

## Table des matières

1. Alphabet grec	3
2. Quelques notations	4
Chapitre 1. Les sommes finis de réels	5
1. Quelques conseils pour commencer	5
2. Quelques sommes remarquables	6
3. Sommes et coefficients binomiaux	6
4. Sommes et inégalités	9
Chapitre 2. Les produits finis de réels	11
Chapitre 3. Dénombrement	13
1. Dénombrabilité	13
2. Cardinal d'un ensemble fini	13
3. Arrangements ou listes sans répétition	14
4. Listes avec répétition	15
5. Combinaisons sans répétition	15
6. Exemples de tirages usuels	16
Chapitre 4. Espaces probabilisés	17
1. Quelques notions sur les tribus d'événements	17
2. Probabilité	18
3. Suite monotone d'événements	19
4. Probabilité conditionnelle	20
Chapitre 5. Généralités sur les variables aléatoires réelles	23
1. Tribu des boréliens <sup>1</sup> de $\mathbf{R}$	23
2. Définition formelle d'une variable aléatoire	23
3. Loi d'une variable aléatoire	24
4. Fonction de répartition d'une variable aléatoire	25
5. Opérations sur les variables aléatoires	25
6. Indépendance de variables aléatoires	25
7. Les variables aléatoires discrètes	26
8. Les variables aléatoires continues	27
Chapitre 6. Séries doubles	31
1. Séries doubles à termes positifs	31
2. Séries doubles à termes quelconques	34
Chapitre 7. Vecteurs aléatoires discrets de $\mathbf{R}^n$	35

---

<sup>1</sup>En hommage au mathématicien français Emile Borel (1871 - 1956)

1. Le cas où $n = 2$ : couples de variables discrètes	35
2. Ajustement linéaire entre deux variables aléatoires	37
3. Le cas où $n > 2$ : vecteurs discrets	38
Chapitre 8. Moments d'une variable aléatoire	41
1. Le cas discret	41
2. Le cas continu	49
Chapitre 9. Indépendance stochastique	53
1. Indépendance d'événements	53
2. Indépendance de variables aléatoires	54
3. Formule de convolution	56
4. <b>Indépendance et moments</b>	57
Chapitre 10. Les lois de probabilités usuelles	59
1. Lois discrètes	59
2. Lois à densité	64
Chapitre 11. Convergences et approximations	73
1. Deux inégalités de concentration	73
2. Convergence en probabilité	73
3. Convergence en loi et approximations	75
Chapitre 12. Estimations	79
Index	83

**1. Alphabet grec**

Majuscule	Minuscule	Nom	Transcription
A	$\alpha$	alpha	a
B	$\beta$	bêta	b
Γ	$\gamma$	gamma	g
Δ	$\delta$	delta	d
E	$\varepsilon$	epsilon	e,é
Z	$\zeta$	dzéta	z
H	$\eta$	êta	ê,é
Θ	$\theta$	thêta	th
I	$\iota$	iota	i
K	$\kappa$	kappa	k ou c
Λ	$\lambda$	lambda	l
M	$\mu$	mu	m
N	$\nu$	nu	n
Ξ	$\xi$	ksi	x
O	$\omicron$	omicron	o
Π	$\pi$	pi	p
P	$\rho$	rhô	r ou rh
Σ	$\sigma$	sigma	s
T	$\tau$	tau	t
Υ	$\upsilon$	upsilon	u ou y
Φ	$\varphi$	phi	ph
X	$\chi$	khi	kh ou ch
Ψ	$\psi$	psi	ps
Ω	$\omega$	oméga	o

## 2. Quelques notations

$\wedge$	connecteur logique "et"
$\vee$	connecteur logique "ou"
$\mathbf{N}_{n_0}$	ensemble des entiers naturels supérieurs ou égaux à $n_0$
$\llbracket 1, n \rrbracket$	intervalle d'entiers compris entre 1 et $n$
$\mathcal{B}(\mathbf{R})$ ou $\mathcal{B}_{\mathbf{R}}$ ou $\mathcal{B}$	tribu de Borel
$\bigsqcup$	union disjointe d'ensembles
$\Omega$	univers
$\mathcal{A}$	tribu ( $\sigma$ – algèbre) des événements sur $\Omega$
$\mathbf{P}$	probabilité sur un espace probabilisable $(\Omega, \mathcal{A})$
$\mathbf{P}_A(B)$	probabilité conditionnelle de $B$ sachant que $A$ est réalisé
$(\Omega, \mathcal{A}, \mathbf{P})$	espace probabilisé
$X(\Omega)$	ensemble des valeurs prises par une variable aléatoire $X$
$[X \in \mathcal{I}]$	événement constitué des $\omega \in \Omega$ tels que $X(\omega) \in \mathcal{I}$
$X \boxplus Y$	somme de deux variables indépendantes
$f_X$	densité de probabilité d'une variable $X$
$F_X$	fonction de répartition d'une variable $X$
$\mathbf{E}(X)$	espérance d'une variable $X$
$m_r(X)$	moment d'ordre $r \in \mathbf{N}$ d'une variable $X$
$\mu_r(X)$	moment centré d'ordre $r \in \mathbf{N}$ de la variable $X$
$\rho_r(X)$	moment factoriel d'ordre $r \in \mathbf{N}^*$ de la variable $X$
$\mathcal{L}_d^p$	ensemble des variables discrètes possédant un moment d'ordre $p \in \mathbf{N}$
$\mathcal{L}_c^p$	ensemble des variables à densité possédant un moment d'ordre $p \in \mathbf{N}$
$\mathbf{V}(X)$	variance de la variable $X$
$\sigma(X)$	écart-type de la variable $X$
$\text{Cov}(X, Y)$	covariance du couple aléatoire $(X, Y)$
$\rho_{X, Y}$ ou $\rho(X, Y)$	coefficient de corrélation du couple aléatoire $(X, Y)$
$X \stackrel{\mathcal{L}}{=} Y$	les variables $X$ et $Y$ suivent la même loi
$\delta_C$	loi de Dirac de paramètre $C$
$\mathcal{U}(\llbracket 1, n \rrbracket)$	loi uniforme sur $\llbracket 1, n \rrbracket$
$\mathcal{U}([a, b])$	loi uniforme sur $[a, b]$
$\mathcal{B}(p)$	loi de Bernoulli de paramètre $p \in ]0, 1[$
$\mathcal{B}(n, p)$	loi binomiale de paramètres $n \in \mathbf{N}$ et $p \in ]0, 1[$
$\mathcal{H}(N, n, p)$	loi hypergéométrique de paramètres $N \in \mathbf{N}$ et $n \in \mathbf{N}$ et $p \in ]0, 1[$ tel que $Np \in \mathbf{N}^*$ et $N(1-p) \in \mathbf{N}^*$ .
$\mathcal{P}(\lambda)$	loi de Poisson de paramètre $\lambda \in \mathbf{R}_+^*$
$\mathcal{G}(p)$	loi géométrique de paramètre $p \in ]0, 1[$
$\mathcal{E}(\lambda)$	loi exponentielle de paramètre $\lambda \in \mathbf{R}_+^*$
$\mathcal{N}(0, 1)$	loi normale centrée réduite
$\mathcal{N}(m, \sigma)$	loi normale de paramètres $m \in \mathbf{R}$ et $\sigma \in \mathbf{R}_+^*$
$(X_n)_{n \geq n_0} \xrightarrow{\mathcal{L}} X$	la suite de variables $(X_n)_{n \geq n_0}$ converge en loi vers la variable $X$
$(X_n)_{n \geq n_0} \xrightarrow{\mathbf{P}} X$	la suite de variables $(X_n)_{n \geq n_0}$ converge en probabilité vers la variable $X$
$\sum_{n \geq n_0} u_n < \infty$	la série de terme général $u_n$ est convergente

## Les sommes finis de réels

### 1. Quelques conseils pour commencer

L'importance de l'outil de sommation, pour effectuer des calculs de probabilité, est telle qu'il est bon de rappeler quelques résultats très importants.

Dans tout ce qui va suivre les suites  $(a_n)_n$  et  $(b_n)_n$  sont deux suites de réels.

#### (1) Changement d'indice

L'ordre des éléments dans une **somme finie** étant **sans importance**, le rôle de l'indice est de surveiller que tous les termes de la suite sont effectivement dans la somme et que chacun ne sera compté qu'une seule fois. Le rôle de l'indice est donc préservé par **bijection**. Par exemple on posera fréquemment les changements affines suivants  $j = k + 1$  ou  $j = k - 1$  mais jamais  $j = k^2$ . Tout cela se généralise par l'égalité :

$$\sum_{k \in K} a_k = \sum_{i \in I} a_{\varphi^{-1}(i)}$$

où l'application  $\varphi : I \rightarrow K$  est une bijection.

#### (2) Les grosses erreurs à éviter :

-  Ne jamais écrire que :  $\sum_{k=p}^q a_k b_k = \left( \sum_{k=p}^q a_k \right) \left( \sum_{k=p}^q b_k \right)$ . On rappelle au passage à notre cher lecteur qu'un **produit de deux sommes simples finies donne une somme double finie**, c'est la formule de Fubini pour le cas fini, pour tout  $(p, q, r, s) \in \mathbb{N}^4$ ,  $p \leq q$ ,  $r \leq s$  :

$$\left( \sum_{k=p}^q a_k \right) \left( \sum_{l=r}^s b_l \right) = \sum_{k=p}^q \sum_{l=r}^s a_k b_l = \sum_{l=r}^s \sum_{k=p}^q a_k b_l$$

-  Ne jamais écrire que :  $\sum_{k=p}^q \frac{a_k}{b_k} = \frac{\sum_{k=p}^q a_k}{\sum_{k=p}^q b_k}$  où pour tout entier  $k$  de l'intervalle  $\llbracket p, q \rrbracket$ ,  $b_k$  est non nul.

#### (3) Inversion de l'ordre de sommation dans une somme double finie dans le cas où les indices sont liés par des inégalités

**Il n'existe pas de règle générale**, mais avant toute chose il faut bien avoir en tête qu'effectuer une inversion n'est pas un acte gratuit, elle doit toujours se justifier par la volonté de diminuer la difficulté des calculs. Voici trois règles fondamentales que vous devez toujours suivre dans l'ordre de leur présentation :

- **Règle 1** : écrire les contraintes de valeurs et d'ordre que subissent les indices.
- **Règle 2** : reproduire sur la première somme à calculer, c'est-à-dire celle écrite la plus à droite, la contrainte d'ordre (inégalité) entre les deux indices.

- **Règle 3** : la contrainte entre les deux indices ayant été précisée, écrire les valeurs extrêmes que prend le second indice.

Par exemple :

$$S = \sum_{i=0}^n \sum_{j=0}^i a_{i,j} = \sum_{j=0}^n \sum_{i=j}^n a_{i,j}$$

(4) **Somme géométrique.**

Lorsque vous calculez  $\sum_{k=0}^n x^k$  vérifiez toujours, et écrivez-le bien sur votre copie, que la raison  $x$  est bien différente de 1 avant de donner le résultat.

## 2. Quelques sommes remarquables

PROPOSITION 1.1. *Pour tout entier naturel  $n$  non nul, nous avons :*

$$\begin{aligned} (1) \quad \sum_{k=1}^n k &= \frac{n(n+1)}{2} \\ (2) \quad \sum_{k=1}^n k^2 &= \frac{n(n+1)(2n+1)}{6} \\ (3) \quad \sum_{k=1}^n k^3 &= \left( \sum_{k=1}^n k \right)^2 = \frac{n^2(n+1)^2}{4} \end{aligned}$$

PROPOSITION 1.2. (*Somme géométrique*)

$$\forall (p, n) \in \mathbf{N}^2, \quad p \leq n, \quad \sum_{k=p}^n x^k = x^p \left( \frac{1 - x^{n-p+1}}{1 - x} \right) \quad \text{si } x \neq 1$$

La formule reste valable pour  $x$  élément de  $\mathbb{C}$ .

PROPOSITION 1.3. (*Identité de Bernoulli*)

$$\forall (a, b) \in \mathbf{R}^2, \quad \forall n \in \mathbf{N}, \quad a^{n+1} - b^{n+1} = (a - b) \sum_{k=0}^n a^{n-k} b^k = (a - b) \sum_{k=0}^n a^k b^{n-k}$$

La formule reste valable pour  $a$  et  $b$  éléments de  $\mathbb{C}$ .

## 3. Sommes et coefficients binomiaux

### 3.1. Coefficient binomial.

DÉFINITION 1.1. (*Coefficient binomial*) Soit  $n$  et  $p$  deux entiers naturels tels que  $p \leq n$ . On appelle *coefficient binomial*, l'entier naturel noté  $\binom{n}{p}$ <sup>1</sup> ou aussi  $C_n^p$  défini par :

$$\binom{n}{p} = \frac{n!}{p!(n-p)!}$$

**Par convention** On posera que  $\binom{n}{p} = 0$  lorsque  $p < 0$  ou lorsque  $p > n$ .

**Interprétation combinatoire :** le nombre de façons de choisir une partie constituée de  $p$  éléments choisis à partir d'un ensemble en comportant  $n$  est égal à  $\binom{n}{p}$ .

PROPRIÉTÉS 1.1. (*des coefficients binomiaux*)

---

<sup>1</sup>Lire " $p$  parmi  $n$ ".

(1) Pour  $p$  et  $n$  deux entiers naturels tels que  $0 \leq p \leq n$  :

(a)  $\binom{n}{p} \in \mathbf{N}$

(b)  $\binom{n}{p} = \frac{n(n-1) \times \cdots \times (n-p+1)}{p!}$

(c)  $\binom{n}{p} = \binom{n}{n-p}$  (*propriété de symétrie*)

(2) Pour  $p$  et  $n$  deux entiers naturels tels que  $1 \leq p \leq n$  :

(a)  $\binom{n}{p} = \frac{n}{p} \binom{n-1}{p-1}$  (*formule d'absorption-extraction*)

(b)  $\binom{n}{p} = \binom{n-1}{p} + \binom{n-1}{p-1}$  (*formule de triangle de Pascal*)

**Interprétation combinatoire :** le nombre de façons de choisir une partie<sup>2</sup> constituée de  $p$  éléments choisis à partir d'un ensemble qui en compte  $n$  est égal à  $\binom{n}{p}$ .

### 3.2. Les trois principales identités sur les coefficients binomiaux.

PROPOSITION 1.4. (1)  $\sum_{k=p}^n \binom{k}{p} = \binom{n+1}{p+1}$  où  $p$  et  $n$  sont deux entiers naturels tels que  $p \leq n$   
(*formule du triangle de Pascal généralisée*).

(2)  $\sum_{k=0}^n \binom{n}{k} a^k b^{n-k} = (a+b)^n$  (*formule du binôme de Newton*) (La formule reste valable dans  $\mathbb{C}$ ).

(3)  $\sum_{k=0}^n \binom{n_1}{k} \binom{n_2}{n-k} = \binom{n_1+n_2}{n}$  où  $n_1, n_2$  et  $n$  sont trois entiers naturels tels que  $n \leq n_1 + n_2$   
(*formule de Vandermonde*).

COROLLAIRE 1.1. Pour tout entier naturel  $n$  :  $\sum_{k=0}^n \binom{n}{k}^2 = \binom{2n}{n}$

**3.3. Conseils pour bien sommer à travers des exemples.** Avant toute chose, il vous sera indispensable de comptabiliser les occurrences<sup>3</sup> de l'indice de sommation  $k$ . En effet votre réaction ne sera pas du tout la même.

3.3.1. *Sommes simples ne contenant qu'un seul coefficient binomial.*

**L'indice de sommation n'apparaît qu'une fois dans le binomial.**

–  $\sum_k \binom{*}{k}$  ou  $\sum_k \binom{k}{*}$

Il n'y a qu'une recette et celle-ci est simple : ouvrez bien vos yeux afin de voir où se situe l'indice de sommation. S'il est en bas vous penserez à la **formule du binôme de Newton**. Sinon ce sera la **formule du triangle de Pascal** qu'il faudra utiliser.

–  $\sum_k k \binom{*}{k}$

Penser à utiliser la formule d'**absorption/extraction** afin de faire disparaître le facteur  $k$ .

<sup>2</sup>Sous-ensemble.

<sup>3</sup>Nombre d'apparitions.

- $\sum_{\mathbf{k}} \mathbf{k} \binom{\mathbf{k}}{*}$  Attention la formule d'absorption/extraction ne s'applique pas directement car le produit  $\mathbf{k} \times \mathbf{k}!$  ne donne rien de spécial. On pensera donc à écrire que :

$$\mathbf{k} \binom{\mathbf{k}}{*} = (\mathbf{k} + 1) \binom{\mathbf{k}}{*} - \binom{\mathbf{k}}{*}$$

et par linéarité de la somme vous aurez à calculer deux sommes qui vous ramèneront à "TPG". Ayez bien en tête qu'il ne faut manipuler que des produits de termes consécutifs. Par exemple que faire avec

$$\sum_{\mathbf{k}=0}^n \mathbf{k} \binom{m - \mathbf{k} - 1}{m - n - 1}$$

car le produit  $\mathbf{k} \times (m - \mathbf{k} - 1)!$  ne donne absolument rien !

Réponse : écrire que  $k = m - (m - k)$ , dans ce cas le produit  $(m - k) \times (m - k - 1)!$  est égal à  $(m - k)!$  La séparation de la somme en deux, suivie de l'emploi des formules d'absorption/extraction et la formule du triangle de Pascal généralisée permettent de refermer.

$$- \sum_{\mathbf{k}} \frac{1}{\mathbf{k} + 1} \binom{*}{\mathbf{k}}$$

Penser à utiliser la formule d'**absorption/extraction** afin de faire disparaître le facteur  $\frac{1}{\mathbf{k} + 1}$ .

**L'indice de sommation apparaît deux fois dans le binomial.**

$$- \sum_{\mathbf{k}} \binom{n + \mathbf{k}}{m + \mathbf{k}} \text{ ou } \sum_{\mathbf{k}} \binom{n - \mathbf{k}}{m - \mathbf{k}}$$

$m$  et  $n$  étant deux entiers naturels.

Vous devez savoir qu'il **n'y a aucun moyen de transformer**  $n + k$  ou  $n - k$  car il n'existe pas de propriété sur les binomiaux le permettant. L'idée va donc consister à faire disparaître  $\mathbf{k}$  dans l'expression du bas en profitant de la propriété de symétrie des coefficients binomiaux. Cela vous ramènera par la suite à utiliser la **formule du triangle de Pascal généralisée**.

**3.3.2. Les sommes simples présentant deux coefficients binomiaux.**

$$- \sum_{\mathbf{k}} \binom{f(\mathbf{k})}{g(\mathbf{k})}$$

Dans cette situation vous n'avez qu'une seule issue de sortie, c'est utiliser la **formule de Vandermonde** (dont on dit qu'elle aurait été en fait trouvée par un chinois en 1303) à la seule condition que la somme  $f(\mathbf{k}) + g(\mathbf{k})$  fasse disparaître  $\mathbf{k}$ . Si ce n'est pas le cas vous devrez utiliser la **formule de symétrie** sur l'un des deux coefficient binomial de manière à ce que la somme des indices du bas des binomiaux fasse disparaître  $\mathbf{k}$ .

$$- \sum_{\mathbf{k}} \binom{m}{\mathbf{k}} \binom{k}{n}$$

Il n'est pas question d'évoquer Vandermonde puisque les indices  $\mathbf{k}$  se croisent et donc se neutralisent mutuellement par simplification des factorielles. Vous obtiendrez après la "création/disparition" d'une factorielle un nouveau produit de binomiaux dont un seul fera apparaître  $\mathbf{k}$ , ce qui vous ramènera à un calcul simple.

$$- \sum_{\mathbf{k}=0}^m \frac{\binom{m}{\mathbf{k}}}{\binom{n}{\mathbf{k}}}$$

Cette somme peut vous faire peur car **il n'y a pas dans le cours d'identité avec des quotients de coefficients binomiaux**, cela doit être clair ! Le seul moyen de vous en sortir est de transformer, à l'aide de factorielles, votre quotient de binomiaux en un autre quotient de binomiaux plus simple à manipuler, et c'est gagné !

#### 4. Sommes et inégalités

Voici les principaux résultats regroupés dans la proposition suivante :

PROPOSITION 1.5. (1) *Pour tout entier naturel  $n$  non nul, et pour tout réels  $x_1, \dots, x_n, y_1, \dots, y_n$  tels que pour tout  $k$  appartenant à  $\llbracket 1, n \rrbracket$  on ait  $x_k \leq y_k$ , alors :*

$$\sum_{k=1}^n x_k \leq \sum_{k=1}^n y_k$$

(2) *Pour tout entier naturel  $n$  non nul et pour tous réels  $x_1, \dots, x_n, y_1, \dots, y_n$  positifs tels que pour tout  $k$  appartenant à  $\llbracket 1, n \rrbracket$  on ait  $x_k \leq y_k$  alors :*

$$\left( \begin{array}{l} \forall k \in \llbracket 1, n \rrbracket, \quad x_k \leq y_k \\ \exists k_0 \in \llbracket 1, n \rrbracket \mid x_{k_0} < y_{k_0} \end{array} \right) \implies \left( \sum_{k=1}^n x_k < \sum_{k=1}^n y_k \right)$$

*Cette propriété est plus commodément exploitée sous la forme suivante :*

$$\left( \begin{array}{l} \forall k \in \llbracket 1, n \rrbracket, \quad x_k \leq y_k \\ \sum_{k=1}^n x_k = \sum_{k=1}^n y_k \end{array} \right) \implies (\forall k \in \llbracket 1, n \rrbracket, \quad x_k = y_k)$$

(3) (**Généralisation de l'inégalité triangulaire**) *Soit  $n$  un entier naturel non nul et  $n$  réels  $x_1, x_2, \dots, x_n$  alors on a :*

$$\left| \sum_{k=1}^n x_k \right| \leq \sum_{k=1}^n |x_k|$$

(4) (**Inégalité de Cauchy-Schwarz**) *Pour tout entier naturel  $n$  non nul et pour tout réel  $x_1, \dots, x_n, y_1, \dots, y_n$  on a :*

$$\left( \sum_{i=1}^n x_i y_i \right)^2 \leq \left( \sum_{i=1}^n x_i^2 \right) \left( \sum_{i=1}^n y_i^2 \right)$$

*Le cas d'égalité étant obtenu lorsque les deux  $n$ -uplets  $(x_1, \dots, x_n)$  et  $(y_1, \dots, y_n)$  sont **proportionnels**.*



## Les produits finis de réels

Rappelons les règles de manipulations élémentaires des produits.

PROPOSITION 2.1. Dans cette proposition  $(x_k)_{k \geq 1}$  et  $(y_k)_{k \geq 1}$  désignent des suites de réels.

- (1)  $\forall (p, n) \in \mathbf{N}^2, p \leq n, \forall a \in \mathbf{R}, \prod_{k=p}^n a = a^{n-p+1}$ .
- (2)  $\forall n \in \mathbf{N}^*, \left( \prod_{k=1}^n x_k \right) \times \left( \prod_{k=1}^n y_k \right) = \prod_{k=1}^n (x_k \times y_k)$ .
- (3)  $\forall (p, n) \in (\mathbf{N}^*)^2, \prod_{k=1}^{n+p} x_k = \left( \prod_{k=1}^n x_k \right) \times \left( \prod_{k=n+1}^{n+p} x_k \right)$  (*associativité*).
- (4)  $\forall n \in \mathbf{N}^*, \frac{\prod_{k=1}^n x_k}{\prod_{k=1}^n y_k} = \prod_{k=1}^n \left( \frac{x_k}{y_k} \right)$  où pour tout  $k \in \llbracket 1, n \rrbracket, y_k \neq 0$ .
- (5)  $\forall (n, m) \in \mathbf{N}^* \times \mathbf{N}, \left( \prod_{k=1}^n x_k \right)^m = \prod_{k=1}^n (x_k)^m$ .
- (6)  $\forall n \in \mathbf{N}^*, \prod_{k=1}^n \frac{x_{k+1}}{x_k} = \frac{x_{n+1}}{x_1}$  où pour tout  $k \in \llbracket 1, n \rrbracket, x_k \neq 0$  (*télescopage*).

PROPOSITION 2.2. (**Lien entre produit et somme**) Pour tout entier naturel  $n$  non nul et pour tout réel  $x_k$  strictement positif lorsque  $k \in \llbracket 1, n \rrbracket$  :

$$\ln \left( \prod_{k=1}^n x_k \right) = \sum_{k=1}^n \ln(x_k)$$

REMARQUE 2.1.  Les signes  $\sum$  et  $\prod$  ne s'inversent pas. C'est à dire que l'on n'écrit pas :  $\sum \prod \dots = \prod \sum \dots$

Les résultats qui vont suivre ne sont pas à connaître par coeur. Cependant vous devez être capable de les retrouver car ils interviennent régulièrement dans les calculs, surtout en analyse.

PROPOSITION 2.3. (**Produits à "trous"**)

- (1)  $\forall n \in \mathbf{N}^*, \prod_{k=1}^n 2k = 2^n n!$
- (2)  $\forall n \in \mathbf{N}, \prod_{k=0}^n (2k+1) = \frac{(2n+1)!}{2^n n!}$ .
- (3)  $\forall (p, n) \in \mathbf{N} \times \mathbf{N}^*, p \leq n-1, \prod_{k=p+1}^n k = \frac{n!}{p!}$ .

PROPOSITION 2.4. (**Produits et inégalités**) Pour tout  $n$  non nul, et pour tous réels  $x_1, \dots, x_n, y_1, \dots, y_n$  positifs tels que pour tout  $k$  appartenant à  $\llbracket 1, n \rrbracket$  on ait  $x_k \leq y_k$ , alors :

$$\prod_{k=1}^n x_k \leq \prod_{k=1}^n y_k$$

## Dénombrément

DÉFINITION 3.1. (**Ensemble fini**) Soit  $E$  un ensemble. On dit que  $E$  est un **ensemble fini** lorsqu'il existe un entier  $n$  naturel non nul, et une bijection  $\varphi : \llbracket 1, n \rrbracket \longrightarrow E$ .

### 1. Dénombrabilité

DÉFINITION 3.2. (**Ensemble dénombrable – Ensemble au plus dénombrable**) On dit d'un ensemble qu'il est **dénombrable** lorsque l'on peut indexer tous ses éléments par des entiers. On dit qu'un ensemble est **au plus dénombrable** lorsqu'il est fini ou dénombrable.

EXEMPLE 3.1. –  $\mathbf{N}, \mathbf{N}^*, \mathbf{Z}, \mathbf{Q}, \mathbf{N}^2$  sont dénombrables.  
 – Toute partie de  $\mathbf{N}$  est finie ou dénombrable.  
 –  $\mathbf{R}$  et  $\mathbf{C}$  ne sont pas dénombrables.

### 2. Cardinal d'un ensemble fini

DÉFINITION 3.3. (**Dénombrer – Cardinal d'un ensemble**) **Dénombrer** un ensemble fini, c'est trouver le nombre d'éléments qu'il contient et que l'on appelle **cardinal**. On le notera  $\text{card}(E)$  ou  $|E|$ .

CONVENTION 3.1. On posera que  $|\emptyset| = 0$ .

PROPRIÉTÉS 3.1. (1) Soit  $A$  un ensemble fini et  $B$  un ensemble tel qu'il existe une **bijection** de  $A$  vers  $B$  (on dit que  $A$  et  $B$  sont **équipotents**). Alors  $B$  est un ensemble fini et  $|A| = |B|$ .

(2) Soit  $A$  et  $B$  deux ensembles tels que  $A$  est fini et  $B \subset A$ , alors :

- (a)  $|B| \leq |A|$
- (b)  $|A - B| = |A| - |B|$
- (c)  $(|A| = |B|) \iff (A = B)$

(3)  $|A \cup B| = |A| + |B| - |A \cap B|$

(4) (**Formule de Poincaré ou du crible**)

$$\forall n \in \mathbf{N}^*, \quad \left| \bigcup_{k=1}^n A_k \right| = \sum_{k=1}^n (-1)^{k-1} \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} |A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}|$$

Notez que l'utilisation de cette formule est la plupart du temps **évitable sauf** dans les problèmes de **rencontres**.

(5)  $|A \uplus B| = |A| + |B|$

(6)  $\left| \biguplus_{k=1}^n A_k \right| = \sum_{k=1}^n |A_k|$

(7)  $|A \times B| = |A| \times |B|$

(8)  $\forall n \in \mathbf{N}^*, \quad \left| \prod_{k=1}^n A_k \right| = \prod_{k=1}^n |A_k|$

DÉFINITION 3.4. (**Partition**) On dit que la famille d'ensembles  $(A_i)_{i \in I}$  constitue une **partition** d'un ensemble  $E$  lorsque :

- (1)  $\forall i \in I, A_i \neq \emptyset$
- (2)  $\forall (i, j) \in I^2, (i \neq j) \implies (A_i \cap A_j = \emptyset)$
- (3)  $\bigsqcup_{i \in I} A_i = E$

PROPOSITION 3.1. (**Lemme des bergers**) Soit  $E$  et  $F$  deux ensembles finis et  $f$  une **application surjective** de  $E$  vers  $F$ . On suppose qu'il existe un entier naturel  $p$  non nul tel que pour tout élément  $y$  de  $F$  on ait  $|f^{-1}(\{y\})| = p$ , alors :

$$|E| = p|F|$$

COROLLAIRE 3.1. Soit  $E$  un ensemble et soit  $(A_i)_{i \in \llbracket 1, n \rrbracket}$  une **partition** de  $E$  formée de  $n$  sous-ensembles finis. Alors  $E$  est un ensemble fini et  $|E| = \sum_{i=1}^n |A_i|$ . Si toutes les cellules de la partition sont **équipotentes**<sup>1</sup> de cardinal commun  $c \in \mathbf{N}^*$ , alors  $|E| = nc$  (par le lemme des bergers<sup>2</sup>).

PROPOSITION 3.2. (**Dénombrement de  $F^E$** ) Soit  $E$  et  $F$  deux ensembles finis, alors l'ensemble des applications de  $E$  dans  $F$  noté  $\mathcal{A}(E, F)$  ou  $F^E$  est fini et :

$$|\mathcal{A}(E, F)| = |F|^{|E|}$$

### 3. Arrangements ou listes sans répétition

DÉFINITION 3.5. (**Arrangement**) Soit  $E$  un ensemble fini non vide possédant  $n$  éléments,  $n \in \mathbf{N}^*$ , et soit  $p$  un entier naturel non nul inférieur ou égal à  $n$ . On appelle **arrangement** de  $p$  éléments de  $E$  tout  **$p$ -uplet** d'éléments de  $E$  deux à deux distincts, c'est-à-dire une **application injective** de  $\llbracket 1, p \rrbracket$  dans  $E$ .

REMARQUE 3.1. Dans un arrangement l'ordre des éléments a de l'importance, ceux-ci ne pouvant être répétés.

PROPOSITION 3.3. (**Dénombrement des arrangements**) Le nombre d'arrangements de  $p$  éléments de  $E$  de cardinal  $n$ , avec  $1 \leq p \leq n$  est égal au produit  $n \times (n-1) \times \cdots \times (n-p+1)$  noté  $A_n^p$ . En particulier  $A_n^p = 0$  quand  $p > n$ .

CONVENTION 3.2. On posera pour tout entier naturel  $p$  que  $A_0^p = 1$ .

PROPOSITION 3.4. (1)  $\forall (n, p) \in (\mathbf{N}^*)^2, p \leq n, A_n^p = \frac{n!}{(n-p)!}$ .

- (2) (**Dénombrement des injections**) Le nombre d'**injections** d'un ensemble  $E$  de cardinal  $p$  vers un ensemble  $F$  de cardinal  $n$ , avec  $p \leq n$ , est égal à  $A_n^p$ .
- (3) (**Dénombrement des permutations**) Si  $|E| = n$ , alors le nombre de **permutations** de  $E$ , c'est-à-dire le nombre de **bijections** de  $E$  dans lui-même est égal à  $A_n^n = n!$

<sup>1</sup>Cela veut dire qu'il existe une *bijection* les cellules prises deux à deux.

<sup>2</sup>C'est un cas particulier du premier point.

#### 4. Listes avec répétition

DÉFINITION 3.6. (*p-liste*) Soit  $E$  un ensemble fini non vide possédant  $n$  éléments et soit  $p$  un entier naturel non nul. On appelle *p-liste* de  $E$  tout  $p$ -uplet d'éléments de  $E$ .

