Chapitre J2 : Variables aléatoires.

I Variables aléatoires et lois

0) Motivation

Souvent, à propos d'une expérience aléatoire, comme le lancer de deux dés, les joueurs ne s'intéressent pas à la vision d'ensemble de l'univers Ω des issues possibles, mais seulement à ce que les fait gagner ou perdre. Par exemple à la valeur de la somme de deux dés. Pour les joueurs, l'univers des issues qu'on a étudié au **J1** est donc moins important que la suite des entiers « somme des deux dés ». Dans le langage des probabilités, le joueur s'intéresse aux propriétés de la *variable aléatoire* « somme des deux dés ».

Pour le mathématicien, qui veut aider les joueurs, et qui étudie le lien entre les objets, autrement dit les *applications*, il s'agit donc d'étudier, du point de vue des probabilités, l'*application* de l'univers Ω dans $\mathbb{N}$ dont les valeurs sont les sommes des résultats des dés.

1) Le langage des variables aléatoires :

a) (i) Définition générale : Soit Ω un ensemble *fini* (univers). Une *variable aléatoire* (abrév. v.a.) X sur Ω est une application $X : \Omega \to E$ où E est un ensemble qui peut être quelconque.

On dira que X est un variable aléatoire réelle si $E \subset \mathbb{R}$.

N.B. Autrement dit une v.a. est une *application quelconque* : simplement, son ensemble de départ est un *univers probabiliste*. Attention, lorsqu'on considère des *univers infinis* il faut être un peu plus précis (cf. deuxième année).

(ii) **Premier exemple** Pour le lancer de deux dés : un rouge et un vert. Une issue est un couple $\omega = (\omega_1, \omega_2)$ où la première coord. est le résultat du dé rouge. L'univers des issues est $\Omega = [\![1,6]\!]^2$.

On peut considérer différentes v.a. comme par exemple :

- la v.a. « résultat du lancer du dé rouge » :

$$\begin{aligned} X_1 : \Omega &\to [\![1,6]\!] \\ (\omega_1, \omega_2) &\mapsto \omega_1 \end{aligned}$$

- la v.a. « résultat du lancer du dé vert » :

$$\begin{aligned} X_2 : \Omega &\to [\![1,6]\!] \\ (\omega_1, \omega_2) &\mapsto \omega_2 \end{aligned}$$

- la v.a. « somme des résultats » :

$$\begin{aligned} S = X_1 + X_2 : \Omega &\to [\![2,12]\!] \\ (\omega_1, \omega_2) &\mapsto \omega_1 + \omega_2 \end{aligned}$$

(iii) **Exemple : indicatrice d'un événement** Si $A \subset \Omega$ est un *événement* d'un univers fini Ω, on appelle *indicatrice* de l'événement A et on note $\mathbf{1}_A$, la fonction caractéristique de A.

Exemple pour le tirage d'un dé : $\Omega = [\![1,6]\!]$. Disons qu'on gagne un point si le numéro tiré est pair et qu'on ne gagne rien sinon, on peut considérer la v.a. qui renvoie 1 si le numéro tiré est pair et 0 si le numéro tiré est impair. C'est la fonction indicatrice de l'événement « le numéro tiré est pair » qui est le sous-ensemble $\{2,4,6\}$ de Ω.

N.B. Derrières les phrases du type « le numéro tiré est pair », on peut mettre un *événement* (comme on l'a fait ici et au **J1**) ou bien comme on vient de le faire une *variable aléatoire*! Mathématiquement cependant les fonctions et les ensembles ne se manipulent pas de la même façon. Les fonctions sont plus souples, plus puissantes...

b) Ensembles associés à une variable aléatoire : langage probabiliste

(i) Préimage : Soit $X : \Omega \to E$ une v.a. Soit $A \subset E$. La préimage $X^{-1}(A) = \{\omega \in \Omega,\ X(\omega) \in A\}$ qui est un sous-ensemble de Ω donc un *événement* se note aussi : l'événement

$$\{X \in A\} \quad \text{ou même} \quad (X \in A)$$

(ii) Préimage d'un singleton : Avec les mêmes notations, pour $x \in E$, la préimage $X^{-1}(\{x\})$ se note simplement :

$$\{X = x\}$$

C'est l'événement « $X = x$ ».

Exemple : Pour S la somme des lancers de deux dés, l'événement « $S = 3$ » est

(iii) De même, si $X : \Omega \to \mathbb{R}$ et $x \in \mathbb{R}$, on note $X \leq x$ par exemple, ce qui en notations plus standard en math. se noterait :

2) Avec des probabilités en plus : la notion de *loi* d'une v.a.

a) Définition : (i) Jusqu'à maintenant, on n'a pas parlé d'une probabilité P sur notre univers Ω. Or le joueur s'intéresse en fait, pour une variable aléatoire $X : \Omega \to E$, d'une part à l'ensemble $X(\Omega)$ des valeurs de la v.a. mais surtout aux *probabilités* des différents événements associés à X.

(ii) **Définition :** Soit (Ω, P) un espace probabilisé fini et soit $X : \Omega \to E$ une v.a. On appelle *loi de* X l'application :

$$\begin{aligned} P_X : \mathcal{P}(E) &\to [0,1], \\ A &\mapsto P(X^{-1}(A)) = P(X \in A). \end{aligned}$$

Prop. : L'application P_X ainsi définie est une *probabilité* sur l'ensemble E et aussi sur l'ensemble $X(\Omega)$.

