

VI. Probabilités

VI.1. Espace des épreuves, variables aléatoires, lois

On choisit en théorie des probabilités de considérer que 100% de chances se traduit par la valeur 1 de la probabilité. Quand un événement A est de probabilité 1, l'événement contraire est de probabilité $1 - 1 = 0$. On dit qu'un événement A de probabilité égale à 1 est *presque sûr*.

Une probabilité P est une mesure de masse 1 sur un ensemble Ω , qui est un ensemble assez riche pour représenter toutes les caractéristiques du phénomène aléatoire qu'on souhaite étudier ; les événements correspondent à des sous-ensembles de Ω ; du point de vue de la technique de la théorie de la mesure, on suppose que ces événements dont on pourra mesurer la probabilité forment une tribu \mathcal{F} de parties de Ω . Pour toute la suite, les *événements* sont les éléments $A \in \mathcal{F}$.

Définition. Un *espace de probabilité* est un espace mesuré (Ω, \mathcal{F}, P) , où P est une mesure de probabilité sur l'espace mesurable (Ω, \mathcal{F}) , c'est-à-dire une mesure qui vérifie la condition de normalisation $P(\Omega) = 1$.

Une variable aléatoire, en général réelle, représente le résultat (nombre réel) d'une certaine expérience aléatoire. Un exemple instructif (mais discutable) est fourni par la fonction « random » des langages de programmation.

Ce qui est discutable et délicat, tout à fait hors du programme de ce cours, est la possibilité qu'une machine déterministe comme un ordinateur puisse mimer un comportement aléatoire. La fonction `random` est le résultat fourni par un générateur *pseudo-aléatoire*. La discussion de la qualité d'un générateur pseudo-aléatoire est un sujet qu'on peut trouver passionnant, et en tout cas fondamental pour tout ce qui concerne la simulation aléatoire.

Ce problème de la réalisation *effective* d'un phénomène aléatoire est en dehors du champ couvert par les mathématiques. Le probabiliste est incapable de dire comment *produire* de l'aléatoire. On peut considérer qu'il se place au delà de cette question : si on lui fournit de l'aléatoire, il sera en mesure d'essayer de modéliser *la globalité* du phénomène, et ce modèle étant admis, de calculer les valeurs moyennes de certaines quantités liées au phénomène, par l'intermédiaire d'intégrales portant sur l'espace Ω de tous les tirages possibles muni de la mesure de probabilité P .

On attend de la fonction `random` qu'elle fournisse une valeur entre 0 et 1, et de telle sorte que pour tout intervalle $I \subset [0, 1]$ de longueur $1/N$ fixé, la « probabilité » que le résultat de la fonction `random` tombe dans I soit égale à $1/N$. On veut aussi que des appels successifs répétés à cette fonction `random` donnent des résultats « indépendants ». D'une certaine façon, c'est tout ce que l'utilisateur demande à la fonction `random`. On va considérer dans le cours de probabilité que le résultat est la valeur $X(\omega)$ d'une fonction X évaluée en un point $\omega \in \Omega$ choisi au hasard (mais encore une fois, le mathématicien ne sait pas faire *un* choix aléatoire qui serait issu des axiomes de la théorie des ensembles et désignerait ce point $\omega \in \Omega$ plutôt qu'un autre).

Définition. Une *variable aléatoire* réelle X sur l'espace de probabilité (Ω, \mathcal{F}, P) est une fonction \mathcal{F} -mesurable réelle sur Ω . La *loi* P_X de la variable X est la probabilité sur \mathbb{R} qui est la mesure image de la probabilité P par l'application mesurable X ,

$$\forall B \in \mathcal{B}_{\mathbb{R}}, \quad P_X(B) = P(X^{-1}(B)) = P(\{\omega \in \Omega : X(\omega) \in B\}) = P(X \in B).$$

La loi P_X est aussi une probabilité.

Ce que nous avons demandé à la fonction aléatoire **random**, considérée comme une variable aléatoire U définie sur un certain espace (Ω, \mathcal{F}, P) , c'est que U suive la loi uniforme sur $[0, 1]$: la probabilité P_U est égale à $\mathbf{1}_{[0,1]}(x) dx$, la mesure de Lebesgue sur $[0, 1]$,

$$\forall A \subset [0, 1], \quad A \in \mathcal{B}_{\mathbb{R}} \Rightarrow P(\{\omega \in \Omega : U(\omega) \in A\}) = \lambda(A),$$

λ étant la mesure de Lebesgue. On dit que U est une *variable uniforme* (sur $[0, 1]$).

