Variables aléatoires discrètes

0. Introduction

Lorsque l'on étudie un phénomène on est amené à étudier des grandeurs numériques ou vectorielles liées à ce dernier. Intuitivement, une variable aléatoire représente un nombre (ou un vecteur) lié à une expérience aléatoire, dont la valeur dépend exclusivement du résultat ω de cette expérience. Il est d'ailleurs fréquent qu'une quantité liée au résultat de l'expérience soit plus importante aux yeux de l'observateur que le résultat de l'expérience lui-même. Au vu de ce qui vient d'être dit, il semble naturel d'introduire la notion de variable aléatoire via la notion d'application de Ω dans E.

1. Variables aléatoires discrètes

Soit $(\Omega,\mathcal{A},\mathbb{P})$ un espace probabilisé et E est un ensemble.

Définition.– Une variable aléatoire discrète (VAD) sur $(\Omega,\mathcal{A})$ à valeurs dans E est une application $X:\Omega\to E$ qui vérifie les propriétés:

1) $X(\Omega)$ est fini ou dénombrable.
2) $\forall A\in\mathcal{P}(E),\ X^{-1}(A)\in\mathcal{A}$.

L'ensemble des VAD sur $(\Omega,\mathcal{A})$ à valeurs dans E est noté $V_E(\Omega,\mathcal{A})$.

Théorème.– Une application $X:\Omega\to E$ est une VAD sur $(\Omega,\mathcal{A})$ si et seulement si :

1) $X(\Omega)$ est fini ou dénombrable.
2) $\forall x\in E,\ X^{-1}(\{x\})\in\mathcal{A}$.

Proposition.–

1) Si $A\in\mathcal{A}$ alors $\mathbf{1}_A$ est une VAD sur $(\Omega,\mathcal{A})$ à valeurs réelles.
2) Toute application constante de Ω dans E est une VAD sur $(\Omega,\mathcal{A})$ à valeurs dans E.

<u>Notations</u> : Soit X une VAD sur $(\Omega,\mathcal{A})$ à valeurs dans E, $A\in\mathcal{P}(E)$ et $x\in E$.

L'ensemble $X^{-1}(A)=\{\omega\in\Omega\mid X(\omega)\in A\}$ est noté $\{X\in A\}$ ou $(X\in A)$ ou encore $X\in A$.

L'ensemble $X^{-1}(\{x\})=\{\omega\in\Omega\mid X(\omega)=x\}$ est noté $\{X=x\}$ ou $(X=x)$ ou encore $X=x$.

Si de plus X est à valeurs <u>réelles</u> alors x est un nombre réel et on adopte les notations suivantes :

L'ensemble $X^{-1}(]-\infty,x])=\{\omega\in\Omega\mid X(\omega)\le x\}$ est noté $\{X\le x\}$ ou $(X\le x)$ ou $X\le x$.

L'ensemble $X^{-1}(]-\infty,x[)=\{\omega\in\Omega\mid X(\omega)< x\}$ est noté $\{X< x\}$ ou $(X< x)$ ou $X< x$.

L'ensemble $X^{-1}([x,+\infty[)=\{\omega\in\Omega\mid X(\omega)\ge x\}$ est noté $\{X\ge x\}$ ou $(X\ge x)$ ou $X\ge x$.

L'ensemble $X^{-1}(]x,+\infty[)=\{\omega\in\Omega\mid X(\omega)> x\}$ est noté $\{X> x\}$ ou $(X> x)$ ou $X> x$.

Proposition.– Soient $X:\Omega\to E$ une VAD à valeurs dans E, $A\in\mathcal{P}(E)$ et $y\in E$.

1) Si $y\notin X(\Omega)$ alors $(X=y)=\varnothing$ et $\mathbb{P}(X=y)=0$
2) $\mathbb{P}(X\in A)=\sum_{x\in A}\mathbb{P}(X=x)=\sum_{x\in A\cap X(\Omega)}\mathbb{P}(X=x)$.
3) $\sum_{x\in E}\mathbb{P}(X=x)=\sum_{x\in X(\Omega)}\mathbb{P}(X=x)=1$.

Proposition.–

1) Si $X:\Omega\to E$ est une VAD sur $(\Omega,\mathcal{A})$ à valeurs dans E et si $f:E\to F$ est une application de E dans l'ensemble F alors $f\circ X$ est une VAD sur $(\Omega,\mathcal{A})$ à valeurs dans F.
2) Si pour tout $k\in[\![1,p]\!]$, $X_k:\Omega\to E_k$ est une VAD sur $(\Omega,\mathcal{A})$ à valeurs dans l'ensemble E_k alors $(X_1,\cdots,X_p)$ est une VAD sur $(\Omega,\mathcal{A})$ à valeurs dans le produit cartésien $E_1\times\cdots\times E_p$.
3) Si pour $k\in[\![1,p]\!]$, $X_k:\Omega\to E_k$ est une VAD sur $(\Omega,\mathcal{A})$ à valeurs dans E_k et si $f:E_1\times\cdots\times E_p\to F$ est une application alors $f\circ(X_1,\cdots,X_p)$ est une VAD sur $(\Omega,\mathcal{A})$ à valeurs dans F.

<u>Remarques</u>

✓ L'application $f\circ X$ est notée $f(X)$. L'application $f\circ(X_1,\cdots,X_p)$ est notée $f(X_1,\cdots,X_p)$.