(iii) **Importance pratique de la notion de loi :** *déplacement du regard par rapport au* **J1**. C'est la connaissance de la loi qui compte pour le joueur : il se moque de l'ensemble Ω ou de X en tant qu'application : c'est le sens de la notation $P(X \in A)$. Par exemple il vaut connaître $P($« la somme des dés est paire »$)$

b) Réduction aux événements élémentaires

(i) **Prop.** Soit X une v.a. sur un espace probabilisé fini (Ω, P). La loi de X est entièrement déterminée par la donnée pour tout $x \in X(\Omega)$ des nombres $P_X(\{x\}) = P(X = x)$.

(ii) *L'essentiel de la preuve :* Pour tout $A \in E$, $P(X \in A) = \sum_{x \in A} P(X = x)$.

En général, c'est ainsi qu'on donnera la loi d'une v.a. : en donnant les $P(X = x)$ cf. exples infra.

(iii) **Réciproque plus théorique mais utile !**

(H) Soient $n \in \mathbb{N}^*$ et E un ensemble et $(x_1, \dots, x_n) \in E^n$ deux à deux distincts. Soit $(p_1, \dots, p_n) \in (\mathbb{R}^+)^n$ tels que : $\sum_{i=1}^{n} p_i = 1$.

Soit enfin Ω un ensemble fini de cardinal au moins n.

(C) Il existe une v.a. $X : \Omega \to E$ et une probabilité P sur Ω telle que la loi de X sur (Ω, P) soit donnée par :

$$\forall\, i \in [\![1, n]\!],\ P(X = x_i) = p_i.$$

(iv) *Preuve du (iii) :* cf. notes manuscrites.

(v) **Terminologie :** une famille (p_i) de réels positifs telle que $\sum_{i=1}^{n} p_i = 1$ s'appelle parfois une *distribution de probabilités*. Le (iii) dit alors que si on se fixe une telle distribution de probabilité, il existe une v.a. X dont la loi suit cette distribution.

3) Lois classiques et importantes :

a) Loi de Bernoulli :

(i) **Définition :** Soit (Ω, P) un espace probabilisé et $X : \Omega \to \{0,1\}$ une v.a. qui ne peut prendre que deux valeurs (ici, par convention, 0 ou 1). En notant $p = P(X = 1)$, on a bien sûr $P(X = 0) = 1 - p$ et on dit que X suit une *loi de Bernoulli de paramètre p*.
On notera aussi : $X \sim \mathcal{B}(p)$.

(ii) **Premier exemple concret :** *n tirages à pile ou face d'une pièce équilibrée :* on note X_i la v.a. « résultat du i-ème lancer ». En notant 1 pour pile et 0 pour face, on sait que (cf. construction du modèle à la fin du **J1**) que : $P(X_i = 1) = P(X_i = 0) = 1/2$ si la pièce est équilibrée.

(iii) **Deuxième exemple concret :** *une étape d'une marche aléatoire sur* $\mathbb{Z}$: on note Y_i la v.a. qui décide si à l'étape i, on va à gauche ou à droite, à valeurs plutôt dans $\{-1, 1\}$. Une telle v.a. Y est plutôt appelée v.a. de Rademacher, mais on a l'équivalence : Y est de Rademacher ssi $X = 2Y - 1$ est une v.a. de Bernoulli.

(iv) **Troisième exemple plus théorique :** pour tout sous-ensemble A de Ω, la v.a. $\mathbf{1}_A$ suit une loi de Bernoulli de paramètre

b) la loi binomiale :
(i) **Avec l'exemple du jeu de pile ou face :** On considère n lancers d'une pièce, qui tombe sur pile avec probabilité $p \in [0,1]$. On se place du point de vue du joueur qui parie toujours sur pile et on s'intéresse à la variable aléatoire S : nombre de succès en n lancers.

Avec la construction faite à la fin du **J1** *:* on dispose d'un espace probabilsé (Ω, P) pour étudier ce problème et donc on peut se demander quelle est la *loi* de S i.e. pour tout $k \in [\![0, n]\!]$, $P(S = k)$.

(ii) *Comment faire ce calcul ?* L'événement $A_k = \{S = k\}$ est l'ensemble des issues formées de k piles et $n - k$ faces.
— Toutes ces issues ont la même probabilité :
— Le nombre d'élément de A_k est :
D'où la formule :

$$\forall\, k \in [\![0, n]\!],\ P(S = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

(iii) **Définition générale :** Soit (Ω, P) un espace probabilisé fini et $X : \Omega \to \mathbb{N}$ une v.a. On dit que X suit une *loi binomiale* de paramètres (n, p) si $X(\Omega) \subset [\![0, n]\!]$ et :

$$\forall\, k \in [\![0, n]\!],\ P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

On notera $X \sim \mathcal{B}(n, p)$.

(iv) **Reformulation :** la construction mentionnée au (i) donne que les n lancers sont *indépendants* au sens défini à la fin du **J1**. On verra ci-dessous qu'on peut dire aussi que les v.a. correspondantes sont *indépendantes*.