REMARQUE 3.2. Dans une *p-liste* l'ordre des éléments a de l'importance ceux-ci pouvant être répétés.

- PROPOSITION 3.5. (1) (*Dénombrement des listes*) Soit  $n$  et  $p$  deux entiers naturels non nuls. Le nombre de  $p$ -listes d'éléments d'un ensemble  $E$  de cardinal  $n$  est égal à  $n^p$ .
- (2) (*Dénombrement des applications*) Le nombre d'applications d'un ensemble de  $p$  éléments vers un ensemble de  $n$  éléments est égal à  $n^p$ .

#### 5. Combinaisons sans répétition

##### 5.1. Généralités.

DÉFINITION 3.7. (*Combinaison*) Soit  $E$  un ensemble de cardinal  $n \in \mathbf{N}^*$  et soit  $p$  un entier naturel tel que  $p \leq n$ . On appelle *combinaison sans répétition* (ou *combinaison*) de  $E$ , toute partie de  $E$  contenant  $p$  éléments.

REMARQUE 3.3. Dans une combinaison sans répétition l'ordre des éléments n'a pas d'importance.

PROPOSITION 3.6. (*Dénombrement des combinaisons*) Le nombre de combinaisons de  $p$  éléments de  $E$  de cardinal  $n$ , avec  $1 \leq p \leq n$  est égal à :

$$\binom{n}{p} = \frac{A_n^p}{p!}$$

PROPOSITION 3.7. (*Cardinal de l'ensemble des parties d'un ensemble*) Soit  $E$  un ensemble fini de cardinal noté  $|E|$  alors :

$$|\mathcal{P}(E)| = 2^{|E|}$$

##### 5.2. Le problèmes des anagrammes.

PROPOSITION 3.8. (*Dénombrement des anagrammes*) Soit  $n \in \mathbf{N}^*$ . Considérons un mot de  $n$  lettres formé de  $n_k$  lettres  $a_k$  pour  $k$  appartenant à  $\llbracket 1, p \rrbracket$  avec  $\sum_{k=1}^p n_k = n$ . Alors le nombre d'anagrammes que l'on peut former à partir de ce mot est égal à :

$$\binom{n}{n_1} \times \binom{n-n_1}{n_2} \times \dots \times \binom{n-n_1-\dots-n_{p-1}}{n_p} = \frac{n!}{n_1! \times n_2! \times \dots \times n_p!}$$

##### 5.3. Cas de suites monotones.

PROPOSITION 3.9. (1) (*Cardinal d'un ensemble de suites strictement croissantes*)

$$|\{(x_1, x_2, \dots, x_p) \in (\mathbf{N}^*)^p \mid 1 \leq x_1 < x_2 < \dots < x_p \leq n\}| = \binom{n}{p}$$

(2) (*Cardinal d'un ensemble de suites non décroissantes*)

$$|\{(x_1, x_2, \dots, x_p) \in (\mathbf{N}^*)^p \mid 1 \leq x_1 \leq x_2 \leq \dots \leq x_p \leq n\}| = \binom{p+n-1}{p}$$

CONSÉQUENCE 3.1. (1)  $\sum_{1 \leq i_1 < i_2 < \dots < i_p \leq n} 1 = \binom{n}{p}$

$$(2) \quad \sum_{1 \leq i_1 \leq i_2 \leq \dots \leq i_p \leq n} 1 = \binom{p+n-1}{p}$$

### 6. Exemples de tirages usuels

PROPOSITION 3.10. (1) *Le nombre de tirages **successifs et sans remise** de  $n$  boules faits dans une urne contenant  $N$  boules **discernables** avec  $n \leq N$  est égal à :*

$$A_N^n$$

(2) *Le nombre de tirages **successifs et avec remise** de  $n$  boules faits dans une urne contenant  $N$  boules **discernables** est égal à :*

$$N^n$$

(3) *Le nombre de **tirages simultanés** de  $n$  boules faits dans une urne contenant  $N$  boules **discernables** avec  $n \leq N$  est égal à :*

$$\binom{N}{n}$$

## Espaces probabilisés

### 1. Quelques notions sur les tribus d'événements

DÉFINITION 4.1. (**Tribu**) Tout ensemble de parties d'un ensemble  $\Omega$ , contenant  $\Omega$ , stable par réunion au plus dénombrable et par passage au complémentaire, s'appelle une **tribu** ou  $\sigma$ -algèbre sur  $\Omega$ , notée  $\mathcal{A}$ . Autrement dit  $\mathcal{A}$  est une tribu si :

- (1)  $\Omega \in \mathcal{A}$ ,
- (2)  $\forall A \in \mathcal{A}, \bar{A} \in \mathcal{A}$ ,
- (3) Pour toute suite  $(A_n)_{n \geq n_0}$  d'événements de  $\mathcal{A}$ ,  $\bigcup_{n=n_0}^{+\infty} A_n \in \mathcal{A}$ .

PROPOSITION 4.1. (1)  $\emptyset \in \mathcal{A}$ ,  
 (2)  $\mathcal{A}$  est stable pour l'**intersection**,  
 (3)  $\mathcal{A}$  est stable pour la différence et la **différence symétrique**.

EXEMPLE 4.1. (1)  $\mathcal{A}_0 = \{\emptyset, \Omega\}$  appelée **tribu grossière**.  
 (2)  $\mathcal{A}_1 = \{\emptyset, A, \bar{A}, \Omega\}$  est la plus petite tribu engendrée par  $A$  notée  $\sigma(A)$ .  
 (3)  $\mathcal{A}_2 = \mathcal{P}(\Omega)$  (c'est la tribu la plus fine).

REMARQUE 4.1. Il est à noter que toute tribu  $\mathcal{A}$  sur  $\Omega$  vérifie  $\mathcal{A}_0 \subset \mathcal{A} \subset \mathcal{A}_2$ .

PROPOSITION 4.2. L'intersection d'une famille au plus dénombrable de tribus de  $\Omega$  est une tribu de  $\Omega$ .

DÉFINITION 4.2. Soit  $i$  un entier d'un ensemble  $I \subset \mathbb{N}$ . On appelle **tribu engendrée** par une famille  $\mathcal{F}$  d'événements de  $\mathcal{P}(\Omega)$ , la plus petite des tribus  $\mathcal{A}_i$  de  $\mathcal{P}(\Omega)$  (au sens de l'inclusion) contenant  $\mathcal{F}$ , c'est-à-dire l'intersection de toutes les tribus sur  $\Omega$  contenant  $\mathcal{F}$ . La tribu engendrée par  $\mathcal{F}$  est notée  $\sigma(\mathcal{F})$ .

PROPOSITION 4.3. (**Argument de minimalité**)  $(\mathcal{F} \subset \mathcal{T} \text{ et } \mathcal{T} \text{ tribu}) \implies (\sigma(\mathcal{F}) \subset \mathcal{T})$ .

REMARQUE 4.2. Il est à noter que  $\sigma(\mathcal{F}) = \sigma(\mathcal{F}') \not\implies \mathcal{F} = \mathcal{F}'$ . Par exemple si  $\mathcal{F} = \{A\}$  (la famille  $\mathcal{F}'$  est réduite à l'événement  $A$ ) et  $\mathcal{F}' = \{\bar{A}\}$  (la famille  $\mathcal{F}$  est réduite à l'événement  $\bar{A}$ ) où  $A \subset \Omega$  alors  $\sigma(\mathcal{F}) = \{\emptyset, A, \bar{A}, \Omega\} = \sigma(\mathcal{F}')$ .

DÉFINITION 4.3. (**Système complet d'événements**) Soit une suite  $(A_k)_{k \in K}$  d'événements liés à une même épreuve d'univers associé  $\Omega$ . On dit que  $(A_k)_{k \in K}$  forme un **système complet d'événements** si  $(A_k)_{k \in K}$  est une **partition** de  $\Omega$ , autrement dit si :

- (1)  $\forall k \in K, A_k \neq \emptyset$ ,
- (2)  $\forall (i, j) \in K^2, (i \neq j) \implies (A_i \cap A_j = \emptyset)$ ,
- (3)  $\bigsqcup_{k \in K} A_k = \Omega$ .

EXEMPLE 4.2. Soit  $\Omega = \{\omega_i \mid i \in I\}$  où  $I$  est un ensemble au plus dénombrable d'indices, alors  $(\{\omega_i\})_{i \in I}$  est un système complet d'événements.

PROPOSITION 4.4. (1) Soit  $(A_k)_{k \in K}$  un système complet d'événements alors :

$$\sigma((A_k)_{k \in K}) = \left\{ \biguplus_{l \in L} A_l \mid L \subset K \right\}$$

(2) Soit  $\Omega = \{\omega_i, i \in I\}$ .  $\mathcal{P}(\Omega)$  est une tribu du fait que  $\mathcal{P}(\Omega) = \left\{ \biguplus_{j \in J} \{\omega_j\} \mid J \subset I \right\}$ .

(3) Soit  $\Omega$  un univers,  $A$  et  $B$  deux événement de  $\mathcal{P}(\Omega)$ , nous avons l'implication :

$$(A \subset B) \implies (\sigma(A) \subset \sigma(B))$$

EXEMPLE 4.3. *Quelle tribu choisir ?*

- **Si  $\Omega$  est fini** : on pourra prendre  $\mathcal{A} = \mathcal{P}(\Omega)$  qui contient un nombre fini d'éléments, dans ce cas les notions d'algèbres et de tribus sont confondues. A ce propos une tribu est toujours une algèbre mais la réciproque n'est pas vraie en général.
- **Si  $\Omega$  est infini dénombrable** : on pourra prendre  $\mathcal{A} = \mathcal{P}(\Omega)$ .

DÉFINITION 4.4. (**Espace probabilisable**) On appelle **espace probabilisable**, tout couple  $(\Omega, \mathcal{A})$ .

## 2. Probabilité

DÉFINITION 4.5. (**Définition axiomatique d'une probabilité**) Soit  $(\Omega, \mathcal{A})$  un espace probabilisable, on appelle **probabilité** sur  $(\Omega, \mathcal{A})$  toute application  $\mathbf{P} : \mathcal{A} \rightarrow [0, 1]$  vérifiant :

(1)  $\mathbf{P}(\Omega) = 1$

(2) Pour toute suite d'événements  $(A_k)_{k \in K}$ , **deux à deux disjoints**  $\sum_k \mathbf{P}(A_k)$  converge et :

$$\mathbf{P}\left(\biguplus_{k \in K} A_k\right) = \sum_{k \in K} \mathbf{P}(A_k)$$

C'est la propriété de  $\sigma$ -**additivité** de  $\mathbf{P}$ .

DÉFINITION 4.6. On appelle **espace probabilisé** tout triplet  $(\Omega, \mathcal{A}, \mathbf{P})$ .

PROPRIÉTÉS 4.1. Soit  $A$  et  $B$  deux événements d'un même espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$ , alors :

(1)  $\mathbf{P}(\emptyset) = 0$

(2) Si  $A \cap B = \emptyset$  alors  $\mathbf{P}(A \uplus B) = \mathbf{P}(A) + \mathbf{P}(B)$

(3)  $\mathbf{P}(\overline{A}) = 1 - \mathbf{P}(A)$

(4)  $\mathbf{P}(A - B) = \mathbf{P}(A) - \mathbf{P}(A \cap B)$

(5) Si  $B \subset A$  alors  $\mathbf{P}(A - B) = \mathbf{P}(A) - \mathbf{P}(B)$

(6) Si  $B \subset A$  alors  $\mathbf{P}(B) \leq \mathbf{P}(A)$  ( $\mathbf{P}$  est une **application croissante**)

(7)  $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$

(8) Soit  $A_1, \dots, A_n$ ,  $n$  événements d'un même espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$ , alors si les événements sont **deux à deux incompatibles** on a :

$$\mathbf{P}\left(\biguplus_{k=1}^n A_k\right) = \sum_{k=1}^n \mathbf{P}(A_k)$$

(9) (**Formule du crible ou de Poincaré**) Soit  $n$  événements  $A_1, \dots, A_n$  définis sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  alors on a :

$$\mathbf{P}\left(\bigcup_{k=1}^n A_k\right) = \sum_{k=1}^n (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_k \leq n} \mathbf{P}(A_{i_1} \cap \dots \cap A_{i_k})$$

REMARQUE 4.3. Si la probabilité  $\mathbf{P}(A_{i_1} \cap \dots \cap A_{i_k})$  ne dépend pas des indices de sommation, on sera amené à utiliser la somme présentée dans le premier résultat de la conséquence **3.1** page **15**.

DÉFINITION 4.7. (**Événement quasi-impossible – Événement quasi-certain**) Un événement  $A$  est qualifié de **quasi impossible** lorsqu'il est différent de l'ensemble vide mais de probabilité nulle. De même lorsqu'un événement  $E$ , différent de l'univers est de probabilité égale à 1, il sera qualifié de **quasi certain**.

DÉFINITION 4.8. (**Probabilité uniforme - Equiprobabilité**) Soit  $(\Omega, \mathcal{A})$  un espace probabilisable. Pour une probabilité  $\mathbf{P}$  sur  $(\Omega, \mathcal{A})$ , posons pour tout élément  $\omega$  de l'univers  $\Omega$ ,  $p_\omega = \mathbf{P}(\{\omega\})$ . On a alors  $\mathbf{P}(\Omega) = \sum_{\omega \in \Omega} p_\omega = 1$  et donc pour tout  $A \in \mathcal{A}$ ,  $\mathbf{P}(A) = \sum_{\omega \in A} p_\omega$ .

Lorsque  $\Omega$  est fini, l'équiprobabilité (ou équirépartition) est définie sur  $\Omega$  par :

$$\forall \omega \in \Omega, \quad p_\omega = \frac{1}{|\Omega|}$$

on parlera de **probabilité uniforme**. Par conséquent :

$$\forall A \in \mathcal{A}, \quad \mathbf{P}(A) = \frac{|A|}{|\Omega|}$$

REMARQUE 4.4. Nous retrouvons la célèbre formule employée dans le cadre de l'équiprobabilité : "la probabilité d'un événement est égale au nombre de cas favorables à  $A$  divisé par le nombre total de cas"<sup>1</sup> (**relation de Laplace**).

EXEMPLE 4.4. (**de situations d'équiprobabilité**) Lorsque l'expérience est décrite à l'aide de la locution "au hasard" nous sommes dans une **situation d'équiprobabilité**. Cependant l'absence de cette locution sous-entend cette situation lorsque l'on voit le mot équilibré, comme par exemple le lancer d'un dé ou d'une pièce équilibrés, alors que le lancer d'un dé pipé ou d'une pièce truquée montre au contraire une situation où tous les numéros et faces n'ont pas la même probabilité d'apparition.

### 3. Suite monotone d'événements

DÉFINITION 4.9. (**Suite monotone d'événements**) Soit  $(\Omega, \mathcal{A}, \mathbf{P})$  un espace probabilisé.

- On dit qu'une d'événements  $(A_n)_{n \in \mathbf{N}}$  définie sur cet espace est **croissante** lorsque pour tout entier naturel  $n$  on a  $A_n \subset A_{n+1}$ .
- On dit qu'une d'événements  $(A_n)_{n \in \mathbf{N}}$  définie sur cet espace est **décroissante** lorsque pour tout entier naturel  $n$  on a  $A_{n+1} \subset A_n$ .

EXEMPLE 4.5. (**de suites monotones à connaître**)

- (1) La suite  $\left(\bigcup_{k=0}^n A_k\right)_{n \in \mathbf{N}}$  est croissante.
- (2) La suite  $\left(\bigcup_{k=n}^{+\infty} A_k\right)_{n \in \mathbf{N}}$  est décroissante.
- (3) La suite  $\left(\bigcap_{k=0}^n A_k\right)_{n \geq 0}$  est décroissante.

<sup>1</sup>C'est le cardinal de l'univers.

(4) La suite  $\left(\bigcap_{k=n}^{+\infty} A_k\right)_{n \in \mathbf{N}}$  est croissante.

PROPOSITION 4.5. (**Théorème de la limite monotone**) Soit  $(A_n)_{n \geq 0}$  une **suite monotone d'événements** définie sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$ .

(1) Si  $(A_n)_{n \geq 0}$  est une **suite croissante** alors :

$$\mathbf{P}\left(\bigcup_{n=0}^{+\infty} A_n\right) = \lim_{n \rightarrow +\infty} \mathbf{P}(A_n)$$

(2) Si  $(A_n)_{n \geq 0}$  est une **suite décroissante** alors :

$$\mathbf{P}\left(\bigcap_{n=0}^{+\infty} A_n\right) = \lim_{n \rightarrow +\infty} \mathbf{P}(A_n)$$

COROLLAIRE 4.1. Pour toute suite  $(A_n)_{n \geq 0}$  d'événements définie sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$ , alors :

$$(1) \mathbf{P}\left(\bigcup_{n=0}^{+\infty} A_n\right) = \lim_{n \rightarrow +\infty} \mathbf{P}\left(\bigcup_{k=0}^n A_k\right)$$

$$(2) \mathbf{P}\left(\bigcap_{n=0}^{+\infty} A_n\right) = \lim_{n \rightarrow +\infty} \mathbf{P}\left(\bigcap_{k=0}^n A_k\right)$$

REMARQUE 4.5. (1) Dans ce corollaire notez bien que la suite d'événements est **quelconque**.

(2) Noter que si  $(A_n)_{n \geq 0}$  est une suite décroissante alors  $(\overline{A_n})_{n \geq 0}$  est une suite croissante et inversement.

PROPOSITION 4.6. (**Inégalité de Boole, hors-programme mais ...**)

(1) (**Cas fini**) Soit  $n$  un entier supérieur ou égal à 2 et  $n$  événements  $A_1, A_2, \dots, A_n$  définis sur un même espace probabilisé, alors :

$$\mathbf{P}\left(\bigcup_{k=1}^n A_k\right) \leq \sum_{k=1}^n \mathbf{P}(A_k)$$

(2) (**Cas infini**) Soit  $(A_n)_{n \geq 0}$  une suite d'événements définie sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$ , alors :

$$\mathbf{P}\left(\bigcup_{n=0}^{+\infty} A_n\right) \leq \sum_{n=0}^{+\infty} \mathbf{P}(A_n)$$

REMARQUE 4.6. (1) Le premier cas se démontre par récurrence, le second en se ramenant au cas fini à l'aide du corollaire 4.1.

(2) Dans le second cas, que la série converge ou diverge vers  $+\infty$ , le résultat est toujours valable.

## 4. Probabilité conditionnelle

### 4.1. Définition et propriétés.

DÉFINITION 4.10. (**Probabilité conditionnelle**) Soit  $A$  un événement défini sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  de **probabilité non nulle**, alors l'application :

$$\begin{aligned} \mathbf{P}_A : \mathcal{A} &\longrightarrow [0, 1] \\ B &\longmapsto \mathbf{P}_A(B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(A)} \end{aligned}$$

est une probabilité sur  $\mathcal{A}$  appelée **probabilité conditionnelle sachant  $A$**  (ou sachant que  $A$  est réalisé)<sup>2</sup>.

REMARQUE 4.7. **N'oubliez jamais de vérifier, sur une copie, que la probabilité de l'événement conditionnant n'est pas nulle.**

Comme une probabilité conditionnelle est avant toute chose une probabilité, nous avons les propriétés suivantes :

PROPRIÉTÉS 4.2. Soit  $(\Omega, \mathcal{A}, \mathbf{P})$  un espace probabilisé et  $B$  un événement de probabilité non nulle.

- (1)  $\mathbf{P}_B(\emptyset) = 0$ .
- (2) Pour tout  $(A_1, A_2, \dots, A_n) \in \mathcal{A}^n$  qui soient **deux à deux disjoints** on a :

$$\mathbf{P}_B\left(\biguplus_{k=1}^n A_k\right) = \sum_{k=1}^n \mathbf{P}_B(A_k)$$

- (3) Pour tout  $A \in \mathcal{A}$  on a  $\mathbf{P}_B(\overline{A}) = 1 - \mathbf{P}_B(A)$ .
- (4) Pour tout  $(A, C) \in \mathcal{A}$  tels que  $C \subset A$  on a  $\mathbf{P}_B(A - C) = \mathbf{P}_B(A) - \mathbf{P}_B(C)$ .
- (5) Pour tout  $(A, C) \in \mathcal{A}$  on a  $\mathbf{P}_B(A \cup C) = \mathbf{P}_B(A) + \mathbf{P}_B(C) - \mathbf{P}_B(A \cap C)$ .
- (6) **(Formule du crible pour une probabilité conditionnelle)**  
Pour tout  $(A_1, A_2, \dots, A_n) \in \mathcal{A}^n$  on a :

$$\mathbf{P}_B\left(\bigcup_{k=1}^n A_k\right) = \sum_{k=1}^n (-1)^{k-1} \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} \mathbf{P}_B(A_{i_1} \cap \dots \cap A_{i_k})$$

PROPOSITION 4.7. **(La formule du double conditionnement)** La probabilité d'un événement  $A$  conditionné par  $B$  puis par  $C$  est donnée par :

$$\mathbf{P}_B(A | C) = \mathbf{P}_{B \cap C}(A)$$

Autrement dit conditionner par  $B$  puis par  $C$  revient à conditionner par  $B \cap C$ .

PROPOSITION 4.8. **(Conditionnement et indépendance)**

- (1) Soit  $A$  et  $B$  deux événements définis sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  tel que  $\mathbf{P}(A) \neq 0$ , alors  $A$  et  $B$  sont indépendants **si, et seulement si**,  $\mathbf{P}_A(B) = \mathbf{P}(B)$ .
- (2) Soit  $A$  et  $B$  deux événements définis sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  tel que  $\mathbf{P}(B) \neq 0$ , alors  $A$  et  $B$  sont indépendants **si, et seulement si**,  $\mathbf{P}_B(A) = \mathbf{P}(A)$ .

#### 4.2. Les trois théorèmes fondamentaux.

PROPOSITION 4.9. **(Formule des probabilités composées ou formule du conditionnement multiple)** Soit  $A_1, \dots, A_n$ ,  $n$  événements définis sur  $(\Omega, \mathcal{A}, \mathbf{P})$  tels que pour tout  $n \in \mathbf{N}_2$ ,  $\mathbf{P}(A_1 \cap \dots \cap A_{n-1}) \neq 0$  alors on a :

$$\mathbf{P}(A_1 \cap \dots \cap A_n) = \mathbf{P}(A_1) \mathbf{P}_{A_1}(A_2) \mathbf{P}_{A_1 \cap A_2}(A_3) \dots \mathbf{P}_{A_1 \cap \dots \cap A_{n-1}}(A_n)$$

REMARQUE 4.8. **(Mode d'emploi)** On utilise cette formule quand on calcule la **probabilité d'une intersection quelconque d'événements**.

PROPOSITION 4.10. **(Formule des probabilités totales)** Soit  $(A_k)_{k \in K}$  un système complet d'événements défini sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  tel que pour tout  $k \in K$ ,  $\mathbf{P}(A_k) \neq 0$  alors on a :

$$\forall B \in \mathcal{A}, \quad \sum_k \mathbf{P}_{A_k}(B) \mathbf{P}(A_k) < \infty \quad \text{et} \quad \mathbf{P}(B) = \sum_{k \in K} \mathbf{P}_{A_k}(B) \mathbf{P}(A_k)$$

---

<sup>2</sup> $\mathbf{P}_A(B)$  peut s'écrire aussi  $\mathbf{P}(B/A)$ , mais c'est à éviter. Cela se lit : "probabilité conditionnelle de  $B$  sachant que  $A$  est réalisé".

REMARQUE 4.9. (*Mode d'emploi*)

- (1) Calcul de la probabilité d'un événement en conditionnant par tous les cas possibles regroupés dans un système complet d'événements.
- (2) Recherche d'une **relation de récurrence** vérifiée par une suite de probabilités  $(p_n)_n$ .

PROPOSITION 4.11. (*La formule de Thomas Bayes*) Soit  $(A_k)_{k \in K}$  un système complet d'événements défini sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  tel que pour tout  $k \in K$ ,  $\mathbf{P}(A_k) \neq 0$ , alors pour tout  $B \in \mathcal{A}$  de probabilité non nulle on a :

$$\forall i \in K, \quad \mathbf{P}_B(A_i) = \frac{\mathbf{P}_{A_i}(B) \mathbf{P}(A_i)}{\sum_{k \in K} \mathbf{P}_{A_k}(B) \mathbf{P}(A_k)}$$

REMARQUE 4.10. (*Mode d'emploi*) On utilise la formule de Bayes lorsque l'on s'intéresse à la **probabilité d'une cause** d'un événement (que l'on considère comme **conséquence acquise**). Le **système complet d'événements à introduire regroupe toutes les causes possibles de la conséquence**.

## Généralités sur les variables aléatoires réelles

Ce chapitre fondamental que vous allez adorer va nous permettre de revenir aux sources de la définition et des propriétés des variables aléatoires.

### 1. Tribu des boréliens<sup>1</sup> de $\mathbb{R}$

DÉFINITION 5.1. (**Boule ouverte et ouvert de  $\mathbb{R}$** ) On appelle **boule ouverte de  $\mathbb{R}$**  centre  $a$  et de rayon  $r > 0$  est l'ensemble noté  $B(a, r)$  égal à  $]a - r, a + r[$ , et qu'un **ouvert de  $\mathbb{R}$**  est un sous-ensemble de  $\mathbb{R}$  noté  $O$  tel qu'il existe en tout point  $a$  de  $O$  une boule ouverte de centre  $a$  et contenue dans  $O$ .

DÉFINITION 5.2. (**Tribu des boréliens de  $\mathbb{R}$** ) La **tribu des boréliens de  $\mathbb{R}$**  notée  $\mathcal{B}(\mathbb{R})$  (voire  $\mathcal{B}_{\mathbb{R}}$  et même  $\mathcal{B}$ ) est la tribu engendrée par l'ensemble des intervalles ouverts  $\mathbb{R}$ . C'est la plus petite tribu de  $\mathbb{R}$  contenant tous les intervalles ouverts de  $\mathbb{R}$ . Les éléments de  $\mathcal{B}(\mathbb{R})$  sont appelés de **boréliens de  $\mathbb{R}$** .

REMARQUE 5.1. On admettra que  $\mathcal{B}(\mathbb{R})$  est strictement incluse dans  $\mathcal{P}(\mathbb{R})$ .

PROPOSITION 5.1. Dans  $\mathbb{R}$ , les ouverts, les fermés, les singletons, les parties finies de  $\mathbb{R}$ , les parties dénombrables de  $\mathbb{R}$  sont des boréliens de  $\mathbb{R}$ . En particulier les intervalles sont des boréliens de  $\mathbb{R}$ .

### 2. Définition formelle d'une variable aléatoire

DÉFINITION 5.3. Soit  $(\Omega, \mathcal{A})$  et  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  deux espaces probabilisables. Une application de  $\Omega$  dans  $\mathbb{R}$  est dite **variable aléatoire à valeurs dans  $\mathbb{R}$**  si, pour tout  $B \in \mathcal{B}(\mathbb{R})$ , l'ensemble  $X^{-1}(B) \in \mathcal{A}$ , où on note :

$$X^{-1}(B) = \{\omega \in \Omega \mid X(\omega) \in B\}$$

*Explication* : soit  $(\Omega, \mathcal{A}, \mathbf{P})$  il faut bien avoir en tête que la plupart du temps ce n'est pas l'issue d'une expérience qui nous intéresse, mais plutôt la valeur que l'on peut lui attribuer. Pour cela il nous faut construire une application de  $\Omega$  dans  $\mathbb{R}$  ce qui va nous imposer de donner un sens aux expressions  $\mathbf{P}(X \in B)$  où  $B$  est un borélien de  $\mathbb{R}$ . Deux questions se posent : pourquoi cet objet bizarre, appelé borélien, et comment donner un sens à  $\mathbf{P}(X \in B)$ ? Pour faire simple, disons que  $\mathbb{R}$  étant un ensemble vaste,  $\mathcal{P}(\mathbb{R})$  l'est encore bien plus et risque donc de contenir des ensembles qui ne soient pas des événements, alors qu'un borélien (construit à partir d'intervalles) nous assure de ne pas tomber dans un "trou noir". Enfin  $X \in B$  étant l'écriture allégée de l'ensemble  $\{\omega \in \Omega \mid X(\omega) \in B\} = X^{-1}(B)$  alors on imposera naturellement qu'il soit un événement, soit donc que  $X^{-1}(B) \in \mathcal{A}$ .

PROPOSITION 5.2. (**Hors programme mais ...**) Soit  $(\Omega, \mathcal{A})$  un espace probabilisable et  $X : \Omega \mapsto \mathbb{R}$ . Pour que  $X$  soit une variable aléatoire  $(\Omega, \mathcal{A})$ , **il faut et il suffit** que soit satisfaite l'une des cinq propriétés équivalentes suivantes :

- (1)  $\forall B \in \mathcal{B}(\mathbb{R}), X^{-1}(B) \in \mathcal{A}$
- (2)  $\forall a \in \mathbb{R}, X^{-1}(]-\infty, a]) \in \mathcal{A}$
- (3)  $\forall a \in \mathbb{R}, X^{-1}(]-\infty, a[) \in \mathcal{A}$

---

<sup>1</sup>En hommage au mathématicien français Emile Borel (1871 - 1956)

$$(4) \forall a \in \mathbf{R}, X^{-1}([a, +\infty[) \in \mathcal{A}$$

$$(5) \forall a \in \mathbf{R}, X^{-1}(]a, +\infty]) \in \mathcal{A}$$

REMARQUE 5.2. Cette proposition est fondamentale lorsque nous aurons à démontrer que  $\varphi \circ X$  est une variable aléatoire dans le cas où  $X$  est une variable à densité et  $\varphi$  est une application continue sur un ouvert contenant  $X(\Omega)$  sans être forcément  $\mathcal{C}^1$  ou strictement monotone.

NOTATION 5.1. On notera  $X(\Omega) = \text{Im } X = \{X(\omega) \mid \omega \in \Omega\}$  l'ensemble des valeurs prises par  $X$ . D'autre part on définit l'écriture des neuf événements :  $[X = x] = X^{-1}(\{x\})$ ,  $[X > a] = X^{-1}(]a, +\infty[)$ ,  $[X \geq a] = X^{-1}([a, +\infty[)$ ,  $[X < a] = X^{-1}(]-\infty, a])$ ,  $[X \leq a] = X^{-1}(]-\infty, a])$ ,  $[a < X < b] = X^{-1}(]a, b])$ ,  $[a \leq X < b] = X^{-1}([a, b])$ ,  $[a < X \leq b] = X^{-1}(]a, b])$ ,  $[a \leq X \leq b] = X^{-1}([a, b])$ .

DÉFINITION 5.4. (Les différents types de variables du programme)

- (1) On dit que  $X$  est une **variable discrète** lorsque l'ensemble  $X(\Omega)$  est **au plus dénombrable**.
- (2) On dit que  $X$  est une **variable absolument continue** lorsque l'ensemble  $X(\Omega)$  est **infini et indénombrable**.

DÉFINITION 5.5. (Tribu associée à une variable) Soit  $X$  une variable aléatoire définie sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$ . On appelle **tribu associée à  $X$** , notée  $\mathcal{A}_X$ , la tribu engendrée par le système complet d'événement  $([X = x])_{x \in X(\Omega)}$ . Autrement dit :

$$\mathcal{A}_X = \sigma\left(\left([X = x]\right)_{x \in X(\Omega)}\right) = \left\{ \bigcup_{x \in A} [X = x] \mid A \subset X(\Omega) \right\}$$

### 3. Loi d'une variable aléatoire

DÉFINITION 5.6. (Loi d'une variable) Soit  $X$  une variable aléatoire définie sur  $(\Omega, \mathcal{A}, \mathbf{P})$ . On appelle la **loi de la variable  $X$**  l'application :

$$\begin{aligned} P_X : \mathcal{B}(\mathbf{R}) &\longrightarrow [0, 1] \\ A &\longmapsto P_X(A) = \mathbf{P}([X \in A]) \end{aligned}$$

REMARQUE 5.3. On voit par cette définition que la loi d'une variable  $X$  ne fait plus référence à l'espace probabilisé sur lequel celle-ci est définie et c'est tant mieux vu le caractère souvent abstrait de  $\Omega$  et surtout de  $\mathcal{A}$  !

PROPOSITION 5.3. L'application  $P_X$  définit une loi sur la tribu  $\mathcal{B}(\mathbf{R})$ .

