A_n) \end{aligned}$$

Si Ω est un ensemble quelconque, $\mathscr{P}(\Omega)$ est une tribu.

1. C'est celle que l'on considère le plus souvent lorsque Ω est fini. On peut alors définir une probabilité sur cette tribu à l'aide d'un système de poids.
2. Lorsque Ω est infini, il y a potentiellement une difficulté relevant de la logique pour définir une mesure de probabilité intéressante sur $\mathscr{P}(\Omega)$. Cela explique pourquoi on se restreint à une tribu d'événements plus peite que $\mathscr{P}(\Omega)$.

En conséquences directes des axiomes

Proposition 3. *1. Si $A \subset B$ alors $\mathbb{P}(A) \le \mathbb{P}(B)$*

2. (Formule de POINCARÉ*)*

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$$

En langage presque courant, on peut lire

1. La probabilité d'un événement impossible est nulle. La probabilité d'un événement certain est 1.
2. Si A_1 et A_2 sont incompatibles alors la probabilité de l'alternative de A_1 et A_2 est la somme des probabilités.
3. Si un événement A implique un événement B, la probabilité de A est inférieure à celle de B.

Exemples graphiques : Carré (HP)

Considérons le carré $\Omega = [0,1[\times [0,1[$ subdivisé en petits carrés $C_{i,j} = \left[\frac{i}{4}, \frac{i+1}{4}\right[\times \left[\frac{j}{4}, \frac{j+1}{4}\right[$, $i, j \in \{0,\dots,3\}$, de côté $\frac{1}{4}$ comme sur la figure 1

L'ensemble $\mathscr{C} = \{C_{i,j}, i, j \in \{0,\dots,3\}\}$ est un ensemble de parties de Ω. Il comporte 16 éléments. Ce **n'est pas** une tribu de parties de Ω.

Par contre, l'ensemble $\mathscr{T}_0$ où chaque élément est une figure T formée d'une union finie de carrés $C_{i,j}$ est une tribu de parties de Ω. Il comporte 2^{16} éléments. Une probabilité $\mathbb{P}_0$ sur cette tribu est donnée par l'application définie par

$$\forall T \in \mathscr{T}_0, \mathbb{P}_0(T) = \frac{\text{aire}(T)}{\text{aire}(\Omega)} = \text{aire}(T)$$

Sur la figure 1, on a

$$\mathbb{P}_0(T_1) = \frac{4}{16}, \mathbb{P}_0(T_2) = \frac{3}{16}, \mathbb{P}_0(T_1 \cap T_2) = \frac{1}{16} \text{ et } \mathbb{P}_0(\overline{(T_1 \cup T_2)}) = \frac{10}{16}$$

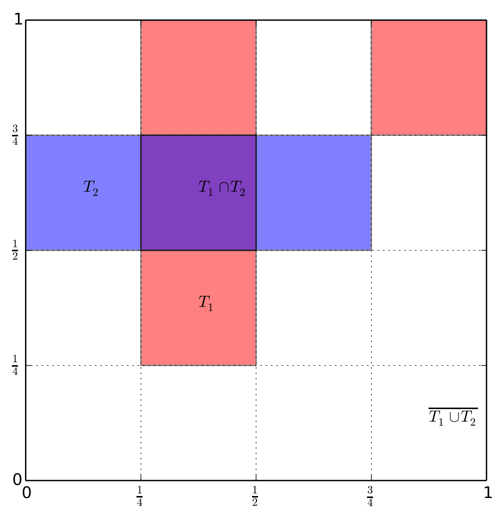

FIGURE 1 – Le carré, T_1 en rouge, T_2 en bleu, des éléments de la tribu $\mathscr{T}_0$

Sur ce même Ω, on peut définir de façon un peu vague, une tribu $\mathscr{T}$ beaucoup plus large, à savoir, l'ensemble des parties du carré admettant une aire [5] et définir sur cette tribu $\mathscr{T}$ la probabilité $\mathbb{P}$ définie par

$$\forall T \in \mathscr{T}, \mathbb{P}(T) = \frac{\text{aire}(T)}{\text{aire}(\Omega)} = \text{aire}(T)$$

Spécifier la probabilité d'un événement A, c'est imposer une mesure de la proportion de A relativement à Ω.

5. C'est beaucoup plus délicat que ça n'en a l'air !

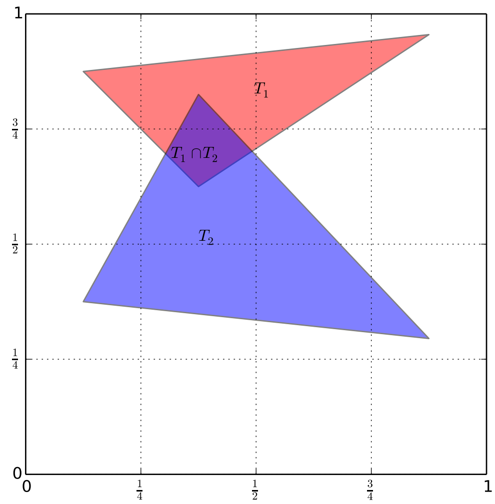

FIGURE 2 – Le carré Ω, quelques éléments de la tribu $\mathscr{T}$

Exemples graphiques : Segment (HP)

Considérons maintenant le segment $\Omega = [0,1[$ subdivisé en petits segments $S_k = \left[\frac{k}{16}, \frac{k+1}{16}\right[, k \in \{0,\dots,15\}$, de longueur $\frac{1}{16}$ comme sur la figure 3. On a épaissi les segments en hauteur pour pouvoir voir les couleurs.

L'ensemble $\mathscr{S} = \{S_k, k \in \{0,\dots,15\}\}$ est un ensemble de parties de Ω. Il comporte 16 éléments. Ce **n'est pas** une tribu de parties de Ω.

Par contre, l'ensemble $\mathscr{T}_0$ où chaque élément est une figure T formée d'une union finie de segments S_k est une tribu de parties de Ω. Il comporte 2^{16} éléments. Une probabilité $\mathbb{P}_0$ sur cette tribu est donnée par l'application définie par

$$\forall T \in \mathscr{T}_0, \mathbb{P}_0(T) = \frac{\text{longueur}(T)}{\text{longueur}(\Omega)} = \text{longueur}(T)$$

Sur la figure 3, on a

$$\mathbb{P}_0(T_1) = \frac{4}{16}, \mathbb{P}_0(T_2) = \frac{3}{16}, \mathbb{P}_0(T_1 \cap T_2) = \frac{1}{16} \text{ et } \mathbb{P}_0(\overline{(T_1 \cup T_2)}) = \frac{10}{16}$$

Comme pour le carré, on peut aussi considérer la tribu $\mathscr{T}$ de toutes les parties T du segment unité admettant une longueur. La longueur longueur(T) d'une telle partie peut-être interprétée comme la probabilité $\mathbb{P}(T)$ de l'événement T.

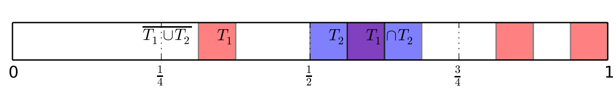

FIGURE 3 – Le segment, T_1 en rouge, T_2 en bleu, des éléments de la tribu $\mathscr{T}_0$

Pouvez vous deviner la transformation qui fait passer du carré au segment et réciproquement ?

Variables aléatoires réelles

Définition 4. *Soit un espace probabilisé* $(\Omega, \mathscr{T}, \mathbb{P})$.

Une variable aléatoire réelle, X, est une application $X : \Omega \to \mathbb{R}$ *telle que pour tout intervalle* $I \in \mathbb{R}$

$$\underbrace{\{X \in I\}}_{\text{notation probabiliste}} := \{\omega \in \Omega, X(\omega) \in I\} = \underbrace{X^{-1}(I)}_{\text{notation ensembliste}}$$

est un événement, i.e. est dans $\mathscr{T}$.

Un exemple graphique (HP)

Reprenons le cas où Ω est le carré $[0,1[\times[0,1[$, $\mathbb{P}$ l'aire et prenons comme v.a., à valeurs dans $[0,1[$, X_1 et X_2 définies par

$$\forall\omega = (\omega_1, \omega_2) \in [0,1[\times[0,1[, X_1(\omega) = \omega_1, X_2(\omega) = \omega_2$$

Sur la figure 4, on a représenté

$$T_1 = \left\{X_2 \in \left[\frac{1}{2}, \frac{3}{4}\right[\right\} \text{ et } T_2 = \left\{X_1 \in \left[\frac{1}{4}, \frac{1}{2}\right[\right\}$$

On a

$$\mathbb{P}(\frac{1}{2} \leq X_1 < \frac{1}{4}) = \frac{1}{4}, \mathbb{P}(\frac{1}{4} \leq X_2 < \frac{1}{2}) = \frac{1}{4},$$
$$\mathbb{P}(\frac{1}{2} \leq X_1 < \frac{1}{4}) = \frac{1}{4} \text{ et } \frac{1}{4} \leq X_2 < \frac{1}{2}) = \frac{1}{16}$$

Une interprétation graphique plus imaginative montre que, si I_1, I_2 sont deux intervalles quelconques de $\mathbb{R}$, alors

- $\{X_1 \in I_1\}$ est un rectangle « vertical »de largeur $\mathbb{P}(X_1 \in I_1)$, de hauteur 1
- $\{X_2 \in I_2\}$ est un rectangle « horizontal »de largeur 1, de hauteur $\mathbb{P}(X_2 \in I_2)$,
- $\{X_1 \in I_1\} \cap \{X_2 \in I_2\}$ est un rectangle de largeur $\mathbb{P}(X_1 \in I_1)$, de hauteur $\mathbb{P}(X_2 \in I_2)$,
- On a donc $\mathbb{P}(X_1 \in I_1 \text{ et } X_2 \in I_2) = \mathbb{P}(X_1 \in I_1).\mathbb{P}(X_2 \in I_2)$.
- Ceci étant vrai pour tout couple d'intervalle (I_1, I_2), les v.a X_1 et X_2 sont indépendantes ?

