
Probabilités - Chapitre 2

Variables aléatoires finies

On suppose dans tout le chapitre qu'il y a un espace probabilisé (Ω, P) avec lequel on travaille, sans le préciser. On suppose en outre que l'univers Ω est de cardinal fini.

I Généralités :

1) Variables aléatoires :

a) Définition :



Définition :

Soit E un ensemble. On appelle **variable aléatoire** toute application X de Ω dans E .

$$X : \begin{array}{l} \Omega \rightarrow E \\ \omega \mapsto X(\omega) \end{array}$$

Le plus souvent, E sera une partie de \mathbb{R} ou de \mathbb{C} , mais on peut aussi avoir des variables aléatoires à valeurs dans \mathbb{R}^n ...

Exemples :

- ▶ Pour modéliser un jeu de "pile" ou "face", avec $\Omega = \{\text{"pile"}, \text{"face"}\}$, on peut introduire une variable aléatoire X définie par $X(\text{"pile"}) = 1$ et $X(\text{"face"}) = 0$.
- ▶ Le lancer d'un dé 6 faces introduit naturellement une variable aléatoire : pour $i \in \llbracket 1, 6 \rrbracket$, $X(\text{"obtenir le chiffre } i\text{"}) = i$.
- ▶ Un lancer de deux dés peut-être traduit par une variable aléatoire

$$X : \omega \mapsto (X_1(\omega), X_2(\omega))$$

où X_1 et X_2 donnent respectivement les résultats du premier dé et du second. On parle alors de **couple de variables aléatoires**.

- ▶ Si on veut modéliser la somme de deux dés, on peut construire une variable aléatoire $S = X_1 + X_2$ où X_1 et X_2 sont des variables modélisant le premier et le second dé....

b) Univers image



Définition :

On appelle **univers image** d'une variable X , noté $X(\Omega)$, l'image de Ω par X . Ainsi,

$$X(\Omega) = \{X(\omega), \omega \in \Omega\}$$

L'univers image est donc l'ensemble des valeurs possibles pour la variable aléatoire.

Exemples :

- ▶ pour pile ou face : $X(\Omega) =$
- ▶ Pour un lancer de deux dés : $X(\Omega) =$
et pour la somme $S(\Omega) =$
- ▶ Soit $c \in \mathbb{R}$ et $X : \Omega \rightarrow \mathbb{R}$ définie pour tout ω par $X(\omega) = c$. Alors $X(\Omega) = \{c\}$. Une telle variable est dite "certaine".

c) Événements associés à une variable aléatoire :

Soit $B \subset E$. On s'intéresse aux résultats de l'expérience qui donnent, par X , une valeur dans B . Autrement dit, on cherche tous les antécédents de B par X (ce qu'on pourrait noter $X^{-1}(B)$). Cet ensemble est sous-ensemble de Ω , donc un événement :



Définition :

Soit X une variable aléatoire réelle et $B \subset \mathbb{R}$. On note $\{X \in B\}$ ou $(X \in B)$ l'événement suivant :

$$(X \in B) = \{\omega \in \Omega \text{ tel que } X(\omega) \in B\}$$

Dans le même esprit, on introduit d'autres notations :

$$(X = a) \text{ pour}$$

$$(X < a) \text{ pour}$$

$$(a \leq X \leq b) \text{ pour}$$

etc.

Par exemple : pour un lancer de deux dés, supposons qu'on s'intéresse à tous les lancers dont la somme est supérieure à 10. Ainsi, $B = [10, +\infty[$ (ou plus simplement $B = \{10, 11, 12\}$). L'événement qui nous intéresse est donc noté

$$(S \in \{10, 11, 12\})$$

ou plus naturellement :

$$(S \geq 10)$$



Proposition 1 :

Soit X une variable aléatoire à univers image fini avec $X(\Omega) = \{x_1, x_2, \dots, x_n\}$ (on suppose les x_i distincts)

Alors les événements $(X = x_1), (X = x_2), \dots, (X = x_n)$ forment un système complet d'événements.

▷ *Preuve* :

◁



REMARQUE :

Le système complet d'événements associé à une variable aléatoire est le plus "fin" que l'on puisse imaginer. C'est à dire que si on connaît tous les $P(X = x_k)$ alors on peut calculer $P(X \in B)$ pour n'importe quel ensemble B en procédant par union.

2) Loi d'une variable aléatoire

a) Définition



Définition :

Soit X une variable aléatoire.

On appelle **loi de X** l'application notée P_X définie pour tout $A \in \mathcal{P}(X(\Omega))$ par

$$P_X(A) = P(X \in A)$$



Propriété 1 :

La loi d'une variable aléatoire est une probabilité sur $X(\Omega)$.

▷ *Preuve* :

◁

En pratique, comme $X(\Omega) = \{x_1, x_2, \dots, x_n\}$ où $n = \text{card}(X(\Omega))$, on a le résultat ci dessous :



Proposition 2 :

Soit X une variable aléatoire avec $X(\Omega) = \{x_1, x_2, \dots, x_n\}$.

Alors la loi de X est entièrement déterminée par la distribution de probabilité

$$(P(X = x_k))_{k \in \llbracket 1, n \rrbracket}$$

▷ *Preuve* : Tout $A \in \mathcal{P}(X(\Omega))$ est une union d'événement de la forme $\{x\}$ avec $x \in X(\Omega)$. Ainsi, on aura $P_X(A) = P(X \in A) = \sum_{x \in A} P(X = x)$ et donc P_X est entièrement déterminée. ◁

Autrement dit, donner la loi d'une variable aléatoire discrète, c'est donner les probabilités des événements du système complet d'événements associé à la variable.

Exemple :

Un forain propose le jeu suivant : pour jouer à ce jeu, on paye 5€. On lance une pièce de monnaie deux fois de suite. La pièce est équilibrée.

- Si la pièce donne deux fois le même résultat, le forain nous donne 6€.
- Si pile sort en premier, puis face, c'est perdu.
- Si face sort en premier, puis pile, le forain nous donne 8 €

Quelle est la loi du gain (mise comprise) ?