On vient donc de montrer que la v.a. S « nombre de succès dans une suite de tirages de Bernoulli indépendants » suit une loi $\mathcal{B}(n, p)$

(v) **Autre exemple très important :** on tire dans une urne, *avec remise* n boules qui peuvent être de deux couleurs, disons noire ou blanche. Si on note p la proportion de boules blanches, la v.a. qui donne le nombre de boules blanches tirées au bout de n tirages suit la loi $\mathcal{B}(n, p)$.
En effet : ces n tirages avec remises sont équivalents à n tirages de Bernoulli *indépendants*.

c) Loi uniforme :

(i) Définition : Si (Ω, P) est un espace probabilisé fini, une v.a. $X : \Omega \to E$ suit une loi uniforme sur $[\![1,n]\!]$ si $X(\Omega) \subset [\![1,n]\!]$ et si :

$$\forall\, k \in [\![1,n]\!],\ P(X=k) = \frac{1}{n}.$$

Plus généralement si $a < b$ sont deux entiers, on dit que X suit une loi uniforme sur $[\![a,b]\!]$ si, et seulement si :

$$\forall\, k \in [\![a,b]\!],\ P(X=k) = \frac{1}{b-a+1}.$$

On note $X \sim \mathcal{U}([\![a,b]\!])$.

(ii) Donnez des exemples de v.a. qui suivent une loi uniforme :

4) Opérations sur les v.a. : calculs des lois correspondantes

a) Composition d'une v.a. avec une fonction

(i) **Définition :** Soit (Ω, P) un espace probabilisé, E et F deux ensembles et $X : \Omega \to E$ une v.a. et $f : E \to F$ une fonction. La composée $f \circ X : \Omega \to F$ est souvent notée $f(X)$ dans les énoncés de probabilité. Elle s'appelle *v.a.* image de X par f.

(ii) **Exemple :** Si $X : \Omega \to \mathbb{R}$ est une v.a. renvoyant un nombre réel, on peut considérer $|X| : \omega \in \Omega \mapsto |X(\omega)|$. En fait $|X| = f \circ X$ où $f = abs$. De même $\sin(X) : \Omega \to \mathbb{R}$, $\omega \mapsto \sin(X(\omega))$.

(iii) **Prop.** Avec les notations du (i), on peut calculer la loi de $f(X)$ à partir de celle de X comme suit :

$$\forall\ y \in f(X(\Omega)),\ P(f(X)=y) = \sum_{x \in f^{-1}(\{y\})} P(X=x).$$

(iv) **Exercice :** Un joueur lance deux fois un dé équilibré. Soit X la v.a. différence entre le résultat du premier lancer et celui du second lancer. Déterminer la loi de X et la loi de $|X|$.

Conseil : il peut être agréable d'utiliser un tableau.

b) Attention pour la somme de deux v.a. quelconques :

Pour X et Y deux v.a. sur Ω à valeurs réelles par exemple, on peut définir la somme $X+Y$.

Attention : la loi de $X+Y$ ne se déduit pas simplement de la loi de X et de celle de Y.

Tout ce qu'on peut dire à ce stade est pour chaque $z \in (X+Y)(\Omega)$, on a :

$$P(X+Y=z) = \sum_{x+y=z} P(\{X=x\} \cap \{Y=y\})$$

Justifiez cette formule !

Mais cela ne se déduit pas des lois de X et Y si on ne sait pas que les événements sont indépendants, ou comme on va le dire ci-dessous que les *v.a.* sont indépendantes. Avant cela, on introduit des éléments de langage supplémentaires pour les couples de v.a.

II Familles de variables aléatoires, indépendance :

1) Le langage des couples de v.a.

a) Une définition évidente :

Déf. Soit Ω un univers fini et X, Y deux v.a. sur Ω, à valeurs dans E et E' respectivement, on appelle *couple (X,Y)* de v.a. la variable aléatoire :

$$\begin{aligned}(X,Y) &: \Omega \to E \times E',\\ &\omega \mapsto (X(\omega), Y(\omega))\end{aligned}$$

Ainsi (X,Y) est une v.a. de Ω dans $E \times E'$.

Exemple : pour le lancer de deux dés, avec X le résultat du premier dé et Y le résultat du second dé.

b) Du point de vue des lois :

(i) **Loi conjointe d'un couple de v.a.**

Définition : Si X et Y sont deux v.a. sur Ω la *loi conjointe* de (X,Y) est la loi du *couple* (X,Y)

(ii) **Lois marginales : d'un couple de v.a.**

Définition : Si X et Y sont deux v.a. sur Ω, les *lois marginales* de (X,Y) sont les lois de X (première loi marginale) et de Y (deuxième loi marginale).

(iii) Lien entre lois marginales et loi conjointe :

Prop. La loi conjointe détermine les lois marginales. La récip. n'est pas vraie en général.

Preuve de la première assertion : comme $\{Y=y\}_{y\in Y(\Omega)}$ est un S.C.E., par la formule des proba. totales, on a formule plus précise suivante :

$$\forall\, x\in X(\Omega),\ P(X=x)=\sum_{y\in Y(\Omega)} P(\{X=x\}\cap\{Y=y\}).$$

(iv) Un contre-exemple pour la réciproque :

Un urne contient trois boules numérotées de 1 à 3. On en tire deux et on note X (resp. Y) la v.a. donnant le premier numéro tiré resp. le second.

Déterminer la loi conjointe de X et Y dans le cas où l'on tire sans remise et dans le cas où l'on tire avec remise : elles sont bien sûr différentes.