Posons maintenant les v.a. D_1 et D_2 définies par

$$D_1 = \begin{cases} 0 & \text{si } 0 \leq X_1 < \frac{1}{4} \\ 1 & \text{si } \frac{1}{4} \leq X_1 < \frac{1}{2} \\ 2 & \text{si } \frac{1}{2} \leq X_1 < \frac{3}{4} \\ 3 & \text{si } \frac{3}{4} \leq X_1 < 1 \end{cases} \text{ et } D_2 = \begin{cases} 0 & \text{si } 0 \leq X_2 < \frac{1}{4} \\ 1 & \text{si } \frac{1}{4} \leq X_2 < \frac{1}{2} \\ 2 & \text{si } \frac{1}{2} \leq X_2 < \frac{3}{4} \\ 3 & \text{si } \frac{3}{4} \leq X_2 < 1 \end{cases}$$

Pour résumer, pour $(i,j) \in \{0,\ldots,3\} \times \{0,\ldots,3\}$, $\omega \in [0,1[^2$, $(D_1(\omega), D_2(\omega)) = (i,j)$ si et seulement si $\omega \in C_{i,j}$. D_1 et D_2 sont des variables prenant un nombre fini de valeurs, uniformément distribuées sur $\{0,\ldots,3\}$ et indépendantes.

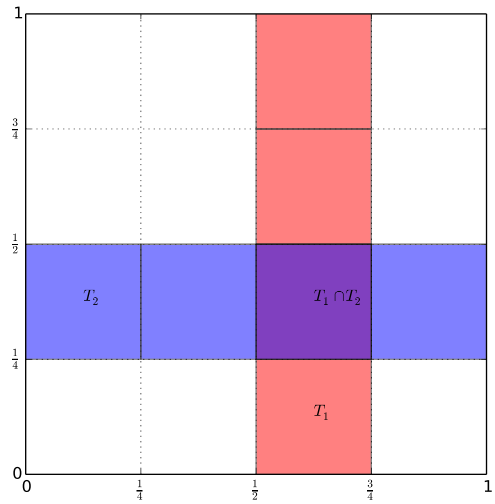

FIGURE 4 – Le carré, T_1 en rouge, T_2 en bleu, $T_1 \cap T_2$

En d'autres termes, on a « construit »un espace [6] pour la modèlisation de l'expérience de tirer indépendamment deux dés à 4 faces numérotées de 0 à 3 [7]. Cet espace est différent de l'espace naturel, fini, où l'on prend $\Omega' = \{0, \dots, 3\}^2$ et $\mathbb{P}'$ la probabilité uniforme sur Ω'.

Les calculs relativement à un espace ou à l'autre donnent les mêmes résultats concernant les expériences à deux dés, le nouvel espace permet en plus de se garder la possibilité de construire d'autres v.a. n'ayant rien à voir avec l'expérience des deux dés.

Variables aléatoires réelles, opérations

Une fois que l'on a une ou plusieurs v.a.r, on peut en fabriquer de nouvelles par toutes les opérations algébriques et compositions avec les fonctions numériques de variable réelle usuelles.

Proposition 5. *Soit X, Y, Z,... des v.a.r*

1. *(Stabilité par CL) Si λ, μ sont des nombres réels alors $\lambda.X + \mu.Y$ est une v.a.r.*
2. *(Stabilité par composition I) Si $h : I \to \mathbb{R}$ est une fonction réelle de variable réelle, X à valeurs dans I, un intervalle de $\mathbb{R}$, alors [8] $Y = h(X)$ définie par*

$$\forall \omega \in \Omega, Y(\omega) = [h(X)](\omega) = h(X(\omega)) = (h \circ X)(\omega)$$

   est une variable aléatoire réelle.

3. *(Stabilité par composition II) Si $h : \mathbb{R}^2 \to \mathbb{R}$ est une fonction réelle de deux variables réelles, $Z = h(X,Y)$ définie par*

$$\forall \omega \in \Omega, Z(\omega) = [h(X,Y)](\omega) = h(X(\omega), Y(\omega))$$

   est une variable aléatoire réelle.

6. En fait deux : $(\Omega, \mathscr{T}_0, \mathbb{P}_0)$ et $(\Omega, \mathscr{T}, \mathbb{P})$, ce dernier n'ayant été que très incomplètement défini.
7. Les amateurs de jeux de rôles apprécieront
8. modulo des hypothèses très faibles sur h, toutes les fonctions h que nous utilisons vérifient ces hypothèses relevant de la logique mathématique la plus fondamentale.

Commentaires sur les hypothèses : on peut prendre h continue, continue par morceaux, monotone, et d'une façon générale, toute fonction que l'on rencontre dans la pratique.

Exemples : Prendre une v.a, prendre son max, son min, mettre au carré, prendre le log, l'exp,....

Exercice 1.— Jet de deux dés à 4 faces, loi de la somme, loi de la différence.

Indication:Faire une décomposition des évenements suivant la valeur prise par le premier dé (*i.e.* formule des probabilités totales)

Variables aléatoires réelles, loi, fonction de répartition

Définition 6. *Soit un espace probabilisé $(\Omega, \mathscr{T}, \mathbb{P})$. Soit X une variable aléatoire réelle. La loi de X est la donnée de la famille de nombres $\{\mathbb{P}(X \in I), I \text{ intervalle de } \mathbb{R}\}$*

Plus modestement, on peut se contenter de la sous-famille des intervalles $]\infty, x]$, $x \in \mathbb{R}$ pour reconstruire la famille complète via les propriétés axiomatiques de $\mathbb{P}$. La loi de X est donc donnée par la connaissance de la fonction de répartition de X,

$$\begin{array}{rccl} F_X : & \mathbb{R} & \to & [0,1] \\ & x & \mapsto & \mathbb{P}(X \leq x) := \mathbb{P}(X \in]-\infty, x]) \end{array}$$

Par exemple, pour $a < b$, on a

$$\mathbb{P}(X \in]a,b]) = F_X(b) - F_X(a), \mathbb{P}(X \in]a,+\infty[) = 1 - F_X(a)$$

Proposition 7. *La fonction de répartition F_X d'une v.a.r X est une fonction $\mathbb{R} \to [0,1]$, vérifiant*

1. *F_X est croissante*
2. $\lim_{x\to-\infty} F_X(x) = 0$, $\lim_{x\to+\infty} F_X(x) = 1$
3. *(Hors programme) F_X est continue à droite en tout point*

1.2 Variables aléatoires réelles, nombre fini de valeurs

Si X est une variable aléatoire (réelle) prenant un nombre fini de valeurs distinctes $x_1, \dots, x_N$, on a revu dans le chapitre de révisions que la loi de X est la donnée des nombres

$$p_1 = \mathbb{P}(X = x_1), \dots, p_N = \mathbb{P}(X = x_N)$$

On a vu sur un exemple à 3 valeurs que la donnée de cette famille est équivalente à la connaissance de la loi de X au sens donné précédemment. Plus précisément :

Si I est un intervalle de $\mathbb{R}$,

$$\mathbb{P}(X \in I) = \sum_{k=1}^{N} 1\!\!1_{\{x_k \in I\}} p_k$$

Démonstration. Le sens facile, c'est de remarquer que chaque singleton $\{x_n\}$ est un intervalle, si on connaît la loi de X au sens étendu, alors on connaît $\mathbb{P}(X \in [x_k, x_k]) = \mathbb{P}(X = x_k)$ et donc on connaît la loi de X au sens première année.

L'autre sens est plus difficile conceptuellement : on connaît une quantité finie de nombres p_k et on arrive à reconstruire les nombres, possiblement en quantité possiblement infinie, $\mathbb{P}(X \in I)$. Il y a beaucoup d'intervalles dans $\mathbb{R}$!

L'assertion « X prend ses valeurs dans l'ensemble fini $\{x_1, \ldots, x_N\}$ »se traduit en termes probabilistes par « la famille finie d'événements $\{X = x_1\}, \ldots, \{X = x_N\}$ forme un *système complet d'événements*, *i.e.*

1. Leur alternative est l'événement certain

$$\cup_{k=1}^{N} \{X = x_k\} = \{X = x_1\} \cup \cdots \cup \{X = x_N\} = \Omega$$

2. Ils sont mutuellement incompatibles.

$$\forall k \neq \ell, \{X = x_k\} \cap \{X = x_\ell\} = \emptyset$$

On peut alors, et c'est une technique qu'on utilisera encore et encore, décomposer un événement sur ce système complet.

On dira « décomposer l'événement A suivant les valeurs possibles de la variable X »ou « conditionner l'événement A aux valeurs possibles de la variable X » lorsque l'on écrit une égalité du type

$$A = \cup_{k=1}^{N} A \cap \{X = x_k\}$$

ou (égalité qui se déduit de la précédente en appliquant $\mathbb{P}$)

$$\mathbb{P}(A) = \sum_{k=1}^{N} \mathbb{P}(A \cap \{X = x_k\})$$

Dans ce cas, si I est un intervalle, $A = \{X \in I\}$, on a

$$\begin{aligned}
\mathbb{P}(X \in I) &= \mathbb{P}(\{X \in I\} \cap \cup_{k=1}^{N} \{X = x_k\}) \\
&= \mathbb{P}(\cup_{k=1}^{N} (\{X \in I\} \cap \{X = x_k\}) \text{ (distributivité)} \\
&= \sum_{k=1}^{N} \mathbb{P}((\{X \in I\} \cap \{X = x_k\})
\end{aligned}$$

Pour $k \in \{1, \ldots, N\}$, que vaut $\mathbb{P}(\{X \in I\} \cap \{X = x_k\})$? Il y a deux possibilités

— Soit $x_k \in I$, auquel cas $\{X \in I\} \cap \{X = x_k\} = \{X = x_k\}$ et $\mathbb{P}(\{X \in I\} \cap \{X = x_k\}) = p_k$,

— Soit $x_k \notin I$, auquel cas $\{X \in I\} \cap \{X = x_k\} = \emptyset$ et $\mathbb{P}(\{X \in I\} \cap \{X = x_k\}) = 0$,

On a donc

$$\mathbb{P}(X \in I) = \sum_{k=1}^{N} 1\!\!1_{\{x_k \in I\}} p_k$$

□

Exemples

Exercice 2.— Si X suit la loi uniforme sur $\left\{\frac{k}{n}, k \in \{0, \ldots, n\}\right\}$ où $n \in \mathbb{N}^*$. Tracer histogramme, fonction de répartition, idem pour $Y = X^2$. Formule de transfert générique pour une fonction pour X. Une remarque sur cette formule ?