Représentation

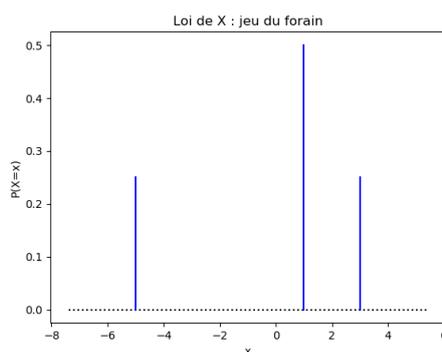
Donner la distribution de probabilité, c'est à dire la liste des $(P(X = x_k))$, suffit pour expliciter une loi.

Dans le cas de variables aléatoires finies à valeur réelles ou complexe, si il n'y a pas de formule, on appréciera de donner un tableau récapitulatif.

Par exemple pour le jeu précédent, on peut résumer la loi de X via le tableau suivant :

| | | | |
|------------|---------------|---------------|---------------|
| x | -5 | 1 | 3 |
| $P(X = x)$ | $\frac{1}{4}$ | $\frac{1}{2}$ | $\frac{1}{4}$ |

On peut aussi faire un diagramme en bâtons représentant la distribution de probabilité de la loi : en chaque x_k possible, on trace un trait vertical de hauteur $P(X = x_k)$:



NOTATION

Si X et Y sont deux variables aléatoires qui ont même loi, on pourra utiliser la notation $X \sim Y$.

À noter :

ET SI X EST À VALEUR VECTORIELLE ?

La définition de loi, même si elle parle d'un $x \in X(\Omega)$, ne précise nulle part que x serait réel ou complexe... il peut donc tout à fait s'agir de vecteur !

Par exemple, si X est le couple (X_1, X_2) où X_1 et X_2 sont les résultats d'un dé, alors $X(\Omega) = \llbracket 1, 6 \rrbracket^2$.

Ainsi, si $x \in X(\Omega)$, alors x est un couple (x_1, x_2) ... et $P(X = x)$ signifie $P((X_1, X_2) = (x_1, x_2))$, ou encore $P(X_1 = x_1 \cap X_2 = x_2)$.

Une partie de ce cours sera consacrée à ces cas particuliers...

b) Loi de Bernoulli :



Définition :

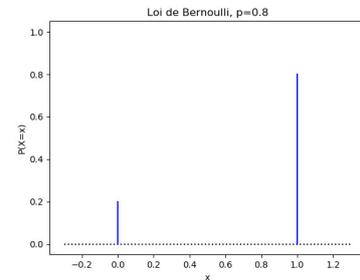
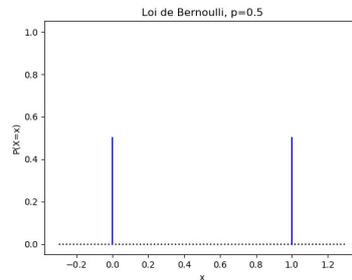
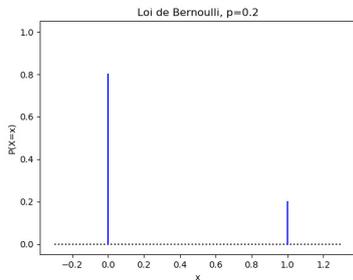
Soit $p \in]0, 1[$. On appelle **variable de Bernoulli** de paramètre p toute v.a.r X vérifiant :

(i) $X(\Omega) = \{0, 1\}$

(ii) Sa loi est donnée par $P(X = 1) = p$ et $P(X = 0) = 1 - p$

Notation : $X \sim \mathcal{B}(p)$

On peut donner son diagramme en bâton :



Méthode :

QUAND UTILISER UNE LOI DE BERNOULLI ?

On utilise typiquement cette loi dans la modélisation d'une expérience n'admettant que deux résultats : succès/echec, positif/negatif, etc.

Par exemple pour pile ou face, avec $X(\text{"pile"}) = 1$ et $X(\text{"face"}) = 0$ et une probabilité p d'avoir "pile", on utilisera naturellement une variable $X \sim \mathcal{B}(p)$.

c) Loi uniforme :

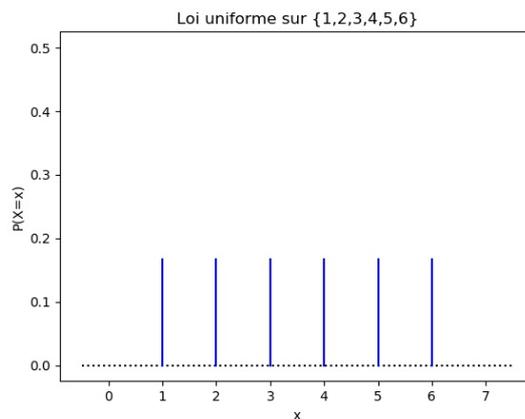


Définition :

On dit qu'une variable aléatoire X suit une **loi uniforme** sur un ensemble fini E , et on note $X \sim \mathcal{U}(E)$, si et seulement si

(i) $X(\Omega) = E$

(ii) Pour tout $x \in E$, $P(X = x) = \frac{1}{n}$ où $n = \text{card}(E)$.



REMARQUE :

Fréquemment, on utilisera des lois uniformes sur des intervalles d'entier $\llbracket a, b \rrbracket$, qu'on notera donc $X \sim \mathcal{U}(\llbracket a, b \rrbracket)$

On aura alors pour tout $k \in \llbracket a, b \rrbracket$, $P(X = k) = \frac{1}{b - a + 1}$

**Méthode :****UTILISATION DE LA LOI UNIFORME**

Cette loi intervient dans des situations semblables à celle utilisant la probabilité uniforme pour modéliser des événements équiprobables.

Par exemple pour modéliser un dé 6 faces équilibré, on utilisera une variable

$$X \sim \mathcal{U}(\llbracket 1; 6 \rrbracket)$$

3) Opérations sur les variables aléatoires**a) Opérations usuelles**

Les variables aléatoires sont des applications, à valeur dans un ensemble E .

Si cet ensemble est par exemple inclus dans \mathbb{R} ou \mathbb{C} , on peut définir naturellement toutes les opérations usuelles qu'on peut avoir dans \mathbb{R} ou \mathbb{C} :

**Propriété 2 :**

Soient X et Y deux variables aléatoires réelles et soit $\alpha \in \mathbb{R}$. Alors :

- (i) $X + Y$ est une variable aléatoire.
- (ii) αX est une variable aléatoire.
- (iii) XY est une variable aléatoire.
- (iv) Si Y n'est jamais nulle, X/Y est une variable aléatoire.