✓ Si $X:\Omega\to E$ est une VAD réelle sur $(\Omega,\mathcal{A})$ alors X^2, $\sin X^2$ et e^X sont des VAD réelles sur $(\Omega,\mathcal{A})$.

Définition.– Un vecteur aléatoire discret X sur $(\Omega,\mathcal{A})$ est une VAD sur $(\Omega,\mathcal{A})$ de la forme $X=(X_1,\cdots,X_p)$ où pour tout $k\in[\![1,p]\!]$, X_k est une VAD sur $(\Omega,\mathcal{A})$ à valeurs dans un ensemble E_k. La VAD X_k est appelée $k^{\text{ième}}$ marginale du vecteur aléatoire X.

<u>Remarque</u> : Soient $(A_1,\cdots,A_p)\in\mathcal{P}(E_1)\times\cdots\times\mathcal{P}(E_p)$ et $(x_1,\cdots,x_p)\in E_1\times\cdots\times E_p$.

$((X_1,\cdots,X_p)\in A_1\times\cdots\times A_p)=(X_1\in A_1)\cap\cdots\cap(X_p\in A_p)$.

$((X_1,\cdots,X_p)=(x_1,\cdots,x_p))=(X_1=x_1)\cap\cdots\cap(X_p=x_p)$.

3.2 Détermination de la loi d'une variable aléatoire discrète

Théorème.– Soient $X : \Omega \to E$ une VAD sur $(\Omega, \mathcal{A})$.

$\forall A \in \mathcal{P}(E)$, $\mathbb{P}_X(A) = \sum_{x \in A} \mathbb{P}(X = x) = \sum_{x \in A \cap X(\Omega)} \mathbb{P}(X = x)$.

La loi $\mathbb{P}_X$ de X est donc parfaitement déterminée par la donnée des $\mathbb{P}(X = x)$, $x \in X(\Omega)$.

Théorème.– Soient $X : \Omega \to E$ une VAD sur $(\Omega, \mathcal{A})$ à valeurs dans E et $f : E \to F$ une application.

1) $f(X)$ est une VAD sur $(\Omega, \mathcal{A})$ à valeurs dans F.
2) $\forall B \in \mathcal{P}(F)$, $\mathbb{P}(f(X) \in B) = \sum_{x \in f^{-1}(B)} \mathbb{P}(X = x)$.
3) $\forall y \in F$, $\mathbb{P}(f(X) = y) = \sum_{x \in f^{-1}(\{y\})} \mathbb{P}(X = x)$.

Proposition.– Soient X, Y des VAD sur $(\Omega, \mathcal{A})$ à valeurs respectivement dans E et F.

1) (X, Y) est une variable aléatoire discrète sur $(\Omega, \mathcal{A})$ à valeurs dans $E \times F$.
2) $\forall x \in E$, $\mathbb{P}(X = x) = \sum_{y \in F} \mathbb{P}((X, Y) = (x, y)) = \sum_{y \in Y(\Omega)} \mathbb{P}((X, Y) = (x, y))$.
3) $\forall y \in F$, $\mathbb{P}(Y = y) = \sum_{x \in E} \mathbb{P}((X, Y) = (x, y)) = \sum_{x \in X(\Omega)} \mathbb{P}((X, Y) = (x, y))$.

Définition.– Soient X, Y des VAD sur $(\Omega, \mathcal{A})$ à valeurs respectivement dans E et F.

1) La loi $\mathbb{P}_{(X,Y)}$ du vecteur aléatoire (X, Y) est une probabilité sur $(E \times F, \mathcal{P}(E \times F))$.
   Elle est appelée loi conjointe du vecteur aléatoire (X, Y).
2) La loi $\mathbb{P}_X$ de la variable aléatoire X est une probabilité sur $(E, \mathcal{P}(E))$.
   Elle est appelée première loi marginale du vecteur aléatoire (X, Y).
3) La loi $\mathbb{P}_Y$ de la variable aléatoire Y est une probabilité sur $(F, \mathcal{P}(F))$.
   Elle est appelée seconde loi marginale du vecteur aléatoire (X, Y).

Remarque :

Les résultats sont encore valables dans le cas de p VAD $X_1, \cdots, X_p$ sur $(\Omega, \mathcal{P}(\Omega))$ à valeurs respectivement dans les ensembles $E_1, \cdots, E_p$. Pour $k \in [\![1, p]\!]$ et $x_k \in E_k$ on a :

$$\mathbb{P}(X = x_k) = \sum_{(x_1, \cdots, x_{k-1}, x_{k+1}, \cdots, x_p) \in E_1 \times \cdots \times E_{k-1} \times E_{k+1} \times \cdots \times E_p} \mathbb{P}((X_1, \cdots, X_p) = (x_1, \cdots x_p)).$$

3.3 Loi conditionnelle

Théorème.– Soient $X : \Omega \to E$ et $Y : \Omega \to F$ des VAD sur $(\Omega, \mathcal{A})$ à valeurs dans E et F.

On considère $A \in \mathcal{P}(E)$ tel que $\mathbb{P}(X \in A) > 0$ et on lui associe l'application $\mathbb{P}_Y^{(X \in A)} : \mathcal{P}(F) \to [0,1]$ définie par : $\forall B \in \mathcal{P}(F)$, $\mathbb{P}_Y^{(X \in A)}(B) = \mathbb{P}((Y \in B) \mid (X \in A))$.

