

Séries, intégrales et probabilités (2)

Thierry MEYRE

Préparation à l'agrégation interne. Année 2021-2022.

Université de Paris. IREM.

<https://www.lpsm.paris/dw/doku.php?id=users:meyre:index>

Table des matières

1	Intégrale de Riemann.	11
1.1	Applications réglées.	11
1.2	Construction de l'intégrale de Riemann	16
1.2.1	Rappels sur les applications linéaires continues	16
1.2.2	Intégrale d'une application en escalier	17
1.2.3	Intégrale d'une application réglée	18
1.2.4	Propriétés élémentaires	20
1.2.5	Intégrale d'une application à valeurs réelles	22
1.3	Outils pratiques de calcul d'une intégrale	26
1.3.1	Utilisation d'une primitive	26
1.3.2	Intégration par parties	30
1.3.3	Changement de variable	32
1.3.4	Exercices suggérés.	34
1.3.5	Applications suggérées.	36
1.4	Sommes de Riemann	37
1.5	Convergences de suites d'applications	40
1.6	Intégrale d'une fonction dépendant d'un paramètre	44
1.7	Arcs paramétrés	45
1.7.1	Rectification d'un arc paramétré.	45
1.7.2	Intégrale curviligne d'une forme différentielle.	46
2	Intégrales impropres.	47
2.1	Définition des intégrales impropres.	47
2.2	Étude de la convergence : cas positif	49
2.2.1	Comparaison avec une intégrale de référence.	50
2.2.2	Intégration des relations de comparaison	52
2.2.3	Comparaison avec une série	55
2.3	Cas général : critère de Cauchy et autres méthodes.	56
2.3.1	Critère de Cauchy et conséquences.	56
2.3.2	Méthodes directes	59
2.4	Calcul des intégrales impropres	60

2.4.1	Intégration par parties généralisée	60
2.4.2	Changement de variable généralisé	62
2.5	Application à l'étude de séries.	62
2.6	Transformée de Laplace à variable réelle	67
2.6.1	Quelques propriétés générales	67
2.6.2	Transformée de Laplace et calcul des probabilités	69
3	Intégration sur un intervalle quelconque	71
3.1	Fonctions intégrables	71
3.1.1	Le cas positif	71
3.1.2	Le cas réel	72
3.1.3	Le cas complexe	73
3.2	Théorèmes de convergence	73
3.3	Intégrale dépendant d'un paramètre	80
3.3.1	Étude de la continuité	80
3.3.2	Étude de la dérivabilité	84
3.4	La fonction Γ d'Euler	86
3.5	Transformation de Laplace à variable complexe	90
4	Intégrales multiples	93
4.1	Intégrale double sur un domaine simple borné	93
4.2	Intégrale double sur un produit d'intervalles quelconques	94
4.2.1	Le cas positif	95
4.2.2	Le cas réel ou complexe	96
4.3	Changement de variable	98
5	Méthodes de calcul approché d'une intégrale	101
5.1	Intégrale d'une fonction polynomiale.	102
5.1.1	Rang et ordre d'une méthode de quadrature	102
5.1.2	Exemples pour $p = 0$	104
5.1.3	Exemples avec une subdivision régulière : méthodes de Newton-Cotes	105
5.2	Interpolation polynomiale.	108
5.2.1	Interpolation de Lagrange et Hermite	109
5.2.2	Évaluation de l'erreur	109
5.3	Intégration numérique	111
5.3.1	Avec un polynôme de Lagrange	111
5.3.2	Avec un polynôme d'Hermite	112
5.3.3	Méthodes composées.	115
5.3.4	Lien avec la méthode de Romberg-Richardson	116
5.4	Méthodes de quadrature de Gauss.	117

5.5	Une méthode probabiliste	118
6	Applications de l'analyse au calcul des grandeurs	119
6.1	Longueur d'un arc de classe C^1	119
6.1.1	Rectification d'un arc paramétré.	119
6.1.2	Calcul pratique	121
6.2	Produit mixte et produit vectoriel dans \mathbb{R}^n	124
6.2.1	Orientation de \mathbb{R}^n	124
6.2.2	Produit mixte dans \mathbb{R}^n euclidien orienté	125
6.2.3	Produit vectoriel dans \mathbb{R}^n euclidien orienté	126
6.2.4	Interprétations géométriques en dimension 3	128
6.3	Théorème du changement de variables	129
6.3.1	Enoncé général	129
6.3.2	Quelques cas particuliers	131
6.4	Formule de Green-Riemann	134
6.5	Aire d'une nappe géométrique	134
6.5.1	Nappe paramétrée, nappe géométrique	134
6.5.2	Aire d'une nappe géométrique	136
7	Espaces de Banach, suites et séries	137
7.1	Définition et premiers exemples	137
7.2	Séries à termes réels positifs	138
7.2.1	Définitions et théorèmes de comparaison	138
7.2.2	Comparaison à une série géométrique	141
7.2.3	Série et intégrale impropre positives.	143
7.3	Séries à valeurs dans un espace de Banach	144
7.4	Espaces de Banach usuels de suites et de fonctions	146
7.5	Suites d'applications à valeurs dans un espace de Banach . . .	149
7.6	Séries d'applications à valeurs dans un espace de Banach . . .	152
8	Exponentielle dans une algèbre de Banach	155
8.1	Définitions et premières propriétés	155
8.2	Exponentielle de matrice	158
8.2.1	Calcul explicite	159
8.2.2	Quelques applications topologiques	163
8.2.3	Application aux systèmes différentiels linéaires	164
9	Espaces préhilbertiens	167
9.1	Produit scalaire sur un \mathbb{K} -espace vectoriel.	167
9.2	Norme euclidienne, norme hermitienne	169
9.3	Orthogonalité et procédé de Gram-Schmidt	171

9.4	Projection orthogonale, meilleure approximation	173
9.5	Inégalité de Bessel et égalité de Parseval	176
10	Séries de Fourier	179
10.1	Polynôme et série trigonométriques	180
10.2	Coefficients et série de Fourier	182
10.3	Le théorème de convergence de Dirichlet	184
10.4	Le théorème de Fejér	188
10.5	Approximation en moyenne quadratique	192
11	Séries entières	197
11.1	L'algèbre des séries entières	197
11.2	Rayon de convergence	198
11.3	Continuité et dérivabilité	202
11.4	Principe des zéros isolés	205
11.5	Formule de Cauchy, égalité de Parseval	207
11.6	Fonctions développables en série entière	209
12	Espaces mesurés. Espaces probabilisés.	213
12.1	Clans et tribus.	213
12.1.1	Définitions.	213
12.1.2	Clans et tribus engendrés.	215
12.2	Mesures positives.	216
12.2.1	Définitions.	216
12.2.2	Propriétés d'une mesure.	217
12.3	Application à la modélisation du hasard.	220
12.3.1	Un modèle simple : le modèle additif.	220
12.3.2	Modèle définitif : les espaces probabilisés.	225
12.4	Probabilités conditionnelles. Indépendance d'événements.	227
12.4.1	Définition des probabilités conditionnelles.	227
12.4.2	Formule de Bayes.	230
12.4.3	Indépendance d'événements.	231
12.5	Exercices.	234
12.5.1	Clans.	234
12.5.2	Formule de Poincaré.	235
12.5.3	Équiprobabilité.	236
12.5.4	Probabilités conditionnelles.	237
12.5.5	Événements indépendants.	238

13 Variables aléatoires réelles	243
13.1 La tribu borélienne réelle	243
13.2 Les variables aléatoires réelles et leurs lois	245
13.3 Fonction de répartition d'une v.a.r.	248
13.4 Variables discrètes	250
14 Variables à densité	255
14.1 Définitions	255
14.2 Exemples classiques de lois à densité	256
14.3 Fonction de répartition	258
14.4 Espérance. Théorème de transfert	260
14.5 Moments d'une variable aléatoire	261
14.6 Exercices sur les fonctions de répartition	262
14.6.1 Minimum de variables exponentielles.	262
14.6.2 Variables amnésiques.	262
14.6.3 Loi du χ^2 à un degré de liberté.	262
14.6.4 Loi gaussienne dans \mathbb{R}	263
14.6.5 Avec une loi de Cauchy.	263
15 Vecteurs aléatoires et indépendance	265
15.1 Les vecteurs aléatoires et leurs lois	266
15.2 Vecteurs aléatoires discrets	268
15.3 Vecteurs aléatoires à densité	268
15.4 Indépendance de p v.a.r.	272
15.5 Produit de convolution de densités	274
15.6 Exercices	277
15.6.1 Calcul de densité marginale.	277
15.6.2 Densité gaussienne en dimension 2.	277
15.6.3 Loi uniforme sur le disque	277
15.6.4 Gauss et Cauchy.	278
15.6.5 Avec des lois exponentielles.	278
15.6.6 Méthode de Box-Muller	278
16 Loi des grands nombres.	281
16.1 Suite indépendante équadistribuée.	282
16.2 Convergences de suites de v.a.r.	285
16.2.1 Événements négligeables, égalité presque sûre.	285
16.2.2 Convergences en probabilité et presque sûre.	286
16.3 Loi des grands nombres	288

17 Loi normale	291
17.1 La loi normale centrée réduite	291
17.2 La loi normale générale	293
17.3 Approximation normale de la loi binomiale	295
18 Le théorème-limite central	299
18.1 Énoncé classique	300
18.2 TLC et fonctions de répartition	301
18.3 Un énoncé sans équidistribution	304
18.4 Test du χ^2 d'ajustement.	305

Probabilités : bibliographie

Ouvrages de référence pour l'agrégation interne

- ESC** Probabilités et Statistiques pour le CAPES et l'Agrégation interne. Escoffier. Éditions Ellipses. 2006
- MEY**¹ Probabilités, cours et exercices corrigés. Agrég. interne, Classes prépa., Licence, Capes. Tome 1. Meyre. Éd. Calvage & Mounet. 2016
- MEY**² Probabilités, cours et exercices corrigés. Agrég. interne, Classes prépa., Licence, Capes. Tome 2. Meyre. Éd. Calvage & Mounet. 2018
- OUV**¹ Probabilités 1, Capes et Agrégation. Ouvrard. Éditions Cassini.

Ouvrages plus difficiles

- BAR** Probabilité. Barbe, Ledoux. EDP Sciences. 2007.
- COT** Exercices de Probabilités. Cottrell et al. Éditions Cassini. 2005.
- FOA** Calcul des Probabilités. Foata, Fuchs. Éditions Dunod. 2ème édition. 1998

Autres ouvrages pour les corrigés des exercices

- BIL** Probability and Measure. Billingsley. Wiley.
- DAC**¹ **cours** Probabilités et Statistiques 1. Problèmes à temps fixe. Dacunha-Castelle, Duflo. 2e édition. Éditions Masson.
- DAC**¹ **exo** Exercices de Probabilités et Statistiques 1. Problèmes à temps fixe. Dacunha-Castelle, Duflo. 3e tirage corrigé. Éditions Masson.
- FEL**¹ An Introduction to Probability Theory and its Applications. Feller. Wiley 3rd edition. Volume I
- FEL**² An Introduction to Probability Theory and its Applications. Feller. Wiley 2nd edition. Volume II
- OUV**² Probabilités 2, Maîtrise et Agrégation. Ouvrard. Éditions Cassini.
- REV** Probabilités. Revuz. Editions Hermann. Collection Méthodes.

ROS Initiation aux Probabilités. Ross. Presses polytechniques et universitaires romandes. Troisième édition.

RUD Analyse réelle et complexe. Rudin. Editions Masson.

Chapitre 12

Espaces mesurés. Espaces probabilisés.

Dans les deux premières sections de ce chapitre, nous allons définir la structure d'espace mesuré. Bien que la théorie abstraite de l'intégration de Lebesgue repose sur cette structure, nous ne la développerons pas ici car elle ne figure pas au programme officiel. En revanche, la structure d'espace mesuré va nous mener au modèle fondamental utilisé en calcul des probabilités, que l'on appelle *espace probabilisé*.

12.1 Clans et tribus.

12.1.1 Définitions.

Définition 12.1.1 Soit E un ensemble. On appelle *clan* sur E un ensemble \mathcal{C} de parties de E tel que les trois conditions suivantes soient satisfaites :

1. $E \in \mathcal{C}$
2. $A \in \mathcal{C} \implies A^c = E - A \in \mathcal{C}$
3. $(A, B) \in \mathcal{C}^2 \implies A \cup B \in \mathcal{C}$

Remarque. Un clan sur E est encore appelé algèbre (de Boole) de parties de E .

De cette définition nous déduisons les propriétés immédiates suivantes :

1. $\emptyset \in \mathcal{C}$
2. $(A, B) \in \mathcal{C}^2 \implies A \cap B \in \mathcal{C}$
3. $(A_1, \dots, A_n) \in \mathcal{C}^n \implies \bigcup_{i=1}^n A_i \in \mathcal{C}$ et $\bigcap_{i=1}^n A_i \in \mathcal{C}$

4. Si A et B sont des ensembles de \mathcal{C} , alors leur différence $A - B = A \cap B^c$ est encore un ensemble de \mathcal{C} . Un cas particulier est celui où $B \subset A$: on parle alors de différence propre.
5. Si A et B sont des ensembles de \mathcal{C} , alors leur différence symétrique $A \Delta B = (A \cup B) - (A \cap B) = (A - B) \cup (B - A)$ est encore un ensemble de \mathcal{C} .

Définition 12.1.2 Soit E un ensemble. On appelle tribu sur E un ensemble \mathcal{T} de parties de E tel que les trois conditions suivantes soient satisfaites :

1. $E \in \mathcal{T}$
2. $A \in \mathcal{T} \implies A^c \in \mathcal{T}$
3. $(A_n)_{n \in \mathbb{N}} \in \mathcal{T}^{\mathbb{N}} \implies \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{T}$

Remarque. Une tribu sur E est encore appelée σ -algèbre (de Boole) de parties de E .

Comme $\emptyset = E^c \in \mathcal{T}$, la stabilité par union dénombrable nous donne aussi la stabilité de \mathcal{T} par union finie. Une tribu est donc un clan stable par passage à l'union dénombrable.

Une tribu est aussi stable par passage à l'intersection dénombrable puisque :

$$\bigcap_{n \in \mathbb{N}} A_n = \left[\bigcup_{n \in \mathbb{N}} A_n^c \right]^c$$

Exemples.

1. L'ensemble $\mathcal{P}(E)$ de toutes les parties de E est la plus grande tribu sur E (au sens de l'inclusion). On l'appelle tribu totale sur E .
2. L'ensemble $\{\emptyset, E\}$ est la plus petite tribu sur E . On l'appelle tribu grossière sur E .
3. Sur \mathbb{R} , l'ensemble \mathcal{C} des unions finies d'intervalles de la forme

$$]-\infty, a[\quad , \quad [a, b[\quad \text{ou} \quad [b, +\infty[$$

est un clan et non une tribu.

Le lecteur pourra montrer d'abord la stabilité de \mathcal{C} par intersection finie avant de prouver la stabilité par passage au complémentaire. Il pourra remarquer aussi que tout élément de \mathcal{C} a un nombre fini de composantes connexes et donc que :

$$\bigcup_{n \in \mathbb{N}} \left[n, n + \frac{1}{2} \right] \notin \mathcal{C}$$

12.1.2 Clans et tribus engendrés.

Proposition 12.1.3 *L'intersection d'une famille non vide $(\mathcal{C}_i)_{i \in I}$ de clans sur E est encore un clan sur E .*

L'intersection d'une famille non vide $(\mathcal{T}_i)_{i \in I}$ de tribus sur E est encore une tribu sur E .

Corollaire 12.1.4 *Soit \mathcal{E} un ensemble quelconque de parties de E . Il existe un plus petit clan sur E (au sens de l'inclusion) qui contienne \mathcal{E} ; on l'appelle clan engendré par \mathcal{E} et on le note $\mathcal{C}(\mathcal{E})$.*

Il existe une plus petite tribu sur E (au sens de l'inclusion) qui contienne \mathcal{E} ; on l'appelle tribu engendrée par \mathcal{E} et on la note $\sigma(\mathcal{E})$.

Démonstration: La famille des clans sur E contenant \mathcal{E} n'est pas vide car elle contient $\mathcal{P}(E)$. L'intersection de tous les clans sur E contenant \mathcal{E} est donc encore un clan, il contient bien sûr \mathcal{E} et est manifestement le plus petit au sens de l'inclusion ayant cette propriété. Un raisonnement similaire s'applique aux tribus sur E . □

Définition 12.1.5 *Soient E et F des ensembles et $f : E \rightarrow F$ une application. Pour tout $A \subset F$, on appelle image réciproque de A par f et l'on note $f^{-1}(A)$ le sous-ensemble de E défini par :*

$$f^{-1}(A) = \{x \in E, f(x) \in A\}$$

Remarques.

1. Nous utiliserons également la notation $\{f \in A\}$ pour l'image réciproque de A par f .
2. Si \mathcal{F} est un ensemble de parties de F , nous noterons $f^{-1}(\mathcal{F})$ l'ensemble des parties de E de la forme $f^{-1}(A)$, avec $A \in \mathcal{F}$.

Proposition 12.1.6 *Soient E et F des ensembles et $f : E \rightarrow F$ une application.*

Si \mathcal{U} est une tribu sur F , alors $f^{-1}(\mathcal{U})$ est une tribu sur E appelée tribu image réciproque de \mathcal{U} par f .

Démonstration: Nous constatons d'abord que $E = f^{-1}(F) \in f^{-1}(\mathcal{U})$. Pour tout $U \in \mathcal{U}$, nous avons l'égalité

$$[f^{-1}(U)]^c = f^{-1}(U^c)$$

d'où la stabilité de $f^{-1}(\mathcal{U})$ par passage au complémentaire.
Enfin, si $(U_n) \in \mathcal{U}^{\mathbb{N}}$, l'égalité

$$\bigcup_{n \in \mathbb{N}} f^{-1}(U_n) = f^{-1} \left(\bigcup_{n \in \mathbb{N}} U_n \right)$$

nous donne la stabilité de $f^{-1}(\mathcal{U})$ par union dénombrable. □

12.2 Mesures positives.

12.2.1 Définitions.

Définition 12.2.1 On appelle espace mesurable un couple (E, \mathcal{T}) , où E est un ensemble et \mathcal{T} une tribu sur E . Les éléments de la tribu \mathcal{T} sont appelés ensembles mesurables.

Définition 12.2.2 Soit (E, \mathcal{T}) un espace mesurable. On appelle mesure positive sur (E, \mathcal{T}) une application $\mu : \mathcal{T} \rightarrow [0, +\infty]$ qui satisfait les deux conditions suivantes :

1. $\mu(\emptyset) = 0$,
2. pour toute suite $(A_n)_{n \in \mathbb{N}^*}$ d'ensembles de \mathcal{T} deux à deux disjoints, on a l'égalité :

$$\mu \left(\bigcup_{n \in \mathbb{N}^*} A_n \right) = \sum_{n \in \mathbb{N}^*} \mu(A_n)$$

Remarques.

1. La 2^e condition intervenant dans cette définition s'appelle σ -additivité.
2. Pour tout ensemble mesurable $A \in \mathcal{T}$, on appelle $\mu(A)$ la mesure de A .
3. Il est fréquent de dire simplement mesure au lieu de mesure positive.

Notation. Pour une suite $(A_n)_{n \in \mathbb{N}^*}$ d'ensembles dont nous aurons préalablement vérifié qu'ils sont 2 à 2 disjoints, nous emploierons la notation :

$$\bigsqcup_{n \in \mathbb{N}^*} A_n = \bigcup_{n \in \mathbb{N}^*} A_n$$

De la sorte, la propriété de σ -additivité d'une mesure se réécrit :

$$\mu \left(\bigsqcup_{n \in \mathbb{N}^*} A_n \right) = \sum_{n \in \mathbb{N}^*} \mu(A_n)$$

Définition 12.2.3 On appelle espace mesuré un triplet (E, \mathcal{T}, μ) , où μ est une mesure sur l'espace mesurable (E, \mathcal{T}) .

Exemples.

1. Soit E un ensemble. Sur l'espace mesurable $(E, \mathcal{P}(E))$, nous définissons une mesure μ , dite de comptage, en convenant que pour toute partie A de E , $\mu(A)$ vaut le cardinal de A (noté dans la suite $\#A$). Le lecteur vérifiera la σ -additivité en distinguant les deux cas où le sous-ensemble $\sum_{n \in \mathbb{N}^*} A_n$ est fini ou infini.
2. Soit E un ensemble non vide et $a \in E$. Sur l'espace mesurable $(E, \mathcal{P}(E))$, nous définissons la mesure de Dirac au point a , notée δ_a , comme suit :

$$\forall A \in \mathcal{P}(E) \quad \delta_a(A) = 1 \text{ si } a \in A \quad ; \quad \delta_a(A) = 0 \text{ si } a \notin A.$$

Le lecteur vérifiera facilement la propriété de σ -additivité en remarquant que, si les $(A_n)_{n \in \mathbb{N}^*}$ sont deux à deux disjoints, alors a est élément au plus d'un d'entre eux.

12.2.2 Propriétés d'une mesure.

Additivité.

Une conséquence immédiate de la définition est la propriété d'*additivité* d'une mesure : si A_1, \dots, A_n sont des ensembles mesurables 2 à 2 disjoints, alors

$$\mu(A_1 \cup \dots \cup A_n) = \mu(A_1) + \dots + \mu(A_n).$$

Il suffit en effet d'appliquer la définition 12.2.2 en prenant $A_k = \emptyset$ pour $k \geq n + 1$.

Cette propriété d'additivité nous donne l'égalité suivante :

$$\forall (A, B) \in \mathcal{T}^2 \quad \mu(A) = \mu(A - B) + \mu(A \cap B) \quad (12.1)$$

En effet, l'ensemble A est égal à la réunion des deux ensembles disjoints $A - B$ et $A \cap B$.

En particulier, si $B \subset A$, nous obtenons $\mu(A) = \mu(A - B) + \mu(B)$, d'où la propriété suivante concernant une différence propre :

$$B \subset A \text{ et } \mu(B) < +\infty \quad \implies \mu(A - B) = \mu(A) - \mu(B)$$

Comme l'ensemble $A \cup B$ est réunion des ensembles disjoints B et $A - B$, nous obtenons en utilisant (12.1) :

$$\forall (A, B) \in \mathcal{T}^2 \quad \mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B).$$

Si $\mu(A \cap B) < +\infty$, nous pouvons également écrire :

$$\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B).$$

Croissance.

$$\forall (A, B) \in \mathcal{T}^2 \quad A \subset B \implies \mu(A) \leq \mu(B)$$

En effet, on a l'égalité : $\mu(B) = \mu(A) + \mu(B - A)$.

Sous- σ -additivité.

$$\forall (A_n) \in \mathcal{T}^{\mathbb{N}^*} \quad \mu \left(\bigcup_{n \in \mathbb{N}^*} A_n \right) \leq \sum_{n \in \mathbb{N}^*} \mu(A_n) \quad (12.2)$$

Démonstration: Posons $U_0 = \emptyset$ et, pour $n \in \mathbb{N}^*$, $U_n = A_1 \cup \dots \cup A_n$ et $B_n = U_n - U_{n-1} = A_n - U_{n-1}$. Par construction, les ensembles B_n sont deux à deux disjoints, on a $B_n \subset A_n$ pour tout $n \in \mathbb{N}^*$ et :

$$\bigcup_{n \in \mathbb{N}^*} B_n = \bigcup_{n \in \mathbb{N}^*} A_n.$$

Nous en déduisons :

$$\mu \left(\bigcup_{n \in \mathbb{N}^*} A_n \right) = \mu \left(\bigcup_{n \in \mathbb{N}^*} B_n \right) = \sum_{n \in \mathbb{N}^*} \mu(B_n) \leq \sum_{n \in \mathbb{N}^*} \mu(A_n)$$

□

Passage à la limite croissante.

Si $(A_n) \in \mathcal{T}^{\mathbb{N}^*}$ est une suite croissante au sens de l'inclusion, c'est-à-dire telle que $A_n \subset A_{n+1}$ pour tout $n \in \mathbb{N}^*$, nous convenons d'adopter la notation suivante :

$$\bigcup_{n \in \mathbb{N}^*} A_n = \lim \uparrow A_n$$

Pour toute suite $(A_n) \in \mathcal{T}^{\mathbb{N}^*}$ croissante au sens de l'inclusion, nous avons alors :

$$\mu(\lim \uparrow A_n) = \lim \uparrow \mu(A_n)$$

Démonstration: Si nous introduisons les mêmes ensembles que dans la démonstration précédente, la croissance de la suite $(A_n) \in \mathcal{T}^{\mathbb{N}^*}$ nous donne ici : $U_n = A_n = B_1 \cup \dots \cup B_n$, d'où $\mu(A_n) = \mu(B_1) + \dots + \mu(B_n)$. Toujours en procédant comme dans la démonstration précédente, nous en déduisons :

$$\mu\left(\bigcup_{n \in \mathbb{N}^*} A_n\right) = \mu\left(\bigcup_{n \in \mathbb{N}^*} B_n\right) = \sum_{n \in \mathbb{N}^*} \mu(B_n) = \lim \uparrow \mu(A_n).$$

□

Passage à la limite décroissante.

Si $(A_n) \in \mathcal{T}^{\mathbb{N}^*}$ est une suite décroissante au sens de l'inclusion, c'est-à-dire telle que $A_{n+1} \subset A_n$ pour tout $n \in \mathbb{N}^*$, nous convenons d'adopter la notation suivante :

$$\bigcap_{n \in \mathbb{N}^*} A_n = \lim \downarrow A_n$$

Pour toute suite $(A_n) \in \mathcal{T}^{\mathbb{N}^*}$ décroissante au sens de l'inclusion, nous avons alors l'implication suivante :

$$\mu(A_1) < +\infty \implies \mu(\lim \downarrow A_n) = \lim \downarrow \mu(A_n)$$

Démonstration: Pour tout $n \in \mathbb{N}^*$, posons $C_n = A_1 - A_n$; puisque $A_n \subset A_1$, nous avons :

$$\mu(A_1) = \mu(C_n) + \mu(A_n).$$

Puisque $\mu(A_1) < +\infty$, nous en déduisons que $\mu(C_n) < +\infty$ et $\mu(A_n) < +\infty$, ainsi que l'égalité :

$$\mu(C_n) = \mu(A_1) - \mu(A_n) \tag{12.3}$$

La suite $(C_n)_{n \in \mathbb{N}^*}$ est croissante et nous avons l'égalité ensembliste :

$$\bigcup_{n \in \mathbb{N}^*} C_n = A_1 - \lim \downarrow A_n$$

Par passage à la limite croissante, puis en utilisant (12.3), nous obtenons donc :

$$\mu(A_1 - \lim \downarrow A_n) = \lim \uparrow \mu(C_n) = \mu(A_1) - \lim \downarrow \mu(A_n).$$

Il nous reste à écrire, en utilisant de nouveau $\mu(A_1) < +\infty$:

$$\mu(\lim \downarrow A_n) = \mu(A_1) - \mu(A_1 - \lim \downarrow A_n)$$

pour obtenir la conclusion voulue.

□

Remarques.

1. Il est facile de constater que la condition $\mu(A_1) < +\infty$ peut être remplacée dans l'énoncé précédent par :

$$\exists n \in \mathbb{N}^* \quad \mu(A_n) < +\infty$$

2. En revanche, si tous les A_n sont de mesure infinie, la propriété précédente peut tomber en défaut, comme le montre le contre-exemple suivant.

Si nous notons μ la mesure de comptage sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$ et si nous prenons $A_n = \{k \in \mathbb{N}, k \geq n\}$, nous avons $\lim \downarrow A_n = \emptyset$ d'où $\mu(\lim \downarrow A_n) = 0$ alors que $\lim \downarrow \mu(A_n) = +\infty$.

12.3 Application à la modélisation du hasard.

L'objet de la théorie des Probabilités est de construire un modèle permettant d'étudier les phénomènes dépendant du hasard. Nous allons procéder en deux étapes en présentant d'abord un modèle simple, mais dont le champ d'application sera limité, puis en le raffinant pour aboutir à notre modèle définitif.