On peut également composer par une fonction :

**Propriété 3 :**

Soit X une variable aléatoire et f une fonction définie sur $X(\Omega)$ à valeur dans un ensemble F .

Alors $f \circ X : \omega \mapsto f(X(\omega))$ est une variable aléatoire à valeur dans F , que l'on note simplement $f(X)$.

**Méthode :****OPÉRATIONS ET LOI**

Il n'y a, en général, pas de formule vraiment pratique pour donner la loi de $X + Y$, de XY ou de $f(X)$: il faut être "astucieux", et on verra des exemples en TD.

Néanmoins, pour $f(X)$, on peut donner la méthode suivante :

1. Commencer par déterminer $F = f(X)(\Omega)$
2. Pour tout $y \in F$, pour tout $\omega \in \Omega$, $f(X)(\omega) = y \Leftrightarrow X(\omega) \in f^{-1}(\{y\})$

On en déduit que $(f(X) = y) = \bigcup_{x \in f^{-1}(\{y\})} (X = x)$ et donc

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

Exemple

Soit $X \sim \mathcal{U}(\llbracket -1, 2 \rrbracket)$.

Quelle est la loi de $Y = 3X + 1$?

Toujours avec $X \sim \mathcal{U}([-1, 2])$, quelle est la loi de $Z = (X - 1)^2$?

Remarque :

Si $X \sim Y$ (c'est à dire si X et Y on même loi), il est clair que $f(X) \sim f(Y)$...

b) Somme de variables aléatoires de Bernoulli

Problème considéré :

Soit $n \in \mathbb{N}^*$. On lance n fois et de manière indépendante une pièce de monnaie qui donne "face" avec la probabilité p . On compte le nombre de faces obtenus, et on note X ce nombre.

Quelle est la loi de X ?

Traduction comme somme de v.a.

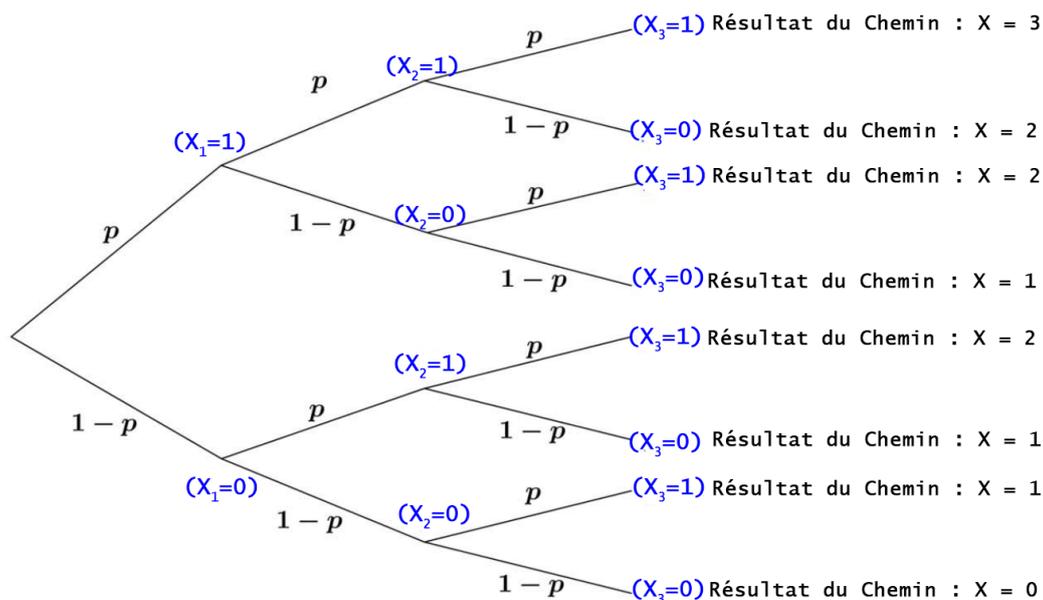
On modélise le résultat du i ème lancé avec une variable aléatoire $X_i \sim \mathcal{B}(p)$: l'événement $(X_i = 1)$ correspond à "obtenir face au i eme lancer", et $(X_i = 0)$ à "obtenir pile au i eme lancer".

La variable X recherchée est donc $X = \sum_{i=1}^n X_i$

Remarquons que comme les lancers sont indépendants, si $i \neq j$, $(X_i = 1)$ et $(X_j = 1)$ sont indépendants (idem avec $(X_i = 0)$ et $(X_j = 1)$, etc)

Détermination de la loi de X :

On peut également voir cela sous forme de “schéma de Bernoulli” et dénombrer les chemins qui nous intéressent :



Le raisonnement est identique : tous les chemins contenant k fois 1 ont même probabilité de survenir, et le nombre de tels chemins correspondent au nombre de façon de choisir de monter k fois parmi les n étapes du schéma.

c) Variable aléatoire binomiale

L'exemple précédent nous permet d'introduire une dernière loi :



Définition :

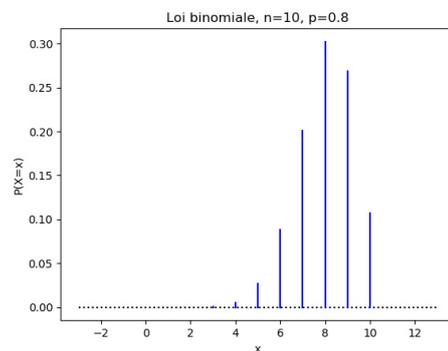
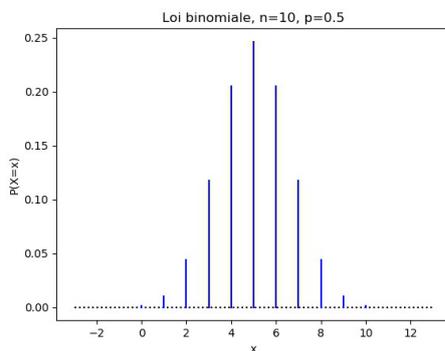
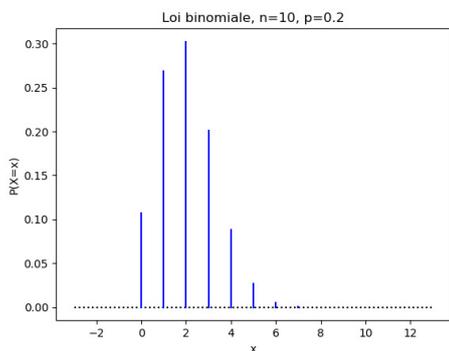
Soit X une variable aléatoire, $n \in \mathbb{N}^*$ et $p \in]0, 1[$.