DÉFINITION 5.7. (Egalité en loi) Soit  $X$  et  $Y$  deux variables aléatoires définies sur le même espace probabilisés, on dit que les deux variables suivent la même loi lorsque  $P_X = P_Y$  ce que l'on note  $X \stackrel{\mathcal{L}}{=} Y$  ou encore  $X \sim Y$ .

PROPOSITION 5.4. Soit  $X$  et  $Y$  deux variables aléatoires définies sur le même espace probabilisés, on dit que les deux variables suivent la même loi **si, et seulement si**, elles ont même fonction de répartition.

REMARQUE 5.4.  L'égalité en loi ne signifie surtout pas que  $X = Y$  ni même que  $\mathbf{P}([X = Y]) = 1$  ! Prenons l'exemple suivant : on joue à "pile ou face" avec une pièce équilibrée, et soient  $X$  et  $Y$  deux variables aléatoires définies comme suit : si on obtient "pile",  $X$  prend la valeur 1 alors que  $Y$  prend la valeur 0 et inversement si l'on obtient "face"  $X$  prend la valeur 0 alors que  $Y$  prend la valeur 1. Nous pouvons clairement dire que  $X$  et  $Y$  suivent la même loi de Bernoulli de paramètre 1/2, mais elles ne sont pas égales car elles ne prennent la même valeur en même temps lors de chaque issue de l'expérience. La seule chose que nous pouvons dire est qu'elles ont même espérance et même variance.

Moralité essentielle **si  $X = Y$  alors  $X \stackrel{\mathcal{L}}{=} Y$  mais la réciproque est fausse en général.**  
En revanche **si  $X \stackrel{\mathcal{L}}{=} Y$  alors  $\mathbf{E}(X) = \mathbf{E}(Y)$  et  $\mathbf{V}(X) = \mathbf{V}(Y)$ .**

#### 4. Fonction de répartition d'une variable aléatoire

DÉFINITION 5.8. (**Fonction de répartition**) On appelle **fonction de répartition** associée à une variable  $X$  la fonction notée  $F_X$  définie par :

$$\forall x \in \mathbb{R}, \quad F_X(x) = \mathbf{P}([X \leq x])$$

PROPRIÉTÉS 5.1. (1)  $F_X$  est non décroissante (i.e. croissante au sens large).

$$(2) \quad \lim_{x \rightarrow -\infty} F_X(x) = 0$$

$$(3) \quad \lim_{x \rightarrow +\infty} F_X(x) = 1$$

(4)  $F_X$  est continue à droite en tout point de  $\mathbb{R}$ .

PROPOSITION 5.5. (**admise**) **Réciproquement** on admet que toute fonction  $F : \mathbb{R} \rightarrow [0, 1]$  vérifiant les propriétés 5.1 pourra être considérée comme la fonction de répartition d'une variable aléatoire discrète  $X$ . C'est à dire qu'il existe une variable aléatoire discrète  $X$  telle que pour tout  $x \in \mathbb{R}$ ,  $F(x) = \mathbf{P}([X \leq x])$ .

PROPOSITION 5.6. En notant  $\lim_{x \rightarrow x_0^-} F_X(x) = F_X(x_0^-)$  qui existe et est finie puisque  $F_X$  est croissante, nous avons :

$$(1) \quad \forall (a, b) \in \mathbb{R}^2, a < b, \quad \mathbf{P}([a < X \leq b]) = F_X(b) - F_X(a)$$

$$(2) \quad \forall (a, b) \in \mathbb{R}^2, a < b, \quad \mathbf{P}([a \leq X \leq b]) = F_X(b) - F_X(a^-)$$

$$(3) \quad \forall (a, b) \in \mathbb{R}^2, a < b, \quad \mathbf{P}([a < X < b]) = F_X(b^-) - F_X(a)$$

$$(4) \quad \forall (a, b) \in \mathbb{R}^2, a < b, \quad \mathbf{P}([a \leq X < b]) = F_X(b^-) - F_X(a^-)$$

$$(5) \quad \forall a \in \mathbb{R}, \quad \mathbf{P}([X = a]) = F_X(a) - F_X(a^-)$$

#### 5. Opérations sur les variables aléatoires

PROPOSITION 5.7. (**hors programme**) L'ensemble  $\mathcal{V}_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbf{P})$  des variables aléatoires réelles définies sur  $(\Omega, \mathcal{A}, \mathbf{P})$  est un espace vectoriel sur  $\mathbb{R}$  stable par multiplication.

PROPOSITION 5.8. (**admise**) Soit  $X$  et  $Y$  deux variables aléatoires réelles sur un même espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$ , alors  $|X|$ ,  $X + Y$ ,  $\alpha X$ ,  $XY$ ,  $\sup(X, Y)$ ,  $\inf(X, Y)$  restent des variables aléatoires réelles.

#### 6. Indépendance de variables aléatoires

DÉFINITION 5.9. (1) (**Cas de deux variables**) Soit  $X$  et  $Y$  deux variables aléatoires sur  $(\Omega, \mathcal{A}, \mathbf{P})$ . On dit que  $X$  et  $Y$  sont **indépendantes** (pour la probabilité  $\mathbf{P}$ ) lorsque pour tout couple  $(A, B)$  de boréliens de  $\mathbb{R}$ , les événements  $[X \in A]$  et  $[Y \in B]$  sont indépendants. Autrement dit :

$$\forall (A, B) \in (\mathcal{B}(\mathbb{R}))^2, \quad \mathbf{P}([X \in A] \cap [Y \in B]) = \mathbf{P}([X \in A]) \mathbf{P}([Y \in B])$$

(2) (**Cas de  $n$  variables**) On dit que  $n$  variables  $(X_i)_{i \in [1, n]}$  définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  sont **indépendantes** lorsque pour tout  $n$ -uplet  $(B_1, \dots, B_n)$  de boréliens de  $\mathcal{B}(\mathbb{R})$  on a :

$$\mathbf{P}([X_1 \in B_1] \cap \dots \cap [X_n \in B_n]) = \mathbf{P}([X_1 \in B_1]) \times \dots \times \mathbf{P}([X_n \in B_n])$$

(3) (**Cas d'une suite de variables aléatoires**) On dit qu'une suite de variables aléatoires discrètes  $(X_n)_{n \in \mathbb{N}}$  définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  est constituée de **variables indépendantes** si toute **sous-suite finie** est formée de variables indépendantes.

(4) Si les variables sont **indépendantes** et toutes de **même loi**, on dit qu'elles sont indépendantes et **identiquement distribuées (i.i.d.)**.

PROPOSITION 5.9. (**Caractérisation des variables indépendantes**) Soit  $X$  et  $Y$  deux variables aléatoires réelles sur  $(\Omega, \mathcal{A}, \mathbf{P})$ , alors  $X$  et  $Y$  sont **indépendantes si, et seulement si** :

$$\forall (x, y) \in \mathbb{R}^2, \quad \mathbf{P}([X \leq x] \cap [Y \leq y]) = \mathbf{P}([X \leq x]) \mathbf{P}([Y \leq y])$$

PROPOSITION 5.10. (**Fonctions de variables indépendantes**) Soit  $X$  et  $Y$  deux variables aléatoires réelles sur  $(\Omega, \mathcal{A}, \mathbf{P})$ ,  $I$  un intervalle contenant  $X(\Omega)$ ,  $J$  un intervalle contenant  $Y(\Omega)$ ,  $\varphi : I \rightarrow \mathbb{R}$  et  $\psi : J \rightarrow \mathbb{R}$ . On suppose que alors  $X$  et  $Y$  sont **indépendantes** et  $\varphi \circ X$  et  $\psi \circ Y$  sont des variables aléatoires. Alors  $\varphi \circ X$  et  $\psi \circ Y$  sont **indépendantes**.

REMARQUE 5.5.  **La réciproque de ce théorème est fautive** comme le prouve l'exemple suivant : soit  $\Omega = \{P, F\}$  muni de la probabilité uniforme (lancé d'une pièce correcte). Posons  $X(P) = 1$ ,  $X(F) = -1$ ,  $Y = X$ ,  $\varphi(x) = x^3$  et  $\psi(y) = y^2$ . On constate que  $\varphi(X)$  et  $\psi(Y)$  sont des variables aléatoires indépendantes car  $\psi(Y) = Y^2 = 1$  (variable certaine). Cependant  $X$  et  $Y$  ne sont pas indépendantes car  $\mathbf{P}([X = 1] \cap [Y = 1]) = \mathbf{P}([X = 1]) = 1/2$  alors que  $\mathbf{P}([X = 1]) \mathbf{P}([Y = 1]) = 1/4$ .

PROPOSITION 5.11. (**Variables entières sans mémoire**) Soit  $X_1, \dots, X_n$   $n$  variables aléatoires réelles sur l'espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$ , alors  $X$  et  $Y$  sont **indépendantes si, et seulement si** :

$$\forall (x_1, \dots, x_n) \in \mathbb{R}^n, \quad \mathbf{P}([X_1 \leq x_1] \cap \dots \cap [X_n \leq x_n]) = \mathbf{P}([X_1 \leq x_1]) \times \dots \times \mathbf{P}([X_n \leq x_n])$$

## 7. Les variables aléatoires discrètes

DÉFINITION 5.10. On dit que  $X$  est une **variable discrète** lorsque l'ensemble  $X(\Omega)$  est **au plus dénombrable**.

DÉFINITION 5.11. (**Support d'une variable discrète**) On appelle **support** d'une variable aléatoire discrète  $X$  l'ensemble noté  $S_X$  défini par :

$$S_X = \{x \in X(\Omega) \mid \mathbf{P}([X = x]) > 0\}$$

PROPOSITION 5.12. (**Système complet d'événements associée à une variable discrète**) Soit  $X$  une variable aléatoire définie sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$ , alors la famille  $([X = x])_{x \in X(\Omega)}$  constitue un système complet d'événements appelé **système complet d'événements associée à  $X$** .

DÉFINITION 5.12. (**Loi d'une variable discrète**) Soit  $X$  une variable aléatoire à valeurs dans  $\{x_1, x_2, \dots, x_n, \dots\}$  On appelle la **loi de la variable  $X$**  l'application :

$$\begin{aligned} P_X : X(\Omega) &\longrightarrow \mathbb{R} \\ x &\longmapsto P_X(x) = \mathbf{P}([X = x]) \end{aligned}$$

Toute loi de probabilité d'une variable aléatoire discrète  $X$  sera **caractérisée** par la donnée de :

- (1)  $X(\Omega)$
- (2)  $\forall x_k \in X(\Omega), \mathbf{P}([X = x_k])$

DÉFINITION 5.13. (**Distribution de probabilité associée à une variable**) La suite réelle  $(p_n)_{n \geq 1}$  définie pour tout  $n \geq 1$  par  $p_n = \mathbf{P}([X = x_n])$  où  $X(\Omega) = \{x_n \mid n \in \mathbb{N}^*\}$ , est appelée **distribution de probabilité associée à la variable  $X$** .

PROPOSITION 5.13. La série  $\sum_{n \geq 1} p_n$  converge et de somme égale à 1.

REMARQUE 5.6. (**Attention faux-ami**) Retenez absolument ceci : on dit que deux variables  $X$  et  $Y$  sont égales si pour toute issue  $\omega$  de  $\Omega$  nous avons  $X(\omega) = Y(\omega)$  et qu'elles suivent la même loi<sup>2</sup> si  $X(\Omega) = Y(\Omega)$  et pour toute réalisation commune  $z_k$  aux variables  $X$  et  $Y$  on a  $\mathbf{P}([X = z_k]) = \mathbf{P}([Y = z_k])$ .

<sup>2</sup>On notera  $X \stackrel{L}{=} Y$ .

PROPOSITION 5.14. (**Fonction de répartition**) La donnée de la fonction de répartition  $F_X$  permet d'expliciter les probabilités  $p_k = \mathbf{P}([X = x_k])$ , et réciproquement la donnée des probabilités  $p_k$  permet d'expliciter la fonction de répartition  $F_X$  qui est une **fonction en escalier sur  $X(\Omega)$** . En prenant par exemple  $X(\Omega) = \llbracket x_1, x_n \rrbracket$ , nous avons :

$$F_X(x) = \begin{cases} 0 & \text{si } x < x_1 \\ \sum_{i=1}^k p_i & \text{si } x \in [x_k, x_{k+1}[ , \text{ où } k \text{ parcourt } \llbracket 1, n-1 \rrbracket \\ 1 & \text{si } x \geq x_n \end{cases}$$

PROPOSITION 5.15. (**Variable discrète fonction d'une autre**) Soit  $X$  une variable aléatoire discrète définie sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  et  $g$  une fonction définie sur  $X(\Omega)$  à valeurs dans  $\mathbf{R}$ , alors  $Y = g \circ X$  noté  $g(X)$  **reste** une variable aléatoire discrète.

PROPOSITION 5.16. Soit  $X$  une variable aléatoire définie sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  et  $g$  une application définie sur  $X(\Omega)$  à valeurs dans  $\mathbf{R}$ . Si  $Y = g \circ X$  alors  $\mathcal{A}_Y \subset \mathcal{A}_X$ .

REMARQUE 5.7. On a tenu à préciser dans cette proposition que les variables de départ et d'arrivée (obtenue après transformation) restent discrètes car **ce n'est pas forcément le cas** si la variable de départ est à **densité**.

PROPOSITION 5.17. (**Caractérisation de la loi d'une variable fonction d'une autre**). Soit  $X$  une variable aléatoire définie sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  et  $g$  une application définie sur  $X(\Omega)$  à valeurs dans  $\mathbf{R}$ . La **caractérisation** de la loi de  $Y = g \circ X$  notée  $g(X)$  sera faite en donnant :

- (1)  $Y(\Omega)$ ,
- (2)  $\forall y \in Y(\Omega)$ ,  $\mathbf{P}([Y = y])$  où  $\mathbf{P}([Y = y]) = \sum_{\substack{x \in X(\Omega) \\ g(x) = y}} \mathbf{P}([X = x])$ .

DÉFINITION 5.14. (**Variable aléatoire entière sans mémoire**) On dit qu'une variable aléatoire  $X$  à valeurs dans  $\mathbf{N}$  est **sans mémoire** lorsque :

$$\forall (m, n) \in \mathbf{N}^2, \quad \mathbf{P}_{[X > n]}([X > m + n]) = \mathbf{P}([X > m])$$

## 8. Les variables aléatoires continues

DÉFINITION 5.15. (**Variable à densité – Densité de probabilité**) Soit  $X$  une variable aléatoire à valeurs réelles définie sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$ . On dit que  $X$  est une **variable à densité** lorsque sa fonction de répartition  $F_X$  est continue sur  $\mathbf{R}$  et de classe  $\mathcal{C}^1$  sur  $\mathbf{R} - I$  où  $I$  est un ensemble fini de points éventuellement vide<sup>3</sup>. Toute fonction  $f_X$  à valeurs positives, qui ne diffère de  $F'_X$  qu'en un nombre fini de points, est appelée **densité de probabilité** de  $X$  (ou "**loi**" de  $X$ ).

PROPOSITION 5.18. Soit  $X$  une variable à densité de fonction de répartition  $F_X$ . Si  $f_X$  est une densité de  $X$  nous avons :

$$\forall x \in \mathbf{R}, \quad F_X(x) = \int_{-\infty}^x f(t) dt$$

- REMARQUE 5.8. (1) L'intégrale de  $f$  ne changeant pas quand on modifie  $f$  en un nombre fini de points, **il n'y a pas d'unicité de la densité**<sup>4</sup>.
- (2) Donner la **loi** d'une variable à densité  $X$ , c'est donner la fonction de répartition voire une **densité** de  $X$ .

<sup>3</sup>On dit que  $F_X$  est de classe  $\mathcal{C}^1$  **presque partout**.

<sup>4</sup>Ainsi donc ne dites JAMAIS "la densité..."

- (3) L'ensemble  $X(\Omega)$  est donné par le **support de  $X$**  noté  $\mathcal{S}_X$  défini par  $\mathcal{S}_X = \{x \in \mathbb{R} \mid f(x) \neq 0\}$ .
- (4) N'ayez crainte, une densité peut être supérieure à 1 ! Ne pas confondre avec  $\int_{-\infty}^{+\infty} f = 1$ .
- (5) Si  $f$  est nulle en dehors d'un intervalle  $I$ , alors  $X$  prend presque sûrement ses valeurs dans  $I$ .
- (6) Si  $f$  et  $g$  sont deux densités d'une même variable à densité alors  $f = g$  p.s.

PROPOSITION 5.19. (**Probabilité ponctuelle**) Si  $X$  est une variable à densité alors :

$$\forall x \in \mathbb{R}, \quad \mathbf{P}([X = x]) = 0$$

Ainsi une variable à densité ne peut être discrète.

PROPOSITION 5.20. (**Caractérisation d'une densité**) Toute fonction réelle  $f$  définie sur  $\mathbb{R}$  positive, continue sur  $\mathbb{R} - I$  où  $I$  est un ensemble fini de points éventuellement vide et telle que  $\int_{-\infty}^{+\infty} f$  converge et vaut 1 est la densité de probabilité d'une variable aléatoire.

PROPOSITION 5.21. (**Calcul de probabilités**) Soit  $X$  une variable aléatoire à densité définie sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  de densité associée  $f$ . Alors pour tout  $a \in \mathbb{R}$  :

$$\mathbf{P}([X \geq a]) = \mathbf{P}([X > a]) = \int_a^{+\infty} f(x) dx$$

et pour tout  $(a, b) \in \mathbb{R}^2$ ,  $a \leq b$  :

$$\mathbf{P}([a \leq X \leq b]) = \mathbf{P}([a < X \leq b]) = \mathbf{P}([a \leq X < b]) = \mathbf{P}([a < X < b]) = \int_a^b f$$

PROPOSITION 5.22. (**Fonction d'une variable à densité**) Soit  $X$  une variable aléatoire à densité définie sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  et soit  $\varphi$  une **fonction de classe  $\mathcal{C}^1$ , strictement monotone sur  $J$  contenant  $X(\Omega)$  et de dérivée non nulle sur  $J$  sauf éventuellement en un nombre fini de points**, alors  $Y = \varphi \circ X$  reste une variable à densité.

REMARQUE 5.9. Dans ce cas en supposant que  $\varphi'$  ne s'annule pas sur  $\varphi(X(\Omega))$  :

$$f_Y(x) = \begin{cases} \frac{f_X(\varphi^{-1}(x))}{|\varphi'(x)|} & \text{si } x \in \varphi(X(\Omega)) \\ 0 & \text{sinon} \end{cases}$$

Cette formule n'est pas à apprendre par coeur, elle doit être retrouvée pour chaque changement de variable, cependant on peut la réciter dans le cas d'un changement affine.

REMARQUE 5.10. Que faire en cas où la proposition 5.22 ne s'applique pas ? Il faut dire que le plus souvent c'est l'absence de monotonie qui posera des problèmes. Par exemple lorsque  $\varphi : x \rightarrow x^2$  ou  $\varphi : x \rightarrow |x|$  etc... Voici dans ce cas la démarche à suivre avec soin :

- (1) Vérifier que  $Y = \varphi \circ X$  reste une variable aléatoire en montrant que pour tout  $x \in \mathbb{R}$ ,  $[Y \leq x] \in \mathcal{A}$ . Pour cela on transformera pour tout  $x \in \mathbb{R}$ ,  $[Y \leq x]$  en une condition équivalente sur  $X$ .
- (2) Une fois assuré que  $Y$  reste une variable aléatoire, on cherchera sa fonction de répartition  $F_Y$  en vérifiant qu'elle est continue sur  $\mathbb{R}$  entier et qu'elle est de classe  $\mathcal{C}^1$  sur  $\mathbb{R}$  sauf éventuellement<sup>5</sup> en un nombre fini de points. Rien ne vous empêche de vérifier, pour la cohérence de vos calculs, que  $\lim_{-\infty} F = 0$  et  $\lim_{+\infty} F = 1$ .

<sup>5</sup>N'oubliez pas l'adverbe, c'est essentiel !

- (3) Si vous reconnaissez une fonction de répartition connue (uniforme ou exponentielle)  $Y$  suit donc la loi reconnue que vous donnerez avec ses paramètres. Sinon on dérivera  $F_Y$  là où c'est possible, disons sur  $\mathbb{R} - I$  ( $I$  ensemble éventuellement vide) en complétant la définition, le cas échéant, de la **future densité** en lui donnant des valeurs arbitraires positives ou nulles aux points de  $I$  si cet ensemble est non vide.

DÉFINITION 5.16. (**Variable à densité sans mémoire**) Soit  $X$  une variable aléatoire à densité définie sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$ . On dit que  $X$  est une **variable sans mémoire** lorsque :

$$\forall (x, y) \in \mathbb{R}^2, \quad \mathbf{P}_{[X > x]}([X > x + y]) = \mathbf{P}([X > y])$$

REMARQUE 5.11. Il est à noter que la définition reste la **même** dans le cas discret.



## Séries doubles

### 1. Séries doubles à termes positifs

#### 1.1. Définitions.

DÉFINITION 6.1. – Une **suite double** (ou **suite à deux indices**)  $u$  est une application de  $\mathbf{N}^2$  dans  $\mathbf{R}$  dont le terme général est noté  $u(i, j)$  ou plus simplement  $u_{i,j}$ .

– Une **série double** est une suite de sommes d'éléments d'une suite double, c'est-à-dire une suite de sommes partielles d'éléments d'une suite double  $\left( \sum_{i=0}^n \sum_{j=0}^m u_{i,j} \right)_{(n,m) \in \mathbf{N}^2}$ .

DÉFINITION 6.2. Une série double est dite à **termes positifs** lorsque pour tout couple  $(i, j) \in \mathbf{N}^2$ , on a :

$$u_{i,j} \geq 0$$

Nous utiliserons parfois l'acronyme "SDATP" pour désigner les *séries doubles à termes positifs*, mais n'en abusez pas sur les copies.

REMARQUE 6.1. Toute série double permet de définir **deux séries simples** si l'un des deux indices est fixé :

$$\sum_i u_{i,j} \quad \text{si } j \text{ est fixé} \quad \text{et} \quad \sum_j u_{i,j} \quad \text{si } i \text{ est fixé}$$

#### 1.2. Les théorèmes de convergences des séries doubles à termes positifs.

PROPOSITION 6.1. (**Interversion des sommations dans les séries doubles : théorème de Fubini-Tonelli**) Soit  $u = (u_{i,j})_{(i,j) \in I \times J}$  une famille à deux indices de **réels positifs** indexée par  $I \times J \subset \mathbf{N}^2$ .

Si les deux conditions suivantes sont réunies :

- pour tout  $i \in I$ , la **série simple**  $\sum_j u_{i,j}$  est convergente de somme  $\sum_{j \in J} u_{i,j} = a_i$ ,
- la **série simple**  $\sum_i a_i$  converge vers un réel  $S$ ,

alors la **série double**  $\sum_{(i,j)} u_{i,j}$  est **convergente**.

On a alors les résultats suivants :

- pour tout  $j \in J$ , la **série simple**  $\sum_i u_{i,j}$  est convergente de somme  $\sum_{i \in I} u_{i,j} = b_j$ ,
- la **série simple**  $\sum_j b_j$  converge vers  $S$ .

– On note  $S = \sum_{(i,j) \in I \times J} u_{i,j} = \sum_{i \in I} \left( \sum_{j \in J} u_{i,j} \right) = \sum_{j \in J} \left( \sum_{i \in I} u_{i,j} \right)$ , et on appelle ce réel somme de la série double de terme général  $u_{i,j}$ .

REMARQUE 6.2. (**essentielle**)

- (1) *Moralité* si les deux conditions du théorème sont vérifiées, on peut donc intervertir les sommations.

- (2) Il ne faut surtout pas vous focaliser sur l'ordre des indices dans les deux conditions à réunir, c'est-à-dire que vous pouvez l'inverser, c'est pourquoi il n'est pas rare de voir le théorème de Fubini écrit à l'aide d'une condition nécessaire et suffisante que voici : soit une suite double à termes positifs, alors les trois assertions suivantes sont équivalentes :

– la série double  $\sum_{i,j} u_{i,j}$  est convergente ;

$$- \left\{ \begin{array}{l} \forall j \in J, \sum_i u_{i,j} \text{ converge} \\ \sum_j \sum_{i \in I} u_{i,j} \text{ converge} \end{array} \right.$$

$$- \left\{ \begin{array}{l} \forall i \in I, \sum_j u_{i,j} \text{ converge} \\ \sum_i \sum_{j \in J} u_{i,j} \text{ converge} \end{array} \right.$$

De plus dans ces conditions on a :

$$\sum_{(i,j) \in I \times J} u_{i,j} = \sum_{i \in I} \left( \sum_{j \in J} u_{i,j} \right) = \sum_{j \in J} \left( \sum_{i \in I} u_{i,j} \right)$$

– Ce théorème d'usage très courant s'étend par une récurrence immédiate à des suites multiples :

$$\sum_{i_1 \in I_1} \sum_{i_2 \in I_2} \cdots \sum_{i_p \in I_p} u_{i_1, \dots, i_p} = \sum_{(i_1, \dots, i_p) \in I_1 \times \cdots \times I_p} u_{i_1, \dots, i_p}$$

- (3) Comme nous le verrons dans un document annexe dans le cas des séries à termes quelconques il se peut que l'on ait :

$$\sum_{i \in I} \left( \sum_{j \in J} u_{i,j} \right) \neq \sum_{j \in J} \left( \sum_{i \in I} u_{i,j} \right)$$

tout en ayant l'existence des deux sommes. Vous devez simplement conclure que la série double  $\sum_{(i,j)} u_{i,j}$  ne converge pas !

REMARQUE 6.3. (très importante)  “partialiser” en calculant  $\lim_{n \rightarrow +\infty} \sum_{i=0}^n \sum_{j=0}^n u_{i,j}$  est certes une solution de facilité, mais ce **n'est pas suffisant** pour conclure que la série double converge en cas où la limite serait constante (pensez au “problème des chemins” rencontré lors de l'étude de la continuité des fonctions de deux variables). Pour bien comprendre cela il suffit de vous dire la somme  $\sum_{i=0}^n \sum_{j=0}^n u_{i,j}$  lorsque  $n$  tend vers l'infini ne représente que les termes de la suite de sommes  $S_{n,m} = \sum_{i=0}^m \sum_{j=0}^n u_{i,j}$  tels que  $n = m$ , donc on ne travaille que dans une direction (suivant la première bissectrice) au lieu de travailler dans toutes les directions avec  $n$  et  $m$  grands. En revanche retenez par coeur que **pour toute série double convergente**  $\sum_{(i,j)} u_{i,j}$  on a :

$$\sum_{(i,j) \in \mathbf{N}^2} u_{i,j} = \lim_{n \rightarrow +\infty} \sum_{i=0}^n \left( \sum_{j=0}^n u_{i,j} \right)$$

Est-ce bien clair ?

COROLLAIRE 6.1. Soit  $\sum_i a_i$  et  $\sum_j b_j$  deux séries simples à termes positifs convergentes. Alors la série double  $\sum_{i,j} (a_i b_j)$  est convergente et :

$$\sum_{(i,j) \in \mathbf{N}^2} a_i b_j = \left( \sum_{i \in \mathbf{N}} a_i \right) \left( \sum_{j \in \mathbf{N}} b_j \right)$$

PROPOSITION 6.2. (**Théorème de comparaison**) Soit  $(u_{i,j})_{(i,j) \in I \times J}$  et  $(v_{i,j})_{(i,j) \in I \times J}$  deux familles de nombres réels positifs tels que :

$$\forall (i,j) \in I \times J, \quad v_{i,j} \leq u_{i,j}$$

Si  $\sum_{(i,j)} u_{i,j}$  converge, alors  $\sum_{(i,j)} v_{i,j}$  converge aussi et on a  $\sum_{(i,j) \in I \times J} v_{i,j} \leq \sum_{(i,j) \in I \times J} u_{i,j}$ .

Par contraposée si  $\sum_{(i,j)} v_{i,j}$  diverge, alors  $\sum_{(i,j)} u_{i,j}$  diverge aussi.

PROPOSITION 6.3. (**Sommation par paquets - résultat admis**) Soit  $\sum_{i,j} u_{i,j}$  où  $(i,j) \in I \times J$ , une série double **absolument convergente**, et  $(A_k)_{k \in K}$  une partition de  $I \times J$ , alors on a :

- pour tout  $k \in K$ , la série  $\sum_{(i,j)} u_{i,j}$  où  $(i,j) \in A_k$  est **absolument convergente** de somme  $\sum_{(i,j) \in A_k} u_{i,j}$ ,
- la série  $\sum_k \left( \sum_{(i,j) \in A_k} u_{i,j} \right)$  est **absolument convergente**,
- enfin on a :

$$\sum_{k \in K} \left( \sum_{(i,j) \in A_k} u_{i,j} \right) = \sum_{(i,j) \in I \times J} u_{i,j}$$

REMARQUE 6.4. (**essentielle**) Nous allons envisager trois partitions de  $\mathbf{N}^2$ , les plus usuelles car les plus naturelles, de manière à sommer par tranches ie par colonnes, par lignes, soit par diagonales.

- $(A_n)_{n \in \mathbf{N}}$  où pour tout  $n \in \mathbf{N}$ ,  $A_n = \{(i,n) \mid i \in \mathbf{N}\}$  et donc  $\mathbf{N}^2 = \bigsqcup_{n \in \mathbf{N}} (\mathbf{N} \times \{n\})$ .

Alors en cas de convergence  $\sum_{(i,j) \in A_n} u_{i,j} = \sum_{i=0}^{+\infty} u_{i,n}$  (sommation suivant la  $(n+1)^{\text{ème}}$  colonne<sup>1</sup>) et :

$$\sum_{(i,j) \in \mathbf{N}^2} u_{i,j} = \sum_{n=0}^{+\infty} \left( \sum_{i=0}^{+\infty} u_{i,n} \right)$$

- $(A_n)_{n \in \mathbf{N}}$  où pour tout  $n \in \mathbf{N}$ ,  $A_n = \{(n,j) \mid j \in \mathbf{N}\}$  et donc  $\mathbf{N}^2 = \bigsqcup_{n \in \mathbf{N}} (\{n\} \times \mathbf{N})$ .

Alors en cas de convergence  $\sum_{(i,j) \in A_n} u_{i,j} = \sum_{j=0}^{+\infty} u_{n,j}$  (sommation suivant la  $(n+1)^{\text{ème}}$  ligne<sup>2</sup>) et :

$$\sum_{(i,j) \in \mathbf{N}^2} u_{i,j} = \sum_{n=0}^{+\infty} \left( \sum_{j=0}^{+\infty} u_{n,j} \right)$$

- $(A_n)_{n \in \mathbf{N}}$  où pour tout  $n \in \mathbf{N}$ ,  $A_n = \{(i,j) \mid i+j = n\}$  et donc  $\mathbf{N}^2 = \bigsqcup_{n \in \mathbf{N}} \{(i,j) \in \mathbf{N}^2 \mid i+j = n\}$ .

Alors en cas de convergence  $\sum_{(i,j) \in A_n} u_{i,j} = \sum_{i=0}^n u_{i,n-i}$  (sommation suivant la  $(n+1)^{\text{ème}}$  diagonale)

<sup>1</sup>puisque l'on change de lignes.

<sup>2</sup>puisque l'on change de colonnes.

et :

$$\sum_{(i,j) \in \mathbf{N}^2} u_{i,j} = \sum_{n=0}^{+\infty} \left( \sum_{i=0}^n u_{i,n-i} \right)$$

REMARQUE 6.5. (1) Les deux premiers choix nous ramènent directement au **théorème de Fubini**, qui est donc un **corollaire** de ce théorème de sommation par paquets.

(2) Ce théorème permet de découper de manière arbitraire l'ensemble  $\mathbf{N}^2$  puis de sommer sur chaque sous-ensemble  $A_n$ , puis d'étudier la somme de la famille constituée par les sommes de chaque paquet d'où le nom du théorème. Vous comprenez bien que se pose mécaniquement le problème de la commutativité de la convergence dû au découpage d'où la nécessité absolue d'avoir la série  $\sum_{(i,j)} u_{i,j}$  absolument convergente.

## 2. Séries doubles à termes quelconques

DÉFINITION 6.3. Soit  $(u_{i,j})_{(i,j) \in I \times J}$  une famille de nombres réels de signe quelconque. On dit que la série double  $\sum_{(i,j)} u_{i,j}$  est **absolument convergente** si  $\sum_{(i,j)} |u_{i,j}|$  converge.