12.3.1 Un modèle simple : le modèle additif.**L'ensemble des résultats possibles.**

Qu'est-ce que le hasard ? Comme nous enseignons les mathématiques et non la philosophie, nous allons adopter une définition empirique, c'est-à-dire basée sur l'expérience.

Nous dirons qu'une expérience (scientifique ou autre, de la vie courante par exemple) dépend du hasard et nous l'appellerons *expérience aléatoire* – du mot latin *alea* qui signifie jet de dés – si elle produit un résultat que l'on ne connaît pas à l'avance mais qui en revanche appartient à un ensemble connu à l'avance. Cet ensemble, appelé *ensemble des résultats possibles* est noté traditionnellement Ω . Il peut être plus ou moins grand suivant l'expérience considérée.

Exemples :

1. Jeu de pile ou face. $\Omega = \{p, f\}$.
2. Jet de dé à 6 faces. $\Omega = \{1, 2, \dots, 6\}$.

3. On joue à pile ou face de façon répétée et l'on s'intéresse au rang d'apparition du premier pile. $\Omega = \mathbb{N}^*$.
4. Durée de vie d'une batterie de voiture. $\Omega = \mathbb{R}_+$ ou $\Omega = [0, T]$ avec T suffisamment grand.
5. Courbe d'évolution de la température sur une journée dans une station météo. $\Omega = C^0([0, T], \mathbb{R})$.

Dans les exemples 1 à 3, Ω est fini ou dénombrable : c'est le cas *discret*. Dans les exemples 4 et 5, Ω est infini non dénombrable : c'est le cas *continu*.

L'ensemble Ω est le premier ingrédient du modèle mathématique que nous allons construire pour représenter le mieux possible notre expérience aléatoire. Nous allons mettre en oeuvre deux ingrédients supplémentaires pour construire notre modèle additif.

L'ensemble des événements.

Une fois faite notre expérience aléatoire, nous nous trouvons devant un certain résultat $\omega \in \Omega$. À part peut-être le cas où Ω est très petit, il arrive très souvent que ce ne soit pas le résultat précis ω qui nous intéresse mais que l'on cherche plutôt à répondre à la question : "est-ce que ω appartient à tel sous-ensemble donné de Ω ?"

Exemple. Durée de vie d'un ordinateur.

Si le constructeur vous garantit la machine pendant trois ans, il a bien sûr d'abord étudié la question : « quelles sont les chances pour que $\omega \leq 3$? » De la réponse à cette question dépendra le prix auquel il va vous facturer la garantie. Dans cet exemple, la question intéressante s'écrit donc : « $\omega \in A$? », avec $A = [3, +\infty[\subset \Omega = \mathbb{R}_+$.

Nous appellerons *événement* un sous-ensemble A de Ω tel qu'un observateur de l'expérience est capable de répondre à la question " $\omega \in A$?"

Exemples. Ω est l'événement certain. \emptyset est l'événement impossible.

Dès que nous allons manipuler des événements, certains opérateurs ensemblistes vont apparaître :

- L'événement contraire de A est son complémentaire $A^c = \Omega - A$.
- L'événement "le résultat est dans A_1 ou A_2 " s'écrit $A_1 \cup A_2$
- L'événement "le résultat est dans A_1 et A_2 " s'écrit $A_1 \cap A_2$

Nous travaillerons donc sur des intersections ou des unions finies d'événements, ainsi que sur des complémentaires d'événements.

Un observateur de l'expérience étant donné, il est donc naturel de supposer que l'ensemble des parties de Ω dont il peut discerner la réalisation ou la non-réalisation après l'expérience a une structure de clan. Nous noterons dans la suite \mathcal{C} le clan des événements discernables par notre observateur.

Remarque. Plusieurs observateurs distincts peuvent être associés à une même expérience. Par exemple, si le résultat de l'expérience aléatoire qui nous intéresse est le temps (mesuré en secondes) mis par une athlète pour courir un 100 mètres (prenons $\Omega = [0; 15]$) et si un premier arbitre n'est muni que d'une montre à trotteuse tandis que le second se réfère à un dispositif électronique, nous avons $\mathcal{C}_1 \subsetneq \mathcal{C}_2$. Ainsi, l'événement $A = [0; 9, 73]$ qui se traduit en langage courant par "elle a battu le record du monde féminin du 9 septembre 2007" vérifie $A \in \mathcal{C}_2$ mais $A \notin \mathcal{C}_1$.

Probabilités.

Nous avons tous l'intuition que certains événements ont plus de chances de se produire que d'autres. Par exemple, dans une loterie, il est plus probable de tirer un billet perdant que de tirer le gros lot.

Pour préciser cette intuition, nous souhaitons associer à un événement donné un nombre réel qui mesure les chances qu'il se produise ; par exemple, si l'événement se produit avec une chance sur deux (pile ou face avec une pièce équilibrée), nous lui associerons le nombre $1/2$.

Nous appellerons donc probabilité une application $P : \mathcal{C} \rightarrow [0, 1]$ vérifiant certaines conditions que nous allons préciser maintenant.

Tout d'abord, Ω étant l'événement certain, nous demanderons $P(\Omega) = 1$ (en langage courant, il y a "100 chances sur 100" qu'il se produise).

Ensuite, il est naturel d'exiger la propriété d'*additivité* :

$$A \in \mathcal{C}, B \in \mathcal{C}, A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B).$$

Notons que la condition $A \cap B = \emptyset$ – les deux événements A et B sont *disjoints* – est indispensable pour ne pas compter deux fois les mêmes résultats possibles.

Une des raisons pour lesquelles il est naturel de faire cette hypothèse d'additivité d'une probabilité s'appelle :

La loi empirique des grands nombres

Si nous pouvons répéter une même expérience un grand nombre de fois dans des conditions identiques et que nous notions $\varphi_n(A)$ le nombre de fois où le résultat a appartenu à l'événement A au cours des n premières répétitions de l'expérience, nous nous attendons à ce que le rapport $\varphi_n(A)/n$ (fréquence de réalisation de l'événement A au cours des n premières répétitions de l'expérience) tende vers une certaine limite lorsque n augmente.

Intuitivement, c'est cette limite que nous aurions envie d'appeler "probabilité que l'événement A se réalise". Si nous admettons qu'une telle limite existe pour tout événement A et que nous la notions $P(A)$, il est facile de constater que nous avons ainsi défini une application $P : \mathcal{C} \rightarrow [0, 1]$, additive et telle que $P(\Omega) = 1$.

Le modèle additif

Nous pouvons donc modéliser une expérience aléatoire par un triplet (Ω, \mathcal{C}, P) , où Ω est un ensemble, \mathcal{C} un clan sur Ω et P une application additive de \mathcal{C} dans $[0, 1]$ telle que $P(\Omega) = 1$.

Ce modèle s'avère en général suffisant lorsque l'expérience aléatoire considérée ne mène qu'à un nombre fini de résultats possibles, c'est-à-dire lorsque le cardinal de Ω vérifie : $\#\Omega < +\infty$.

Lorsque Ω est fini, nous prendrons pratiquement toujours $\mathcal{C} = \mathcal{P}(\Omega)$, qui est le seul clan sur Ω contenant tous les singletons : ce clan d'événements correspond à un observateur qui discerne le résultat précis ω de l'expérience.

Nous allons maintenant considérer un cas particulier qui a de nombreuses applications pratiques.

Exemple. Probabilité uniforme sur un ensemble Ω fini.

Nous nous plaçons dans le cas où $\#\Omega < +\infty$ et nous prenons pour ensemble des événements $\mathcal{C} = \mathcal{P}(\Omega)$.

Il existe une unique probabilité sur (Ω, \mathcal{C}) qui attribue les mêmes chances à tous les résultats possibles (on dit alors qu'on est en situation d'équiprobabilité). Elle est appelée *probabilité uniforme* sur Ω et est définie par :

$$\forall \omega \in \Omega, \quad P(\{\omega\}) = \frac{1}{\#\Omega}.$$

Pour tout événement $A \subset \Omega$, on a alors :

$$P(A) = \frac{\#A}{\#\Omega} = \frac{\text{nombre de cas favorables}}{\text{nombre de cas possibles}}.$$

Nous terminons ce sous-paragraphe par une proposition caractérisant les probabilités sur $(\Omega, \mathcal{P}(\Omega))$ lorsque l'ensemble des résultats possibles Ω est fini.

Proposition 12.3.1 *Soit $n \in \mathbb{N}^*$ et $\Omega = \{\omega_1, \dots, \omega_n\}$ un ensemble fini. Il existe une bijection entre l'ensemble des probabilités P sur $(\Omega, \mathcal{P}(\Omega))$ et l'ensemble des n -uplets $(p_k)_{1 \leq k \leq n} \in (\mathbb{R}_+)^n$ tels que :*

$$\sum_{1 \leq k \leq n} p_k = 1.$$

Cette bijection est donnée par :

$$\forall k \in \{1, \dots, n\} \quad p_k = P(\{\omega_k\}).$$

Démonstration: Dans le sens direct, il est évident que si P est une probabilité sur $(\Omega, \mathcal{P}(\Omega))$, alors le n -uplet $(p_k)_{1 \leq k \leq n}$ ainsi défini a les propriétés voulues.

Réciproquement, un tel n -uplet étant donné, s'il existe une probabilité P qui convienne, alors elle satisfait nécessairement la relation suivante par additivité :

$$\forall A \in \mathcal{P}(\Omega) \quad P(A) = \sum_{k/\omega_k \in A} p_k.$$

Il est facile de conclure en vérifiant que l'on a bien défini ainsi une probabilité sur $(\Omega, \mathcal{P}(\Omega))$. \square

Exemples.

1. Le cas le plus simple d'espace de probabilité fini non trivial est donné par $\#\Omega = 2$ et correspond à la modélisation du jeu de pile ou face (avec une pièce éventuellement biaisée). Il y a alors bijection entre les probabilités sur $(\Omega, \mathcal{P}(\Omega))$ et le segment $[0, 1]$: celle-ci est donnée par la probabilité $p \in [0, 1]$ que la pièce tombe sur pile. Formellement, on écrit en général $\Omega = \{0, 1\}$ (pile étant représenté par 1 et face par 0) et la probabilité définie par $P(\{1\}) = p$ et donc $P(\{0\}) = 1 - p$ est appelée probabilité de Bernoulli de paramètre $p \in [0, 1]$.

2. Pour tout $n \in \mathbb{N}^*$ et tout $p \in [0, 1]$, on définit sur $\Omega = \{0, 1, \dots, n\}$ (qui est donc de cardinal $n + 1$) la probabilité binomiale de paramètres n et p , notée $P = B_{n,p}$, par

$$\forall k \in \{0, 1, \dots, n\} \quad B_{n,p}(\{k\}) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Nous verrons plus loin dans quel type de situation la probabilité binomiale apparaît.

12.3.2 Modèle définitif : les espaces probabilisés.

Le modèle développé dans le paragraphe précédent s'applique bien au cas Ω fini, qui couvre un certain nombre de situations courantes : jeux de hasard, sondages d'opinion, contrôles de qualité etc. Néanmoins, même avec un jeu aussi simple que le pile ou face, on ne peut pas toujours se restreindre au cas $\#\Omega < +\infty$: si nous étudions par exemple le rang d'apparition du premier pile lors de jets successifs, on est amené à prendre $\Omega = \mathbb{N}^*$ puisqu'il est impossible de borner à l'avance ce rang par un entier fixé.

Lorsque le temps intervient, nous voyons apparaître naturellement des espaces encore plus "gros" (infinis non dénombrables) :

- durée de vie d'un composant électronique : $\Omega = \mathbb{R}_+$
- évolution du cours d'un actif financier : $\Omega = C^0([0, T], \mathbb{R})$
- déplacement d'une particule : $\Omega = C^0([0, T], \mathbb{R}^3)$

Le modèle du paragraphe précédent présente d'autres insuffisances : la structure de clan pour l'ensemble \mathcal{C} des événements et l'additivité d'une probabilité sont des hypothèses trop faibles pour nous permettre de mener effectivement des calculs lorsque Ω est infini, comme nous allons le voir sur le simple exemple suivant.

Si nous revenons au rang du premier pile lors de jets successifs d'une pièce équilibrée, une première question qui se pose est de savoir si ce rang est fini : peut-on obtenir face indéfiniment ?

Intuitivement, cela paraît extrêmement improbable et l'on s'attend à ce que la probabilité en soit nulle. Plus précisément, il est naturel de faire le raisonnement suivant : l'événement $A_n = \{ \text{on a obtenu face lors des } n \text{ premiers jets} \}$ est de probabilité $1/2^n$ et donc l'événement $A = \{ \text{on obtient face indéfiniment} \}$ a pour probabilité $\lim 1/2^n = 0$.

Pour rendre ce raisonnement rigoureux, il faudrait déjà que $A = \bigcap_n A_n$ soit effectivement un événement, ce qui n'est pas assuré sous la seule hypothèse : l'ensemble des événements est un clan. Nous avons besoin de supposer que

l'ensemble des événements est une tribu.

En outre, nous voudrions avoir : $A = \lim \downarrow A_n \Rightarrow P(A) = \lim \downarrow P(A_n)$. En passant aux événements complémentaires, cela revient à demander la propriété $P(\lim \uparrow B_n) = \lim \uparrow P(B_n)$. Or, pour une application additive comme P , cette propriété est équivalente à la σ -additivité : pour le voir, il suffit d'écrire, pour une suite (A_n) d'événements deux à deux disjoints

$$\bigcup_{n \in \mathbb{N}^*} A_n = \lim \uparrow \bigcup_{k=1}^n A_k.$$

Pour pouvoir mener nos calculs, nous allons donc imposer des contraintes supplémentaires sur les éléments constitutifs de notre modèle : l'ensemble des événements doit être une tribu (ou σ -algèbre), que nous noterons désormais plutôt \mathcal{A} , et P une application σ -additive. Nous en arrivons donc à la définition suivante, qui décrit le modèle que nous utiliserons désormais.

Définition 12.3.2 Une probabilité sur un espace mesurable (Ω, \mathcal{A}) est une mesure positive sur cet espace telle que $P(\Omega) = 1$.

Le triplet (Ω, \mathcal{A}, P) est alors appelé espace probabilisé (ou espace de probabilité).

En d'autres termes, un espace probabilisé est un espace mesuré de masse totale égale à 1.

Notons que dans le cas $\#\Omega < +\infty$, cette définition est équivalente à celle du paragraphe précédent. En effet, dans ce cas, $\mathcal{A} \subset \mathcal{P}(\Omega)$ est fini si bien que \mathcal{A} algèbre et \mathcal{A} σ -algèbre sont équivalents, de même que P additive et P σ -additive. Comme nous n'utiliserons le modèle additif que dans le cas $\#\Omega < +\infty$, nous le considérerons comme un cas particulier de notre modèle général.

Contrairement à ce qui se passait dans le cas $\#\Omega < +\infty$, dans le cas général, nous serons amenés à considérer des tribus d'événements $\mathcal{A} \subsetneq \mathcal{P}(\Omega)$. En effet, plus la tribu \mathcal{A} est grande, plus il est difficile de construire une mesure de probabilité sur celle-ci. Lorsque Ω est infini non dénombrable, $\mathcal{P}(\Omega)$ est en général trop grand pour que l'on puisse construire sur cette tribu une mesure de probabilité qui mène à un modèle satisfaisant. En fait, on a souvent recours à une politique minimaliste en prenant pour \mathcal{A} la plus petite tribu (c'est-à-dire la tribu engendrée) qui contienne les événements que nous voulons pouvoir observer.

Dans le cas où Ω est infini dénombrable, nous avons la caractérisation suivante d'une probabilité sur $(\Omega, \mathcal{P}(\Omega))$, que nous admettrons.

Proposition 12.3.3 Soit $\Omega = \{\omega_n, n \in \mathbb{N}\}$ un ensemble infini dénombrable. Il existe une bijection entre l'ensemble des probabilités P sur $(\Omega, \mathcal{P}(\Omega))$ et l'ensemble des suites $(p_n)_{n \in \mathbb{N}} \in (\mathbb{R}_+)^{\mathbb{N}}$ telles que :

$$\sum_{n \in \mathbb{N}} p_n = 1.$$

Cette bijection est donnée par :

$$\forall n \in \mathbb{N} \quad p_n = P(\{\omega_n\}).$$

Par σ -additivité, nous avons alors la relation :

$$\forall A \in \mathcal{P}(\Omega) \quad P(A) = \sum_{k/\omega_k \in A} p_k.$$

Exemples. Sur l'espace mesurable $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$, on définit :

1. pour tout $p \in]0, 1[$, la probabilité géométrique de paramètre p par

$$\forall n \in \mathbb{N} \quad P(\{n\}) = (1 - p) p^n.$$

2. pour tout $\lambda > 0$, la probabilité de Poisson de paramètre λ par

$$\forall n \in \mathbb{N} \quad P(\{n\}) = e^{-\lambda} \frac{\lambda^n}{n!}.$$

□

Le choix d'une probabilité P sur (Ω, \mathcal{A}) qui “colle au plus près” à la réalité d'une expérience aléatoire est une question suffisamment vaste pour faire l'objet d'une branche entière des mathématiques : la statistique.

Le fait qu'un même espace probabilisable (Ω, \mathcal{A}) puisse être muni de probabilités différentes correspondant à des “hasards différents” est illustré par le *paradoxe de Bertrand*, que vous trouverez en ligne par exemple sur Wikipédia.

12.4 Probabilités conditionnelles. Indépendance d'événements.

12.4.1 Définition des probabilités conditionnelles.

La notion de probabilité conditionnelle apparaît naturellement lorsqu'on possède une information partielle sur le résultat d'une expérience aléatoire.

Exemple : À la Foire du Trône, un jeu de loterie est constitué d'une urne contenant 80 billets perdants et 20 billets gagnants, dont 2 seulement donnent droit au gros lot.

Je tire un billet « au hasard » (cette expression sous-entend que l'on est en situation d'équiprobabilité) ; avec quelle probabilité vais-je gagner le gros lot ? Nous pouvons représenter cette expérience aléatoire par le modèle suivant :

- $\Omega = \{b_1, \dots, b_{100}\}$ où b_1 et b_2 représentent les billets qui donnent droit au gros lot, b_3, \dots, b_{20} représentent des billets gagnants ordinaires et b_{21}, \dots, b_{100} des billets perdants ;
- $\mathcal{A} = \mathcal{P}(\Omega)$;
- la probabilité P est uniforme : $\forall i = 1, 2, \dots, 100 \quad P(\{b_i\}) = 1/100$.

L'événement qui nous intéresse est $A = \{b_1, b_2\}$ et nous obtenons donc immédiatement : $P(A) = 2/100$.

Imaginons maintenant que juste après mon tirage et avant que je ne découvre mon billet, le forain — qui a un signe de reconnaissance secret — m'affirme : « Vous avez tiré un billet gagnant ! ». Sachant cela, je ne vais plus calculer ma probabilité de la même façon.

Muni de ce renseignement, je sais que le résultat de l'expérience appartient à l'ensemble $E = \{b_1, \dots, b_{20}\}$. Comme il y a symétrie entre tous les billets, ils sont encore équiprobables, d'où le nouveau calcul de ma probabilité :

$$\frac{\text{nombre de cas favorables}}{\text{nombre de cas possibles}} = \frac{2}{20} = \frac{1}{10}.$$

Nous constatons que la valeur de la probabilité n'est plus la même car, disposant d'une information supplémentaire, nous avons en fait modifié le calcul comme suit :

$$P'(A) = \frac{\#(A \cap E)}{\#E} = \frac{P(A \cap E)}{P(E)}.$$

P' est appelée probabilité conditionnelle sachant E (i.e. sachant mon information partielle). De façon plus générale, nous avons la proposition suivante :

Proposition 12.4.1 Soit (Ω, \mathcal{A}, P) un espace de probabilité et $E \in \mathcal{A}$ un événement tel que $P(E) > 0$. L'application $P^E : \mathcal{A} \rightarrow [0, 1]$ définie par :

$$\forall A \in \mathcal{A}, \quad P^E(A) = \frac{P(A \cap E)}{P(E)},$$

est une probabilité. On l'appelle probabilité conditionnelle à l'événement E ou probabilité sachant E . On note sa valeur sur un événement A comme suit :

$$P(A|E) = P^E(A) = \frac{P(A \cap E)}{P(E)}.$$

Démonstration : C'est bien une application de \mathcal{A} dans $[0, 1]$ puisqu'on a :

$$A \cap E \subset E \Rightarrow 0 \leq P(A \cap E) \leq P(E).$$

De plus, on vérifie immédiatement : $P^E(\emptyset) = 0$ et $P^E(\Omega) = 1$.

Il nous reste à montrer que l'application P^E est σ -additive, ce qui résulte du calcul suivant, valable pour toute suite $(A_n) \in \mathcal{A}^{\mathbb{N}}$ constituée d'événements 2 à 2 disjoints :

$$\frac{P\left(\left(\biguplus_n A_n\right) \cap E\right)}{P(E)} = \frac{P\left(\biguplus_n (A_n \cap E)\right)}{P(E)} = \frac{\sum_n P(A_n \cap E)}{P(E)} = \sum_n \frac{P(A_n \cap E)}{P(E)} \quad \square$$

Remarque : Dans cette définition de la probabilité conditionnelle, l'idée est que, étant donnée l'information partielle E , l'événement intéressant n'est plus l'événement A de départ mais plutôt $A \cap E$. L'application $A \mapsto P(A \cap E)$ est bien positive et σ -additive mais n'a pas la bonne masse totale (i.e. la valeur 1 sur Ω). Pour en faire une probabilité, il suffit de la renormaliser en divisant par $P(E)$ mais ceci n'est possible que si $P(E) > 0$.

Exemple : On change les règles de la loterie : il y a toujours 20 billets gagnants sur 100 mais pour obtenir le gros lot, il faut d'abord tirer un billet gagnant puis il faut faire tourner une roue divisée en 5 parties égales dont une seule donne droit au gros lot. Quelle est la probabilité d'obtenir celui-ci ? Écrire le modèle serait un peu plus long : $\Omega = \{b_1, \dots, b_{100}\} \times \{r_1, \dots, r_5\}$ etc. Pour gagner du temps, passons directement au calcul, en écrivant les événements informellement :

$$A = \{\text{le joueur gagne le gros lot}\}; E = \{\text{le joueur tire un billet gagnant}\}.$$

Nous pouvons alors mener le calcul de la façon suivante :

$$P(A) = P(A \cap E) = P(A|E)P(E) = \frac{1}{5} \times \frac{20}{100} = \frac{4}{100}.$$

Nous venons d'utiliser la formule $P(A \cap E) = P(A|E)P(E)$; plus généralement nous avons la proposition suivante.

Proposition 12.4.2 (Formule des probabilités composées) *Considérons $(A_1, \dots, A_n) \in \mathcal{A}^n$ avec $P(A_1 \cap \dots \cap A_{n-1}) > 0$. Alors, on a l'égalité :*

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

Démonstration: Le membre de droite s'écrit par définition d'une probabilité conditionnelle :

$$P(A_1) \times \frac{P(A_1 \cap A_2)}{P(A_1)} \times \frac{P(A_1 \cap A_2 \cap A_3)}{P(A_1 \cap A_2)} \times \cdots \times \frac{P(A_1 \cap \cdots \cap A_n)}{P(A_1 \cap \cdots \cap A_{n-1})},$$

et l'on conclut par simplification. \square

12.4.2 Formule de Bayes.

Proposition 12.4.3 (Formule des probabilités totales) *Si E_1, \dots, E_n sont des événements tels que $P(E_i) > 0$ pour tout $1 \leq i \leq n$ et formant une partition de l'ensemble Ω (i.e. $\biguplus_{1 \leq i \leq n} E_i = \Omega$), alors on a pour tout événement $A \in \mathcal{A}$ l'égalité suivante :*

$$P(A) = \sum_{i=1}^n P(A|E_i)P(E_i).$$

Démonstration: Ceci résulte des égalités successives :

$$\begin{aligned} P(A) &= P\left(A \cap \left(\biguplus_{1 \leq i \leq n} E_i\right)\right) = P\left(\biguplus_{1 \leq i \leq n} (A \cap E_i)\right) \\ &= \sum_{1 \leq i \leq n} P(A \cap E_i) = \sum_{1 \leq i \leq n} P(A|E_i)P(E_i) \end{aligned}$$

\square

Notons en particulier que la formule des probabilités totales se réécrit dans le cas $n = 2$:

Soient deux événements E et A avec $0 < P(E) < 1$. Alors, on a l'égalité :

$$P(A) = P(A|E)P(E) + P(A|E^c)P(E^c)$$

Nous déduisons de la formule des probabilités totales le théorème suivant :

Théorème 12.4.4 (Formule de Bayes) *Si E_1, \dots, E_n sont des événements de probabilités strictement positives, formant une partition de l'ensemble Ω et si A est un événement tel que $P(A) > 0$, alors on a l'égalité :*

$$P(E_1|A) = \frac{P(A|E_1)P(E_1)}{\sum_{i=1}^n P(A|E_i)P(E_i)}.$$

Démonstration : Il suffit d'écrire les égalités :

$$P(E_1|A) = \frac{P(A \cap E_1)}{P(A)} = \frac{P(A|E_1)P(E_1)}{P(A)}$$

puis d'appliquer la formule des probabilités totales au dénominateur. \square

Remarque : Revenons au cas particulier $n = 2$. La formule de Bayes s'écrit alors :

$$P(E|A) = \frac{P(A|E)P(E)}{P(A|E)P(E) + P(A|E^c)P(E^c)}$$

Notons qu'il n'est donc pas possible de calculer $P(E|A)$ à partir de la seule donnée de $P(A|E)$; il faut connaître d'autres quantités pour faire ce calcul.

Exercice : Une maladie rare touche 1 personne sur 10000 dans la population française. Quand cette maladie est présente, un test sanguin permet de la détecter dans 99% des cas. En revanche, ce test produit des faux positifs dans 1 cas sur 1000.

Le test d'une personnes est positif. Quelle est la probabilité qu'elle soit vraiment atteinte de la maladie? Que pensez-vous de la qualité de ce test sanguin?

Solution : Définissons les événements :

$$E = \{\text{la personne est atteinte de la maladie}\}; A = \{\text{le test est positif}\}$$

L'énoncé nous fournit les données suivantes :

$$P(E) = 10^{-4}, P(A|E) = 0,99, P(A|E^c) = 10^{-3}$$

La formule de Bayes nous permet alors de calculer :

$$P(E|A) = \frac{0,99 \times 10^{-4}}{0,99 \times 10^{-4} + 10^{-3} \times 0,9999} \sim \frac{1}{11}$$

Conclusion : À moins de vouloir provoquer beaucoup de panique inutile, le test est bon pour la poubelle!

12.4.3 Indépendance d'événements.

Définissons les événements $A = \{\text{au moins un ascenseur du bâtiment dans lequel nous nous trouvons sera en panne demain}\}$ et $B = \{\text{l'indice CAC 40 va s'effondrer demain}\}$. Intuitivement, la réalisation (ou non-réalisation) de l'événement A n'a aucune influence sur les chances que l'événement B se produise, ce qui s'écrit $P(A|B) = P(A)$ ou encore $P(A \cap B) = P(A)P(B)$. De tels événements seront dits *indépendants*.

Définition 12.4.5 Deux événements A et B sont dits indépendants s'ils vérifient l'égalité :

$$P(A \cap B) = P(A)P(B)$$

La notion d'indépendance est essentielle en calcul des probabilités ; elle fera souvent partie de nos hypothèses en modélisation. Par exemple, pour étudier la durée de vie d'un circuit électronique constitué de plusieurs composants élémentaires, nous supposerons souvent que ces différents composants tombent en panne indépendamment les uns des autres.

Remarque. Il est facile de vérifier que si le couple (A, B) est constitué d'événements indépendants, alors il en est de même pour les couples suivants :

$$(A, B^c), (A^c, B), (A^c, B^c).$$

Exemple. On tire une carte au hasard dans un jeu de 52 et l'on considère les événements $A = \{\text{c'est un roi}\}$ et $B = \{\text{c'est un coeur}\}$. Nous calculons facilement :

$$P(A \cap B) = \frac{1}{52}, P(A) = \frac{4}{52} = \frac{1}{13}, P(B) = \frac{13}{52} = \frac{1}{4}.$$

Nous en déduisons que les événements A et B sont indépendants, conformément à la définition précédente.