On dit que X suit la **loi binomiale de paramètres n et p** et on note $X \sim \mathcal{B}(n, p)$ si et seulement si :

(i) $X(\Omega) = \llbracket 0, n \rrbracket$

(ii) $P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$

Quelques diagrammes en bâton :



Méthode :

QUAND UTILISER LA LOI BINOMIALE :

La loi binomiale s'utilise quand il s'agit d'étudier la répétition avec indépendance d'une même situation "succès/échec" où l'on compte les succès.

L'hypothèse d'indépendance est très importante dans cette répétition, comme le calcul fait en préambule l'a montré...

Exemple :

Un chapeau contient 4 lapins angora et 9 lapins de garenne. On extrait dix fois un lapin du chapeau, en remettant le lapin obtenu entre chaque tirage.

Soit N le nombre de lapin angora extrait du chapeau.

Quelle est la loi de N ? Quelle est la probabilité d'obtenir une seule fois un lapin angora ?

II Espérance et Variance

Deux caractéristiques sont souvent étudiées pour les variables aléatoires et correspondent à des besoins concrets en statistique et en analyse de données.

1) Espérance

a) Définition

**Définition :**

Soit $X : \Omega \rightarrow E$ une v.a.r. avec $X(\Omega)$ un ensemble fini et E un espace vectoriel.

On appelle **espérance** de X la quantité noté $E(X)$ définie par

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

**REMARQUE :**

L'espérance d'une v.a. est la moyenne pondérée des résultats possibles : les coefficients sont les probabilités d'avoir ces résultats.

On dit que l'espérance est une caractéristique de "position" de la variable aléatoire : c'est la valeur moyenne de la variable.

Exemple

On reprend l'exemple du jeu du forain :

b) Espérance des lois usuelles :

Exemple de loi uniforme :

On lance un dé 6 faces équilibrés. Quel est le résultat moyen obtenu ?

Cas d'une Bernoulli

Soit $X \sim \mathcal{B}(p)$, alors

Ainsi :

| |
|---------------------------------------------|
| Si $X \sim \mathcal{B}(p)$, alors $E(X) =$ |
|---------------------------------------------|

Cas d'une binomiale :

Soit $X \sim \mathcal{B}(n, p)$, alors

Ainsi (à retenir) :

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

Cas d'une loi certaine :

Soit X suivant une loi certaine, c'est à dire qu'il existe $C \in \mathbb{R}$ tel que $X(\omega) = C$ pour tout ω .
Alors $E(X) = CP(X = C) = C$.

Ainsi (à retenir) :

| |
|---------------------------------------|
| Si X est certaine, alors $E(X) = X$ |
|---------------------------------------|

2) Propriétés de l'espérance :

a) Théorème du transfert :

Soit X une v.a.r. et f une fonction à valeur définie sur un ensemble contenant $X(\Omega)$.

On souhaite calculer $E(f(X))$.

D'après la définition, il nous faut trouver la loi de $f(X)$, c'est à dire commencer par chercher $f(X)(\Omega)$, puis calculer

$$E(f(X)) = \sum_{y \in f(X)(\Omega)} y P(f(X) = y)$$

Sauf que loi de $f(X)$ n'est pas simple! Heureusement, on a le résultat ci dessous :



Theorème 1 : Theorème du transfert

Soit X une variable aléatoire et $f : X(\Omega) \rightarrow E$ une fonction. Alors :

$$\begin{aligned} E(f(X)) &= \sum_{x \in X(\Omega)} f(x) P(X = x) \\ &= \sum_{k \in I} f(x_k) P(X = x_k) \end{aligned}$$

▷ *Preuve* :

Remarquons déjà que pour tout $y \in f(X)(\Omega)$, l'événement $(f(X) = y)$ correspond à l'événement " X est un des antécédents de y ", c'est à dire $(X \in f^{-1}(\{y\}))$.

Comme nos variables ont un nombre fini de valeurs, on peut donc écrire ce dernier ensemble sous la forme $\bigcup_{x \in f^{-1}(\{y\})} (X = x)$.

$$\text{Donc } P(f(X) = y) = \sum_{x \in f^{-1}(\{y\})} P(X = x)$$

Ainsi,

$$\begin{aligned} E(f(X)) &= \sum_{y \in f(X)(\Omega)} y P(f(X) = y) \\ &= \sum_{y \in f(X)(\Omega)} y \sum_{x \in f^{-1}(\{y\})} P(X = x) \\ &= \sum_{y \in f(X)(\Omega)} \sum_{x \in f^{-1}(\{y\})} y P(X = x) \end{aligned}$$

Or si $x \in f^{-1}(\{y\})$, alors $y = f(x)$ et donc

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

La somme est faite pour tout les y possibles, puis pour tout les $x \in f^{-1}(\{y\})$. Or les $f^{-1}(\{y\})$ forment une partition de tous les $x \in X(\Omega)$: la somme est donc finalement faite sur tout les $x \in X(\Omega)$!