$\mathbb{P}_Y^{(X \in A)}$ est une probabilité sur $(F, \mathcal{P}(F))$ et est appelée loi conditionnelle de Y sachant $X \in A$.

4. Quelques situations fondamentales

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé.

4.1 Variables aléatoires de Bernoulli

Définition.– Une variable aléatoire de Bernoulli de paramètre $p \in [0,1]$ sur $(\Omega, \mathcal{A})$ est une variable aléatoire discrète $X : \Omega \to \mathbb{N}$ à valeurs entières vérifiant : $\mathbb{P}(X = 0) = 1 - p$ et $\mathbb{P}(X = 1) = p$.

Remarque : Les variables aléatoires de Bernoulli interviennent dans les expériences aléatoires ne comportant que deux issues : « échec » et « succès ». Une telle expérience aléatoire est appelée expérience de Bernoulli. L'issue « échec » est souvent codée par 0 et l'issue « succès » par 1.

Proposition.–

1) Si A est un évènement de $\mathcal{A}$ alors $\mathbf{1}_A$ est une variable aléatoire de Bernoulli de paramètre $p = \mathbb{P}(A)$.
2) Si X est une variable aléatoire de Bernoulli de paramètre p sur $(\Omega, \mathcal{A})$ alors $X \rightsquigarrow \mathcal{B}(p)$ et X est égale presque sûrement à $\mathbf{1}_{(X=1)}$.

4.2 Répétition de n expériences de Bernouilli mutuellement indépendantes

Théorème.– Soient $X_1, \cdots, X_n : \Omega \to \mathbb{N}$ des VAD à valeurs dans $\mathbb{N}$ mutuellement indépendantes et suivant toutes une loi de Bernouilli de paramètre $p \in [0,1]$. Dans ces conditions :

$$S_n = \sum_{k=1}^{n} X_k \rightsquigarrow \mathcal{B}(n, p).$$

Corollaire1.– Lors de la répétition de n expériences de Bernouilli mutuellement indépendantes et de même paramètre p, le nombre de succès obtenus suit une loi binomiale de paramètres n et p.

Corollaire2.– Si $A_1, \cdots, A_n$ sont des évènements de $\mathcal{A}$ mutuellement indépendants pour $\mathbb{P}$ et de même probabilité $p \in [0,1]$ alors $S_n = \sum_{k=1}^{n} 1_{A_k} \rightsquigarrow \mathcal{B}(n,p)$.

4.3 Répétition d'une infinité d'expériences de Bernouilli indépendantes

Théorème.– Soit $(X_n)_{n\geq 1}$ est une suite de VAD à valeurs dans $\mathbb{N}$ mutuellement indépendantes et suivant toutes une loi de Bernouilli de paramètre $p \in]0,1]$.

Soit $T : \Omega \to \mathbb{N}$ la VAD à valeurs dans $\mathbb{N}$ définie de la façon suivante : $T(\omega) = \min\{k \in \mathbb{N}^* \mid X_k(\omega) = 1\}$ si $\omega \in \bigcup_{n\in\mathbb{N}^*}(X_n = 1)$ et $T(\omega) = 0$ si $\omega \notin \bigcup_{n\in\mathbb{N}^*}(X_n = 1)$.

Dans ces conditions : $T \rightsquigarrow \mathcal{G}(p)$.

Corollaire1.– Lors de la répétition d'une infinité d'expériences de Bernouilli mutuellement indépendantes et de même paramètre p, le rang du premier succès obtenu suit une loi géométrique de paramètre p.

Corollaire2.– Soit $(A_n)_{n\geq 1}$ est une suite d'évènements de $\mathcal{A}$ mutuellement indépendants pour $\mathbb{P}$ et de même probabilité $p \in [0,1]$. Soit $T : \Omega \to \mathbb{N}$ la variable aléatoire définie de la façon suivante : $T(\omega) = \min\{k \in \mathbb{N}^* \mid \omega \in A_k\}$ si $\omega \in \bigcup_{n\in\mathbb{N}^*} A_n$ et $T(\omega) = 0$ si $\omega \notin \bigcup_{n\in\mathbb{N}^*} A_n$. Alors : $T \rightsquigarrow \mathcal{G}(p)$.

4.4 Caractérisation de la loi géométrique

Proposition.– Soit $X : \Omega \to \mathbb{N}$ une variable aléatoire discrète à valeurs dans $\mathbb{N}$ telle que $\mathbb{P}(X = 0) = 0$. La VAD X suit une loi géométrique de paramètre p pour un certain $p \in]0,1]$ si et seulement si : $\forall n \in \mathbb{N}, \forall k \in \mathbb{N}, \ \mathbb{P}(X > n + k \mid X > n) = \mathbb{P}(X > k)$.

4.5 Théorème de Poisson

Théorème.– (Poisson)

Soient $\lambda \in \mathbb{R}^{*+}$ et (p_n) une suite d'éléments de $[0,1]$ telle que $\lim np_n = \lambda$.

Soit (X_n) une suite de VAD sur $(\Omega, \mathcal{A})$ à valeurs dans $\mathbb{N}$ telle que pour tout $n \in \mathbb{N}$, $X_n \rightsquigarrow \mathcal{B}(n, p_n)$.