Exercice 3.— Tracer, à la machine, histogramme et fonction de répartition pour la loi binomiale $\mathscr{B}(n,p)$.
1. Prendre pour p fixé, des valeurs de n augmentant. Observer l'évolution des graphiques.
2. Fixer $\lambda > 0$. Prendre des valeurs de n augmentant et $p = \frac{\lambda}{n}$

Correction Ex.–2 *c.f.* Fig. 5

Pour une fonction $h : \mathbb{R} \to \mathbb{R}$, on a $\mathbb{E}(h(X)) = \dfrac{1}{n+1} \displaystyle\sum_{k=0}^{n} h\left(\frac{k}{n}\right)$. Il s'agit d'une somme de RIEMANN pour la fonction h. Si h est $\mathscr{C}^0$ sur $[0,1]$ et N est très grand, $\mathbb{E}(h(X))$ vaut approximativement $\displaystyle\int_0^1 h(x)\,dx$.

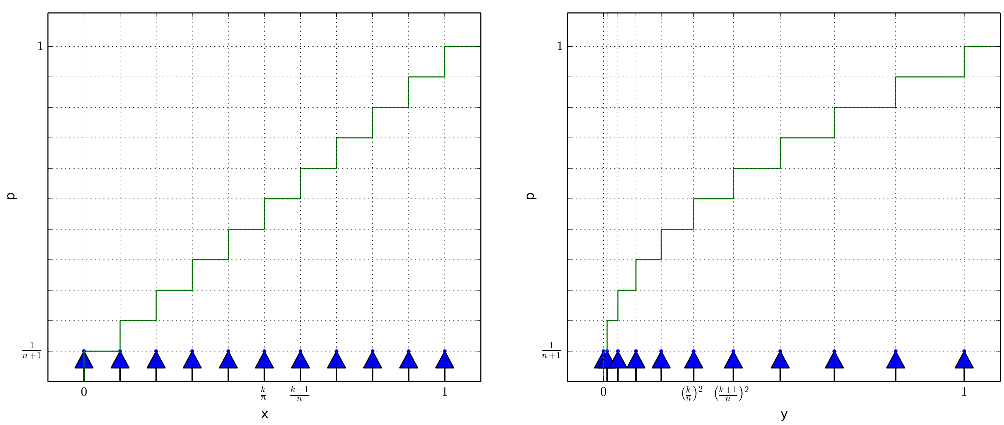

FIGURE 5 – $X \sim \mathscr{U}_{\{\frac{k}{n}, k \in \{0,\dots,n\}\}}$, sa distribution et la distribution de $Y = X^2$, fonctions de répartition.

v.a de BERNOULLI

Définition 8. *Une variable aléatoire de* BERNOULLI *est une variable aléatoire ne pouvant prendre que les valeurs* 0 *(échec,faux) ou* 1 *(succès,vrai).*

— Une indicatrice d'événement $1\!\!1_A$ est une v.a de BERNOULLI et réciproquement. On a $A = \{1\!\!1_A = 1\}$.
— Toute v.a.r prenant un nombre fini de valeurs est combinaison linéaire de v.a de BERNOULLI. Si X est à valeurs dans $\{x_1, \dots, x_K\}$, en posant $A_k = \{X = x_k\}$, on a

$$X = \sum_{k=1}^{K} x_k 1\!\!1_{A_k}$$

1.3 V.a. uniforme sur $[0,1]$

Définition et loi

On présente maintenant l'exemple fondamental de v.a.r pouvant prendre une infinité de valeurs.

Définition 9. *Soit U une variable aléatoire réelle. On dit que U suit la* loi uniforme *sur* $[0,1]$ *si, pour tout intervalle I de* $\mathbb{R}$*,*

$$\mathbb{P}(U \in I) = \textit{longueur}(I \cap [0,1])$$

Dans ce cas, on note $U \sim \mathscr{U}_{[0,1]}$.

U est p.s. à valeurs dans $[0,1]$. En effet, on a

$$\mathbb{P}(U \notin [0,1]) = \mathbb{P}(U \in]-\infty,0]) + \mathbb{P}(U \in [1,-\infty[) = 0+0 = 0$$

On peut comparer le résultat suivant (admis, difficile) avec le théorème plus faible du chapitre de révisions.

Théorème 10 (Existence d'un modèle probabiliste fondamental)**.** *Il existe un ensemble Ω, une tribu $\mathscr{T}$ sur Ω, une probabilité $\mathbb{P}$ sur $\mathscr{T}$ et une v.a. $U : \Omega \to \mathbb{R}$ tels que*

$$U \sim \mathscr{U}_{[0,1]}$$

Ayant à disposition un tel modèle $(\Omega, \mathscr{T}, \mathbb{P})$, on peut aller plus loin (*c.f.* fin de ce cours pour l'indépendance, *c.f.* derniers cours de l'année pour des idées de construction) :

Il existe une suite $(U_n)_{n\in\mathbb{N}}$ de v.a. sur $(\Omega, \mathscr{T}, \mathbb{P})$, indépendantes, distribuées uniformément sur $[0,1]$. Sur cet espace, à l'aide de cette suite de v.a. uniformes et indépendantes, on peut, en utilisant les techniques de simulation/construction de v.a. décrites dans les sections **Simulations et construction de v.a prenant un nombre fini de valeurs** et **Lois conditionnelles, Simulation informatique,** construire une chaine de MARKOV de matrice de transition donnée ou les modèles de tirage au sort avec ou sans remise.

Fonction de répartition d'une loi uniforme

— Calcul et graphe. On a, si $U \sim \mathscr{U}_{[0,1]}$,

$$\forall u \in \mathbb{R}, F_U(u) = \begin{cases} 0 & \text{si } u \leq 0 \\ u & \text{si } 0 \leq u \leq 1 \\ 1 & \text{si } u \geq 1 \end{cases}.$$

— Fonction de répartition du carré d'une telle variable ? Soit $V = U^2$. V est p.s. à valeurs dans $[0,1]$, on a, pour $0 \leq v \leq 1$,

$$\mathbb{P}(V \leq v) = \mathbb{P}(U^2 \leq v \text{ et } U \in [0,1]) = \mathbb{P}(0 \leq U \leq \sqrt{v}) = \sqrt{v}$$

et donc

$$\forall v \in \mathbb{R}, F_V(v) = \begin{cases} 0 & \text{si } v \leq 0 \\ \sqrt{v} & \text{si } 0 \leq v \leq 1 \\ 1 & \text{si } v \geq 1 \end{cases}.$$

— *c.f.* fig. 6.

Remarque sur la probabilité de prendre une valeur « exactement ».

$$\forall x \in \mathbb{R}, \mathbb{P}(U = x) = \mathbb{P}(U \in [x,x]) = 0$$

En conséquence, U est aussi p.s. à valeurs dans $]0,1[$.

V.a. uniforme sur un intervalle borné

Définition 11. *Soit X une variable aléatoire réelle, a, $b \in \mathbb{R}$, $a < b$. On dit que X suit la* loi uniforme *sur $[a,b]$ si, pour tout intervalle I de $\mathbb{R}$,*

$$\mathbb{P}(X \in I) = \frac{longueur(I \cap [a,b])}{longueur([a,b])} = \frac{1}{b-a} longueur(I \cap [a,b])$$

Dans ce cas, on note $X \sim \mathscr{U}_{[a,b]}$.

1. Si $U \sim \mathscr{U}_{[0,1]}$ alors en posant $X = a + (b-a)U$, $X \sim \mathscr{U}_{[a,b]}$.
2. Si $X \sim \mathscr{U}_{[a,b]}$ alors en posant $U = \frac{X-a}{b-a}$, $U \sim \mathscr{U}_{[0,1]}$.

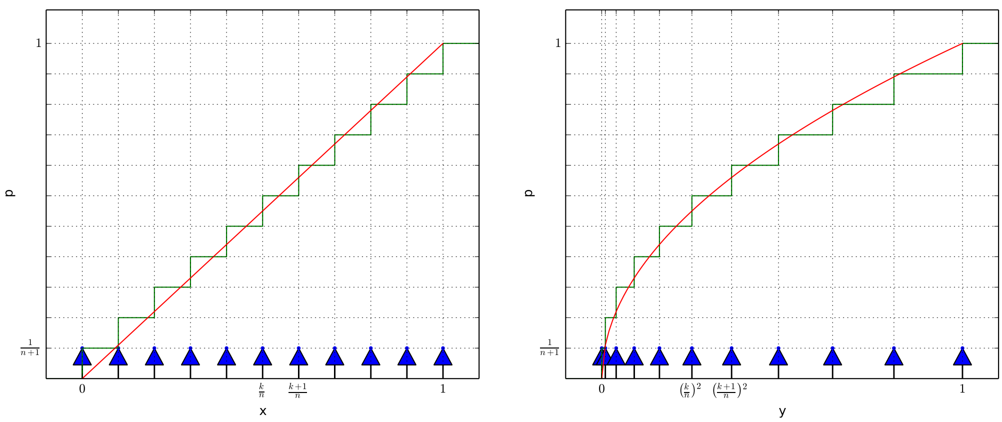

FIGURE 6 – $X \sim \mathscr{U}_{[0,1]\}}$, $Y = X^2$, fonctions de répartition, comparaison avec le cas discret.

Simulations et construction de v.a prenant un nombre fini de valeurs.

Pour simuler (informatiquement) des variables aléatoires, les langages informatiques disposent en général d'un *générateur* fondamental de nombres aléatoires. L'instruction s'appelle toujours `rand()` ou `random()`.

Pour nous ce sera `np.random.rand()` où `np` est l'alias de `numpy` Chaque appel à ce générateur retourne un nombre aléatoire de distribution uniforme sur $[0,1[$, indépendant des précédents. Cela signifie que si l'on fait plusieurs appels successifs, la suite des nombres obtenus a les mêmes propriétés statistiques (distribution, moyenne, ...) qu'une suite (abstraite) de valeurs prises par $U_1, \dots, U_n$ où $U_1, \dots, U_n$ sont des v.a indépendantes et

$$\forall k \in \{1, \dots, n\}, U_k \sim \mathscr{U}_{[0,1]}$$

La question qui se pose est la suivante : à partir d'une v.a. uniforme sur $[0,1]$, U, comment construire une v.a. (prenant un nombre fini de valeurs) ayant une loi discrète finie donnée. Reformulé pour Python, la question est : comment réécrire une fonction du type `np.random.choice` en utilisant `np.random.rand` ?

Supposons que l'on doive simuler une v.a. X prenant trois valeurs distinctes (non nécessairement numériques), x_1, x_2 et x_3. L'information dont on dispose est la distribution de X

$$p_1 := \mathbb{P}(X = x_1),\ p_2 := \mathbb{P}(X = x_2),\ p_3 := \mathbb{P}(X = x_3),$$

La méthode est simple (graphique) On découpe l'intervalle $[0,1]$ en trois intervalles disjoints I_1, I_2 et I_3 de longueurs respectives p_1, p_2 et p_3. On tire au sort U grâce au générateur uniforme. Si $U \in I_1$, on prend $X = x_1$, si $U \in I_2$, on prend $X = x_2$,...