Plus généralement, dans un tel tirage, la figure et la couleur de la carte sont indépendantes.

Définition 12.4.6 Soient $(A_1, \dots, A_n) \in \mathcal{A}^n$ des événements.

Nous dirons que ces événements sont 2 à 2 indépendants si pour tout $(i, j) \in \{1, \dots, n\}$ avec $i \neq j$, on a $P(A_i \cap A_j) = P(A_i)P(A_j)$.

Nous dirons que ces événements sont mutuellement indépendants (ou encore indépendants dans leur ensemble) si pour toute sous-famille d'indices $\{i_1, \dots, i_k\} \subset \{1, \dots, n\}$, on a :

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \times P(A_{i_2}) \times \dots \times P(A_{i_k}).$$

Remarque. Des événements mutuellement indépendants sont clairement indépendants 2 à 2 mais la réciproque est fautive, comme nous allons le voir dans le contre-exemple ci-dessous. Autrement dit, la propriété de mutuelle indépendance est plus forte que la propriété d'indépendance 2 à 2.

Lorsqu'un énoncé dit que des événements A_1, \dots, A_n sont indépendants sans autre précision, c'est toujours l'indépendance mutuelle, c'est-à-dire la plus forte, qu'il faut comprendre.

Contre-exemple: On jette une pièce équilibrée 2 fois de suite et l'on considère les événements : $A = \{\text{on obtient pile au premier jet}\}$, $B = \{\text{on obtient pile au second jet}\}$, $C = \{\text{les résultats des deux jets sont différents}\}$. Il est facile de calculer :

$$P(A) = P(B) = P(C) = 1/2$$

et

$$P(A \cap B) = P(A \cap C) = P(B \cap C) = 1/4.$$

Nous en déduisons que les événements A , B et C sont indépendants 2 à 2. En revanche, ces événements ne sont pas mutuellement indépendants puisque :

$$P(A \cap B \cap C) = P(\emptyset) = 0 \neq P(A)P(B)P(C).$$

□

Proposition 12.4.7 Des événements $(A_1, \dots, A_n) \in \mathcal{A}^n$ sont mutuellement indépendants si et seulement si on a

$$P\left(\bigcap_{i=1}^n C_i\right) = \prod_{i=1}^n P(C_i),$$

pour tous $C_i \in \{\Omega, A_i, A_i^c, \emptyset\}$, $1 \leq i \leq n$.

Le sens direct est établi dans [OUV¹ 56-57]. Le sens réciproque est laissé à titre d'exercice.

Nous allons maintenant mettre en évidence une situation générale dans laquelle la *probabilité binomiale* apparaît naturellement.

Proposition 12.4.8 Soient $A_i, 1 \leq i \leq n$, des événements indépendants ayant tous la même probabilité $p \in [0, 1]$. Pour tout $0 \leq k \leq n$, la probabilité que k exactement d'entre eux soient réalisés est égale à :

$$B_{n,p}(\{k\}) = \binom{n}{k} p^k (1-p)^{n-k}$$

Démonstration: Notons Θ l'ensemble des parties de $\{1, \dots, n\}$ qui sont de cardinal k , de sorte que $\#\Theta = \binom{n}{k}$.

L'événement dont nous voulons calculer la probabilité est le suivant :

$$\bigcup_{\theta \in \Theta} \left[\left(\bigcap_{i \in \theta} A_i \right) \cap \left(\bigcap_{i \notin \theta} A_i^c \right) \right]$$

Un instant de réflexion montre qu'il s'agit d'une union d'événements deux à deux disjoints, si bien que la probabilité cherchée s'écrit :

$$\sum_{\theta \in \Theta} P \left[\left(\bigcap_{i \in \theta} A_i \right) \cap \left(\bigcap_{i \notin \theta} A_i^c \right) \right]$$

À cause de l'hypothèse d'indépendance des $A_i, 1 \leq i \leq n$, cette probabilité s'écrit encore :

$$\sum_{\theta \in \Theta} p^k (1-p)^{n-k} = \binom{n}{k} p^k (1-p)^{n-k}$$

□

Après avoir défini l'indépendance d'une famille finie d'événements, nous terminons ce paragraphe en passant au cas d'une famille infinie dénombrable.

Définition 12.4.9 Une suite d'événements $(A_n)_{n \in \mathbb{N}}$ est dite indépendante si pour toute sous-famille finie d'indices $\{i_1, \dots, i_k\} \subset \mathbb{N}$ (avec $k \in \mathbb{N}^*$), on a l'égalité :

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \times P(A_{i_2}) \times \dots \times P(A_{i_k}).$$

12.5 Exercices.

12.5.1 Clans.

Indicatrices

Soit A un clan sur Ω ; on définit la fonction $\mathbf{1}_A : \Omega \rightarrow \{0, 1\}$ par :

$\mathbf{1}_A(\omega) = 0$ si $\omega \notin A$, $\mathbf{1}_A(\omega) = 1$ si $\omega \in A$.

Montrer les égalités suivantes : $\mathbf{1}_{A \cap B} = \inf(\mathbf{1}_A, \mathbf{1}_B) = \mathbf{1}_A \mathbf{1}_B$,

$\mathbf{1}_{A \cup B} = \sup(\mathbf{1}_A, \mathbf{1}_B) = \mathbf{1}_A + \mathbf{1}_B - \mathbf{1}_A \mathbf{1}_B$,

$\mathbf{1}_{A^c} = 1 - \mathbf{1}_A$,

ainsi que l'équivalence : $\forall \omega \in \Omega, \mathbf{1}_A(\omega) \leq \mathbf{1}_B(\omega) \Leftrightarrow A \subset B$.

On rappelle que la différence symétrique est définie par :

$A \Delta B = (A \cup B) \setminus (A \cap B) = (A^c \cap B) \cup (A \cap B^c)$.

Montrer que :

$\mathbf{1}_{A \Delta B} = |\mathbf{1}_A - \mathbf{1}_B| = \mathbf{1}_A + \mathbf{1}_B \pmod{2} = \mathbf{1}_A + \mathbf{1}_B - 2 \mathbf{1}_A \mathbf{1}_B$.

En déduire que la différence symétrique est commutative, associative et que l'intersection est distributive par rapport à la différence symétrique.

Clan engendré par une partition.

Soit (P_1, \dots, P_n) une partition de Ω . Décrire le clan \mathcal{C} qu'elle engendre.

Réunions d'algèbres.

Montrer que la réunion d'une suite croissante d'algèbres est une algèbre. Est-ce que la réunion d'une suite quelconque d'algèbres est une algèbre ?

***Description de l'algèbre engendrée.**

Soit \mathcal{E} une famille de parties d'un ensemble Ω . Montrer que l'algèbre engendrée par \mathcal{E} est égale à l'ensemble des réunions finies d'intersections finies d'éléments de \mathcal{E} ou de complémentaires d'éléments de \mathcal{E} (parties de la forme $\bigcup_{i=1}^n \bigcap_{j=1}^{m_i} E_{i,j}$, où $\forall i, j$ $E_{i,j} \in \mathcal{E}$ ou $(E_{i,j})^c \in \mathcal{E}$).

Remarque : Il est impossible de décrire de façon analogue la σ -algèbre engendrée par \mathcal{E} .

Qu'est-ce qui ne se généralise pas dans la démonstration précédente ?

12.5.2 Formule de Poincaré.

[FOA 17]

Démonstration par récurrence.

Soit \mathcal{A} une tribu sur Ω et P une probabilité sur \mathcal{A} . On se donne une famille finie (A_1, \dots, A_n) d'éléments de \mathcal{A} .

Etablir par récurrence la formule suivante, dite de Poincaré :

$$P(\cup_{i=1}^n A_i) = \sum_{1 \leq i \leq n} P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i \cap A_j) + \dots + (-1)^{n-1} P(A_1 \cap \dots \cap A_n).$$

Application : Indicatrice d'Euler

[FOA 37-38]

On choisit au hasard un nombre parmi $\{1, \dots, n\}$, tous les choix étant équiprobables. Soit p un entier inférieur à n . On note A_p l'événement "le nombre choisi est divisible par p ".

— Calculer $P(A_p)$ lorsque p divise n .

- Soit $n = p_1^{\alpha_1} \cdots p_r^{\alpha_r}$ la décomposition en facteurs premiers de l'entier n . Que représente l'événement $(A_{p_1} \cup \cdots \cup A_{p_r})^c$?
- On note $\phi(n)$ le nombre d'entiers strictement inférieurs à n et premiers avec n . Montrer que l'on a

$$\phi(n) = n \prod_{p \text{ premier, } p|n} \left(1 - \frac{1}{p}\right).$$

Application : Facteur distrait.

[FOA 38-39]

Un facteur dispose de n lettres adressées à n destinataires distincts. Il fait la distribution au hasard entre ces n personnes. Quelle est la probabilité de l'événement "une lettre au moins parvient à son destinataire" ? En donner une approximation pour n assez grand.

12.5.3 Équiprobabilité.

Trois dés.

[COT 5-6]

On jette trois dés à 6 faces bien équilibrés.

- Calculer la probabilité d'obtenir au moins un as.
- Que vaut la probabilité d'obtenir au moins deux faces portant le même chiffre ?
- Calculer la probabilité que la somme des points marqués sur les trois faces soit paire.

Happy birthday to you !

[FOA 31]

Quelle est la probabilité que dans votre promotion deux agrégatifs au moins aient leur anniversaire le même jour ?

Remarque : Pour simplifier, on ne s'occupera pas du 29 Février.

Boules et cases.

[DAC¹ exo page 7]

Soit Ω l'ensemble des configurations que l'on peut obtenir en répartissant r boules indiscernables dans n cases numérotées.

1. Montrer que le cardinal de Ω est égal au nombre de solutions $(r_1, \dots, r_n) \in \mathbb{N}^n$ de l'équation : $r_1 + r_2 + \dots + r_n = r$.
2. Montrer que ce nombre vaut C_{n+r-1}^r de 3 façons différentes :
 - (a) par récurrence sur n à l'aide de l'identité : $C_n^p = C_{n-1}^{p-1} + C_{n-2}^{p-1} + \dots + C_{p-1}^{p-1}$.
 - (b) en développant en série entière les deux membres de l'égalité : $\frac{1}{(1-t)^n} = \left(\sum_{k=0}^{\infty} t^k \right)^n$.
 - (c) en identifiant Ω à l'ensemble des façons d'ordonner en ligne r boules indiscernables et $n - 1$ séparations (indiscernables aussi).
3. On répartit au hasard r boules indiscernables dans n cases. Calculer la probabilité qu'aucune case ne soit vide.

12.5.4 Probabilités conditionnelles.

Maladie rare.

[COT 11]

On considère une certaine maladie qui touche 1/1000e de la population. Un laboratoire d'analyse de sang assure avec une fiabilité de 99% la détection de cette maladie lorsqu'elle est effectivement présente. Cependant, le test indique aussi un résultat faussement positif pour 0,2% des personnes réellement saines à qui on l'applique.

Quelle est la probabilité qu'une personne soit vraiment malade sachant que son test est positif? Commenter le résultat.

Le rouge et le noir.

[ROS 66-67]

On considère 3 cartes à jouer de même forme mais de couleurs différentes : la première est noire des deux côtés, la seconde rouge des deux côtés, tandis que la troisième a une face noire et une face rouge.

On mélange les trois cartes au fond d'un chapeau puis on en tire une carte au hasard, dont on ne montre qu'une face. Sachant que cette face est rouge, quelle est la probabilité que l'autre face soit noire ?

Loi de succession de Laplace.

[COT 15-16]

On dispose de $N + 1$ urnes, numérotées de 0 à N . L'urne numéro k contient k boules rouges et $N - k$ boules blanches. On choisit une urne au hasard.

Sans connaître son numéro, on en tire n fois de suite une boule, avec remise après chaque tirage. Quelle est la probabilité que le $(n+1)^e$ tirage donne encore une boule rouge sachant que, au cours des n premiers tirages, seules des boules rouges ont été tirées? Calculer la limite de cette probabilité lorsque $N \rightarrow \infty$.

Taux de panne

[COT 16-18]

Soit T une variable aléatoire prenant ses valeurs dans \mathbb{N} telle que pour tout $n \in \mathbb{N}$, $P(T \geq n) > 0$. On appelle taux de panne la suite $\theta(n) = P(T = n | T \geq n)$, $n \in \mathbb{N}$.

1. Calculer $P(T = n)$ en fonction des $\theta(k)$, $k \leq n$.
2. Etablir qu'une suite de réels $\theta(k)$ convient comme taux de panne **ssi** :

$$\forall k \in \mathbb{N} \quad 0 \leq \theta(k) < 1 \quad \text{et} \quad \sum_{k=0}^{\infty} \theta(k) = \infty$$

3. Montrer que T suit une loi géométrique **ssi** $\theta(k) = \text{constante}$ pour tout $k \in \mathbb{N}$. (On dit que T suit une loi géométrique de paramètre $a \in]0, 1[$ si pour tout $k \in \mathbb{N}$, $P(T = k) = a(1-a)^k$.)

12.5.5 Événements indépendants.

Événement auto-indépendant.

Montrer qu'un événement A est indépendant de lui-même **ssi** $P(A) = 0$ ou $P(A) = 1$.

Bruit qui court.

[COT 13]

La personne I_1 reçoit l'information 0 ou 1 et la transmet telle quelle à I_2 avec la probabilité p , I_2 de même à I_3 , etc... I_n la transmet au monde entier. On suppose que les n personnes I_1, \dots, I_n sont indépendantes. Quelle est la probabilité p_n que le monde reçoive la bonne information? Calculer $\lim_{n \rightarrow \infty} p_n$.

Indicatrice d'Euler (bis).

On choisit au hasard un nombre parmi $\{1, \dots, n\}$, tous les choix étant équiprobables. Soit p un entier inférieur à n . On note A_p l'événement "le nombre choisi est divisible par p ".

1. Calculer $P(A_p)$ lorsque p divise n .
2. On suppose que (p_1, \dots, p_k) sont des diviseurs premiers distincts de n , montrer que les événements A_{p_1}, \dots, A_{p_k} sont indépendants.
3. On note $\phi(n)$ le nombre d'entiers strictement inférieurs à n et premiers avec n . Montrer que l'on a

$$\phi(n) = n \prod_{p \text{ premier, } p|n} \left(1 - \frac{1}{p}\right).$$

Corrigé :

1. Puisque p divise n (notation : $p|n$), il existe un entier m tel que $n = mp$; on a alors : $A_p = \{p, 2p, \dots, mp\}$. On en déduit : $P(A_p) = \frac{m}{n} = \frac{1}{p}$.
2. Considérant une sous-famille $A_{p_{i_1}}, \dots, A_{p_{i_l}}$, nous remarquons que

$$\bigcap_{j=1}^l A_{p_{i_j}} = A_{\prod_{j=1}^l p_{i_j}},$$

puisque les p_{i_j} sont premiers. Pour la même raison, $(\prod_{j=1}^l p_{i_j})|n$ donc on peut appliquer a. pour obtenir :

$$P(\bigcap_{j=1}^l A_{p_{i_j}}) = \frac{1}{\prod_{j=1}^l p_{i_j}} = \prod_{j=1}^l P(A_{p_{i_j}}),$$

ce qui prouve l'indépendance demandée.

3. Prenons ici p_1, \dots, p_k **tous** les diviseurs premiers de n . D'après b), on a :

$$P(\bigcap_{i=1}^k A_{p_i}^c) = \prod_{p \text{ premier, } p|n} \left(1 - \frac{1}{p}\right).$$

Or un entier appartient à $\bigcap_{i=1}^k A_{p_i}^c$ **ssi** il n'est divisible par aucun des diviseurs premiers de n , autrement dit **ssi** il est premier avec n . En passant aux probabilités, on en déduit :

$$\prod_{p \text{ premier, } p|n} \left(1 - \frac{1}{p}\right) = \frac{\phi(n)}{n}.$$

Principe de Hardy-Weinberg.

[MEY 43-44]

Les gènes se présentent le plus souvent en paires et sous deux formes alléliques A,a, ce qui donne trois génotypes AA,Aa,aa. Ces génotypes s'expriment par différents phénotypes ; par exemple, dans le cas de l'albinisme, les individus aa sont albinos, contrairement aux individus AA ou Aa.

Chaque individu reçoit au hasard un gène de chacun de ses parents et donc chaque allèle composant le gène d'un des parents a la probabilité 1/2 de passer à l'enfant. Les génotypes des parents sont indépendants. On note p, q, r les probabilités qu'un adulte, dans une population donnée, ait les génotypes AA,Aa,aa.

1. Calculer les probabilités P, Q, R qu'un enfant ait les génotypes AA,Aa,aa (on pourra poser $\theta = p + q/2$).
2. Montrer que $Q^2 = 4PR$.
3. A quelle condition sur p, q, r a-t-on $P = p, Q = q, R = r$? Montrer qu'elle est toujours réalisée dès la 2ème génération.

Corrigé :

1. Nous prenons pour espace Ω l'ensemble des triplets (génotype de l'enfant, génotype du père, génotype de la mère).

L'évènement « L'enfant a un génotype AA » s'écrit donc :

$$\{(AA, AA, AA); (AA, Aa, AA); (AA, AA, Aa); (AA, Aa, Aa)\}$$

et , d'après les lois de la génétique rappelées dans l'énoncé, admet pour probabilité :

$$P = p^2 + pq/2 + pq/2 + q^2/4 = \theta^2.$$

Un raisonnement symétrique (en échangeant les lettres A et a) nous donne :

$$R = r^2 + rq/2 + rq/2 + q^2/4 = (1 - \theta)^2.$$

Enfin, puisque $P + Q + R = 1$, on obtient : $Q = 2\theta(1 - \theta)$.

2. Cette relation résulte immédiatement de ce qui précède.
3. D'après la question précédente, on a nécessairement $q^2 = 4pr$.

Réciproquement, si cette condition est vérifiée, les égalités obtenues dans la première question nous donnent :

$$P = p^2 + pq + q^2/4 = p^2 + pq + pr = p(p + q + r) = p$$

$$\text{et } R = r^2 + rq + q^2/4 = r^2 + rq + rp = r(p + q + r) = r.$$

Il en résulte $Q = q$ puisqu'on a : $P + Q + R = p + q + r = 1$.

Finalement, l'égalité $q^2 = 4pr$ est une condition nécessaire et suffisante de "stationnarité" des probabilités des différents génotypes au cours des générations.

Ainsi, le résultat du 2) nous dit que dans tous les cas, la stationnarité s'établit dès la deuxième génération.

Chapitre 13

Variables aléatoires réelles

13.1 La tribu borélienne réelle

Définition 13.1.1 Si (E, d) est un espace métrique, on appelle tribu borélienne sur E et l'on note $\mathcal{B}(E)$ la tribu engendrée par l'ensemble des ouverts de E . Les éléments de $\mathcal{B}(E)$ sont appelés ensembles boréliens de E .

Remarque. Nous constatons immédiatement que $\mathcal{B}(E)$ est également la tribu engendrée par l'ensemble des fermés de E . \square

Si l'espace métrique sous-jacent est égal à \mathbb{R} muni de la distance usuelle, nous parlerons de tribu borélienne réelle et nous la noterons donc $\mathcal{B}(\mathbb{R})$.

On ne sait pas décrire explicitement tous les boréliens réels mais il sera suffisant pour la suite de ce cours de savoir que $\mathcal{B}(\mathbb{R})$ contient tous les intervalles réels (fermés, ouverts, semi-ouverts) et leurs unions finies ou dénombrables. On peut construire un sous-ensemble réel qui n'est pas borélien mais ce n'est pas si facile... On peut également démontrer que $\#\mathcal{B}(\mathbb{R}) = \#\mathbb{R}$.

Proposition 13.1.2 La tribu $\mathcal{B}(\mathbb{R})$ est engendrée par chacun des ensembles suivants de parties réelles :

- $\mathcal{E}_1 = \{]a, b[, \text{ avec } (a, b) \in \mathbb{Q}^2 \text{ et } a < b\}$
- $\mathcal{E}_2 = \{]a, b] , \text{ avec } (a, b) \in \mathbb{Q}^2 \text{ et } a < b\}$
- $\mathcal{E}_3 = \{[a, b] , \text{ avec } (a, b) \in \mathbb{Q}^2 \text{ et } a \leq b\}$
- $\mathcal{E}_4 = \{] - \infty, a], \quad a \in \mathbb{Q}\}$
- $\mathcal{E}_5 = \{] - \infty, a[, \quad a \in \mathbb{Q}\}$

Démonstration:

1. Puisque tous les éléments de \mathcal{E}_1 sont des ouverts, on a $\sigma(\mathcal{E}_1) \subset \mathcal{B}(\mathbb{R})$. Prouvons l'inclusion inverse.

Tout ouvert de \mathbb{R} est une union de boules ouvertes, c'est-à-dire d'intervalles de la forme $]a, b[$, avec $(a, b) \in \mathbb{R}^2$ et $a < b$.

Par densité de \mathbb{Q} dans \mathbb{R} , nous pouvons trouver une suite strictement décroissante $(a_n) \in \mathbb{Q}^{\mathbb{N}}$ qui converge vers a et une suite strictement croissante $(b_n) \in \mathbb{Q}^{\mathbb{N}}$ qui converge vers b , de sorte que :

$$]a, b[= \bigcup_{n \in \mathbb{N}}]a_n, b_n[.$$

Par conséquent, tout ouvert de \mathbb{R} est une union d'intervalles ouverts de la forme $]a, b[$, avec $(a, b) \in \mathbb{Q}^2$ et $a < b$. L'ensemble des intervalles ouverts à extrémités rationnelles étant dénombrable, nous avons prouvé que tout ouvert de \mathbb{R} s'écrivait comme une union dénombrable d'éléments de \mathcal{E}_1 , d'où $\mathcal{B}(\mathbb{R}) \subset \sigma(\mathcal{E}_1)$.

2. On montre que $\sigma(\mathcal{E}_2) = \sigma(\mathcal{E}_1)$ en utilisant des égalités de la forme :

$$]a, b] = \bigcap_{n \in \mathbb{N}^*}]a, b + \frac{1}{n}[\quad ; \quad]a, b[= \bigcup_{n \in \mathbb{N}^*}]a, b - \frac{1}{n}]$$

3. On montre que $\sigma(\mathcal{E}_3) = \sigma(\mathcal{E}_2)$ en utilisant des égalités de la forme :

$$[a, b] = \bigcap_{n \in \mathbb{N}^*}]a - \frac{1}{n}, b] \quad ; \quad]a, b] = \bigcup_{n \in \mathbb{N}^*} [a - \frac{1}{n}, b]$$

4. On montre que $\sigma(\mathcal{E}_4) = \sigma(\mathcal{E}_2)$ en utilisant des égalités de la forme :

$$]a, b] =] - \infty, b] -] - \infty, a] \quad ; \quad] - \infty, a] = \bigcup_{n \in \mathbb{N}^*}] - n, a]$$

5. On montre que $\sigma(\mathcal{E}_5) = \sigma(\mathcal{E}_4)$ en utilisant des égalités de la forme :

$$] - \infty, a] = \bigcap_{n \in \mathbb{N}^*}] - \infty, a + \frac{1}{n}[\quad ; \quad] - \infty, a[= \bigcup_{n \in \mathbb{N}^*}] - \infty, a - \frac{1}{n}].$$

□

Remarque. Par passage aux complémentaires, nous constatons immédiatement que la tribu $\mathcal{B}(\mathbb{R})$ est aussi engendrée par les demi-droites de la forme $[a, +\infty[$ (respectivement $]a, +\infty[$), avec $a \in \mathbb{Q}$.

13.2 Les variables aléatoires réelles et leurs lois

Dans toute la suite de ce chapitre, nous considérons un espace probabilisé (Ω, \mathcal{A}, P) .

Dans de très nombreuses situations aléatoires, nous allons nous intéresser à une fonction du résultat ω de l'expérience plutôt qu'à ce résultat lui-même.

Exemple : Un joueur lance une fléchette en visant le centre d'une cible. Si nous prenons pour unité le rayon de la cible, nous pouvons modéliser ceci comme une expérience aléatoire dont le résultat est un élément ω du disque unité fermé :

$$\Omega = \bar{D}(0, 1) = \{\omega \in \mathbb{R}^2, \omega_1^2 + \omega_2^2 \leq 1\}.$$

En réalité, la seule chose qui nous intéresse vraiment n'est pas la position exacte $\omega = (\omega_1, \omega_2)$ de l'impact de la fléchette sur la cible mais seulement sa distance au centre de la cible, qui nous est donnée par la variable aléatoire réelle :

$$\begin{aligned} X : \quad \Omega &\longrightarrow \mathbb{R} \\ \omega = (\omega_1, \omega_2) &\longmapsto \sqrt{\omega_1^2 + \omega_2^2} \end{aligned}$$

Pour mesurer la qualité de la performance du joueur, on va se poser des questions telles que "est-ce que $X(\omega) \in B$?", où B est un sous-ensemble de $[0, 1]$. Par exemple, on peut décider que le joueur a gagné si $X(\omega) \in [0, \frac{1}{10}]$.

Nous allons donc poser la définition suivante.

Définition 13.2.1 *On appelle variable aléatoire réelle (en abrégé v.a.r.) définie sur (Ω, \mathcal{A}, P) toute application $X : \Omega \rightarrow \mathbb{R}$ vérifiant :*

$$\forall B \in \mathcal{B}(\mathbb{R}) \quad X^{-1}(B) \in \mathcal{A}. \quad (13.1)$$

Rappelons que $X^{-1}(B)$ est la partie de Ω définie par $\{\omega \in \Omega, X(\omega) \in B\}$ et que nous la notons encore sous la forme abrégée $\{X \in B\}$.

La définition précédente nous assure que, pour tout borélien réel B , cela a un sens de calculer la probabilité $P(\{X \in B\})$ – que l'on note plus simplement $P(X \in B)$ – puisque $\{X \in B\} \in \mathcal{A}$.

Pour vérifier qu'une application $X : \Omega \rightarrow \mathbb{R}$ est une variable aléatoire réelle, nous serons souvent amenés à utiliser les deux propositions suivantes, que nous admettrons.

Proposition 13.2.2 *Soit \mathcal{E} un ensemble de parties de \mathbb{R} qui engendre la tribu $\mathcal{B}(\mathbb{R})$. Alors une application $X : \Omega \rightarrow \mathbb{R}$ est une variable aléatoire réelle sur (Ω, \mathcal{A}, P) si et seulement si la condition suivante est satisfaite :*

$$\forall B \in \mathcal{E} \quad X^{-1}(B) \in \mathcal{A}.$$

Nous pourrions appliquer cette proposition avec chacune des classes $\mathcal{E}_1, \dots, \mathcal{E}_5$ définies dans la proposition 13.1.2

Par exemple, le lecteur pourra montrer en utilisant la classe \mathcal{E}_4 que, si X et Y sont deux variables aléatoires réelles sur (Ω, \mathcal{A}, P) , alors il en est de même de $\inf(X, Y)$ et $\sup(X, Y)$.

Proposition 13.2.3 *Muni des lois $+$, \times , \cdot , l'ensemble des variables aléatoires réelles définies sur (Ω, \mathcal{A}, P) est une algèbre sur \mathbb{R} .*

Connaître les quantités de la forme $P(X \in B)$ est très intéressant car cela mesure les chances que la valeur prise par la variable X “tombe” dans tel ou tel sous-ensemble borélien réel (par exemple, typiquement, un intervalle).

Ainsi, si nous revenons à l'exemple précédent, une information telle que $P(X \in [0, \frac{1}{10}]) = 0,95$ nous dit que le joueur est très habile.

Théorème et définition 13.2.4 *Considérons un espace probabilisé (Ω, \mathcal{A}, P) et une variable aléatoire réelle $X : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$.*

Alors l'application $P_X : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$ définie par :

$$\forall B \in \mathcal{B}(\mathbb{R}) \quad P_X(B) = P(X \in B)$$

est une mesure de probabilité sur l'espace mesurable $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, appelée loi de probabilité (ou simplement loi) de la v.a.r. X .