Dans ces conditions : $\forall k \in \mathbb{N}, \ \lim\limits_{n\to+\infty} \mathbb{P}(X_n = k) = e^{-\lambda}\dfrac{\lambda^k}{k!}$.

<u>Remarque</u> :Si $\lim np_n = \lambda$ alors la probabilité binomiale de paramètre n et p_n peut être approchée, pour n suffisamment grand, par une probabilité de Poisson de paramètre λ.

5. Pour aller plus loin (HP)

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé.

5.1 Minimum et maximum de deux variables aléatoires indépendantes

Proposition.– Soient $X, Y : \Omega \to \mathbb{N}$ des variables aléatoires discrètes et <u>indépendantes</u> sur $(\Omega, \mathcal{A}, \mathbb{P})$.

Si $X \rightsquigarrow \mathcal{G}(p)$ et si $Y \rightsquigarrow \mathcal{G}(p)$ alors $\min(X,Y) \rightsquigarrow \mathcal{G}\left(1 - (1-p)^2\right)$.

5.2 Somme de variables aléatoires indépendantes

Proposition.– Si $X, Y : \Omega \to \mathbb{R}$ sont des VAD <u>indépendantes</u> sur $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs réelles alors : $\forall x \in \mathbb{R}, \ \mathbb{P}(X+Y = x) = \sum_{t\in\mathbb{R}} \mathbb{P}(X = t)\mathbb{P}(Y = x - t)$.

Proposition.– Soient $X, Y : \Omega \to \mathbb{N}$ des VAD à valeurs dans $\mathbb{N}$ <u>indépendantes</u> sur $(\Omega, \mathcal{A}, \mathbb{P})$.

1) Si $X \rightsquigarrow \mathcal{B}(n_1, p)$ et si $Y \rightsquigarrow \mathcal{B}(n_2, p)$ alors $X + Y \rightsquigarrow \mathcal{B}(n_1 + n_2, p)$.
2) Si $X \rightsquigarrow \mathcal{P}(\lambda_1)$ et si $Y \rightsquigarrow \mathcal{P}(\lambda_2)$ alors $X + Y \rightsquigarrow \mathcal{P}(\lambda_1 + \lambda_2)$.

Proposition.– Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires discrètes sur $(\Omega, \mathcal{A})$ à valeurs dans $\mathbb{N}$. On suppose que les $X_n, n \in \mathbb{N}^*$ sont mutuellement indépendantes et que pour tout $n \in \mathbb{N}^*$, $X_n \rightsquigarrow \mathcal{G}(p)$ La loi de la variable aléatoire $S_n = X_1 + \cdots + X_n$ est donnée par les égalités : $\mathbb{P}(S_n = k) = 0$ si $k \in [0, n-1]$ et $\mathbb{P}(S_n = k) = \binom{k-1}{n-1} p^n (1-p)^{k-n}$ si $k \geq n$.

La loi de S_n est appelée loi binomiale négative de paramètres n et p.

5.3 Echantillonnage et loi hypergéométrique

Le problème typique de l'échantillonnage est le suivant. On souhaite connaitre les intentions de vote d'une population comportant m individus. Cette population se compose de m_a personnes qui souhaitent voter pour le candidat A et de m_b personnes souhaitant voter pour le candidat B avec $m_a + m_b = m$.

On effectue un sondage sur n personnes tirées « au hasard » et on se demande quelle est la loi suivie par le nombre de personnes déclarant vouloir voter pour A.

La formulation la plus neutre de ce problème est celle de l'urne. On considère une urne contenant m_a boules blanches et m_b boules noires et on tire n boules dans cet urne. Le tirage peut se faire « avec remise » ou « sans remise ».

Proposition.– On considère des entiers naturels m_a, m_b et n tels que $n \leq m_a$ et $n \leq m_b$ et une urne U contenant m_a boules blanches et m_b boules noires.

1) Lors d'un tirage avec remise de n boules dans l'urne U, le nombre total de boules blanches obtenues suit une loi binomiale de paramètres n et p avec $p = \dfrac{m_a}{m_a + m_b}$.
2) Lors d'un tirage sans remise de n boules dans l'urne U, le nombre total de boules blanches obtenues suit une loi hypergéométrique de paramètres m_a, m_b et n.

Jacob Bernoulli (1654-1705)

Jacob Bernoulli est l'ainé de la famille des Bernoulli, famille de mathématiciens suisse d'origine belge. Il commence par étudier la théologie et la philosophie mais assez rapidement ses centres d'intérêt le portent vers l'astronomie et les mathématiques. Après ses études il entreprend de nombreux voyages en France, en Angleterre et aux Pays-Bas et en profite pour échanger avec de nombreux mathématiciens. De retour à Bâle, il devient professeur à l'université et occupera ce poste jusqu'à son décès. Son frère est aussi mathématicien. Le deux frères collaborent dans un premier temps mais se fâchent vite, se disputant la priorité des découvertes. Jacob Bernouilli s'est principalement intéressé au calcul différentiel, aux séries, au calcul des probabilités et à l'étude de certaines courbes. On lui doit, entre autre, la loi faible des grands nombres pour le jeu de pile ou face (ce que l'on appelle aujourd'hui le schéma de Bernoulli), une preuve rigoureuse de la convergence de la série $\sum \frac{1}{n^2}$ et la résolution de certaines équations différentielles liées aux courbes que sont le lemniscate (de Bernoulli !), la cycloïde et la spirale logarithmique. On lui doit aussi les nombres et les polynômes qui portent son nom. Pour l'anecdote, Jacob Bernoulli aimait particulièrement la spirale logarithmique et ses propriétés d'invariance. A tel point qu'il demanda à ce que l'on en grave une sur sa tombe. Le graveur, piètre mathématicien se trompa et dessina une spirale d'Archimède...