En d'autres termes, on pose

$$I_1 = [0, p_1[,\ I_2 = [p_1, p_1 + p_2[,\ I_3 = [p_1 + p_2, 1[\ \text{ et}$$

$$X = \begin{cases} x_1 & \text{si } U \in I_1 \\ x_2 & \text{si } U \in I_2 \\ x_3 & \text{si } U \in I_3 \end{cases}$$

En termes de fonction sur Ω,

$$\forall \omega \in \Omega, X(\omega) = \begin{cases} x_1 & \text{si } U(\omega) \in I_1 \\ x_2 & \text{si } U(\omega) \in I_2 \\ x_3 & \text{si } U(\omega) \in I_3 \end{cases}$$

On a $\{X = x_1\} = \{U \in I_1\}$ et donc $\mathbb{P}(X = x_1) = \mathbb{P}(U \in I_1) = p_1$, etc...

En Python cela donne la fonction suivante

```
def VA3finie(p):
    """
    VA3finie(p): liste de 3 nombres positifs sommant à 1
    retourne un tirage aléatoire d'un nombre entre 0 et 2 avec
    la distribution p
    P(VAfinie(p)=k)=p[k]
    """
    u=np.random.rand() #tire un nombre uniformement entre 0 et 1
    if(u<=p[0]):
        return 0
    if(u<=p[0]+p[1]):
        return 1
    if(u<=p[0]+p[1]+p[2]):#RQ: ne sert à rien, automatiquement vrai si p correct
        return 2
    return -1 #On n'arrive jamais à ce point si p correct
```

Pour finir, généralisons au cas de K valeurs. Si X prend un nombre fini de valeurs numériques ordonnées $x_1 < x_2 < \cdots < x_K$, avec probabilités p_1, p_2,..p_K, ce mécanisme peut s'exprimer à l'aide de la fonction de répartition de X. Si F (ne dépendant que de la distribution connue de X) est définie par

$$\forall x \in \mathbb{R}, F(x) := \mathbb{P}(X \leq x)$$

(On pose $x_0 := -\infty$, $F(x_0) := 0$)

$$X = x_k \text{ si } U \in I_k := [F(x_{k-1}), F(x_k)[$$

Ce qui donne en Python :

```
def VAfinie(p):
    """
    VAfinie(p): p liste de nombres positifs sommant à 1
    retourne un tirage aléatoire d'un nombre entre 0 et len(p)-1 avec
    la distribution p
    P(VAfinie(p)=k)=p[k]
    """
    u=np.random.rand() #tire un nombre uniformement entre 0 et 1
    q=0.0
    for y in range(len(p)) :
        q=q+p[y]
        if(u<=q) :
            return y
    return -1 #si erreurr!!!
```

2 Variables aléatoires réelles, espérance

2.1 Espérance et v.a.r intégrables

On admet qu'il existe une opération $\mathbb{E}$, agissant dans un premier temps sur les variables aléatoires réelles *positives*, dont les valeurs sont soit un nombre réel, soit le symbole ∞, et telle que

1. $\mathbb{E}(\mathbb{1}) = 1, \mathbb{E}(0) = 0$
2. Si $A \in \mathscr{T}$, $\mathbb{E}(\mathbb{1}_A) = \mathbb{P}(A)$,
3. Si X est une variable aléatoire réelle prenant des valeurs positives alors $\mathbb{E}(X)$ a toujours un sens. [9] Il s'agit soit d'un nombre réel positif, soit du symbole « infini » (∞).
4. Dans ce cas, on a l'équivalence $\mathbb{E}(X) = 0$ si et seulement si $\mathbb{P}(X = 0) = 1$.
5. (Croissance de l'espérance) Si X et Y sont deux variables aléatoires réelles prenant des valeurs positives et si $X \leq Y$, alors [10]
$$0 \leq \mathbb{E}(X) \leq \mathbb{E}(Y)$$

1. (Additivité de l'espérance) Si X et Y sont deux variables aléatoires réelles prenant des valeurs positives, alors, avec l'extension naturelle de cette écriture au cas où le symbole $+\infty$ apparaît,
$$\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$$
2. (Homogénéité de l'espérance) Si X est une variable aléatoire réelle prenant des valeurs positives et $\lambda \in [0, +\infty[$, alors, avec conventions particulières pour gérer le cas $\mathbb{E}(X) = \infty$, on a
$$\mathbb{E}(\lambda.X) = \lambda.\mathbb{E}(X)$$

1. Ce qui est à comprendre c'est que $\mathbb{E}$ est un genre de moyenne. On moyenne sur tous les « ω »
2. $\mathbb{E}$ se comporte à beaucoup d'égards (linéarité, positivité) comme le symbole $\sum$ ou le symbole $\int$, ce qui va ressortir lors de diverses formules de transfert.

Variables aléatoires réelles intégrables

Définition 12. *Soit X une variable aléatoire réelle.*

1. *On dit que X est* intégrable *ou que X* admet une espérance *si* $\mathbb{E}(|X|) \neq \infty$.
2. *Si X est intégrable, $X^+ = \max(X, 0)$ et $X^- = -\min(X, 0)$ le sont aussi, sont à valeurs positives, on a*
$$X = X^+ - X^-, \; |X| = X^+ + X^-,$$
et on pose
$$\mathbb{E}(X) := \mathbb{E}(X^+) - \mathbb{E}(X^-)$$

9. La formule de Cavalieri
$$\mathbb{E}(X) = \int_0^{+\infty} \mathbb{P}(X > t)\, dt$$
vue en exercice dans le chapitre précédent peut servir de définition pour le cas le plus général. En particulier, si X prend un nombre fini de valeurs réelles positives, $\mathbb{E}(X)$ désigne l'espérance de X au sens de la première année.

10. avec l'extension naturelle de cette écriture au cas où l'un des deux vaut ∞

RQ : ceci n'est pas sans rappeler des éléments de démonstration du théorème $ACV \Rightarrow CV$. C'est exprès, comme on le verra dans le chapitre sur les v.a. à densité. On a les règles de calcul suivantes.

Proposition 13. *Soit X, Y des variables aléatoires réelles* intégrables *et* λ, μ *deux nombres réels,*

1. *(Linéarité de l'espérance)* $Z = \lambda.X + \mu.Y$ *est une variable aléatoire réelle intégrable et*

$$\mathbb{E}(Z) = \mathbb{E}(\lambda.X + \mu.Y) = \lambda.\mathbb{E}(X) + \mu.\mathbb{E}(Y)$$

2. *(Croissance) Si* $X \leq Y$ *alors* $\mathbb{E}(X) \leq \mathbb{E}(Y)$ *avec égalité ssi* $\mathbb{P}(X = Y) = 1$.
3. *(Inégalité triangulaire)*

$$|\mathbb{E}(\lambda.X + \mu.Y)| \leq |\lambda|.\mathbb{E}(|X|) + |\mu|.\mathbb{E}(|Y|)$$

La façon dont se distribue $h(X)$ ou $h(X,Y)$ pour une fonction h quelconque est un de nos problèmes centraux.

Nombre fini de valeurs.

Dans le cas d'une v.a.r. X prenant un nombre fini de valeurs, on retrouve, par linéarité de $\mathbb{E}$, la formule de première année

Si X est une v.a.r., à valeurs dans l'ensemble fini $\{x_k, k \in \{1, \dots, K\}\} \subset \mathbb{R}$, alors

1. $X = \sum_{k=1}^{K} x_k 1\!\!1_{\{X = x_k\}}$
2. X admet une espérance,

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

Exercice 4.— Soit n un entier naturel non nul, soit X une variable suivant la loi $\mathscr{U}_{\{0, \frac{1}{n}, \dots, \frac{n-1}{n}, \frac{n}{n}\}}$. Soit $f : [0,1] \to \mathbb{R}$ une fonction continue.
1. Donner une formule pour $S_n(f) = \mathbb{E}(f(X))$. Appliquer cette formule à l'exemple $f : x \mapsto x^2$.
2. Quelle est, en général, la limite de $S_n(f)$ lorsque $n \to +\infty$.
3. Quelle est la valeur de cette limite pour l'exemple proposé ?
c.f. figure 6.

L'espérance d'une fonction d'une variable uniforme

Supposons que $U \sim \mathscr{U}_{[0,1]}$. On admet que l'on a la formule de transfert suivante :

Si $h : \mathbb{R} \to \mathbb{R}$ est une fonction, on a

$$\mathbb{E}(h(U)) = \int_{-\infty}^{+\infty} h(u) 1\!\!1_{\{u \in [0,1]\}} \, du = \int_0^1 h(u) \, du$$

pourvu que cette dernière intégrale soit absolument convergente.

On convient que $\mathbb{E}(h(U)) = +\infty$ si h est positive et l'intégrale diverge vers $+\infty$.

Exemples :

1. Montrer que le formule de transfert en question est vraie au cas où h est constante par morceaux, *i.e.* h est de la forme $h = \sum_{k=1}^{K} x_k 1\!\!1_{I_k}$ où les I_k sont des intervalles disjoints de $\mathbb{R}$. (On pourra vérifier la formule dans le cas d'un seul intervalle et argumenter par linéarité des opérations).
2. Calculer $\mathbb{E}(U^2)$, $E(\frac{1}{1+U})$, $\mathbb{E}(\frac{1}{\sqrt{1-U^2}})$.

3. Calculer $\mathbb{E}(e^{-\lambda.U})$ où $\lambda \in \mathbb{R}$.
4. Les v.a.r $\ln U$, $\frac{1}{U}$, $\frac{1}{\sqrt{U}}$ sont-elles intégrables ?
5. Soit V une v.a.r uniforme sur $[0,1]$. Donner l'espérance de $U+V$, de $U-V$.
6. Soit $a<b$, $a,b \in \mathbb{R}$, V une v.a.r uniforme sur $[a,b]$. Donner la formule de transfert générique donnant $\mathbb{E}(h(V))$.