Démonstration: Il est immédiat d'après la définition que P_X est une application de $\mathcal{B}(\mathbb{R})$ dans $[0, 1]$ telle que $P_X(\emptyset) = P(X^{-1}(\emptyset)) = P(\emptyset) = 0$ et $P_X(\mathbb{R}) = P(X^{-1}(\mathbb{R})) = P(\Omega) = 1$.

Il nous reste donc à prouver la σ -additivité de P_X . Pour cela, nous considérons une suite $(B_n)_{n \in \mathbb{N}} \in \mathcal{B}(\mathbb{R})^{\mathbb{N}}$ telle que $i \neq j \Rightarrow B_i \cap B_j = \emptyset$.

Nous allons montrer dans un premier temps que :

$$X^{-1}\left(\bigsqcup_{n \in \mathbb{N}} B_n\right) = \bigsqcup_{n \in \mathbb{N}} X^{-1}(B_n).$$

En effet, on a :

$$X^{-1}\left(\bigcup_{n \in \mathbb{N}} B_n\right) = \{\omega \in \Omega, \exists n \in \mathbb{N}, X(\omega) \in B_n\} = \bigcup_{n \in \mathbb{N}} X^{-1}(B_n)$$

et, pour $i \neq j$,

$$X^{-1}(B_i) \cap X^{-1}(B_j) = X^{-1}(B_i \cap B_j) = X^{-1}(\emptyset) = \emptyset,$$

d'où le résultat voulu.

Nous déduisons de cette première étape les égalités suivantes :

$$\begin{aligned} P_X \left(\bigcup_{n \in \mathbb{N}} B_n \right) &= P \left(X^{-1} \left(\bigcup_{n \in \mathbb{N}} B_n \right) \right) = P \left(\bigcup_{n \in \mathbb{N}} X^{-1}(B_n) \right) \\ &= \sum_{n \in \mathbb{N}} P(X^{-1}(B_n)) = \sum_{n \in \mathbb{N}} P_X(B_n), \end{aligned}$$

d'où la conclusion. □

Remarque. C'est la condition (13.1) définissant une variable aléatoire réelle X sur (Ω, \mathcal{A}, P) qui nous a permis de "transporter" la probabilité P définie sur son ensemble de départ en une nouvelle probabilité P_X sur son ensemble d'arrivée $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. On dit parfois que P_X est la probabilité image de la probabilité P par la variable aléatoire X .

La loi de X nous dit donc avec quelles chances cette v.a.r. prend ses valeurs dans tel ou tel sous-ensemble borélien réel, en particulier dans les intervalles.

Exemple. Soit $a \in \mathbb{R}$ fixé et la variable constante $X \equiv a$. On a donc $P_X(\{a\}) = 1$. C'est la loi la plus simple qui soit : elle est constituée d'une unique masse ponctuelle de poids 1 située au point a . On l'appelle mesure de Dirac au point a et on la note δ_a . C'est la mesure définie sur tout $A \in \mathcal{B}(\mathbb{R})$ par $\delta_a(A) = 1$ si $a \in A$ et $\delta_a(A) = 0$ si $a \notin A$.

Remarque. Il est fréquent d'utiliser de façon interchangeable les termes *mesure de probabilité* et *loi de probabilité* sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

En effet, par définition même, la loi de probabilité d'une v.a.r. est bien une mesure de probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Mais réciproquement, nous pouvons faire apparaître n'importe quelle mesure de probabilité μ sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ comme la loi d'une certaine variable aléatoire que nous allons construire maintenant.

Prenons pour espace probabilisé $(\Omega, \mathcal{A}, P) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$ et définissons X comme étant l'application identité sur $\Omega = \mathbb{R}$. Le lecteur vérifiera aisément que X est une variable aléatoire réelle et que la loi de X est égale à μ .

13.3 Fonction de répartition d'une variable aléatoire réelle

A priori, pour déterminer la loi d'une v.a.r. X , il faut connaître la valeur de $P(X \in B)$ pour tout borélien réel B . En fait, nous allons voir dans ce paragraphe qu'il suffit de connaître $P(X \in B)$ pour des boréliens réels bien particuliers : les demi-droites $B =] - \infty, x]$, $x \in \mathbb{R}$.

Nous ne démontrerons pas complètement ce résultat car nous admettrons le résultat intermédiaire suivant : la loi d'une v.a.r. X est entièrement déterminée par ses valeurs sur tous les intervalles réels.

Définition 13.3.1 On appelle fonction de répartition (en abrégé *f.r.*) de la v.a.r. X l'application $F_X : \mathbb{R} \rightarrow [0, 1]$ définie par :

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

Remarque. On peut très bien parler de fonction de répartition de la loi de X car en réalité F_X ne dépend que de P_X :

$$\forall x \in \mathbb{R} \quad F_X(x) = P(X \in] - \infty, x]) = P_X(] - \infty, x]).$$

Ainsi, une autre façon de présenter cette notion consiste à définir d'abord la fonction de répartition d'une mesure de probabilité μ sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ par :

$$\forall x \in \mathbb{R} \quad F_\mu(x) = \mu(] - \infty, x])$$

puis à dire que la fonction de répartition d'une variable aléatoire réelle est égale par définition à celle de sa loi de probabilité. \square

Le théorème suivant nous dit que la connaissance de la fonction de répartition d'une loi de probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ détermine cette loi de façon unique.

Théorème 13.3.2 Si deux variables aléatoires réelles X et Y ont même fonction de répartition, alors elles ont même loi :

$$F_X = F_Y \implies P_X = P_Y$$

Autrement dit, l'application qui à une loi associe sa fonction de répartition est injective.

Démonstration: Connaître F_X , c'est connaître la valeur de la loi P_X sur les demi-droites $] -\infty, x]$, $x \in \mathbb{R}$. Nous allons montrer que ceci suffit à calculer la valeur de P_X sur n'importe quel intervalle réel I et nous admettrons que la loi P_X est alors parfaitement déterminée sur tous les boréliens réels B . Nous énumérons les différentes formes possibles pour l'intervalle I et donnons à chaque fois une formule de calcul de $P_X(I)$ à partir de la connaissance de F_X .

Si I est une demi-droite de la forme $] -\infty, x]$, c'est un cas évident puisque par définition $P_X(] -\infty, x]) = F_X(x)$.

Si I est une demi-droite ouverte de la forme $] -\infty, x[$, par passage à la limite croissante dans une probabilité, nous pouvons écrire l'égalité :

$$P_X(] -\infty, x[) = \lim_n \uparrow P_X(] -\infty, x - \frac{1}{n}]) = \lim_n \uparrow F_X(x - \frac{1}{n})$$

Par passage au complémentaire dans ces deux premiers cas, nous traitons les demi-droites de la forme $]x, +\infty[$ ou $[x, +\infty[$.

$$P_X(]x, +\infty[) = 1 - P_X(] -\infty, x]) = 1 - F_X(x)$$

$$P_X([x, +\infty[) = 1 - P_X(] -\infty, x]) = 1 - \lim_n \uparrow F_X(x - \frac{1}{n})$$

Nous pouvons maintenant traiter les cas où I est un intervalle borné, en le faisant apparaître comme différence propre de demi-droites. Soient a et b deux réels tels que $a \leq b$.

$$P_X(]a, b]) = P_X(] -\infty, b]) - P_X(] -\infty, a]) = F_X(b) - F_X(a)$$

De la même façon, on trouve :

$$P_X([a, b]) = P_X(] -\infty, b]) - P_X(] -\infty, a]) = F_X(b) - \lim_n \uparrow F_X(a - \frac{1}{n})$$

Nous laissons au lecteur le soin de traiter de façon similaire les cas restants. \square

Comment utiliser ce théorème ? Dans la suite, nous citerons un certain nombre de lois de probabilité "classiques", c'est-à-dire d'un usage fréquent en calcul des probabilités. Pour la plupart d'entre elles, il est possible de calculer explicitement les fonctions de répartition correspondantes. Une façon de déterminer la loi d'une v.a.r. X sera de calculer sa fonction de répartition. Si nous reconnaissons alors la fonction de répartition d'une loi classique, en vertu du théorème précédent, nous pourrions affirmer que la loi de X est égale à cette

loi classique.

Nous terminons cette section par un théorème permettant de caractériser les fonctions de répartition d'une loi de probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ ou encore les fonctions de répartition des variables aléatoires réelles, ce qui revient au même d'après la remarque qui termine la section précédente.

Théorème 13.3.3 *Considérons une application $F : \mathbb{R} \rightarrow \mathbb{R}$. Alors F est la fonction de répartition d'une loi de probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ si et seulement si elle possède simultanément les quatre propriétés suivantes : F est croissante, continue à droite et vérifie*

$$\lim_{t \rightarrow -\infty} F(t) = 0 \text{ et } \lim_{t \rightarrow +\infty} F(t) = 1.$$

Démonstration: Le sens direct sera démontré en cours. Le sens réciproque est admis. \square

La section et le chapitre suivants concernent deux types particulièrement intéressants de variables aléatoires réelles : les variables discrètes et les variables admettant une densité. Notons que ces deux types sont loin de couvrir toutes les possibilités concernant les v.a.r. Ce sont simplement deux cas particuliers importants et eux seuls figurent au programme officiel.

13.4 Variables discrètes

Une variable aléatoire réelle définie sur (Ω, \mathcal{A}, P) est dite discrète si le sous-ensemble réel $X(\Omega)$ est fini ou dénombrable. Autrement dit, une v.a.r. est dite discrète si elle ne prend qu'un nombre fini ou dénombrable de valeurs. C'est un cas particulièrement simple et si nous notons $X(\Omega) = \{x_0, \dots, x_n, \dots\}$ (dans le cas dénombrable), alors la loi de la variable aléatoire X est entièrement déterminée par la donnée de la suite $(p_n)_{n \in \mathbb{N}}$ définie par :

$$\forall n \in \mathbb{N} \quad p_n = P(X = x_n).$$

Cette suite est bien sûr à valeurs positives ou nulles et telle que

$$\sum_{n=0}^{+\infty} p_n = 1$$

et le lecteur pourra faire le rapprochement avec la proposition 12.3.3.

Le lecteur est invité à se référer à [DAN 418-436 (ou 431-452 pour l'édition 2016)] qui traite de façon exhaustive les notions du programme officiel qui sont relatives aux variables aléatoires discrètes, en particulier les lois

discrètes usuelles suivantes : Bernoulli, binomiale, hypergéométrique, géométrique, Poisson.

Nous n'écrivons ici que la définition de l'espérance d'une variable discrète et le *théorème de transfert* dans le cas discret.

Définition 13.4.1 Soit X une variable aléatoire réelle discrète telle que $X(\Omega) = \{x_0, \dots, x_n, \dots\}$. On appelle espérance de la variable X le nombre réel :

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

dès lors que cette série est absolument convergente. On dit alors que la variable X admet une espérance (ou encore qu'elle est intégrable).

Théorème 13.4.2 (de transfert, cas discret) Soit $\psi : \mathbb{R} \rightarrow \mathbb{R}$ une application ; alors la variable aléatoire $\psi(X)$ admet pour espérance le nombre réel

$$E[\psi(X)] = \sum_n \psi(x_n) P(X = x_n)$$

dès que cette série est absolument convergente.

Un outil très utile pour étudier les variables aléatoires à valeurs dans \mathbb{N} est présenté maintenant sous la forme d'un problème corrigé.

Problème : Fonctions génératrices On considère une variable aléatoire X à valeurs dans \mathbb{N} .

1. Montrer que pour tout $s \in [-1, 1]$, la variable aléatoire s^X admet une espérance (on pose $s^0 = 1$ pour tout s) et que l'on a : $E[s^X] = \sum_{n=0}^{\infty} P(X = n) s^n$.
On notera $G_X(s) = E[s^X]$ et l'on appellera G_X la fonction génératrice de X .
2. Montrer que X et Y ont même loi si et seulement si $G_X = G_Y$.
3. Calculer les fonctions génératrices des lois suivantes :
 - (a) Bernoulli de paramètre $p \in [0, 1]$.
 - (b) Géométrique de paramètre $a \in]0, 1[$.
 - (c) Poisson de paramètre λ .
4. Si X et Y sont deux variables indépendantes, calculer G_{X+Y} en fonction de G_X et G_Y . En déduire :
 - (a) la fonction génératrice d'une loi binomiale de paramètres $(n, p) \in \mathbb{N}^* \times [0, 1]$.

- (b) la loi de la somme de deux variables indépendantes suivant des lois de Poisson de paramètres respectifs λ et μ .
5. Montrer que si X est intégrable, alors G_X est dérivable sur $[-1, 1]$, de dérivée : $G'_X(s) = E[Xs^{X-1}]$.
En déduire les espérances des lois introduites dans la question 3.

Corrigé :

1. Pour tout $s \in [-1, 1]$, nous avons l'inégalité $|P(X = n)s^n| \leq P(X = n)$ qui entraîne l'absolue convergence de la série

$$\sum_{n=0}^{\infty} P(X = n)s^n$$

donc s^X est intégrable et admet pour espérance

$$E[s^X] = \sum_{n=0}^{\infty} P(X = n)s^n.$$

2. L'égalité obtenue dans la question précédente nous donne le développement en série entière de G_X sur $[-1, 1]$. En particulier, G_X admet des dérivées à tous les ordres au point $s = 0$ et l'on a, pour tout $n \in \mathbb{N}$:

$$P(X = n) = \frac{G_X^{(n)}(0)}{n!}$$

Il en résulte que $G_X = G_Y \implies P_X = P_Y$; la réciproque est évidente. La fonction génératrice de X caractérise donc la loi de X .

3. (a) $G_X(s) = 1 - p + ps$.
(b) $G_X(s) = \sum_{n=0}^{+\infty} (1 - a)a^n s^n = \frac{1-a}{1-as}$.
(c) $G_X(s) = \sum_{n=0}^{+\infty} e^{-\lambda} \frac{\lambda^n}{n!} s^n = e^{-\lambda(1-s)}$.
4. On a par définition : $G_{X+Y}(s) = E[s^{X+Y}] = E[s^X s^Y]$.

Si X et Y sont indépendantes, on en déduit, en utilisant une propriété des variables discrètes indépendantes énoncée dans [DAN 427] :

$$\forall s \in [-1, 1] \quad G_{X+Y}(s) = E[s^X]E[s^Y] = G_X(s)G_Y(s)$$

On peut également faire une preuve directe sans utiliser cette propriété, en considérant le produit de Cauchy de deux séries entières absolument convergentes [DAN 430].

- (a) Une récurrence immédiate prouve que si X_1, \dots, X_n sont des variables aléatoires à valeurs dans \mathbb{N} indépendantes, alors :

$$\forall s \in [-1, 1] \quad G_{X_1 + \dots + X_n}(s) = G_{X_1}(s) \cdots G_{X_n}(s).$$

En particulier, si X_1, \dots, X_n suivent la loi de Bernoulli de paramètre p , alors $X_1 + \dots + X_n$ suit la loi binomiale de paramètres (n, p) , dont la fonction génératrice vaut donc (en utilisant 3(a)) : $G(s) = (1 - p + ps)^n$.

- (b) Si $X \sim \mathcal{P}(\lambda)$ et $Y \sim \mathcal{P}(\mu)$ sont indépendantes, alors l'égalité obtenue en 3(c) nous permet d'écrire :

$$G_{X+Y}(s) = G_X(s)G_Y(s) = e^{-\lambda(1-s)}e^{-\mu(1-s)} = e^{-(\lambda+\mu)(1-s)}.$$

On reconnaît la fonction génératrice de la loi $\mathcal{P}(\lambda + \mu)$, ce qui d'après la question 2 prouve que $X + Y$ suit la loi de Poisson de paramètre $\lambda + \mu$.

5. D'après la première question, la série entière qui définit la fonction génératrice de la variable aléatoire X :

$$G_X(s) = \sum_{n=0}^{\infty} P(X = n)s^n$$

a un rayon de convergence $R \geq 1$. Nous en déduisons que G_X est dérivable sur $] -1, 1[$ et que :

$$\forall s \in] -1, 1[\quad G'_X(s) = \sum_{n=1}^{\infty} P(X = n)ns^{n-1}.$$

Notre hypothèse d'intégrabilité sur X équivaut à la convergence de la série à termes positifs suivante :

$$\sum_{n=1}^{\infty} nP(X = n).$$

Un résultat classique sur les séries entières nous permet d'en déduire que la série :

$$\sum_{n=1}^{\infty} P(X = n)ns^{n-1}$$

converge uniformément sur $[0,1]$ vers une application continue. En particulier, $G'_X(s)$ admet une limite quand $s \nearrow 1$ et nous savons que cela entraîne la dérivabilité à gauche de G_X au point 1, ainsi que l'égalité :

$$G'_X(1) = \lim_{s \nearrow 1} G'_X(s) = \lim_{s \nearrow 1} \sum_{n=1}^{\infty} P(X = n)ns^{n-1} = \sum_{n=1}^{\infty} nP(X = n) = E[X].$$

Dans les trois exemples suivants, nous allons donc utiliser l'égalité :

$$E[X] = G'_X(1).$$

Si $X \sim \text{Bernoulli}(p)$, $G'_X(s) = p$ donc $E[X] = p$.

Si $X \sim \text{Géom}(a)$, $G'_X(s) = \frac{a(1-a)}{(1-as)^2}$ donc $E[X] = \frac{a}{1-a}$.

Enfin, pour $X \sim \mathcal{P}(\lambda)$, $G'_X(s) = \lambda e^{-\lambda(1-s)}$ d'où $E[X] = \lambda$.

Chapitre 14

Variables à densité

À l'extrême opposé d'une variable aléatoire discrète, dont la loi n'est faite que de masses ponctuelles, la loi d'une variable aléatoire absolument continue correspond à une répartition continue de masse sur la droite réelle et en particulier n'admet aucune masse ponctuelle. Plus précisément, si nous reprenons l'analogie entre une loi sur \mathbb{R} et une répartition de masse sur un fil infiniment long et infiniment mince de masse totale 1, nous allons maintenant nous intéresser à la situation correspondant à un fil admettant une densité de masse linéique, à savoir une application f à valeurs dans \mathbb{R}_+ telle que la masse du segment élémentaire $[x, x + dx]$ soit $f(x) dx$.

Nous rencontrerons des variables aléatoires absolument continues dans de nombreuses situations de modélisation où interviennent des grandeurs continues comme le temps (1er instant de panne d'une machine, temps d'attente à un guichet etc.), l'espace (mesure d'une longueur perturbée par de petites erreurs aléatoires) ou autre (pression atmosphérique mesurée dans une station météo).

14.1 Définitions

Définition 14.1.1 On appelle densité de probabilité sur \mathbb{R} toute application $f \in C_M(\mathbb{R}, \mathbb{R}_+)$ telle que :

$$\int_{-\infty}^{+\infty} f(x) dx = 1.$$

Définition 14.1.2 Soit f une densité de probabilité sur \mathbb{R} . On dit qu'une variable aléatoire réelle X admet f pour densité si, pour tout intervalle réel I , on a l'égalité :

$$P(X \in I) = \int_I f(x) dx.$$

Remarques.

1. Nous avons admis que la connaissance des quantités $P(X \in I)$ pour tout intervalle réel I déterminait entièrement la loi de X . Ainsi, si X admet la densité f , sa loi P_X est entièrement déterminée. On dit que P_X est la loi de densité f ou encore que X suit la loi de densité f .
2. Il n'est pas nécessaire de vérifier à l'avance que f est une densité de probabilité sur \mathbb{R} mais seulement que $f \in C_M(\mathbb{R}, \mathbb{R}_+)$. En effet, en appliquant l'égalité précédente avec $I = \mathbb{R}$, on en déduit que f est bien une densité de probabilité.

14.2 Exemples classiques de lois à densité

La loi uniforme sur $[a, b]$

C'est la loi de densité $f(x) = \frac{1}{b-a} \mathbf{1}_{[a,b]}(x)$ (on vérifie immédiatement que $f \in C_M(\mathbb{R}, \mathbb{R}_+)$ et $\int_{\mathbb{R}} f(x) dx = 1$).

Cette loi est notée $U([a, b])$. Elle correspond à l'expression "tirer un nombre au hasard entre a et b ". La fonction 'random' d'une calculatrice ou d'un ordinateur est censée simuler une telle loi pour $a = 0$ et $b = 1$.

Si la v.a.r. X admet f pour densité, on dit qu'elle suit la loi uniforme sur $[a, b]$ et l'on note $X \sim U([a, b])$.

La loi exponentielle

Soit $\lambda > 0$ fixé. L'application f_λ définie pour tout $x \in \mathbb{R}$ par :

$$f_\lambda(x) = \lambda e^{-\lambda x} \mathbf{1}_{\mathbb{R}_+^*}(x)$$

est une densité de probabilité sur \mathbb{R} .

En effet, on constate facilement que $f_\lambda \in C_M(\mathbb{R}, \mathbb{R}_+)$ et nous calculons :

$$\int_{-\infty}^{+\infty} f_\lambda(x) dx = \int_0^{+\infty} \lambda e^{-\lambda x} dx = [-e^{-\lambda x}]_0^{+\infty} = 1$$

On appelle loi exponentielle de paramètre λ la loi de densité f_λ . Comme nous le verrons dans l'exercice 14.6.2, cette loi est caractérisée par son "absence de mémoire". Par exemple, la durée de vie d'un matériel idéal qui ne subirait aucun vieillissement est une variable aléatoire exponentielle.

La loi gaussienne (ou loi normale)

Nous montrerons plus tard que l'application f définie par :

$$\forall x \in \mathbb{R}, \quad f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

vérifie bien $\int_{\mathbb{R}} f(x) dx = 1$. Comme il est clair que $f \in C_M(\mathbb{R}, \mathbb{R}_+)$ (elle est même continue), l'application f représentée ci-dessous est une densité de probabilité appelée *densité gaussienne (ou normale) centrée réduite*. La loi associée est notée $\mathcal{N}(0, 1)$.

Si nous fixons deux paramètres $\mu \in \mathbb{R}$ et $\sigma > 0$, nous déduisons de ce qui précède que l'application $f_{\mu, \sigma}$ définie par :

$$\forall x \in \mathbb{R}, \quad f_{\mu, \sigma}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

est aussi une densité de probabilité (il suffit de faire un simple changement de variable dans une intégrale pour le voir). On l'appelle densité gaussienne (ou normale) de paramètres μ et σ^2 . La loi associée est notée $\mathcal{N}(\mu, \sigma^2)$.

La loi de Cauchy

Pour tout $a > 0$ fixé, nous définissons l'application f_a par :

$$\forall x \in \mathbb{R}, \quad f(x) = \frac{a}{\pi} \frac{1}{a^2 + x^2}.$$

On a bien sûr $f \in C_M(\mathbb{R}, \mathbb{R}_+)$ (elle est même continue) et l'on calcule :

$$\int_{-\infty}^{+\infty} f_a(x) dx = \frac{1}{\pi} \left[\arctan \frac{x}{a} \right]_{-\infty}^{+\infty} = 1$$

L'application f_a est donc une densité de probabilité appelée densité de Cauchy de paramètre a . La loi associée est notée $C(a)$.

La loi gamma

Pour $a > 0$ et $\lambda > 0$, nous définissons l'application $g_{a, \lambda}$ par :

$$g_{a, \lambda}(x) = \frac{\lambda^a}{\Gamma(a)} e^{-\lambda x} x^{a-1} \mathbf{1}_{\mathbb{R}_+^*}(x).$$

Dans le cas $a \geq 1$ et $\lambda > 0$, on vérifie immédiatement que $g_{a, \lambda} \in C_M(\mathbb{R}, \mathbb{R}_+)$ et l'on calcule, avec le changement de variable $y = \lambda x$,

$$\int_{-\infty}^{+\infty} g_{a, \lambda}(x) dx = \frac{1}{\Gamma(a)} \int_0^{+\infty} e^{-\lambda x} (\lambda x)^{a-1} \lambda dx = \frac{1}{\Gamma(a)} \int_0^{+\infty} e^{-y} y^{a-1} dy = 1.$$

L'application $g_{a,\lambda}$ est donc une densité de probabilité appelée densité gamma de paramètres a et λ . La loi associée est notée $\gamma(a, \lambda)$.

Remarque. Dans le cas $(a, \lambda) \in]0, 1[\times \mathbb{R}_+^*$, on a $g_{a,\lambda} \notin C_M(\mathbb{R}, \mathbb{R}_+)$ puisque :

$$\lim_{x \rightarrow 0^+} g_{a,\lambda}(x) = \lim_{x \rightarrow 0^+} \frac{\lambda^a}{\Gamma(a)} e^{-\lambda x} x^{a-1} = +\infty.$$

L'application $g_{a,\lambda}$ n'est donc pas une densité de probabilité sur \mathbb{R} , au sens de la Définition 14.1.1, page 255. Néanmoins, on a $g_{a,\lambda} \in C_M(\mathbb{R}_+^*, \mathbb{R}_+)$ et $\int_{\mathbb{R}_+^*} g_{a,\lambda} = 1$. Nous dirons que $g_{a,\lambda}$ est une densité de probabilité sur \mathbb{R}_+^* , permettant de définir la loi $\gamma(a, \lambda)$ dans le cas général $(a, \lambda) \in (\mathbb{R}_+^*)^2$.

Cette difficulté technique n'aura aucune incidence quand nous étudierons la loi $\gamma(a, \lambda)$ dans la suite car nous prendrons toujours soin de nous ramener à des calculs d'intégrales sur \mathbb{R}_+^* .

14.3 Fonction de répartition

Proposition 14.3.1 *Soit $f \in C_M(\mathbb{R}, \mathbb{R}_+)$ et X une variable aléatoire réelle. L'application f est une densité de probabilité et la variable X admet f pour densité si et seulement si la fonction de répartition de X est donnée par :*

$$\forall x \in \mathbb{R}, \quad F_X(x) = \int_{-\infty}^x f(t) dt. \quad (14.1)$$

Démonstration: Dans le sens direct, cela résulte de la définition 14.1.2 avec $I =]-\infty, x]$.

Dans le sens réciproque, en faisant tendre x vers l'infini dans l'égalité (14.1), on obtient grâce au théorème 13.3.3, page 250,

$$\int_{-\infty}^{+\infty} f(t) dt = 1,$$

et donc f est bien une densité de probabilité. En outre, la variable X suit bien la loi de densité f puisque la fonction de répartition caractérise la loi d'une variable aléatoire réelle. □

À titre d'application, nous allons démontrer la proposition suivante.

Proposition 14.3.2 *Soit X une v.a.r. admettant la densité f . Nous définissons la v.a.r. $Y = \sigma X + \mu$, où $\mu \in \mathbb{R}$ et $\sigma > 0$ sont deux paramètres fixés. Alors Y admet la densité $f_{\mu,\sigma}$ définie par la formule :*

$$\forall x \in \mathbb{R}, \quad f_{\mu,\sigma}(x) = \frac{1}{\sigma} f\left(\frac{x - \mu}{\sigma}\right) \quad (14.2)$$

Démonstration: Calculons la f.r. de Y pour tout $t \in \mathbb{R}$:

$$F_Y(t) = P(Y \leq t) = P(\sigma X + \mu \leq t) = P\left(X \leq \frac{t - \mu}{\sigma}\right).$$

Puisque la variable X admet la densité f , nous en déduisons :

$$\forall t \in \mathbb{R}, \quad F_Y(t) = \int_{-\infty}^{\frac{t-\mu}{\sigma}} f(x) dx.$$

Dans cette intégrale, nous effectuons le changement de variable $y = \sigma x + \mu \Leftrightarrow x = \frac{y-\mu}{\sigma}$, ce qui nous donne :

$$\forall t \in \mathbb{R}, \quad F_Y(t) = \int_{-\infty}^t f\left(\frac{y-\mu}{\sigma}\right) \frac{1}{\sigma} dy.$$

La proposition 14.3.1 nous permet d'en déduire que l'application $f_{\mu,\sigma}$ définie par la formule (14.2) est une densité de probabilité et que Y suit la loi de densité $f_{\mu,\sigma}$. □

Proposition 14.3.3 *Soit X une variable aléatoire réelle de fonction de répartition F_X et $f \in C(\mathbb{R}, \mathbb{R}_+)$. Alors la variable X admet f pour densité si et seulement si F_X est dérivable sur \mathbb{R} avec :*

$$\forall x \in \mathbb{R}, \quad F'_X(x) = f(x).$$

Démonstration: Supposons que X admet la densité f . Alors la fonction de répartition de X est donnée par :

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

Comme f est continue, nous en déduisons que F_X est dérivable sur \mathbb{R} et que $F'_X = f$.