Denis Poisson (1781-1840)

Mathématicien français, Denis Poisson est né à Pithiviers. Son père qui souhaite l'aider à obtenir une bonne situation sociale le dirige vers la médecine. Peu intéressé par une carrière médicale, Poisson s'en détourne assez vite pour finalement entré à l'école Polytechnique en 1798. Il s'y révèle un étudiant brillant et écrit deux mémoires importants, l'un sur la méthode d'élimination de Bezout et l'autre sur les équations aux différences finies. Recommandé par Laplace il devient très vite répétiteur puis professeur à l'école Polytechnique en remplacement de Fourier. Poisson a publié près de quatre cents articles. Il étudie le mouvement du pendule, l'effet de petites perturbations sur les mouvements planétaires, met au point les lois de l'électrostatique, est à l'origine des premiers travaux sur les séries de Fourier et écrit en 1837 un important mémoire sur les probabilités. Ce traité ne fut que peu remarqué par les contemporains de Poisson mais eut une grande influence par la suite.

Astronome au bureau des longitudes, membre de l'académie des sciences, Professeur à l'école Polytechnique et Professeur de mécanique à la faculté des sciences, ce bourreau de travail qu'était Poisson a déclaré un jour : « la vie n'est bonne qu'à deux choses : découvrir et enseigner les mathématiques ».

Moments d'une variable aléatoire discrète réelle

Dans tout le chapitre $(\Omega, \mathcal{A}, \mathbb{P})$ est un espace probabilisé.

1. Espérance d'une variable aléatoire discrète

1.1 Le cas des variables aléatoires à valeurs positives

Définition.– Si $X : \Omega \to \overline{\mathbb{R}^+}$ est une VAD sur $(\Omega, \mathcal{A})$ à valeurs positives alors l'espérance $\mathbb{E}(X)$ de X est l'élément de $[0, \overline{+\infty}]$ défini par : $\mathbb{E}(X) = \sum_{t \in \mathbb{R}^+} t\mathbb{P}(X = t)$.

Théorème.– (de transfert cas positif)

Si $X : \Omega \to E$ est une VAD sur $(\Omega, \mathcal{A})$ à valeurs dans E et si $f : E \to \mathbb{R}^+$ est une application à valeurs positives alors la VAD $f(X)$ est à valeurs positives et $\mathbb{E}(f(X)) = \sum_{x \in E} f(x)\mathbb{P}(X = x)$.

Proposition.– Si $X : \Omega \to \mathbb{R}$ est une VAD sur $(\Omega, \mathcal{A})$ à valeurs réelles alors $\mathbb{E}(|X|) = \sum_{t \in \mathbb{R}} |t|\mathbb{P}(X = t)$.

Proposition.– Soient $X, Y : \Omega \to \overline{\mathbb{R}^+}$ des VAD sur $(\Omega, \mathcal{A})$ à valeurs positives.

1) Si X et Y ont la même loi $(X \sim Y)$ alors $\mathbb{E}(X) = \mathbb{E}(Y)$.
2) Si $(\alpha, \beta) \in \left(\overline{\mathbb{R}^+}\right)^2$ alors $\mathbb{E}(\alpha X + \beta Y) = \alpha\mathbb{E}(X) + \beta\mathbb{E}(Y)$.
3) Si $0 \le X \le Y$ alors $0 \le \mathbb{E}(X) \le \mathbb{E}(Y) \le +\infty$.

1.2 Le cas des variables aléatoires à valeurs réelles d'espérance finie

Définition.– Une VAD <u>réelle</u> X sur $(\Omega, \mathcal{A})$ est dite d'espérance finie si et seulement si $\mathbb{E}(|X|) < +\infty$.

L'ensemble des VAD réelles sur $(\Omega, \mathcal{A})$ qui sont d'espérance finie est noté $L^1_d(\Omega, \mathcal{A}, P)$. On a donc : $L^1_d(\Omega, \mathcal{A}, P) = \{X \in V_{\mathbb{R}}(\Omega, \mathcal{A}) \mid \mathbb{E}(|X|) < +\infty\}$.

L'espérance de $X \in L^1_d(\Omega, \mathcal{A}, P)$ est le réel $\mathbb{E}(X)$ défini par : $\mathbb{E}(X) = \sum_{t \in \mathbb{R}} t\mathbb{P}(X = t)$.

<u>Remarque</u> :

Si l'espérance d'une variable aléatoire discrète à valeurs dans $\mathbb{R}^+$ est toujours définie il n'en est pas de même de l'espérance d'une variable aléatoire réelle X qui elle n'est définie que si $\mathbb{E}(|X|) < +\infty$.

1.3 Espace $L^1_d(\Omega, \mathcal{A}, \mathbb{P})$

Théorème.– (de transfert)

Soient $X : \Omega \to E$ une VAD sur $(\Omega, \mathcal{A})$ à valeurs dans E et $f : E \to \mathbb{R}$ une application.