On a vu qu'une mesure finie sur \mathbb{R} est complètement caractérisée par la connaissance de la mesure de tous les intervalles $]-\infty, t]$, ce qui justifie l'importance de la définition qui suit.

Définition. Si X est une variable aléatoire sur (Ω, \mathcal{F}, P) , sa *fonction de répartition* est la fonction F_X définie par

$$\forall t \in \mathbb{R}, \quad F_X(t) = P(X \leq t) = P_X(]-\infty, t]).$$

La fonction P_X caractérise complètement la loi P_X de la v.a. X (où v.a. est une abréviation pour « variable aléatoire »).

Proposition. La fonction de répartition F_X d'une v.a. X est croissante au sens large, continue à droite et

$$\lim_{t \rightarrow -\infty} F_X(t) = 0, \quad \lim_{t \rightarrow +\infty} F_X(t) = 1.$$

Preuve. — On a déjà prouvé la continuité à droite, essentiellement comme on prouve la limite en $-\infty$: si t_n tend en décroissant vers $-\infty$, l'ensemble $\{X \leq t_n\}$ décroît vers vide, et comme la mesure est finie, la régularité décroissante est valable,

$$\lim_n F_X(t_n) = \lim_n P(X \leq t_n) = P(\cap_n \{X \leq t_n\}) = P(\emptyset) = 0;$$

si t_n tend en croissant vers $+\infty$, les ensembles $\{X \leq t_n\}$ tendent en croissant vers Ω , et dans ce cas

$$\lim_n F_X(t_n) = \lim_n P(X \leq t_n) = P(\cup_n \{X \leq t_n\}) = P(\Omega) = 1.$$

Exemples de lois de probabilité et de fonctions de répartition.

— Si U est uniforme sur $[0, 1]$, on a $F_U(t) = t$ pour tout $t \in [0, 1]$; on a $F_U(t) = 0$ si $t < 0$ et $F_U(t) = 1$ si $t > 1$.

— La loi exponentielle de paramètre $\lambda > 0$ est $\lambda \mathbf{1}_{x \geq 0} e^{-\lambda x} dx$; elle intervient pour modéliser certains types de temps d'attente. La fonction de répartition est égale à $1 - e^{-\lambda t}$ quand $t \geq 0$, et égale à 0 pour tout $t < 0$. On a vu qu'on obtient la densité de la loi en dérivant la fonction de répartition.

— Une variable aléatoire gaussienne centrée réduite G suit la loi $(2\pi)^{-1/2} e^{-x^2/2} dx$. La fonction de répartition de la gaussienne,

$$F_G(t) = \int_{-\infty}^t e^{-x^2/2} \frac{dx}{\sqrt{2\pi}}$$

ne se calcule pas à partir des fonctions élémentaires déjà connues.

— Une variable nulle a pour loi la mesure de Dirac δ_0 en 0 ; la fonction de répartition vaut 1 pour $t \geq 0$, 0 sinon.

— Une variable de Bernoulli de paramètre p vaut 1 avec probabilité $p \in [0, 1]$, et 0 avec probabilité $1 - p$. La fonction de répartition présente deux sauts, un saut de hauteur $1 - p$ en 0 et un saut de hauteur p au point 1 ; la hauteur du saut est égale à la probabilité que la v.a. X prenne précisément cette valeur où le saut de F_X a lieu,

$$P(X = t_0) = P_X(\{t_0\}) = F_X(t_0) - \lim_{t \nearrow t_0} F_X(t).$$

Si U est uniforme et si I un intervalle de longueur p contenu dans $[0, 1]$, la variable aléatoire $\omega \rightarrow \mathbf{1}_{U(\omega) \in I}$ suit la loi de Bernoulli de paramètre p : on peut fabriquer des lois connues à partir d'autres lois connues. C'est l'un des thèmes de la *simulation*.

Exercice. Si X a une fonction de répartition F_X continue, alors la variable aléatoire

$$F_X(X) : \omega \in \Omega \rightarrow F_X(X(\omega))$$

suit la loi uniforme sur $[0, 1]$.

Solution. — La fonction F_X est continue, et $F_X(x)$ varie de 0 à 1 quand x décrit la droite réelle \mathbb{R} , donc F_X prend toutes les valeurs t telles que $0 < t < 1$; comme F_X est croissante et continue, l'ensemble $\{x : F_X(x) \leq t\}$ est un intervalle fermé $]-\infty, x(t)]$ et $F_X(x(t)) = t$. Ainsi,

$$(F_X(X(\omega)) \leq t) \Leftrightarrow (X(\omega) \leq x(t)),$$

donc si on considère la variable aléatoire $Y = F_X(X)$,

$$F_Y(t) = P(F_X(X) \leq t) = P(X \leq x(t)) = F_X(x(t)) = t.$$

Ainsi, la fonction de répartition de Y est celle des v.a. uniformes sur $[0, 1]$, d'où le résultat.