Dans le sens réciproque, puisque F_X est une primitive de f , nous avons pour tous réels $w \leq x$:

$$\int_w^x f(t) dt = F_X(x) - F_X(w).$$

Si nous faisons tendre w vers $-\infty$ dans l'égalité précédente, nous obtenons grâce au théorème 13.3.3 :

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

Ceci nous permet de conclure, en vertu de la proposition 14.3.1. □

Remarque. Cet énoncé admet une généralisation au cas f continue par morceaux. Le lecteur pourra consulter [DAN 408] à ce sujet. □

Voici une application de la proposition 14.3.3 : Soit X v.a.r. de loi uniforme sur l'intervalle $] -\frac{\pi}{2}, \frac{\pi}{2}[$. Que vaut la loi de $Y = \tan X$?
Calculons la f.r. de Y pour tout $t \in \mathbb{R}$:

$$F_Y(t) = P(Y \leq t) = P(X \leq \arctan t) = \int_{-\frac{\pi}{2}}^{\arctan t} \frac{1}{\pi} dx = \frac{\arctan t + \frac{\pi}{2}}{\pi}.$$

Cette fonction étant de classe C^1 sur \mathbb{R} , en calculant $f = F'_Y$, nous déduisons de la proposition précédente que Y admet pour densité l'application f définie par $f(x) = \frac{1}{\pi} \frac{1}{1+x^2}$ pour tout $x \in \mathbb{R}$.
Autrement dit, Y suit la loi de Cauchy de paramètre 1.

Remarque. La fonction de répartition d'une v.a.r. de densité f s'écrivant $F_X(t) = \int_{-\infty}^t f(x) dx$, c'est en particulier une application continue, d'où le qualificatif « continue » donné à la v.a.r.

Pourquoi précise-t-on v.a.r. *absolument* continue ? Parce que les v.a.r. admettant une densité ne sont qu'un cas particulier de variables dont la f.r. est continue. C'est néanmoins le seul type de variables continues que nous examinerons dans le cadre de ce cours.

14.4 Espérance. Théorème de transfert

Définition 14.4.1 Soit X une variable aléatoire réelle de densité f . On appelle espérance de X le nombre réel

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

dès lors que cette intégrale impropre est absolument convergente.

Nous énonçons simplement ici un important résultat sur lequel nous reviendrons plus précisément dans le chapitre suivant dans le cas des vecteurs aléatoires.

Théorème 14.4.2 (de transfert) Soit X une variable aléatoire réelle de densité f . Pour toute application $\psi \in C_M(\mathbb{R}, \mathbb{R})$ telle que $|\psi|f$ soit intégrable sur \mathbb{R} , la variable aléatoire $\psi(X)$ admet pour espérance :

$$E[\psi(X)] = \int_{\mathbb{R}} \psi(x)f(x) dx.$$

Remarque. La variable aléatoire $\psi(X)$ n'admet pas forcément de densité, comme nous le constatons en prenant $\psi \equiv 0$. \square

Le théorème de transfert admet une sorte de réciproque (hors programme) que nous énonçons maintenant.

Proposition 14.4.3 *Soient X une variable aléatoire réelle et $f : \mathbb{R} \rightarrow \mathbb{R}_+$ une application intégrable sur \mathbb{R} . Alors, il y a équivalence entre les deux propositions suivantes :*

1. *L'application f est une densité de probabilité sur \mathbb{R} et la variable aléatoire réelle X admet f pour densité.*
2. *Pour toute application $\psi \in C^0(\mathbb{R}, \mathbb{R})$ bornée, la variable aléatoire $\psi(X)$ admet pour espérance :*

$$E[\psi(X)] = \int_{\mathbb{R}} \psi(x)f(x) dx.$$

Le lecteur est invité à consulter [DAN 438-450] pour des exemples et propriétés de variables à densité (que Dantzer appelle des variables *continues*).

14.5 Moments d'une variable aléatoire

Le lecteur pourra étudier dans le livre d'ESCOFFIER ou dans le tome 1 d'OUVRARD (par exemple) les notions de *moments*, *variance* et *écart-type* d'une variable aléatoire réelle en distinguant le cas discret et le cas où la variable admet une densité.

Une application importante de la notion de variance est l'inégalité de Bienaymé-Tchebychev, qui est une clé dans la démonstration de la loi faible des grands nombres. Dans l'énoncé suivant, la v.a.r. X est soit discrète, soit à densité.

Proposition 14.5.1 (Inégalité de Bienaymé-Tchebychev) *Considérons une variable aléatoire réelle X admettant un moment d'ordre 2.*

Pour tout $t > 0$, nous avons :

$$P(|X - E[X]| \geq t) \leq \frac{\text{Var} X}{t^2}$$

Dans les ouvrages cités ci-dessus, le lecteur trouvera encore les notions de *covariance* et de *coefficient de corrélation* pour un couple de variables aléatoires. Ces notions sont introduites en distinguant à nouveau le cas discret et le cas où les variables admettent des densités, conformément au programme officiel.

14.6 Exercices sur les fonctions de répartition

Dans cette section, nous regroupons quelques exercices sur les variables aléatoires réelles qui peuvent tous être résolus en utilisant la notion de fonction de répartition. En outre, les exercices 14.6.3, 14.6.4 et 14.6.5 peuvent être résolus par une seconde méthode : utiliser la proposition 14.4.3 en faisant un changement de variable dans une intégrale sur \mathbb{R} .

14.6.1 Minimum de variables exponentielles.

Soient X_1, \dots, X_n des v.a.r. suivant des lois exponentielles de paramètres respectifs $\lambda_1, \dots, \lambda_n$. On suppose que ces v.a.r. sont indépendantes, ce qui équivaut à :

$\forall (x_i)_{1 \leq i \leq n} \in \mathbb{R}^n$, les événements $\{X_i \leq x_i\}, 1 \leq i \leq n$, sont mutuellement indépendants.

Calculer la loi de $X = \min_{1 \leq i \leq n} X_i$.

14.6.2 Variables amnésiques.

[COT 91-92] Soit T une v.a. à valeurs réelles telle que, pour tous $s, t \geq 0$, on ait :

$$P(T > t + s) = P(T > t)P(T > s).$$

Le but de cet exercice est de montrer que, soit $P(T > 0) = 0$, soit T suit une loi exponentielle.

1. Montrer que si $P(T > 0) > 0$, alors pour tout $t > 0$, $P(T > t) > 0$.
2. On définit alors l'application $f(t) = \ln P(T > t)$, $t > 0$.

Montrer que $f(x) = xf(1)$ pour tout x rationnel positif, puis pour tout x réel positif.

3. Conclure.

14.6.3 Loi du χ^2 à un degré de liberté.

[COT 83 et 86]

Soit $X \sim \mathcal{N}(0, 1)$. Montrer que la variable aléatoire réelle $Y = X^2$ admet pour densité l'application $f : \mathbb{R} \rightarrow \mathbb{R}_+$ définie par :

$$\forall y \in \mathbb{R} \quad f(y) = \frac{1}{\sqrt{2\pi y}} \exp\left(-\frac{y}{2}\right) \mathbf{1}_{\mathbb{R}_+^*}(y).$$

Cette densité définit la loi du χ^2 à un degré de liberté.

14.6.4 Loi gaussienne dans \mathbb{R} .

Soit $X \sim \mathcal{N}(0, 1)$. Montrer que la variable aléatoire réelle $Y = m + \sigma X$, où $m \in \mathbb{R}$ et $\sigma > 0$, admet une densité que l'on calculera. Comment s'appelle la loi de Y ?

14.6.5 Avec une loi de Cauchy.

[COT 52-53] Soit X une variable aléatoire réelle suivant une loi de Cauchy de paramètre 1.

1. Quelle est la loi de aX , où $a \in \mathbb{R}^*$?
2. Montrer que $Y = \ln |X|$ admet la densité : $p(y) = \frac{1}{\pi \cosh y}$.

Chapitre 15

Vecteurs aléatoires et indépendance

Remarque préliminaire : Les expressions *vecteur aléatoire de dimension n* et *variable aléatoire à valeurs dans \mathbb{R}^n* sont synonymes. Il est donc approprié de s'intéresser aux vecteurs aléatoires dans le cadre d'une leçon telle que « Variables aléatoires possédant une densité. Exemples. »

Pour introduire la notion de vecteur aléatoire, nous nous intéressons à quelques situations particulières :

- Nous choisissons une personne au hasard dans la population française ; soit X son âge et Y son taux de cholestérol. Est-ce qu'il y a un lien entre les v.a.r. X et Y ou bien prennent-elles leurs valeurs de façon indépendante ?
- Nous observons la trajectoire de la fusée Ariane en repérant sa position à différents instants $t_i, 1 \leq i \leq p$, grâce à ses coordonnées $(X_{t_i}, Y_{t_i}, Z_{t_i}), 1 \leq i \leq p$ dans un certain repère orthonormé. Pour tenir compte des erreurs entre la trajectoire idéale et la trajectoire réelle, chacune de ces coordonnées est considérée comme une v.a.r. Une idée de la trajectoire réelle nous est alors donnée par le "grand vecteur" $(X_{t_1}, Y_{t_1}, Z_{t_1}, \dots, X_{t_p}, Y_{t_p}, Z_{t_p})$ qui est aléatoire.
- Dans une chaîne de production, le responsable du contrôle qualité prélève p pièces au hasard sur la production de la journée et en mesure les masses respectives X_1, \dots, X_p . Comment peut-il estimer le poids moyen d'une pièce produite dans la journée ou encore l'écart-type de la masse d'une pièce ?

Dans toutes ces situations, nous constatons notre besoin d'étudier plusieurs variables aléatoires réelles simultanément et non plus une par une. De fa-

çon générale, nous sommes amenés à étudier des n -uplets dont chacune des composantes est une variable aléatoire réelle.

15.1 Les vecteurs aléatoires et leurs lois

Dans toute la suite, nous considérerons un espace probabilisé (Ω, \mathcal{A}, P) .

Avant de définir les vecteurs aléatoires à valeurs dans \mathbb{R}^n , nous rappelons que la tribu borélienne sur \mathbb{R}^n est la tribu engendrée par l'ensemble des ouverts de \mathbb{R}^n (pour la topologie usuelle).

Proposition 15.1.1 *Soit $n \in \mathbb{N}^*$. La tribu $\mathcal{B}(\mathbb{R}^n)$ est engendrée par l'ensemble des pavés ouverts à extrémités rationnelles, c'est-à-dire par :*

$$\mathcal{E} = \left\{ \prod_{i=1}^n]a_i, b_i[\text{ avec } (a_1, \dots, a_n, b_1, \dots, b_n) \in \mathbb{Q}^{2n} \right\}$$

Démonstration: Puisque ces pavés sont ouverts, on a $\sigma(\mathcal{E}) \subset \mathcal{B}(\mathbb{R}^n)$. Prouvons l'inclusion inverse.

Si nous munissons \mathbb{R}^n de la norme infinie ($\|(x_1, \dots, x_n)\|_\infty = \sup_{1 \leq i \leq n} |x_i|$), nous constatons que les boules ouvertes sont des pavés ouverts. On conclut alors en utilisant le même schéma de démonstration que pour le premier point de la proposition 13.1.2

□

Remarque. On montre facilement que la tribu $\mathcal{B}(\mathbb{R}^n)$ est aussi engendrée par l'ensemble des pavés fermés à extrémités rationnelles, c'est-à-dire par :

$$\mathcal{E}' = \left\{ \prod_{i=1}^n [a_i, b_i] \text{ avec } (a_1, \dots, a_n, b_1, \dots, b_n) \in \mathbb{Q}^{2n} \right\}$$

De façon similaire à ce qui a été fait dans le chapitre précédent à propos des variables aléatoires réelles, nous définissons les vecteurs aléatoires à valeurs dans \mathbb{R}^n comme suit.

Définition 15.1.2 *On appelle variable aléatoire à valeurs dans \mathbb{R}^n (ou vecteur aléatoire à valeurs dans \mathbb{R}^n , ou vecteur aléatoire de dimension n) défini sur (Ω, \mathcal{A}, P) toute application $X : \Omega \rightarrow \mathbb{R}^n$ vérifiant :*

$$\forall B \in \mathcal{B}(\mathbb{R}^n) \quad X^{-1}(B) \in \mathcal{A}. \quad (15.1)$$

Un résultat général de la théorie de la mesure permet de démontrer la proposition suivante, que nous admettrons.

Proposition 15.1.3 *Soit \mathcal{E} un ensemble de parties de \mathbb{R}^n qui engendre la tribu $\mathcal{B}(\mathbb{R}^n)$. Alors une application $X : \Omega \rightarrow \mathbb{R}^n$ est un vecteur aléatoire n -dimensionnel sur (Ω, \mathcal{A}, P) si et seulement si la condition suivante est satisfaite :*

$$\forall B \in \mathcal{E} \quad X^{-1}(B) \in \mathcal{A}.$$

Cette proposition admet un corollaire important, qui sera notre outil principal pour démontrer qu'une application $X : \Omega \rightarrow \mathbb{R}^n$ est un vecteur aléatoire défini sur (Ω, \mathcal{A}, P) . Avant de l'énoncer, nous introduisons quelques notations. Pour tout $1 \leq i \leq n$, nous appellerons $\phi_i : \mathbb{R}^n \rightarrow \mathbb{R}$ la i -ème projection canonique, définie par :

$$\forall x = (x_1, \dots, x_n) \in \mathbb{R}^n \quad \phi_i(x) = x_i.$$

Pour tout $1 \leq i \leq n$, nous considérons alors l'application $X_i = \phi_i \circ X : \Omega \rightarrow \mathbb{R}$ de sorte que nous pouvons écrire l'application X à valeurs dans \mathbb{R}^n sous la forme $X = (X_1, \dots, X_n)$.

Corollaire 15.1.4 *L'application $X : \Omega \rightarrow \mathbb{R}^n$ est un vecteur aléatoire de dimension n sur (Ω, \mathcal{A}, P) si et seulement si pour tout $1 \leq i \leq n$, l'application $X_i = \phi_i \circ X : \Omega \rightarrow \mathbb{R}$ est une variable aléatoire réelle sur (Ω, \mathcal{A}, P) .*

Démonstration: Supposons que X est un vecteur aléatoire n -dimensionnel et montrons que pour tout $1 \leq i \leq n$ et tout intervalle réel $]a, b[$, nous avons $X_i^{-1}(]a, b[) \in \mathcal{A}$, ce qui suffira à établir le sens direct de notre corollaire d'après la proposition 13.2.2. Or nous avons l'égalité :

$$X_i^{-1}(]a, b[) = X^{-1}(\mathbb{R} \times \dots \underbrace{]a, b[}_{i\text{-ème position}} \dots \times \mathbb{R}) \in \mathcal{A}$$

puisque $\mathbb{R} \times \dots]a, b[\dots \times \mathbb{R}$ est un ouvert donc un borélien de \mathbb{R}^n .

Passons au sens réciproque. D'après la proposition précédente, il suffit de vérifier que pour tout pavé ouvert $\prod_{i=1}^n]a_i, b_i[$, nous avons :

$$X^{-1} \left(\prod_{i=1}^n]a_i, b_i[\right) \in \mathcal{A}.$$

Or ceci résulte de l'égalité :

$$X^{-1} \left(\prod_{i=1}^n]a_i, b_i[\right) = \bigcap_{i=1}^n \underbrace{X_i^{-1}(]a_i, b_i[)}_{\in \mathcal{A}}.$$

□

Remarque. Autrement dit, un vecteur aléatoire de dimension n n'est autre qu'un n -uplet dont toutes les composantes sont des variables aléatoires réelles. Une conséquence immédiate de ce corollaire est que l'ensemble des vecteurs aléatoires n -dimensionnels définis sur (Ω, \mathcal{A}, P) est un \mathbb{R} -espace vectoriel.

Nous continuons à généraliser ce qui a été fait dans le chapitre précédent à propos des variables aléatoires réelles, en définissant la loi d'un vecteur aléatoire à valeurs dans \mathbb{R}^n comme suit.

Théorème et définition 15.1.5 *Considérons un espace probabilisé (Ω, \mathcal{A}, P) et un vecteur aléatoire n -dimensionnel $X : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$.*

Alors l'application $P_X : \mathcal{B}(\mathbb{R}^n) \rightarrow [0, 1]$ définie par :

$$\forall B \in \mathcal{B}(\mathbb{R}^n) \quad P_X(B) = P(X \in B)$$

est une mesure de probabilité sur l'espace mesurable $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, appelée loi de probabilité (ou simplement loi) du vecteur aléatoire X .

Démonstration: Tout-à-fait similaire à celle qui a été faite dans le cas des variables aléatoires réelles. □

15.2 Vecteurs aléatoires discrets

Définition 15.2.1 *Un vecteur aléatoire n -dimensionnel $X = (X_1, \dots, X_n)$ est dit discret si toutes ses composantes X_1, \dots, X_n sont des variables aléatoires réelles discrètes.*

Pour tout $1 \leq i \leq n$, nous notons $E_i = X_i(\Omega)$ qui est donc un sous-ensemble réel fini ou dénombrable. Notons qu'alors $X(\Omega) \subset E_1 \times \dots \times E_n$ est un sous-ensemble de \mathbb{R}^n également fini ou dénombrable. Le lecteur pourra montrer que connaître la loi d'un vecteur aléatoire discret équivaut à connaître toutes les quantités suivantes :

$$P(X_1 = x_1, \dots, X_n = x_n), \quad x_1 \in E_1, \dots, x_n \in E_n.$$

Nous renvoyons le lecteur à [MEY] IV.3, IV.5, V et VII pour les notions liées aux vecteurs aléatoires discrets.

15.3 Vecteurs aléatoires à densité

Dans cette section, nous écrirons les énoncés dans le cas de la dimension 2 pour alléger les écritures mais tout se généralise en dimension n . Un vecteur aléatoire de dimension 2 est encore appelé *couple aléatoire*.

Définition 15.3.1 On appelle densité de probabilité sur \mathbb{R}^2 toute application $f : \mathbb{R}^2 \rightarrow \mathbb{R}_+$ intégrable et telle que :

$$\int_{\mathbb{R}^2} f(x, y) \, dx \, dy = 1.$$

Exemple. Si $D \subset \mathbb{R}^2$ admet une aire non nulle et finie, alors l'application $f : \mathbb{R}^2 \rightarrow \mathbb{R}_+$ définie par :

$$\forall (x, y) \in \mathbb{R}^2 \quad f(x, y) = \frac{1}{\text{aire}(D)} \mathbf{1}_D(x, y)$$

est une densité de probabilité appelée densité uniforme sur D .

Définition 15.3.2 On dit qu'un couple aléatoire (X, Y) admet f pour densité si pour tous intervalles réels I et J , on a l'égalité :

$$P(X \in I, Y \in J) = \int_{I \times J} f(x, y) \, dx \, dy.$$

Remarques.

1. L'égalité précédente s'écrit évidemment de façon équivalente :

$$P((X, Y) \in I \times J) = \int_{I \times J} f(x, y) \, dx \, dy.$$

Un résultat de théorie de la mesure nous dit alors que

$$P((X, Y) \in B) = \int_B f(x, y) \, dx \, dy,$$

pour tout sous-ensemble $B \subset \mathbb{R}^2$ pour lequel l'intégrale du membre de droite a un sens (par exemple un *domaine simple* du plan).

2. Il n'est pas nécessaire de vérifier à l'avance que f est une densité de probabilité sur \mathbb{R}^2 mais seulement que f est une application positive et intégrable sur \mathbb{R}^2 . En effet, en appliquant l'égalité précédente avec $I = J = \mathbb{R}$, on en déduit que f est bien une densité de probabilité.

Théorème 15.3.3 (de transfert) Soit (X, Y) un couple aléatoire de densité f et $\psi : \mathbb{R}^2 \rightarrow \mathbb{R}$ une application telle que $|\psi|f$ soit intégrable sur \mathbb{R}^2 . Alors $\psi(X, Y)$ est une variable aléatoire réelle qui admet pour espérance :

$$E[\psi(X, Y)] = \int_{\mathbb{R}^2} \psi(x, y) f(x, y) \, dx \, dy.$$

Conformément au programme, nous admettons ce théorème. Remarquons simplement que $\psi(X, Y)$ n'est pas nécessairement une variable à densité, comme nous le voyons en prenant $\psi \equiv 0$.

En fait, on peut raffiner le théorème précédent en énonçant une équivalence¹.

Proposition 15.3.4 *Soient (X, Y) un couple aléatoire et $f : \mathbb{R}^2 \rightarrow \mathbb{R}_+$ une application intégrable sur \mathbb{R}^2 . Alors, il y a équivalence entre les propositions :*

1. *L'application f est une densité de probabilité sur \mathbb{R}^2 et le couple aléatoire (X, Y) admet f pour densité.*
2. *Pour toute application $\psi : \mathbb{R}^2 \rightarrow \mathbb{R}$ telle que $|\psi|f$ soit intégrable sur \mathbb{R}^2 , on a l'égalité :*

$$E[\psi(X, Y)] = \int_{\mathbb{R}^2} \psi(x, y) f(x, y) dx dy.$$

Démonstration: L'implication $1 \Rightarrow 2$ n'est autre que le théorème de transfert. Montrons donc $2 \Rightarrow 1$.

En appliquant une première fois notre égalité avec $\psi \equiv 1$, nous obtenons que f est bien une densité de probabilité sur \mathbb{R}^2 .

Pour tous intervalles réels I et J , nous appliquons maintenant l'égalité avec $\psi = \mathbf{1}_{I \times J}$. Remarquons d'abord que

$$E[\mathbf{1}_{I \times J}(X, Y)] = E[\mathbf{1}_{\{X \in I, Y \in J\}}] = P(X \in I, Y \in J).$$

Nous obtenons donc :

$$P(X \in I, Y \in J) = \int_{I \times J} f(x, y) dx dy,$$

d'où la conclusion. □

Remarque. Pour éviter d'avoir à vérifier l'intégrabilité de $|\psi|f$, nous admettrons que l'équivalence reste valable en remplaçant la proposition 2. par : Pour toute application $\psi \in C^0(\mathbb{R}^2, \mathbb{R})$ bornée, on a l'égalité :

$$E[\psi(X, Y)] = \int_{\mathbb{R}^2} \psi(x, y) f(x, y) dx dy.$$

En utilisant le théorème du changement de variable dans des intégrales doubles, on prouve [FOA 197-198] le résultat suivant qui est très utile pour effectuer des calculs pratiques de densités de couples aléatoires.

1. Cette équivalence n'est pas au programme officiel du concours interne.

Proposition 15.3.5 *Soit Φ un C^1 -difféomorphisme de U sur V , où U et V sont deux ouverts de \mathbb{R}^2 . Considérons un couple aléatoire (X, Y) admettant une densité f nulle en dehors de U (donc $P[(X, Y) \in U] = 1$) et définissons un nouveau couple aléatoire en posant $(S, T) = \Phi(X, Y)$.*

Alors (S, T) admet pour densité l'application g définie par :

$$\forall (s, t) \in \mathbb{R}^2, \quad g(s, t) = f \circ \Phi^{-1}(s, t) |J_{\Phi^{-1}}(s, t)| \mathbf{1}_V(s, t),$$

où $J_{\Phi^{-1}}(s, t)$ désigne le déterminant jacobien du difféomorphisme Φ^{-1} .

Le lecteur, après avoir lu la section suivante sur l'indépendance de variables aléatoires, pourra utiliser ce corollaire pour résoudre les exercices 15.6.4 et 15.6.5.

Nous terminons cette section par un résultat sur l'existence de densités pour les v.a.r. coordonnées d'un vecteur aléatoire admettant une densité.

Proposition 15.3.6 *Si le couple aléatoire (X, Y) admet une densité f sur \mathbb{R}^2 , alors les variables aléatoires réelles X et Y admettent respectivement des densités de probabilité sur \mathbb{R} données par :*

$$f_X(x) = \int_{\mathbb{R}} f(x, y) dy \quad , \quad f_Y(y) = \int_{\mathbb{R}} f(x, y) dx,$$

dès lors que ces applications f_X et f_Y sont continues par morceaux. On les appelle les densités marginales.

Démonstration: Nous faisons la démonstration pour la variable X ; la démonstration pour Y s'en déduit par symétrie.

Soit I un intervalle réel quelconque; nous avons alors les égalités :

$$P(X \in I) = P((X, Y) \in I \times \mathbb{R}) = \int_{I \times \mathbb{R}} f(x, y) dx dy.$$

Si nous définissons l'application $f_X : \mathbb{R} \rightarrow \mathbb{R}_+$ par :

$$\forall x \in \mathbb{R} \quad f_X(x) = \int_{\mathbb{R}} f(x, y) dy,$$

alors le théorème de Fubini-Tonelli implique que f_X est intégrable sur \mathbb{R} et

$$\int_{I \times \mathbb{R}} f(x, y) dx dy = \int_I f_X(x) dx.$$

Finalement, nous avons bien montré que pour tout intervalle réel I ,

$$P(X \in I) = \int_I f_X(x) dx.$$

□

Remarque. Ainsi, si un vecteur aléatoire admet une densité, alors chacune de ses composantes est une variable aléatoire à densité. Il est à noter que la réciproque est fautive. En effet, soit X une variable aléatoire réelle à densité (nous pouvons prendre $X \sim \mathcal{N}(0, 1)$ par exemple). Nous allons prouver que le couple aléatoire (X, X) n'admet aucune densité, ce qui nous fournira un contre-exemple.

Notons $\Delta = \{(x, y) \in \mathbb{R}^2, x = y\}$ la première diagonale du plan. Nous avons bien sûr $P[(X, X) \in \Delta] = 1$.

Raisonnons par l'absurde en supposant que le couple aléatoire (X, X) admet une densité f . Nous avons alors :

$$P[(X, X) \in \Delta] = \int_{\Delta} f(x, y) dx dy = \int_{\mathbb{R}^2} \mathbf{1}_{x=y} f(x, y) dx dy.$$

D'après le théorème de Fubini-Tonelli, cette dernière intégrale vaut :

$$\int_{\mathbb{R}} \left(\int_{\mathbb{R}} \mathbf{1}_{x=y} f(x, y) dx \right) dy = \int_{\mathbb{R}} \left(\int_y^y f(x, y) dx \right) dy = 0,$$

d'où une contradiction. □

15.4 Indépendance de p v.a.r.

Dans toute cette section, on fixe un entier $p \geq 2$.

Définition 15.4.1 On dit que p variables aléatoires réelles X_1, \dots, X_p sont indépendantes (sous-entendu dans leur ensemble) si pour tous intervalles réels I_1, \dots, I_p , on a l'égalité :

$$P(X_1 \in I_1, \dots, X_p \in I_p) = P(X_1 \in I_1) \cdots P(X_p \in I_p)$$

Un instant de réflexion permet de constater que cette condition équivaut à : Pour tous interv. réels I_1, \dots, I_p , les événements $\{X_1 \in I_1\}, \dots, \{X_p \in I_p\}$ sont indépendants dans leur ensemble.

Proposition 15.4.2 On considère p variables aléatoires réelles X_1, \dots, X_p indépendantes et admettant des densités respectives f_{X_1}, \dots, f_{X_p} .

Alors le vecteur aléatoire (X_1, \dots, X_p) admet pour densité l'application $f : \mathbb{R}^p \rightarrow \mathbb{R}_+$ définie par :

$$\forall (x_1, \dots, x_p) \in \mathbb{R}^p \quad f(x_1, \dots, x_p) = f_{X_1}(x_1) \cdots f_{X_p}(x_p)$$

Remarque. L'existence de cette densité est déjà un résultat intéressant en soi. En outre, sous nos hypothèses, nous obtenons une formule explicite de cette densité.