Pourtant les lois marginales sont les mêmes dans les deux cas : loi uniforme sur $[\![1,3]\!]$ pour X comme pour Y.

c) Loi conditionnelle :

(i) Déf. Soit (X,Y) un couple de v.a. sur un univers fini (Ω,P). Pour tout $y\in Y(\Omega)$ tel que $P(Y=y)\neq 0$, la *loi conditionnelle* de X sachant $\{Y=y\}$ est la loi de X sur l'espace probabilisé $(\Omega,P_{\{Y=y\}})$ où $P_{\{Y=y\}}$ désigne la probabilité conditionnelle relative à l'événement $\{Y=y\}$.

Elle est donc déterminée par la donnée, pour tout $x\in X(\Omega)$ de :

$$P_{\{Y=y\}}(X=x)=P(X=x\mid Y=y).$$

(ii) Un intérêt dans la modélisation : il arrive souvent qu'on modélise un couple de v.a. (X,Y) en se donnant la loi de Y puis la loi de X conditionnée par celle de Y. Voir par exemple l'exercice sur le secrétaire de la planche : X est le nombre de correspondant que le secrétaire arrive à joindre, puis Y le nombre de personnes qu'il joint au second appel...

2) Indépendance de deux v.a.

a) Définition : Soit (Ω,P) un espace probabilisé fini. Soient $X:\Omega\to E$ et $Y:\Omega\to F$ deux v.a. On dit que X et Y sont indépendantes ssi pour toute partie A de E et toute partie B de F les événements $(X\in A)$ et $(Y\in B)$ sont indépendants autrement dit ssi :

$$\forall\, A\ \in\mathcal{P}(E),\ \forall\, B\in\mathcal{P}(F),\ P\left((X\in A)\cap(Y\in B)\right)=P(X\in A).P(Y\in B).$$

b) Caractérisation en se réduisant aux « valeurs » :

Prop. Avec les notations du a), X et Y sont indépendantes ssi :

$$\forall\, x\in E,\ \forall\, y\in F,\ P\left((X=x)\cap(Y=y)\right)=P(X=x).P(Y=y).$$

Preuve du sens non évident : Ecrire les événements $\{X\in A\}$ et $\{Y\in B\}$ comme deux réunions.

c) Lien avec la notion d'événements indépendants :

(i) **Prop. (qui pourrait aussi servir de déf. des événements indépendants)** Si (Ω,P) est un espace probabilisé et A et B sont deux événements dans Ω, A et B sont indépendants ssi les v.a. $\mathbf{1}_A$ et $\mathbf{1}_B$ sont indépendantes.

(ii) **Traductions pour deux v.a. suivant une loi de Bernoulli** Deux v.a. X et Y suivant un loi de Bernoulli à valeurs dans $\{0,1\}$ sont indépendantes ssi les événements $(X=1)$ et $(Y=1)$ sont indépendants.

3) Fonctions de deux variables aléatoires indépendantes :

a) Cas où on compose chaque v.a. par une fonction :

Prop. Si X et Y sont deux v.a. indépendantes (*abrév. v.a.i.*) sur Ω alors pour toutes fonctions f et g telles que cela ait un sens, les v.a. $f(X)$ et $g(Y)$ sont indépendantes.

Idée pour la preuve : L'essentiel $P(f(X) \in A) = P(X \in f^{-1}(A))$.

b) Cas de la somme de deux v.a.i.

Prop. Si X et Y sont deux v.a. indépendantes sur (Ω, P) alors :

$$\forall\, z \in (X+Y)(\Omega),\ P(X+Y=z) = \sum_{(x,y)|x+y=z} P(X=x).P(Y=y).$$

Preuve : Comparer à la fonction donnée à la fin du § I.

Remarque pour les v.a. à valeurs dans $\mathbb{N}$:

La formule précédente devient, si $X(\Omega)+Y(\Omega) \subset [\![0,N]\!]$:

$$\forall\, k \in [\![0,N]\!],\ P(X+Y=k) = \sum_{i=0}^{k} P(X=i).P(Y=k-i).$$

Question : à quoi vous fait penser cette formule réécrite sous la forme :

$$P_{X+Y}(k) = \sum_{i=0}^{k} P_X(i)P_Y(k-i)$$

Terminologie : On dira que, si X et Y sont indépendantes, P_{X+Y} est le *produit de convolution* de P_X et P_Y.

c) Exemple (exercice) :

Soient deux v.a.i. X et Y de même loi uniforme $\mathcal{U}([\![0,n]\!])$ sur Ω (**N.B. J'ai changé $[\![1,n]\!]$ en $[\![0,n]\!]$ pour que le résultat soit plus joli)**. Déterminer la loi suivie par $X+Y$ qui s'appelle *loi triangulaire*. Justifier cette terminologie en représentant le graphe de $k \mapsto P_{X+Y}(k)$.

d) Polynôme codant une loi de v.a. à valeurs dans $\mathbb{N}$ (fonction génératrice)

Ce paragraphe d) n'est PAS au programme de cette semaine. Semaine suivante avec les polynômes.

(i) **Déf.** Pour X une v.a. à valeurs dans $[\![0,n]\!]$, posons $G_X(T) = \sum_{k=0}^{n} P(X=k)T^k \in \mathbb{R}[T]$ (on note T l'indéterminée).

(ii) **Question :** Pour X et Y deux v.a. indép. à valeurs dans $\mathbb{N}$, relier : G_{X+Y} avec G_X et G_Y.