1) $f(X) \in L^d_1(\Omega, \mathcal{A}, \mathbb{P}) \Leftrightarrow \sum_{x \in E} |f(x)|\mathbb{P}(X = x) < +\infty$.
2) Si $f(X) \in L^d_1(\Omega, \mathcal{A}, \mathbb{P})$ alors $E(f(X)) = \sum_{x \in E} f(x)\mathbb{P}(X = x)$.

Théorème.–

1) $L^1_d(\Omega, \mathcal{A}, \mathbb{P})$ est un $\mathbb{R}$-espace vectoriel et l'espérance est une forme linéaire sur $L^1_d(\Omega, \mathcal{A}, \mathbb{P})$.
2) Si $X \in L^1_d(\Omega, \mathcal{A}, \mathbb{P})$ et si $X \ge 0$ alors $\mathbb{E}(X) \ge 0$.
3) Si $X \le Y$ et si X et Y sont dans $L^1_d(\Omega, \mathcal{A}, \mathbb{P})$ alors $\mathbb{E}(X) \le \mathbb{E}(Y)$.

Proposition.– Soient $X, Y : \Omega \to \mathbb{R}$ des VAD réelles sur $(\Omega, \mathcal{A})$.

1) Si $X \in L^1_d(\Omega, \mathcal{A}, \mathbb{P})$ alors $|\mathbb{E}(X)| \le \mathbb{E}(|X|)$.
2) Si $X \sim Y$ et si $Y \in L^1_d(\Omega, \mathcal{A}, \mathbb{P})$ alors $X \in L^1_d(\Omega, \mathcal{A}, \mathbb{P})$ et $\mathbb{E}(X) = \mathbb{E}(Y)$.
3) Si $|X| \le Y$ et si $Y \in L^1_d(\Omega, \mathcal{A}, \mathbb{P})$ alors $X \in L^1_d(\Omega, \mathcal{A}, \mathbb{P})$ et $|\mathbb{E}(X)| \le \mathbb{E}(Y)$.

Théorème.– Soient X, Y dans $L^1_d(\Omega, \mathcal{A}, \mathbb{P})$.

Si X et Y sont indépendantes pour $\mathbb{P}$ alors $XY \in L^1_d(\Omega, \mathcal{A}, \mathbb{P})$ et $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$.

1.4 Exemples fondamentaux

Proposition.– Soit X une VAD réelle sur $(\Omega, \mathcal{A})$.

1) Si $X = \mathbf{1}_A$ avec $A \in \mathcal{A}$ alors $X \in L^1_d(\Omega, \mathcal{A}, \mathbb{P})$ et $\mathbb{E}(X) = \mathbb{P}(A)$.
2) Si X est constante égale au réel a alors $X \in L^1_d(\Omega, \mathcal{A}, \mathbb{P})$ et $\mathbb{E}(X) = a$.
4) Si X est presque sûrement égale au réel a alors $X \in L^1_d(\Omega, \mathcal{A}, \mathbb{P})$ et $\mathbb{E}(X) = a$.
5) Si X est bornée alors $X \in L^1_d(\Omega, \mathcal{A}, \mathbb{P})$.

Proposition.– Soit X une VAD réelle sur $(\Omega, \mathcal{A})$.

1) Si $X(\Omega) = \{x_1, \cdots, x_p\}$ avec $x_1, \cdots, x_p$ deux à deux distincts alors $X \in L^1_d(\Omega, \mathcal{A}, \mathbb{P})$ et $\mathbb{E}(X) = \sum_{k=1}^{p} x_k \mathbb{P}(X = x_k)$.

2) Si $X(\Omega)=\{x_n, n\in\mathbb{N}\}$ et si les x_n, $n\in\mathbb{N}$ sont deux à deux distincts alors $X\in L^1_d(\Omega,\mathcal{A},\mathbb{P})$ si et seulement si la série $\sum x_n\mathbb{P}(X=x_n)$ est absolument convergente et lorsque ces conditions sont réunies on a : $\mathbb{E}(X)=\sum_{n=0}^{+\infty} x_n\mathbb{P}(X=x_n)$.

Proposition.– Soit $X:\Omega\to\mathbb{N}$ une variable aléatoire discrète sur $(\Omega,\mathcal{A})$ à valeurs dans $\mathbb{N}$.

Si $X \leadsto \mathcal{B}(p)$ alors $X\in L^1_d(\Omega,\mathcal{A},\mathbb{P})$ et $\mathbb{E}(X)=p$.

Si $X \leadsto \mathcal{B}(n,p)$ alors $X\in L^1_d(\Omega,\mathcal{A},\mathbb{P})$ et $\mathbb{E}(X)=np$.

Si $X \leadsto \mathcal{G}(p)$ alors $X\in L^1_d(\Omega,\mathcal{A},\mathbb{P})$ et $\mathbb{E}(X)=\frac{1}{p}$.

Si $X \leadsto \mathcal{P}(\lambda)$ alors $X\in L^1_d(\Omega,\mathcal{A},\mathbb{P})$ et $\mathbb{E}(X)=\lambda$.

1.5 Moments d'ordre p

Définition.– Soient X une VAD réelle sur $(\Omega,\mathcal{A})$ et $p\in\mathbb{N}^*$.