Variables intégrables, Inégalité de MARKOV

Théorème 14 (Inégalité de MARKOV)**.** *Si X est une variable aléatoire réelle,* positive*, alors, pour tout seuil $\lambda > 0$,*

$$\mathbb{P}(X \geq \lambda) \leq \frac{\mathbb{E}(X)}{\lambda}$$

Démonstration. On peut se limiter au cas où X est intégrable. Comme X est positive, on a

$$1\!\!1_{\{X \geq \lambda\}} \leq \frac{X}{\lambda}$$

Par croissance et linéarité de l'espérance, il vient

$$\mathbb{E}(1\!\!1_{\{X \geq \lambda\}}) \leq \frac{\mathbb{E}(X)}{\lambda}$$

□

2.2 Variables de carré intégrable

Variables de carré intégrable, définition

Définition 15. *Soit X une v.a.r. On dit que X est de carré intégrable ou admet une variance si X^2 est intégrable, i.e. admet une espérance.*

Exercice 5.— Soit $U \sim \mathscr{U}_{[0,1]}$.
1. U est de carré intégrable ?
2. $\frac{1}{\sqrt{U}}$ est intégrable sans être de carré intégrable ?

Variables de carré intégrable, CAUCHY–SCHWARZ

Théorème 16 (Inégalité de CAUCHY–SCHWARZ)**.** *Si X et Y sont deux v.a.r de carré intégrable alors $X.Y$ est intégrable et on a*

$$|\mathbb{E}(X.Y)| \leq \mathbb{E}(|X|.|Y|) \leq \mathbb{E}(|X|^2)^{\frac{1}{2}}\mathbb{E}(|Y|^2)^{\frac{1}{2}}$$

Démonstration. On commence par noter que si X et Y sont de carré intégrable alors toute combinaison linéaire de X et Y l'est aussi. En effet, concernant la somme,

$$0 \leq (X+Y)^2 \leq 4X^2 + 4Y^2$$

et comme par hypothèse, $\mathbb{E}(X^2)$ et $\mathbb{E}(Y^2)$ sont finis, 4 fois leur somme l'est aussi et, par croissance de l'espérance, $\mathbb{E}((X+Y)^2$ est finie.

Si λ, μ deux nombres réels, alors $\lambda.X$ et $\mu.Y$ sont aussi de carré intégrable et leur somme aussi.

Comme $X.Y = \frac{1}{4}((X+Y)^2 - (X-Y)^2)$, on en déduit que XY est intégrable.

Considérons maintenant la fonction $f : \lambda \in \mathbb{R} \mapsto \mathbb{E}((X+\lambda.Y)^2)$. Cette fonction est bien définie, à valeurs *réelles*.

On a, en développant le carré et en utilisant la linéarité de l'espérance d'une part et en utilisant la positivité de l'espérance, que, pour *tout* $\lambda \in \mathbb{R}$,

$$0 \leq \mathbb{E}((X+\lambda.Y)^2) = \mathbb{E}(X^2) + 2\lambda\mathbb{E}(XY) + \lambda^2\mathbb{E}(Y^2)$$

La fonction f est donc une fonction polynômiale, de degré au plus 2, positive.

1. Si $\mathbb{E}(Y^2) \neq 0$, f est de degré 2 et son discriminant (réduit) est *négatif* :

$$(\mathbb{E}(XY))^2 - \mathbb{E}(X^2).\mathbb{E}(Y^2) \leq 0,$$

ce qui, une fois réarrangé donne l'inégalité attendue.

2. Si $\mathbb{E}(Y^2) = 0$, on a affaire à une fonction affine positive et donc sa pente est nulle : $\mathbb{E}(XY) = 0$ et l'inégalité est aussi vérifiée.

L'inégalité intermédiaire avec $\mathbb{E}(|X|.|Y|)$ provient de l'application de l'inégalité démontrée dans le cas positif aux variables $|X|$ et $|Y|$ qui sont elles aussi de carré intégrable. □

Une conséquence en est la suivante :

Proposition 17. *Si X est de carré intégrable, elle est intégrable et*

$$|\mathbb{E}(X)| \leq \mathbb{E}(|X|) \leq \mathbb{E}(X^2)^{\frac{1}{2}}$$

Démonstration. On applique le théorème précédent à X et $Y = \mathbb{1}$. Y est de carré intégrable et $\mathbb{E}(Y^2) = \mathbb{E}(Y) = 1$. □

Variables de carré intégrable, Variance et covariance

Proposition-Définition 18.

1. *Si X est de carré intégrable* [11]*, sa variance est définie par*

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

2. *Son* écart-type*,* $\sigma(X) = \mathbb{V}(X)^{\frac{1}{2}}$.
3. *Si X et Y sont de carré intégrable, leur covariance est définie par*

$$\mathbb{C}ov(X,Y) = \mathbb{E}((X - \mathbb{E}(X)).(Y - \mathbb{E}(Y)))$$

4. *On a* $|\mathbb{C}ov(X,Y)| \leq \sigma(X).\sigma(Y)$.

11. La terminologie officielle est « X admet une variance ».

Formule de KOENIG–HUYGENS

Proposition-Définition 19. *1. Si X admet une variance, sa variance vaut*

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

2. Si X et Y admettent une variance, leur covariance vaut

$$\mathbb{C}ov(X,Y) = \mathbb{E}(X.Y) - \mathbb{E}(X).\mathbb{E}(Y)$$

Démonstration. On ne démontre que la première égalité, la seconde relevant de la même technique de calcul. On a, en développant le carré, par linéarité de l'espérance et le fait que $\mathbb{E}(C) = C$ pour une constante C,

$$\begin{aligned} \mathbb{V}(X) &= \mathbb{E}(X^2 - 2X.\mathbb{E}(X) + \mathbb{E}(X)^2) \\ &= \mathbb{E}(X^2) - 2\mathbb{E}(X).\mathbb{E}(X) + \mathbb{E}(X)^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2 \end{aligned}$$

□

Variables de carré intégrable, centrage, normalisation

Une variable aléatoire *positive* est d'espérance nulle si et seulement si $\mathbb{P}(X = 0) = 1$. (On dit que X est nulle *presque sûrement*.)

Une variable aléatoire est de variance nulle, $\mathbb{V}(X) = 0$, si et seulement si $\mathbb{P}(X = \mathbb{E}(X)) = 1$. Autrement dit ssi X est une constante presque sûrement.

Supposons que X est une variable aléatoire admettant une variance.

— si C une constante (*i.e.* une variable ne prenant presque sûrement qu'une valeur), on a

$$\mathbb{V}(X + C) = \mathbb{V}(X)$$

— si λ est une constante, $\mathbb{V}(\lambda.X) = \lambda^2\mathbb{V}(X)$.

On peut se servir de ces deux propriétés pour centrer, puis normaliser une v.a. non constante X, de carré intégrable.

$$X^* = \frac{X - \mathbb{E}(X)}{\sigma(X)}$$

est une variable d'espérance nulle (centrée), de variance 1. On dit qu'elle est *centrée-réduite*.

Remarquer que si X est mesurée dans une certaine unité, alors $\mathbb{E}(X)$, $\sigma(X)$ sont dans la même unité et X^* est sans unité.

Exercice 6.—
1. Donner espérance et variance d'une v.a $U \sim \mathscr{U}_{[0,1]}$.
2. Donner espérance et variance d'une v.a $V \sim \mathscr{U}_{[a,b]}$.
3. Quelle est la centrée réduite de U ? de V ?

Exercice 7.— Si $X \sim \mathscr{B}\left(n, \frac{1}{2}\right)$. Donner l'histogramme de X^*. Dans quel ensemble fini X^* prend-elle ses valeurs ?

Exercice 8.— Même question si $X \sim \mathscr{U}_{\{0,\ldots,2n\}}$ où $n \in \mathbb{N}^*$.

La règle de la somme

Proposition 20. *Si X et Y sont de carré intégrable alors*

$$\mathbb{V}(X+Y) = \mathbb{V}(X) + \mathbb{V}(Y) + 2\,\mathbb{C}ov(X,Y)$$

Exercice 9.— Donner une formule pour $\mathbb{V}(\sum_{n=1}^{N} X_n)$ où les X_n sont de carré intégrable.

Variables de carré intégrable, BIENAYMÉ–TCHEBYCHEFF

On a vu que l'espérance d'une v.a.r est une sorte de moyenne des valeurs que peut prendre X. Les valeurs de X se répartissent donc de part et d'autre $\mathbb{E}(X)$.

La valeur de la variance mesure quant à elle, assez grossièrement, la *façon* dont les valeurs prises par X se répartissent autour de $\mathbb{E}(X)$.

L'inégalité de BIENAYMÉ–TCHEBYCHEFF est une traduction de l'idée de répartition de X autour de son espérance.

Théorème 21 (Inégalité de BIENAYMÉ–TCHEBYCHEFF)**.** *Si X est de carré intégrable, on a, pour tout seuil $\lambda > 0$,*

$$\mathbb{P}(|X - \mathbb{E}(X)| > \lambda) \leq \frac{\mathbb{V}(X)}{\lambda^2}$$

Démonstration. Il suffit d'appliquer l'inégalité de MARKOV à la variable intégrable positive $Z = (X - \mathbb{E}(X))^2$ et de remarquer que les événements $(|X - \mathbb{E}(X)| > \lambda)$ et $(Z > \lambda^2)$ sont égaux. On a $\mathbb{E}(Z) = \mathbb{V}(X)$. $\square$

Théorème 22 (Inégalité de BIENAYMÉ–TCHEBYCHEFF)**.** *Si X est de carré intégrable, non constante, on a, pour tout seuil $\lambda > 0$,*

$$\mathbb{P}(|X - \mathbb{E}(X)| > \lambda.\sigma(X)) \leq \frac{1}{\lambda^2}$$

Notamment, si $\lambda = 2$, on obtient qu'avec probabilité supérieure à 75%, X est distant de la constante $\mathbb{E}(X)$ de moins de 2 écart-types. **Ceci, quelle que soit la distribution de X.**

Cette inégalité n'est pertinente que pour $\lambda > 1$.

Exercice 10.— Sur $n = 100$ palourdes pêchées en Bretagne sud, on effectue la mesure de leur diamètre. On note d_k le diamètre de la palourde portant le numéro $k \in \{1,\ldots,n\}$. On regroupe ces mesures en un tableau et on calcule (via Excel par exemple) leur moyenne $\overline{d} = \frac{1}{n}\sum_{k=1}^{n} d_k$ et l'écart-type

$$\sigma = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(d_k - \overline{d})^2}$$

1. On prend une palourde au hasard et on note D son diamètre. Calculer $\mathbb{E}(D)$ et $\mathbb{V}(D)$ en fonction de $\overline{d}$ et σ.
2.a. Soit $\alpha > 0$. Enoncer l'inégalité de BIENAYMÉ–TCHEBYCHEFF pour D au seuil $\alpha\sigma$.
2.b. Comment choisir α pour être sûr que moins d'un quart des palourdes ont un diamètre qui diffère du diamètre moyen de plus de $\alpha.\sigma$ en valeur absolue.