(iii) Application : *cf. planche*

4) Indépendance de n variables aléatoires :

a) Définition :

(i) **Déf.** Des v.a. $X_1, \dots, X_n$ définies sur un espace probabilisé (Ω, P) et à valeurs chacune dans un ensemble E_i sont dites indépendantes (ou pour insister *mutuellement indépendantes*) ssi pour toutes parties A_1 de E_1, A_2 de E_2,..., A_n de E_n les événements $(X_1 \in A_1)$, $(X_2 \in A_2), \dots, (X_n \in A_n)$ sont mutuellements indépendants.

(ii) **Caract.** (comme dans le cas de deux v.a.) Pour un espace probabilisé fini, on peut, dans la déf. précédente, se contenter de vérifier l'indép. à partir des valeurs élémentaires $(X_1 = x_1), \dots, (X_n = x_n)$.

b) Familles indép. déduites d'une famille de v.a. (mutuellement) indépendantes :

(i) **Prop. (sous famille)** Si $X_1, \dots, X_n$ sont des v.a. indépendantes, alors toute sous famille $X_{i_1}, \dots, X_{i_k}$ forme une famille de v.a. indépendantes.

(ii) *Preuve admise* Là encore il faut *calculer* : pourquoi suffit-il de mq $X_1, \dots, X_{n-1}$ sont indép. ?

(iii) **Corollaire :** Des v.a.indépendantes sont en part. indép. deux à deux, mais la récip. est fausse.

Exercice : Fabriquer un contre-exemple à partir du contre-exemple vu pour les événements au **J1** : il s'agit de passer d'événements à des v.a., pour cela on a des v.a. naturelles les ...

(iv) **Prop. (division en deux familles)** Si $X_1, \dots, X_n$ sont des v.a. indépendantes alors pour tout $r \in [\![1, n]\!]$ les deux v.a. $(X_1, \dots, X_r) : \Omega \to E_1 \times \cdots \times E_r$ et $(X_{r+1}, \dots, X_n) : \Omega \to E_{r+1} \times \cdots \times E_n$ sont indépendantes.

c) Passage à des fonctions de ces v.a.

(i) **Prop.** Si $X_1, \dots, X_n$ sont des v.a. indép. avec X_i à valeur dans E_i et si $f_1, \dots, f_n$ sont des fonctions avec f_i définie sur E_i, alors $f_1(X_1), \dots, f_n(X_n)$ sont indép.

Preuve : idem cas $n = 2$.

(ii) **Prop. (plus nouvelle)** Si $X_1, \dots, X_n$ sont indép. et si f est une fonction de r variables et g une fonction de $n - r$ variables telles que $f(X_1, \dots, X_r)$ et $g(X_{r+1}, \dots, X_n)$ soient bien définies, alors les v.a. $f(X_1, \dots, X_r)$ et $g(X_{r+1}, \dots, X_n)$ sont indépendantes.

Par exemple avec ces hyp. $X_1 + \cdots + X_r$ et $X_{r+1} + \cdots + X_n$ sont indép.

(iii) **Version plus générale du (ii), lemme des coalitions :** Si $X_1, \dots, X_n$ sont des v.a. indép. et si $I_1, \dots, I_k$ sont des sous-ensembles non vides disjoints de $[\![1, n]\!]$ et si on considère pour tout $j \in [\![1, k]\!]$ une v.a. Y_j qui n'est fonction que des X_i pour $i \in I_j$ alors $Y_1, \dots, Y_k$ seront indépendantes.

Les preuves des énoncés du b) et du c) ne sont pas exigibles.

d) Modélisation : les familles de v.a. indépendantes de même loi (v.a.i.i.d)

(i) **Scholie :** La notion d'indépendance est un modèle mathématique. Notamment on est très intéressé par l'idée de répétition d'une certaine expérience et donc par l'étude d'une suite (finie pour nous) de v.a. indépendantes de même loi.

Du point de vue mathématique, la *construction* d'un espace probabilisé (Ω, P) sur lequel on peut définir une telle suite X_i se fait à partir à l'aide des proba. produits.(cf. fin du **J1**).

Ainsi à la fin du J1, on a construit un univers Ω avec une suite de v.a. de Bernoulli $(X_i)_{i \in [\![1,n]\!]}$ mutuellement indépendantes.

Dans la suite, cette construction sera le plus souvent implicite et on aura beaucoup d'énoncés commençant par : *soit une suite de v.a. (mutuellement) indépendantes et de même loi*, ce qui se dit aussi *v.a. indépendantes et identiquement distribuées* et s'abrège en *v.a.i.i.d.*.

(ii) **Un exemple bien connu :**

Prop. Si $X_1, \dots, X_n$ sont n v.a. suivant une loi de Bernoulli $\mathcal{B}(p)$, mutuellement indépendantes, alors $X_1 + \cdots + X_n$ suit une loi $\mathcal{B}(n, p)$.

(iii) **Généralisation de l'exemple précédent :**

On rappelle que la loi de Bernoulli $\mathcal{B}(p)$ est un cas particulier de la loi binomiale : c'est la loi

Exercice : Si $X_1, \dots, X_n$ sont n v.a.i. suivant une loi binomiale $\mathcal{B}(m_i, p)$ pour $i \in [\![1, n]\!]$, alors la somme $X_1 + \cdots + X_n$ suit une loi $\mathcal{B}(m_1 + \cdots + m_n, p)$.