Ainsi

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

◁

Exemple :

Soit $X \sim \mathcal{U}([-1, 2])$. On veut calculer $E(X^2)$:



REMARQUES :

- Grâce à ce théorème, c'est $P(X = x_k)$ dans la formule, pas $P(f(X) = y_k)$. C'est de là que vient le nom "théorème du transfert" : le f , au lieu d'être dans la probabilité, est "transféré" sur la valeur devant la proba... Il suffit donc d'avoir la loi de X , peu importe f .
- Le théorème du transfert est par exemple utilisé pour calculer les "moments d'ordre n " d'une variable aléatoire réelle ou complexe, c'est à dire $E(X^n)$. Les "moments" sont un peu comme les dérivée n ième pour les v.a.r. et il existe des formules (un peu comme les formules de Taylor...) qui permettent de retrouver la loi de X à partir de ces moments...
- Ce théorème, ainsi que sa preuve, n'impose pas à ce que la variable soit à valeurs réelles : on peut donc tout à fait utiliser des fonctions f définie sur un espace vectoriel à valeur dans un espace vectoriel par exemple ! Tant qu'on peut faire des sommes avec les éléments de E , l'espérance à un sens....

b) Linéarité :



Theorème 2 :

Soient X et Y deux variables aléatoires à valeur dans le même ensemble E .
Alors pour tout λ et μ réel, on a

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

▷ *Preuve* : Posons $Z = (X, Y)$: alors Z est à valeur dans E^2 .

On pose $f : \begin{matrix} E^2 & \rightarrow & E \\ (x, y) & \mapsto & \lambda x + \mu y \end{matrix}$.

Alors $E(\lambda X + \mu Y) = E(f(Z)) = \sum_{z \in E^2} f(z)P(Z = z)$

En notant $z = (x, y)$, on a alors $P(Z = z) = P((X, Y) = (x, y)) = P(X = x \cap Y = y)$
D'où

$$\begin{aligned} E(f(Z)) &= \sum_{(x,y) \in E^2} (\lambda x + \mu y)P(X = x \cap Y = y) \\ &= \lambda \sum_{(x,y) \in E^2} xP(X = x \cap Y = y) + \mu \sum_{(x,y) \in E^2} yP(X = x \cap Y = y) \end{aligned}$$

Enfin,

$$\begin{aligned} \sum_{(x,y) \in E^2} xP(X = x \cap Y = y) &= \sum_{x \in E} \sum_{y \in E} xP(X = x \cap Y = y) \\ &= \sum_{x \in E} x \underbrace{\sum_{y \in E} P(X = x \cap Y = y)}_{=P(X=x)(\text{proba totale})} \\ &= E(X) \end{aligned}$$

De même $\sum_{(x,y) \in E^2} yP(X = x \cap Y = y) = E(Y)$ d'où

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

◀

**Corolaire 1 :**

Soit $n \geq 2$ et soient X_1, X_2, \dots, X_n n variables aléatoires, alors

$$E\left(\sum_{k=1}^n X_k\right) = \sum_{k=1}^n E(X_k)$$

▷ *Preuve* : C'est une récurrence facile, en regroupant les sommes. ◁

Exemples

► Soit X une variable aléatoire et $m = E(X)$. En posant $Y = X - m$ on obtient

$$E(Y) = E(X) - E(m) = m - m = 0$$

On dit alors que la variable Y est “centrée”.

► Soit $X \sim \mathcal{B}(n, p)$. Alors on peut écrire que $X = \sum_{k=1}^n X_k$ avec $X_k \sim \mathcal{B}(p)$

Ainsi $E(X) =$

c) Croissance de l'espérance**Propriété 4 :**

Soient X et Y deux variables aléatoires telles que pour tout $\omega \in \Omega$, $X(\omega) \leq Y(\omega)$.
Alors $E(X) \leq E(Y)$

▷ *Preuve* :

◁

3) Variance**a) Variance et écart type****Définition :**

On appelle **variance** d'une v.a.r X le “moment centré d'ordre 2”, c'est à dire

$$V(X) = E((X - E(X))^2)$$

On appelle **écart-type** de X la quantité $\sigma_X = \sqrt{V(X)}$.

**REMARQUE :**

Interprétation

- $|X - E(X)|$ donne la distance de X à $E(X)$. Cette distance se manipule assez mal (à cause de la valeur absolue...) on préfère regarder cette quantité au carré : $(X - E(X))^2$, puis regarder la valeur moyenne de cette distance quadratique : $E((X - E(X))^2)$.
Ainsi, la variance traduit la tendance d'une v.a.r. à s'écarter de sa moyenne, à "varier" autour de sa moyenne.
Si on prend la racine de cette distance quadratique, on retombe sur une distance "tout court", qui traduit l'écart moyen ("écart type") entre X et sa moyenne.
- La variance est une caractéristique dite de "dispersion". Elle traduit la façon dont une variable aléatoire se disperse autour de la moyenne.

**Proposition 3 : Formule de Koenig-Huygens**

Soit X une v.a.r., alors on a

$$V(X) = E(X^2) - E(X)^2$$

▷ *Preuve* :

◁

b) Exemples**Variance d'une Bernoulli**

Si $X \sim \mathcal{B}(p)$, $E(X) = p$, donc

A retenir :

$$\boxed{\text{Si } X \sim \mathcal{B}(p), V(X) = p(1 - p)}$$

Variance d'une loi Binomiale

Le calcul de la variance d'une loi binomiale ressemble beaucoup à ce qu'on a fait pour l'espérance. On admet que

$$\boxed{\text{si } X \sim \mathcal{B}(n, p), \text{ alors } V(X) = np(1 - p)}$$

Variance d'une loi certaine :**Proposition 4 :**

Soit X une v.a.r. Alors X est une variable certaine si et seulement si $V(X) = 0$

▷ *Preuve* :

◁

c) Opérations et variance :



Proposition 5 :

Soit X une v.a.r et $a, b \in \mathbb{R}$, alors $V(aX + b) = a^2V(X)$

▷ Preuve :

◁

Exemple d'application :



Propriété 5 :

Soit X une v.a.r. admettant une espérance m et un écart type $\sigma_X \neq 0$,
alors $X^* = \frac{X - m}{\sigma_X}$ a une espérance nulle et un écart type de 1.
Une telle variable est dite “centrée réduite”

▷ Preuve :

◁

d) Inégalité de Bienaymé-Tchebitchev

L'inégalité suivante permet d'encadrer, en terme de probabilité, l'espérance :



Proposition 6 :

Soit X une v.a. d'espérance m et d'écart type $\sigma \neq 0$, alors

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

Avant de démontrer cette proposition, commençons par l'interpréter :

Notons déjà que $|X - m| \geq \varepsilon$ signifie que X s'éloigne de plus de ε de sa moyenne.