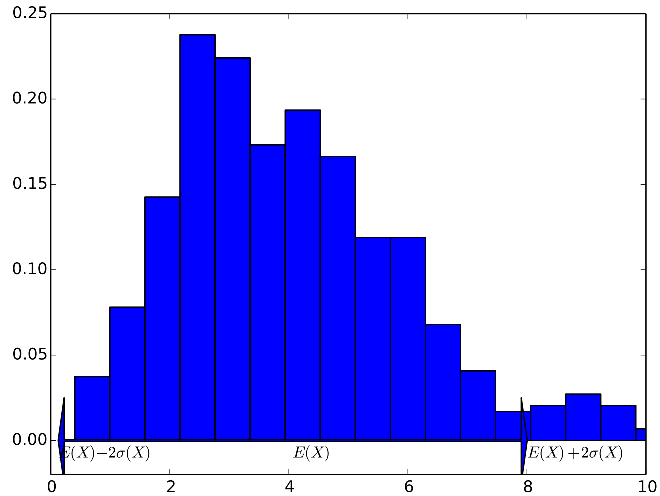

FIGURE 7 – BIENAYMÉ–TCHEBYCHEFF : Au moins 75% de la population est à moins de 2 écarts-types de la moyenne.

3. En fait, on met nos n palourdes dans un sac que l'on secoue, ce qui fait que les plus grandes palourdes vont plus facilement être tirées. On admet que la palourde numéro k est tirée avec probabilité

$$p_k = \frac{d_k}{n.\overline{d}}$$

3.a. Vérifier que $\sum_{k=1}^{n} p_k = 1$.
3.b. On note D' le diamètre de la palourde tirée au sort suivant ce mode. Pourquoi, heuristiquement, $\mathbb{E}(D') \geq \mathbb{E}(D)$?
3.c. Donner une formule pour $\mathbb{E}(D')$ en fonction de $\mathbb{E}(D^2)$.
3.d. Démontrer mathématiquement l'inégalité heuristique.

Variables de carré intégrable, coeff. de corrélation et régression linéaire

Le problème de la régression linéaire est le suivant. On dispose de deux variables X (non constante) et Y, de carré intégrable. Faire la régression linéaire de Y sur X consiste à trouver la variable $\hat{Y}$, fonction affine de X « la plus proche »de Y *au sens des moindres carrés*, *i.e.* trouver deux réels a et b tels que

$$\mathbb{E}((Y-(aX+b))^2)$$

soit minimal. On pose alors $\hat{Y} = aX+b$.

Le calcul montre que

Théorème 23.

$$\hat{Y} = \frac{\mathbb{C}ov(X,Y)}{\mathbb{V}(X)}(X-\mathbb{E}(X))+\mathbb{E}(Y)$$

ou alors, en variables centrées et réduites, si X et Y sont non constantes,

$$\hat{Y^*} = \rho(X,Y)X^*$$

où $\rho(X,Y) = \frac{\mathbb{C}ov(X,Y)}{\sigma(X)\sigma(Y)}$ est le coefficient de corrélation linéaire.

Démonstration. Posons, pour $a,b \in \mathbb{R}$, $f(a,b) = \mathbb{E}((Y-(aX+b))^2)$ et développons cette expression. On a

$$f(a,b) \quad = \quad \mathbb{E}(Y^2)+a^2\mathbb{E}(X^2)+b^2-2a\mathbb{E}(X.Y)-2b\mathbb{E}(Y)+2ab\mathbb{E}(X)$$

Ce serait plus agréable si les variables X et Y étaient centrées, il y aurait deux termes de moins. Reprenons l'expression de départ et forçons le centrage.

$$\begin{aligned} f(a,b) &= \mathbb{E}((\underbrace{Y-\mathbb{E}(Y)}_{=\tilde{Y}}-(\underbrace{a}_{=\alpha}\underbrace{(X-\mathbb{E}(X)}_{=\tilde{X}}+\underbrace{b-\mathbb{E}(Y)+a\mathbb{E}(X)}_{=\beta}))^2) \\ &= \mathbb{E}((\tilde{Y}-(\alpha\tilde{X}+\beta))^2) \\ &= g(\alpha,\beta) \end{aligned}$$

où on a posé $\alpha = a$, $\beta = b-\mathbb{E}(Y)+a\mathbb{E}(X)$. Trouver un couple (a,b) qui rende $f(a,b)$ minimal revient à trouver un couple (α,β) qui rende $g(\alpha,\beta)$ minimal.

Développons l'expression de $g(\alpha,\beta)$, comme nous avons développé $f(a,b)$, le calcul est le même. On a

$$\begin{aligned} g(\alpha,\beta) &= \underbrace{\mathbb{E}(\tilde{Y}^2)}_{=\mathbb{V}(Y)}+a^2\underbrace{\mathbb{E}(\tilde{X}^2)}_{=\mathbb{V}(X)}+b^2-2a\underbrace{\mathbb{E}(\tilde{X}.\tilde{Y})}_{=\mathbb{C}ov(X,Y)}-2b\underbrace{\mathbb{E}(\tilde{Y})}_{=0}+2ab\underbrace{\mathbb{E}(\tilde{X})}_{=0} \\ &= \mathbb{V}(Y)+\beta^2+\alpha^2\mathbb{V}(X)-2\alpha\,\mathbb{C}ov(X,Y) \\ \text{(forme canonique)} &= \underbrace{\beta^2}_{\geq 0}+\underbrace{\mathbb{V}(X)(\alpha-\frac{\mathbb{C}ov(X,Y)}{\mathbb{V}(X)})^2}_{\geq 0}+\frac{\mathbb{V}(Y)\mathbb{V}(X)-\mathbb{C}ov(X,Y)^2}{\mathbb{V}(X)} \\ &\geq g(\frac{\mathbb{C}ov(X,Y)}{\mathbb{V}(X)},0) \end{aligned}$$

Ceci démontre que le minimum de $g(\alpha,\beta)$ est (uniquement) atteint en $(\frac{\mathbb{C}ov(X,Y)}{\mathbb{V}(X)},0)$ et donc que le minimum de f est atteint en

$$a = \frac{\mathbb{C}ov(X,Y)}{\mathbb{V}(X)},\, b = \mathbb{E}(Y)-a.\mathbb{E}(X)$$

Ce qui est ce que l'on cherche. On peut noter que le minimum vaut

$$\frac{\mathbb{V}(Y)\mathbb{V}(X)-\mathbb{C}ov(X,Y)^2}{\mathbb{V}(X)}$$

Cette quantité est positive, par l'inégalité de CAUCHY–SCHWARZ et ne vaut 0 qu'en cas d'égalité dans cette inégalité, *i.e.* lorsque Y est fonction affine de X. □

Pour trouver rapidement la formule de $\hat{Y}$, on cherche a et b tels que $\mathbb{C}ov(X,\hat{Y}) = \mathbb{C}ov(X,Y)$ et $\mathbb{E}(\hat{Y}) = \mathbb{E}(Y)$.

1. Si $\mathbb{C}ov(X,Y) = 0$ ($\rho(X,Y) = 0$), on dit que X et Y sont *non corrélées*.
2. Si $\rho(X,Y) > 0$, on dit que X et Y sont *positivement corrélées*.
3. Si $\rho(X,Y) < 0$, on dit que X et Y sont *négativement corrélées*.

Exercice 11.— Soit $X \sim \mathscr{U}_{[0,1]}$, $Y = X^2$.
1. Donner les valeurs de $\mathbb{V}(X)$, $\mathbb{V}(Y)$, $\mathbb{C}ov(X,Y)$.
2. Donner une formule pour $\hat{Y}$ la régression linéaire de Y sur X.
3. En Python, tirer 100 nombres x_i suivant la loi de X de façon indépendante. Tracer les nuages de points (x_i, y_i) le nuage $(x_i, \hat{y}_i)$ et tracer la droite des moindres carrés calculée statistiquement.

Correction Ex.–11

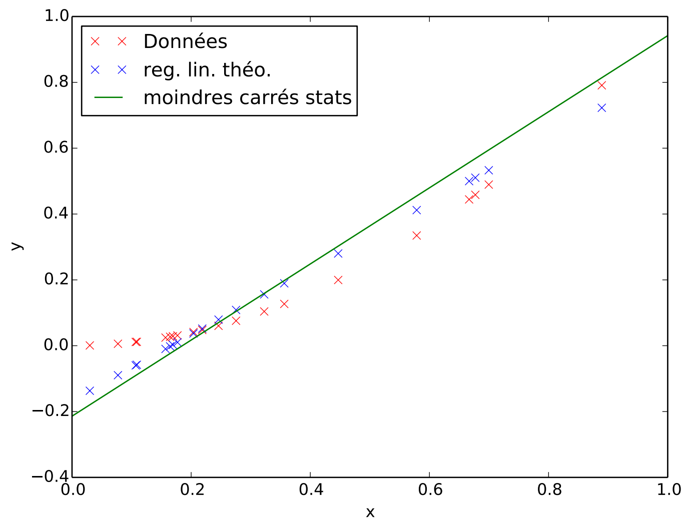

FIGURE 8 – Une série de 20 nombres x tirés indépendamment suivant $X \sim \mathscr{U}_{[0,1]}$, $Y = X^2$, les (x_i, y_i), les $(x_i, \hat{y}_i)$, la droite des moindres carrés du nuage (x_i, y_i)

Proposition 24. *Si X et Y sont non corrélées, on a*

$$\mathbb{V}(X+Y) = \mathbb{V}(X) + \mathbb{V}(Y)$$

Si $X_1, \ldots, X_n$ sont deux à deux non corrélées, i.e.

$$\forall i, j \in \{1, \ldots, n\}, i \neq j \Rightarrow \mathbb{C}ov(X_i, X_j) = 0$$

alors

$$\mathbb{V}(X_1 + \cdots + X_n) = \mathbb{V}(X_1) + \cdots + \mathbb{V}(X_n)$$

2.3 Indépendance

Indépendance d'événements

Définition 25. *Soit $(\Omega, \mathscr{T}, \mathbb{P})$ un espace probabilisé*

1. *Soit $A, B \in \mathscr{T}$ deux événements. On dit que A et B sont indépendants si $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.*
2. *Soit $A_1, \ldots, A_n \in \mathscr{T}$ une famille d'événements. On dit qu'ils sont (mutuellement) indépendants si pour toute sous-famille $A_{i_1}, \ldots, A_{i_k}$, $i_1 < i_2 < \cdots < i_k$,*
$$\mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \cap \ldots \mathbb{P}(A_{i_k})$$
3. *Soit $A_1, \ldots, A_n, \cdots \in \mathscr{T}$ une famille (infinie) d'événements. On dit qu'ils sont (mutuellement) indépendants si toute sous-famille finie est indépendante.*

Exercice 12.— Montrer que si A et B sont indépendants, A et $\overline{B}$ le sont aussi.