Deux preuves : **(M1)** *Preuve par récurrence à l'aide du cas $n = 2$.* On utilisera la :

Formule de Chu-Vandermonde (**Hors Programme** : *pour les concours, rejustifier à chaque fois)* si m, n, p sont trois entiers naturels tels que $p \leq m + n$, on a :

$$\sum_{k=0}^{p} \binom{m}{k}\binom{n}{p-k} = \binom{m+n}{p}.$$

(M2) Sans calcul ! : en choisissant pour chaque X_i une somme de v.a. de Bernoulli indépendantes, dans une grande famille de v.a. de Bernoulli mutuellement indépendantes (On peut le faire car le résultat ne dépend que de la loi de chaque X_i).

III Espérance et variance :

1) Espérance :

« Ce qui m'étonne, dit Dieu, c'est l'Espérance.
Et je n'en reviens pas.
⋮
L'Espérance voit ce qui n'est pas encore et qui sera.
Elle aime ce qui n'est pas encore et qui sera. »

Peguy, extrait du Porche de la deuxième vertu.

a) Définition :

(i) Soit (Ω, P) un espace probabilisé *fini* et $X : \Omega \to E \subset \mathbb{R}$ une v.a. On appelle *espérance de* X, le nombre $E(X)$ défini par :

$$E(X) = \sum_{x \in X(\Omega)} x.P(X = x).$$

(ii) Exemple concret : on joue à pile ou face avec une pièce équilibrée et on gagne un euro pour pile et on perd un euro pour face. Autrement dit on a une v.a. gain X qui vaut 1 pour l'issue pile et -1 pour l'issue face. L'espérance vérifie : $E(X) =$

Interprétation intuitive *pour le joueur ?*

Cas d'une pièce légèrement déséquilibrée ?

(iii) **Scholie :** calculer l'espérance, c'est calculer la *moyenne* des valeurs de X possibles pondérée par les probabilité des issues. *Pour l'instant*, pas grand chose à voir avec ce qui dit Peguy : pourquoi diable *ce qui sera* devrait être proche de la moyenne des possibles ? La réponse sera donnée au § V avec la démonstration de *la loi des grands nombres*. Sans elle, les assurances, les casinos, ne pourraient pas gagner leur vie.

(iv) **Remarque :** l'espérance d'une v.a. ne dépend que de sa *loi*, ce qui va nous amener naturellement à calculer l'espérance pour les lois usuelles, ci-dessous.

(iv) **Exemple trivial : espérance d'une v.a. constante** Si X est une v.a. constante égale à a, alors $E(X) =$

b) Premières propriétés de l'espérance :

(i) **Prop : Expression de l'espérance à partir des proba. des événements élémentaires**
Avec les notations de la déf. du a) on a aussi la formule :

$$E(X) = \sum_{\omega \in \Omega} X(\omega).P(\{\omega\}).$$

Idée de la preuve : Pour chaque $x \in X(\Omega)$, écrire $P(X = x)$ en fonction des $P(\{\omega\})$ pour $X(\omega) = x$. □

(ii) **Prop. (linéarité de l'espérance)** L'application E est une forme linéaire sur $\mathcal{F}(\Omega, \mathbb{R})$. Autrement dit pour tout couple (X, Y) de v.a. réelles définies sur Ω et tout $(\lambda, \mu) \in \mathbb{R}^2$,

$$E(\lambda X + \mu Y) = \lambda E(X) + \mu E(Y)$$

Idée de la preuve : appliquer le (i). *Intérêt de la formule du (i) :* Le X n'est plus dans la proba !

(iii) **Prop. (positivité de l'espérance)** Si X est une v.a. réelle positive, alors $E(X) \geq 0$.

(iv) **Linéaire et positive entraîne croissante** si X et Y sont deux v.a. et $X \leq Y$ alors $E(X) \leq E(Y)$.

(v) **Cas des v.a. positives d'espérance nulle :** Si X est une v.a. positive sur Ω et que $E(X) = 0$ alors : attention cela n'entraîne pas X nulle partout... mais l'ensemble des $\omega \in \Omega$ tel que $X(\omega) \neq 0$ est de probabilité nulle.

Terminologie dans ce cas, on dit que X est nulle presque sûrement.

(vi) **A quoi vous font penser les prop. précédentes ?**

b) Ecriture d'une v.a. comme somme de fonctions caractéristiques (indicatrices) : très utile pour l'espérance

(i) **Remarque :** Toute v.a. $X : \Omega \to E$, d'ensemble image $X(\Omega) = \{x_1, \dots, x_n\}$ peut s'écrire (et cela même si les $x_1, \dots, x_n$ ne sont pas distincts) :

$$X = \sum_{k=1}^{n} x_k \mathbf{1}_{A_k},$$

où A_k est l'événement $X = x_k$, et la linéarité de l'espérance (ou sa définition même si $x_1, \dots, x_n$ sont deux à deux distincts) entraîne que :

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

(ii) Cette décomposition d'une v.a. en somme d'indicatrices est très utile pour certains calculs d'espérances. Exemple : nombre moyen de point fixes d'une permutation (ou problème des couples).

c) Calcul de l'espérance d'une composée : théorème de transfert

(i) **Prop.** Soit (Ω, P) un espace probabilisé fini. Soit $X : \Omega \to \mathbb{R}$ une v.a. réelle et $u : X(\Omega) \to \mathbb{R}$. Alors l'espérance de la v.a. $u(X)$ est donnée par :

$$\boxed{E(u(X)) = \sum_{x \in X(\Omega)} u(x) P(X = x).}$$

(ii) *Preuve facile avec la décomposition en indicatrices :*

pour $X = \sum_{k=1}^{n} x_k \mathbf{1}_{A_k}$ alors $u(X) = \sum_{i=1}^{n} u(x_k) \mathbf{1}_{A_k}$ et $E(u(X)) = \sum_{k=1}^{n} u(x_k) P(A_k)$ ce qui donne la conclusion. □

d) Variables aléatoires centrées :

(i) **Déf.** Une v.a. X est dite *centrée* ssi $E(X) = 0$

(ii) Exemple d'un jeu *équitable.*

(iii) **Prop.** Si X est une v.a. quelconque sur un espace probabilisé fini, alors $X - E(X)$ est une v.a. centrée.