Cette inégalité signifie que la probabilité que cela ait lieu est majorée par une quantité dépendant de la variance.

Concrètement, plus la variance est petite, moins cela va arriver.

On retrouve ainsi cette notion de dispersion.

Pour montrer cette inégalité, on va d'abord montrer l'inégalité de Markov :

 **Proposition 7 : Inégalité de Markov**

Soit Y une variable aléatoire positive admettant une espérance, et soit $a > 0$.

Alors $P(Y \geq a) \leq \frac{E(Y)}{a}$

III Couples de variables aléatoires

1) Généralités

a) Définition :



Définition :

Soit $(\Omega, \mathcal{P}(\Omega), P)$ un espace probabilisé fini. Un **couple de variables aléatoires** est une application $Z : \begin{cases} \Omega & \rightarrow E^2 \\ \omega & \mapsto (X(\omega), Y(\omega)) \end{cases}$

On note simplement $Z = (X, Y)$, et X et Y sont donc des variables aléatoires "classiques". L'univers image de Z est $Z(\Omega) = \{(X(\omega), Y(\omega)); \omega \in \Omega\} \subset X(\Omega) \times Y(\Omega)$: c'est l'ensemble des couples de résultats possibles formés à partir des résultats de X et de Y .

b) Exemple :

On lance 2 dés : un dé 6 faces et un dé 4 faces. Soit X la variable modélisant le dé 6 faces et Y modélisant le dé 4 faces. La v.a. qui modélise le résultat du lancer des deux dés est un couple $Z = (X, Y)$ avec $Z(\Omega) = \llbracket 1; 6 \rrbracket \times \llbracket 1; 4 \rrbracket$.

2) Loi d'un couple de variables aléatoires / loi conjointe :

a) Définition

La définition de loi ne change pas pour un couple, mais pour marquer le côté "couple" on parlera de **loi conjointe**.

Concrètement, on peut donner la caractérisation suivante :



Propriété 6 :

Soit $Z = (X, Y)$ un couple de variables aléatoires. La loi du couple est donnée par l'application

$$\begin{aligned} X(\Omega) \times Y(\Omega) &\rightarrow [0, 1] \\ (x, y) &\mapsto P(X = x, Y = y) \end{aligned}$$

où la notation $P(X = x, Y = y)$ signifie $P((X = x) \cap (Y = y))$

Autrement dit, donner la loi d'un couple de variables aléatoires, c'est donner les probabilités d'obtention de tous les couples possibles.

A noter qu'on ne part pas de $Z(\Omega)$ mais de $X(\Omega) \times Y(\Omega)$, quitte à donner des probabilités nulles.

On peut présenter la loi d'un couple de variables aléatoires sous la forme d'un tableau à double entrée :

| | | | |
|------------------|-----|-----------------------|-----|
| $X \backslash Y$ | ... | y_j | ... |
| \vdots | ... | ... | ... |
| x_i | ... | $P(X = x_i, Y = y_j)$ | ... |
| \vdots | ... | ... | ... |

b) Exemples

- On reprend l'exemple des deux dés. Toutes les combinaisons sont équiprobables, et il y a 24 possibilités (6 possibilités pour X , 4 pour Y). Ainsi, pour tout $(i, j) \in \llbracket 1; 6 \rrbracket \times \llbracket 1; 4 \rrbracket$, $P(X = i \cap Y = j) = \frac{1}{24}$.

- On lance 2 dés 4 faces de manière indépendante.

On note X le résultat du premier lancer, Y le second.

On pose $M = \max(X, Y)$ et on s'intéresse à la loi du couple $Z = (X, M)$.

Notons déjà que $X(\Omega) \times M(\Omega) = \llbracket 1; 4 \rrbracket \times \llbracket 1; 4 \rrbracket$.

| | | | | |
|-----------------|---|---|---|---|
| $X \setminus M$ | 1 | 2 | 3 | 4 |
| 1 | | | | |
| 2 | | | | |
| 3 | | | | |
| 4 | | | | |

3) Lois marginales

a) Définition



Définition :

Soit $Z = (X, Y)$ un couple de variables aléatoires. On dit que X et Y sont les **variables aléatoires marginales** de Z , et les lois de X et de Y sont appelées les **lois marginales**.

Exemple :

Dans le lancer de deux dés (un dé 6 faces, modélisé par X et un 4 faces modélisé par Y), le couple est $Z = (X, Y)$. Les lois marginales sont $X \sim \mathcal{U}(\llbracket 1; 6 \rrbracket)$ et $Y \sim \mathcal{U}(\llbracket 1; 4 \rrbracket)$.

b) De la loi conjointe aux lois marginales

On reprend l'exemple du couple $Z = (X, M)$. Les variables marginales sont X et M . Il est clair que $X \sim \mathcal{U}(\llbracket 1; 4 \rrbracket)$ mais quelle est la loi de M ?

L'idée est d'utiliser la loi conjointe en utilisant les probabilités totales avec le système complet d'événements $(X = 1), (X = 2), (X = 3)$ et $(X = 4)$:

$$P(M = k) = \sum_{i=1}^4 P(M = k, X = i)$$

Ainsi, $P(M = k)$ est obtenu en faisant la somme de la k -ième colonne dans le tableau de la loi conjointe. De la même façon, la loi de Y est obtenu en faisant la somme des lignes.

| | | | | | |
|-----------------|----------------|----------------|----------------|----------------|------------|
| $X \setminus M$ | 1 | 2 | 3 | 4 | Loi de X |
| 1 | $\frac{1}{16}$ | $\frac{1}{16}$ | $\frac{1}{16}$ | $\frac{1}{16}$ | |
| 2 | 0 | $\frac{2}{16}$ | $\frac{1}{16}$ | $\frac{1}{16}$ | |
| 3 | 0 | 0 | $\frac{3}{16}$ | $\frac{1}{16}$ | |
| 4 | 0 | 0 | 0 | $\frac{4}{16}$ | |
| Loi de M | | | | | |

Les lois des variables marginales apparaissent donc en marge du tableau de la loi du couple, d'où l'adjectif "marginal"...