1) On dit que X admet un moment d'ordre p si et seulement si X^p est d'espérance finie.
   Si tel est le cas alors le réel $\mathbb{E}(X^p)=\sum_{t\in\mathbb{R}} t^p\mathbb{P}(X=t)$ est appelé moment d'ordre p de X.
2) L'ensemble des VAD réelles sur $(\Omega,\mathcal{A})$ qui admettent un moment d'ordre p est notée $L^p_d(\Omega,\mathcal{A},\mathbb{P})$. On a donc : $L^p_d(\Omega,\mathcal{A},\mathbb{P})=\left\{X\in V_{\mathbb{R}}(\Omega,\mathcal{A}) \mid \mathbb{E}\left(|X|^p\right)<+\infty\right\}$.

Proposition.– Soit X une VAD réelle sur $(\Omega,\mathcal{A})$.

1) Si X est constante égale au réel a alors X admet un moment d'ordre p égal à a^p.
2) Si X est presque sûrement égale au réel a alors X admet un moment d'ordre p égal à a^p.
3) Si X est bornée alors X admet un moment d'ordre p.

2. Espace $L^2_d(\Omega,\mathcal{A},P)$, covariance, variance et écart type

2.1 Variables aléatoires discrètes ayant un moment d'ordre deux

Théorème.–

1) Si X et Y sont dans $L^2_d(\Omega,\mathcal{A},\mathbb{P})$ alors $XY\in L^1_d(\Omega,\mathcal{A},\mathbb{P})$.
2) L'ensemble $L^2_d(\Omega,\mathcal{A},\mathbb{P})$ est un $\mathbb{R}$-espace vectoriel
3) L'application $\left((X,Y)\mapsto\mathbb{E}(XY)\right)$ est un semi-produit scalaire sur $L^2_d(\Omega,\mathcal{A},\mathbb{P})$.

Théorème.–

1) Si X et Y sont dans $L^2_d(\Omega,\mathcal{A},\mathbb{P})$ alors $XY\in L^1_d(\Omega,\mathcal{A},\mathbb{P})$ et $|\mathbb{E}(XY)|\le\sqrt{\mathbb{E}(X^2)}\sqrt{\mathbb{E}(Y^2)}$.
2) $L^2_d(\Omega,\mathcal{A},\mathbb{P})\subset L^1_d(\Omega,\mathcal{A},\mathbb{P})$ et si $X\in L^2_d(\Omega,\mathcal{A},\mathbb{P})$ alors $|\mathbb{E}(X)|\le\sqrt{\mathbb{E}(X^2)}$.

2.2 Covariance de deux variables aléatoires de $L^2_d(\Omega,\mathcal{A},\mathbb{P})$

Définition.– La covariance de deux variables aléatoires X et Y de $L^2_d(\Omega,\mathcal{A},\mathbb{P})$ est le réel $\mathrm{Cov}(X,Y)$ défini par $\mathrm{Cov}(X,Y)=E(C_XC_Y)$ où $C_X=X-E(X)$ et $C_Y=Y-E(Y)$.

Proposition.– L'application $\left((X,Y)\mapsto\mathrm{Cov}(X,Y)\right)$ est un semi-produit scalaire sur $L^2_d(\Omega,\mathcal{A},\mathbb{P})$.

Proposition.– Soient X,Y dans $L^2_d(\Omega,\mathcal{A},\mathbb{P})$.

1) $\mathrm{Cov}(X,Y)=\mathbb{E}(XY)-\mathbb{E}(X)\mathbb{E}(Y)$.
2) Si X est presque sûrement égale au réel a alors $\mathrm{cov}(X,Y)=0$.
3) Si X et Y sont indépendantes pour $\mathbb{P}$ alors $\mathrm{Cov}(X,Y)=0$.

2.3 Variance et écart type d'une variable aléatoire de $L^2_d(\Omega,\mathcal{A},\mathbb{P})$

Définition.– Soit $X\in L^2_d(\Omega,\mathcal{A},\mathbb{P})$.

1) La variance de X est le réel positif $\mathbb{V}(X)$ défini par $\mathbb{V}(X)=\mathbb{E}(C_X^2)$ où $C_X=X-\mathbb{E}(X)$.
2) L'écart type de X est le réel positif $\sigma(X)$ défini par $\sigma(X)=\sqrt{\mathbb{V}(X)}$.

Proposition.– Soit $X\in L^2_d(\Omega,\mathcal{A},\mathbb{P})$.

1) $\mathbb{V}(X)=\mathrm{cov}(X,X)=\mathbb{E}(X^2)-\mathbb{E}(X)^2$.
2) Si X est presque sûrement égale à a alors $\mathbb{V}(X)=0$.

Proposition.– Soient X,Y dans $L^2_d(\Omega,\mathcal{A},\mathbb{P})$.

1) Si X et Y ont la même loi $(X\sim Y)$ alors $\mathbb{V}(X)=\mathbb{V}(Y)$.
2) $|\mathrm{Cov}(XY)|\le\sigma_X\sigma_Y$.

Proposition.– Soient X,Y dans $L^2_d(\Omega,\mathcal{A},\mathbb{P})$ et $(a,b)\in\mathbb{R}^2$.

1) $\mathbb{V}(X+Y)=\mathbb{V}(X)+\mathbb{V}(Y)+2\mathrm{Cov}(X,Y)$.
2) $\mathbb{V}(aX+b)=a^2\mathbb{V}(X)$ et $\sigma(aX+b)=|a|\sigma(X)$.
3) Si X et Y sont indépendantes pour P alors $\mathbb{V}(X+Y)=\mathbb{V}(X)+\mathbb{V}(Y)$.