2) Calcul d'espérance pour les lois usuelles :

a) Loi uniforme :

Prop. Si $X \sim \mathcal{U}(\llbracket a, b \rrbracket)$ alors $E(X) =$

b) Loi de Bernoulli :

Prop. Si $X \sim \mathcal{B}(p)$ alors $E(X) =$

c) Loi binomiale :

Prop. Si $X \sim \mathcal{B}(n, p)$ alors $E(X) =$

Attention ici deux méthodes de preuve. **(M1)** Calcul direct : demande un calcul sur les binomiaux.

Outil indispensable pour les calculs sur les binomiaux en proba. : pour $k \in \llbracket 1, n \rrbracket$ $k\binom{n}{k} = n\binom{n-1}{k-1}$

(M2) Méthode la plus efficace : voir X comme somme de v.a. de Bernoulli indép.

3) Contrôle avec l'espérance : inégalité de Markov

Andreï Andreïevitch Markov, mathématicien russe (1856–1922)

Prop. Soit X une v.a. réelle *positive* sur un espace probabilisé fini (Ω, P). Alors :

$$\forall\, a > 0,\ P(\{X\ \geq a\}) \leq \frac{E(X)}{a}.$$

Idée de preuve : Y penser plutôt comme $E(X) \geq aP(\{X \geq a\})$...

4) Variance :

a) Définition :

(i) Idée : on veut mesurer l'écart d'une v.a. à sa moyenne. Pour cela on pourrait utiliser $|X - E(X)|$. Cependant, à cause de la commodité des distances euclidiennes (pour les problèmes de minimisation (cf. **H1**), on préfère considérer $(X - E(X))^2$. On veut savoir combien vaut cet écart « en moyenne », d'où la déf. :

(ii) Déf. : Soit X une v.a. réelle. On définit sa *variance* $V(X)$ par :

$$\boxed{V(X) = E\left((X - E(X))^2\right).}$$

b) Comment on calcule en pratique la variance :

En pratique, on utilise plutôt la formule suivante pour le calcul de $V(X)$.

(i) **Caract. (Formule de König-Huygens)** Avec les notations précédentes :

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

Preuve facile : linéarité de l'espérance.

(ii) **Remarque :** *Une façon d'interpréter l'inégalité $E(X)^2 \leq E(X)^2$ avec Cauchy-Schwarz*
On considère la FBS symétrique positive sur les v.a. X, Y : $(X|Y) = \sum_{\omega\in\Omega} X(\omega)Y(\omega)P(\{\omega\})$ (p.s. à poids).
On applique alors Cauchy-Schwarz à $(X|1)$.
De la sorte cette inég. sur l'espérance est l'analogue de $(\int_0^1 f)^2 \leq \int_0^1 f^2$.

(iii) **Prop. Variance d'une fonction affine de X** Si $(a, b) \in \mathbb{R}^2$ est un couple de réels et si X est une v.a. réelle alors :

$$\boxed{V(aX + b) = a^2 V(X).}$$

c) Calcul de variances de lois usuelles :

(i) Loi uniforme $\mathcal{U}([\![1, n]\!])$: pour $X \sim \mathcal{U}([\![1, n]\!])$.

On obtient $V(X) = \dfrac{n^2 - 1}{12}$.

Remarque : Comment en déduire $V(Y)$ si $Y \sim \mathcal{U}([\![a, b]\!])$?

Idée : A partir de Y, on peut par translation se ramener au premier cas, et on sait que la variance ne change pas par translation.

(ii) Loi de Bernoulli

Prop. Si $X \sim \mathcal{B}(p)$ alors $V(X) =$

(iii) Loi binomiale : première approche

Plutôt que de calculer $E(X^2)$, il est ici plus commode de calculer $E(X(X-1))$.

En effet $E(X^2) = \sum_{k=0}^{n} k^2 \binom{n}{k} p^k (1-p)^{n-k}$, alors que $E(X(X-1)) = \sum_{k=0}^{n} k(k-1)\binom{n}{k} p^k (1-p)^{n-k}$.

Cependant, la méthode la plus commode, comme pour l'espérance, est de voir X comme $X = X_1 + \cdots + X_n$ avec X_i v.a. de Bernoulli indépendantes : ceci sera fait au § IV.

5) Ecart type et v.a. centrée réduite :

a) Déf. L'écart type de X est par déf. donné par $\sigma(X) = \sqrt{V(X)}$.

Pourquoi est-il naturel de prendre la *racine carrée* de la variance ?

b) (i) Déf. Une v.a. réelle X telle que $E(X) = 0$ et $V(X) = 1$ est appelée une *v.a. centrée réduite.*

(ii) Prop. Si X est une v.a. de variance non nulle, alors la v.a. $X^* = \dfrac{X - E(X)}{\sigma(X)}$ est centrée réduite.