On peut résumer ce résultat dans le théorème suivant :



Theorème 3 :

Soit $Z = (X, Y)$ un couple de variables aléatoires avec $X(\Omega) = \{x_1, x_2, \dots, x_n\}$, $Y(\Omega) = \{y_1, y_2, \dots, y_p\}$. Les lois marginales sont données par :

$$\text{somme de la } i\text{ème ligne : } \forall i \in \llbracket 1, n \rrbracket, P(X = x_i) = \sum_{j=1}^p P(X = x_i, Y = y_j)$$

$$\text{somme de la } j\text{ème colonne : } \forall j \in \llbracket 1, p \rrbracket, P(Y = y_j) = \sum_{i=1}^n P(X = x_i, Y = y_j)$$

▷ *Preuve* : On procède comme dans l'exemple en utilisant le système complet associé à Y pour déterminer la loi de X , et celui associé à X pour déterminer la loi de Y . ◁

IV Lien entre variables aléatoires

1) Loi conditionnelle

a) Définition

Soit X une variable aléatoire et soit A un événement de probabilité non nulle. On appelle **loi conditionnelle de X sachant A** l'application :

$$\begin{cases} X(\Omega) & \rightarrow [0, 1] \\ x & \mapsto P_A(X = x) = \frac{P(X = x \cap A)}{P(A)} \end{cases}$$

b) Exemple :

On dispose de deux dés : un dé 6 et un dé 20. On lance une pièce équilibrée. Si pile tombe, on lance le dé 6, si c'est face, on lance le dé 20. On note X le résultat obtenu. Soit A l'événement "obtenir pile". La loi de X sachant A est Sachant \bar{A} , X suit une loi

2) Covariance

a) Définition

On cherche à caractériser la façon dont deux variables X et Y s'éloignent simultanément de leur valeur moyenne.



Définition :

Soit (X, Y) un couple de v.a. On appelle **covariance** de X et de Y le réel (qui peut être positif, négatif ou nul)

$$cov(X, Y) = E((X - E(X))(Y - E(Y)))$$

En utilisant la linéarité de l'espérance, on a :

$$E((X - E(X))(Y - E(Y))) = E(XY - E(X)Y - E(Y)X + E(X)E(Y)) = E(XY) - 2E(X)E(Y) + E(X)E(Y) = E(XY) - E(Y)E(X)$$

Ce qui donne une formule plus simple :



Proposition 8 :

Formule de Koenig-Huygens :

$$cov(X, Y) = E(XY) - E(X)E(Y)$$

"espérance du produit - produit des espérances"

b) Exemple et interprétation

On reprend l'exemple (X, M) : on a déjà la loi conjointe, ce qui nous permet de calculer facilement $E(XM)$:

$$E(XM) = 1/16 + 2/16 + 3/16 + 4/16 + 8/16 + 6/16 + 8/16 + 27/16 + 12/16 + 64/16$$

$$E(XM) = 1/16(1 + 2 + 3 + 4 + 8 + 6 + 8 + 27 + 12 + 64) = 135/16$$

D'autre part, les lois marginales nous donne facilement les espérances :

$$E(M) = 1/16(1 + 6 + 15 + 28) = 40/16 \text{ et } E(X) = 5/2$$

$$\text{D'où } cov(X, M) = 135/16 - 100/16 = 35/16.$$

Interprétation :

Si la variance est positive, cela signifie que X et Y varient "dans le même sens" : si X est grand, Y est grand, et vice-versa. C'est le cas de cet exemple : si X est grand, le maximum M a plus de chance d'être grand. Une variance négative traduit une tendance à se comporter à l'inverse, et une variance nulle traduit un comportement équilibré.

c) Quelques propriétés

En utilisant la linéarité de l'espérance, on montre très facilement les propriétés suivantes :



Propriété 7 : Bilinéarité

Soient X, Y et Z trois variables aléatoires, soient α, β deux réels, alors :

i) $cov(X, Y) = cov(Y, X)$

ii) $cov(\alpha X, Y) = \alpha cov(X, Y)$ et $cov(X, \beta Y) = cov(X, \beta Y)$

iii) $cov(X + Y, Z) = cov(X, Z) + cov(Y, Z)$ et $cov(X, Y + Z) = cov(X, Y) + cov(X, Z)$



Propriété 8 : effet des translations

Soient X, Y deux variables aléatoires, soient a, b deux réels, alors :

$$cov(X + a, Y + b) = cov(X, Y)$$



Propriété 9 : Variance d'une somme

Soient X et Y deux variables aléatoires.

$$V(X + Y) = V(X) + V(Y) + 2cov(X, Y)$$

▷ Preuve :

3) Indépendance

a) Définition

 **Définition :**

On dit que deux variables aléatoires sont indépendantes si pour tout couple $(x, y) \in X(\Omega) \times Y(\Omega)$, $(X = x)$ et $(Y = y)$ sont indépendants.
Autrement dit $P(X = x, Y = y) = P(X = x)P(Y = y)$.

On note alors $X \perp\!\!\!\perp Y$.

Caractérisation via la loi conjointe :

La loi conjointe permet rapidement de voir si des variables sont indépendantes, puisque les cases donnent $P(X = x \cap Y = y)$: il suffit de vérifier que chaque case est le produit des valeurs en marge.

b) Composition

 **Proposition 9 :**

Soient X et Y deux variables aléatoires indépendantes, et f et g deux fonctions définies sur $X(\Omega)$ et $Y(\Omega)$. Alors $f(X)$ et $g(Y)$ sont indépendantes

▷ *Preuve* : Même idée que le théorème du transfert : on passe par les images réciproques... ◁

Exemple :

si X et Y indépendants, X et $-Y$ sont indépendants, X^2 et Y indépendants, $\cos(X)$ et e^Y indépendants, etc.

c) Indépendance et covariance :

Lemme 1

Si X et Y sont deux variables aléatoires indépendantes, alors $E(XY) = E(X)E(Y)$.

▷ *Preuve* :

◁

 **Proposition 10 : Indépendance et variance**

Si X et Y sont indépendantes, alors $cov(X, Y) = 0$ et $V(X + Y) = V(X) + V(Y)$.

▷ *Preuve* :

◁



REMARQUES :

Attention à la réciproque !

- On peut très bien avoir X et Y non indépendants, mais malgré tout avoir une covariance nulle...
- Si la covariance de X et Y est nulle, on dit que les variables sont **décorrélées**. Ce n'est pas la même chose qu'indépendant donc...

