

Image d'une probabilité, variables aléatoires

1 Fonctions mesurables

Quand en mathématiques une nouvelle structure est introduite, comme celle d'espace vectoriel, ou comme présentement celle d'espace de probabilité, une démarche féconde est de rechercher les transformations qui préservent cette structure. Pour les espaces vectoriels, ce sont les applications linéaires. Pour les espaces de probabilité, ce sont les "fonctions mesurables" qu'on va introduire dans un instant. Le cas particulier important en sera les "variables aléatoires". Auparavant, adoptons la notation suivante:

Définition si E et F sont des ensembles quelconques, si f est une fonction définie sur E et à valeurs dans F , et si enfin B est un sous ensemble de F , l'ensemble A des x de E tels que $f(x)$ soit dans B sera désormais noté par $A = f^{-1}(B)$. Nous l'appellerons l'**image inverse** de B par f .

Insistons sur le fait que f n'est pas nécessairement injective ni surjective. On vérifie facilement que:

Proposition Si B_1 et B_2 sont des sous ensembles de F alors on a

$$f^{-1}(B_1 \cup B_2) = f^{-1}(B_1) \cup f^{-1}(B_2) \text{ et } f^{-1}(B_1 \cap B_2) = f^{-1}(B_1) \cap f^{-1}(B_2).$$

La même propriété est vraie même avec une famille infinie de B .

Définition Soit alors deux espaces Ω et Ω_1 , chacun muni d'une tribu \mathcal{A} et \mathcal{A}_1 , et soit f une fonction définie sur Ω à valeurs dans Ω_1 . On dit que f est une **fonction mesurable** si pour tout $B \in \mathcal{A}_1$, alors $A = f^{-1}(B)$ est un élément de \mathcal{A} .

Dans ces conditions, on voit facilement que:

Proposition L'ensemble des parties A de la tribu Ω qui sont de la forme $f^{-1}(B)$, avec $B \in \mathcal{A}_1$, est une tribu. On la note parfois $f^{-1}(\mathcal{A}_1)$. Comme f est mesurable, c'est donc une sous tribu de \mathcal{A} .

Montrer qu'une fonction est mesurable est généralement facile grâce au théorème suivant, dont la démonstration est hors programme.

Théorème 4.1 Soit \mathcal{F} une famille de parties de Ω_1 telle que la tribu \mathcal{A}_1 soit la plus petite qui contienne \mathcal{F} . Soit f une fonction de Ω à valeurs dans Ω_1 . Soit \mathcal{A} une tribu sur Ω . Alors f est mesurable pour ce couple de tribus si et seulement si pour tout

$B \in \mathcal{F}$ alors $f^{-1}(B) \in \mathcal{A}$.

Illustrons ceci par un exemple important en l'appliquant au cas où $(\Omega, \mathcal{A}) = (\Omega_1, \mathcal{A}_1) = (\mathbf{R}, \mathcal{B})$, pour montrer que

Proposition *Toute fonction continue f de \mathbf{R} dans \mathbf{R} est mesurable.*

Démonstration Pour cela, on applique le théorème au cas où \mathcal{F} est l'ensemble de tous les intervalles ouverts: par définition de la tribu \mathcal{B} de Borel, l'hypothèse du théorème est vérifiée. Ensuite, on sait d'après le cours d'analyse que l'image inverse d'un intervalle ouvert par une fonction continue est une réunion finie ou dénombrable d'intervalles ouverts, et est donc un borélien.

Démonstration La partie "seulement si" découle des définitions. Pour la partie "si", l'art est de considérer la tribu \mathcal{T} de parties de Ω engendrée par tous les $f^{-1}(B)$ lorsque B parcourt \mathcal{F} ainsi que

$$\mathcal{T}_1 = \{B \subset \Omega_1; f^{-1}(B) \in \mathcal{T}\}.$$

A son tour, \mathcal{T}_1 est une tribu de parties de Ω_1 (ce point se vérifie directement facilement), et elle contient \mathcal{F} , et donc elle contient la tribu \mathcal{A}_1 . D'où

$$f^{-1}(\mathcal{T}_1) \supset f^{-1}(\mathcal{A}_1) = \mathcal{A}.$$

Mais comme par définition de \mathcal{T}_1 on a

$$f^{-1}(\mathcal{T}_1) \subset \mathcal{T},$$

on en tire que $\mathcal{T} = \mathcal{A}$, ce qui est l'égalité cherchée.

2 Image d'une probabilité.

Définition Si (Ω, \mathcal{A}) est muni d'une probabilité, alors la fonction mesurable f permet de définir de façon naturelle une probabilité P_1 sur $(\Omega_1, \mathcal{A}_1)$ ainsi: pour tout $B \in \mathcal{A}_1$

$$P_1(B) = P(f^{-1}(B)).$$

La probabilité P_1 ainsi fabriquée est appelée l'**image** de la probabilité P par la fonction mesurable f . On parle aussi de la probabilité P_1 **transportée** de P par f . On la note traditionnellement $P_1 = f_*P$. D'autres la notent plus correctement Pf^{-1} , mais c'est moins commode.