Espérance

Définition. L'espérance d'une variable aléatoire positive X sur (Ω, \mathcal{F}, P) est l'intégrale de X par rapport à P , elle est notée EX ,

$$(*) \quad EX = \int_{\Omega} X(\omega) dP(\omega).$$

Une variable aléatoire X est intégrable si $E|X| < +\infty$ et dans ce cas son espérance EX est donnée par la formule (*) précédente.

Calculs d'espérances à partir de la loi

Si f est une fonction borélienne sur \mathbb{R} et X une variable aléatoire sur (Ω, \mathcal{F}, P) on obtient par composition une fonction \mathcal{F} -mesurable $f(X) : \omega \rightarrow f(X(\omega))$, c'est-à-dire une nouvelle variable aléatoire $f(X)$, qui est une « fonction de la variable aléatoire X ».

Proposition. Si f est une fonction borélienne de \mathbb{R} dans $[0, +\infty]$, on a

$$E f(X) = \int_{\Omega} f(X(\omega)) dP(\omega) = \int_{\mathbb{R}} f(x) dP_X(x).$$

De plus, si f est réelle et $f(X)$ intégrable, l'espérance de $f(X)$ se calcule par la même formule.

Preuve. — Il suffit d'appliquer ce qui a été fait pour le calcul de l'intégrale par rapport à une mesure image.

Si f est l'application identique, $f(x) = x$ pour tout $x \in \mathbb{R}$, alors $f(X) = X$ et on trouve quand X est intégrable

$$E X = \int_{\mathbb{R}} x dP_X(x).$$

En appliquant avec $f(x) = x^2$, on trouve le moment d'ordre deux,

$$E X^2 = \int_{\mathbb{R}} x^2 dP_X(x).$$

Exemple. Il est clair intuitivement que l'espérance d'une variable U uniforme sur $[0, 1]$ est égale à $1/2$, mais on va le confirmer par le calcul,

$$E U = \int_{\mathbb{R}} x \mathbf{1}_{[0,1]}(x) dx = \int_0^1 x dx = \left[\frac{x^2}{2} \right]_0^1 = 1/2.$$

Pour une gaussienne centrée réduite G on trouve $E G = 0$ par parité, et

$$E |G| = \int_{\mathbb{R}} |x| e^{-x^2/2} \frac{dx}{\sqrt{2\pi}} = 2 \int_0^{+\infty} x e^{-x^2/2} \frac{dx}{\sqrt{2\pi}} = \sqrt{\frac{2}{\pi}}.$$

On avait calculé en exemple

$$E G^2 = \int_{\mathbb{R}} x^2 e^{-x^2/2} \frac{dx}{\sqrt{2\pi}} = 1.$$

Définition. La fonction caractéristique φ_X d'une variable aléatoire réelle X est la fonction à valeurs complexes définie sur \mathbb{R} par

$$\forall t \in \mathbb{R}, \quad \varphi_X(t) = E e^{itX}.$$

On verra que la loi F_X est caractérisée par la fonction caractéristique φ_X (injectivité de la transformation de Fourier).

Exemple. Pour une gaussienne centrée réduite G , on a d'après un calcul fait précédemment

$$\varphi_G(t) = \int_{\mathbb{R}} e^{itx} e^{-x^2/2} \frac{dx}{\sqrt{2\pi}} = e^{-t^2/2}.$$

Couple de v.a.

Si X, Y sont deux v.a. définies sur (Ω, \mathcal{F}, P) , on peut considérer l'application mesurable « couple »

$$\omega \in \Omega \rightarrow (X(\omega), Y(\omega)) \in \mathbb{R}^2$$

qui est mesurable à valeurs dans $(\mathbb{R}^2, \mathcal{B}_{\mathbb{R}} \otimes \mathcal{B}_{\mathbb{R}}) = (\mathbb{R}^2, \mathcal{B}_{\mathbb{R}^2})$. La *loi jointe* du couple (X, Y) est la probabilité $P_{(X,Y)}$ sur $(\mathbb{R}^2, \mathcal{B}_{\mathbb{R}} \otimes \mathcal{B}_{\mathbb{R}}) = (\mathbb{R}^2, \mathcal{B}_{\mathbb{R}^2})$ qui est la mesure image de la probabilité P par l'application couple (X, Y) ,

$$\forall B \in \mathcal{B}_{\mathbb{R}^2}, \quad P_{(X,Y)}(B) = P((X, Y) \in B).$$

VI.2. Indépendance

VI.2.1. Indépendance de deux tribus ou de v.a.