Exercice 13.— Montrer que si A est indépendant de lui-même alors $\mathbb{P}(A)=1$ ou $\mathbb{P}(A)=0$.

V.a. indépendantes : deux définitions équivalentes

Définition 26 (Première définition)**.** *1. Deux v.a.r X et Y sont* indépendantes *si pour tous intervalles I, J de $\mathbb{R}$,*

$$\mathbb{P}(X\in I \text{ et } Y\in J)=\mathbb{P}(X\in I).\mathbb{P}(Y\in J)$$

2. De façon analogue, les v.a.r $X_1,\dots,X_n$ sont dites (mutuellement) indépendantes *si pour tous intervalles $I_1,\dots,I_n$,*

$$\mathbb{P}(X_1\in I_1 \text{ et } \dots X_n\in I_n)=\mathbb{P}(X_1\in I_1)\dots\mathbb{P}(X_n\in I_n)$$

Traduisons ceci en termes d'espérance en utilisant la relation fondamentale $\mathbb{P}(A)=\mathbb{E}(1\!\!1_A)$.
X et Y sont indépendantes si pour tous intervalles I et J,

$$\mathbb{E}(1\!\!1_{\{X\in I\}}.1\!\!1_{\{Y\in J\}})=\mathbb{E}(1\!\!1_{\{X\in I\}}).\mathbb{E}(1\!\!1_{\{X\in J\}})$$

Prenons maintenant une fonction f de la forme $f=\sum_{k=1}^K f_k 1\!\!1_{I_k}$ et $g=\sum_{\ell=1}^L g_\ell 1\!\!1_{J_\ell}$ où $(I_k)_{1\le k\le K}$ et $(J_\ell)_{1\le\ell\le L}$ sont deux familles finies d'intervalles de $\mathbb{R}$. Un peu d'algèbre (on pourra tester avec des familles de deux intervalles) utilisant uniquement la linéarité de l'espérance montre que

$$\mathbb{E}(f(X).g(Y))=\mathbb{E}(f(X)).\mathbb{E}(g(Y))$$

On a bien évidemment un énoncé similaire pour une famille $X_1,\dots,X_n$ de variables indépendantes en prenant des fonctions $f_1,\dots,f_n$ en escalier sur $\mathbb{R}$:

$$\mathbb{E}(\prod_{\nu=1}^n f_\nu(X_\nu))=\prod_{\nu=1}^n\mathbb{E}(f_\nu(X_\nu))$$

Définition 27 (Deuxième définition équivalente)**.** *1. Deux v.a.r X et Y sont* indépendantes *si pour toutes fonctions $f,g:\mathbb{R}\to\mathbb{R}$*

$$\mathbb{E}(f(X).g(Y))=\mathbb{E}(f(X)).\mathbb{E}(g(Y))$$

pourvu que ces espérances soient bien définies.

2. De façon analogue, les v.a.r $X_1,\dots,X_n$ sont dites (mutuellement) indépendantes *si pour toutes fonctions $f_1,\dots,f_n:\mathbb{R}\to\mathbb{R}$,*

$$\mathbb{E}(\prod_{\nu=1}^n f_\nu(X_\nu))=\prod_{\nu=1}^n\mathbb{E}(f_\nu(X_\nu))$$

dès que cette formule à un sens.

Lemmes d'indépendance

On a les résultats suivants

Lemme 28 (des coalitions)**.** *Si $X_1,\dots,X_n,Y_1,\dots,Y_p$ sont indépendantes et si $u:\mathbb{R}^n\to\mathbb{R}$, $v:\mathbb{R}^p\to R$ alors, en posant*

$$U:=u(X_1,\dots,X_n) \text{ et } V:=v(Y_1,\dots,Y_p),$$

les variables U et V sont indépendantes.

Lemme 29. *Si $X_1,\dots,X_n$ sont indépendantes et si $u_1,\dots,u_n : \mathbb{R} \to \mathbb{R}$ alors, en posant*

$$U_1 := u_1(X_1), \dots U_n := u_n(X_n),$$

les variables $U_1,\dots,U_n$ sont indépendantes.

Exercice 14.— Soit E une. v.a.r à valeurs dans $]0,+\infty[$ admettant une variance, α une v.a.r à valeurs dans $\left]0,\frac{\pi}{2}\right[$, $m>0$, $g>0$ deux constantes. On suppose que E et α sont indépendantes.

On pose $D = \frac{2.E.\sin(2\alpha)}{m.g}$ la distance parcourue par une sagaie lancée avec énergie (cinétique) initiale E et angle de tir α. On pose $S = \sin(2\alpha)$.
1. Montrer que S est à valeurs dans $]0,1]$ et admet une variance.
2. Exprimer l'espérance de D en fonction de m, g, $\mathbb{E}(E)$ et $\mathbb{E}(S)$.
3. Exprimer la variance de D en fonction de m, g, $\mathbb{E}(E)$, $\mathbb{E}(S)$, $\mathbb{V}(E)$ et $\mathbb{V}(S)$.
4. Montrer la formule (propagation de l'incertitude relative en physique)

$$\frac{\mathbb{V}(D)}{\mathbb{E}(D)^2} = \frac{\mathbb{V}(E)}{\mathbb{E}(E)^2}.\frac{\mathbb{V}(S)}{\mathbb{E}(S)^2} + \frac{\mathbb{V}(E)}{\mathbb{E}(E)^2} + \frac{\mathbb{V}(S)}{\mathbb{E}(S)^2}$$

Ces quantités dépendent-elles de m et g ?
5. (Optionnel). Calculer $\mathbb{E}(S)$, $\mathbb{V}(S)$ sous l'hypothèse $\alpha \sim \mathscr{U}_{]0,\frac{\pi}{2}[}$.

Exercice 15.— Un chasseur dispose d'une seule sagaie (de masse $m = 1kg$), la cible est un troupeau de bisons à une distance entre 20 et 30m.

Il développe au lancer une énergie E comprise entre $E_{min} = 50J$ et $E_{max} = 350J$, qu'il ne contrôle pas.

Son angle de tir α, lui même incontrolé, est compris entre $\frac{\pi}{8}$ et $\frac{3\pi}{8}$.

Comme notre individu ne contrôle rien (il ne va pas survivre longtemps), on suppose qu'énergie et angle de tir sont indépendants.
1. Que valent espérance et écart-type de la distance (on arrondira aux entiers les plus proches) à laquelle la sagaie est lancée ? Quelle information obtient-on en appliquant l'inégalité de BIENAYMÉ–TCHEBYCHEFF au seuil de deux écarts-types ?

Indication:On rappelle qu'avec une énergie de lancer E, un angle de tir α, le point d'impact est à distance $d = \frac{2.E.\sin(2\alpha)}{m.g}$. On pourra interpréter « incontrolé » comme une uniformité dans le tirage au sort.
2. Tracer l'histogramme de la distance pour $NS = 100000$ simulations de lancer. (Faire 50 « bins »).
3. Evaluer numériquement la probabilité de toucher un animal. Vérifier la cohérence de moyenne et écart-type de vos simulations avec espérance et écart-type théoriques.

Corrélation et indépendance

Proposition 30. *1. Si X et Y sont deux variables de carré intégrable* indépendantes, *elles sont non corrélées : $\mathbb{C}ov(X,Y) = 0$.*

2. **La réciproque est très fausse, contrairement à ce que croit le monde entier.**

Démonstration. 1. Si X et Y sont indépendantes, en prenant $f(x) = x - \mathbb{E}(X)$ et $g(y) = y - \mathbb{E}(Y)$, $f(X).G(Y)$, $f(X)$, $g(Y)$ sont intégrables et on a donc par indépendance $\mathbb{C}ov(X,Y) = \mathbb{E}(f(X).g(Y)) = \mathbb{E}(f(X))\mathbb{E}(g(Y)) = 0$

2. Prenons un exemple assez élémentaire. Soit X une v.a.r uniformément distribuée sur l'ensemble fini $\{-1,0,+1\}$ et $Y = X^2$. On a alors

$$\mathbb{E}(X) = 0, \mathbb{E}(Y) = \frac{2}{3}, \mathbb{E}(XY) = \mathbb{E}(X^3) = 0 \text{ et } \mathbb{C}ov(X,Y) = 0$$

Ces deux variables ne sont pas indépendantes car

$$\mathbb{P}(X = 1 \text{ et } Y^2 = 1) = \mathbb{P}(X = 1) = \frac{1}{3}, \mathbb{P}(X = 1)\mathbb{P}(Y^2 = 1) = \frac{2}{9}$$

□

2.4 Probabilités conditionnelles

Ce qui a été dit dans la section de rappels de première année, s'applique tel quel au cadre théorique général. On a notamment un renforcement de la formule des probabilités totales énoncée en termes de probabilités conditionnelles. Nous nous en servirons dans le chapitre sur les variables discrètes, à la fin de l'année.

Théorème 31. *Si $B_1,\dots,B_n,\dots$ est une famille (possiblement indexée par $n \in \mathbb{N}^*$ d'événements formant un système complet d'événements incompatibles, alors, pour tout événement A*

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}(A|B_1).\mathbb{P}(B_1) + \cdots + \mathbb{P}(A|B_n).\mathbb{P}(B_n) + \dots \\ &= \lim_{N\to+\infty} \sum_{n=1}^{N} \mathbb{P}(A|B_n).\mathbb{P}(B_n)\end{aligned}$$

Rq : Si $\mathbb{P}(B_n) = 0$, la probabilité conditionnelle $\mathbb{P}(A|B_n)$ n'est pas bien définie. On prend cependant la convention $\mathbb{P}(A|B_n).\mathbb{P}(B_n) = 0$ (Zéro super absorbant) dans la formule.