Démonstration: Pour tous intervalles réels I_1, \dots, I_p , on a l'égalité :

$$P(X_1 \in I_1, \dots, X_p \in I_p) = \prod_{i=1}^p P(X_i \in I_i) = \prod_{i=1}^p \int_{I_i} f_{X_i}(x_i) dx_i$$

Comme les applications $f_{X_i}, 1 \leq i \leq p$ sont toutes positives, le théorème de Fubini nous permet d'en déduire :

$$P((X_1, \dots, X_p) \in I_1 \times \dots \times I_p) = \int_{I_1 \times \dots \times I_p} f(x_1, \dots, x_p) dx_1 \dots dx_p,$$

où l'application f est définie comme dans l'énoncé, d'où la conclusion. \square

Réciproquement, nous avons le résultat suivant :

Proposition 15.4.3 *Supposons que le vecteur aléatoire (X_1, \dots, X_p) admet pour densité une application $f : \mathbb{R}^p \rightarrow \mathbb{R}_+$ à variables séparables, c'est-à-dire vérifiant*

$$\forall (x_1, \dots, x_p) \in \mathbb{R}^p, \quad f(x_1, \dots, x_p) = f_1(x_1) \cdots f_p(x_p),$$

où les applications $f_i : \mathbb{R} \rightarrow \mathbb{R}_+$ sont intégrables. Alors :

1. Les variables aléatoires réelles X_1, \dots, X_p sont indépendantes.
2. Pour tout $1 \leq i \leq p$, la variable aléatoire réelle X_i admet une densité de la forme

$$f_{X_i}(x_i) = c_i f_i(x_i),$$

où les c_i sont des constantes positives telles que $\prod_{i=1}^p c_i = 1$.

Démonstration: Dans un souci d'alléger les notations, nous l'écrivons dans le cas $p = 2$ mais elle se généralise immédiatement.

D'après la proposition 15.3.6, la variable aléatoire réelle X_1 admet pour densité

$$f_{X_1}(x_1) = \int_{\mathbb{R}} f(x_1, x_2) dx_2 = \int_{\mathbb{R}} f_1(x_1) f_2(x_2) dx_2 = c_1 f_1(x_1),$$

où l'on a posé $c_1 := \int_{\mathbb{R}} f_2(x_2) dx_2 \in \mathbb{R}_+$ puisque f_2 a été supposée intégrable. Par un raisonnement symétrique, on montre que X_2 admet une densité de la forme $f_{X_2}(x_2) = c_2 f_2(x_2)$.

D'une part, nous avons :

$$\int_{\mathbb{R}^2} f_{X_1}(x_1) f_{X_2}(x_2) dx_1 dx_2 = \left(\int_{\mathbb{R}} f_{X_1}(x_1) dx_1 \right) \left(\int_{\mathbb{R}} f_{X_2}(x_2) dx_2 \right) = 1$$

puisque f_{X_1} et f_{X_2} sont des densités de probabilité sur \mathbb{R} .

D'autre part, nous avons, en utilisant les égalités précédentes :

$$\int_{\mathbb{R}^2} f_{X_1}(x_1) f_{X_2}(x_2) dx_1 dx_2 = c_1 c_2 \int_{\mathbb{R}^2} f_1(x_1) f_2(x_2) dx_1 dx_2 = c_1 c_2$$

puisque $f(x_1, x_2) = f_1(x_1) f_2(x_2)$ est une densité de probabilité sur \mathbb{R}^2 .

Nous en déduisons l'égalité $c_1 c_2 = 1$. Notons qu'elle a pour conséquence :

$$\forall (x_1, x_2) \in \mathbb{R}^2 \quad f(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2) \quad (15.2)$$

Il nous reste à démontrer l'indépendance des v.a.r. X_1 et X_2 .

Pour tous intervalles réels I_1, I_2 , nous avons l'égalité :

$$P(X_1 \in I_1, X_2 \in I_2) = P((X_1, X_2) \in I_1 \times I_2) = \int_{I_1 \times I_2} f(x_1, x_2) dx_1 dx_2$$

D'après l'égalité (15.2), nous avons donc :

$$P(X_1 \in I_1, X_2 \in I_2) = \int_{I_1 \times I_2} f_{X_1}(x_1) f_{X_2}(x_2) dx_1 dx_2$$

Comme les applications f_{X_1} et f_{X_2} sont positives, le théorème de Fubini nous permet d'en déduire :

$$P(X_1 \in I_1, X_2 \in I_2) = \left(\int_{I_1} f_{X_1}(x_1) dx_1 \right) \left(\int_{I_2} f_{X_2}(x_2) dx_2 \right) = P(X_1 \in I_1) P(X_2 \in I_2)$$

d'où la conclusion. \square

15.5 Produit de convolution de densités

Définition 15.5.1 Soient f et g des densités de probabilité sur \mathbb{R} . On appelle produit de convolution des densités f et g , noté $f * g$, l'application de \mathbb{R} dans $\overline{\mathbb{R}}_+$ définie comme suit :

$$\forall x \in \mathbb{R}, \quad f * g(x) = \int_{-\infty}^{+\infty} f(x-y) g(y) dy.$$

Proposition 15.5.2 Soient X et Y deux variables aléatoires réelles indépendantes admettant des densités respectives f et g . Si $f * g \in C_M(\mathbb{R}, \mathbb{R})$, alors la variable aléatoire réelle $X + Y$ admet pour densité $f * g$.

Démonstration: D'après la Proposition 15.4.2, le couple aléatoire (X, Y) admet pour densité $f(x, y) = f_X(x)f_Y(y)$ sur \mathbb{R}^2 .

Appliquons la Proposition 15.3.5 avec $U = V = \mathbb{R}^2$ et $\Phi(x, y) = (x + y, y)$ si bien que $\Phi^{-1}(s, t) = (s - t, t)$ et $J_{\Phi^{-1}}(s, t) = 1$.

Nous en déduisons que le couple aléatoire $(X + Y, Y)$ admet la densité $g(s, t) = f(s - t, t) = f_X(s - t)f_Y(t)$ sur \mathbb{R}^2 .

La Proposition 15.3.6 nous permet alors de conclure en calculant la première densité marginale de ce couple aléatoire. \square

Remarque. Le lecteur trouvera dans [OUV¹ 203-204] une démonstration alternative, bien adaptée au programme officiel.

Proposition 15.5.3 Soient X et Y deux v.a.r. indépendantes de lois respectives $\gamma(a, \lambda)$ et $\gamma(b, \lambda)$ avec $(a, b, \lambda) \in (\mathbb{R}_+^*)^3$. Alors, $X + Y \sim \gamma(a + b, \lambda)$.

Démonstration: Notons comme suit la densité de la loi $\gamma(a, \lambda)$:

$$g_{a,\lambda}(x) = \frac{\lambda^a}{\Gamma(a)} e^{-\lambda x} x^{a-1} \mathbf{1}_{\mathbb{R}_+^*}(x).$$

Nous allons montrer que :

$$\boxed{g_{a,\lambda} * g_{b,\lambda} = g_{a+b,\lambda}} \quad (15.3)$$

Ceci nous permettra de conclure puisque la variable aléatoire $X + Y$ admet alors $g_{a+b,\lambda}$ pour densité.

Par définition du produit de convolution, nous avons :

$$\forall x \in \mathbb{R}, g_{a,\lambda} * g_{b,\lambda}(x) = \int_{-\infty}^{+\infty} g_{a,\lambda}(x - y)g_{b,\lambda}(y)dy$$

En tenant compte des indicatrices de \mathbb{R}_+^* intervenant dans les densités, nous constatons que cette intégrale est nulle pour $x \leq 0$ et que

$$\forall x \in \mathbb{R}_+^*, g_{a,\lambda} * g_{b,\lambda}(x) = \int_0^x \frac{\lambda^a}{\Gamma(a)} e^{-\lambda(x-y)} (x - y)^{a-1} \frac{\lambda^b}{\Gamma(b)} e^{-\lambda y} y^{b-1} dy$$

Ainsi, nous avons :

$$g_{a,\lambda} * g_{b,\lambda}(x) = \begin{cases} \frac{\lambda^{a+b}}{\Gamma(a)\Gamma(b)} e^{-\lambda x} \int_0^x (x - y)^{a-1} y^{b-1} dy & \text{si } x > 0 \\ 0 & \text{sinon} \end{cases} \quad (15.4)$$

Arrêtons nous un instant sur cette dernière intégrale. En posant le changement de variable $y = xt \iff t = \frac{y}{x}$, nous obtenons :

$$\int_0^x (x-y)^{a-1} y^{b-1} dy = \int_0^1 x^{a-1} (1-t)^{a-1} x^{b-1} t^{b-1} x dt = x^{a+b-1} B(b, a)$$

avec, par définition de la fonction Bêta, $B(b, a) = \int_0^1 t^{b-1} (1-t)^{a-1} dt$. On sait que $B(b, a) = B(a, b)$ (il suffit de poser le changement de variable $s = 1-t$) donc l'égalité (15.4) se réécrit finalement :

$$\forall x \in \mathbb{R}, \quad g_{a,\lambda} * g_{b,\lambda}(x) = \frac{\lambda^{a+b}}{\Gamma(a)\Gamma(b)} e^{-\lambda x} x^{a+b-1} B(a, b) \mathbf{1}_{\mathbb{R}_+^*}(x)$$

Puisque $g_{a,\lambda} * g_{b,\lambda} \in C_M(\mathbb{R}, \mathbb{R})$, la Proposition 15.5.2 nous dit que cette application est une densité de probabilité. D'autre part, nous constatons que $g_{a,\lambda} * g_{b,\lambda}$ est proportionnelle à la densité $g_{a+b,\lambda}$. Nous en déduisons que ces densités sont égales : nous obtenons donc bien l'égalité (15.3). En outre, la constante de proportionnalité valant à la fois 1 et $\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} B(a, b)$, nous avons « en bonus » l'égalité : $B(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$. \square

Application 15.5.4 (Loi d'Erlang) Si X_1, \dots, X_n sont indépendantes de même loi $\mathcal{E}(\lambda)$, alors $X_1 + \dots + X_n$ admet la densité :

$$\frac{\lambda^n}{(n-1)!} x^{n-1} e^{-\lambda x} \mathbf{1}_{\mathbb{R}_+^*}(x)$$

(Cette densité définit la loi d'Erlang de paramètres n et λ : elle a été introduite historiquement pour étudier le nombre de clients contactant simultanément un centre d'appels téléphoniques.)

Démonstration: On vérifie que $\gamma(1, \lambda) = \mathcal{E}(\lambda)$ et la proposition précédente nous permet d'en déduire que $[\mathcal{E}(\lambda)]^{*n} = \gamma(n, \lambda)$ qui a bien la densité annoncée. \square

Application 15.5.5 (Loi $\chi^2(n)$) Si X_1, \dots, X_n sont indépendantes de même loi $\mathcal{N}(0, 1)$, alors $X_1^2 + \dots + X_n^2$ admet la densité :

$$\frac{1}{2^{n/2} \Gamma\left(\frac{n}{2}\right)} x^{n/2-1} e^{-x/2} \mathbf{1}_{\mathbb{R}_+^*}(x)$$

(Cette densité définit la loi du Khi-deux à n degrés de liberté, notée $\chi^2(n)$, qui permet de construire un test statistique très utilisé.)

Démonstration: Si $Y \sim \mathcal{N}(0, 1)$, alors $Y^2 \sim \gamma\left(\frac{1}{2}, \frac{1}{2}\right)$: cela se vérifie en calculant la fonction de répartition de Y^2 .

Comme les v.a.r. X_1^2, \dots, X_n^2 sont indépendantes, leur somme suit la loi $[\gamma\left(\frac{1}{2}, \frac{1}{2}\right)]^{*n} = \gamma\left(\frac{n}{2}, \frac{1}{2}\right)$ dont on reconnaît la densité. \square

15.6 Exercices

15.6.1 Calcul de densité marginale.

Soit (X, Y) un couple aléatoire à valeurs dans \mathbb{R}^2 , admettant pour densité :

$$p(x, y) = \frac{1}{\pi} \frac{ye^{-y}}{x^2 + y^2} \mathbf{1}_{\mathbb{R}_+^*}(y).$$

Vérifier que p est bien une densité sur \mathbb{R}^2 et calculer la densité marginale de la variable aléatoire réelle Y .

15.6.2 Densité gaussienne en dimension 2.

Soit (X, Y) un vecteur aléatoire à valeurs dans \mathbb{R}^2 de densité :

$$p(x, y) = \frac{1}{2\pi\sigma\tau\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{x^2}{\sigma^2} - 2\rho\frac{xy}{\sigma\tau} + \frac{y^2}{\tau^2}\right)\right\},$$

où σ et τ sont deux réels positifs strictement et $|\rho| < 1$.

Montrer, par un calcul de densités marginales, que :

$$X \sim \mathcal{N}(0, \sigma^2) \text{ et } Y \sim \mathcal{N}(0, \tau^2).$$

15.6.3 Loi uniforme sur le disque

[OUV¹ 223]

Soit (X, Y) un couple aléatoire qui suit la loi uniforme sur le disque $D(0, 1) = \{(x, y) \in \mathbb{R}^2, x^2 + y^2 \leq 1\}$. Calculer les densités marginales de X et Y et en déduire que ces deux variables aléatoires réelles ne sont pas indépendantes.

Remarque. Avec la notion de covariance qui apparaîtra dans la suite du cours, il est facile de montrer que $\text{Cov}(X, Y) = 0$: on dit que les variables X et Y sont décorrélées. Cet exercice nous fournit donc un exemple de variables décorrélées mais non indépendantes.

15.6.4 Gauss et Cauchy.

[FOA 201]

On considère deux v.a.r. indépendantes S et T de même loi $\mathcal{N}(0, 1)$.

Montrer que le couple $(\frac{S}{T}, T)$ admet une densité que l'on calculera.

En déduire, par calcul de densité marginale, que $\frac{S}{T}$ suit une loi de Cauchy de paramètre 1.

15.6.5 Avec des lois exponentielles.

[OUV² 75-77]

Soient $X \sim \text{Exp}(\lambda)$ et $Y \sim \text{Exp}(\mu)$ deux variables aléatoires réelles indépendantes ($\lambda > 0, \mu > 0$).

On définit les variables aléatoires réelles $U = \inf(X, Y)$ et $V = |X - Y|$.

Montrer que U et V sont indépendantes.

15.6.6 Méthode de Box-Muller

Soient X et Y deux variables aléatoires réelles indépendantes de même loi uniforme sur $]0, 1[$. Nous définissons :

$$(S, T) = (\sqrt{-2 \ln X} \cos(2\pi Y), \sqrt{-2 \ln X} \sin(2\pi Y))$$

Montrer que S et T sont deux variables indépendantes de même loi $\mathcal{N}(0, 1)$.

Corrigé. Montrons que nous définissons bien un C^1 -difféomorphisme de l'ouvert $U =]0, 1[^2$ sur l'ouvert $V = \mathbb{R}^2 \setminus (\mathbb{R}_+ \times \{0\})$ en posant :

$$\Phi(x, y) = (\sqrt{-2 \ln x} \cos(2\pi y), \sqrt{-2 \ln x} \sin(2\pi y))$$

En remarquant que $\rho(x) = \sqrt{-2 \ln x}$ et $\theta(y) = 2\pi y$ ne sont autre que les coordonnées polaires du point $\Phi(x, y)$, le lecteur constatera que Φ est bien une bijection de U sur V .

En outre, il est immédiat de vérifier que Φ est de classe C^1 . Il nous reste donc à prouver que Φ^{-1} est bien de classe C^1 mais nous rencontrons une difficulté pour expliciter cette bijection réciproque. En effet, si nous posons $\Phi(x, y) = (s, t) \in V$, nous calculons facilement

$$x = \exp\left(-\frac{s^2 + t^2}{2}\right)$$

mais il n'est pas possible de donner une formule pour y , sauf à « couper en morceaux » l'ensemble V . Nous préférons procéder indirectement en utilisant

un théorème de calcul différentiel.

Pour ce faire, nous calculons d'abord le déterminant jacobien de Φ , et non de Φ^{-1} comme d'habitude :

$$J_{\Phi}(x, y) = \begin{vmatrix} -\frac{\cos(2\pi y)}{x\sqrt{-2\ln x}} & -2\pi\sqrt{-2\ln x}\sin(2\pi y) \\ -\frac{\sin(2\pi y)}{x\sqrt{-2\ln x}} & 2\pi\sqrt{-2\ln x}\cos(2\pi y) \end{vmatrix} = -\frac{2\pi}{x}$$

Ce déterminant jacobien ne s'annulant jamais sur U , le *théorème d'inversion globale* affirme que Φ^{-1} est bien de classe C^1 sur V .

En outre, nous avons l'égalité :

$$J_{\Phi^{-1}}(s, t) = \frac{1}{J_{\Phi}(x, y)} = -\frac{x}{2\pi} = -\frac{1}{2\pi} \exp\left(-\frac{s^2 + t^2}{2}\right)$$

Nous pouvons maintenant appliquer la Proposition 15.3.5 pour obtenir la densité du couple (S, T) :

$$\frac{1}{2\pi} \exp\left(-\frac{s^2 + t^2}{2}\right) \mathbf{1}_V(s, t)$$

Si nous ôtons $\mathbf{1}_V(s, t)$ de cette formule, nous ne modifions la densité que sur la demi-droite $\mathbb{R}_+ \times \{0\}$, qui est de « surface nulle » : cela ne change donc rien aux valeurs des intégrales doubles dans lesquelles cette fonction intervient et il est donc correct de dire que (S, T) admet pour densité :

$$g(s, t) = \frac{1}{2\pi} \exp\left(-\frac{s^2 + t^2}{2}\right)$$

Cette fonction étant à variables séparables, la Proposition 15.4.3 nous permet de conclure.

Chapitre 16

Loi des grands nombres.

En ce qui concerne les *théorèmes limites* (c'est-à-dire les théorèmes qui énoncent des résultats asymptotiques) en calcul des probabilités, le programme comprend la loi faible des grands nombres avec sa démonstration par l'inégalité de Bienaymé-Tchebychev, ainsi que deux théorèmes admis : la loi forte des grands nombres et le théorème central limite. Nous traiterons ce dernier théorème, encore appelé *théorème-limite central*, et ses conséquences dans le chapitre suivant, en nous concentrant dans le présent chapitre sur les deux énoncés (faible et fort) de loi des grands nombres.

Introduction Un célèbre fabricant de piles électriques affirme : “elles durent deux fois plus longtemps”. Pour déterminer si c'est de la publicité mensongère ou non, il faudrait connaître l'espérance de vie d'une pile de ce type et la comparer à celle d'une pile “ordinaire”. Mais comment procéder dans la pratique pour évaluer, au moins approximativement, la quantité abstraite $E[X]$, où X est la durée de vie d'une pile en fonctionnement continu, que nous modélisons par une variable aléatoire réelle ?

Une idée qui vient naturellement à l'esprit est de tester indépendamment les unes des autres N piles du même type et, en notant $X_i(\omega)$ la durée de vie de la i -ème pile observée à la fin de cette expérience, de faire l'approximation suivante pour N suffisamment grand :

$$\frac{X_1(\omega) + \cdots + X_N(\omega)}{N} \approx E[X].$$

Pour justifier cette démarche, nous aurons besoin de théorèmes limites en probabilité. Le premier, appelé *loi des grands nombres*, montre que l'on a bien la convergence suivante dans un sens à préciser :

$$\lim_{n \rightarrow \infty} \frac{X_1 + \cdots + X_n}{n} = E[X].$$

Le second, appelé théorème central limite, précisera à quelle vitesse la convergence a lieu dans la loi des grands nombres, ce qui a une importance énorme dans les applications pratiques (statistique, méthodes de Monte-Carlo...)

Notons que pour donner un sens mathématique rigoureux à la limite précédente, il ne suffit pas d'avoir un "grand nombre" de variables aléatoires réelles X_1, \dots, X_N modélisant les durées de vie des piles testées mais une suite $(X_i)_{i \in \mathbb{N}^*}$ de variables aléatoires réelles. Notre hypothèse de base est que les durées de vie des différentes piles sont indépendantes les unes des autres mais qu'elles ont toute le même comportement aléatoire, plus précisément la même loi de probabilité. Avant d'aborder les théorèmes limites, nous aurons donc besoin d'étudier quelques propriétés d'une suite $(X_i)_{i \in \mathbb{N}^*}$ de variables aléatoires réelles indépendantes ayant toutes même loi.

Dans toute la suite de ce chapitre, les variables aléatoires considérées sont définies sur un espace de probabilité (Ω, \mathcal{A}, P) .

16.1 Suite indépendante équadistribuée.

Définition 16.1.1 Une suite de variables aléatoires réelles $(X_i)_{i \in \mathbb{N}^*}$ est dite indépendante si pour tout $k \in \mathbb{N}^*$ et pour tous indices $1 \leq i_1 < i_2 < \dots < i_k$, les variables aléatoires réelles X_{i_1}, \dots, X_{i_k} sont indépendantes.

Une suite de variables aléatoires réelles $(X_i)_{i \in \mathbb{N}^*}$ est dite identiquement distribuée (ou équadistribuée) si toutes les variables aléatoires réelles $X_i, i \in \mathbb{N}^*$ ont même loi de probabilité.

Une suite $(X_i)_{i \in \mathbb{N}^*}$ possédant ces deux propriétés est donc dite indépendante identiquement distribuée (i.i.d. en abrégé).

Nous retrouverons cette hypothèse tout au long de ce chapitre. Elle correspond à une expérience répétée à l'infini dans des conditions identiques (bien sûr, dans la pratique, elle n'est répétée qu'un "grand nombre" de fois).

Par exemple, un jeu de pile ou face répété à l'infini est modélisé par une suite i.i.d. de variables de Bernoulli de paramètre p .

Proposition 16.1.2 Soit $(X_i)_{i \in \mathbb{N}^*}$ une suite i.i.d. et $f \in C(\mathbb{R}, \mathbb{R})$. Alors la suite $(f(X_i))_{i \in \mathbb{N}^*}$ est encore i.i.d.

Démonstration: Montrons d'abord que la suite $(f(X_i))_{i \in \mathbb{N}^*}$ est indépendante, c'est-à-dire que pour tout $k \in \mathbb{N}^*$ et tous indices $1 \leq i_1 < \dots < i_k$, les variables aléatoires réelles $f(X_{i_1}), \dots, f(X_{i_k})$ sont indépendantes.

Nous considérons donc des intervalles ouverts réels arbitraires I_1, \dots, I_k et nous calculons la probabilité suivante :

$$P[f(X_{i_1}) \in I_1, \dots, f(X_{i_k}) \in I_k] = P[X_{i_1} \in f^{-1}(I_1), \dots, X_{i_k} \in f^{-1}(I_k)].$$

Remarquons que les sous-ensembles réels $f^{-1}(I_1), \dots, f^{-1}(I_k)$ sont ouverts en tant qu'images réciproques d'ouverts par une application continue ; ce sont donc des éléments de la tribu borélienne $\mathcal{B}(\mathbb{R})$. Or, par hypothèse, les variables X_{i_1}, \dots, X_{i_k} sont indépendantes, si bien que nous pouvons encore écrire la probabilité précédente sous forme du produit :

$$P[X_{i_1} \in f^{-1}(I_1)] \cdots P[X_{i_k} \in f^{-1}(I_k)] = P[f(X_{i_1}) \in I_1] \cdots P[f(X_{i_k}) \in I_k],$$

ce qui nous permet de conclure.

Passons à la démonstration de l'équidistribution de la suite $(f(X_i))_{i \in \mathbb{N}^*}$.

En notant P_{X_i} la loi de probabilité de la variable aléatoire réelle X_i , nous avons par hypothèse :

$$\forall i \in \mathbb{N}^* \quad P_{X_i} = P_{X_1}.$$

Pour tout $i \in \mathbb{N}^*$ et tout intervalle ouvert réel I , nous avons donc :

$$P[f(X_i) \in I] = P[X_i \in f^{-1}(I)] = P_{X_i}[f^{-1}(I)] = P_{X_1}[f^{-1}(I)] = P[f(X_1) \in I],$$

ce qui s'écrit encore :

$$P_{f(X_i)}(I) = P_{f(X_1)}(I).$$

Une loi de probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ étant entièrement déterminée par ses valeurs sur les intervalles ouverts, nous avons ainsi démontré :

$$\forall i \in \mathbb{N}^* \quad P_{f(X_i)} = P_{f(X_1)}.$$

□

Remarque. Si nous avons considéré une suite $(f_i)_{i \in \mathbb{N}^*}$ d'applications continues de \mathbb{R} dans \mathbb{R} , alors la suite $(f_i(X_i))_{i \in \mathbb{N}^*}$ serait encore indépendante mais n'aurait aucune raison d'être identiquement distribuée. Le lecteur pourra adapter la première partie de la démonstration précédente à titre d'exercice. Il est très facile de trouver un contre-exemple en ce qui concerne l'équidistribution.

Notations : Dans la suite, nous noterons $L^1(\Omega, \mathcal{A}, P)$ l'ensemble des variables aléatoires réelles définies sur l'espace de probabilité (Ω, \mathcal{A}, P) et admettant une espérance.

De même, nous noterons $L^2(\Omega, \mathcal{A}, P)$ l'ensemble des variables aléatoires réelles X définies sur l'espace de probabilité (Ω, \mathcal{A}, P) et admettant un moment d'ordre deux, c'est-à-dire telles que X^2 admet une espérance.

Nous ferons l'abus de langage courant consistant à appeler *variable de variance finie* une variable admettant un moment d'ordre deux. C'est un abus

de langage parce que si X n'est pas une variable de variance finie, cela ne signifie pas qu'elle a une variance infinie mais que sa variance n'est pas définie du tout ! C'est le cas par exemple si X suit une loi de Cauchy.

On démontre l'inclusion $L^2(\Omega, \mathcal{A}, P) \subset L^1(\Omega, \mathcal{A}, P)$ en utilisant l'inégalité $|x| \leq x^2 + 1$ valable pour tout réel x .

Proposition 16.1.3 *Soit $(X_i)_{i \in \mathbb{N}^*}$ une suite identiquement distribuée (mais pas nécessairement indépendante).*

Alors nous avons l'équivalence suivante :

$$[\forall i \in \mathbb{N}^* \quad X_i \in L^1(\Omega, \mathcal{A}, P)] \iff X_1 \in L^1(\Omega, \mathcal{A}, P).$$

Nous dirons alors que la suite (X_i) est dans $L^1(\Omega, \mathcal{A}, P)$ et, dans ce cas, toutes les variables aléatoires réelles ont la même espérance $m = E[X_i]$, $i \in \mathbb{N}^$.*

De même, nous avons l'équivalence :

$$[\forall i \in \mathbb{N}^* \quad X_i \in L^2(\Omega, \mathcal{A}, P)] \iff X_1 \in L^2(\Omega, \mathcal{A}, P).$$

Nous dirons alors que la suite (X_i) est dans $L^2(\Omega, \mathcal{A}, P)$ et, dans ce cas, toutes les variables aléatoires réelles ont la même variance $\sigma^2 = \text{Var}X_i$, $i \in \mathbb{N}^$.*

Démonstration: Nous ne ferons la démonstration que dans le cas où la loi commune aux variables X_i admet une densité de probabilité g .

Le lecteur pourra à titre d'exercice faire la preuve dans le cas où cette loi commune est discrète. Nous admettrons cette proposition dans les autres cas. D'après le théorème de transfert, nous avons pour tout $i \in \mathbb{N}^*$:

$$E[|X_i|] = \int_{-\infty}^{+\infty} |x|g(x) dx = E[|X_1|],$$

ce qui nous donne la première équivalence.

En outre, si l'intégrale impropre précédente est convergente, ce qui équivaut à :

$$\forall i \in \mathbb{N}^* \quad X_i \in L^1(\Omega, \mathcal{A}, P),$$

le théorème de transfert nous donne encore :

$$E[X_i] = \int_{-\infty}^{+\infty} x g(x) dx = E[X_1]$$

et nous posons $m := E[X_1]$.