PROPOSITION 6.4. (**Interversion des sommations dans les séries doubles à termes quelconques : théorème de Fubini-Tonelli**) Soit  $u = (u_{i,j})_{(i,j) \in I \times J}$  une famille à deux indices de réels indexée par  $I \times J \subset \mathbf{N}^2$  telle que la série double  $\sum_{(i,j)} |u_{i,j}|$  converge. Dans cette condition la série  $\sum_{(i,j)} u_{i,j}$  converge et :

- Pour tout  $i \in I$ , la **série simple**  $\sum_j u_{i,j}$  est **convergente** ;
- Pour tout  $j \in J$ , la **série simple**  $\sum_i u_{i,j}$  est **convergente** ;
- on a l'égalité :

$$\sum_{(i,j) \in I \times J} u_{i,j} = \sum_{i \in I} \left( \sum_{j \in J} u_{i,j} \right) = \sum_{j \in J} \left( \sum_{i \in I} u_{i,j} \right)$$

REMARQUE 6.6. Ce théorème s'applique **systématiquement** lors de la **présentation** du **théorème de transfert**, donnant l'espérance d'une variable fonction d'un couple de variables. Maintenant si ces dernières sont positives, on se ramènera au théorème de Fubini dans le cas positif.

COROLLAIRE 6.2. Soit  $\sum_i a_i$  et  $\sum_j b_j$  deux séries simples absolument convergentes. Alors la série double  $\sum_{i,j} (a_i b_j)$  est absolument convergente et :

$$\sum_{(i,j) \in \mathbf{N}^2} a_i b_j = \left( \sum_{i \in \mathbf{N}} a_i \right) \left( \sum_{j \in \mathbf{N}} b_j \right)$$

## Vecteurs aléatoires discrets de $\mathbb{R}^n$

### 1. Le cas où $n = 2$ : couples de variables discrètes

DÉFINITION 7.1. (**Couple de variables aléatoires**) Soit  $X$  et  $Y$  deux variables aléatoires discrètes définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$ . On appelle **couple de variables aléatoires** l'application :

$$\begin{aligned} C : \Omega &\longrightarrow \mathbb{R}^2 \\ \omega &\longmapsto C(\omega) = (X(\omega), Y(\omega)) \end{aligned}$$

DÉFINITION 7.2. (**Loi d'un couple**) Soit  $X$  et  $Y$  deux variables aléatoires discrètes définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$ . On pose  $X(\Omega) = \{x_1, \dots, x_n, \dots\}$  et  $Y(\Omega) = \{y_1, \dots, y_m, \dots\}$ . On appelle la **loi du couple**  $C = (X, Y)$  ou **loi conjointe** de  $X$  et  $Y$  l'application :

$$\begin{aligned} P_C : X(\Omega) \times Y(\Omega) &\longrightarrow [0, 1] \\ (x_i, y_j) &\longmapsto P_C(x_i, y_j) = \mathbf{P}([X = x_i] \cap [Y = y_j]) = p_{i,j} \end{aligned}$$

#### Une caractérisation de la loi d'un couple

On caractérise la loi d'un couple  $C = (X, Y)$  en donnant  $C(\Omega)$  ou bien le produit cartésien  $X(\Omega) \times Y(\Omega)$  puis la probabilité de chaque couple de réalisations  $(x_i, y_j)$  du produit cartésien.

PROPOSITION 7.1. L'ensemble des couples  $\{(x_i, y_j) \mid x_i \in X(\Omega) \text{ et } y_j \in Y(\Omega)\}$  constitue un **système complet d'événements**.

DÉFINITION 7.3. (**Tribu associée à un couple**) La tribu  $\mathcal{A}_C$  engendrée par le système complet d'événements associée à  $C$  est appelée la **tribu associée à  $C$** .

$$\mathcal{A}_C = \left\{ \bigcup_{(x,y) \in I} ([X = x] \cap [Y = y]), I \subset X(\Omega) \times Y(\Omega) \right\}$$

PROPOSITION 7.2. Etant donné un couple de variables discrètes  $C = (X, Y)$ , nous avons :

$$\mathcal{A}_X \subset \mathcal{A}_C \quad \text{et} \quad \mathcal{A}_Y \subset \mathcal{A}_C$$

PROPOSITION 7.3. (**Fonction d'un couple de variables discrètes**) Soit  $(X, Y)$  un couple de variables discrètes définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  et soit  $g$  une fonction de  $\mathbb{R}^2$  dans  $\mathbb{R}$  définie sur  $X(\Omega) \times Y(\Omega)$ . Alors  $Z = g(X, Y)$  **reste** une variable aléatoire discrète sur  $\Omega$  et  $\mathcal{A}_Z \subset \mathcal{A}_C$ .

PROPOSITION 7.4. (**Caractérisation de la loi d'une fonction d'un couple aléatoire discret**) Notons  $Z = g(X, Y)$  dont les composantes  $X$  et  $Y$  sont toutes des variables aléatoires discrètes, définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$ . **Caractériser** la loi de la variable  $Z$ , c'est donner :

- (1)  $Z(\Omega) = g(X, Y)(\Omega)$  et
- (2) pour tout  $z$  de  $Z(\Omega)$  :

$$\mathbf{P}([Z = z]) = \sum_{\substack{(x,y) \in X(\Omega) \times Y(\Omega) \\ g(x,y) = z}} \mathbf{P}([X = x] \cap [Y = y])$$

PROPOSITION 7.5. *Pour toute partie  $I$  de  $Z(\Omega)$  :*

$$\mathbf{P}([Z \in I]) = \sum_{\substack{(x,y) \in X(\Omega) \times Y(\Omega) \\ g(x,y) \in I}} \mathbf{P}([X = x] \cap [Y = y])$$

COROLLAIRE 7.1. *Si  $X, Y$  sont des variables aléatoires discrètes, alors les applications suivantes de  $\Omega$  dans  $\mathbf{R}$  sont des variables aléatoires discrètes :*

- (1)  $\forall (\alpha, \beta) \in \mathbf{R}^2, \alpha X + \beta Y : \omega \in \Omega \mapsto (\alpha X + \beta Y)(\omega) = \alpha X(\omega) + \beta Y(\omega) \in \mathbf{R}$
- (2)  $X \times Y : \omega \in \Omega \mapsto (X \times Y)(\omega) = X(\omega) \times Y(\omega) \in \mathbf{R}$
- (3)  $\frac{X}{Y} : \omega \in \Omega \mapsto \frac{X}{Y}(\omega) = \frac{X(\omega)}{Y(\omega)} \in \mathbf{R}$  avec  $\forall \omega \in \Omega, Y(\omega) \neq 0$
- (4)  $\inf(X, Y) : \omega \in \Omega \mapsto \inf(X, Y)(\omega) = \min(X(\omega), Y(\omega)) \in \mathbf{R}$
- (5)  $\sup(X, Y) : \omega \in \Omega \mapsto \sup(X, Y)(\omega) = \max(X(\omega), Y(\omega)) \in \mathbf{R}$

REMARQUE 7.1. – Pour la loi de  $\sup(X, Y)$ , on calculera d'abord la probabilité  $\mathbf{P}([\sup(X, Y) \leq x])^1$  avant d'obtenir la probabilité ponctuelle.

– Pour la loi de  $\inf(X, Y)$ , on calculera d'abord la probabilité  $\mathbf{P}([\inf(X, Y) > x])^2$  avant d'obtenir la probabilité ponctuelle.

– Nous avons :

- $\inf(X, Y) = \frac{1}{2}(X + Y - |X - Y|)$  et
- $\sup(X, Y) = \frac{1}{2}(X + Y + |X - Y|)$  ainsi
- $\sup(X, Y) + \inf(X, Y) \stackrel{\mathcal{L}}{=} X + Y$  et
- $\sup(X, Y) - \inf(X, Y) \stackrel{\mathcal{L}}{=} |X - Y|$ .

DÉFINITION 7.4. (**Lois marginales**) *A tout couple de variables aléatoires correspond deux lois marginales celles de  $X$  et de  $Y$ .*

- (1) **La loi de  $X$  ou première loi marginale du couple** est l'application :

$$P_X : \begin{array}{ll} X(\Omega) & \longrightarrow [0, 1] \\ x_i & \longmapsto P_X(x_i) = \mathbf{P}([X = x_i]) \end{array}$$

- (2) **La loi de  $Y$  ou deuxième loi marginale du couple** est l'application :

$$P_Y : \begin{array}{ll} Y(\Omega) & \longrightarrow [0, 1] \\ y_j & \longmapsto P_Y(y_j) = \mathbf{P}([Y = y_j]) \end{array}$$

PROPOSITION 7.6. *Pour tout couple  $C = (X, Y)$  de variables aléatoires discrètes définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  on a :*

$$\forall x_i \in X(\Omega), \quad \mathbf{P}([X = x_i]) = p_{i\bullet} = \sum_{y_j \in Y(\Omega)} p_{i,j}$$

$$\text{et } \forall y_j \in Y(\Omega), \quad \mathbf{P}([Y = y_j]) = p_{\bullet j} = \sum_{x_i \in X(\Omega)} p_{i,j}$$

DÉFINITION 7.5. (**Lois conditionnelles**) *A tout couple correspond deux lois conditionnelles.*

<sup>1</sup>sup d'un couple et fonction de répartition "font bon ménage".

<sup>2</sup>inf d'un couple et antirépartition font "bon ménage".

- (1) Pour tout  $y_j$  de  $Y(\Omega)$  tel que  $\mathbf{P}([Y = y_j]) \neq 0$ , on appelle **loi conditionnelle de  $X$  sachant que  $[Y = \mathbf{y}_j]$**  l'application :

$$\begin{aligned} \mathbf{P}_{[Y=y_j]} : X(\Omega) &\longrightarrow [0, 1] \\ x_i &\longmapsto \mathbf{P}_{[Y=y_j]}([X = x_i]) = \frac{p_{i,j}}{p_{\bullet j}} \end{aligned}$$

- (2) Pour tout  $x_i$  de  $X(\Omega)$  tel que  $\mathbf{P}([X = x_i]) \neq 0$ , on appelle **loi conditionnelle de  $Y$  sachant que  $[X = \mathbf{x}_i]$**  l'application :

$$\begin{aligned} \mathbf{P}_{[X=x_i]} : Y(\Omega) &\longrightarrow [0, 1] \\ y_j &\longmapsto \mathbf{P}_{[X=x_i]}([Y = y_j]) = \frac{p_{i,j}}{p_{i\bullet}} \end{aligned}$$

Lorsque les deux variables constituant le couple  $C$  sont discrètes **finies**, on peut représenter la loi du couple et les deux lois marginales à l'aide d'un **tableau de contingence à deux entrées**.

$X \downarrow Y \longrightarrow$	$y_1$	$y_2$	$\cdots$	$y_m$	$\downarrow \mathbf{P}([X = x_i])$
$x_1$	$p_{1,1}$	$p_{1,2}$	$\cdots$	$p_{1,m}$	$p_{1\bullet}$
$x_2$	$p_{2,1}$	$p_{2,2}$	$\cdots$	$p_{2,m}$	$p_{2\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$
$x_{n-1}$	$p_{n-1,1}$	$p_{n-1,2}$	$\cdots$	$p_{n-1,m}$	$p_{n-1\bullet}$
$x_n$	$p_{n,1}$	$p_{n,2}$	$\cdots$	$p_{n,m}$	$p_{n\bullet}$
$\mathbf{P}([Y = y_j]) \rightarrow$	$p_{\bullet 1}$	$p_{\bullet 2}$	$\cdots$	$p_{\bullet m}$	$\mathbf{1}$

REMARQUE 7.2. Il est **toujours** possible de déterminer, les lois marginales à partir de la loi du couple. Mais la réciproque est **fausse sauf en cas d'indépendance des variables du couple**.

## 2. Ajustement linéaire entre deux variables aléatoires

Il nous est arrivé par le passé de rencontrer des couples de variables aléatoires dépendantes  $(X, Y)$  (non constantes sinon le cas est sans intérêt), mais nous ne nous sommes pas posé, jusqu'à maintenant, la question de la relation existante éventuellement entre celles-ci. À savoir se demander s'il existe une fonction  $f$  telle que  $Y = f(X)$ . Dans quel but ? Dans celui de prévoir  $Y$ , par exemple, une fois  $X$  connue. On dira dans le cas où  $f$  est trouvée, que  $X$  est la **variable explicative** et  $Y$  la **variable expliquée**. Une autre façon de poser le problème, consiste à dire que l'on s'intéresse à l'éventuelle nature de l'influence de  $X$  sur  $Y$ . Signalons que nous pourrions inverser la situation en s'interrogeant sur l'influence de  $Y$  sur  $X$ . Tout d'abord signalons que nous nous placerons dans le cas où  $X$  et  $Y$  **admettent un moment d'ordre deux**, soit lorsque  $(X, Y) \in \mathcal{L}^2$ . C'est sous ces hypothèses que l'on envisage classiquement le **problème de régression linéaire de  $Y$  en  $X$**  qui consiste à déterminer la fonction  $f$  affine définie par  $f(t) = at + b$  telle que le couple  $(a, b)$ , lorsqu'il parcourt  $\mathbf{R}^2$ , minimise la fonction  $S$  de deux variables  $a$  et  $b$  définie sur  $\mathbf{R}^2$  par :

$$S(a, b) = \mathbf{E} \left( (Y - aX - b)^2 \right)$$

Le problème de regression consiste à chercher à exprimer une variable  $X_1$  en fonction d'une autre  $X_2$  afin de voir si des observations fournies par  $X_2$  permettraient d'expliquer  $X_1$ .

PROPOSITION 7.7. La **droite de régression de  $Y$  en  $X$**  notée  $\Delta$  a pour équation :

$$\hat{Y} = \left( \frac{\text{Cov}(X, Y)}{\mathbf{V}(X)} \right) X + \left( \mathbf{E}(Y) - \frac{\mathbf{E}(X) \text{Cov}(X, Y)}{\mathbf{V}(X)} \right) = \left( \frac{\text{Cov}(X, Y)}{\mathbf{V}(X)} \right) (X - \mathbf{E}(X)) + \mathbf{E}(Y)$$

Et en inversant les rôles symétriques de  $X$  et  $Y$  nous obtenons :

$$\hat{X} = \left( \frac{\text{Cov}(X, Y)}{\mathbf{V}(Y)} \right) (Y - \mathbf{E}(Y)) + \mathbf{E}(X)$$

qui est l'équation de la **droite de régression de  $X$  en  $Y$**  notée  $\Delta'$ .

REMARQUE 7.3. (**hors programme**) Dans le cas général on montre que parmi toutes les fonctions de régression  $f : \mathbf{R} \rightarrow \mathbf{R}$ , celle minimisant  $\mathbf{E} \left( (Y - f(X))^2 \right)$  est définie sur le support de  $X$  par :

$$f(x) = \mathbf{E}(Y \mid [X = x])$$

### 3. Le cas où $n > 2$ : vecteurs discrets

Dans cette section nous généralisons tout ce qui a été vu auparavant.

DÉFINITION 7.6. (**Vecteur aléatoire**) Soit  $X_1, \dots, X_n$   $n$  variables aléatoires discrètes définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$ . On appelle **vecteur aléatoire de dimension  $n$**  l'application :

$$\begin{aligned} V : \Omega &\longmapsto \mathbf{R}^n \\ \omega &\longmapsto V(\omega) = (X_1(\omega), \dots, X_n(\omega)) \end{aligned}$$

DÉFINITION 7.7. (**Loi d'un vecteur aléatoire**) Soit  $n$  variables aléatoires discrètes  $X_1, \dots, X_n$  définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$ . On appelle la **loi du vecteur**  $V = (X_1, \dots, X_n)$  l'application :

$$\begin{aligned} P_V : X_1(\Omega) \times \dots \times X_n(\Omega) &\longrightarrow [0, 1] \\ (x_1, \dots, x_n) &\longmapsto P_V(x_1, \dots, x_n) = \mathbf{P}([X_1 = x_1] \cap \dots \cap [X_n = x_n]) \end{aligned}$$

PROPOSITION 7.8. (**Caractérisation de la loi d'un vecteur de  $\mathbf{R}^n$** ) Caractériser la loi du vecteur  $V = (X_1, \dots, X_n)$  dont les composantes  $X_k$ ,  $k \in \llbracket 1, n \rrbracket$  sont toutes des variables aléatoires discrètes, définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$ , c'est donner :

- (1)  $V(\Omega) = (X_1, \dots, X_n)(\Omega)$  qui est inclus dans  $\prod_{i=1}^n X_i(\Omega)$ . Maintenant vous pouvez vous contenter de donner l'univers-image  $X_i(\Omega)$  pour tout  $i$  de  $\{1, 2, \dots, n\}$ .
- (2)  $\forall x_1 \in X_1(\Omega), \dots, \forall x_n \in X_n(\Omega), \mathbf{P}([X_1 = x_1] \cap \dots \cap [X_n = x_n])$ .

DÉFINITION 7.8. (**Lois marginales**) Les  $n$  lois marginales de dimension un se définissent de la même manière que dans le cas d'un couple. A savoir pour tout  $k \in \llbracket 1, n \rrbracket$  la  $k^{\text{ème}}$  loi marginale du vecteur  $\mathbf{V} = (X_1, \dots, X_n)$  est l'application :

$$\begin{aligned} P_{X_k} : X_k(\Omega) &\longrightarrow [0, 1] \\ x_k &\longmapsto P_{X_k}(x_k) = \mathbf{P}([X_k = x_k]) \end{aligned}$$

PROPOSITION 7.9. Pour tout couple  $V = (X_1, \dots, X_n)$  de variables aléatoires discrètes définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  on a pour tout  $k$  de  $\llbracket 1, n \rrbracket$  et pour tout  $x_k$  de  $X_k(\Omega)$  :

$$P([X_k = x_k]) = \sum_{(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) \in \prod_{m \in \llbracket 1, n \rrbracket - \{i\}} X_m(\Omega)} \mathbf{P} \left( \bigcap_{j=1}^n [X_j = x_j] \right)$$

PROPOSITION 7.10. (**Fonction d'un vecteur aléatoire discret**) Soit  $V = (X_1, \dots, X_n)$  est une famille de variables aléatoires toutes définies sur le même espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  et  $\varphi$  une fonction définie sur une partie de  $\mathbf{R}^n$  contenant  $\prod_{k=1}^n X_k(\Omega)$ , alors  $\varphi \circ V$  reste une variable discrète.

PROPOSITION 7.11. (**Tribu associée**) Soit  $V = (X_1, \dots, X_n)$  est une famille de variables aléatoires toutes définies sur le même espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  et  $\varphi$  une fonction définie sur une partie de  $\mathbf{R}^n$  contenant  $\prod_{k=1}^n X_k(\Omega)$ , alors :

$$\mathcal{A}_{\varphi(V)} \subset \mathcal{A}_{(X_1, \dots, X_n)}$$

PROPOSITION 7.12. (*Caractérisation de la loi d'une fonction d'un vecteur aléatoire discret*) Soit  $V = (X_1, \dots, X_n)$  est une famille de variables aléatoires toutes définies sur le même espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  et  $\varphi$  une fonction définie sur une partie de  $\mathbf{R}^n$  contenant  $\prod_{k=1}^n X_k(\Omega)$ , alors la loi de  $Z = \varphi(V)$  est

caractérisée par la donnée de :

- $Z(\Omega) \subset \text{Im } \varphi$ ,
- et par celles de :

$$\forall z \in Z(\Omega), \quad \mathbf{P}([Z = z]) = \sum_{\substack{(x_1, \dots, x_n) \in X_1(\Omega) \times \dots \times X_n(\Omega) \\ \varphi(x_1, \dots, x_n) = z}} \mathbf{P}([X_1 = x_1] \cap \dots \cap [X_n = x_n])$$

PROPOSITION 7.13. (*Théorème de transfert*) Soit  $(X_1, \dots, X_n)$  est une famille de variables aléatoires toutes définies sur le même espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  et  $\varphi$  une fonction définie sur une partie de  $\mathbf{R}^n$  contenant  $\prod_{k=1}^n X_k(\Omega)$ , alors l'espérance de  $\varphi(X_1, \dots, X_n)$  existe si, et seulement si, la série de terme général :

$$\varphi(x_1, \dots, x_n) \mathbf{P}([X_1 = x_1] \cap \dots \cap [X_n = x_n]) \quad \text{où} \quad (x_1, \dots, x_n) \in \prod_{k=1}^n X_k(\Omega)$$

converge absolument, et dans ce cas :

$$\mathbf{E}(\varphi(X_1, \dots, X_n)) = \sum_{(x_1, \dots, x_n) \in X_1(\Omega) \times \dots \times X_n(\Omega)} \varphi(x_1, \dots, x_n) \mathbf{P}([X_1 = x_1] \cap \dots \cap [X_n = x_n])$$

REMARQUE 7.4. Bien qu'au programme officiel, l'utilisation de ce théorème admis n'est jamais tombé en concours, peut-être à cause de la série multiple en jeu.



## Moments d'une variable aléatoire

### 1. Le cas discret

DÉFINITION 8.1. (*Espérance mathématique ou espérance*) Soit  $X$  une variable aléatoire discrète définie sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$ .  $X$  admet une *espérance* lorsque la série de terme général  $x_k \mathbf{P}([X = x_k])$ , où  $x_k \in X(\Omega)$ , est absolument convergente. En cas de convergence de la série rencontrée, on appelle espérance de  $X$  le **nombre réel** noté  $\mathbf{E}(X)$  défini par :

$$\mathbf{E}(X) = \sum_{x_k \in X(\Omega)} x_k \mathbf{P}([X = x_k])$$

Interprétation : l'espérance d'une variable aléatoire est, lorsqu'elle existe, la moyenne des valeurs de cette variable pondérée par leurs probabilités de réalisation.

PROPOSITION 8.1. (*hors programme*) L'ensemble des variables discrètes admettant un moment d'ordre un est noté  $\mathcal{L}_d^1$ , muni des opérations  $+$  (interne) et  $\bullet$  (externe) est un espace vectoriel.

REMARQUE 8.1. (1) La convergence absolue est **suffisante** pour que la série soit **commutativement convergente** car si la série n'est pas à termes de signe fixé, sommer les éléments dans n'importe quel ordre peut provoquer de drôles de surprises concernant la convergence. Penser à la série de terme général  $\frac{(-1)^k}{k}$  qui converge, de somme  $-\ln 2$  alors qu'en regroupant les termes de la somme  $\sum_{k \geq 1} \frac{(-1)^k}{k}$  en deux paquets disjoints, celui des termes positifs (d'indices pairs) et celui des termes négatifs (d'indices impairs) nous "tombons" sur les séries  $\sum_{k \geq 1} \frac{1}{2k}$  et  $\sum_{k \geq 0} \frac{1}{2k+1}$  qui divergent !

- (2) Si  $X(\Omega) \subset \mathbb{R}_+$ , la série étant à **termes positifs**, il sera totalement **inutile** de travailler en valeur absolue.
- (3) Si  $X(\Omega)$  est un **ensemble fini**, l'existence de l'espérance de  $X$  ne pose absolument aucun problème. On écrira dans ce cas **directement** son expression sans aucune réserve de convergence à faire.

#### 1.1. Le théorème de transfert : son intitulé.

PROPOSITION 8.2. (*Théorème de transfert*)

- (1) (*Cas d'une fonction d'une seule variable aléatoire*) Soit  $X$  une variable aléatoire discrète définie sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  et soit  $g$  une fonction définie sur  $X(\Omega)$  à valeurs dans  $\mathbb{R}$ , alors  $g(X)$  **reste** une variable aléatoire discrète définie sur  $(\Omega, \mathcal{A}, \mathbf{P})$  et  $g(X)$  admet une espérance si et seulement si la série de terme général  $g(x_k) \mathbf{P}([X = x_k])$ , où  $x_k \in X(\Omega)$ , est absolument convergente. En cas de convergence l'espérance de  $g(X)$  est le nombre réel noté  $\mathbf{E}(g(X))$  défini par :

$$\mathbf{E}(g(X)) = \sum_{x_k \in X(\Omega)} g(x_k) \mathbf{P}([X = x_k])$$

- (2) (**Cas d'une fonction de deux variables aléatoires**) Soit  $X$  et  $Y$  deux variables aléatoires discrètes définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$ . On posera que  $X(\Omega) = \{x_i, i \in \mathbf{N}\}$  et  $Y(\Omega) = \{y_j, j \in \mathbf{N}\}$ . Soit  $\varphi$  une fonction de deux variables définie en tout point du produit cartésien  $X(\Omega) \times Y(\Omega)$ . Alors  $\mathbf{E}(\varphi(X, Y))$  est définie si, et seulement si la série de terme général  $\varphi(x_i, y_j) p_{i,j}$  est absolument convergente, auquel cas :

$$\mathbf{E}(\varphi(X, Y)) = \sum_{i=0}^{+\infty} \sum_{j=0}^{+\infty} \varphi(x_i, y_j) \mathbf{P}([X = x_i] \cap [Y = y_j]) = \sum_{j=0}^{+\infty} \sum_{i=0}^{+\infty} \varphi(x_i, y_j) \mathbf{P}([X = x_i] \cap [Y = y_j])$$

EXEMPLE 8.1. En cas d'existence :

- Pour entier naturel  $r$  on a  $\mathbf{E}(X^r) = \sum_{x_i \in X(\Omega)} x_i^r \mathbf{P}([X = x_i])$
- $\mathbf{E}(X + Y) = \sum_{x_i \in X(\Omega)} \sum_{y_j \in Y(\Omega)} (x_i + y_j) \mathbf{P}([X = x_i] \cap [Y = y_j])$
- $\mathbf{E}(X - Y) = \sum_{x_i \in X(\Omega)} \sum_{y_j \in Y(\Omega)} (x_i - y_j) \mathbf{P}([X = x_i] \cap [Y = y_j])$
- $\mathbf{E}(XY) = \sum_{x_i \in X(\Omega)} \sum_{y_j \in Y(\Omega)} x_i y_j \mathbf{P}([X = x_i] \cap [Y = y_j])$
- $\mathbf{E}\left(\frac{X}{Y}\right) = \sum_{x_i \in X(\Omega)} \sum_{y_j \in Y(\Omega)} \frac{x_i}{y_j} \mathbf{P}([X = x_i] \cap [Y = y_j])$  avec pour tout  $y_j \in Y(\Omega)$ ,  $y_j \neq 0$

## 1.2. Applications du théorème de transfert.

PROPOSITION 8.3. (1) (**Espérance de la transformation affine d'une variable discrète**)

Soit  $a$  et  $b$  deux réels tels que  $a$  soit non nul et  $X$  une variable aléatoire définie sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$ , alors  $aX + b$  admet une espérance **si, et seulement si**,  $X$  admet une espérance. Si l'une des deux conditions est réalisée, alors :

$$\mathbf{E}(aX + b) = a\mathbf{E}(X) + b$$

- (2)  $\mathbf{E}(X)$  existe **si, et seulement si**,  $\mathbf{E}(|X|)$  existe.

- (3) (**Espérance d'une variable bornée**) Soit  $X$  une variable aléatoire définie sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$ . S'il existe un couple  $(a, b) \in \mathbf{R}^2$  tel que  $a \leq X \leq b$  alors  $X$  admet une espérance avec  $a \leq \mathbf{E}(X) \leq b$ .

- (4) (**Linéarité de l'espérance**)

- (a) Soit  $X$  et  $Y$  deux variables aléatoires définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  **admettant chacune une espérance** et soit  $a, b$  deux réels quelconques. Alors  $aX + bY$  admet une espérance qui vaut :

$$\mathbf{E}(aX + bY) = a\mathbf{E}(X) + b\mathbf{E}(Y)$$

- (b) **Plus généralement** si  $X_1, X_2, \dots, X_n$  sont  $n$  variables discrètes **admettant chacune une espérance**, alors pour tout  $(a_1, a_2, \dots, a_n) \in \mathbf{R}^n$ ,  $\sum_{k=1}^n a_k X_k$  admet une espérance égale à :

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

- (5) (**Espérance et ordre**) Soit  $X$  et  $Y$  deux variables aléatoires définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  **admettant chacune une espérance**. Alors :

- (a) Si  $X \geq 0$  p.s. (i.e.  $\mathbf{P}([X \geq 0]) = 1$ ) alors  $\mathbf{E}(X) \geq 0$ .

- (b) Si  $X \leq Y$  p.s. (i.e.  $\mathbf{P}([X \leq Y]) = 1$ ) alors  $\mathbf{E}(X) \leq \mathbf{E}(Y)$  (on dit que l'espérance est un opérateur croissant).
- (c) Si  $X = Y$  p.s. (i.e.  $\mathbf{P}([X = Y]) = 1$ ) alors  $\mathbf{E}(X) = \mathbf{E}(Y)$ .
- (d) Si  $Y \in \mathcal{L}_d^1$  et  $|X| \leq Y$  p.s. alors  $X \in \mathcal{L}_d^1$  et  $\mathbf{E}(|X|) \leq \mathbf{E}(Y)$ .
- (e) Si  $X \in \mathcal{L}_d^1$  alors  $|\mathbf{E}(X)| \leq \mathbf{E}(|X|)$ .

REMARQUE 8.2. Les résultats 5.a, 5.b et 5.c restent valables lorsque respectivement  $X \geq 0$  (ie  $\forall \omega, X(\omega) \geq 0$ )  $X \leq Y$  (ie  $\forall \omega, X(\omega) \leq Y(\omega)$ ) et  $X = Y$  (ie  $\forall \omega, X(\omega) = Y(\omega)$ ).

### 1.3. Les différents moments.

DÉFINITION 8.2. (1) (**Moment d'ordre  $r \in \mathbf{N}$** ) Soit  $r$  un entier naturel et  $X$  une variable aléatoire définie sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$ . On dit que  $X$  admet un **moment d'ordre  $r$**  lorsque la série de terme général  $x_k^r \mathbf{P}([X = x_k])$ , où  $x_k \in X(\Omega)$ , est absolument convergente. En cas de convergence le moment d'ordre  $r$  de  $X$  est le **nombre réel** noté  $m_r(X)$  défini par :

$$m_r(X) = \sum_{x_k \in X(\Omega)} x_k^r \mathbf{P}([X = x_k])$$

- (2) (**Moment centré d'ordre  $r \in \mathbf{N}$** ) Soit  $r$  un entier naturel et  $X$  une variable aléatoire définie sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  admettant une espérance. On dit que  $X$  admet un **moment centré d'ordre  $r$**  lorsque la série de terme général  $(x_k - \mathbf{E}(X))^r \mathbf{P}([X = x_k])$ , où  $x_k \in X(\Omega)$ , est absolument convergente. En cas de convergence le moment centré d'ordre  $r$  de  $X$  est le **nombre réel** noté  $\mu_r(X)$  défini par :

$$\mu_r(X) = \sum_{x_k \in X(\Omega)} (x_k - \mathbf{E}(X))^r \mathbf{P}([X = x_k])$$

- (3) (**Moment factoriel d'ordre  $r \in \mathbf{N}^*$** ) Soit  $r$  un entier naturel non nul et  $X$  une variable aléatoire définie sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$ . On dit que  $X$  admet un **moment factoriel d'ordre  $r$**  lorsque la série de terme général  $x_k(x_k - 1) \times \cdots \times (x_k - r + 1) \mathbf{P}([X = x_k])$ , où  $x_k \in X(\Omega)$ , est absolument convergente. En cas de convergence le moment factoriel d'ordre  $r$  de  $X$  est le **nombre réel** noté  $\rho_r(X)$  défini par :

$$\rho_r(X) = \sum_{x_k \in X(\Omega)} x_k(x_k - 1) \times \cdots \times (x_k - r + 1) \mathbf{P}([X = x_k])$$

PROPOSITION 8.4. (**hors programme**) L'ensemble des variables discrètes admettant un moment d'ordre  $p$  est noté  $\mathcal{L}_d^p$ , muni des opérations  $+$  (interne) et  $\bullet$  (externe) est un espace vectoriel.

REMARQUE 8.3. (1) Les moments factoriels sont surtout employés dans le cas où la variable aléatoire  $X$  prend ses valeurs dans  $\mathbf{N}$ , factorielle oblige.

- (2) Plutôt que d'étudier la convergence de la série de terme général  $x_k(x_k - 1) \times \cdots \times (x_k - r + 1) \mathbf{P}([X = x_k])$  il suffit de voir si la variable  $X$  admet un moment d'ordre  $r$ . Car le développement du produit définissant le terme général précédent nous amène à une combinaison linéaire de termes généraux de séries convergentes d'après ce que nous verrons dans la proposition 8.5.
- (3) Les définitions 8.2 associées au **théorème de transfert**, nous permettent de dire qu'en cas d'existence :

- $\forall r \in \mathbf{N}, m_r(X) = \mathbf{E}(X^r)$
- $\forall r \in \mathbf{N}, \mu_r(X) = \mathbf{E}((X - \mathbf{E}(X))^r)$
- $\forall r \in \mathbf{N}^*, \rho_r(X) = \mathbf{E}(X(X - 1) \times \cdots \times (X - r + 1))$

- (4) L'espérance d'une variable aléatoire est son **moment d'ordre un**.