Exercice 16.— Soit ε une v.a. uniformément distribuée sur $\{-1,+1\}$
1. Soit $U \sim \mathscr{U}_{[0,1]}$. On suppose que U et ε sont indépendantes et on pose $V = \varepsilon.U$. Montrer que $V \sim \mathscr{U}_{[-1,1]}$.
Indication:Pour un intervalle I de $\mathbb{R}$ quelconque, calculer $\mathbb{P}(V \in I)$ en conditionnant sur les valeurs de ε.
3. Soit W une v.a.r telle que

1. Sachant $\{\varepsilon = +1\}$, W est uniformément distribuée sur $[0,1]$,
2. sachant $\{\varepsilon = -1\}$, W est uniformément distribuée sur $[-1,0]$,

3.a. Montrer que $W \sim \mathscr{U}_{[-1,1]}$.
Indication:Comme précédemment, calculer $\mathbb{P}(W \in I)$ en conditionnant sur les valeurs de ε, la difficulté est de traduire les « sachant », *i.e.* le concept de loi conditionnelle
3.b. Montrer que $|W| = \varepsilon.W \sim \mathscr{U}_{[0,1]}$ et que $|W|$ et ε sont indépendantes.

Lois conditionnelles, Simulation informatique

Les modèles probabilistes décrivant un problème physique sont spécifiés via la donnée de variables aléatoires fondamentales et les relations qu'elles entretiennent entre elles :

1. On peut imposer des relations fonctionnelles, par exemple $Z = f(X,Y)$;
2. on peut imposer de l'indépendance ;
3. on peut donner des lois conditionnelles.

C'est ce dernier point que l'on développe maintenant.

Donnons un exemple simple de mélange de populations. Une population de drosophiles est composée à $p\%$ d'individus ailés et $q\%$ d'individus possédant des ailes légèrement atrophiées et $r\%$ des ailes vestigiales, inutiles. On a $p+q+r=1$. On s'intéresse à leur durée de vie V.

1. Sachant qu'une drosophile est normalement ailée, sa durée de vie V (en heures) suit une loi uniforme [12] sur $[0,100]$;
2. Sachant qu'une drosophile a des ailes atrophiées, sa durée de vie V (en heures) suit une loi uniforme sur $[0,75]$;
3. Sachant qu'une drosophile a des ailes vestigiales, sa durée de vie V (en heures) suit une loi uniforme sur $[0,50]$;

On prend une drosophile au hasard (uniforme sur la population), quelle est la loi de la durée de vie ? Son espérance ? Comment simule-t-on une telle variable aléatoire ?

Commençons par la simulation informatique. On peut définir des variables aléatoires (informatique), `Vn()`, resp. `Va()`, resp. `Vv()`, retournant une durée de vie dans le cas normalement ailé, resp. atrophié, resp. vestigial. Ce sont des tirages uniformes simples.

Pour définir la variable aléatoire (informatique), `V(p)` retournant une durée de vie de drosophile pour notre population, il suffit de

1. Tirer au sort le type d'aile que la drosophile porte en respectant les paramètres p, q, r.
2. suivant les cas, normalement ailé, resp. atrophié, resp. vestigial, retourner la valeur calculée par `Vn()`,resp. `Va()`, resp. `Vv()`.

Listing 1 – python/droso-conditionnelles.py

```
import numpy as np

def Vn(): #Simule temps de vie, cas normal. ailé
    return 100*np.random.rand()

def Va(): #Simule temps de vie, cas atrophié
    return 75*np.random.rand()

def Vv(): #Simule temps de vie, cas vestigial
    return 50*np.random.rand()

global aile_n;aile_n=0 #définit les codes
global aile_a;aile_a=1
global aile_v;aile_v=2

def T(p,q): # retourne le type d'aile dans pop. de param. p,q
    u=np.random.rand()
    if u<=p: #Si celle tirée au sort est ailée
        return aile_n
    elif u<=p+q: #Si celle tirée au sort est atrophiée
        return aile_a
    return aile_v #sinon retourne temps pour vestigiale
```

12. Cette hypothèse n'est pas biologiquement raisonnable, l'exemple ne vaut que pour illustrer le propos

```
def V(p,q): #Simule temps de vie, cas d'une population de param. p,q
    t=T(p,q) #On tire le type d'aile
    #Sachant le type, on peut retourner une valeur tirée suivant loi adéquate
    if t==aile_n:
        return Vn() #temps pour normalement ailee
    if t==aile_a:
        return Va() #temps pour atrophiée
    if t==aile_v:
        return Vv() #temps pour vestigiale
```

Pour mener les calculs, introduisons les v.a. appropriées. Soit T la v.a donnant le type de la drosphile tirée au sort. T est à valeurs dans l'ensemble fini $\{\text{norm. ailé}, \text{atrophié}, \text{vestigial}\} = \{n, a, v\}$. On suppose T distribuée à l'aide des proportions dans la population, *i.e.*

$$\mathbb{P}(T = n) = p, \mathbb{P}(T = a) = q \text{ et } \mathbb{P}(T = v) = r$$

Concernant l'espérance, on peut mener un calcul basé sur le principe des espérances conditionnelles, en conditionnant sur la valeur de T

$$\begin{aligned} \mathbb{E}(V) &= \mathbb{E}(V|T = n)\mathbb{P}(T = n) + \mathbb{E}(V|T = a)\mathbb{P}(T = a) + \mathbb{E}(V|T = v)\mathbb{P}(T = v) \\ &= 100p + 75q + 50r \end{aligned}$$

Si on ne connait pas ce principe (HP), on peut tout de même s'en sortir en suivant ce que fait la simulation informatique. Soit U une v.a. uniforme sur $[0, 1]$, indépendante de T et G_n, G_a et G_v trois fonctions telles que $Vn = G_n(U)$, resp. $Va = G_a(U)$, resp. $Vv = G_v(U)$, a pour loi la loi conditionnelle de V sachant la drosophile normale ailée, resp. atrophiée, resp. vestigiale. On pose alors (définition d'une fonction sur Ω en trois morceaux)

$$V = \begin{cases} G_n(U) & \text{sur } \{T = n\} \\ G_a(U) & \text{sur } \{T = a\} \\ G_v(U) & \text{sur } \{T = v\} \end{cases}$$

Cette v.a.r V a les lois conditionnelles voulues. On a alors, en remarquant que

$$1\!\!1 = 1\!\!1_{\{T=n\}} + 1\!\!1_{\{T=a\}} + 1\!\!1_{\{T=v\}}$$

et que

$$1\!\!1_{\{T=n\}}.V = 1\!\!1_{\{T=n\}}.G_n(U),\ 1\!\!1_{\{T=a\}}.V = 1\!\!1_{\{T=a\}}.G_a(U) \text{ et } 1\!\!1_{\{T=v\}}.V = 1\!\!1_{\{T=v\}}.G_v(U)$$

que

$$\begin{aligned} \mathbb{E}(V) &= \mathbb{E}(1\!\!1.V) = \mathbb{E}(1\!\!1_{\{T=n\}}.V + 1\!\!1_{\{T=a\}}.V + 1\!\!1_{\{T=v\}}.V) \\ &= \mathbb{E}(1\!\!1_{\{T=n\}}.V) + \mathbb{E}(1\!\!1_{\{T=a\}}.V) + \mathbb{E}(1\!\!1_{\{T=v\}}.V) \\ &= \mathbb{E}(1\!\!1_{\{T=n\}}.G_n(U)) + \mathbb{E}(1\!\!1_{\{T=a\}}.G_n(U)) + \mathbb{E}(1\!\!1_{\{T=v\}}.G_n(U)) \end{aligned}$$

Comme, par indépendance de T et U,

$$\mathbb{E}(1\!\!1_{\{T=n\}}.G_n(U)) = \mathbb{E}(1\!\!1_{\{T=n\}}).\mathbb{E}(G_n(U)) = p.50$$

et similairement pour les autres termes de la somme, on a le résultat attendu.

Concernant la loi de V, le résultat sera plus facilement exprimable à l'issue du prochain chapitre mais on peut donner la formule de transfert générique pour V en effectuant un calcul similaire au précédent, pour une fonction $h : \mathbb{R} \to \mathbb{R}$ positive ou telle que $\mathbb{E}(h(V))$ admet une espérance, on a

$$\mathbb{E}(h(V)) = \int_{-\infty}^{+\infty} h(v).\delta_V(v)\, dv$$

où

$$\delta_V(v) = \begin{cases} \frac{p}{100} + \frac{q}{75} + \frac{r}{50} & \text{si } 0 \leq v \leq 50 \\ \frac{p}{100} + \frac{q}{75} & \text{si } 50 < v \leq 75 \\ \frac{p}{100} & \text{si } 75 < v \leq 100 \\ 0 & \text{sinon} \end{cases}$$

En effet, introduisons, comme dans le calcul de $\mathbb{E}(V)$ les variables T, U et les fonctions G_n, G_a et G_v. On a alors, en remarquant que

$$1\!\!1_{\{T=n\}}.h(V) = 1\!\!1_{\{T=n\}}.h(G_n(U)),\ 1\!\!1_{\{T=a\}}.h(V) = 1\!\!1_{\{T=a\}}.h(G_a(U)) \text{ et } 1\!\!1_{\{T=v\}}.h(V) = 1\!\!1_{\{T=v\}}.h(G_v(U))$$

que

$$\mathbb{E}(h(V)) \;=\; \mathbb{E}(1\!\!1_{\{T=n\}}.h(G_n(U))) + \mathbb{E}(1\!\!1_{\{T=a\}}.h(G_n(U))) + \mathbb{E}(1\!\!1_{\{T=v\}}.h(G_n(U)))$$

Comme, par indépendance de T et U,

$$\mathbb{E}(1\!\!1_{\{T=n\}}.h(G_n(U))) = \mathbb{E}(1\!\!1_{\{T=n\}}).\mathbb{E}(h(G_n(U))) = p.\frac{1}{100}\int_0^{100} h(v)\, dv$$

et similairement pour les autres termes de la somme,

$$\mathbb{E}(1\!\!1_{\{T=a\}}.h(G_a(U))) = q.\frac{1}{75}\int_0^{75} h(v)\, dv = q.\frac{1}{75}\int_0^{100} h(v).1\!\!1_{\{0\leq v\leq 75\}}\, dv$$

et

$$\mathbb{E}(1\!\!1_{\{T=v\}}.h(G_v(U))) = r.\frac{1}{50}\int_0^{50} h(v)\, dv = r.\frac{1}{50}\int_0^{100} h(v).1\!\!1_{\{0\leq v\leq 50\}}\, dv$$

En regroupant les termes, on obtient

$$\mathbb{E}(h(V)) = \int_0^{100} h(v)\left(\frac{p}{100} + \frac{q}{75}1\!\!1_{\{0\leq v\leq 75\}} + \frac{r}{50}1\!\!1_{\{0\leq v\leq 50\}}\right)\, dv$$

Ce qui peut se traduire facilement en le résultat attendu.