De façon similaire, nous obtenons la deuxième équivalence en écrivant, pour tout $i \in \mathbb{N}^*$:

$$E[X_i^2] = \int_{-\infty}^{+\infty} x^2 g(x) dx = E[X_1^2].$$

Si cette espérance est finie, en gardant la notation $m = E[X_i]$ (qui a un sens puisque $L^2(\Omega, \mathcal{A}, P) \subset L^1(\Omega, \mathcal{A}, P)$), nous avons pour tout $i \in \mathbb{N}^*$:

$$\text{Var}X_i = E[X_i^2] - m^2 = E[X_1^2] - m^2 = \text{Var}X_1 = \sigma^2.$$

□

16.2 Convergences de suites de variables aléatoires réelles

16.2.1 Événements négligeables, égalité presque sûre.

Définition 16.2.1 Un événement $A \in \mathcal{A}$ est dit négligeable si $P(A) = 0$.

En termes courants, un événement négligeable n'a donc "aucune chance de se produire", il est "extrêmement improbable" mais il ne faudrait pas en conclure pour autant qu'il est impossible : seul l'événement \emptyset est qualifié d'impossible. Pour illustrer cette distinction, prenons l'exemple d'un nombre réel tiré au hasard uniformément entre 0 et 1. Nous pouvons modéliser cette expérience en prenant $\Omega = [0, 1]$, $\mathcal{A} = \mathcal{B}([0, 1])$, qui est la plus petite tribu sur Ω contenant tous les sous-intervalles de $[0, 1]$, et en prenant pour P la *mesure de Lebesgue*, qui est (nous admettons son existence et son unicité) l'unique probabilité sur (Ω, \mathcal{A}) telle que :

$$\forall [a, b] \subset [0, 1] \quad P([a, b]) = b - a.$$

Soit $x \in [0, 1]$ arbitraire fixé. Nous constatons que l'événement $A = \{x\} = [x, x]$ est négligeable. Il n'est néanmoins pas impossible qu'un nombre tiré au hasard entre 0 et 1 soit égal à x .

Proposition 16.2.2 Toute union finie ou dénombrable d'événements négligeables est encore un événement négligeable.

Démonstration: Soit $(A_n)_{n \in \mathbb{N}} \in \mathcal{A}^{\mathbb{N}}$ une suite d'événements tels que :

$$\forall n \in \mathbb{N} \quad P(A_n) = 0.$$

Par propriété de sous- σ -additivité de la mesure de probabilité P (cf. formule (12.2) page 218), nous avons alors :

$$P\left(\bigcup_{n \in \mathbb{N}} A_n\right) \leq \sum_{n \in \mathbb{N}} P(A_n) = 0,$$

d'où la conclusion. □

Remarque. Cela n'a plus aucune raison d'être vrai pour une union quelconque. Ainsi, si nous reprenons l'exemple précédent, nous avons :

$$\Omega = \bigcup_{x \in [0,1]} \{x\},$$

qui n'est évidemment pas négligeable.

Définition 16.2.3 *Un événement A est dit presque sûr si l'événement complémentaire A^c est négligeable.*

Deux variables aléatoires réelles X et Y sont dites égales presque sûrement si l'événement $\{\omega \in \Omega, X(\omega) = Y(\omega)\}$ est presque sûr.

Un événement presque sûr est donc un événement de probabilité 1.

Deux variables aléatoires réelles sont égales presque sûrement si et seulement si l'événement $\{X \neq Y\}$ est négligeable. À titre d'exercice, le lecteur pourra en déduire que l'égalité presque sûre est une relation d'équivalence sur l'ensemble des variables aléatoires réelles définies sur (Ω, \mathcal{A}, P) .

16.2.2 Convergences en probabilité et presque sûre.

Nous définissons maintenant deux types de convergence concernant les suites de variables aléatoires réelles.

Définition 16.2.4 *Soient $(Y_n)_{n \in \mathbb{N}^*}$ et Y des variables aléatoires réelles définies sur un espace de probabilité (Ω, \mathcal{A}, P) .*

On dit que la suite (Y_n) converge presque sûrement vers Y et l'on note $Y_n \xrightarrow{p.s.} Y$ s'il existe un événement négligeable $N \in \mathcal{A}$ tel que :

$$\forall \omega \notin N \quad Y_n(\omega) \longrightarrow Y(\omega).$$

On dit que la suite (Y_n) converge en probabilité vers Y et l'on note $Y_n \xrightarrow[n \rightarrow +\infty]{(P)} Y$ si :

$$\forall \epsilon > 0 \quad P(|Y_n - Y| > \epsilon) \longrightarrow 0.$$

Proposition 16.2.5 *Une limite en probabilité est unique modulo l'égalité presque sûre.*

Démonstration: Supposons que la suite (Y_n) converge en probabilité à la fois vers Y et Z . Pour tout $\epsilon > 0$, nous écrivons alors l'inclusion (facile à vérifier en passant aux événements complémentaires et en appliquant l'inégalité triangulaire) :

$$\{|Y - Z| > \epsilon\} \subset \{|Y_n - Y| > \frac{\epsilon}{2}\} \cup \{|Y_n - Z| > \frac{\epsilon}{2}\}.$$

Par passage aux probabilités puis en faisant tendre n vers l'infini, nous en déduisons : $P(|Y - Z| > \epsilon) = 0$.

Comme $\epsilon > 0$ était arbitraire, nous en déduisons que pour tout $p \in \mathbb{N}^*$, $P(|Y - Z| > \frac{1}{p}) = 0$.

Nous concluons alors en écrivant :

$$P(Y \neq Z) = P\left(\lim \uparrow \left\{ |Y - Z| > \frac{1}{p} \right\}\right) = \lim \uparrow P\left(|Y - Z| > \frac{1}{p}\right) = 0.$$

□

Il serait facile de démontrer directement qu'une limite au sens de la convergence presque sûre est unique modulo l'égalité presque sûre (vous pouvez le faire à titre d'exercice) mais cela va résulter de la proposition suivante.

Proposition 16.2.6 *Si (Y_n) converge presque sûrement vers Y , alors (Y_n) converge en probabilité vers Y .*

Démonstration: Le lecteur est invité à lire [DAN 452].

□

Remarque. La réciproque est fautive : le lecteur pourra s'en convaincre en étudiant le contre-exemple suivant.

Nous définissons une suite $(I_n)_{n \in \mathbb{N}^*}$ de sous-intervalles de $[0, 1]$ comme suit :

$$\forall k \in \mathbb{N} \quad \forall j \in \{0, 1, \dots, 2^k - 1\} \quad I_{2^{k+j}} = \left] \frac{j}{2^k}, \frac{j+1}{2^k} \right].$$

Nous prenons alors pour espace de probabilité $(\Omega, \mathcal{A}, P) = ([0, 1], \mathcal{B}([0, 1]), \lambda_{[0,1]})$ et nous définissons pour tout $n \in \mathbb{N}^*$ la variable aléatoire $Y_n = \mathbf{1}_{I_n}$.

Nous constatons alors que $Y_n \xrightarrow[n \rightarrow +\infty]{(P)} 0$ mais que la suite (Y_n) ne converge pas presque sûrement puisque :

$$\forall \omega \in \Omega \quad \limsup Y_n(\omega) = 1 \text{ et } \liminf Y_n(\omega) = 0.$$

16.3 Loi des grands nombres

Théorème 16.3.1 (Loi faible des grands nombres) Soit $(X_i)_{i \in \mathbb{N}^*}$ une suite indépendante dans $L^2(\Omega, \mathcal{A}, P)$. Nous supposons que les variables X_i , $i \in \mathbb{N}^*$ ont toutes même espérance, notée m , et même variance. Pour tout $n \in \mathbb{N}^*$, nous définissons la moyenne empirique d'ordre n :

$$\bar{X}_n = \frac{X_1 + \cdots + X_n}{n}.$$

Alors nous avons la convergence en probabilité suivante :

$$\bar{X}_n \xrightarrow[n \rightarrow +\infty]{(P)} m.$$

Démonstration: Pour tout $n \in \mathbb{N}^*$, définissons la variable $S_n = \sum_{i=1}^n X_i$ de sorte que $\bar{X}_n = S_n/n$. Comme $L^2(\Omega, \mathcal{A}, P)$ est un espace vectoriel, nous avons $S_n \in L^2(\Omega, \mathcal{A}, P)$ et nous calculons immédiatement $E[S_n] = nm$ puis, en utilisant l'indépendance des variables X_i ,

$$\text{Var}S_n = \sum_{i=1}^n \text{Var}X_i = n\sigma^2,$$

où l'on a noté σ^2 la variance commune aux variables X_i .

Nous en déduisons que, pour tout $n \in \mathbb{N}^*$, $E[\bar{X}_n] = m$ et $\text{Var}\bar{X}_n = \frac{\sigma^2}{n}$.

Pour tout $\epsilon > 0$, l'inégalité de Bienaymé-Tchebychev nous permet alors d'écrire :

$$P(|\bar{X}_n - m| > \epsilon) \leq \frac{1}{\epsilon^2} \text{Var}\bar{X}_n = \frac{\sigma^2}{n\epsilon^2} \rightarrow 0,$$

d'où la conclusion. □

Remarques.

1. La loi des grands nombres nous dit que, conformément à notre intuition, la moyenne empirique (observable après l'expérience) converge vers la moyenne probabiliste théorique, c'est-à-dire l'espérance.
2. Ce théorème s'applique en particulier à toute suite (X_i) i.i.d. dans $L^2(\Omega, \mathcal{A}, P)$. Au prix d'une démonstration plus difficile, on peut démontrer alors qu'il se généralise au cas d'une suite (X_i) i.i.d. dans $L^1(\Omega, \mathcal{A}, P)$. Cependant, le théorème suivant va nous conduire à une conclusion encore plus forte sous la même hypothèse d'indépendance et équidistribution dans $L^1(\Omega, \mathcal{A}, P)$.

Théorème 16.3.2 (Loi forte des grands nombres, Kolmogorov) *Soit $(X_i)_{i \in \mathbb{N}^*}$ une suite indépendante identiquement distribuée. Alors il y a équivalence entre les deux conditions suivantes :*

1. *La suite (\bar{X}_n) converge presque sûrement vers une certaine variable aléatoire X à valeurs dans \mathbb{R} .*
2. *La suite (X_i) est dans $L^1(\Omega, \mathcal{A}, P)$.*

En supposant ces conditions satisfaites et en notant m l'espérance commune aux variables X_i , nous avons alors $X = m$ presque sûrement, c'est-à-dire :

$$\bar{X}_n \xrightarrow{p.s.} m = E[X_i].$$

Ce théorème, que nous admettons, porte le nom de loi *forte* des grands nombres parce que la convergence presque sûre est plus forte que la convergence en probabilité qui apparaissait dans le théorème précédent.

En outre, il nous donne une condition *nécessaire et suffisante* de convergence presque sûre de la suite des moyennes empiriques.

Chapitre 17

Loi normale

17.1 La loi normale centrée réduite

Proposition 17.1.1 (Intégrale de Gauss)

$$\int_{-\infty}^{+\infty} e^{-x^2} dx = \sqrt{\pi}$$

[GOURDON 163,335] démontre cette égalité par 3 méthodes différentes :

- dérivation d’une intégrale dépendant d’un paramètre
- inégalité de convexité, changement de variables et formule de Wallis
- passage en coordonnées polaires

□

Nous suivons maintenant la présentation de Foata et Fuchs dans leur ouvrage “Calcul des probabilités” chez Dunod, 2ème édition, pages 178 à 181.

Proposition et définition 17.1.2 *L’application $g : \mathbb{R} \rightarrow \mathbb{R}_+$ définie par :*

$$\forall x \in \mathbb{R} \quad g(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

est une densité de probabilité. La loi de densité g est appelée loi de Gauss ou loi normale (centrée réduite) et notée $\mathcal{N}(0, 1)$.

Gauss a introduit cette loi en 1809 à propos d’un problème statistique d’estimation de paramètre.

Le graphe de la densité de Gauss g est une «courbe en cloche» aplatie (voir figure page ??). On le trace facilement en notant que f est paire, qu’elle

admet un maximum global en $x = 0$ égal à $\frac{1}{\sqrt{2\pi}} \simeq 0,399$ et que la courbe admet deux points d'inflexion en $x = -1$ et $x = 1$.

Si $X \sim \mathcal{N}(0, 1)$, on calcule immédiatement $E[X] = 0$ par le théorème de transfert en utilisant la parité de g et

$$\text{Var}X = E[X^2] = \int_{-\infty}^{+\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 1$$

par intégration par parties. C'est pourquoi on parle de loi normale *centrée réduite*.

La fonction de répartition de la loi de Gauss est donnée par

$$\forall x \in \mathbb{R} \quad \Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

On vérifie facilement que Φ est de classe C^∞ , strictement croissante et donc bijective de \mathbb{R} sur son image $]0, 1[$, telle que $\Phi(-x) = 1 - \Phi(x)$ pour tout $x \in \mathbb{R}$.

Son graphe est une «courbe en S» assez étalée, symétrique par rapport à $(0, 1/2)$, où elle admet un point d'inflexion et où la pente de sa tangente vaut $\frac{1}{\sqrt{2\pi}} \simeq 0,399$.

On n'a pas de formule plus explicite pour Φ mais cette fonction et sa bijection réciproque Φ^{-1} sont tabulées, notamment pour des applications statistiques. Une valeur que nous utiliserons dans la suite est $\Phi(1,96) \approx 0,975$.

Si $X \sim \mathcal{N}(0, 1)$, alors $P(|X| \leq x) = 2\Phi(x) - 1$ pour tout $x \geq 0$, de sorte que $P(|X| \leq 1,96) \approx 0,95$.

On peut démontrer l'équivalence suivante lorsque $x \rightarrow +\infty$:

$$1 - \Phi(x) \sim \frac{1}{x} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$$

Comme cela tend «vite» vers 0, on dit que la loi de Gauss a une queue de distribution peu épaisse.

Proposition 17.1.3 *La loi de Gauss admet des moments de tous les ordres. Si $X \sim \mathcal{N}(0, 1)$, nous avons, pour tout $n \in \mathbb{N}$:*

$$E[X^{2n+1}] = 0 \quad ; \quad E[X^{2n}] = \frac{(2n)!}{2^n n!}$$

Démonstration: Pour tout $k \in \mathbb{N}^*$, nous avons $|x|^k g(x) = o(1/x^2)$ quand $x \rightarrow \pm\infty$ donc X admet un moment d'ordre k donné par :

$$E[X^k] = \int_{-\infty}^{+\infty} x^k \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dt$$

Lorsque $k = 2n + 1$, cette intégrale est nulle pour des raisons de parité.

On démontre par intégration par parties que $E[X^{2n}] = (2n - 1)E[X^{2n-2}]$ pour tout $n \in \mathbb{N}^*$, ce qui nous donne par récurrence la formule annoncée puisque $E[X^0] = 1$ par convention. □

Pour simuler numériquement la loi de Gauss, on utilise la proposition suivante, qui se démontre grâce à un changement de variable dans une intégrale double. [OUV² 67-68]

Proposition 17.1.4 (Méthode de Box-Muller) Soient U_1 et U_2 deux v.a.r. indépendantes, de même loi uniforme sur $[0,1]$. Nous posons :

$$X := \sqrt{-2 \ln U_1} \cos(2\pi U_2) \quad , \quad Y := \sqrt{-2 \ln U_1} \sin(2\pi U_2).$$

Alors, les v.a.r. X et Y sont indépendantes et de même loi $\mathcal{N}(0,1)$.

17.2 La loi normale générale

Proposition et définition 17.2.1 Soient $\mu \in \mathbb{R}$ et $\sigma > 0$ deux paramètres fixés. L'application $g_{\mu,\sigma^2} : \mathbb{R} \rightarrow \mathbb{R}_+$ définie par :

$$\forall x \in \mathbb{R} \quad g_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

est une densité de probabilité. La loi de densité g_{μ,σ^2} est appelée loi normale de paramètres (μ, σ^2) et notée $\mathcal{N}(\mu, \sigma^2)$.

On vérifie que g_{μ,σ^2} est bien une densité en effectuant un changement de variable affine dans une intégrale simple, puis en utilisant la proposition 17.1.2. Notons que pour $\mu = 0$ et $\sigma^2 = 1$, on retrouve bien la densité de la loi $\mathcal{N}(0,1)$ telle qu'elle a été définie dans la sous-section précédente. Nous allons maintenant préciser les rapports entre la loi normale générale et la loi normale centrée réduite.

Proposition 17.2.2 Soient $\mu \in \mathbb{R}$ et $\sigma > 0$ deux paramètres fixés. On considère deux variables aléatoires réelles X et Y telles que $Y = \mu + \sigma X$. Alors on a l'équivalence suivante :

$$X \sim \mathcal{N}(0,1) \iff Y \sim \mathcal{N}(\mu, \sigma^2)$$

Démonstration: On vérifie facilement que cette équivalence résulte de la proposition 14.3.2. □

Pour simuler numériquement la loi $\mathcal{N}(\mu, \sigma^2)$, on applique d'abord la méthode de Box-Muller pour simuler $X \sim \mathcal{N}(0, 1)$ puis l'on calcule $Y = \mu + \sigma X$.

Corollaire 17.2.3 Soit $Y \sim \mathcal{N}(\mu, \sigma^2)$. Alors $E[Y] = \mu$ et $\text{Var}Y = \sigma^2$

Ce résultat justifie que le premier paramètre d'une loi normale soit traditionnellement noté μ ou m (comme «moyenne» probabiliste, c'est-à-dire espérance) et le second σ^2 puisqu'il est égal à la variance de la loi.

Corollaire 17.2.4 La fonction de répartition Φ_{μ, σ^2} de la loi $\mathcal{N}(\mu, \sigma^2)$ est donnée par :

$$\forall x \in \mathbb{R} \quad \Phi_{\mu, \sigma^2}(x) = \Phi\left(\frac{x - \mu}{\sigma}\right)$$

Démonstration: En reprenant les notations de la proposition 17.2.2, nous avons, pour tout $x \in \mathbb{R}$:

$$\Phi_{\mu, \sigma^2}(x) = P(Y \leq x) = P(\mu + \sigma X \leq x) = P\left(X \leq \frac{x - \mu}{\sigma}\right) = \Phi\left(\frac{x - \mu}{\sigma}\right)$$

□

La proposition suivante énonce la stabilité de la loi normale par produit de convolution.

Proposition 17.2.5 Soient X et Y deux v.a.r. indépendantes de lois respectives $\mathcal{N}(m_1, \sigma_1^2)$ et $\mathcal{N}(m_2, \sigma_2^2)$, avec $(m_1, m_2) \in \mathbb{R}^2$ et $(\sigma_1, \sigma_2) \in (\mathbb{R}_+^*)^2$. Alors la v.a.r. $X + Y$ suit la loi $\mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$.

Le résultat essentiel établi par cette proposition est que la somme de deux variables gaussiennes indépendantes est encore gaussienne. Il est alors facile de déterminer ses paramètres en calculant l'espérance et la variance.

Démonstration: Commençons par réduire le problème au cas où X est centrée réduite et Y centrée en supposant démontrée la proposition suivante : si $X' \sim \mathcal{N}(0, 1)$ et $Y' \sim \mathcal{N}(0, s^2)$, avec $s > 0$, sont indépendantes, alors $X' + Y' \sim \mathcal{N}(0, 1 + s^2)$. Nous pouvons alors en déduire la proposition dans le cas général en définissant les variables aléatoires réelles :

$$X' = \frac{X - m_1}{\sigma_1} \quad , \quad Y' = \frac{Y - m_2}{\sigma_1} \quad ,$$

si bien que les hypothèses précédentes sont satisfaites avec $s = \sigma_2/\sigma_1$ et :

$$X + Y = m_1 + \sigma_1 X' + m_2 + \sigma_1 Y' = m_1 + m_2 + \sigma_1 (X' + Y').$$

Par transformation affine de la variable $X' + Y' \sim \mathcal{N}(0, 1 + (\sigma_2^2/\sigma_1^2))$, nous obtenons alors la conclusion dans le cas général.

Nous supposons donc désormais $m_1 = m_2 = 0$, $\sigma_1 = 1$ et $\sigma_2 = s > 0$. D'après la Proposition 15.5.2, page 274, la variable aléatoire réelle $X + Y$ admet une densité proportionnelle à :

$$\int_{-\infty}^{+\infty} e^{-\frac{(x-y)^2}{2}} e^{-\frac{y^2}{2s^2}} dy = \int_{-\infty}^{+\infty} \exp\left(-\frac{s^2 x^2 - 2s^2 xy + (1+s^2)y^2}{2s^2}\right) dy.$$

Cette dernière expression s'écrit encore :

$$\int_{-\infty}^{+\infty} \exp\left(-\frac{(1+s^2)\left(y - \frac{s^2}{1+s^2}x\right)^2}{2s^2}\right) \exp\left(\left(\frac{s^2}{1+s^2} - 1\right)\frac{x^2}{2}\right) dy,$$

ou encore, puisque l'intégration porte sur la variable y ,

$$\exp\left(-\frac{x^2}{2(1+s^2)}\right) \int_{-\infty}^{+\infty} \exp\left(-\frac{(1+s^2)\left(y - \frac{s^2}{1+s^2}x\right)^2}{2s^2}\right) dy.$$

Nous constatons que cette dernière intégrale est une constante en effectuant, à x fixé, le simple changement de variable $u = y - \frac{s^2}{1+s^2}x$.

Nous avons donc prouvé que la variable aléatoire réelle $X + Y$ admet une densité proportionnelle à $\exp(-x^2/(2(1+s^2)))$. Nous en déduisons que la seule constante de proportionnalité possible pour que ce soit effectivement une densité de probabilité vaut $1/\sqrt{2\pi(1+s^2)}$ et que $X + Y \sim \mathcal{N}(0, 1 + s^2)$. \square

Pour une présentation de la loi normale dans \mathbb{R}^d , encore appelée loi de Laplace-Gauss en dimension d , le lecteur pourra lire [ESC 147ss] qui en fait une présentation cadrant bien avec le programme officiel.

17.3 Approximation normale de la loi binomiale

Historiquement, Abraham de Moivre, mathématicien anglais d'origine française, avait mis en évidence dès 1733 une approximation normale de la loi binomiale en étudiant un modèle de pile ou face avec une pièce équilibrée.

Sa preuve reposait sur des calculs laborieux d'estimation des coefficients binomiaux.

Pierre-Simon de Laplace avait généralisé ce résultat en 1820 au cas d'une pièce éventuellement biaisée. Il avait également mis en évidence le lien entre la loi normale et la théorie des erreurs d'observation, dont nous reparlerons dans le chapitre suivant.

Nous suivons ici la présentation d'[OUVRARD tome 1, 228-229].

Théorème 17.3.1 (de Moivre, Laplace) *Considérons $p \in]0, 1[$ et, pour tout $n \in \mathbb{N}^*$, une variable aléatoire réelle S_n qui suit la loi binomiale $B(n, p)$. Nous noterons \tilde{S}_n la variable centrée réduite associée à S_n , i.e.*

$$\tilde{S}_n = \frac{S_n - E[S_n]}{\sigma(S_n)} = \frac{S_n - np}{\sqrt{np(1-p)}}.$$

Alors nous avons la convergence suivante, uniforme sur tous les intervalles réels I :

$$\sup_I \left| P(\tilde{S}_n \in I) - \int_I \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt \right| \xrightarrow{n \rightarrow +\infty} 0.$$

Autrement dit, quand n est «grand», \tilde{S}_n suit une loi approximativement égale à la loi normale centrée réduite, ce que nous notons $\tilde{S}_n \approx \mathcal{N}(0, 1)$.

En utilisant un changement de variable affine dans une intégrale simple, nous pouvons dire de façon équivalente que $S_n \approx \mathcal{N}(np, np(1-p))$, toujours pour n «grand». Plus précisément, nous avons le résultat suivant :

Corollaire 17.3.2 *Considérons $p \in]0, 1[$ et, pour tout $n \in \mathbb{N}^*$, une variable aléatoire réelle S_n qui suit la loi binomiale $B(n, p)$. Alors nous avons la convergence suivante, uniforme sur tous les intervalles réels I :*

$$\sup_I \left| P(S_n \in I) - \int_I \frac{1}{\sqrt{2\pi np(1-p)}} \exp\left(-\frac{(t - np)^2}{2np(1-p)}\right) dt \right| \xrightarrow{n \rightarrow +\infty} 0.$$

Pour pouvoir utiliser cette approximation gaussienne de la loi binomiale, il nous reste à préciser ce que signifie n «grand» ! Le théorème de Berry¹-Esseen², dont nous donnerons l'énoncé général dans le chapitre suivant, met en évidence le rôle joué par la quantité $np(1-p)$ dans la qualité de l'approximation.

1. Andrew C. Berry.

2. Carl-Gustav Esseen (1918-2001), mathématicien suédois.

Proposition 17.3.3 (Berry-Esseen, cas binomial) *Nous reprenons les hypothèses et les notations du théorème 17.3.1 et nous notons en outre $F_{\tilde{S}_n}$ la fonction de répartition de la variable \tilde{S}_n et Φ la fonction de répartition de la loi $\mathcal{N}(0, 1)$. Nous avons alors, pour tout $n \in \mathbb{N}^*$:*

$$\sup_{x \in \mathbb{R}} |F_{\tilde{S}_n}(x) - \Phi(x)| \leq \frac{p^2 + (1-p)^2}{\sqrt{np(1-p)}}.$$

Si $np(1-p)$ est suffisamment grand, le théorème de Berry-Esseen garantit que l'approximation de la loi binomiale par la loi normale sera bonne. Dans la pratique, certaines règles empiriques existent, variables d'ailleurs d'un ouvrage à l'autre. Voici celle qui est adoptée par Dacunha-Castelle et Duflo :

$$B(n, p) \approx \mathcal{N}(np, np(1-p)) \text{ dès que } np(1-p) > 18.$$

Voici une autre règle, adoptée dans le programme officiel de probabilités au lycée.

$$B(n, p) \approx \mathcal{N}(np, np(1-p)) \text{ dès que } n \geq 30, np \geq 5 \text{ et } n(1-p) \geq 5.$$

Si les conditions précédentes ne sont pas satisfaites, on pourra envisager l'approximation de la loi binomiale par la loi de Poisson. Voici une règle empirique énoncée par OUVRARD :

$$B(n, p) \approx \mathcal{P}(np) \text{ dès que } n \geq 30 \text{ et } p \leq 0,1.$$

Notons que si $p \geq 0,9$, l'approximation poissonnienne est aussi utilisable en comptant les échecs à la place des succès (ou les faces au lieu des piles!).

Le lecteur pourra trouver des applications en statistique de l'approximation gaussienne d'une loi binomiale dans [MEY 131-132] : Exemple 3.7 (intervalle de confiance dans un sondage) et Exercice 2 (intervalle de fluctuation pour le nombre de clients se connectant à un serveur informatique).

Chapitre 18

Le théorème-limite central

Ce théorème-limite est considéré comme central en théorie des probabilités pour les deux raisons suivantes :

1. Il précise à quelle vitesse a lieu la convergence énoncée par la loi des grands nombres, ce qui a une importance cruciale pour les applications (intervalle de confiance en statistique, méthode de Monte-Carlo pour le calcul approché d'une intégrale etc.)
2. Il permet de répondre à la question suivante : Pourquoi les lois normales (ou gaussiennes) jouent-elles un rôle central dans la modélisation de phénomènes aussi divers qu'une erreur de mesure en Physique, une donnée psychométrique en Psychologie différentielle, une fluctuation dans l'évolution d'un cours de la bourse en Finance ou de la taille d'une population animale en Biologie ?
Pourquoi le champ d'application des lois de Gauss est-il si large qu'on les a qualifiées de lois *normales* ? Après tout, la densité gaussienne n'a pas une forme si simple que cela ; pourquoi donc la considérer comme plus normale que les autres ?

Le théorème-limite central (ou *théorème central limite* selon la terminologie du programme officiel) nous dit *en gros* que **si l'on additionne beaucoup de variables indépendantes de variances finies, alors le résultat est une variable approximativement gaussienne.**

La précision de ce résultat sous des hypothèses aussi faibles est surprenante car si l'on sait qu'une loi est gaussienne – i.e. de la forme $\mathcal{N}(m, \sigma^2)$ – alors on dispose de procédures statistiques qui permettent d'estimer les deux paramètres m et σ^2 et donc d'identifier complètement la loi (approximative) du phénomène étudié. Or notre hypothèse de départ sur les lois des variables indépendantes que l'on additionne est très faible : on demande simplement qu'elle soient de variances finies.