(5) La variance d'une variable aléatoire est son **moment centré d'ordre deux**.

PROPOSITION 8.5. Soit  $X$  une variable aléatoire définie sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$ .

- (1) **Si**  $X$  admet un moment d'ordre  $r + 1$ , **alors**  $X$  admet un moment d'ordre  $p$  pour tout  $p \in \llbracket 1, r \rrbracket$ .
- (2) **Si**  $X$  admet un moment centré d'ordre  $r + 1$ , **alors**  $X$  admet un moment centré d'ordre  $p$  pour tout  $p \in \llbracket 1, r \rrbracket$ .
- (3)  $X$  admet un moment d'ordre  $r$  **si, et seulement si**,  $X$  admet un moment centré d'ordre  $r$ .

COROLLAIRE 8.1. **Si**  $\mathbf{E}(X^2)$  existe **alors**  $\mathbf{E}(X)$  existe aussi. Conséquence  $\mathcal{L}_d^2 \subset \mathcal{L}_d^1$ ; Le premier point de la proposition 8.5 donne que  $\forall k \geq 2, \dots \subset \mathcal{L}_d^k \subset \dots \subset \mathcal{L}_d^2 \subset \mathcal{L}_d^1$ .

PROPOSITION 8.6. (**Moments d'ordre un et deux et antirépartition – hors programme mais connaître la démonstration**)

- (1) Soit  $X$  une variable aléatoire à valeurs entières positives définie sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$ .  $X$  admet une espérance **si, et seulement si**, la série de terme général  $\mathbf{P}([X > n])$ , où  $n \in \mathbf{N}$ , converge et en cas d'existence :

$$\mathbf{E}(X) = \sum_{n=0}^{+\infty} \mathbf{P}([X > n])$$

- (2) Soit  $X$  une variable aléatoire à valeurs entières positives définie sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$ .  $X$  admet un moment d'ordre deux **si, et seulement si**, la série de terme général  $(2n + 1) \mathbf{P}([X > n])$ , où  $n \in \mathbf{N}$ , converge et en cas d'existence :

$$\mathbf{E}(X^2) = \sum_{n=0}^{+\infty} (2n + 1) \mathbf{P}([X > n])$$

#### 1.4. Variance.

DÉFINITION 8.3. (**Variance – Ecart-type**) Soit  $X$  une variable aléatoire définie sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  **admettant une espérance**. On appelle **variance** de  $X$  l'espérance de  $(X - \mathbf{E}(X))^2$  lorsqu'elle existe que l'on note  $\mathbf{V}(X)$ . Autrement dit sous réserve d'existence :

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

La variance traduit la **dispersion d'une variable aléatoire autour de sa moyenne**. Définie à partir d'un carré, sa dimension n'est pas celle de  $X$ , c'est pourquoi nous utiliserons souvent sa racine carrée.

DÉFINITION 8.4. **L'écart-type** d'une variable aléatoire  $X$  définie sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  est le nombre réel noté  $\sigma(X)$  défini par :

$$\sigma(X) = \sqrt{\mathbf{V}(X)}$$

PROPOSITION 8.7. (**Le théorème de Huygens-Koenig**) Soit  $X$  une variable aléatoire définie sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  alors  $X$  admet une variance **si, et seulement si**,  $X^2$  admet une espérance. Si l'une des deux conditions est vérifiée, alors :

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

REMARQUE 8.4. De cette proposition découle que  $\mathbf{E}(X^2) \geq (\mathbf{E}(X))^2$  par la **positivité** de la variance.

PROPOSITION 8.8. (**Transformation affine**) Soit  $(a, b) \in \mathbb{R}^* \times \mathbb{R}$  et  $X$  une variable aléatoire définie sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$ . Alors  $aX + b$  admet un moment centrée d'ordre deux **si, et seulement si**,  $X$  admet un moment centrée d'ordre deux. Si l'une des deux conditions est vérifiée, alors :

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

DÉFINITION 8.5. (1) (**Variable centrée**) Une variable aléatoire discrète  $X$  définie sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  **admettant une espérance** est dite **centrée** lorsque :

$$\mathbf{E}(X) = 0$$

(2) (**Variable réduite**) Une variable aléatoire discrète  $X$  définie sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  **admettant une variance** est dite **réduite** si :

$$\mathbf{V}(X) = 1$$

(3) (**Variable centrée et réduite**) Soit une variable aléatoire discrète  $X$  définie sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  **admettant une variance non nulle** (et donc une espérance quelconque). On appelle **variable aléatoire centrée et réduite associée à  $X$**  la variable notée  $X^*$  définie par :

$$X^* = \frac{X - \mathbf{E}(X)}{\sigma(X)}$$

### 1.5. Espérances conditionnelles.

DÉFINITION 8.6. (**Espérance conditionnelle (I)**) Soit  $X$  une variable aléatoire discrète définie sur espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$ . Soit  $(A_n)_{n \in \mathbf{N}}$  un **système complet d'événements de probabilités non nulles**. On dit que  $X$  admet une **espérance conditionnelle sachant  $A_n$**  notée  $\mathbf{E}(X | A_n)$  lorsque la série simple  $\sum_{x_i} x_i \mathbf{P}_{A_n}([X = x_i])$  est absolument convergente. En cas de convergence :

$$\mathbf{E}(X | A_n) = \sum_{x_i \in X(\Omega)} x_i \mathbf{P}_{A_n}([X = x_i])$$

PROPOSITION 8.9. (**Formule de l'espérance totale**) Soit  $X$  une variable aléatoire discrète définie sur espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$ . Soit  $(A_n)_{n \in \mathbf{N}}$  un **système complet d'événements de probabilités non nulles**.  $X$  admet une espérance **si, et seulement si**,  $X$  admet une espérance pour la probabilité conditionnelle  $\mathbf{P}_{A_n}$  notée  $\mathbf{E}(X | A_n)$  et la série simple  $\sum_{n \geq 0} \mathbf{E}(X | A_n) \mathbf{P}(A_n)$  converge absolument. En cas de convergence on a :

$$\mathbf{E}(X) = \sum_{n=0}^{+\infty} \mathbf{E}(X | A_n) \mathbf{P}(A_n)$$

DÉFINITION 8.7. (**Espérance conditionnelle (II)**)

(1) Pour tout  $x_i$  de  $X(\Omega)$  tel que  $\mathbf{P}([X = x_i]) \neq 0$ , on appelle **espérance conditionnelle de  $Y$  sachant que  $[X = x_i]$**  l'espérance mathématique de  $Y$  pour la loi conditionnelle de  $Y$  si l'événement  $[X = x_i]$  est réalisé, le réel noté  $\mathbf{E}(Y | [X = x_i])$  et définie par :

$$\mathbf{E}(Y | [X = x_i]) = \sum_{j \in J} y_j \frac{p_{i,j}}{p_{i \bullet}}$$

en cas de convergence absolue de la série rencontrée.

(2) Pour tout  $y_j$  de  $Y(\Omega)$  tel que  $\mathbf{P}([Y = y_j]) \neq 0$ , on appelle **espérance conditionnelle de  $X$  sachant que  $[Y = y_j]$** , l'espérance mathématique de  $X$  pour la loi conditionnelle de  $X$  si l'événement  $[Y = y_j]$  est réalisé, le réel noté  $\mathbf{E}(X | [Y = y_j])$  et définie par :

$$\mathbf{E}(X | [Y = y_j]) = \sum_{i \in I} x_i \frac{p_{i,j}}{p_{\bullet j}}$$

en cas de convergence absolue de la série rencontrée.

PROPOSITION 8.10. (**Cas de l'indépendance**) Soit  $X$  et  $Y$  deux variables aléatoires **indépendantes** définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$ . Alors en cas d'existence :

$$\mathbf{E}(X | [Y = y_j]) = \mathbf{E}(X) \quad \text{et} \quad \mathbf{E}(Y | [X = x_i]) = \mathbf{E}(Y)$$

PROPRIÉTÉS 8.1. Soit  $X, Y, Z$  et  $T$  quatre variables discrètes,  $g$  une fonction de deux variables et  $c$  une constante.

- (1)  $\mathbf{E}(c | [X = c]) = c$ .
- (2)  $\mathbf{E}(Z + T | [X = x]) = \mathbf{E}(Z | [X = x]) + \mathbf{E}(T | [X = x])$ .
- (3)  $\mathbf{E}(cY | [X = x]) = c\mathbf{E}(Y | [X = x])$ .
- (4)  $\mathbf{E}(g(X, Y) | [X = x]) = \mathbf{E}(g(x, Y) | [X = x])$ .
- (5)  $|\mathbf{E}(Y | [X = x])| \leq \mathbf{E}(|Y| | [X = x])$ .
- (6) Si  $Z \leq Y$  alors  $\mathbf{E}(Z | [X = x]) \leq \mathbf{E}(Y | [X = x])$ .

### 1.6. Covariance – Matrice de covariance-variance.

DÉFINITION 8.8. (**Covariance**) Soit  $X, Y \in \mathcal{L}_d^2$ , on appelle **covariance** de  $X$  et  $Y$  que l'on note  $\text{Cov}(X, Y)$  l'espérance de  $(X - \mathbf{E}(X))(Y - \mathbf{E}(Y))$ . Autrement dit :

$$\text{Cov}(X, Y) = \mathbf{E}((X - \mathbf{E}(X))(Y - \mathbf{E}(Y)))$$

En particulier :

$$\text{Cov}(X, X) = \mathbf{V}(X)$$

REMARQUE 8.5. Par sa définition, la covariance donne une certaine idée du lien existant entre les deux variables  $X$  et  $Y$ .

PROPRIÉTÉS 8.2. (1) Soit  $X, Y \in \mathcal{L}_d^2$  alors :

$$\text{Cov}(X, Y) = \mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y)$$

C'est la **formule de Koenig** pour la covariance.

(2) Si  $X, Y$  sont deux variables **indépendantes** définies sur un même espace probabilisé, alors :

$$\text{Cov}(X, Y) = 0$$

On dit que les variables sont non corrélées. La réciproque est **fausse sauf pour les variables au plus binaires ; par exemple le cas de variables de Bernoulli**.

(3) La covariance est une **forme bilinéaire symétrique et positive** sur l'espace vectoriel  $\mathcal{L}_d^2$ , c'est à dire pour tout  $X, Y, Z \in \mathcal{L}_d^2$  et  $\lambda, \mu \in \mathbb{R}$ , on a :

- (a)  $\text{Cov}(X, Y) = \text{Cov}(Y, X)$  (**symétrie**)
- (b)  $\text{Cov}(\lambda X + \mu Y, Z) = \lambda \text{Cov}(X, Z) + \mu \text{Cov}(Y, Z)$  (**linéarité à gauche** et par symétrie la **bilinéarité**)
- (c)  $\text{Cov}(X, X) = \mathbf{V}(X) \geq 0$  (**positivité**)

(4) Soit  $X_1, \dots, X_n, Y_1, \dots, Y_m \in \mathcal{L}_d^2$  et  $\lambda_1, \dots, \lambda_n, \mu_1, \dots, \mu_m \in \mathbb{R}$  alors :

$$\text{Cov}\left(\sum_{i=1}^n \lambda_i X_i, \sum_{j=1}^m \mu_j Y_j\right) = \sum_{(i,j) \in [1,n] \times [1,m]} \lambda_i \mu_j \text{Cov}(X_i, Y_j)$$

(5) Soit  $X_1, \dots, X_n \in \mathcal{L}_d^2$  et  $\lambda_1, \dots, \lambda_n \in \mathbb{R}$  alors :

$$\mathbf{V} \left( \sum_{i=1}^n \lambda_i X_i \right) = \sum_{i=1}^n \lambda_i^2 \mathbf{V}(X_i) + 2 \sum_{1 \leq i < j \leq n} \lambda_i \lambda_j \text{Cov}(X_i, X_j)$$

(6) Soit  $X, Y \in \mathcal{L}_d^2$  alors on a les identités de polarisation :

$$(a) \quad \mathbf{V}(X) + \mathbf{V}(Y) = \frac{1}{2} (\mathbf{V}(X+Y) + \mathbf{V}(X-Y))$$

$$(b) \quad \text{Cov}(X, Y) = \frac{1}{4} (\mathbf{V}(X+Y) - \mathbf{V}(X-Y))$$

$$(c) \quad \text{Cov}(X, Y) = \frac{1}{2} (\mathbf{V}(X+Y) - \mathbf{V}(X) - \mathbf{V}(Y))$$

$$(d) \quad \text{Cov}(X, Y) = \frac{1}{2} (\mathbf{V}(X) + \mathbf{V}(Y) - \mathbf{V}(X-Y))$$

(7) Soit  $X_1, \dots, X_n \in \mathcal{L}_d^2$  **indépendantes** et  $n$  réels  $\lambda_1, \dots, \lambda_n$ , alors :

$$\mathbf{V} \left( \sum_{i=1}^n \lambda_i X_i \right) = \sum_{i=1}^n \lambda_i^2 \mathbf{V}(X_i)$$

Attention la réciproque est **fausse**.

REMARQUE 8.6. Les identités (6b), (6c) et (6d) sont appelées **identités de polarisation**.

REMARQUE 8.7. Puisque nous sommes en train de parler de covariance, il est bon de savoir qu'elle permet de définir un produit scalaire **si les variables sont centrées**. Retenez que l'application :

$$\begin{aligned} (\mathcal{L}_d^2)^2 &\longrightarrow \mathbb{R} \\ (X, Y) &\longmapsto \mathbf{E}(XY) \end{aligned}$$

définit un **produit scalaire sur  $\mathcal{L}_d^2$** .

DÉFINITION 8.9. (**Matrice de covariance**) Soit  $V = (X_1, X_2, \dots, X_n)$  un vecteur aléatoire dont les composantes  $X_k$  sont dans  $\mathcal{L}_d^2$  et définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$ . On appelle **matrice de covariance** la **matrice symétrique réelle**  $S = (s_{i,j})_{(i,j) \in [1,n]^2}$  définie par :

$$\forall (i, j) \in [1, n]^2, \quad s_{i,j} = \text{Cov}(X_i, X_j)$$

soit :

$$S = \begin{bmatrix} \mathbf{V}(X_1) & \text{Cov}(X_1, X_2) & \cdots & \text{Cov}(X_1, X_n) \\ \text{Cov}(X_2, X_1) & \mathbf{V}(X_2) & \ddots & \vdots \\ \vdots & \ddots & \ddots & \text{Cov}(X_{n-1}, X_n) \\ \text{Cov}(X_n, X_1) & \cdots & \text{Cov}(X_n, X_{n-1}) & \mathbf{V}(X_n) \end{bmatrix}$$

PROPOSITION 8.11. Si les composantes de  $V$  sont deux à deux indépendantes, alors :

$$S = \text{diag}(\mathbf{V}(X_1), \dots, \mathbf{V}(X_n))$$

PROPOSITION 8.12. (**Utilité de S**) Soit  $U = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \in \mathcal{M}_{n,1}(\mathbb{R})$  et  $V = (X_1, X_2, \dots, X_n)$  un vecteur

aléatoire discret possédant une matrice de covariance. Alors la variable discrète  $\sum_{k=1}^n a_k X_k$  admet un moment

d'ordre deux et on a :

$$\mathbf{V}\left(\sum_{k=1}^n a_k X_k\right) = {}^t U S U$$

La quantité  ${}^t U S U$  s'appelle **forme quadratique associée à la matrice S**. Elle est **positive** par positivité de la variance et donc  $S$  admet des **valeurs propres positives** et comme  $\mathbf{V}\left(\sum_{i=1}^n a_i X_i\right) = 0$  n'entraîne pas que  $\sum_{i=1}^n a_i X_i = 0$  on dit que **la matrice de covariance est semi-définie positive**.

REMARQUE 8.8. La proposition précédente montre que la connaissance de la matrice de covariance d'un vecteur  $V$  permet le calcul de  $\mathbf{V}(\varphi(V))$  pour toute forme linéaire  $\varphi : (t_1, \dots, t_n) \in \mathbb{R}^n \mapsto \varphi(t_1, \dots, t_n) = \sum_{i=1}^n \alpha_i t_i \in \mathbb{R}$  (où  $(\alpha_1, \dots, \alpha_n) \in \mathbb{R}^n$ ).

### 1.7. Coefficient de corrélation linéaire.

DÉFINITION 8.10. (**Coefficient de corrélation linéaire d'un couple de variables**) Soit  $X, Y \in \mathcal{L}_d^2$ , **non constantes**, donc de variances non nulles, définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$ . On appelle **coefficient de corrélation linéaire de X et Y, de Bravais-Pearson, le nombre réel noté  $\rho_{X,Y}$  ou  $\rho(X, Y)$  défini par :**

$$\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sigma(X)\sigma(Y)}$$

REMARQUE 8.9. Le réel  $\rho_{X,Y}$  "mesure" grossièrement le **degré de liaison** entre les deux variables  $X$  et  $Y$ .

PROPOSITION 8.13. (**Inégalité de Cauchy-Schwarz**) Soit  $X, Y \in \mathcal{L}_d^2$ , alors la variable  $XY$  admet une espérance et on a :

$$|\mathbf{E}(XY)| \leq \sqrt{\mathbf{E}(X^2)}\sqrt{\mathbf{E}(Y^2)}$$

avec **égalité si, et seulement si**, il existe deux réels  $a$  et  $b$  dont l'un au moins n'est pas nul tel que :

$$\mathbf{P}([aX = bY]) = 1$$

REMARQUE 8.10.  L'existence de l'espérance de la variable  $XY$  résulte directement du fait que :

$$|XY| \leq \frac{1}{2}(X^2 + Y^2)$$

en partant de l'inégalité  $(|X| - |Y|)^2 \geq 0$  que vous développez.

PROPRIÉTÉS 8.3. (1)  $\rho_{X,Y} \in [-1, 1]$  (découle directement de l'inégalité de Cauchy-Schwarz).

(2)  $\rho_{X,Y} = \pm 1$  **si, et seulement si**,  $X, Y$  sont presque sûrement liées par une **relation affine** d'où **le nom** de coefficient de corrélation linéaire. Autrement dit on a l'équivalence :

$$(\exists (a, b) \in \mathbb{R}^* \times \mathbb{R} \mid \mathbf{P}([Y = aX + b]) = 1) \iff (\rho_{X,Y} = \pm 1)$$

(3) **Si**  $X, Y$  sont **indépendantes** alors  $\rho_{X,Y} = 0$  (la réciproque est **fausse**) on dit que les **variables sont non corrélées**.

## 2. Le cas continu

DÉFINITION 8.11. (**Espérance**) Soit  $X$  une variable aléatoire à densité définie sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  de densité associée  $f$ . Alors  $X$  admet une **espérance** lorsque l'intégrale  $\int_{-\infty}^{+\infty} xf(x) dx$  est absolument convergente. En cas de convergence, on appelle espérance de  $X$  le nombre réel noté  $\mathbf{E}(X)$  défini par :

$$\mathbf{E}(X) = \int_{-\infty}^{+\infty} xf(x) dx$$

REMARQUE 8.11. Notez bien que **la convergence absolue n'est pas nécessaire** puisque l'intégrande ne change pas de signe sur  $\mathbf{R}_+$  et sur  $\mathbf{R}_-$ . En revanche cela ne serait pas le cas, par exemple, pour  $\int_{-\infty}^{+\infty} (x-1)f(x) dx$  où il faudrait travailler en valeur absolue.

PROPOSITION 8.14. (**Espérance et antirépartition – hors programme et à redémontrer avant emploi**) Dans le cas où  $X$  est une variable à densité à valeurs positives de fonction de répartition  $F_X$ , alors en cas de convergence :

$$\mathbf{E}(X) = \int_0^{+\infty} \mathbf{P}([X > x]) dx = \int_0^{+\infty} (1 - F_X(x)) dx$$

PROPOSITION 8.15. (**Théorème de transfert – admis**) Soit  $X$  une variable aléatoire à densité sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  de densité associée  $f$ . Soit  $\varphi$  une fonction continue sur un intervalle  $I$  contenant  $X(\Omega)$  et d'extrémités  $a$  et  $b$  avec  $-\infty \leq a < b \leq +\infty$  sauf éventuellement en un nombre fini de points, alors  $\varphi(X)$  admet une espérance notée  $\mathbf{E}(\varphi(X))$  si, et seulement si,  $\int_a^b \varphi(x)f(x) dx$  est absolument convergente. En cas de convergence :

$$\mathbf{E}(\varphi(X)) = \int_a^b \varphi(x)f(x) dx$$

PROPOSITION 8.16. (**admise**) Soit  $X$  et  $Y$  deux variables aléatoires à densité définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  admettant chacune une espérance. Alors :

- (1) Si  $X \geq 0$  p.s. alors  $\mathbf{E}(X) \geq 0$ .
- (2) Si  $X \leq Y$  p.s. alors  $\mathbf{E}(X) \leq \mathbf{E}(Y)$  (on dit que l'espérance est un opérateur croissant).

REMARQUE 8.12. Ces résultats restent valables lorsque  $X \geq 0$  (ie  $\forall \omega, X(\omega) \geq 0$ ) respectivement  $X \leq Y$  (ie  $\forall \omega, X(\omega) \leq Y(\omega)$ ).

PROPOSITION 8.17. (**Espérance de la transformation affine d'une variable à densité**) Soit  $X$  une variable aléatoire à densité définie sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$ , alors pour tout couple  $(a, b) \in \mathbf{R}^* \times \mathbf{R}$ ,  $aX + b$  admet une espérance si, et seulement si,  $X$  admet une espérance. Si l'une des deux conditions est réalisée, alors :

$$\mathbf{E}(aX + b) = a\mathbf{E}(X) + b$$

PROPOSITION 8.18. (**Linéarité de l'espérance – admise**) Soit  $n$  variables aléatoires à densité  $X_1, X_2, \dots, X_n$ , définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  et admettant chacune une espérance. Soit  $n$  réels  $\lambda_1, \dots, \lambda_n$ , alors on admet que  $\sum_{k=1}^n \lambda_k X_k$  possède une espérance qui vaut :

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

DÉFINITION 8.12. (**Moment d'ordre  $r \in \mathbf{N}$** ) Soit  $r$  un entier naturel et  $X$  une variable aléatoire à densité définie sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  de densité associée  $f$ . Alors  $X$  admet un **moment d'ordre  $r$**  si l'intégrale  $\int_{-\infty}^{+\infty} x^r f(x) dx$  est convergente. En cas de convergence, on appelle moment d'ordre  $r$  de  $X$  le nombre réel noté  $m_r(X)$  défini par :

$$m_r(X) = \int_{-\infty}^{+\infty} x^r f(x) dx$$

REMARQUE 8.13. Notez bien que **la convergence absolue n'est pas nécessaire** puisque l'intégrande ne change pas de signe sur  $\mathbf{R}_+$  et sur  $\mathbf{R}_-$ , et ce, quelle que soit la parité de  $r$ .

REMARQUE 8.14. (**Hors programme**) Pour tout entier naturel  $r$  non nul, on notera  $\mathcal{L}_c^r$  l'espace vectoriel des variables à densité admettant un moment d'ordre  $r$ .

DÉFINITION 8.13. (**Moment centré d'ordre  $r \in \mathbf{N}$** ) Soit  $r$  un entier naturel et  $X$  une variable aléatoire à densité définie sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  possédant une **espérance** et de densité associée  $f$ . Alors  $X$  admet un **moment centré d'ordre  $r$**  si l'intégrale  $\int_{-\infty}^{+\infty} (x - \mathbf{E}(X))^r f(x) dx$  est absolument convergente. En cas de convergence, on appelle moment centré d'ordre  $r$  de  $X$  le nombre réel noté  $\mu_r(X)$  défini par :

$$\mu_r(X) = \int_{-\infty}^{+\infty} (x - \mathbf{E}(X))^r f(x) dx$$

REMARQUE 8.15. D'après ces définitions, l'espérance de  $X$  est le moment d'ordre un de  $X$  et sa variance le **moment centré d'ordre deux**.

PROPOSITION 8.19. Soit  $X$  une variable aléatoire à densité définie sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$ .

- (1) **Si** la variable  $X$  admet un moment d'ordre  $r + 1$ , **alors**  $X$  admet un moment d'ordre  $p$  pour tout  $p \in \llbracket 1, r \rrbracket$ . En corollaire l'existence de  $\mathbf{E}(X^2)$  entraîne celle de  $\mathbf{E}(X)$ .
- (2) **Si** la variable  $X$  admet un moment centré d'ordre  $r + 1$ , **alors**  $X$  admet un moment centré d'ordre  $p$  pour tout  $p \in \llbracket 1, r \rrbracket$ .
- (3) La variable  $X$  admet un moment d'ordre  $r$  **si, et seulement si**,  $X$  admet un moment centré d'ordre  $r$ .

PROPOSITION 8.20. (**fondamentale**) Une variable presque sûrement bornée admet des moments de tous ordres.

PROPOSITION 8.21. (**Moment d'ordre un et deux et antirépartition — hors programme et à redémontrer avant emploi**) Dans le cas où  $X$  est une variable à densité à **valeurs positives** de fonction de répartition  $F_X$ , alors en cas de convergence :

- (1) (**Espérance et antirépartition**)

$$\mathbf{E}(X) = \int_0^{+\infty} \mathbf{P}([X > x]) dx = \int_0^{+\infty} (1 - F_X(x)) dx$$

- (2) (**Moment d'ordre deux et antirépartition**)

$$\mathbf{E}(X^2) = \int_0^{+\infty} 2x \mathbf{P}([X > x]) dx = \int_0^{+\infty} 2x(1 - F_X(x)) dx$$

DÉFINITION 8.14. (**Variance — Écart-type**) Soit  $X$  une variable aléatoire à densité définie sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  possédant une **espérance** et de densité associée  $f$ .

- On dit que  $X$  admet une **variance** si l'intégrale  $\int_{-\infty}^{+\infty} (x - \mathbf{E}(X))^2 f(x) dx$  est convergente. En cas de convergence, on appelle variance de  $X$  le nombre réel noté  $\mathbf{V}(X)$  défini par :

$$\mathbf{V}(X) = \int_{-\infty}^{+\infty} (x - \mathbf{E}(X))^2 f(x) dx$$

- On appelle **écart-type** de  $X$  le nombre réel noté  $\sigma(X)$  défini par :

$$\sigma(X) = \sqrt{\mathbf{V}(X)}$$

PROPOSITION 8.22. (**Théorème de Huygens-Koenig**)  $X$  admet une variance **si, et seulement si**,  $X$  admet un moment d'ordre deux. Si l'une des deux conditions est remplie alors :

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

PROPOSITION 8.23. (**Variance de la transformation affine d'une variable à densité**) Soit  $X$  une variable aléatoire à densité définie sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$ , alors pour tout couple  $(a, b) \in \mathbf{R}^* \times \mathbf{R}$ ,  $aX + b$  admet un moment d'ordre deux **si, et seulement si**,  $X$  admet un moment d'ordre deux. Si l'une des deux conditions est réalisée, alors :

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

DÉFINITION 8.15. (**Variable à densité centrée réduite**) Soit  $X$  une variable aléatoire à densité définie sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  possédant une espérance et de variance non nulle. Alors :

$$X^* = \frac{X - \mathbf{E}(X)}{\sigma(X)}$$

est appelée **variable centrée réduite** associée à  $X$ .

REMARQUE 8.16. La définition reste la même dans le cas discret :

- Une variable à densité est **centrée** lorsque son espérance est **nulle**.
- Une variable à densité est **réduite** lorsque sa variance est égale à **1**.



## Indépendance stochastique

### 1. Indépendance d'événements

DÉFINITION 9.1. (1) (**Cas de deux événements**) Soit  $A$  et  $B$  deux événements d'un même espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$ , alors  $A$  et  $B$  sont **indépendants** lorsque :

$$\mathbf{P}(A \cap B) = \mathbf{P}(A) \mathbf{P}(B)$$

(2) (**Cas de  $n$  événements**)

(a) On dit que  $n$  événements  $A_1, \dots, A_n$  d'un même espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  sont **deux à deux indépendants** lorsque :

$$\forall (i, j) \in \llbracket 1, n \rrbracket^2, \quad (i \neq j) \implies (\mathbf{P}(A_i \cap A_j) = \mathbf{P}(A_i) \mathbf{P}(A_j))$$

(b) On dit que  $n$  événements  $A_1, \dots, A_n$  d'un même espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  sont **mutuellement indépendants**<sup>1</sup> lorsque :

$$\forall I \in \mathcal{P}(\llbracket 1, n \rrbracket), \quad I \neq \emptyset, \quad \mathbf{P}\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} \mathbf{P}(A_i)$$

**NB** : Il y a  $\sum_{k=2}^n \binom{n}{k} = 2^n - n - 1$  vérifications à faire.

(3) (**Cas d'une suite d'événements**) Soit  $(A_n)_{n \geq 0}$  une suite d'événements définie sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$ .

(a) On dit que  $(A_n)_{n \geq 0}$  est une **suite d'événements deux à deux indépendants** lorsque :

$$\forall (i, j) \in \mathbf{N}^2, \quad (i \neq j) \implies (\mathbf{P}(A_i \cap A_j) = \mathbf{P}(A_i) \mathbf{P}(A_j))$$

(b) On dit que  $(A_n)_{n \geq 0}$  est une **suite d'événements mutuellement indépendants** lorsque :

$$\forall I \in \mathcal{P}(\mathbf{N}), \quad I \text{ fini}, \quad I \neq \emptyset, \quad \mathbf{P}\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} \mathbf{P}(A_i)$$

REMARQUE 9.1. **L'indépendance mutuelle entraîne l'indépendance deux à deux, mais la réciproque est fautive.**

PROPOSITION 9.1. (1) Soit  $A$  et  $B$  deux événements **indépendants** définis sur un même espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$ , alors les propositions suivantes sont **équivalentes** :

- Les événements  $A$  et  $\overline{B}$  sont indépendants ;
- Les événements  $\overline{A}$  et  $B$  sont indépendants ;
- Les événements  $\overline{A}$  et  $\overline{B}$  sont indépendants.

<sup>1</sup>On travaille par paquets d'événements.

- (2) Soit  $(A_n)_{n \geq 1}$  une suite d'événements deux à deux indépendants (resp. mutuellement indépendants), alors **il en est de même** de la suite d'événements  $(B_n)_{n \geq 1}$  définie pour tout  $n \in \mathbf{N}^*$  par  $B_n = A_n$  ou  $\overline{A_n}$ .