Cette fonction P_1 sur \mathcal{A}_1 est bien une probabilité. En effet,

$$P_1(\Omega_1) = P(f^{-1}(\Omega_1)) = P(\Omega) = 1;$$

De plus si B_1 et B_2 sont des parties disjointes de Ω_1 , alors $f^{-1}(B_1)$ et $f^{-1}(B_2)$ sont alors des parties disjointes de Ω . Cela permet de vérifier facilement l'axiome d'additivité dénombrable pour P_1 .

3 Les variables aléatoires réelles et leurs lois.

Nous appliquons les concepts précédents, qui étaient bien abstraits, au cas où l'espace d'arrivée $(\Omega_1, \mathcal{A}_1)$ est $(\mathbb{R}, \mathcal{B})$. Dans ce cadre, une fonction mesurable de Ω à valeurs dans \mathbb{R} prend le nom de *variable aléatoire réelle*, ou de *variable aléatoire* si le contexte est clair (on pourra ensuite considérer des variables aléatoires à valeurs dans \mathbb{R} ou dans \mathbb{R}^n quand on aura précisé de quelle tribu équiper \mathbb{R}^n).

Définition Une **variable aléatoire réelle** est une fonction mesurable d'une tribu (Ω, \mathcal{A}) dans la tribu $(\mathbb{R}, \mathcal{B})$ où \mathcal{B} est l'ensemble des boréliens de \mathbb{R} .

Plutôt que de noter la variable aléatoire f , la tradition est de la noter par une lettre majuscule comme X . En dépit du nom de "variable aléatoire," qu'on garde pour des raisons historiques, X est donc une fonction réelle définie sur Ω . L'avantage de travailler dans \mathbb{R} est que grâce au Théorème 2.12, on sait comment sont faites les probabilités sur \mathbb{R} et donc les probabilités transportées par les variables aléatoires. On abandonne d'ailleurs également pour $P_1 = X_*P$ ce nom de probabilité transportée de P par la variable aléatoire X , on la note plutôt P_X et on l'appelle la *loi de la variable aléatoire* X : c'est une probabilité sur \mathbb{R} .

Définition Si X est une variable aléatoire réelle définie sur un espace probabilisé (Ω, \mathcal{A}, P) , l'application P_X définie de l'ensemble des boréliens de \mathcal{B} dans $[0, 1]$ par $P_X(B) = P(X^{-1}(B))$ est une probabilité sur \mathbb{R} appelé **loi de la variable aléatoire** X .

Quant à la fonction de répartition F_{P_X} , il est plus simple de la noter F_X . Donc on a

$$F_X(x) = P(\{\omega \in \Omega ; X(\omega) \leq x\});$$

ici encore, il est plus simple d'écrire $F_X(x) = P(X \leq x)$.

Définition Si X est une variable aléatoire réelle, la fonction F_X définie sur \mathbb{R} par

$$F_X(x) = P(\{\omega \in \Omega ; X(\omega) \leq x\})$$

est la **fonction de répartition** de la variable aléatoire X .

Enfin, v.a. est une abréviation courante pour "variable aléatoire".

A propos du schéma Succès Echec fini d'ordre N , nous avons déjà rencontré la variable aléatoire X qui était le nombre de succès en N expériences pour laquelle nous

avons vu que $P(X = k) = C_N^k p^k (1-p)^{N-k}$. C'est donc dire que la loi de X est la loi discrète concentrée sur les entiers $0, 1, \dots, N$ et égale à

$$(1-p)^N \delta_0 + N(1-p)^{N-1} p \delta_1 + \dots + C_N^k p^k (1-p)^{N-k} \delta_k + \dots + p^N \delta_N$$

(Rappelons que δ_k est la probabilité de Dirac concentrée en k).

Plus généralement:

Définition Soit (Ω, \mathcal{A}, P) un espace de probabilité. Une variable aléatoire X sur Ω ne prenant qu'un nombre fini de valeurs $a_1 < a_2 < \dots < a_N$ sera dite **étagée**.

Les parties $X^{-1}(\{a_j\}) = A_j$ de Ω sont des éléments de \mathcal{A} , puisque les $\{a_j\}$ sont des intervalles, d'un type un peu particulier, et donc des boréliens. Les A_j sont deux à deux disjoints, et si on introduit leurs indicateurs, on peut écrire

$$X = a_1 \mathbf{1}_{A_1} + \dots + a_N \mathbf{1}_{A_N}.$$

Si $p_j = P(A_j)$ on voit que la loi de X est

$$P_X = p_1 \delta_{a_1} + \dots + p_N \delta_{a_N}.$$

Une autre manière de dire la même chose est d'écrire $P(X = a_j) = p_j$ pour tout j .

Il y a un certain nombre de lois de probabilités qu'on rencontre souvent dans la nature que nous pourrions présenter maintenant, mais il est préférable de définir quelques caractéristiques des variables aléatoires avant pour pouvoir présenter une carte d'identité plus complète de chacune de ces lois classiques.

Exercices sur la section 4.

1. Soit X une variable aléatoire de densité $ax^{a-1} \mathbf{1}_{]0,1[}(x)$. Calculer l'image de sa loi par $x \mapsto x/(1-x)$. Méthode: calculer la fonction de répartition de $Y = X/(1-X)$ et dériver celle-ci.
2. Soit X une variable aléatoire suivant la loi de Cauchy, c'est-à-dire de densité $\frac{1}{\pi(1+x^2)}$. Calculer l'image de sa loi par $x \mapsto 1/x$.