18.1 Énoncé classique

Conformément au programme officiel, nous admettons ce théorème.

Théorème 18.1.1 (Théorème-limite central) *Soit $(X_i)_{i \in \mathbb{N}^*}$ une suite i.i.d. de variables aléatoires réelles dans $L^2(\Omega, \mathcal{A}, P)$; notons m leur espérance commune et σ^2 leur variance commune et posons, pour tout $n \in \mathbb{N}^*$, $S_n = X_1 + \dots + X_n$. Enfin, soit Z une variable aléatoire réelle de loi $\mathcal{N}(0, 1)$. Alors, pour tout intervalle réel I , nous avons la convergence suivante :*

$$P\left(\frac{S_n - nm}{\sigma\sqrt{n}} \in I\right) \xrightarrow[n \rightarrow +\infty]{} P(Z \in I). \quad (18.1)$$

On dit que la suite de variables aléatoires réelles $((S_n - nm)/(\sigma\sqrt{n}))_{n \in \mathbb{N}^*}$ converge en loi vers Z (ou encore vers la loi normale $\mathcal{N}(0, 1)$) et l'on note :

$$\frac{S_n - nm}{\sigma\sqrt{n}} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1)$$

Remarques.

1. La variable aléatoire réelle qui apparaît dans le membre de gauche est la variable centrée réduite associée à S_n par transformation affine. On vérifie en effet immédiatement que $E[S_n] = nm$ et $\text{Var}S_n = n\sigma^2$. Pour alléger les écritures, nous utiliserons la notation :

$$\tilde{S}_n := \frac{S_n - nm}{\sigma\sqrt{n}}$$

2. Si nous notons \bar{X}_n la moyenne empirique d'ordre n associée à la suite $(X_i)_{i \in \mathbb{N}^*}$, alors le théorème-limite central s'écrit encore, pour tout intervalle réel I ,

$$P\left(\frac{\sqrt{n}}{\sigma}(\bar{X}_n - m) \in I\right) \xrightarrow[n \rightarrow +\infty]{} P(Z \in I).$$

Ainsi, si $I = [a, b]$, nous obtenons en utilisant la densité de la loi normale centrée réduite :

$$P\left(a \leq \frac{\sqrt{n}}{\sigma}(\bar{X}_n - m) \leq b\right) \xrightarrow[n \rightarrow +\infty]{} \int_a^b \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt.$$

Si nous prenons $-a = b = 1,96$, nous obtenons (d'après la fin du paragraphe précédent) que lorsque n est «grand», il y a approximativement 95 chances sur 100 pour que l'on ait : $|\bar{X}_n - m| \leq 1,96\sigma/\sqrt{n}$.

C'est en ce sens que le théorème-limite central implique que l'erreur commise en faisant une approximation de m par \bar{X}_n , comme nous y invite la loi des grands nombres, est d'ordre de grandeur σ/\sqrt{n} pour n «grand». On dit de façon équivalente que la convergence énoncée par la loi des grands nombres a lieu avec une vitesse d'ordre de grandeur \sqrt{n}/σ .

3. La convergence énoncée par la loi des grands nombres était valable sous l'hypothèse $(X_i)_{i \in \mathbb{N}^*}$ i.i.d. dans $L^1(\Omega, \mathcal{A}, P)$ mais pour préciser sa vitesse grâce au théorème-limite central, nous avons besoin d'une hypothèse supplémentaire : la suite $(X_i)_{i \in \mathbb{N}^*}$ est supposée dans $L^2(\Omega, \mathcal{A}, P)$.
4. En prenant $I = [a, b]$ (pour fixer les idées) et en faisant un simple changement de variable dans une intégrale, on montre que la suite $((S_n - nm)/\sqrt{n})_{n \in \mathbb{N}^*}$ converge en loi vers la loi normale $\mathcal{N}(0, \sigma^2)$.
5. La convergence en loi est une notion plus difficile à manipuler que la convergence en probabilité ou la convergence presque sûre. Elle est la plus faible de ces trois types de convergences : on peut montrer que, pour des variables aléatoires réelles quelconques $(Y_n)_{n \in \mathbb{N}^*}$ et Y , on a les implications

$$Y_n \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} Y \implies Y_n \xrightarrow[n \rightarrow +\infty]{(P)} Y \implies Y_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} Y.$$

Malgré la difficulté de cette notion, on est obligé d'en passer par la convergence en loi tout simplement parce que les types plus forts de convergence ne permettent pas d'obtenir un théorème du genre théorème-limite central : sous les hypothèses de ce théorème, on peut en effet prouver qu'il n'existe aucun $\alpha > 0$ tel que la suite $(n^\alpha(\bar{X}_n - m))_{n \in \mathbb{N}^*}$ converge en probabilité vers une variable aléatoire non presque sûrement nulle.

18.2 TLC et fonctions de répartition

Corollaire 18.2.1 *Sous les hypothèses du théorème-limite central et en notant Φ la fonction de répartition de la loi $\mathcal{N}(0, 1)$, nous avons la convergence suivante pour tout $x \in \mathbb{R}$:*

$$P \left(\frac{S_n - nm}{\sigma\sqrt{n}} \leq x \right) \xrightarrow[n \rightarrow +\infty]{} \Phi(x). \quad (18.2)$$

Démonstration: Il suffit d'appliquer le théorème 18.1.1 avec $I =] - \infty, x]$.
□

Remarques.

1. On peut reformuler cet énoncé en disant que, si l'on note F_n la fonction de répartition de la variable aléatoire réelle \tilde{S}_n , alors la suite $(F_n)_{n \in \mathbb{N}^*}$ converge simplement sur \mathbb{R} vers la fonction de répartition Φ de la loi normale centrée réduite.
2. Le corollaire reste vrai si l'on remplace l'inégalité large par une inégalité stricte. Il suffit en effet de prendre $I =] - \infty, x[$ dans le théorème précédent. En nous souvenant des propriétés des fonctions de répartition, nous pouvons écrire ceci sous la forme :

$$\forall x \in \mathbb{R} \quad F_n(x-) \xrightarrow[n \rightarrow +\infty]{} \Phi(x) \quad (18.3)$$

3. On peut montrer que cet énoncé 18.2.1 n'est pas seulement un corollaire du théorème 18.1.1 mais lui est en fait équivalent. C'est pourquoi le lecteur trouvera le théorème-central limite énoncé sous cette forme (convergence simple de fonctions de répartition) dans plusieurs ouvrages. De même, la convergence (18.3) est en fait équivalente à l'énoncé du théorème-limite central.

Les convergences (18.2) et (18.3), qui sont donc équivalentes à l'énoncé du théorème-limite central, peuvent être améliorées. En effet, un critère de convergence uniforme appelé *second théorème de Dini*, permet de démontrer le résultat suivant, que nous admettrons.

Lemme 18.2.2 *Soit $(F_n)_{n \in \mathbb{N}^*}$ une suite de fonctions de répartition qui converge simplement sur \mathbb{R} vers une fonction de répartition continue F . Alors $(F_n)_{n \in \mathbb{N}^*}$ converge uniformément vers F sur \mathbb{R} .*

La fonction de répartition Φ de la loi $\mathcal{N}(0, 1)$ étant continue (c'est d'ailleurs vrai pour toute fonction de répartition d'une loi qui admet une densité de probabilité), nous déduisons immédiatement de ce lemme que les convergences (18.2) et (18.3) sont en réalité uniformes sur \mathbb{R} . Par exemple, nous avons, en notant F_n la fonction de répartition de la variable aléatoire réelle \tilde{S}_n :

$$\sup_{x \in \mathbb{R}} |F_n(x) - \Phi(x)| \xrightarrow[n \rightarrow +\infty]{} 0 \quad (18.4)$$

Il est alors facile d'en déduire la proposition suivante, que le lecteur pourra prouver à titre d'exercice, en utilisant des égalités telles que :

$$P\left(\tilde{S}_n \in [a, b]\right) = F_n(b) - F_n(a-), \quad P\left(\tilde{S}_n \in]a, b[\right) = F_n(b-) - F_n(a)$$

Proposition 18.2.3 *Sous les hypothèses et avec les notations du théorème-limite central, la convergence (18.1) est uniforme sur tous les intervalles réels I , ce qui s'écrit :*

$$\sup_I \left| P \left(\frac{S_n - nm}{\sigma\sqrt{n}} \in I \right) - P(Z \in I) \right| \xrightarrow{n \rightarrow +\infty} 0 \quad (18.5)$$

Lorsque n tend vers l'infini, la loi de la variable \tilde{S}_n et la loi de la variable Z , c'est-à-dire la loi $\mathcal{N}(0, 1)$, deviennent arbitrairement proches uniformément sur tous les intervalles réels I . C'est en ce sens que l'on peut dire que, lorsque n est suffisamment grand, la loi de la variable \tilde{S}_n est approximativement gaussienne centrée réduite, ce qu'on écrira (informellement) $\tilde{S}_n \approx \mathcal{N}(0, 1)$. Revenons maintenant au fait que les convergences (18.2) et (18.3) ont lieu uniformément sur \mathbb{R} . Si nous notons F_{S_n} la fonction de répartition de la variable S_n et $\Phi_{nm, n\sigma^2}$ la fonction de répartition de la loi $\mathcal{N}(nm, n\sigma^2)$, on démontre facilement que, pour tout $n \in \mathbb{N}^*$ et tout $x \in \mathbb{R}$:

$$F_{S_n}(x) = F_n \left(\frac{x - nm}{\sigma\sqrt{n}} \right) \text{ et } \Phi_{nm, n\sigma^2}(x) = \Phi \left(\frac{x - nm}{\sigma\sqrt{n}} \right).$$

Nous en déduisons les convergences uniformes suivantes :

$$\sup_{x \in \mathbb{R}} |F_{S_n}(x) - \Phi_{nm, n\sigma^2}(x)| \xrightarrow{n \rightarrow +\infty} 0, \quad \sup_{x \in \mathbb{R}} |F_{S_n}(x-) - \Phi_{nm, n\sigma^2}(x)| \xrightarrow{n \rightarrow +\infty} 0.$$

Par un raisonnement similaire à ce qui a été fait précédemment, nous pouvons en déduire que, lorsque n tend vers l'infini, la loi de la variable S_n et la loi $\mathcal{N}(nm, n\sigma^2)$ deviennent arbitrairement proches uniformément sur tous les intervalles réels I . C'est en ce sens que l'on peut dire que, lorsque n est suffisamment grand, la loi de la variable S_n vaut approximativement la loi $\mathcal{N}(nm, n\sigma^2)$, ce que nous écrivons : $S_n \approx \mathcal{N}(nm, n\sigma^2)$.

Il nous reste à préciser ce que signifie n «grand» ! Comme nous l'avons fait dans le chapitre précédent à propos du théorème de De Moivre-Laplace, nous utilisons le *théorème de Berry-Esseen*, ici dans son énoncé général, lequel nous donne des précisions sur la vitesse à laquelle a lieu la convergence énoncée par le théorème-limite central.

Le lecteur pourra retrouver cet énoncé dans “L'essentiel en théorie des probabilités” par Jean JACOD et Philip PROTTER chez Cassini, page 191 (corriger le membre de droite en recentrant la variable X_1). On ne connaît pas la valeur optimale de la constante C intervenant ci-dessous mais on sait que l'on peut prendre $C \leq 0,8$.

Théorème 18.2.4 (Berry-Esseen) Soit $(X_i)_{i \in \mathbb{N}^*}$ une suite indépendante identiquement distribuée de variables aléatoires réelles dans $L^3(\Omega, \mathcal{A}, P)$. En reprenant les notations de (18.4), il existe une constante $C > 0$ telle que :

$$\forall n \in \mathbb{N}^* \quad \sup_{x \in \mathbb{R}} |F_n(x) - \Phi(x)| \leq C \frac{E[|X_1 - E[X_1]|^3]}{\sigma^3 \sqrt{n}}$$

Pour savoir à quelle vitesse avait lieu la convergence dans la loi des grands nombres, nous avons dû rajouter une condition L^2 et la réponse nous a alors été fournie par le théorème-limite central. Pour savoir à quelle vitesse avait lieu la convergence énoncée par ce dernier, nous avons dû rajouter une condition L^3 et la réponse nous a alors été fournie par le théorème de Berry-Esseen. Remarquons que les questions de vitesse de convergence s'arrêtent là car le théorème de Berry-Esseen n'est pas un théorème asymptotique : la majoration qu'il énonce est valable pour tout entier $n \in \mathbb{N}^*$, même petit.

18.3 Un énoncé sans équidistribution

Nous avons annoncé dans l'introduction que, selon le théorème-limite central, si l'on additionne beaucoup de variables indépendantes admettant un moment d'ordre deux, alors le résultat est une variable approximativement gaussienne. C'est ce que nous venons de montrer mais à une objection près : jusqu'à présent, nous n'avons pas seulement supposé les variables indépendantes et dans $L^2(\Omega, \mathcal{A}, P)$ mais nous leur avons également demandé d'avoir toutes la même loi. Nous dirons simplement à ce sujet qu'il existe divers raffinements du théorème-limite central qui permettent, sous certaines conditions, de s'affranchir de l'hypothèse d'équidistribution. Nous donnons un tel énoncé à titre d'exemple, en l'admettant.

Proposition 18.3.1 (TLC, suite bornée dans $L^{2+\epsilon}$) Soit $\epsilon > 0$ et $(X_i)_{i \in \mathbb{N}^*}$ une suite de variables aléatoires réelles indépendantes mais non nécessairement de même loi, formant une famille bornée dans $L^{2+\epsilon}(\Omega, \mathcal{A}, P)$, i.e. telle que :

$$\sup_{i \in \mathbb{N}^*} E[|X_i|^{2+\epsilon}] < +\infty.$$

Par commodité, nous supposons $E[X_i] = 0$ et nous notons $\sigma_i^2 = \text{Var} X_i$ pour tout $i \in \mathbb{N}^*$.

Alors, si $\sum_{i=1}^{+\infty} \sigma_i^2 = +\infty$, nous avons la convergence en loi suivante :

$$\frac{S_n}{\sqrt{\sum_{i=1}^n \sigma_i^2}} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

18.4 Test du χ^2 d'ajustement.

La construction de ce test statistique d'usage très courant fait appel à une version multidimensionnelle du théorème-limite central qui est hors programme. Nous admettrons donc certains résultats dans la suite, en présentant les idées essentielles permettant la mise en oeuvre de ce test.

Nous considérons n répétitions d'une expérience aléatoire qui produit un résultat dans un ensemble fini, par exemple $\llbracket 1; k \rrbracket$.

Nous avons des raisons de supposer que la loi sur $\llbracket 1; k \rrbracket$ qui gouverne cette expérience est donnée par $p = (p^1, \dots, p^k)$, avec $p^i \geq 0$ pour tout $i \in \llbracket 1; k \rrbracket$ et $p^1 + \dots + p^k = 1$, mais nous voudrions confirmer ou infirmer cette hypothèse au regard des valeurs $(x_1, \dots, x_n) = (X_1(\omega), \dots, X_n(\omega))$ observées au cours des n répétitions de l'expérience, c'est-à-dire tester l'*ajustement* de notre modèle à la réalité expérimentale.

Pour ce faire, une première étape consiste à introduire une sorte de distance entre les lois de probabilité sur l'ensemble $\llbracket 1; k \rrbracket$, l'idée étant de regarder ensuite si la loi empirique (définie ci-dessous et calculable à partir du résultat de l'expérience) est proche ou éloignée de la loi théorique p dont nous avons fait l'hypothèse.

Définition 18.4.1 Soient p et q deux lois de probabilité sur $\llbracket 1; k \rrbracket$.

On appelle «distance» du χ^2 entre p et q la quantité :

$$d_{\chi^2}(p, q) = \sum_{i=1}^k \frac{(p^i - q^i)^2}{p^i}$$

Le mot distance est entre guillemets car ce n'est pas du tout une distance au sens des espaces métriques : elle n'est visiblement pas symétrique et il est facile de constater qu'elle ne vérifie pas non plus l'inégalité triangulaire. En réalité, son seul rapport avec une distance est la propriété suivante :

$$d_{\chi^2}(p, q) = 0 \Leftrightarrow p = q$$

Remarquons que cette pseudo-distance a tendance à surévaluer les différences entre p et q sur les entiers i où p_i est petit : nous chercherons à limiter ce phénomène dans la suite en imposant des conditions telles que $np_i \geq 5$ pour tout $i \in \llbracket 1; k \rrbracket$.

La deuxième étape de construction du test qui va nous permettre de confirmer ou infirmer notre hypothèse consiste à comparer la loi théorique p avec la loi empirique \bar{p}_n que nous définissons comme suit :

Définition 18.4.2 Si (X_1, \dots, X_n) est la variable aléatoire modélisant les n répétitions de notre expérience, nous posons :

$$\forall i \in \llbracket 1; k \rrbracket, \quad N_n^i = \sum_{j=1}^n \mathbf{1}_{\{X_j=i\}}$$

Nous appelons alors loi empirique, notée \bar{p}_n , la loi sur $\llbracket 1; k \rrbracket$ définie par :

$$\forall i \in \llbracket 1; k \rrbracket, \quad \bar{p}_n^i = \frac{N_n^i}{n}$$

Notons que la valeur de \bar{p}_n dépend du résultat ω de l'expérience, d'où le qualificatif empirique. En toute rigueur, c'est d'ailleurs $\bar{p}_n(\omega)$ (et non pas \bar{p}_n) qui est une loi de probabilité sur $\llbracket 1; k \rrbracket$.

Définition 18.4.3 On appelle χ^2 d'ajustement la variable aléatoire suivante :

$$nd_{\chi^2}(p, \bar{p}_n) = n \sum_{i=1}^k \frac{(p^i - \bar{p}_n^i)^2}{p^i} = \sum_{i=1}^k \frac{(np^i - N_n^i)^2}{np^i}$$

Nous rappelons maintenant la définition d'une loi de probabilité sur \mathbb{R} qui a été introduite dans l'Application 15.5.5, page 276, et identifiée à une loi Gamma, ce qui nous fournit sa densité.

Définition 18.4.4 Considérons $Z = (Z_1, \dots, Z_d)$ un vecteur aléatoire dont les composantes sont indépendantes et de même loi $\mathcal{N}(0, 1)$.

Alors la loi de la variable aléatoire positive $\|Z\|^2 = Z_1^2 + \dots + Z_d^2$ est appelée loi du χ^2 à d degrés de liberté et notée $\chi^2(\mathbf{d})$.

Cette loi est en fait égale à la loi $\gamma(\frac{1}{2}, \frac{d}{2})$ et admet donc pour densité :

$$g_{\frac{1}{2}, \frac{d}{2}}(x) = \frac{1}{2^{\frac{d}{2}} \Gamma(\frac{d}{2})} e^{-\frac{x}{2}} x^{\frac{d}{2}-1} \mathbf{1}_{\mathbb{R}_+^*}(x)$$

Le résultat essentiel qui va nous permettre de construire le test dit du χ^2 est le suivant :

Proposition 18.4.5 (Pearson) Si pour tout $n \in \mathbb{N}^*$, le vecteur aléatoire (X_1, \dots, X_n) suit la loi $p^{\otimes n}$, alors la convergence suivante a lieu :

$$nd_{\chi^2}(p, \bar{p}_n) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \chi^2(\mathbf{k} - \mathbf{1})$$

Autrement dit, si nous notons F_{k-1} la fonction de répartition de la loi $\chi^2(\mathbf{k} - \mathbf{1})$, nous avons la convergence suivante pour tout $x \in \mathbb{R}$:

$$P(nd_{\chi^2}(p, \bar{p}_n) \leq x) \xrightarrow[n \rightarrow +\infty]{} F_{k-1}(x)$$

Comme annoncé plus haut, la démonstration de ce résultat nécessite la version multidimensionnelle du théorème-limite central, hors programme, donc nous l'admettons. Remarquons simplement que si (X_1, \dots, X_n) suit la loi $p^{\otimes n}$, c'est-à-dire si les variables X_i sont indépendantes et de même loi p , alors le TLC (en dimension 1) implique la convergence suivante :

$$\frac{N_n^i - np^i}{\sqrt{np^i}} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1 - p_i)$$

Il est alors assez intuitif que la loi limite du χ^2 d'ajustement, c'est-à-dire de la variable suivante :

$$nd_{\chi^2}(p, \bar{p}_n) = \sum_{i=1}^k \left(\frac{N_n^i - np^i}{\sqrt{np^i}} \right)^2$$

soit une loi du χ^2 mais on pourrait penser que celle-ci a k degrés de liberté, alors que la vraie loi limite est $\chi^2(\mathbf{k} - \mathbf{1})$. En fait, la «perte» d'un degré de liberté peut se comprendre en constatant que nos variables ne sont pas totalement «libres» puisqu'il existe entre elles la relation linéaire suivante :

$$\sum_{i=1}^k (N_n^i - np^i) = 0$$

La proposition de Pearson nous dit donc que, si l'expérience est bien gouvernée par la loi p supposée, alors le χ^2 d'ajustement suit une loi proche de $\chi^2(\mathbf{k} - \mathbf{1})$ lorsque le nombre n de répétitions de l'expérience devient «grand».

En revanche, si l'expérience est en réalité régie par une loi $q \neq p$, alors il existe $1 \leq i \leq k$ tel que $q^i \neq p^i$ et la loi des grands nombres implique la convergence suivante lorsque n tend vers l'infini :

$$d_{\chi^2}(p, \bar{p}_n) = \sum_{i=1}^k \frac{(p^i - \bar{p}_n^i)^2}{p^i} \xrightarrow{\text{p.s.}} d_{\chi^2}(p, q) > 0$$

d'où nous déduisons le comportement asymptotique du χ^2 d'ajustement :

$$nd_{\chi^2}(p, \bar{p}_n) \xrightarrow{\text{p.s.}} +\infty \quad (18.6)$$

C'est la différence entre ces deux comportements asymptotiques qui va nous permettre de tester l'hypothèse H_0 : «L'expérience est gouvernée par la loi p » contre l'hypothèse alternative H_1 : «L'expérience est gouvernée par une loi $q \neq p$ ».

Passons à la construction effective du test.

Si nous notons F_{k-1} la fonction de répartition de la loi $\chi^2(\mathbf{k} - \mathbf{1})$, nous prouvons facilement que F_{k-1} est une bijection de \mathbb{R}_+ sur $[0, 1[$ en constatant que la densité de la loi $\chi^2(\mathbf{k} - \mathbf{1})$ est strictement positive sur \mathbb{R}_+^* et nulle sur \mathbb{R}_- . Par conséquent, pour tout $\alpha \in]0, 1]$, il existe un unique $c_\alpha \in \mathbb{R}_+$ tel que $\chi^2(\mathbf{k} - \mathbf{1})(]c_\alpha, +\infty[) = \alpha$ et l'on a $c_\alpha = F_{k-1}^{-1}(1 - \alpha)$.

Si l'hypothèse H_0 est réalisée, donc si l'expérience est régie par la loi p , la proposition de Pearson implique :

$$P_p(nd_{\chi^2}(p, \bar{p}_n) > c_\alpha) \xrightarrow{n \rightarrow +\infty} \alpha \quad (18.7)$$

En revanche, si c'est l'hypothèse H_1 qui est réalisée, donc si l'expérience est gouvernée par une loi $q \neq p$, alors le comportement asymptotique (18.6) entraîne la convergence suivante lorsque n tend vers l'infini :

$$P_q(nd_{\chi^2}(p, \bar{p}_n) > c_\alpha) \xrightarrow{n \rightarrow +\infty} 1 \quad (18.8)$$

Nous pratiquons donc notre test comme suit :

Nous choisissons une valeur $\alpha \in]0, 1]$ (typiquement α est petit car, comme nous allons le voir, il représente le niveau d'erreur du test) et nous en déduisons la valeur c_α . Pour le résultat ω de l'expérience que nous observons, nous calculons la valeur du χ^2 d'ajustement :

$$nd_{\chi^2}(p, \bar{p}_n(\omega)) = \sum_{i=1}^k \frac{(np^i - N_n^i(\omega))^2}{np^i}$$

Nous comparons alors cette valeur à c_α pour conclure :

- Si $nd_{\chi^2}(p, \bar{p}_n(\omega)) > c_\alpha$, alors nous rejetons l'hypothèse H_0 .
- Si $nd_{\chi^2}(p, \bar{p}_n(\omega)) \leq c_\alpha$, alors nous acceptons l'hypothèse H_0 .

De façon générale, lorsque nous pratiquons un test statistique, notre conclusion peut être erronée de deux façons différentes :

Erreur de 1ère espèce : Je rejette l'hypothèse H_0 alors qu'elle est satisfaite en réalité.

Erreur de 2nde espèce : J'accepte l'hypothèse H_0 alors qu'elle n'est pas satisfaite en réalité.

Dans de nombreuses situations pratiques, ces deux types d'erreurs ne sont pas symétriques, l'erreur de première espèce étant considérée comme plus grave que l'autre :

Par exemple, si je teste le câble d'un ascenseur supposé pouvoir accueillir 10 personnes (750kg) et si je note M la masse critique à partir de laquelle le câble casse, je choisirai H_0 : « $M \leq 750$ » et H_1 : « $M > 750$ ».

L'erreur de 1ère espèce conduirait des usagers de l'ascenseur à un grand plongeon : c'est ce risque que je veux absolument maîtriser. L'erreur de 2nde espèce conduirait à des réparations inutiles sur l'ascenseur : je veux l'éviter mais elle est moins grave que la première.

Dans le test du χ^2 que nous venons de construire, la convergence (18.7) se traduit comme suit : la probabilité de commettre une erreur de 1ère espèce est asymptotiquement égale à α . On dit qu'on a construit un test de niveau d'erreur asymptotique α ou de *niveau de confiance asymptotique* $1 - \alpha$.

Quant à la convergence (18.8), sa traduction est plus vague : lorsque n devient «grand», la probabilité de commettre une erreur de 2nde espèce devient «petite» mais nous ne maîtrisons pas la vitesse à laquelle cette convergence se produit. On dit que la *puissance* du test, c'est-à-dire la probabilité de rejeter l'hypothèse H_0 quand elle n'est effectivement pas satisfaite dans la réalité, tend vers 1 lorsque n tend vers l'infini.

Nous terminons cette introduction au test du χ^2 en le généralisant au cas où l'ensemble E des résultats possibles pour l'expérience est infini. Nous pouvons alors adapter notre méthode comme suit :

Nous choisissons une partition finie (E_1, \dots, E_k) de l'ensemble E . Si ν est la loi sur E supposée gouverner l'expérience, nous posons $p^i = \nu(E_i)$ et nous comptons maintenant le nombre de fois où l'on tombe dans la classe E_i au cours des n répétitions de l'expérience :

$$N_n^i(\omega) = \sum_{j=1}^n \mathbf{1}_{\{X_j(\omega) \in E_i\}}$$

Le reste du test se déroule comme précédemment.

Notons qu'avec cette méthode, nous ne testons en réalité que l'adéquation aux valeurs de ν sur les différentes classes E_i .

Attention, le choix des classes E_i n'est pas du tout innocent ! Une règle d'or, liée à la remarque faite à la suite de la définition de d_{χ^2} , est que les *effectifs théoriques* np_i (espérance de N^i dans le cas où l'expérience est vraiment gouvernée par la loi ν) des classes E_i ne doivent jamais être trop «petits»

Dans la pratique, on impose souvent que la condition suivante soit satisfaite :

$$\boxed{\forall i = \llbracket 1; k \rrbracket \quad np^i \geq 5}$$

D'autres règles pratiques existent : par exemple, on a constaté qu'il était préférable de choisir les classes E_i de sorte que leurs effectifs théoriques respectifs np_i soient à peu près tous équivalents. Nous n'insisterons pas plus sur ce sujet des règles pratiques qui concernent plus les orfèvres du test du χ^2 que les agrégatifs !

ISBN 978-2-91-635238-1



9 782916 352381