- REMARQUE 9.2. (1) Ne pas confondre événements indépendants et événements incompatibles.
- (2) Si  $A$  est un événement tel que  $\mathbf{P}(A) = 0$  ou  $\mathbf{P}(A) = 1$ , alors il est indépendant de tout événement, y compris de lui-même. C'est le cas en particulier pour  $\Omega$  et  $\emptyset$ .
- (3) Deux événements incompatibles  $A$  et  $B$  de probabilités strictement positives ne sont jamais indépendants puisque  $\mathbf{P}(A \cap B) = 0$  alors que  $\mathbf{P}(A)\mathbf{P}(B) \neq 0$ .
- (4) L'indépendance stochastique de deux événements dépend toujours de l'espace probabilisé choisi, ce n'est donc pas une propriété intrinsèques aux événements.
- (5) **L'indépendance stochastique n'est pas transitive.** Autrement dit soit  $A, B, C$  trois événements définis sur un même espace probabilisé, alors :

$$\left( \begin{array}{l} A \text{ et } B \text{ indépendants} \\ B \text{ et } C \text{ indépendants} \end{array} \right) \not\Rightarrow (A \text{ et } C \text{ indépendants})$$

## 2. Indépendance de variables aléatoires

### 2.1. Le cas général.

- DÉFINITION 9.2. (1) (**Cas de deux variables**) On dit que deux variables aléatoires discrètes  $X_1$  et  $X_2$  définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  sont **indépendantes** pour la probabilité  $\mathbf{P}$  lorsque pour tout couple  $(B_1, B_2)$  de boréliens de  $\mathbf{R}$  on a :

$$\mathbf{P}([X_1 \in B_1] \cap [X_2 \in B_2]) = \mathbf{P}([X_1 \in B_1]) \times \mathbf{P}([X_2 \in B_2])$$

- (2) (**Cas de  $n$  variables**) On dit que  $n$  variables aléatoires  $(X_i)_{i \in [1, n]}$  définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  sont :
- (a) **Indépendantes deux à deux** lorsque pour tout couple  $(i, j) \in \llbracket 1, n \rrbracket^2$ ,  $i \neq j$  les variables  $X_i$  et  $X_j$  sont indépendantes.
- (b) **Mutuellement indépendantes** pour tout  $k \in \llbracket 1, n \rrbracket$  et  $(i_1, \dots, i_k) \in \llbracket 1, n \rrbracket^k$  avec  $i_1 < \dots < i_k$  et quel que soit les boréliens  $B_{i_1}, \dots, B_{i_k}$  on a :

$$\mathbf{P}([X_{i_1} \in B_{i_1}] \cap \dots \cap [X_{i_k} \in B_{i_k}]) = \mathbf{P}([X_{i_1} \in B_{i_1}]) \times \dots \times \mathbf{P}([X_{i_k} \in B_{i_k}])$$

- (3) (**Cas d'une suite de variables aléatoires**) On dit qu'une suite de variables aléatoires discrètes  $(X_n)_{n \in \mathbf{N}}$  définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  est constituée de **variables indépendantes** lorsque **toute sous-suite finie** est formée de variables indépendantes.
- (4) Lorsque les variables  $(X_n)_{n \in \mathbf{N}}$  sont **indépendantes** et toutes de **même loi**, on dit qu'elles sont **indépendantes et identiquement distribuées (i.i.d.)** et l'on écrit que les **variables sont i.i.d.**

PROPOSITION 9.2. (**admis**) Soit  $V = (X_1, \dots, X_n)$  un vecteur aléatoire de variables aléatoires réelles discrètes mutuellement indépendantes définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$ . Alors quel que soit l'entier  $p$  de  $\llbracket 2, n-1 \rrbracket$  les tribus respectivement associées à  $(X_1, \dots, X_p)$  et  $(X_{p+1}, \dots, X_n)$  sont indépendantes.

PROPOSITION 9.3. (**Lemme des coalitions - admis**) Soit une famille  $(X_1, \dots, X_n, X_{n+1}, \dots, X_{n+m})$  de **variables aléatoires indépendantes** toutes définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$ . Alors **toute fonction** de  $(X_1, X_2, \dots, X_n)$  est **indépendante** de **toute fonction** de  $(X_{n+1}, X_{n+2}, \dots, X_{n+m})$ . On peut **généraliser en mélangeant** les variables  $X_k$  et en prenant **autant de fonctions que l'on veut**.

## 2.2. Le cas discret.

DÉFINITION 9.3. (1) (**Cas de deux variables**) On dit que deux variables aléatoires discrètes  $X_1$  et  $X_2$  définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  sont **indépendantes** lorsque pour tout couple  $(x_1, x_2)$  de  $X_1(\Omega) \times X_2(\Omega)$  on a :

$$\mathbf{P}([X_1 = x_1] \cap [X_2 = x_2]) = \mathbf{P}([X_1 = x_1]) \times \mathbf{P}([X_2 = x_2])$$

(2) (**Cas de  $n$  variables**) On dit que  $n$  variables aléatoires discrètes  $(X_i)_{i \in [1, n]}$  définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  sont **indépendantes** lorsque pour tout  $k$ -uplet  $(x_1, \dots, x_n)$  de  $\prod_{i=1}^n X_i(\Omega)$  on a :

$$\mathbf{P}([X_1 = x_1] \cap \dots \cap [X_n = x_n]) = \mathbf{P}([X_1 = x_1]) \times \dots \times \mathbf{P}([X_n = x_n])$$

(3) (**Cas d'une suite de variables aléatoires**) On dit qu'une suite de variables aléatoires discrètes  $(X_n)_{n \in \mathbf{N}}$  définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  est constituée de **variables indépendantes** lorsque **toute sous-suite finie** est formée de variables indépendantes.

(4) Lorsque les variables  $(X_n)_{n \in \mathbf{N}}$  sont **indépendantes** et toutes de **même loi**, on dit qu'elles sont **indépendantes et identiquement distribuées (i.i.d.)** et l'on écrit que les **variables sont i.i.d.**

REMARQUE 9.3. (1) Dans le cas de l'indépendance on dira que "la loi du vecteur  $V(X_1, \dots, X_n)$  est égale au produit des lois marginales<sup>2</sup>" et on notera que  $P_V = \bigotimes_{i \in [1, n]} P_{X_i}$ .

(2) On notera que la définition de l'indépendance de  $n$  variables aléatoires n'est pas en accord avec l'indépendance de  $n$  événements. En effet le lecteur aura sûrement constaté que l'on ne travaille pas par paquets de variables constituant des sous-familles de  $(X_1, \dots, X_n)$  mais sur toutes les variables dans leur ensemble.

PROPOSITION 9.4. (**Lemme des coalitions - admis**) Soit une famille  $(X_1, \dots, X_n, X_{n+1}, \dots, X_{n+m})$  de **variables aléatoires indépendantes** toutes définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$ . Alors **toute fonction** de  $(X_1, X_2, \dots, X_n)$  est **indépendante** de **toute fonction** de  $(X_{n+1}, X_{n+2}, \dots, X_{n+m})$ . On peut **généraliser en mélangeant** les variables  $X_k$  et en prenant **autant de fonctions que l'on veut**.

REMARQUE 9.4.  Faites très attention la réciproque de l'implication :

$$(X \text{ et } Y \text{ indépendantes}) \implies (\varphi(X) \text{ et } \psi(Y) \text{ indépendantes})$$

est **FAUSSE!** En effet prenez  $Y = X^2$  et  $\varphi$  et  $\psi$  deux fonctions constantes différentes.

## 2.3. Le cas continu.

DÉFINITION 9.4. (**Variables à densité indépendantes**)

(1) (**Indépendance de deux variables à densité**) Soit  $X$  et  $Y$  deux variables à densité définies sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$ . On dit qu'elles sont **indépendantes** lorsque :

$$\forall (x, y) \in \mathbb{R}^2, \quad \mathbf{P}([X \leq x] \cap [Y \leq y]) = \mathbf{P}([X \leq x]) \mathbf{P}([Y \leq y])$$

(2) (**Indépendance de  $n$  variables à densité**) Soit  $X_1, \dots, X_n$ ,  $n$  variables à densité définies sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$ . On dit qu'elles sont **mutuellement indépendantes ou indépendantes** lorsque pour tout  $(x_1, \dots, x_n) \in \mathbb{R}^n$  :

$$\mathbf{P}([X_1 \leq x_1] \cap \dots \cap [X_n \leq x_n]) = \mathbf{P}([X_1 \leq x_1]) \times \dots \times \mathbf{P}([X_n \leq x_n])$$

<sup>2</sup>Autrement dit pour un couple  $(X, Y)$ ,  $\forall x_i \in X(\Omega)$ ,  $\forall y_j \in Y(\Omega)$ ,  $p_{i,j} = p_{i \bullet} \times p_{\bullet j}$ .

- (3) (**Indépendance d'une suite de variables à densité**) Soit  $(X_n)_n$  une suite de variables à densité définies sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$ . On dit que cette suite est constituée de variables **indépendantes** lorsque pour toute partie  $I$  finie non vide de  $\mathbf{N}$ ,  $(X_n)_{n \in I}$  est constituée de variables indépendantes.

PROPOSITION 9.5. (**Lemme des coalitions**) Soit  $n+m$  variables à densité  $X_1, \dots, X_n, X_{n+1}, \dots, X_{n+m}$  définies sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$ , qui soient **indépendantes**, alors toute fonction de  $X_1, \dots, X_n$  est **indépendante** de toute fonction de  $X_{n+1}, \dots, X_{n+m}$ .

### 3. Formule de convolution

Que soit dans le cas discret ou le cas continu, la formule de convolution permet d'obtenir la loi de la somme de deux lois indépendantes.

#### 3.1. Le cas discret.

PROPOSITION 9.6. Soit  $C = (X, Y)$  un couple de deux variables aléatoires discrètes définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$ . On note  $S = X + Y$  et  $I_s = \{(x_i, y_j) \in X(\Omega) \times Y(\Omega) \mid x_i + y_j = s\}$ . Alors on a  $S(\Omega) \subset \{x_i + y_j \mid x_i \in X(\Omega), y_j \in Y(\Omega)\}$  et :

$$\forall s \in S(\Omega), \quad \mathbf{P}([S = s]) = \sum_{(x_i, y_j) \in I_s} \mathbf{P}([X = x_i] \cap [Y = y_j])$$

Si de plus  $X$  et  $Y$  sont **indépendantes** :

$$\forall s \in S(\Omega), \quad \mathbf{P}([S = s]) = \sum_{\substack{x_i \in X(\Omega) \\ s - x_i \in Y(\Omega)}} \mathbf{P}([X = x_i] \cap [Y = s - x_i]) = \sum_{\substack{y_j \in Y(\Omega) \\ s - y_j \in X(\Omega)}} \mathbf{P}([X = s - y_j] \cap [Y = y_j])$$

On dira que la loi de  $X + Y$  est la **convolée** de la loi de  $X$  et de celle de  $Y$  et on note  $P_{X+Y} = P_X * P_Y = P_Y * P_X$ .

REMARQUE 9.5. On passe d'une somme à l'autre en échangeant les rôles joués par  $X$  et  $Y$ .

#### 3.2. Le cas continu.

PROPOSITION 9.7. (**Somme de deux variables à densité indépendantes - admise**) Soit  $X$  et  $Y$  deux variables aléatoires à densité indépendantes sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  de densités respectives  $f_X$  et  $f_Y$ . Si la fonction  $h : x \mapsto h(x) = \int_{-\infty}^{+\infty} f_X(t) f_Y(x-t) dt$  ou bien  $h : x \mapsto h(x) = \int_{-\infty}^{+\infty} f_X(x-t) f_Y(t) dt$  est définie et continue presque partout, alors on admet  $X + Y$  est une **variable à densité** et  $h$  en est une densité. On dira que la loi de  $X + Y$  est la **convolée** de la loi de  $X$  et de celle de  $Y$  et on note  $h = f_{X+Y} = f_X * f_Y = f_Y * f_X$ .

REMARQUE 9.6. On passe de la première à la seconde intégrale par le changement de variable affine  $u \mapsto x - t$  où  $x$  est fixé dans  $\mathbf{R}$ .

COROLLAIRE 9.1. Dans le cas fréquent où  $X$  et  $Y$  sont à **valeurs positives** (ex : lois expos, lois gammas), il suffit de remarquer que  $f_Y(x-t)$  est nul si  $x$  dépasse  $t$  et que  $f_X(t)$  est nul si  $t$  est négatif. Ainsi le produit  $f_X(t) f_Y(x-t)$  ne peut être non nul que pour  $x$  vérifiant  $0 \leq t \leq x$ , ce qui n'est évidemment jamais vérifié si  $x$  est négatif. On fera donc uniquement le calcul de  $\int_0^x f_X(t) f_Y(x-t) dt$  pour  $x \geq 0$ ,  $h$  étant nulle sur  $\mathbf{R}_-^*$ .

$$h(x) = \left( \int_0^x f_X(t) f_Y(x-t) dt \right) \mathbf{1}_{\mathbf{R}_+}(x) = \left( \int_0^x f_X(x-t) f_Y(t) dt \right) \mathbf{1}_{\mathbf{R}_+}(x)$$

REMARQUE 9.7. **importante** : On peut démontrer que si  $(X, Y) \in (\mathcal{L}_c^2)^2$  ou si au moins une des deux densités  $f_X, f_Y$  est bornée alors  $f_{X+Y}$  est continue presque partout.

REMARQUE 9.8.  (**Faux-ami**) La somme de variables à densité **ne redonne pas** forcément une variable à densité. Pensez tout simplement à  $X - X$  !

#### 4. Indépendance et moments

PROPOSITION 9.8. (**admis en continu**)

- (1) Soit  $X$  et  $Y$  deux variables aléatoires à densité **indépendantes** définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  **admettant chacune une espérance**. Alors  $XY$  admet une espérance qui vaut :

$$\mathbf{E}(XY) = \mathbf{E}(X) \times \mathbf{E}(Y)$$

- (2) Soit  $X_1, X_2, \dots, X_n$ ,  $n$  variables aléatoires **finies et indépendantes** définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  **admettant chacune une espérance**, alors  $\prod_{k=1}^n X_k$  admet une espérance et on montre par récurrence que :

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

PROPOSITION 9.9. (**admis en continu**) Soit  $X$  et  $Y$  deux variables aléatoires à densité **indépendantes** définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  et soit  $\varphi$  une fonction définie sur  $X(\Omega)$  et sur  $Y(\Omega)$ . Alors si  $\varphi(X)$  et  $\varphi(Y)$  **restent** des variables aléatoires à densité alors elles sont **indépendantes** et **si elles admettent chacune une espérance**,  $\varphi(X) \times \varphi(Y)$  admet aussi une espérance qui vaut :

$$\mathbf{E}(\varphi(X) \times \varphi(Y)) = \mathbf{E}(\varphi(X)) \times \mathbf{E}(\varphi(Y))$$

On peut même pousser la remarque avec deux fonctions  $\varphi$  et  $\psi$  définies sur  $X(\Omega)$  (respectivement  $Y(\Omega)$ ) telles que  $\varphi(X)$  et  $\psi(Y)$  **admettent chacune une espérance**. Alors  $\varphi(X) \times \psi(Y)$  admet une espérance qui vaut :

$$\mathbf{E}(\varphi(X) \times \psi(Y)) = \mathbf{E}(\varphi(X)) \times \mathbf{E}(\psi(Y))$$

PROPOSITION 9.10. (**admise en continu**) Soit  $X_1, X_2, \dots, X_n$ ,  $n$  variables aléatoires à densité **indépendantes** définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  **admettant chacune une variance** et  $\alpha_1, \alpha_2, \dots, \alpha_n$ ,  $n$  réels. Alors la variable  $\sum_{k=1}^n \alpha_k X_k$  possède une variance qui vaut :

$$\mathbf{V}\left(\sum_{k=1}^n \alpha_k X_k\right) = \sum_{k=1}^n \alpha_k^2 \mathbf{V}(X_k)$$



## Les lois de probabilités usuelles

### 1. Lois discrètes

#### 1.1. Loi de Dirac ou loi d'une variable quasi-certaine.

DÉFINITION 10.1. (**Loi de Dirac**) Une variable aléatoire  $X$  suit une **loi de Dirac** de paramètre  $C \in \mathbb{R}$  si  $\mathbf{P}([X = C]) = 1$  ce que l'on note  $X \hookrightarrow \delta_C$ . On dit écrit que  $X = C$  p.s. (**presque surement**), on dit aussi que  $X$  est la variable **quasi-certaine** égale à  $C$ .

EXEMPLE 10.1. Les variables  $\mathbf{1}_\emptyset$  et  $\mathbf{1}_\Omega$  sont des variables de Dirac.

La loi de Dirac est la plus simple des lois existantes, elle est associée à un phénomène déterministe<sup>1</sup> dont le résultat de toute expérience permettant de l'étudier est le même, égal à  $C$ .

PROPOSITION 10.1. (**Espérance et variance**) Si  $X \hookrightarrow \delta_C$  alors :

$$\mathbf{E}(X) = C \quad \text{et} \quad \mathbf{V}(X) = 0$$

DÉFINITION 10.2. (**hors programme**) Une loi est qualifiée de **loi dégénérée** lorsque la variance est nulle.

REMARQUE 10.1. (**essentielle**) Une variable à variance nulle n'est pas forcément presque surement nulle! En revanche une variable ayant un moment d'ordre deux nul est presque surement nulle.

PROPOSITION 10.2. Nous avons l'équivalence :

$$(\mathbf{V}(X) = 0) \iff (X \hookrightarrow \delta_C)$$

#### 1.2. Variable uniforme

##### . Modèle probabiliste de l'urne

- Soit une urne contenant  $n$  boules numérotées de 1 à  $n$ . On tire une boule de l'urne.
- On introduit  $\mathbf{X}$  une variable aléatoire associée au **numéro de la boule obtenue**.

DÉFINITION 10.3. Soit  $a$  et  $b$  deux entiers relatifs tels que  $a \leq b$  On dit qu'une variable aléatoire  $X$  suit la **loi uniforme** de paramètre  $\llbracket a, b \rrbracket$  ce que l'on note  $X \hookrightarrow \mathcal{U}(\llbracket a, b \rrbracket)$  lorsque :

$$X(\Omega) = \llbracket a, b \rrbracket \quad \text{et} \quad \forall k \in \llbracket a, b \rrbracket, \quad \mathbf{P}([X = k]) = \frac{1}{b - a + 1}$$

DÉFINITION 10.4. (**Cas particulier**) On dit que suit la **loi uniforme** de paramètre  $\llbracket 1, n \rrbracket$ , ce que l'on note  $X \hookrightarrow \mathcal{U}(\llbracket 1, n \rrbracket)$ , lorsque :

- (1)  $X(\Omega) = \llbracket 1, n \rrbracket$
- (2)  $\forall k \in \llbracket 1, n \rrbracket, \mathbf{P}([X = k]) = \frac{1}{n}$

---

<sup>1</sup>**Un phénomène déterministe** est un phénomène où la connaissance de la situation initiale et d'un ensemble de règles d'évolution permet de déterminer entièrement quelle sera la situation future en fonction du temps.

PROPOSITION 10.3. (**Transformation affine**)

$$(X \hookrightarrow \mathcal{U}(\llbracket a, b \rrbracket)) \iff (X - a + 1 \hookrightarrow \mathcal{U}(\llbracket 1, b - a + 1 \rrbracket))$$

PROPOSITION 10.4. (**Espérance et variance**) Une variable uniforme admet des moments de tous ordres en tant que variable finie.

- (1) En particulier, si  $X \hookrightarrow \mathcal{U}(\llbracket 1, n \rrbracket)$  alors  $\mathbf{E}(X) = \frac{n+1}{2}$  et  $\mathbf{V}(X) = \frac{n^2-1}{12}$ .
- (2) Si  $X \hookrightarrow \mathcal{U}(\llbracket a, b \rrbracket)$  alors  $\mathbf{E}(X) = \frac{a+b}{2}$  et  $\mathbf{V}(X) = \frac{(b-a+1)^2-1}{12}$

### 1.3. Loi de Bernoulli ou loi indicatrice

#### . Modèle probabiliste général : épreuve de Bernoulli

- On réalise une expérience pouvant amener deux issues possibles, l'une interprétée comme un succès et l'autre comme un échec.
- On définit la variable aléatoire  $\mathbf{X}$  prenant la valeur 1 en cas de succès et 0 lors d'un échec.

#### Cas particulier : modèle probabiliste de l'urne

Soit une urne contenant des boules blanches et des boules noires en proportions respectives  $p$  et  $q$  avec  $0 < p < 1$  et  $p + q = 1$ . On tire **une** boule de l'urne.

On introduit  $\mathbf{X}$  une variable aléatoire associée à la **couleur de la boule obtenue**. Les résultats potentiels seront "codés" :

- obtenir une boule blanche sera assimilé au succès et correspondra à l'événement  $[X = 1]$  ;
- obtenir une boule noire sera assimilé à l'échec et correspondra à l'événement  $[X = 0]$ .

DÉFINITION 10.5. (**Loi de Bernoulli**) On dit que  $X$  suit la **loi de Bernoulli** de paramètre  $p$  ce que l'on note  $X \hookrightarrow \mathcal{B}(p)$ , lorsque :

- (1)  $X(\Omega) = \{0, 1\}$
- (2)  $\mathbf{P}([X = 1]) = p$ ,  $\mathbf{P}([X = 0]) = 1 - p$

REMARQUE 10.2. Une telle variable  $X$  est la **variable indicatrice** de l'événement  $X^{-1}(\{1\})$ .

PROPOSITION 10.5. (**Espérance et variance**) Si  $X \hookrightarrow \mathcal{B}(p)$  alors :

$$\mathbf{E}(X) = p \quad \text{et} \quad \mathbf{V}(X) = p(1-p)$$

REMARQUE 10.3. (**fondamentale**) Les calculs des moments de  $X$  sont facilités par le fait que pour tout  $n \geq 1$ ,  $X^n = X$ .

DÉFINITION 10.6. (**Variable indicatrice**) On appelle **variable indicatrice (ou indicateur) d'un événement**  $A$  la variable de Bernoulli notée  $\mathbf{1}_A$  définie par :

$$\mathbf{1}_A = \begin{cases} 1 & \text{si l'événement } A \text{ est réalisé} \\ 0 & \text{sinon} \end{cases}$$

En particulier :  $\mathbf{1}_\Omega = 1$  et  $\mathbf{1}_\emptyset = 0$ .

PROPOSITION 10.6. Nous avons clairement :

$$\mathbf{E}(\mathbf{1}_A) = \mathbf{P}(A) \quad \text{et} \quad \mathbf{V}(\mathbf{1}_A) = \mathbf{P}(A)(1 - \mathbf{P}(A))$$

REMARQUE 10.4. L'égalité  $\mathbf{E}(\mathbf{1}_A) = \mathbf{P}(A)$ , permettant de transformer une probabilité en une espérance et inversement, associée aux **propriétés de croissance et de linéarité** de l'opérateur espérance nous permettent d'obtenir de nombreuses inégalités probabilistes comme celles de Markov, Bienaymé-Tchebychev, Laplace, Chernoff...

PROPRIÉTÉS 10.1. (1)  $\mathbf{1}_\Omega = 1$

(2)  $\mathbf{1}_\emptyset = 0$

(3)  $\mathbf{1}_{\bar{A}} = 1 - \mathbf{1}_A$

(4)  $\mathbf{1}_{A \cap B} = \mathbf{1}_A \mathbf{1}_B$  (cela se généralise au cas de  $n$  événements par récurrence)

(5)  $\forall n \in \mathbf{N}^*, (\mathbf{1}_A)^n = \mathbf{1}_A$

(6)  $\mathbf{1}_{A \cup B} = \mathbf{1}_A + \mathbf{1}_B - \mathbf{1}_A \mathbf{1}_B$  (cela se généralise au cas de  $n$  événements par récurrence)

(7)  $\mathbf{1}_{A \Delta B} = |\mathbf{1}_A - \mathbf{1}_B|$

CONSÉQUENCE 10.1. Si  $X \hookrightarrow \mathcal{B}(p)$  alors pour tout  $n \in \mathbf{N}^*$ ,  $X^n = X$  où  $X \hookrightarrow \mathcal{B}(p)$

PROPOSITION 10.7. (**Décomposition canonique d'une variable aléatoire discrète**) Toute variable aléatoire  $X$  discrète finie définie sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  peut se présenter de façon unique sous la forme  $X = \sum_{k \in K} x_k \mathbf{1}_{A_k}$  où la suite  $(x_k)_{k \in K}$  est **strictement croissante** et les  $A_k = \{\omega \in \Omega \mid X(\omega) = x_k\}$  non vides forment une **partition** de  $\Omega$ .

REMARQUE 10.5. Cette proposition permet de démontrer que l'espace vectoriel des variables discrètes finies prenant  $n$  valeurs distinctes est de dimension  $n$ .

#### 1.4. Loi binomiale ou loi des tirages avec remise.

##### 1.4.1. Séquences de $n$ tirages bernoulliens.

DÉFINITION 10.7. On effectue une succession de  $n$  épreuves de Bernoulli **indépendantes** et de **même paramètre**  $p$  (probabilité du succès). L'univers du schéma de Bernoulli sera  $\Omega^n$  où  $\Omega = \{E, S\}$  ( $S$  pour succès et  $E$  pour échec)

DÉFINITION 10.8. (**avec le modèle de l'urne**) Soit une urne contenant une proportion  $p$  de boules blanches et une proportion  $q$  de boules noires. Après l'épreuve  $E_1$  consistant à extraire une boule de l'urne, on remet la boule tirée dans l'urne et on procède à un second tirage dans des conditions identiques, etc... On obtient ainsi une séquence  $E_1, \dots, E_n$  de tirages dits "**tirages bernoulliens**" ou avec remise.

REMARQUE 10.6. Ceci pourrait se généraliser à une suite infinie de tirages bernoulliens.

##### 1.4.2. Loi binomiale

###### Modèle probabiliste général

- On effectue  $n$  essais indépendants d'une épreuve aléatoire n'ayant que deux issues possibles : **soit** le succès avec une probabilité  $p \in ]0, 1[$  **soit** l'échec avec la probabilité  $q = 1 - p$ .
- On introduit  $X$  une variable aléatoire associée au nombre de succès obtenus au cours des  $n$  essais.

###### Modèle probabiliste de l'urne

- On tire  $n$  boules, **avec remise** après chaque tirage, dans une urne bicolore contenant des boules blanches en proportion  $p \in ]0, 1[$  et des boules noires en proportion  $q$  tel que  $p + q = 1$ .
- On introduit  $X$  une variable aléatoire associée au **nombre de boules blanches obtenues au cours des  $n$  tirages**.

DÉFINITION 10.9. On dit qu'une variable aléatoire  $X$  suit la **loi binomiale** de paramètres  $n$  et  $p$  ce que l'on note  $X \hookrightarrow \mathcal{B}(n, p)$ , lorsque :

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

(2)  $\forall k \in \llbracket 0, n \rrbracket, \mathbf{P}([X = k]) = \binom{n}{k} p^k q^{n-k}$

PROPOSITION 10.8. (**Espérance et variance**) Si  $X \hookrightarrow \mathcal{B}(n, p)$  alors :

$$\mathbf{E}(X) = np \quad \text{et} \quad \mathbf{V}(X) = np(1 - p)$$

PROPOSITION 10.9. (*Stabilité de la loi binomiale pour la somme de variables indépendantes et de même paramètre  $p$* )

- (1) (**Cas de deux variables**) Soit  $X$  et  $Y$  deux variables aléatoires **indépendantes** définies sur le même espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  et telles que  $X \hookrightarrow \mathcal{B}(n, p)$  et  $Y \hookrightarrow \mathcal{B}(n', p)$  alors :

$$X + Y \hookrightarrow \mathcal{B}(n + n', p)$$

- (2) (**Cas de  $n$  variables**) Soit  $X_1, X_2, \dots, X_m$ ,  $m$  variables aléatoires **indépendantes** définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  telles que pour tout  $k \in \llbracket 1, m \rrbracket$ ,  $X_k \hookrightarrow \mathcal{B}(n_k, p)$ . Alors :

$$\sum_{k=1}^m X_k \hookrightarrow \mathcal{B}\left(\sum_{k=1}^m n_k, p\right)$$

- (3) (**Somme de  $n$  variables de Bernoulli indépendantes de même paramètre  $p$** ) Soit  $X_1, X_2, \dots, X_n$ ,  $n$  variables aléatoires **indépendantes** définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  telles que pour tout  $k \in \llbracket 1, n \rrbracket$ ,  $X_k \hookrightarrow \mathcal{B}(p)$ . Alors :

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

### 1.5. Loi hypergéométrique ou loi des tirages sans remise

#### . Modèle probabiliste de l'urne

- Soit une urne bicolore contenant  $N$  boules dont une proportion  $p$ ,  $p \in ]0, 1[$  de boules blanches et une proportion  $q$ ,  $q \in ]0, 1[$  de boules noires telles que  $p + q = 1$ , avec  $Np$ ,  $Nq$  deux entiers naturels non nuls.
- On tire  $n$  boules de l'urne **une à une et sans remise**. On introduit  $X$  une variable aléatoire associée au **nombre de boules blanches obtenues au cours des  $n$  tirages**.

DÉFINITION 10.10. (**Loi hypergéométrique**) On dit qu'une variable aléatoire  $X$  suit la **loi hypergéométrique** de paramètres  $N, n, p$  ce que l'on note  $X \hookrightarrow \mathcal{H}(N, n, p)$ , lorsque :

- (1)  $X(\Omega) \subset \llbracket 0, n \rrbracket$  plus précisément  $X(\Omega) = \llbracket \max(0, n - N(1 - p)), \min(Np, n) \rrbracket$
- (2)  $\forall k \in \llbracket 0, n \rrbracket$ ,  $\mathbf{P}([X = k]) = \frac{\binom{Np}{k} \binom{N(1-p)}{n-k}}{\binom{N}{n}}$

PROPOSITION 10.10. (**Espérance et variance**) Si  $X \hookrightarrow \mathcal{H}(N, n, p)$  alors :

$$\mathbf{E}(X) = np \quad \text{et} \quad \mathbf{V}(X) = \left(\frac{N-n}{N-1}\right) np(1-p)$$

REMARQUE 10.7. (1) Le réel  $\frac{N-n}{N-1}$  s'appelle le **coefficient d'exhaustivité** de la loi.

- (2)  $\lim_{N \rightarrow +\infty} \mathbf{V}(X) = np(1-p)$  ainsi lorsque le nombre de boules de l'urne croit indéfiniment, l'influence de la remise ou de la non remise est **négligeable**.

- (3) Le nom de la loi qui peut vous paraître bizarre vient du fait que l'on retrouve le rapport  $\frac{\binom{n_1}{k} \binom{n_2}{n-k}}{\binom{n_1+n_2}{n}}$  dans le développement en série entière d'une fonction tabulée : la **fonction hypergéométrique**.

### 1.6. Loi géométrique ou loi du temps d'attente du premier succès

#### . Modèle probabiliste général

- On considère une suite d'épreuves de Bernoulli **indépendantes** sans limitation de nombre et de **même paramètre  $p$**  jusqu'à obtenir un succès (de probabilité  $p \in ]0, 1[$ ) **pour la première fois**.
- On introduit  $X$  une variable aléatoire associée au **rang d'apparition du premier succès**.

**Modèle probabiliste de l'urne**

- On effectue une succession illimitée de tirages d'une boule avec remise à chaque fois, à partir d'une urne bicolore constituée d'une proportion  $p$  de boules blanches et  $q$  de boules noires avec  $p \in ]0, 1[$  et  $p + q = 1$ , jusqu'à obtenir une boule blanche pour la première fois.
- On introduit  $X$  une variable aléatoire associée au **rang d'apparition de la première boule blanche**.

REMARQUE 10.8. *Les résultats des tirages obtenus sont indépendants.*

DÉFINITION 10.11. (**Loi géométrique**) On dit qu'une variable aléatoire suit la **loi géométrique** de paramètre  $p \in ]0, 1[$ , ce que l'on note  $X \hookrightarrow \mathcal{G}(p)$ , lorsque :

- (1)  $X(\Omega) = \mathbf{N}^*$
- (2)  $\forall k \in \mathbf{N}^*, \mathbf{P}([X = k]) = q^{k-1}p$

REMARQUE 10.9. *Le nom de la loi vient du fait que les éléments de la suite  $(\mathbf{P}([X = k]))_{k \in \mathbf{N}^*}$  sont en progression géométrique.*

PROPOSITION 10.11. (**Espérance et variance**) Si  $X \hookrightarrow \mathcal{G}(p)$  alors :

$$\mathbf{E}(X) = \frac{1}{p} \quad \text{et} \quad \mathbf{V}(X) = \frac{q}{p^2}$$

PROPOSITION 10.12. (**La loi géométrique et l'antirépartition**) La loi géométrique et l'antirépartition font très "bon ménage", en effet l'événement  $[X > k]$  est réalisé **si, et seulement si**, les  $k$  premiers essais n'ont donné que des échecs. Alors dans ce cas pour tout  $k \in \mathbf{N}$ ,  $\mathbf{P}([X > k]) = q^k$ .

PROPOSITION 10.13. (**Une caractérisation de la loi géométrique : l'absence de mémoire**)

$$(X \hookrightarrow \mathcal{G}(p)) \iff (\forall (m, n) \in \mathbf{N}^2, \mathbf{P}_{[X > n]}([X > m + n]) = \mathbf{P}([X > m]))$$

REMARQUE 10.10. (1) Pour démontrer qu'une variable  $X$  suit une loi géométrique, il suffit de prouver que  $X(\Omega) = \mathbf{N}^*$  et que pour tout entier naturel  $k$  non nul, le rapport  $\frac{\mathbf{P}([X = k + 1])}{\mathbf{P}([X = k])}$  est égal à une constante réelle comprise entre 0 et 1.

- (2) Nous verrons dans le chapitre concernant les variables absolument continues (VARAD), que la loi exponentielle a la même caractéristique d'absence de mémoire. En effet nous démontrerons à titre d'exercice que la partie entière d'une variable exponentielle nous donnera une variable géométrique. On peut même dire qu'une variable exponentielle est aux variables à densité ce qu'est une variable géométrique pour les variables discrètes.