6) Contrôle par la variance : inégalité de Bienaymé-Tchebychev

Irenée-Jules Bienaymé, mathématicien français (1796–1878)

Prop. Toute v.a. réelle X sur un espace probabilisé fini (Ω, P) vérifie :

$$\forall\, \varepsilon > 0,\ P(|X - E(X)| \geq \varepsilon) \leq \frac{V(X)}{\varepsilon^2}.$$

Idée de la preuve : appliquer l'inégalité de Markov à

Applications : cf. § V et planche.

IV Covariance :

1) Définition, calcul :

a) **Déf.** Soit X et Y deux v.a. réelles sur un même espace probabilisé fini (Ω, P). On appelle *covariance* de X et Y et on note $\mathrm{Cov}(X,Y)$ l'espérance du produit des variables centrées associées à X et Y autrement dit :

$$\mathrm{Cov}(X,Y) = E\left((X - E(X)).(Y - E(Y))\right).$$

b) **Prop. (une première formule pratique pour le calcul)** Avec les hyp. du a), on a l'égalité :

$$\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y).$$

c) **Remarque pour le calcul de $E(XY)$** : conséquence du théorème de transfert

Avec les mêmes hyp. :

$$E(XY) = \sum_{(x,y)\in X(\Omega)\times Y(\Omega)} xyP(X = x, Y = y).$$

2) Propriétés de la covariance :

a) Forme bilinéaire symétrique positive :

Prop. L'application $(X,Y) \mapsto \mathrm{Cov}(X,Y)$ est une F.B.S. positive sur l'e.v. des variables aléatoires réelles

Remarque : retenir aussi que par déf. $\mathrm{Cov}(X,X) = V(X)$.

b) Application aux calculs de variance :

(i) **Prop. (variance d'une somme de deux v.a.)** Soit (X,Y) un couple de v.a. réelles. Alors

$$V(X+Y) = V(X) + V(Y) + 2\,\mathrm{Cov}(X,Y).$$

Remarque : Plus généralement, par prop. déjà connues de la variance et la covariance, pour toute combinaison linéaire :

$$V(aX + bY) = a^2 V(X) + b^2 V(Y) + 2ab\,\mathrm{Cov}(X,Y).$$

(ii) **Généralisation pour n v.a.** De même si $X_1, \dots, X_n$ sont n v.a. réelles :

$$V(X_1 + \cdots + X_n) = \sum_{k=1}^{n} V(X_k) + 2 \sum_{1 \le i < j \le n} \mathrm{Cov}(V_i, V_j).$$

c) Application au contrôle de la covariance : inégalité de Cauchy-Schwarz
(i) Idée :

La preuve standard de l'inég. de Cauchy-Schwarz s'applique aux FBS positives, on perd seulement la CNS d'égalité

(ii) On a donc ici :
(Cauchy-Schwarz) si X et Y sont deux v.a. réelles :

$$|\mathrm{Cov}(X, Y)| \le \sigma(X).\sigma(Y).$$

(iii) **Remarque :** *exercice*
On peut déterminer ici aussi une C.N.S d'égalité : l'inég. du (ii) est une égalité ssi il existe $(a, b) \in \mathbb{R}^2$ tels que $P(Y = aX + b) = 1$ autrement dit l'égalité $Y = aX + b$ est « presque sûre ».
3) Cas de v.a. indépendantes :
a) Prop. Si X et Y sont deux v.a. indépendantes alors $\mathrm{Cov}(X, Y) = 0$ et donc par le résultat du 2) b) :

$$V(X + Y) = V(X) + V(Y)$$

b) Déf. Deux v.a. X et Y telles que $\mathrm{Cov}(X, Y) = 0$ sont dites *non corrélées.*

La prop. du a) dit que des v.a. indép. sont non corrélées, la récip. est fausse.

c) Application du a) Si $X_1, \dots, X_n$ sont des v.a. *deux à deux* indépendantes, alors :

$$V(X_1 + \cdots + X_n) = V(X_1) + \cdots + V(X_n).$$

d) Application au calcul de la variance pour une v.a. X de loi $\mathcal{B}(n, p)$
En écrivant $X = X_1 + \cdots + X_n$ avec X_i mutuellement indép. suivant $\mathcal{B}(p)$, on a :

$$V(X) = \sum_{i=1}^{n} V(X_i) = \sum_{i=1}^{n} p(1-p) = np(1-p).$$

V Application à la loi faible des grands nombres :
On considère une suite (X_n) de v.a.i.i.d. (mutuellement indépendantes identiquement distribuées i.e. de même loi qui est celle d'une v.a. notée X pour simplifier) sur un espace probabilisé fini (Ω, P).
On note $S_n = X_1 + \cdots + X_n$ de sorte que S_n/n représente la moyenne des valeurs de X_i. **Idée :** Les différentes versions de la loi des grands nombres disent que S_n/n tend, en des sens divers vers $E(X)$. Autrement dit, l'espérance abstraite correspond, pour n grand, à la valeur moyenne statistique des X_i

Formalisation : *exercice fait en classe.*
a) Pour $X_1, \dots, X_n$ des v.a. indépendantes, que dire de $V(X_1 + \cdots + X_n)$?
b)
c) En déduire le **Théorème** suivant, appelé *loi faible des grands nombres* : avec les hypothèses et notations précédentes :

$$\forall\, \varepsilon > 0,\ P\left(\left|\frac{S_n}{n} - E(X)\right| < \varepsilon\right) \underset{n \to +\infty}{\longrightarrow} 1.$$