Proposition.– Soient $X_1,\cdots,X_p$ dans $L^2_d(\Omega,\mathcal{A},\mathbb{P})$.

1) $\mathbb{V}\left(\sum_{k=1}^{p} X_k\right) = \sum_{k=1}^{p}\mathbb{V}(X_k) + 2\sum_{1\leq i<j\leq n}\mathrm{Cov}(X_i,X_j)$.

2) Si $X_1,\cdots,X_p$ sont deux à deux indépendantes alors $\mathbb{V}\left(\sum_{k=1}^{p} X_k\right) = \sum_{k=1}^{p}\mathbb{V}(X_k)$.

2.4 Exemples fondamentaux

Proposition.– Soit X une VAD réelle sur $(\Omega,\mathcal{A})$.

1) Si $X = \mathbb{1}_A$ avec $A \in \mathcal{A}$ alors $X \in L^2_d(\Omega,\mathcal{A},\mathbb{P})$ et $\mathbb{V}(X) = \mathbb{P}(A)(1-\mathbb{P}(A))$.

2) Si X est constante égale au réel a alors $X \in L^2_d(\Omega,\mathcal{A},\mathbb{P})$ et $\mathbb{V}(X) = 0$.

3) Si X est presque sûrement égale au réel a alors $X \in L^2_d(\Omega,\mathcal{A},\mathbb{P})$ et $\mathbb{V}(X) = 0$.

4) Si X est bornée alors $X \in L^2_d(\Omega,\mathcal{A},\mathbb{P})$.

Proposition.– Soit $X : \Omega \to \mathbb{N}$ une variable aléatoire discrète sur $(\Omega,\mathcal{A})$ à valeurs dans $\mathbb{N}$.

1) Si $X \rightsquigarrow \mathcal{B}(p)$ alors $X \in L^2_d(\Omega,\mathcal{A},\mathbb{P})$ et $\mathbb{V}(X) = p(1-p)$.

2) Si $X \rightsquigarrow \mathcal{B}(n,p)$ alors $X \in L^2_d(\Omega,\mathcal{A},\mathbb{P})$ et $\mathbb{V}(X) = np(1-p)$.

3) Si $X \rightsquigarrow \mathcal{G}(p)$ alors $X \in L^2_d(\Omega,\mathcal{A},\mathbb{P})$ et $\mathbb{V}(X) = \dfrac{1-p}{p^2}$.

4) Si $X \rightsquigarrow \mathcal{P}(\lambda)$ alors $X \in L^2_d(\Omega,\mathcal{A},\mathbb{P})$ et $\mathbb{V}(X) = \lambda$.

Définition.– Soit $X : \Omega \to \mathbb{R}$ une VAD réelle sur $(\Omega,\mathcal{A})$.

La variable aléatoire X est dite centrée si et seulement si elle est d'espérance nulle.

La variable aléatoire est dite réduite si et seulement si elle est de variance égale à un.

Proposition.– Soit $X : \Omega \to \mathbb{R}$ une VAD réelle sur $(\Omega,\mathcal{A})$.

1) Si $X \in L^1_d(\Omega,\mathcal{A},\mathbb{P})$ alors $C_X = X - \mathbb{E}(X)$ est une VAD réelle centrée, dite associée à X.

2) Si $X \in L^2_d(\Omega,\mathcal{A},\mathbb{P})$ et si $\mathbb{V}(X) > 0$ alors $R_X = \dfrac{1}{\sigma(X)}(X - \mathbb{E}(X))$ est une VAD centrée réduite, dite associée à X.

3. Loi faible des grands nombres

Proposition.– (Inégalités de Markov et Tchebychev)

1) Si $X \in V_{\mathbb{R}}(\Omega,\mathcal{A})$ alors : $\forall \varepsilon > 0$, $\mathbb{P}(|X| \geq \varepsilon) \leq \dfrac{\mathbb{E}(|X|)}{\varepsilon}$. (Markov)

2) Si $X \in L^2_d(\Omega,\mathcal{A},P)$ alors : $\forall \varepsilon > 0$, $\mathbb{P}(|X - \mathbb{E}(X)| \geq \varepsilon) \leq \dfrac{\mathbb{V}(X)}{\varepsilon^2}$. (Tchebychev)

Théorème.– (Loi faible des grands nombres)

Soit $(X_n)_{n\geq 1}$ une suite d'éléments de $L^2_d(\Omega,\mathcal{A},\mathbb{P})$. On pose : $S_n = \sum_{k=1}^{n} X_k$.

Si les X_n, $n \geq 1$ sont deux à deux indépendantes pour la probabilité $\mathbb{P}$ et si elles ont la même loi alors les X_n, $n \geq 1$ ont la même espérance m, le même écart type σ et on a :

1) $\forall \varepsilon > 0$, $\mathbb{P}\left(\left|\frac{S_n}{n} - m\right| \geq \varepsilon\right) \leq \dfrac{\sigma^2}{n\varepsilon^2}$.

2) En particulier : $\forall \varepsilon > 0$, $\lim\limits_{n\to+\infty} \mathbb{P}\left(\left|\frac{S_n}{n} - m\right| \geq \varepsilon\right) = 0$