V Vecteurs aléatoires

Il s'agit d'une généralisation des couples aléatoires. Pas de nouveau concept, juste une extension rapide.

1) Généralités

a) Définitions

Soit $n \in \mathbb{N}$. On appelle **vecteur aléatoire** tout n -uplet $Z = (X_1, X_2, \dots, X_n)$ où X_1, X_2, \dots, X_n sont n variables aléatoires réelles.

On définit également la **loi conjointe** de Z comme étant l'application suivante :

$$\begin{cases} X_1(\Omega) \times \dots \times X_n(\Omega) & \rightarrow [0, 1] \\ (x_1, \dots, x_n) & \mapsto P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) \end{cases}$$

Le terme **loi marginale** désigne à nouveau les lois de X_1, X_2, \dots, X_n . On ne peut néanmoins plus utiliser de tableau pour représenter facilement les lois....

Remarque :

On définit de la même façon les "suites de variables aléatoires" quand on définit une application de \mathbb{N} dans l'ensemble des variables aléatoires : $n \mapsto X_n$

b) Exemples

Les vecteurs et suites interviennent dès qu'on répète une opération plus de 2 fois : lancer n fois d'une pièce de monnaie, lancer de 3 dés, etc.

2) Variance et espérance d'une somme de variables aléatoires

a) Espérance



Theorème 4 :

Soit X_1, X_2, \dots, X_n n variables aléatoires. Alors

$$E(X_1 + X_2 + \dots + X_n) = E(X_1) + E(X_2) + \dots + E(X_n)$$

b) Variance d'une somme de n variables aléatoires



Theorème 5 :

Soit X_1, X_2, \dots, X_n n variables aléatoires. Alors

$$V(X_1 + X_2 + \dots + X_n) = V(X_1) + V(X_2) + \dots + V(X_n) + 2 \sum_{1 \leq i < j \leq n} cov(X_i, X_j)$$

Autrement dit : on additionne les variances, puis tous les "doubles" covariances (même principe que $(a + b + c)^2$: somme des carrés et doubles produits)

▷ *Preuve* : (idée de la démo) : Par récurrence, même principe que pour l'espérance . ◁

3) Indépendance

a) Indépendance mutuelle



Définition :

On dit que n variables aléatoires X_1, X_2, \dots, X_n sont **mutuellement indépendantes** si et seulement si

$$\forall (x_1, x_2, \dots, x_n) \in X_1(\Omega) \times X_2(\Omega) \times \dots \times X_n(\Omega),$$

$$P(X_1 = x_1 \cap X_2 = x_2 \cap \dots \cap X_n = x_n) = P(X_1 = x_1)P(X_2 = x_2) \dots P(X_n = x_n)$$

Autrement dit : la loi conjointe est le produit des lois marginales.

Remarques :

1. En général, l'indépendance mutuelle est une hypothèse de travail : on demande rarement de la démontrer.
2. On omet souvent le mot "mutuelle" et on dit simplement que X_1, X_2, \dots, X_n sont indépendantes.
3. On parle parfois d'indépendance 2 à 2 lorsque les variables sont indépendantes lorsqu'on les prends 2 par 2 : c'est une hypothèse moins forte que l'indépendance mutuelle, car des variables peuvent être indépendantes deux à deux mais pas indépendantes mutuellement.

b) Sous familles et "collections"



Proposition 11 :

Soit X_1, X_2, \dots, X_n n variables aléatoires indépendantes. Alors toute sous famille est une famille de variables aléatoires indépendantes.

▷ *Preuve* : admis

◁

De ce théorème, on tire immédiatement :



Corolaire 2 :

L'indépendance mutuelle implique l'indépendance 2 à 2.

▷ *Preuve* : L'indépendance 2 à 2 signifie qu'en prenant une sous famille de deux éléments, cette famille est indépendante. Si la famille globale est indépendante, le théorème précédent donne directement le résultat.

◁

Enfin, le résultat suivant, parfois appelé "lemme des coalitions", est fréquemment utilisé :



Theorème 6 : lemme des coalitions

Soit X_1, X_2, \dots, X_n n variables aléatoires indépendantes. Soit $p < n$ et soit f et g deux fonctions définies respectivement sur $X_1(\Omega) \times \dots \times X_p(\Omega)$ et $X_{p+1}(\Omega) \times \dots \times X_n(\Omega)$, alors $f(X_1, X_2, \dots, X_p)$ et $g(X_{p+1}, \dots, X_n)$ sont indépendantes.

▷ *Preuve* : Admis.

◁

Par exemple si X, Y, Z et T sont indépendantes, alors $\min(X, Y)$ et $\max(Z, T)$ sont indépendantes, $X + Y$ et $Z + T$ sont indépendantes, XZ et YT sont indépendantes, etc.

c) Indépendance et variance



Theorème 7 :

Si les variables X_1, X_2, \dots, X_n sont deux à deux indépendantes, alors

$$V(X_1 + X_2 + \dots + X_n) = V(X_1) + \dots + V(X_n)$$

Ce théorème reste bien sûr vrai avec "mutuellement indépendantes".

▷ *Preuve* : On sait que $V(X_1 + X_2 + \dots + X_n) = V(X_1) + \dots + V(X_n) + 2 \sum_{1 \leq i < j \leq n} cov(X_i, X_j)$.

Or pour tout (i, j) avec $i < j$, X_i et X_j sont indépendantes, donc $cov(X_i, X_j) = 0$ et on a l'égalité proposée. ◁

Exemple :

Si $S \sim \mathcal{B}(n, p)$, on écrit $S = \sum_{i=1}^n X_i$ avec $X_i \sim \mathcal{B}(p)$ et les X_i indépendants.

On a donc $V(S) = \sum_{i=1}^n V(X_i) = \sum_{i=1}^n p(1-p) = np(1-p)$.

On aurait très bien pu faire la preuve de cette formule "à la main" avec la définition et le théorème du transfert, mais c'est un peu plus compliqué....

Remarque :

En réalité, cette propriété de somme est valable pour des variables décorrélées : l'hypothèse d'indépendance est à peine trop forte, mais c'est souvent dans ce contexte que ce sera utilisé.