**1.7. Loi de Poisson ou loi des phénomènes rares****. Modèle probabiliste**

Il n'y en a pas !

Cela veut dire concrètement que si l'on modélise un phénomène aléatoire par cette loi, l'énoncé devra vous le préciser, ce qui sera pour le lecteur un soulagement !

Il faut savoir que l'on utilise cette loi lorsque l'on veut compter le **nombre d'apparitions d'un phénomène rare sur un grand nombre d'observations, durant un intervalle de temps fixé**, comme par exemple le nombre d'accidents dans un atelier, le nombre de pannes d'un appareil, le nombre d'embouteillages téléphoniques, la nombre de crimes etc... Elle s'applique souvent dans les problèmes de files d'attente.

Cette loi peut intervenir aussi en tant que **loi-limite** d'une loi binomiale, sous certaines conditions que nous préciserons ultérieurement, mais disons que  **$n$  devra être grand et  $p$  petit**.

DÉFINITION 10.12. (**Loi de Poisson**) On dit qu'une variable aléatoire suit la **loi de Poisson** de paramètre  $\lambda \in \mathbf{R}_+^*$ , ce que l'on note  $X \hookrightarrow \mathcal{P}(\lambda)$ , lorsque :

$$(1) X(\Omega) = \mathbf{N}$$

$$(2) \forall k \in \mathbf{N}, \mathbf{P}([X = k]) = \frac{e^{-\lambda} \lambda^k}{k!}.$$

PROPOSITION 10.14. (1) (**Espérance et variance**) Si  $X \hookrightarrow \mathcal{P}(\lambda)$  alors :

$$\mathbf{E}(X) = \lambda \quad \text{et} \quad \mathbf{V}(X) = \lambda$$

(2) (**Stabilité de la loi de Poisson pour la somme de variables indépendantes**)

(a) (**Cas de deux variables**) Soit  $X$  et  $Y$  deux variables aléatoires discrètes et **indépendantes** définies sur le même espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  et telles que  $X \hookrightarrow \mathcal{P}(\lambda)$  et  $Y \hookrightarrow \mathcal{P}(\lambda')$  alors :

$$X + Y \hookrightarrow \mathcal{P}(\lambda + \lambda')$$

(b) (**Cas de  $n$  variables**) Soit  $X_1, X_2, \dots, X_n$ ,  $n$  variables aléatoires **indépendantes** définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  telles que pour tout  $k \in \llbracket 1, n \rrbracket$ ,  $X_k \hookrightarrow \mathcal{P}(\lambda_k)$ . Alors :

$$\sum_{k=1}^n X_k \hookrightarrow \mathcal{P}\left(\sum_{k=1}^n \lambda_k\right)$$

DÉFINITION 10.13. (**Processus poissonniens**) Modèle mathématique décrivant et caractérisant la réalisation d'événements aléatoires indépendants se succédant au cours du temps : des arrivées, des naissances, des pannes ou défaillances ou incidents, des désintégrations, des décès ... Parmi tous les processus, celui de Poisson peut se caractériser de diverses façons, dont la principale pour l'utilisateur est d'être "sans vieillissement" (ou sans mémoire).

EXEMPLE 10.2. – **Standart téléphonique** : Le nombre d'appels téléphoniques reçus par un standart au cours d'un intervalle de temps  $[0, T]$  fixé peut être approché par une loi de Poisson.

– **Péage sur une autoroute** : Le nombre de voitures traversant le péage d'une autoroute dans un intervalle de temps  $[t, t + \theta]$ ,  $\theta \in \mathbf{R}_+^*$  et en supposant le processus homogène dans le temps, ne dépend que de la longueur  $\theta$  de l'intervalle choisi et suit approximativement une loi de Poisson.

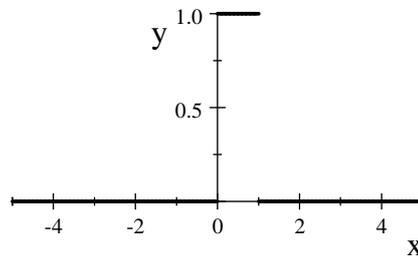
## 2. Loïs à densité

### 2.1. Loi uniforme $\mathcal{U}(I)$ .

DÉFINITION 10.14. (**Loi uniforme**) Soit  $(a, b) \in \mathbf{R}^2$ , tel que  $a < b$ ,  $I = (a, b)$  (ouvert ou fermé en  $a$  et/ou en  $b$ ). On dit que  $X$  une variable à densité définie sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  suit la **loi uniforme sur  $I$** , ce qu'on note  $X \hookrightarrow \mathcal{U}(I)$ , si une densité  $f$  associée à  $X$  est définie par :

$$f(x) = \begin{cases} 0 & \text{si } x \notin I \\ \frac{1}{b-a} & \text{si } x \in I \end{cases}$$

EXEMPLE 10.3. de représentation graphique d'une densité

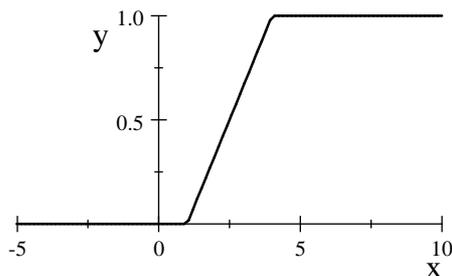


Tracé d'une densité d'une variable suivant la loi uniforme d'extrémités 0 et 1

PROPOSITION 10.15. (**Fonction de répartition d'une variable uniforme**) Nous prendrons l'exemple du cas classique où  $I = [a, b]$ .

$$F_X(x) = \begin{cases} 0 & \text{si } x < a \\ \frac{x-a}{b-a} & \text{si } x \in [a, b] \\ 1 & \text{si } x > b \end{cases}$$

EXEMPLE 10.4. de représentation graphique d'une fonction de répartition



Tracé de la fonction de répartition de  $X \hookrightarrow \mathcal{U}([1, 4])$

PROPOSITION 10.16. (1) (**Transformation affine**) Nous avons l'équivalence :

$$(X \hookrightarrow \mathcal{U}([a, b])) \iff ((-X) \hookrightarrow \mathcal{U}([-b, -a]))$$

(2) (**Espérance et variance**) Si  $X \hookrightarrow \mathcal{U}(I)$  alors :

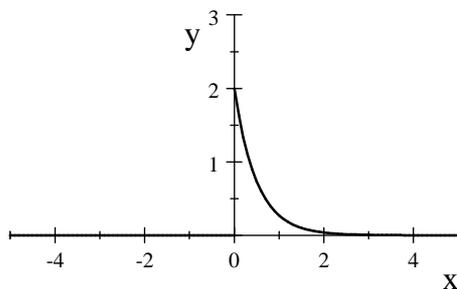
$$\mathbf{E}(X) = \frac{a+b}{2} \quad \text{et} \quad \mathbf{V}(X) = \frac{(b-a)^2}{12}$$

## 2.2. Loi exponentielle $\varepsilon(\lambda)$ .

DÉFINITION 10.15. (**Loi exponentielle**) Soit  $\lambda > 0$ . On dit que  $X$  une variable aléatoire à densité définie sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  suit la **loi exponentielle de paramètre  $\lambda$** , ce qu'on note  $X \hookrightarrow \varepsilon(\lambda)$ , si une densité  $f$  associée à  $X$  est définie par :

$$f(x) = \begin{cases} 0 & \text{si } x < 0 \\ \lambda e^{-\lambda x} & \text{si } x \in \mathbf{R}_+ \end{cases}$$

EXEMPLE 10.5. de représentation graphique d'une densité

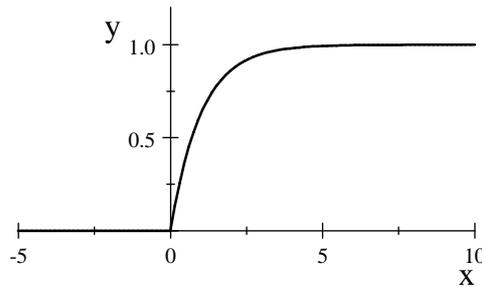


Tracé d'une densité de  $X \hookrightarrow \varepsilon(2)$

PROPOSITION 10.17. (*Fonction de répartition d'une variable exponentielle*) Si  $X \hookrightarrow \varepsilon(\lambda)$  alors :

$$F_X(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1 - e^{-\lambda x} & \text{si } x \in \mathbb{R}_+ \end{cases}$$

EXEMPLE 10.6. de représentation graphique d'une fonction de répartition



Tracé de la fonction de répartition de  $X \hookrightarrow \varepsilon(1)$

PROPOSITION 10.18. (*Transformation affine*)

$$\forall a > 0, (X \hookrightarrow \varepsilon(\lambda)) \iff \left( aX \hookrightarrow \varepsilon\left(\frac{\lambda}{a}\right) \right)$$

PROPOSITION 10.19. (*Espérance et variance*) Si  $X \hookrightarrow \varepsilon(\lambda)$  alors :

$$\mathbf{E}(X) = \frac{1}{\lambda} \quad \text{et} \quad \mathbf{V}(X) = \frac{1}{\lambda^2}$$

REMARQUE 10.11. La fonction  $r : x \mapsto r(x) = 1 - F_X(x) = \mathbf{P}([X > x])$  est appelée **fonction de survie**. Dans une application typique de médecine, d'électronique... ,  $r(x)$  représente la probabilité d'attendre plus de  $x$  unités de temps avant l'apparition d'un certain phénomène,  $1/\lambda$  est le temps moyen d'attente de ce phénomène aléatoire.

PROPOSITION 10.20. (*Caractérisation d'une loi exponentielle : la propriété d'absence de mémoire*)

$$(X \hookrightarrow \varepsilon(\lambda)) \iff (\forall (x, y) \in \mathbb{R}_+^2, \mathbf{P}_{[X > x]}([X > x + y]) = \mathbf{P}([X > y]))$$

- **Interprétation** : On met en marche un appareil et on note  $X$  sa durée de fonctionnement avant la première panne. Le résultat précédent s'exprime en disant qu'il n'y a pas de phénomène d'usure : la probabilité qu'ayant déjà fonctionné pendant une durée  $x$ , l'appareil fonctionne encore pendant une durée de vie  $y$ , est égale à la probabilité que l'appareil fonctionne pendant une durée  $y$  à partir de sa mise en marche. Cette propriété est donc une propriété de **non-vieillesse ou d'absence de mémoire** ("l'appareil ne se souvient pas d'avoir vieilli").
- **Lien entre la loi exponentielle et la loi de Poisson** : La loi exponentielle joue un rôle important dans la théorie des files d'attente<sup>2</sup>. On montre facilement que si le nombre d'apparition  $Y$  d'un événement sur un intervalle de temps  $x$  suit une loi de Poisson de paramètre  $\lambda x$  alors le temps  $X$  séparant deux événements consécutifs suit la loi exponentielle de paramètre  $\lambda$ . En effet nous avons clairement  $[Y = 0] = [X > x]$  ce qui donne :  $\mathbf{P}([Y = 0]) = \mathbf{P}([X > x])$ , c'est-à-dire que  $e^{-\lambda x} = 1 - F_X(x)$  et

<sup>2</sup>Théorie permettant de prendre en compte et de modéliser les goulots d'étranglement dans les processus.

donc  $F_X(x) = 1 - e^{-\lambda x}$  qui est bien la fonction de répartition de la loi exponentielle de paramètre  $\lambda$ . La loi exponentielle est donc naturellement associée à la loi de Poisson dans de tels cas. Elle intervient couramment dans des problèmes de fiabilité (temps séparant deux pannes consécutives, durée de vie de composants...).

### 2.3. Loïs de Laplace-Gauss ou loïs normales.

DÉFINITION 10.16. (*Intégrale de Gauss*) On appelle *intégrale de Gauss* :

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 1$$

REMARQUE 10.12. (1) La définition aurait pu s'appeler proposition, du fait que ce résultat, admis en EC, se démontre parfaitement.

(2) Il est utile de connaître une variante de cette intégrale (obtenue par changement de variable) :

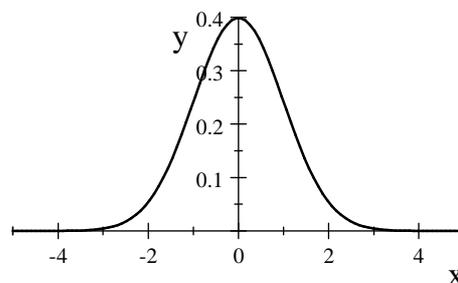
$$\int_{-\infty}^{+\infty} e^{-x^2} dx = \sqrt{\pi}$$

#### 2.3.1. La loi normale centrée et réduite $\mathcal{N}(0, 1)$ .

DÉFINITION 10.17. (*Loi normale centrée et réduite*) On dit que  $X$  une variable aléatoire à densité définie sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  suit la *loi normale centrée réduite*, ce qu'on note  $X \hookrightarrow \mathcal{N}(0, 1)$ , si une densité  $f$  associée à  $X$  est définie par :

$$\forall x \in \mathbb{R}, \quad f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

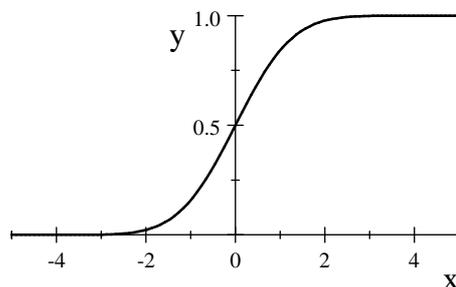
REMARQUE 10.13. Du fait de la parité de  $f$  on parle de *densité de distribution bilatérale*.



Tracé d'une densité d'une variable normale centrée réduite

DÉFINITION 10.18. (*Fonction de répartition d'une variable normale centrée et réduite*) Nous la noterons  $\Phi$ . Elle est définie par :

$$\forall x \in \mathbb{R}, \quad \Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$



Tracé de la fonction de répartition d'une variable centrée réduite

REMARQUE 10.14. L'intégrale  $\int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$  sera tabulée du fait que l'on ne peut expliciter une primitive de  $x \mapsto e^{-x^2/2}$  à l'aide de fonctions usuelles. On parle de la **table de la fonction de répartition de la loi normale centrée et réduite**.

PROPRIÉTÉS 10.2. (**Centre de symétrie**)

$$\forall x \in \mathbb{R}, \quad \Phi(-x) = 1 - \Phi(x)$$

REMARQUE 10.15. Cette relation est **valable** pour toute fonction de répartition de variable à densité à densité **paire** (on parle de **distribution bilatérale**).

PROPOSITION 10.21. (**Espérance et variance**) Si  $X \hookrightarrow \mathcal{N}(0, 1)$  alors :

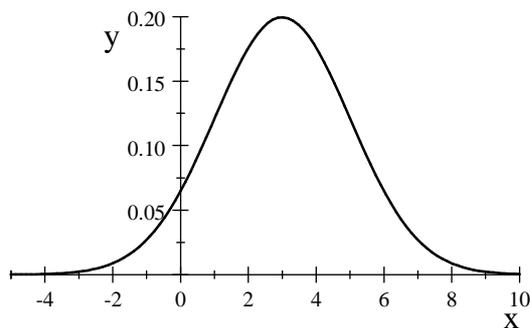
$$\mathbf{E}(X) = 0 \quad \text{et} \quad \mathbf{V}(X) = 1$$

2.3.2. **Loi normale de paramètres quelconques**  $\mathcal{N}(m, \sigma^2)$  où  $m \in \mathbb{R}$  et  $\sigma \in \mathbb{R}_+^*$ .

DÉFINITION 10.19. (**Loi normale quelconque**) On dit que  $X$  une variable aléatoire à densité définie sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  suit la **loi normale de paramètres**  $m \in \mathbb{R}$  et  $\sigma^2 \in \mathbb{R}_+^*$ , ce qu'on note  $X \hookrightarrow \mathcal{N}(m, \sigma^2)$ , si une densité  $f$  associée à  $X$  est définie par :

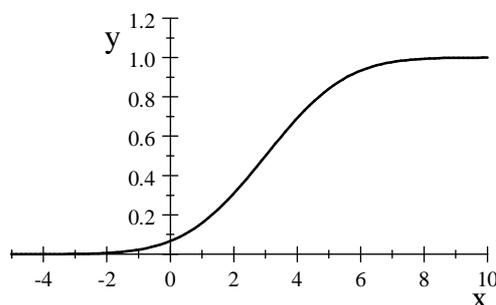
$$\forall x \in \mathbb{R}, \quad f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}$$

EXEMPLE 10.7. de représentation graphique d'une densité



Tracé d'une densité de  $X \hookrightarrow \mathcal{N}(3, 2)$

EXEMPLE 10.8. *de représentation graphique d'une fonction de répartition*



Tracer de la fonction de répartition de  $X \hookrightarrow \mathcal{N}(3, 2)$

On peut passer très facilement d'une variable normale de paramètres quelconques à une variable normale centrée réduite, en utilisant le théorème fondamental suivant :

PROPOSITION 10.22. (**Le théorème fondamental du changement de variable affine**) Nous avons l'équivalence :

$$(X \hookrightarrow \mathcal{N}(m, \sigma^2)) \iff \left( Y = \frac{X - m}{\sigma} \hookrightarrow \mathcal{N}(0, 1) \right)$$

PROPOSITION 10.23. (**Espérance et variance**) Si  $X \hookrightarrow \mathcal{N}(m, \sigma^2)$  alors :

$$\mathbf{E}(X) = m \quad \text{et} \quad \mathbf{V}(X) = \sigma^2$$

PROPOSITION 10.24. (**Transformation affine**) Soit  $(a, b) \in \mathbb{R}^* \times \mathbb{R}$ , nous avons l'équivalence :

$$(X \hookrightarrow \mathcal{N}(m, \sigma^2)) \iff (aX + b \hookrightarrow \mathcal{N}(am + b, a^2\sigma^2))$$

PROPOSITION 10.25. (**Stabilité de la loi normale pour la somme de variables indépendantes**)

(1) (**Cas de deux variables**) Soit  $X \hookrightarrow \mathcal{N}(m_1, \sigma_1^2)$  et  $Y \hookrightarrow \mathcal{N}(m_2, \sigma_2^2)$  définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  avec  $X, Y$  **indépendantes** alors :

$$X + Y \hookrightarrow \mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$$

(2) (**Cas de n variables**) Soit  $X_1, X_2, \dots, X_n$ ,  $n$  variables aléatoires **indépendantes** définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  telles que pour tout  $k \in \llbracket 1, n \rrbracket$ ,  $X_k \hookrightarrow \mathcal{N}(m_k, \sigma_k^2)$ . Alors on montre par récurrence que :

$$\sum_{k=1}^n X_k \hookrightarrow \mathcal{N}\left(\sum_{k=1}^n m_k, \sum_{k=1}^n \sigma_k^2\right)$$

## 2.4. Lois gamma.

### 2.4.1. La fonction gamma d'Euler.

DÉFINITION 10.20. (**Fonction gamma**) On appelle **fonction gamma**, la fonction notée  $\Gamma$  définie sur  $\mathbb{R}_+^*$  à valeurs dans  $\mathbb{R}_+^*$  par :

$$\Gamma(x) = \int_0^{+\infty} e^{-t} t^{x-1} dt$$

PROPOSITION 10.26. (**Propriétés de la fonction gamma**)

(1) La fonction  $\Gamma$  vérifie la relation fonctionnelle :  $\forall x > 0, \Gamma(x + 1) = x\Gamma(x)$

$$(2) \forall n \in \mathbf{N}^*, \Gamma(n) = (n-1)!$$

$$(3) \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

$$(4) \forall n \in \mathbf{N}, \Gamma\left(n + \frac{1}{2}\right) = \frac{(2n)!\sqrt{\pi}}{4^n n!} \text{ (à savoir retrouver)}$$

#### 2.4.2. La loi petit gamma.

DÉFINITION 10.21. (**Loi petit gamma**) Soit  $t$  un réel strictement positif. On dit que  $X$  une variable aléatoire à densité définie sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  suit la **loi petit gamma de paramètre  $t$** , ce qu'on note  $X \hookrightarrow \gamma(t)$ , si une densité  $f_X$  associée à  $X$  est définie par :

$$f_X(x) = \begin{cases} \frac{e^{-x} x^{t-1}}{\Gamma(t)} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}$$

REMARQUE 10.16. En général on ne sait pas expliciter la fonction de répartition sans intégrale. En revanche c'est possible quand  $t$  est un entier naturel, surtout **petit**, où l'on procède par **intégration par parties**.

PROPOSITION 10.27. (**Espérance et variance**) Si  $X \hookrightarrow \gamma(t)$  alors :

$$\mathbf{E}(X) = t \quad \text{et} \quad \mathbf{V}(X) = t$$

PROPOSITION 10.28. (**Stabilité de la loi petit gamma pour la somme de variables indépendantes**)

(1) Soit  $X$  et  $Y$  deux variables à densité **indépendantes** définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  alors :

$$(X \hookrightarrow \gamma(t_1) \quad \text{et} \quad Y \hookrightarrow \gamma(t_2)) \implies (X + Y \hookrightarrow \gamma(t_1 + t_2))$$

(2) (**Cas de  $n$  variables**) Soit  $X_1, X_2, \dots, X_n$ ,  $n$  variables aléatoires **indépendantes** définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  telles que pour tout  $k \in \llbracket 1, n \rrbracket$ ,  $X_k \hookrightarrow \gamma(t_k)$ . Alors on montre par récurrence que :

$$\sum_{k=1}^n X_k \hookrightarrow \gamma\left(\sum_{k=1}^n t_k\right)$$

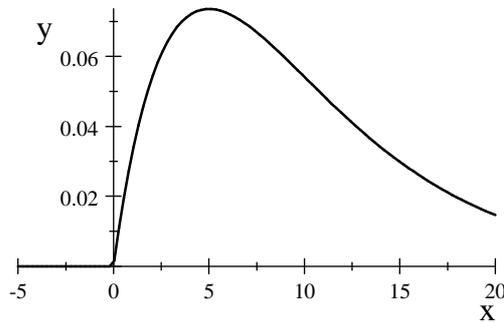
#### 2.4.3. La loi grand gamma.

DÉFINITION 10.22. (**Loi grand gamma**) Soit  $b$  et  $t$  deux réels strictement positifs. On dit que  $X$  une variable aléatoire à densité définie sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  suit la **loi grand gamma de paramètres  $b$  et  $t$** , ce qu'on note  $X \hookrightarrow \Gamma(b, t)$ , si une densité  $f$  associée à  $X$  est définie par :

$$f_X(x) = \begin{cases} \frac{e^{-\frac{x}{b}} x^{t-1}}{\Gamma(t) b^t} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}$$

La distribution gamma est une **généralisation de la loi exponentielle**. En effet si la loi exponentielle correspond à la distribution de probabilité du temps séparant l'apparition de deux événements donnés, la loi gamma fournit la distribution du temps qui s'écoule entre la  $k^{\text{ème}}$  et la  $(k+r)^{\text{ème}}$  apparition de l'événement. La loi gamma est appliquée comme modèle de probabilité pour prévoir la durée de vie des appareils qui subissent une usure.

EXEMPLE 10.9. *de représentation graphique d'une densité*



Tracé d'une densité de  $X \hookrightarrow \Gamma(2, 5)$

DÉFINITION 10.23. *La loi  $\Gamma(1, t)$  est dite **loi gamma à un paramètre**  $t$ .*

PROPOSITION 10.29.

(1) *(La loi gamma et la loi exponentielle)*

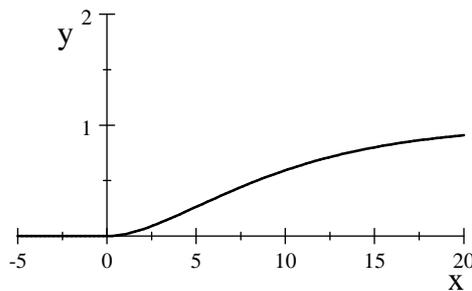
$$\forall b \in \mathbb{R}_+^*, \quad \Gamma(b, 1) = \varepsilon\left(\frac{1}{b}\right)$$

(2) *(Lien entre les lois  $\gamma$  et  $\Gamma$ )* Soit  $b$  et  $t$  deux réels strictement positifs alors on a l'équivalence :

$$(X \hookrightarrow \gamma(t)) \iff (Y = bX \hookrightarrow \Gamma(b, t))$$

REMARQUE 10.17. *En général on ne sait pas expliciter la fonction de répartition sans intégrale. En revanche c'est possible quand  $t$  est un entier naturel, surtout **petit**, où l'on procède par **intégration par parties**.*

EXEMPLE 10.10. *de représentation graphique d'une fonction de répartition*



Tracé de la fonction de répartition de  $X \hookrightarrow \Gamma(2, 5)$

PROPOSITION 10.30. *(Espérance et variance)* Si  $X \hookrightarrow \Gamma(b, t)$  alors :

$$\mathbf{E}(X) = bt \quad \text{et} \quad \mathbf{V}(X) = b^2t$$

PROPOSITION 10.31. *(Stabilité de la loi gamma pour la somme de variables indépendantes)*

- (1) (**Cas de deux variables**) Soit  $X, Y$  deux variables à densité **indépendantes** définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  alors :

$$(X \hookrightarrow \Gamma(b, t_1) \text{ et } Y \hookrightarrow \Gamma(b, t_2)) \implies (X + Y \hookrightarrow \Gamma(b, t_1 + t_2))$$

- (2) (**Cas de  $n$  variables aléatoires**) On peut généraliser le résultat précédent par récurrence sur  $n \in \mathbf{N}^*$ . Soit  $X_1, \dots, X_n$ ,  $n$  variables aléatoires à densité **indépendantes** définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  telles que pour tout  $k \in \llbracket 1, n \rrbracket$ ,  $X_k \hookrightarrow \Gamma(b, t_k)$ , alors on montre par récurrence que :

$$\sum_{k=1}^n X_k \hookrightarrow \Gamma\left(b, \sum_{k=1}^n t_k\right)$$

- (3) (**Cas particulier important : somme de variables exponentielles indépendantes, loi d'Erlang**) Soit  $X_1, \dots, X_n$ ,  $n$  variables aléatoires à densité **indépendantes** définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  telles que pour tout  $k \in \llbracket 1, n \rrbracket$ ,  $X_k \hookrightarrow \varepsilon(\lambda)$ , on a selon le résultat précédent de cette proposition :

$$\sum_{k=1}^n X_k \hookrightarrow \Gamma\left(\frac{1}{\lambda}, n\right)$$

C'est la **loi d'Erlang**.

REMARQUE 10.18. Ne parlez surtout pas de stabilité de la loi exponentielle, c'est rédhibitoire !

## Convergences et approximations

L'objet de ce chapitre consiste à étudier le comportement asymptotique (que nous précisons ci-dessous en restant dans le cadre du programme) d'une suite de variables aléatoires.

### 1. Deux inégalités de concentration

PROPOSITION 11.1. (*Inégalité de Markov*) Soit  $X$  une variable aléatoire définie sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  **positive** et **admettant une espérance**, alors on a :

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

REMARQUE 11.1. L'inégalité traduit formellement l'idée que "moins une variable est dispersée, plus ses valeurs sont centrées autour de la moyenne et inversement". Elle est qualifiée de **grossière** car elle est applicable **quelle que soit la nature et la loi de  $X$** , et il n'est pas rare de voir, en l'appliquant, une probabilité majorée par un réel supérieur à 1 ! Elle sert principalement à démontrer qu'une suite de variable aléatoire **converge en probabilité** vers une autre variable.

COROLLAIRE 11.1. (1) On a pour  $X$  une variable telle que  $|X|$  **admette une espérance**, alors on a :

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

(2) Soit  $X$  une variable admettant un moment d'ordre deux alors on a :

$$\forall a > 0, \quad \mathbf{P}(|X| \geq a) \leq \frac{\mathbf{E}(X^2)}{a^2}$$

(3) (*Inégalité de Bienaymé-Tchebychev*) Soit  $X$  une variable **admettant un moment d'ordre deux**, alors on a :

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

REMARQUE 11.2. Si la variance de  $X$  est nulle on sait que la variable est quasi-certaine égale à son espérance. Donc pour tout  $\varepsilon > 0$ ,  $\mathbf{P}(|X - \mathbf{E}(X)| \geq \varepsilon) = 0$  entraîne que  $\mathbf{P}(|X - \mathbf{E}(X)| \geq \varepsilon) = 0$  et l'inégalité de Bienaymé-Tchebychev est vérifiée.

### 2. Convergence en probabilité

DÉFINITION 11.1. (*Converge en probabilité*) Soit  $(X_n)_{n \in \mathbf{N}}$  une suite de variables aléatoires définies sur un même espace probabilisé et d'autre part une variable aléatoire  $X$  définie sur ce même espace. On dit que  $(X_n)_{n \in \mathbf{N}}$  **converge en probabilité** vers  $X$  si<sup>1</sup> :

$$\forall \varepsilon > 0, \quad \lim_{n \rightarrow +\infty} \mathbf{P}(|X_n - X| \geq \varepsilon) = 0$$

---

<sup>1</sup>La probabilité que  $X_n$  s'éloigne de  $X$  d'au moins  $\varepsilon$  tend vers 0 quand  $n$  est grand.

ce qui revient à :

$$\forall \varepsilon > 0, \quad \lim_{n \rightarrow +\infty} \mathbf{P} (|X_n - X| < \varepsilon) = 1$$

NOTATION 11.1. On note  $(X_n)_{n \in \mathbf{N}} \xrightarrow{\mathbf{P}} X$ .

REMARQUE 11.3. (1)  $\left( (X_n)_{n \in \mathbf{N}} \xrightarrow{\mathbf{P}} X \right) \iff \left( (X_n - X)_{n \in \mathbf{N}} \xrightarrow{\mathbf{P}} 0 \right)$ .

Néanmoins on n'a pas forcément les résultats  $\lim_{n \rightarrow +\infty} \mathbf{E}(X_n) = \mathbf{E}(X)$  et  $\lim_{n \rightarrow +\infty} \mathbf{V}(X_n - X) = 0$ .

(2) Il faut bien que vous ayez en tête que cette définition formalise une notion de **probabilité de rapprochement entre les valeurs de deux variables aléatoires lorsque le nombre d'expériences est très grand**. C'est-à-dire qu'elle traduit le fait que, lorsque le nombre d'expériences est extrêmement grand, il y a une forte probabilité pour que  $X_n$  et  $X$  fournissent la **même valeur** pour une même issue  $\omega$  de l'expérience.

PROPOSITION 11.2. Soit  $(X_n)_{n \in \mathbf{N}}$  une suite de variables aléatoires définies sur un même espace probabilisé, alors une **condition suffisante** pour que la suite converge en loi vers la constante réelle  $a$  (on notera  $(X_n)_{n \in \mathbf{N}} \xrightarrow{\mathbf{P}} a$ ) est  $\lim_{n \rightarrow +\infty} \mathbf{E}(X_n) = a$  et  $\lim_{n \rightarrow +\infty} \mathbf{V}(X_n) = 0$ .

PROPOSITION 11.3. (**Théorème de Khintchine**) Soit  $(X_k)_{k \in \mathbf{N}^*}$  une suite de variables aléatoires définies sur un même espace probabilisé **i.i.d.** ayant une espérance commune  $m$  et une variance commune  $\sigma^2$ . Posons  $Z_n = \frac{X_1 + \dots + X_n}{n}$  alors :

$$\forall \varepsilon > 0, \quad \lim_{n \rightarrow +\infty} \mathbf{P} (|Z_n - m| \geq \varepsilon) = 0$$

Ce qui s'interprète en disant que la suite  $(Z_n)_{n \in \mathbf{N}^*}$  **converge en probabilité** vers la variable **quasi-certaine** égale à  $m$ . C'est la **loi faible des grands nombres**.

REMARQUE 11.4. La loi faible est à opposer à la loi forte qui est hors programme, où il s'agit de montrer que  $\mathbf{P} \left( \left\{ \omega \in \Omega \mid \lim_{n \rightarrow +\infty} Z_n(\omega) = \mathbf{E}(X_1) \right\} \right) = 1$ . Le qualificatif de "forte" provient du fait que l'on a pour tout  $\omega \in \Omega$ ,  $\lim_{n \rightarrow +\infty} Z_n(\omega) = \mathbf{E}(X_1)$  p.s. alors que dans la loi faible on n'exige pas cette contrainte pour  $Z_n$ , puisque c'est la probabilité de l'écart absolu entre  $Z_n$  et  $\mathbf{E}(X_1)$  dépassant un seuil strictement positif qui doit tendre vers 0, ce n'est pas la même chose !