Définition. On suppose donné un espace de probabilité (Ω, \mathcal{F}, P) . Deux événements $A, B \in \mathcal{F}$ sont *indépendants* si $P(A \cap B) = P(A)P(B)$.

On vérifie qu'alors A^c et B , A et B^c , A^c et B^c sont également indépendants, autrement dit, tous les événements de la tribu (à quatre éléments $\emptyset, A, A^c, \Omega$) engendrée par A sont indépendants des événements engendrés par B .

Définition. Deux sous-tribus $\mathcal{A}, \mathcal{B} \subset \mathcal{F}$ sont indépendantes si tous les événements $A \in \mathcal{A}$ et $B \in \mathcal{B}$ sont indépendants.

Si X est une variable aléatoire sur Ω , la *sous-tribu engendrée par X* , notée $\sigma(X) \subset \mathcal{F}$ est la tribu $X^{-1}(\mathcal{B})$, image inverse par X de la tribu borélienne. Si X_1, \dots, X_n sont des v.a. sur Ω , la tribu $\sigma(X_1, \dots, X_n)$ engendrée par X_1, \dots, X_n est la plus petite tribu contenant toutes les tribus $\sigma(X_j)$, $j = 1, \dots, n$, ou de façon équivalente, contenant tous les ensembles $(X_j \in B_j)$, $j = 1, \dots, n$ et B_j borélien de \mathbb{R} .

Définition. Deux v.a. X et Y sont indépendantes si les tribus engendrées $\sigma(X)$ et $\sigma(Y)$ sont indépendantes.

Comme tous les événements de la tribu $\sigma(X)$ sont de la forme $\{X \in A\}$, et de même pour $\sigma(Y)$, cela revient à dire que

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

Exemple. Si $\Omega = [0, 1]^2$ et si P est la mesure de Lebesgue sur le carré, les variables aléatoires X et Y définies par $X(x, y) = x$ et $Y(x, y) = y$ sont indépendantes.

Proposition. Deux variables aléatoires réelles X, Y sur (Ω, \mathcal{F}, P) sont indépendantes si et seulement si la loi jointe est le produit tensoriel des lois,

$$P_{(X,Y)} = P_X \otimes P_Y.$$

Vérification. — Les deux v.a. sont indépendantes si et seulement si pour tous les sous-ensembles boréliens A, B de \mathbb{R} , on a

$$P_X(A)P_Y(B) = P(X \in A)P(Y \in B) = P((X \in A) \& (Y \in B)) = P_{(X,Y)}((X, Y) \in A \times B),$$

et l'égalité entre les termes extrêmes est la caractérisation de la mesure produit $P_X \otimes P_Y$.

Choix de la probabilité, statistiques

Si on considère un dé imparfait, mais dont on a des raisons de supposer qu'il possède un axe de rotation rendant quatre faces un peu rectangulaires, déduites les unes des autres par successions de quart de tours, équiprobables, et deux faces carrées déduites l'une de l'autre par une symétrie orthogonale par rapport au plan orthogonal à l'axe de rotation, elles deux équiprobables également, on peut envisager que la situation d'un lancer est décrite par un paramètre unique, la probabilité p de tomber sur une face carrée ; représentons par un 1 le fait de tomber sur une face carrée (et 0 figure les rectangles ; pour un cube idéal, $p = 1/3$). Si p est inconnu, et si on envisage d'estimer la valeur de p en réalisant N tirages successifs indépendants, l'espace Ω qui décrit l'expérience est assez clair, c'est $\Omega = \{0, 1\}^N$, mais il n'a pas de probabilité naturelle : on peut considérer toute la famille $(P_x)_{x \in [0,1]}$, dépendant du paramètre $x = p \in [0, 1]$, des lois sur Ω qui sont définies par la probabilité des singletons

$$P_x(\{\varepsilon_1, \dots, \varepsilon_N\}) = x^k (1-x)^{N-k}$$

où $k = \sum_{j=1}^N \varepsilon_j$ est le nombre de 1 dans la suite $\omega = (\varepsilon_1, \dots, \varepsilon_N) \in \Omega$. Un problème d'estimation statistique sera de chercher à proposer une valeur vraisemblable pour p , au vu d'une expérience effective de N lancers, et de proposer un « intervalle de confiance ».

Accessoirement, ce modèle des $(P_x)_{x \in [0,1]}$ est celui qui permet d'expliquer le mieux la preuve dite *de Bernstein* pour le théorème d'approximation polynomiale de Weierstrass.