PROPOSITION 11.4. (**Théorème d'or de Bernoulli**) C'est un **cas particulier** de la loi faible des grands nombres lorsque les **épreuves sont bernoulliennes** de même paramètre  $p$ . Ainsi  $Z_n$  représente la **fréquence des succès**. Et plus précisément nous avons :

$$\forall \varepsilon > 0, \quad \mathbf{P} (|Z_n - p| \geq \varepsilon) \leq \frac{p(1-p)}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2}$$

REMARQUE 11.5. (1)  Retenez que  $p(1-p) \leq 1/4$ . Cela vient des variations de  $x \mapsto x(1-x)$  sur le segment  $[0, 1]$ .

(2) Il faut bien interpréter ce théorème en se disant qu'il délivre une estimation de la probabilité que de grands écarts se produisent entre la fréquence observée et la probabilité "vraie" d'un succès. A mesure que le nombre d'essais augmente, la probabilité que la fréquence  $Z_n$  s'éloigne de la "vraie" valeur de  $p$  est majorée par une quantité qui est de plus en plus petite.

(3) Ce théorème justifie l'approche fréquentiste des probabilités : la probabilité d'un événement étant la limite (en probabilité) de la fréquence empirique de son nombre d'observations dans une suite d'expériences indépendantes et reproductibles.

### 3. Convergence en loi et approximations

DÉFINITION 11.2. (**Converge en loi**) On dit que la suite de variables aléatoires  $(X_n)_{n \in \mathbf{N}}$  de fonctions de répartition associées  $(F_{X_n})_{n \in \mathbf{N}}$  **converge en loi** vers une variable  $X$  de fonction de répartition associée  $F_X$ , si en tout point  $x$  où  $F_X$  est continue :

$$\lim_{n \rightarrow +\infty} F_{X_n}(x) = F_X(x)$$

On note  $(X_n)_{n \in \mathbf{N}} \xrightarrow{\mathcal{L}} X$ .

REMARQUE 11.6.  Contrairement à la convergence en probabilité :

$$\left( (X_n)_{n \in \mathbf{N}} \xrightarrow{\mathcal{L}} X \right) \not\iff \left( (X_n - X)_{n \in \mathbf{N}} \xrightarrow{\mathcal{L}} 0 \right)$$

PROPOSITION 11.5. Soit  $(X_n)_{n \in \mathbf{N}}$  une suite de variables aléatoires définies sur un même espace probabilisé **convergente en loi** vers une variable  $X$ . Soit  $a$  et  $b$  deux points de continuité de  $F_X$  tels que  $a < b$ , alors :

$$\lim_{n \rightarrow +\infty} \mathbf{P}([a < X_n \leq b]) = \mathbf{P}([a < X \leq b])$$

PROPOSITION 11.6. (**Uniquement valable en discret**) Soit  $(X_n)_{n \in \mathbf{N}}$  une suite de variables discrètes et  $X$  une variables aléatoires discrète définies sur le même espace probabilisé<sup>2</sup>, telle que pour tout  $n \in \mathbf{N}$ ,  $X_n(\Omega) \subset \mathbf{N}$  et  $X(\Omega) \subset \mathbf{N}$  alors nous avons l'**équivalence** suivante :

$$\left( (X_n)_{n \in \mathbf{N}} \xrightarrow{\mathcal{L}} X \right) \iff \left( \forall k \in \mathbf{N}, \lim_{n \rightarrow +\infty} \mathbf{P}([X_n = k]) = \mathbf{P}([X = k]) \right)$$

REMARQUE 11.7. (1) Cette proposition est **uniquement valable en discret**, c'est-à-dire que la suite  $(X_n)_{n \in \mathbf{N}}$  et  $X$  doivent être discrètes. En cas de doute, revenir à la définition.

(2) Cette proposition est valable si les variables ne prennent pas forcément des valeurs entières.

PROPOSITION 11.7. (**Approximation de la loi hypergéométrique par la loi binomiale**) Soit  $(X_k)_{k \in \mathbf{N}}$  une suite de variables aléatoires définies sur un même espace probabilisé telle que pour tout  $k \in \mathbf{N}$ ,  $X_k \hookrightarrow \mathcal{H}(N_k, n, p)$  où  $n \in \mathbf{N}$  et  $p \in [0, 1]$  sont fixés. Supposons que pour tout  $k \in \mathbf{N}$ ,  $N_k p$  et  $N_k(1-p)$  soit un entier naturel et  $\lim_{k \rightarrow +\infty} N_k = +\infty$ . Alors  $(X_k)_{k \in \mathbf{N}} \xrightarrow{\mathcal{L}} X$  où  $X \hookrightarrow \mathcal{B}(n, p)$ .

REMARQUE 11.8. Dans la pratique on pourra approximer la loi  $\mathcal{H}(N_k, n, p)$  par la loi  $\mathcal{B}(n, p)$  dès que  $N \geq 10n$ .

PROPOSITION 11.8. (**Approximation de la loi binomiale par la loi de Poisson**) Soit  $(X_n)_{n \in \mathbf{N}}$  une suite de variables aléatoires définies sur un même espace probabilisé telle que pour tout  $n \in \mathbf{N}$ ,  $X_n \hookrightarrow \mathcal{B}(n, p_n)$  avec  $\lim_{n \rightarrow +\infty} np_n = \lambda$ ,  $\lambda > 0$ . Alors  $(X_n)_{n \in \mathbf{N}} \xrightarrow{\mathcal{L}} X$  où  $X \hookrightarrow \mathcal{P}(\lambda)$ .

REMARQUE 11.9. Dans la pratique on pourra approximer la loi  $\mathcal{B}(n, p)$  par la loi  $\mathcal{P}(np)$  dès que  $n \geq 30$ ,  $p \leq 0,1$ .

PROPOSITION 11.9. (**Le théorème de la limite centrée (TCL) ou "central limit" version somme – admis**) Soit  $(X_n)_{n \in \mathbf{N}^*}$  une suite de variables aléatoires définies sur un même espace probabilisé. On suppose que les variables sont indépendantes de même loi (pas forcément connue) admettant une espérance  $m$  et une variance  $\sigma^2$  non nulle.

<sup>2</sup>C'est ce que dit le programme officiel, mais sachez que la convergence en loi ne fait référence qu'à la loi des variables (contrairement à la convergence en probabilité). On peut donc très bien définir la convergence en loi d'une suite de variables qui ne seraient pas définies sur le même espace.

Posons pour tout  $n \in \mathbf{N}^*$ ,  $S_n = \sum_{k=1}^n X_k$ ,  $S_n^* = \frac{S_n - \mathbf{E}(S_n)}{\sigma(S_n)}$ , alors  $(S_n^*)_{n \in \mathbf{N}^*} \xrightarrow{\mathcal{L}} N$  où  $N \hookrightarrow \mathcal{N}(0, 1)$ .

Autrement dit :

$$\forall (a, b) \in \mathbf{R}^2, \quad \lim_{n \rightarrow +\infty} \mathbf{P} \left( \left[ a \leq \frac{S_n - \mathbf{E}(S_n)}{\sigma(S_n)} \leq b \right] \right) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

REMARQUE 11.10. (1) On pourrait aussi écrire que :

$$\left( \frac{S_n - nm}{\sqrt{n\sigma^2}} \right)_{n \in \mathbf{N}^*} \xrightarrow{\mathcal{L}} N \quad \text{où } N \hookrightarrow \mathcal{N}(0, 1)$$

(2) Dans le cas particulier où  $(X_n)_{n \in \mathbf{N}^*}$  est une suite de variables aléatoires **i.i.d.** de même **loi normale**  $\mathcal{N}(m, \sigma^2)$  alors  $\frac{S_n - nm}{\sqrt{n\sigma^2}} \hookrightarrow \mathcal{N}(0, 1)$  qui est la loi exacte bien sûr, on ne parle pas de convergence mais d'identité en loi dans le théorème de la limite centrée.

(3) Il existe une **“version moyenne” du théorème de la limite centrée** en posant pour tout

$$n \in \mathbf{N}^*, \quad \overline{S}_n = \frac{1}{n} \sum_{k=1}^n X_k, \quad \text{alors } \left( \overline{S}_n^* \right)_{n \in \mathbf{N}^*} \xrightarrow{\mathcal{L}} N \quad \text{où } N \hookrightarrow \mathcal{N}(0, 1) \quad \text{où } \overline{S}_n^* = \frac{\overline{S}_n - \mathbf{E}(\overline{S}_n)}{\sigma(\overline{S}_n)}.$$

PROPOSITION 11.10. (**Théorème de Moivre-Laplace : approximation d'une loi binomiale par une loi normale – admis**) Soit  $(X_n)_{n \in \mathbf{N}^*}$  une suite de variables aléatoires **indépendantes** définies sur un même espace probabilisé. On suppose que pour tout entier naturel  $n$  non nul  $X_n \hookrightarrow \mathcal{B}(p)$ , ainsi nous savons que la somme  $S_n = \sum_{k=1}^n X_k \hookrightarrow \mathcal{B}(n, p)$ . Alors  $(S_n^*)_{n \in \mathbf{N}^*} \xrightarrow{\mathcal{L}} N$  où  $N \hookrightarrow \mathcal{N}(0, 1)$  avec pour tout

$n \in \mathbf{N}^*$ ,  $S_n^* = \frac{S_n - np}{\sqrt{np(1-p)}}$ . Autrement dit :

$$\forall (a, b) \in \mathbf{R}^2, \quad \lim_{n \rightarrow +\infty} \mathbf{P} \left( \left[ a \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq b \right] \right) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

REMARQUE 11.11. Dans la pratique on pourra approximer la loi  $\mathcal{B}(n, p)$  par la loi  $\mathcal{N}(np, \sqrt{np(1-p)})$  dès que  $n \geq 30$ ,  $np \geq 5$  et  $n(1-p) \geq 5$  ou bien dès que  $n \geq 200$  et  $p \simeq 1/2$ .

PROPOSITION 11.11. (**Approximation d'une loi de Poisson par une loi normale**) Soit  $(X_n)_{n \in \mathbf{N}^*}$  une suite de variables aléatoires définies sur un même espace probabilisé **i.i.d.** de même **loi de Poisson** de paramètre  $\lambda > 0$ . On pose pour  $n \in \mathbf{N}^*$ ,  $S_n = \sum_{k=1}^n X_k$ . Donc  $S_n \hookrightarrow \mathcal{P}(n\lambda)$  et  $(S_n^*)_{n \in \mathbf{N}^*} \xrightarrow{\mathcal{L}} N$  où

$N \hookrightarrow \mathcal{N}(0, 1)$  avec pour tout  $n \in \mathbf{N}^*$ ,  $S_n^* = \frac{S_n - n\lambda}{\sqrt{n\lambda}}$ . Autrement dit :

$$\forall (a, b) \in \mathbf{R}^2, \quad \lim_{n \rightarrow +\infty} \mathbf{P} \left( \left[ a \leq \frac{S_n - n\lambda}{\sqrt{n\lambda}} \leq b \right] \right) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

Dans la pratique on pourra approximer la loi  $\mathcal{P}(\lambda)$  par la loi  $\mathcal{N}(\lambda, \sqrt{\lambda})$  dès que  $\lambda \geq 15$ .

PROPOSITION 11.12. (**Les lois qui peuvent être ou non approximées par une loi normale**)

Loi	Approximation	Explication
$\mathcal{U}(\llbracket 1, n \rrbracket)$	non	Une variable uniforme n'est la somme de rien du tout!
$\mathcal{B}(p)$	non	Une variable bernoullienne n'est la somme de rien du tout!
$\mathcal{B}(n, p)$	<b>oui</b>	Une variable binomiale est la somme de variable de Bernoulli i.i.d. ou bien la somme de variables binomiales par stabilité

$\mathcal{H}(N, n, p)$	<i>non</i>	<i>Une variable hypergéométrique n'est la somme de rien du tout!</i>
$\mathcal{P}(\lambda)$	<b><i>oui</i></b>	<i>Une variable de Poisson est la somme de variables de Poisson par stabilité</i>
$\mathcal{P}(r, p)$	<b><i>oui</i></b>	<i>Une variable de Pascal est la somme de variables géométriques</i>
$\mathcal{B}_-(r, p)$	<b><i>oui</i></b>	<i>Une variable binomiale négative est la somme de variables géométriques sur <math>\mathbf{N}</math></i>
$\mathcal{U}([a, b])$	<i>non</i>	<i>Une variable uniforme n'est la somme de rien du tout!</i>
$\varepsilon(\lambda)$	<i>non</i>	<i>Une variable exponentielle n'est la somme de rien du tout!</i>
$\gamma(t)$	<b><i>oui</i></b>	<i>Une variable petit gamma est la somme de variables petit gamma par stabilité</i>
$\Gamma(b, t)$	<b><i>oui</i></b>	<i>Une variable grand gamma est la somme de variables grand gamma par stabilité</i>



## Estimations

DÉFINITION 12.1. (**Echantillon**) Soit  $P$  une loi de probabilité sur  $\mathbb{R}$ . On appelle **échantillon** de loi  $P$  un  $n$ -uplet  $(X_1, X_2, \dots, X_n)$  de variables aléatoires **indépendantes et de même loi**  $P$ .

DÉFINITION 12.2. (**Estimateur**) On appelle **estimateur du paramètre  $g(\theta)$**  (ou **statistique**), toute **fonction** de l'échantillon  $(X_1, X_2, \dots, X_n)$  qui soit indépendante de  $g(\theta)$ .

DÉFINITION 12.3. (**Estimation de  $g(\theta)$** ) C'est la valeur de  $\varphi_n(x_1, \dots, x_n)$  réalisation de  $\varphi_n(X_1, \dots, X_n)$  à partir d'un échantillon  $(X_1, \dots, X_n)$ .

DÉFINITION 12.4. (**Estimateur convergent de  $g(\theta)$** ) On dit l'**estimateur**  $(T_n)_{n \in \mathbf{N}^*}$  est **convergent** (ou **consistant**) si pour tout  $\theta \in \Theta$ , la suite  $(T_n)_{n \in \mathbf{N}^*}$  **converge en probabilité** vers  $g(\theta)$ . Autrement dit :

$$\forall \varepsilon > 0, \quad \lim_{n \rightarrow +\infty} \mathbf{P}(|T_n - g(\theta)| > \varepsilon) = 0$$

Outils de démonstration :

L'inégalité de Markov si  $g(\theta) \neq \mathbf{E}_\theta(T_n)$  en supposant de plus que  $\lim_{n \rightarrow +\infty} \mathbf{E}_\theta\left((T_n - g(\theta))^2\right) = 0$ , et l'inégalité de Bienaymé-Tchebychev si  $g(\theta) = \mathbf{E}_\theta(T_n)$  en supposant de plus que  $\lim_{n \rightarrow +\infty} \mathbf{V}_\theta(T_n) = 0$ .

REMARQUE 12.1. **Un estimateur convergent s'écarte donc du paramètre avec une faible probabilité, si la taille de l'échantillon est assez grande.**

DÉFINITION 12.5. (**Biais**)

(1) On appelle **biais** de l'estimateur  $T_n$  par rapport à  $g(\theta)$  l'application :

$$\begin{aligned} b_{T_n} : \Theta &\longrightarrow \mathbb{R} \\ \theta &\longmapsto b_{T_n}(\theta) = \mathbf{E}_\theta(T_n - g(\theta)) \end{aligned}$$

(2) L'**estimateur** est dit **sans biais** si  $b_{T_n}(\theta) = 0$ .

(3) L'**estimateur** est dit **asymptotiquement sans biais** si  $\lim_{n \rightarrow +\infty} b_{T_n}(\theta)$  est nulle.

PROPRIÉTÉS 12.1.  $\forall \theta \in \Theta, b_{T_n}(\theta) = \mathbf{E}_\theta(T_n) - g(\theta)$ .

DÉFINITION 12.6. (**Erreur quadratique ou risque quadratique**) On appelle **erreur quadratique** de  $T_n$  par rapport à  $\theta$  (ou **risque quadratique** de  $T_n$ ) l'application notée  $r_{T_n}(\theta)$  définie sur  $\Theta$  par :

$$r_{T_n}(\theta) = \mathbf{E}_\theta\left((T_n - g(\theta))^2\right)$$

PROPOSITION 12.1. Si le **risque quadratique** de  $T_n$  par rapport à  $g(\theta)$  tend vers 0 quand  $n$  tend vers l'infini, alors  $(T_n)_{n \in \mathbf{N}^*}$  est un **estimateur convergent** de  $g(\theta)$ .

PROPRIÉTÉS 12.2. L'**erreur quadratique** de  $T_n$  par rapport à  $g(\theta)$  est la somme de la variance de  $T_n$  et du carré du **biais**. Autrement dit :

$$\forall \theta \in \Theta, \quad r_{T_n}(\theta) = \mathbf{V}_\theta(T_n) + (b_{T_n}(\theta))^2$$

COROLLAIRE 12.1.  $\forall \theta \in \Theta, r_{T_n}(\theta) \geq \mathbf{V}_\theta(T_n)$ .

REMARQUE 12.2. *Quand un estimateur est sans biais, l'erreur quadratique est égale à la variance. La proposition suivante, conséquence immédiate des précédentes est souvent utilisé pour démontrer qu'un estimateur est convergent.*

PROPOSITION 12.2. *Si un estimateur est sans biais ou asymptotiquement sans biais et si sa variance tend vers 0, alors il est **convergent**.*

REMARQUE 12.3. (**Comparaison de deux estimateurs**) *Si deux estimateurs sont disponibles pour le même paramètre  $\theta$ , on dira que l'un est meilleur que l'autre si son erreur quadratique par rapport à  $\theta$  est **inférieure**.*

PROPOSITION 12.3. (**Estimateur d'une proportion  $p$** ) *On étudie un caractère dans une population donnée de **proportion inconnue  $p$** . Alors  $F_n$  la **fréquence observée** du caractère étudié dans un échantillon de taille  $n$  est un estimateur **sans biais et convergent de  $p$**  avec  $F_n = \frac{1}{n} \sum_{k=1}^n X_k$  où pour tout  $k \in \llbracket 1, n \rrbracket$ ,  $X_k \hookrightarrow \mathcal{B}(p)$  **indépendantes**.*

PROPOSITION 12.4. (**Estimateur d'une moyenne  $\mu$** ) *On étudie un caractère dans une population donnée de **moyenne inconnue  $\mu$** . Alors  $M_n$  la **moyenne observée** du caractère étudié dans un échantillon de taille  $n$  est un estimateur **sans biais et convergent de  $\mu$**  avec  $M_n = \frac{1}{n} \sum_{k=1}^n X_k$  où pour tout  $k \in \llbracket 1, n \rrbracket$ ,  $\mathbf{E}(X_k) = \mu$  et  $\mathbf{V}(X_k) = \sigma^2$ .*

DÉFINITION 12.7. (**Variance empirique**) *On appelle **variance empirique de l'échantillon** la variable aléatoire :*

$$S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \overline{X_n})^2$$

PROPOSITION 12.5.  $\mathbf{E}(S_n^2) = \left(\frac{n-1}{n}\right) \sigma^2$ .

PROPOSITION 12.6. (**Estimateur d'une variance**) *On appelle **variance empirique non biaisé**, l'estimateur :*

$$V_n = \left(\frac{n}{n-1}\right) S_n^2$$

DÉFINITION 12.8. (**Intervalle de confiance d'un paramètre  $g(\theta)$  au risque  $\alpha$** ) *L'intervalle  $[U_n; V_n]$  est un **intervalle de confiance de  $\theta$  au niveau  $1 - \alpha$**  si :*

$$\mathbf{P}([U_n, V_n] \ni g(\theta)) \geq 1 - \alpha$$

où  $\alpha$  est appelé le **risque** et  $U_n, V_n$  sont des statistiques observables de  $g(\theta)$ , c'est-à-dire indépendante du paramètre à estimer.

PROPOSITION 12.7. *L'intervalle :*

$$IC_\alpha(p) = \left[ \underbrace{F_n - u_{1-\alpha/2} \sqrt{\frac{F_n(1-F_n)}{n}}}_{U_n:VAR} ; \underbrace{F_n + u_{1-\alpha/2} \sqrt{\frac{F_n(1-F_n)}{n}}}_{V_n:VAR} \right]$$

avec pour conditions de validité, théorème de la limite centrée oblige,  $\min(na_n, nb_n, n(1-a_n), n(1-b_n)) \geq 5$ ,  $\varepsilon_\alpha$  étant choisi tel que  $\Phi(u_{1-\alpha/2}) = 1 - \alpha/2$ , où  $\Phi$  est la fonction de répartition de la loi  $\mathcal{N}(0, 1)$ , est un

*intervalle de confiance au niveau  $1 - \alpha$  (ou au risque  $\alpha$ ) pour la proportion  $p$ .*

Pour  $\alpha = 5\%$  :

$$IC_{5\%}(p) \subset \left[ \underbrace{F_n - \frac{1.96}{2\sqrt{n}}}_{U_n:VAR} ; \underbrace{F_n + \frac{1.96}{2\sqrt{n}}}_{V_n:VAR} \right]$$

PROPOSITION 12.8. *Si  $\mu$  est la moyenne d'une loi normale de paramètre  $\sigma$  connu, un intervalle de confiance pour  $\mu$  au niveau  $1 - \alpha$  (ou au risque  $\mu$ ) est :*

$$IC_{\alpha}(\mu) = \left[ \underbrace{\bar{X}_n - u_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}}_{U_n:VAR} ; \underbrace{\bar{X}_n + u_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}}_{V_n:VAR} \right]$$

où  $\bar{X}_n$  est la moyenne observée et  $\varepsilon_{\alpha}$  est tel que  $\Phi(u_{1-\alpha/2}) = 1 - \frac{\alpha}{2}$  avec  $n \geq 30$  pour conditions de validité, théorème de la limite centrée oblige.

Pour  $\alpha = 5\%$  et  $n \geq 30$  :

$$IC_{5\%}(\mu) = \left[ \underbrace{\bar{X}_n - 1.96 \frac{\sigma}{\sqrt{n}}}_{U_n:VAR} ; \underbrace{\bar{X}_n + 1.96 \frac{\sigma}{\sqrt{n}}}_{V_n:VAR} \right]$$



# Index

- écart-type
  - cas continu, 51
  - cas discret, 44
- échantillon, 79
- égalité de Cauchy-Schwarz, 9, 48
- équiprobabilité, 19
- équirépartition, 19
- événement
  - quasi certain, 19
  - quasi impossible, 19
- ajustement linéaire entre deux variables aléatoires, 37
- approximation
  - d'une loi binomiale par une loi de Poisson, 75
  - d'une loi binomiale par une loi normale, 76
  - d'une loi de Poisson par une loi normale, 76
  - d'une loi hypergéométrique par une loi binomiale, 75
- argument de minimalité, 17
- arrangement
  - définition, 14
- arrangements
  - dénombrement, 14
- biais, 79
- borélien de  $\mathbb{R}$ , 23
- boule ouverte de  $\mathbb{R}$ , 23
- caractérisation d'une densité, 28
- cardinal, 13
- changement d'indice, 5
- coefficient de corrélation linéaire, 48
- combinaison, 15
  - dénombrement, 15
- combinaison sans répétition, 15
- comparaison de deux estimateurs, 80
- conditionnement et indépendance, 21
- convergence en loi, 75
- convergence en probabilité, 73
- convolée de deux lois à densité, 56
- convolée de deux lois discrètes, 56
- couple discret
  - ajustement linéaire, 37
  - définition, 35
  - fonction d'un couple, 35
- loi
  - caractérisation, 35
  - définition, 35
  - loi conjointe, 35
  - lois conditionnelles, 36
  - lois marginales, 36
  - tableau de contingence à deux entrées, 37
  - tribu, 35
- covariance, 46
  - formule de Koenig, 46
- décomposition canonique d'une VARD, 61
- dénombrer, 13
- degré de liaison entre deux variables aléatoires, 48
- densité
  - d'une somme de deux variables indépendantes, 56
- densité de probabilité, 27
- distribution bilatérale, 68
- distribution de probabilité, 26
- droite de régression
  - de X en Y, 37
  - de Y en X, 37
- ensemble au plus dénombrable, 13
- ensemble dénombrable, 13
- ensemble fini, 13
- erreur quadratique, 79
- espérance
  - antirépartition
    - dans le cas continu, 49, 50
    - dans le cas discret, 44
  - conditionnelle, 45
  - définition
    - cas continu, 49
    - cas discret, 41
  - linéarité
    - cas continu, 49
    - cas discret, 42
- ordre
  - cas continu, 49
  - cas discret, 42
- produit de variables indépendantes, 57
- transformation affine
  - cas continu, 49

- cas discret, 42
- variable bornée, 42
- espace
  - probabilisé, 18
  - probabilisable, 18
- estimateur, 79
  - asymptotiquement sans biais, 79
  - consistant, 79
  - convergent, 79
  - sans biais, 79
- estimateur
  - d'une moyenne, 80
  - d'une proportion  $p$ , 80
- fonction gamma, 69
- fonction
  - d'un couple
    - cas discret, 35
  - d'un vecteur aléatoire, 38
  - d'une variable aléatoire
    - cas continu, 28
    - cas discret, 27
  - de survie, 66
  - gamma, 69
- fonction de régression, 38
- fonction de répartition
  - variable à densité, 27
  - variable discrète, 25
- fonction en escalier, 27
- fonction hypergéométrique, 62
- fonctions de variables indépendantes, 26
- forme bilinéaire symétrique, positive, 46
- formule
  - de l'espérance totale, 45
  - de Thomas Bayes, 22
  - des probabilités composées, 21
  - des probabilités totales, 21
  - du crible ou de Poincaré
    - dénombrement, 13
    - probabilités, 19
    - probabilités conditionnelles, 21
  - du double conditionnement, 21
- identités de polarisation, 47
- inégalité de Boole, 20
- inégalité de Cauchy-Schwarz
  - pour les sommes de réels, 9
  - variables aléatoires, 48
- inégalité de Markov, 73
- inégalités de Bienaymé-Tchebychev, 73
- indépendance
  - de  $n$  événements, 53
  - deux événements, 53
  - deux à deux, 53
  - deux variables aléatoires discrètes, 55
  - deux variables aléatoires quelconques, 54
  - mutuelle, 53
  - $n$  variables discrètes, 55
  - $n$  variables quelconques, 54
  - suite d'événements, 53
  - suite de variables discrètes, 55
  - suite de variables quelconques, 54
  - suite de variables à densité, 56
  - variables à densité, 55
- intégrale de Gauss, 67
- intervalle de confiance, 80
- lemme de coalition
  - cas continu, 56
  - cas discret, 55
  - cas général, 54
- lemme des bergers, 14
- loi
  - d'une fonction d'un couple discret
    - caractérisation, 35
  - d'une fonction d'un vecteur discret
    - caractérisation, 39
  - gamma, 70
  - gamma à un seul paramètre, 70
- loi binomiale
  - définition, 61
  - espérance, 61
  - stabilité, 62
  - stabilité pour la somme de variables indépendantes, 61
  - variance, 61
- loi d'Erlang, 72
- loi d'une fonction d'un vecteur aléatoire discret, 39
- loi d'une variable
  - caractérisation, 26
- loi d'une variable aléatoire
  - cas continu, 27
  - cas discret, 26
  - cas général, 24
- loi de Bernoulli
  - définition, 60
  - espérance, 60
  - variance, 60
- loi de Dirac
  - définition, 59
  - espérance, 59
  - variance, 59
- loi de Poisson
  - définition, 63
  - espérance, 64
  - stabilité, 64
  - stabilité pour la somme, 64
  - variance, 64
- loi exponentielle
  - caractérisation, 66
  - définition, 65
  - espérance, 66
  - fonction de répartition, 66
  - généralisation, 70
  - variance, 66
- loi exponentielle et loi de Poisson, 66
- loi faible des grands nombres, 74
- loi forte des grands nombres, 74

- loi géométrique
  - absence de mémoire, 63
  - antirépartition, 63
  - définition, 63
  - espérance, 63
  - variance, 63
- loi gamma et loi exponentielle , 71
- loi grand gamma
  - définition, 70
  - espérance, 71
  - stabilité, 71
  - variance, 71
- loi hypergéométrique
  - coefficient d'exhaustivité, 62
  - définition, 62
  - espérance, 62
  - variance, 62
- loi normale
  - stabilité, 69
- loi normale centrée et réduite
  - définition, 67
  - espérance, 68
  - fonction de répartition, 67
  - variance, 68
- loi normale quelconque
  - définition, 68
  - espérance, 69
  - stabilité pour la somme, 69
  - transformation affine, 69
  - variance, 69
- loi petit gamma
  - définition, 70
  - espérance, 70
  - stabilité, 70
  - stabilité pour la somme, 70
  - variance, 70
- loi uniforme
  - définition
    - cas continu, 64
    - cas discret, 59
  - espérance
    - cas continu, 65
    - cas discret, 60
  - fonction de répartition
    - cas continu, 65
  - transformation affine
    - cas continu, 65
    - cas discret, 60
  - variance
    - cas continu, 65
    - cas discret, 60
- matrice de covariance, 47
- moment centré d'ordre  $r$ 
  - cas continu, 50
  - VARD, 43
- moment d'ordre deux
  - antirépartition
    - cas discret, 44
    - et antirépartition
      - cas continu, 50
  - moment d'ordre  $r$ 
    - cas continu, 50
    - VARD, 43
  - moment factoriel d'ordre  $r$  d'une VARD, 43
- nombre d'anagrammes, 15
- nombre d'applications, 15
- nombre d'injections, 14
- nombre de permutations, 14
- ouvert de  $\mathbb{R}$ , 23
- p-liste
  - définition, 15
  - dénombrement, 15
- partition, 14
- probabilité
  - conditionnelle, 21
  - définition axiomatique, 18
  - uniforme, 19
- produit
  - à trous, 11
- produit de convolution, 56
- produit et inégalités, 12
- relation affine entre les deux variables, 48
- risque quadratique, 79
- série double
  - théorème de comparaison, 33
  - théorème de Fubini termes positifs, 31
  - théorème de Fubini termes quelconques, 34
  - théorème de sommation par paquets, 33
- sigma - algèbre, 17
- somme de  $n$  variables de Bernoulli indépendantes, 61
- stabilité pour la somme de variables indépendantes
  - loi binomiale, 62
  - loi de Poisson, 64
  - loi grand gamma, 71
  - loi normale, 69
  - loi petit gamma, 70
- statistique, 79
- suite monotone d'événements, 19
- suites croissantes au sens large
  - dénombrement, 15
- suites strictement croissantes
  - dénombrement, 15
- support d'une variable aléatoire
  - cas continu, 28
  - cas discret, 26
- système complet d'événements, 17
- système complet d'événements associée à une variable discrète, 26
- théorème d'or de Bernoulli, 74
- théorème de Huygens-Koenig
  - cas continu, 51

- cas discret, 44
- théorème de Khintchine, 74
- théorème de la limite centrée
  - "version moyenne", 76
  - "version somme", 75
- théorème de Moivre-Laplace, 76
- théorème de transfert
  - cas continu, 49
  - cas des vecteurs discrets, 39
  - cas discret, 41
- tirages
  - bernoulliens, 61
  - simultanés, 16
  - successifs et avec remise, 16
  - successifs et sans remise, 16
- tribu
  - associée à un couple discret, 35
  - associée à une variable aléatoire, 24
  - définition, 17
  - des boréliens de  $\mathbb{R}$ , 23
  - engendrée, 17
  - grossière, 17
- variable à densité, 27
- variable absolument continue, 24
- variable centrée
  - cas continu, 51
  - cas discret, 45
- variable centrée réduite
  - cas continu, 51
  - cas discret, 45
- variable discrète, 24, 26
- variable fonction d'une autre
  - cas discret, 27
- variable indicatrice, 60
- variable réduite
  - cas continu, 51
  - cas discret, 45
- variable sans mémoire
  - cas continu, 29
  - cas discret, 26
- variables discrètes non corrélées, 48
- variance
  - définition
    - cas continu, 51
    - cas discret, 44
  - empirique, 80
  - empirique non biaisé, 80
  - transformation affine
    - cas continu, 51
    - cas discret, 45
- vecteur aléatoire
  - caractérisation, 38
  - définition, 38
  - fonction d'un vecteur aléatoire discret, 38
  - loi, 38
  - loi d'une fonction d'un vecteur aléatoire discret, 39
  - lois marginales, 38